



Università di Genova

SCUOLA DI SCIENZE MATEMATICHE, FISICHE E
NATURALI

Corso di Laurea Magistrale in Matematica

Verso una definizione di correlazione
topologica: misurare la differenza tra
omologia persistente monoparametrica e
multiparametrica

Relatore

Dott. Ulderico Fugacci

Correlatore

Prof. Matteo Penegini

Candidata

Isabella Mastroianni

Anno Accademico 2023/24
Genova, 25 settembre 2024

Sommario

La mia Tesi di Laurea Magistrale costituisce un primo passo nel tentativo di comprendere — sia da un punto di vista intuitivo che teorico — il motivo per cui la persistenza multiparametrica presenti una maggior capacità discriminante rispetto alla persistenza monoparametrica, e come tale differenza permetta di carpire informazioni sulla dipendenza reciproca tra le funzioni che originano la persistenza. Lo scopo ultimo è proprio quello di convogliare le informazioni ricavate in modo da definire un concetto di “correlazione topologica” tra due funzioni. L’efficacia di questa idea è supportata dal fatto che, in svariate situazioni di studio, si osserva che maggiore è il “legame” tra due funzioni filtranti considerate e minore è l’informazione aggiuntiva che la persistenza multiparametrica fornisce rispetto a quella monoparametrica. Più in dettaglio, nella mia Tesi individuo due problemi: da un lato metto in luce l’evidenza di una differenza tra la capacità discriminante dell’omologia persistente monoparametrica e di quella multiparametrica, di cui non esistono ancora spiegazioni teoriche in letteratura; dall’altro formalizzo l’esigenza di definire un concetto di correlazione tra funzioni che sia il più generale possibile. Come soluzioni a tali problemi propongo due nuove definizioni, tra loro connesse: la differenza topologica e la correlazione topologica. Le dimostrazioni di alcune loro proprietà elementari e il loro utilizzo in svariate esempi consentono di validare i concetti introdotti e di discuterne l’uso in contesti applicativi, delinendo anche possibili sviluppi e scenari futuri.

Contributi originali della mia Tesi di Laurea Magistrale:

- Studio approfondito dello Stato dell’Arte dell’omologia persistente — sia monoparametrica che multiparametrica — e riorganizzazione degli argomenti coinvolti utilizzando un linguaggio comune ad entrambe le teorie.
- Studio delle distanze tra moduli di persistenza monoparametrica e biparametrica.
- Sviluppo di esempi che manifestino la differenza di potere discriminante tra multipersistenza e monopersistenza.

- Definizione del concetto di differenza topologica $\Delta(\Phi_1, \Phi_2)$ tra due funzioni $\Phi_1, \Phi_2 : X \rightarrow \mathbb{R}^2$ e di correlazione $\Delta(\Phi)$ tra le componenti di una funzione $\Phi : X \rightarrow \mathbb{R}^2$.
- Definizione e prime proprietà del concetto di correlazione topologica $corr(f, g)$ per due classi di funzioni $f = \{f_i\}_{i=1, \dots, N}$ e $g = \{g_i\}_{i=1, \dots, N}$, con $f_i, g_i : X_i \rightarrow \mathbb{R}$.
- Analisi delle prime proprietà delle nozioni introdotte.
- Sviluppo e studio di vari esempi a supporto delle nozioni introdotte.
- Delineazione di possibili applicazioni e possibili sviluppi futuri, di carattere topologico, geometrico, algebrico e implementativo-computazionale.

Abstract

My Master’s Thesis constitutes a first step in the attempt to understand — both from an intuitive and theoretical point of view — why multi-parameter persistence has a greater discriminative capacity than single-parameter persistence, and how this difference allows one to extract information on the mutual dependence between the functions that originate persistence. The ultimate aim is precisely to channel the information obtained in such a way as to define a concept of “topological correlation” between two functions. The efficacy of this idea is supported by the fact that, in a variety of study situations, it is observed that the greater the “link” between two filter functions considered, the less additional information a multi-parameter persistence provides with respect to a single-parameter one. More in detail, in my Thesis I identify two problems: on the one hand I highlight the evidence of a difference between the discriminating capacity of monoparametric and multi-parametric persistent homology, for which no theoretical explanation exists yet in the literature; on the other hand I formalise the need to define a concept of correlation between functions that is as general as possible. As solutions to these problems I propose two new, interrelated definitions: topological difference and topological correlation. Demonstrations of some of their elementary properties and their use in various examples allow the validation of the introduced concepts and to discuss their use in application contexts, also outlining possible future developments and scenarios.

Original contributions of my Master’s Thesis:

- In-depth study of the State of the Art of persistent homology — both monoparameter and multiparameter — and reorganization of the topics involved using a unified language for both theories.
- Study of the distances between monoparameter and biparameter persistence modules.
- Development of examples that highlight the difference in discriminative power between multipersistence and monopersistence.
- Definition of the concept of topological difference $\Delta(\Phi_1, \Phi_2)$ between two functions $\Phi_1, \Phi_2 : X \rightarrow \mathbb{R}^2$, and of correlation $\Delta(\Phi)$ between the components of a function $\Phi : X \rightarrow \mathbb{R}^2$.
- Definition of the concept of topological correlation $corr(f, g)$ for two classes of functions $f = \{f_i\}_{i=1, \dots, N}$ and $g = \{g_i\}_{i=1, \dots, N}$, with $f_i, g_i : X_i \rightarrow \mathbb{R}$.
- Analysis of the initial properties of the introduced concepts.

- Development and study of various examples supporting the introduced concepts.
- Outlining of possible applications and possible future developments of a topological, geometric, algebraic, and computational-implementation nature.

Introduzione

Durante il mio percorso di Laurea - triennale e magistrale - in Matematica, sono rimasta affascinata fin da subito dalla topologia e dalla geometria. In particolare, quando mi sono imbattuta nel concetto di omologia, ho avuto chiaro il motivo per cui apprezzassi tali ambiti matematici: in queste discipline è particolarmente evidente ciò che caratterizza tutta la matematica, cioè l'estrema libertà di immaginazione garantita da un rigoroso formalismo. Più in dettaglio la topologia e la geometria consentono di studiare con estrema precisione ed efficacia una moltitudine di oggetti di varia natura, permettendo di classificarli. Per questi motivi, la topologia, e più precisamente la **topologia algebrica**, si sta rivelando efficace nell'ambito dell'Analisi di Dati.

In un mondo in cui la raccolta e l'analisi di grandi moli di dati stanno assumendo un ruolo sempre più determinante, la **Topological Data Analysis**, abbreviata in TDA, sta guadagnando un ruolo di sempre maggior rilievo. L'idea vincente che dà fondamento alla TDA è la seguente. I dati di interesse — anche se astratti e non dotati, originariamente, di una topologia — vengono ulteriormente arricchiti dotandoli di una **struttura topologica**; *gli invarianti topologici e geometrici dello spazio così ottenuto corrispondono a caratteristiche intrinseche, dette features, del dataset*, che sono rilevanti per la comprensione dello stesso.

Tale approccio è utilizzato nell'analisi di dati di praticamente ogni tipologia, molto spesso nuvole di punti, ma anche grafi, funzioni, e così via.

L'**omologia persistente** rappresenta lo strumento più valido ed efficace per indagare le strutture topologiche provenienti da dataset. Essa infatti *permette di individuare le caratteristiche del dato che persistono nel tempo o, più in generale, nei diversi livelli di analisi considerati*. Più in dettaglio, l'omologia persistente analizza i cambiamenti omologici di un dataset rispetto a uno o più parametri reali; nel primo caso, si parla di persistenza *monoparametrica* mentre nel secondo di persistenza *multiparametrica*. Grazie alle sue salde radici teoriche, la TDA e, in particolar modo, l'omologia persistente hanno permesso di ottenere notevoli risultati in ambito applicativo.

Data la sua recente introduzione, l'omologia persistente multiparametrica presenta ancora alcune interessanti sfide sul piano teorico. Inoltre, nonostante di fronte a dataset dipendenti da più parametri l'applicazione della multipersistenza garantisca risultati migliori della monopersistenza, essa non ha ancora avuto una larga diffusione in ambito applicativo. Se inizialmente gli aspetti computazionali costituivano il principale ostacolo, sono ormai stati sviluppati — sebbene recentemente — strumenti che permettono di ridurre la complessità. Ciò che ancora rappresenta un freno al suo utilizzo è che rimane ancora da comprendere, sia da un punto di vista intuitivo che teorico, *quale sia il significato delle maggiori informazioni ritrovate grazie all'approccio multiparametrico rispetto alle sole informazioni ottenute con quello monoparametrico*. Tra le varie sue capacità, l'omologia persistente permette di confrontare efficientemente ed efficacemente due funzioni. Date ad esempio due funzioni $\Phi_1 := (f_1, g_1), \Phi_2 := (f_2, g_2) : X \rightarrow \mathbb{R}^2$, definite su uno spazio topologico X , l'omologia persistente permette di confrontarle in due modi.

Approccio monoparametrico: si studia l'omologia persistente della filtrazione indotta dalle singole componenti f_1, f_2, g_1, g_2 — che sono funzioni a valori in \mathbb{R} e, pertanto, daranno origine a gruppi di omologia persistente monoparametrica;

Approccio multiparametrico: si studia l'omologia persistente della filtrazione indotta dalle stesse Φ_1, Φ_2 — che sono funzioni a valori in \mathbb{R}^2 e, pertanto, daranno origine a gruppi di omologia persistente multiparametrica.

La teoria multiparametrica garantisce una maggior capacità discriminante rispetto a quella monoparametrica, in quanto risulta molto facile costruire esempi in cui Φ_1 e Φ_2 sono indistinguibili per il primo approccio ma non per il secondo, come avremo modo di vedere nella Tesi. In particolare, *cerchiamo di trovare una motivazione teorica che giustifichi tale gap informativo e una sua formalizzazione*.

Sorprendentemente, la risposta a questa prima domanda permetterà di rispondere ad una seconda, dall'importante valore sia applicativo che teorico: *Cosa si intende per correlazione/dipendenza di funzioni? È possibile definire formalmente un tale concetto di correlazione funzionale?* Tali domande nascono dall'esigenza comune a svariati ambiti applicativi di ridurre il numero di funzioni considerate durante l'analisi di dati (ad esempio, lo studio della compatibilità tra molecole confrontandone funzioni fisico-chimiche come la carica elettrostatica e l'idrofobicità).

La mia Tesi di Laurea è strutturata come segue. Nel Capitolo 1 verranno introdotti i *concetti fondamentali* su cui si basa tut-

to il resto del lavoro, ovvero l'**omologia singolare** e i **moduli graduati e multi-graduati**.

Nel Capitolo 2 e nel Capitolo 3, verrà presentato lo *Stato dell'Arte della teoria della persistenza*. In particolare: introdurremo le definizioni e le proprietà fondamentali di **moduli di persistenza e omologia persistente**; analizzeremo in dettaglio i *casi monoparametrico e biparametrico*, fornendo una panoramica generale dei loro *invarianti e descrittori* e soffermandoci su quelli che assumeranno un ruolo centrale nel resto dell'elaborato.

Nel Capitolo 4 *confronteremo la persistenza monoparametrica e multiparametrica*, mettendo in luce il fatto che, da un lato, la teoria monoparametrica risulta più ricca di invarianti e descrittori completi, mentre dall'altro la teoria multiparametrica risulta più informativa, presentando una maggior efficacia discriminante. Tuttavia, attualmente non esiste un modo formale per quantificare e misurare il **gap informativo** tra i due casi. Evidenzieremo tali aspetti tramite esempi in cui confronteremo moduli di persistenza monoparametrici e biparametrici.

Nel Capitolo 5 indagheremo a che punto si trovi la letteratura attuale, in svariati ambiti matematici, nella definizione di un concetto di **correlazione tra funzioni**; procederemo a partire dalla *definizione elementare di correlazione* per poi addentrarci in nozioni più complesse, come quello di *cross-correlation tra segnali*. Tuttavia, vedremo che al momento non esiste una definizione di correlazione tra funzioni generiche su spazi topologici. Questo capitolo, almeno per ora, può sembrare una deviazione immotivata dal percorso delineato nei capitoli precedenti, vedremo invece che esso costituirà la chiave di lettura necessaria per dare una definizione che permetta di misurare il gap informativo tra persistenza biparametrica e monoparametrica.

Il Capitolo 6 rappresenta un capitolo cruciale del mio lavoro di Tesi, poiché lì verranno *introdotte due definizioni*, connesse tra loro, che risponderanno ai due problemi formulati nei capitoli precedenti: la **differenza topologica**, che permetterà di misurare il già menzionato gap informativo, e la **correlazione topologica**, che permetterà di stabilire quando due funzioni continue definite su uno spazio topologico compatto e a valori reali sono correlate tra loro. L'idea alla base delle definizioni è la seguente: due funzioni con un'alta correlazione topologica indurranno un basso gap informativo tra i moduli di persistenza monoparametrici e biparametrici ad esse associati.

Infine, nel Capitolo 7 vengono presentati possibili sviluppi futuri, divisi per ambito — sviluppi topologici, geometrici, algebrici e implementativo-computazionali —, e possibili contesti applicativi per le definizioni date.

Per concludere, sono presenti due appendici.

Nell'Appendice A *si ripercorrerà brevemente quanto visto nei capitoli da 2 a 5* per adattare i vari concetti al **contesto dei complessi simpliciali**.

L'Appendice B invece è dedicata all'**Analisi Dati Topologica**, contesto da cui nasce questo lavoro di Tesi: illustreremo le *idee principali* e porteremo avanti una carrellata degli *strumenti più importanti* utilizzati in tale ambito.

All'irriducibile persistenza delle idee e dei legami umani

Indice

Introduzione	7
1 Prerequisiti teorici	15
1.1 Omologia singolare	15
1.1.1 Omologia singolare a coefficienti in \mathbb{Z}	16
1.1.2 Omologia singolare a coefficienti arbitrari	20
1.1.3 Calcolo dei gruppi di omologia	22
1.2 Moduli graduati e multi-graduati	24
1.2.1 Moduli graduati	24
1.2.2 Moduli multi-graduati	25
2 Moduli di persistenza	27
2.1 Definizione categoriale	27
2.2 Definizione algebrica	28
2.3 Relazione tra le definizioni categoriale ed algebrica	28
3 Omologia persistente	33
3.1 Definizione e prime proprietà	34
3.2 Caso monparametrico	40
3.2.1 Invarianti e descrittori	41
3.2.2 Distanza di bottleneck e stabilità	46
3.3 Caso multiparametrico	52
3.3.1 Invarianti e descrittori	53
3.3.2 Distanza di matching e stabilità	55
4 Problema 1: mono- vs multi-persistenza	61
4.1 Formulazione del Problema 1	61
4.1.1 Invarianti e descrittori a confronto	62
4.1.2 Efficacia discriminante a confronto	63
4.1.3 Formulazione del problema	65
4.2 Esempi	65

5	Problema 2: correlazione tra funzioni	71
5.1	Formulazione del Problema 2	71
5.1.1	Correlazione tra variabili	71
5.1.2	Formulazione del problema	73
5.2	Correlazione tra funzioni	74
5.2.1	Stato dell'Arte	74
5.2.2	Esistono definizioni più generali?	79
6	Due soluzioni connesse	81
6.1	Soluzione 1: differenza topologica	81
6.1.1	Definizione	81
6.1.2	Prime proprietà	82
6.1.3	Esempi	83
6.2	Soluzione 2: correlazione topologica tra le componenti di una funzione	83
6.2.1	Definizione	83
6.2.2	Prime proprietà	84
6.2.3	Esempi	86
6.3	Soluzione 2: correlazione topologica tra collezioni di funzioni .	87
6.3.1	Definizione	87
6.3.2	Esempi	87
7	Conclusioni e sviluppi futuri	89
7.1	Sviluppi topologici	90
7.2	Sviluppi algebrici	90
7.3	Sviluppi geometrici	91
7.4	Sviluppi computazionali e implementativi	92
7.5	Possibili applicazioni	93
A	Omologia persistente di complessi simpliciali	95
A.1	Complessi simpliciali	95
A.2	Omologia simpliciale	97
A.3	Filtrazioni e persistenza simpliciali	98
A.3.1	Problema 1 e Problema 2	103
B	Analisi Dati Topologica	107
	Bibliografia	113
	Ringraziamenti	115

Capitolo 1

Prerequisiti teorici: omologia singolare e moduli graduati e multigraduati

Prima di addentrarci nella trattazione centrale della Tesi, presentiamo le strutture matematiche e le rispettive proprietà su cui si fondano tutti i concetti che affronteremo. In particolare, in questo capitolo analizzeremo l'omologia singolare e i moduli graduati.

1.1 Omologia singolare

Due concetti con cui in un modo o nell'altro tutti i matematici hanno familiarità sono il funtore π_0 e il funtore π_1 ; dato uno spazio topologico X , $\pi_0(X)$ è l'*insieme delle sue componenti connesse*, mentre $\pi_1(X, x_0)$ è il primo gruppo di omotopia di X rispetto ad un punto base $x_0 \in X$ fissato, e viene anche detto *gruppo fondamentale*. Essi sono invarianti topologici, in particolare, se due spazi topologici hanno diversi π_0 o diversi gruppi fondamentali π_1 , allora tali spazi non sono omeomorfi. Cosa descrivono tali invarianti? Intuitivamente, dato uno spazio topologico X , $\pi_0(X)$ “conta” il numero di componenti connesse di X , mentre il gruppo fondamentale $\pi_1(X, x_0)$ “conta” il numero di “buchi” di X . Tuttavia, tali invarianti non sono molto raffinati nel momento in cui si lavora con spazi topologici di “dimensione” alta, ad esempio non sono in grado di distinguere le sfere S^n con $n \geq 2$. Pertanto, si definiscono i cosiddetti *gruppi di omotopia superiore* $\pi_n(X)$, tramite le mappe dal cubo I^n di dimensione n a X e le omotopie $I^n \times I \rightarrow X$ esistenti tra tali mappe. Purtroppo, calcolare tali gruppi risulta molto complesso e dispendioso, anche per spazi “semplici” come le sfere S^n . È qui che entrano in gioco i gruppi di omologia (singolare): essi costituiscono un'alternativa più semplice dal punto di vista computazionale.

1.1.1 Omologia singolare a coefficienti in \mathbb{Z}

Per definire l'omologia, partiamo dalla stessa idea alla base della classificazione delle superfici (compatte e connesse). Infatti, esse sono triangolabili e la classificazione si basa proprio sulle loro possibili triangolazioni. Ora generalizzeremo il concetto di triangolo, ottenendo delle strutture, possibilmente di qualunque dimensione, che permettano di suddividere uno spazio topologico interamente grazie ad esse, così come avviene per i triangoli nelle triangolazioni di spazi triangolabili.

Consideriamo \mathbb{R}^n , con $n \geq 1$, dotato della topologia euclidea e poniamo:

$$v_0 := (0, \dots, 0) \in \mathbb{R}^d \quad (1.1)$$

$$\forall i = 1, \dots, n \ v_i := (0, \dots, 0, 1, 0, \dots, 0), \quad (1.2)$$

con l'entrata pari a 1 in posizione i .

Definizione 1.1.1. *Definiamo*

$$\Delta_n := \left\{ \sum_{i=0}^n \lambda_i v_i \mid \forall_{i=1, \dots, n} \lambda_i \in \mathbb{R}, \lambda_i \geq 0, \sum_{i=0}^n \lambda_i = 1 \right\}$$

l' n -esimo **simplexso standard**, che denotiamo $\Delta_n = [v_0, v_1, \dots, v_n]$.

Alcuni simplexsi standard sono rappresentati in Figura 1.1.

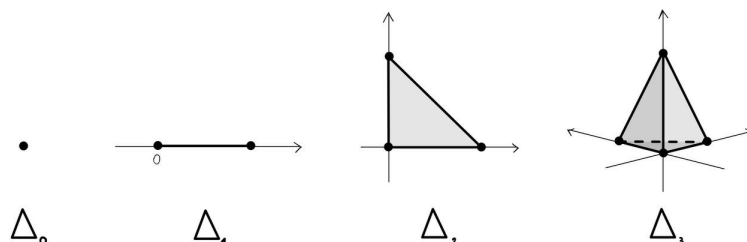


Figura 1.1: Esempi di simplexsi standard.

A partire da questi, si può definire l'*omologia simpliciale*, per cui rimandiamo a [17] e di cui parleremo nell'Appendice A. Noi invece ci concentreremo sull'*omologia singolare*, che si basa su simplexsi più astratti, i *simplexsi singolari*, in modo da poter lavorare non solo con \mathbb{R}^n , ma con generici spazi topologici.

Sia X uno spazio topologico.

Definizione 1.1.2. *Un n -simplexso singolare su X è una mappa continua $\sigma : \Delta_n \rightarrow X$.*

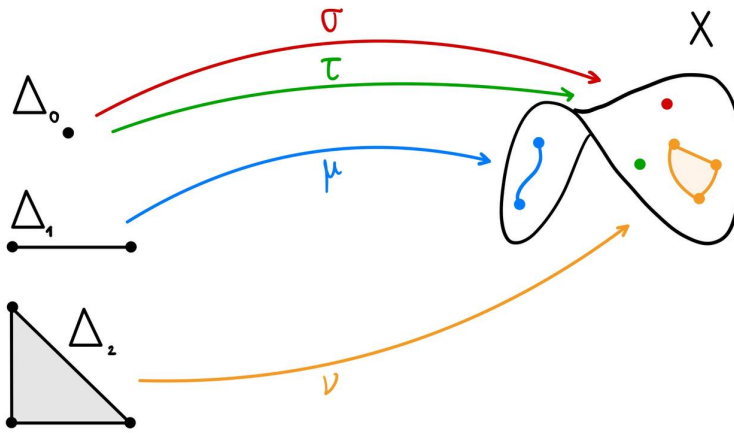


Figura 1.2: Rappresentazione di alcuni semplici singolari definiti su uno spazio topologico X .

Alcuni semplici singolari sono rappresentati in Figura 1.2.

Osservazione 1.1.3. *L'unica condizione richiesta è la continuità della funzione considerata, pertanto nulla garantisce che venga mantenuta la struttura topologica di Δ_n su X ; ad esempio possono verificarsi situazioni come quella rappresentata in Figura 1.3. È proprio per questo che si utilizza il termine*

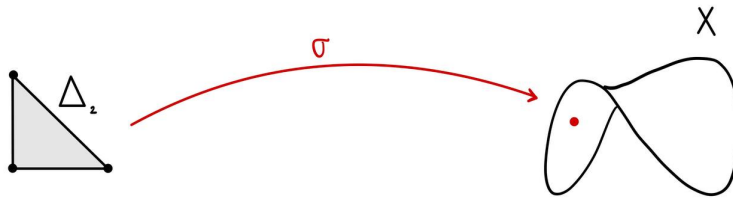


Figura 1.3: I semplici singolari devono il loro nome al fatto che possono verificarsi situazioni come quella rappresentata in figura.

“singolari”.

A questo punto, l'idea è quella di definire una sequenza di gruppi abeliani a partire dai semplici singolari, chiamata *complesso di catene*, di cui potremo considerare l'omologia, e otterremo così i *gruppi di omologia singolare*.

Definizione 1.1.4. *Per ogni $p \in \mathbb{N}$, denotiamo con $C_p(X)$ il gruppo abeliano libero generato dai p -semplici singolari su X . Ogni elemento $\omega \in C_p(X)$ si scriverà come*

$$\sum_{i \in I} k_i \sigma_i,$$

con I un insieme finito di indici e per ogni $i \in I$, $k_i \in \mathbb{Z}$ e σ_i un p -simplex singolare su X . Chiamiamo gli elementi di $C_p(X)$ le p -**catene** di X .

Per ottenere una sequenza di gruppi abeliani, ci occorre definire degli omomorfismi di gruppi abeliani $C_p(X) \rightarrow C_{p-1}(X)$ tali che la loro composizione sia nulla.

Definizione 1.1.5. Siano $p \in \mathbb{N}$ e $\sigma : \Delta_p \rightarrow X$ un p -simpleso singolare su X . Se $p \geq 1$, per ogni $i = 0, \dots, p$, consideriamo

$$\sigma|_{[v_0, \dots, \hat{v}_i, \dots, v_n]},$$

cioè restringiamo il dominio di σ da Δ_p al simpleso generato da tutti i vertici di Δ_p ad eccezione dell' i -esimo.

Definiamo poi

$$\partial_p^X \sigma := \sum_{i=0}^p (-1)^i \sigma|_{[v_0, \dots, \hat{v}_i, \dots, v_n]}.$$

Infine, dato $\omega = \sum_{i \in I} k_i \sigma_i \in C_p(X)$, definiamo

$$\partial_p^X \omega := \sum_{i=1}^n k_i \partial_p^X \sigma_i.$$

Invece, se $p = 0$ poniamo $C_{-1}(X) := \mathbf{0}$ e per ogni 0-catena ω definiamo $\partial_0^X \omega := 0$.

Abbiamo così ottenuto per ogni $p \in \mathbb{N}$ il p -esimo **operatore di bordo** su X $\partial_p^X : C_p(X) \rightarrow C_{p-1}(X)$.

Per alleggerire la notazione, quando sarà chiaro dal contesto, ometteremo di indicare X nella notazione dell'operatore di bordo. Intuitivamente, come rappresentato in Figura 1.4, dove con $\partial^{(i), X} \sigma$ indichiamo $\sigma|_{[v_0, \dots, \hat{v}_i, \dots, v_n]}$, l'operatore di bordo è una mappa che manda ogni p -simpleso σ in una $(p-1)$ -catena che ha coefficienti non nulli solo in corrispondenza dei $(p-1)$ -simplessi che costituiscono il bordo di σ , ed è a questo che deve il suo nome. Chiaramente, in quest'ottica acquisisce senso anche la definizione di $C_{-1}(X) = \mathbf{0}$ e $\partial_0 = \mathbf{0}$.

Si può dimostrare il seguente lemma.

Lemma 1.1.6. Per ogni $p \in \mathbb{N}$, $\partial_{p+1} \circ \partial_p = 0$.

Per una dimostrazione, rimandiamo a [17].

Grazie alle definizioni date e al Lemma 1.1.6, abbiamo perciò ottenuto una sequenza di gruppi abeliani $C(X) := \{(C_p(X), \partial_p)\}_{p \in \mathbb{N}}$, detta **complesso di catene** di X :

$$\dots \xrightarrow{\partial_{p+2}} C_{p+1}(X) \xrightarrow{\partial_{p+1}} C_p(X) \xrightarrow{\partial_p} C_{p-1}(X) \xrightarrow{\partial_{p-1}} \dots \xrightarrow{\partial_1} C_0(X) \xrightarrow{\partial_0} \mathbf{0}.$$

In particolare, possiamo dare le seguenti definizioni.

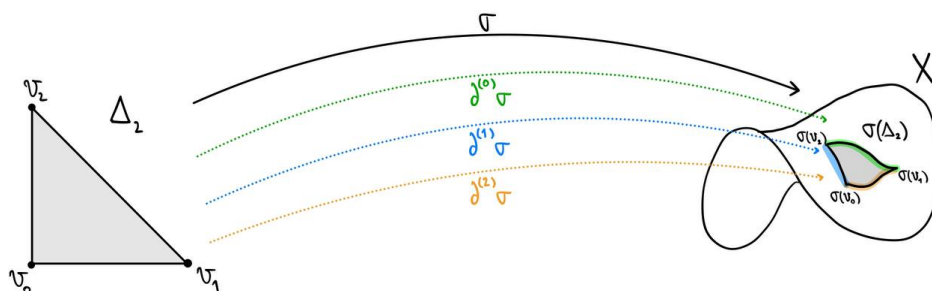


Figura 1.4: Rappresentazione intuitiva dell'operatore di bordo. Con $\partial^{(i)}\sigma$ indichiamo $\sigma|_{[v_0, \dots, \hat{v}_i, \dots, v_n]}$.

Definizione 1.1.7. Per ogni $p \in \mathbb{N}$,

- il sottogruppo $Z_p(X) := \ker \partial_p$ di $C_p(X)$ è chiamato **gruppo dei p -cicli** di X ;
- il sottogruppo $B_p(X) := \text{im} \partial_{p+1}$ di $C_n(X)$ è chiamato **gruppo dei p -bordi** di X ;
- il **p -esimo gruppo di omologia singolare** di X a coefficienti in \mathbb{Z} è il quoziente

$$H_p(X, \mathbb{Z}) := H_p(X) := Z_p(X)/B_p(X).$$

Intuitivamente, possiamo affermare che il p -esimo gruppo di omologia di X identifica p -cicli che non sono bordi di $(p+1)$ -cicli, cioè i “buchi di dimensione p ” di X .

Osservazione 1.1.8. Il quoziente che definisce i gruppi di omologia è ben posto grazie al Lemma 1.1.6 e poiché siamo in un contesto abeliano.

Presenteremo alcuni esempi nella Sezione 1.1.3.

Ora possiamo compiere un ulteriore passo avanti, rendendo gli H_p dei funtori. Per prima cosa, dobbiamo a tal fine definire dei morfismi tra complessi di catene. Possiamo definire tali morfismi a partire da funzioni continue.

Siano X e Y due spazi topologici e sia $f : X \rightarrow Y$ una funzione continua.

Definizione 1.1.9. Per ogni $p \in \mathbb{N}$, la mappa

$$f_{\#} : C_p(X) \rightarrow C_p(Y) \mid \sum k_i \sigma_i \mapsto \sum k_i (f \circ \sigma_i)$$

è un **morfismo di complessi di catene**.

Si può dimostrare che i morfismi di complessi di catene commutano con l'operatore di bordo, cioè

$$\partial^Y f_{\#} = f_{\#} \partial^X,$$

da cui segue che in realtà $f_{\#}$ induce anche un morfismo a livello di omologia:

$$\forall p \in \mathbb{N} \ f_*^{(p)} : H_p(X) \rightarrow H_p(Y) \mid [\omega] \mapsto [f_{\#}\omega],$$

dove $[\cdot]$ indica la classe nel quoziente considerato.

Si dimostra inoltre che $(\cdot)_*$ soddisfa le proprietà funtoriali sui morfismi. Pertanto, per ogni $p \in \mathbb{N}$ abbiamo un funtore dalla categoria degli spazi topologici a quella dei gruppi abeliani H_p , che ad ogni spazio topologico X associa $H_p(X)$ e ad ogni mappa continua $f : X \rightarrow Y$ associa $f_*^{(p)}$.

In particolare, grazie a questa struttura funtoriale si riesce a dimostrare che:

- i gruppi di omologia singolare sono invarianti topologici,
- i gruppi di omologia singolare sono invarianti per equivalenza omotopica¹

Possiamo infine dare la seguente definizione.

Definizione 1.1.10. Per ogni n definiamo $b_p(X) := \text{rk} H_p(X) \in \mathbb{N}$ il p -esimo numero di Betti di X . Definiamo inoltre la **caratteristica di Eulero** di X $\chi(X) := \sum_{p \in \mathbb{N}} (-1)^p b_p(X)$.

Si dimostra che sia i numeri di Betti che, di conseguenza, la caratteristica di Eulero sono invarianti topologici (numerici).

1.1.2 Omologia singolare a coefficienti arbitrari

Finora abbiamo lavorato con l'omologia singolare a coefficienti in \mathbb{Z} , poiché i coefficienti delle p -catene venivano scelti tra gli interi. In realtà, è possibile lavorare anche con una diversa scelta dei coefficienti e, spesso, in realtà lavorare a coefficienti in \mathbb{Z} si rivela impossibile nella pratica. Invece, lavorare ad esempio a coefficienti in un campo — come faremo anche noi nel resto della trattazione — risulta molto vantaggioso. Un primo vantaggio dell'approccio tramite coefficienti in un campo è che i gruppi di omologia che si ottengono sono in realtà spazi vettoriali a coefficienti in quel campo; di conseguenza, i numeri di Betti saranno invece le dimensioni di tali spazi vettoriali. Inoltre, grazie al Teorema dei coefficienti universali, si esibisce un forte legame tra i gruppi di omologia a coefficienti in \mathbb{Z} e quelli a coefficienti in un qualsiasi gruppo abeliano, rendendo così possibile lavorare con i coefficienti più vantaggiosi di volta in volta.

¹così come il gruppo fondamentale.

Siano X uno spazio topologico e G un gruppo abeliano. Generalizziamo le definizioni viste in precedenza nel caso dei coefficienti in \mathbb{Z} nel modo più naturale possibile.

Definizione 1.1.11. *Per ogni $p \in \mathbb{N}$ poniamo*

- per $p \geq 1$,

$$C_p(X, G) := \left\{ \sum_{i \in I} g_i \sigma_i \mid I \text{ finito, } \forall i \in I \ g_i \in G, \sigma_i \text{ } p\text{-simpl. sing.} \right\}$$

il gruppo delle p -catene a coefficienti in G di X ;

- per $p \geq 1$, *definiamo l'operatore di bordo di X $\partial_p^{X,G} : C_p(X, G) \rightarrow C_{p-1}(X, G)$ in modo totalmente analogo alla Definizione 1.1.5;*
- per $p = 0$, *poniamo $C_{-1}(X, G) := \mathbf{0}$ e $\partial_0^{X,G} := \mathbf{0}$.*

Nuovamente, si dimostra l'analogo del Lemma 1.1.6 e otteniamo così un complesso di catene $C(X, G) := \{(C_p(X, G), \partial_p^{X,G})\}_{p \in \mathbb{N}}$. La sua omologia è chiamata **omologia singolare di X a coefficienti in G** .

Osservazione 1.1.12. *Se $G = \mathbb{Z}$, allora per ogni $p \in \mathbb{Z}$ $H_p(X, \mathbb{Z}) = H_p(X)$.*

Le proprietà viste per l'omologia a coefficienti in \mathbb{Z} valgono anche in questo contesto. Inoltre, si può dimostrare il seguente teorema.

Teorema 1.1.13 (Coefficienti universali). *Sia X uno spazio topologico connesso e sia G un gruppo abeliano. Per ogni $p \in \mathbb{N}$ con $p \geq 1$, abbiamo una successione esatta corta della forma seguente*

$$\mathbf{0} \rightarrow H_p(X) \otimes G \rightarrow H_p(X, G) \rightarrow \text{Tor}(H_{p-1}(X), G) \rightarrow \mathbf{0}.$$

Pertanto, per calcolare gli $H_p(X, G)$ è sufficiente calcolare $H_p(X)$, e viceversa.

Osservazione 1.1.14. *Se $G = \mathbb{Q}, \mathbb{C}, \mathbb{R}$ allora per ogni $p \in \mathbb{N}$, $H_p(X, G) \cong H_p(X) \otimes G$, perciò in questi casi il calcolo risulta ancora più semplice.*

Osservazione 1.1.15. *Per garantire il Teorema 1.2.3, che ci servirà nel Capitolo 3, avremo bisogno di lavorare con l'omologia a coefficienti in un campo. Per questo motivo, se non diversamente specificato, lavoreremo con l'omologia a coefficienti in $\mathbb{Z}/2\mathbb{Z}$.*

1.1.3 Gruppi di omologia: come calcolarli e quali informazioni topologiche descrivono?

Nelle due sezioni precedenti abbiamo descritto la costruzione e le prime proprietà dei gruppi di omologia, potendone apprezzare la natura estremamente algebrica e astratta. In realtà, l'omologia descrive caratteristiche topologiche degli spazi considerati, che possono essere diverse a seconda della scelta dei coefficienti. Ora torniamo a lavorare a coefficienti in \mathbb{Z} , ma sappiamo che possiamo facilmente adattare la trattazione a coefficienti qualunque, grazie al Teorema 1.1.13.

Le caratteristiche topologiche più immediate da ricavare tramite l'omologia sono quelle riguardanti la connessione. Infatti, non è difficile dimostrare la seguente proposizione.

Proposizione 1.1.16. *Sia $p \in \mathbb{N}$.*

- Se X è un punto, allora

$$H_p(X) = \begin{cases} \mathbb{Z} & p = 0, \\ \mathbf{0} & \text{altrimenti}; \end{cases}$$

- Se X è connesso per archi, allora $H_0(X) = \mathbb{Z}$;
- Se X è contrattile allora

$$H_p(X) = \begin{cases} \mathbb{Z} & p = 0, \\ \mathbf{0} & \text{altrimenti}. \end{cases}$$

Inoltre, esiste un legame tra il gruppo fondamentale e il primo gruppo di omologia, grazie al seguente teorema.

Teorema 1.1.17 (Hurwitz). *Sia X uno spazio topologico e sia $x_0 \in X$ un punto fissato. Esiste un omomorfismo di gruppi $\eta : \pi_1(X, x_0) \rightarrow H_1(X)$. Se inoltre X è connesso per archi, allora η è surgettivo e si ha che $\ker \eta = [\pi_1(X, x_0), \pi_1(X, x_0)]$, cioè il suo kernel coincide con il gruppo dei commutatori di $\pi_1(X, x_0)$.*

Segue che, se X è connesso per archi allora

$$H_1(X) = \pi_1(X, x_0) / [\pi_1(X, x_0), \pi_1(X, x_0)],$$

cioè in tal caso si ha che $H_1(X)$ è l'abelianizzato di $\pi_1(X)$. In particolare, per quanto visto finora e grazie al Teorema 1.1.17, possiamo concludere che:

- per una circonferenza S^1 , $H_1(S^1) = \mathbb{Z}$;
- per l'unione ad un punto $S^1 \vee S^1$ di due circonferenze, $H_1(S^1 \vee S^1) = \mathbb{Z} \times \mathbb{Z}$;

- per un toro T , $H_1(T) = \mathbb{Z} \times \mathbb{Z} = H_1(S^1 \vee S^1)$;
- per una somma connessa di g copie di un toro, $H_1(T \# \dots \# T) = \mathbb{Z}^{2g}$;
- per il piano proiettivo reale $\mathbb{P}_{\mathbb{R}}^2$, $H_1(\mathbb{P}_{\mathbb{R}}^2) = \mathbb{Z}/2\mathbb{Z}$;
- per una somma connessa di g copie del piano proiettivo reale,

$$H_1(\mathbb{P}_{\mathbb{R}}^2 \# \dots \# \mathbb{P}_{\mathbb{R}}^2) = \mathbb{Z}^{2g-1} \oplus \mathbb{Z}/2\mathbb{Z}.$$

Come accennato all'inizio della sezione, il calcolo dei gruppi di omologia superiore risulta più agevole rispetto a quello dei gruppi di omotopia superiore. Questo perché per i gruppi di omologia vale il teorema di Mayer-Vietoris [17], che possiamo pensare come all'analogo del Teorema di Seifert van Kampen per il calcolo del gruppo fondamentale. Ad esempio, grazie al teorema di Mayer-Vietoris si possono trovare facilmente i seguenti gruppi di omologia:

- per la sfera S^n ,

$$H_p(S^n) = \begin{cases} \mathbb{Z} & p = 0, n \\ \mathbf{0} & \text{altrimenti;} \end{cases}$$

- per l'unione ad un punto di g copie di S^1 ,

$$H_p(S^1 \vee S^1 \vee \dots \vee S^1) = \begin{cases} \mathbb{Z} & p = k \\ \mathbb{Z}^g & p = 1 \\ \mathbf{0} & \text{altrimenti;} \end{cases}$$

- per una somma connessa di g copie del toro T ,

$$H_p(T \# T \# \dots \# T) = \begin{cases} \mathbb{Z} & p = 0 \\ \mathbb{Z}^{2g} & p = 1 \\ \mathbf{0} & \text{altrimenti;} \end{cases}$$

- per una somma connessa di g copie del piano proiettivo reale $\mathbb{P}_{\mathbb{R}}^2$,

$$H_p(\mathbb{P}_{\mathbb{R}}^2 \# \mathbb{P}_{\mathbb{R}}^2 \# \dots \# \mathbb{P}_{\mathbb{R}}^2) = \begin{cases} \mathbb{Z} & p = 0 \\ \mathbb{Z}^{2g-1} \oplus \mathbb{Z}/2\mathbb{Z} & p = 1 \\ \mathbf{0} & \text{altrimenti;} \end{cases}$$

- dato che la retta proiettiva complessa $\mathbb{P}_{\mathbb{C}}^1$ è omeomorfa alla sfera S^2 , abbiamo

$$H_p(\mathbb{P}_{\mathbb{C}}^1) = H_p(S^2) = \begin{cases} \mathbb{Z} & p = 0, 2 \\ \mathbf{0} & \text{altrimenti.} \end{cases}$$

Inoltre, sempre grazie al Teorema di Mayer-Vietoris, si può dimostrare il seguente risultato.

Teorema 1.1.18 (Invarianza del dominio, Brouwer). *Per $n, m \in \mathbb{N}$, \mathbb{R}^n è omeomorfo a \mathbb{R}^m se e solo se $n = m$.*

Questo risultato generalizza la capacità di π_0 e π_1 di distinguere tra \mathbb{R}^1 , \mathbb{R}^2 ed \mathbb{R}^3 a tutti gli \mathbb{R}^n . Inoltre, un suo corollario è il ben noto Teorema di punto fisso di Brouwer.

Corollario 1.1.19 (Punto fisso, Brouwer). *Per $n \in \mathbb{N}$ con $n \geq 2$, posto D^n il disco in \mathbb{R}^n di raggio 1, si ha che se $f : D^n \rightarrow D^n$ è una funzione continua allora esiste un punto fisso rispetto ad f , cioè esiste $x_0 \in D^n$ tale che $f(x_0) = x_0$.*

1.2 Moduli graduati e multi-graduati

In questa sezione ci occuperemo di definire delle strutture algebriche che si riveleranno fondamentali nel seguito della trattazione: i moduli graduati e multi-graduati. Vedremo infatti che i gruppi di omologia provenienti da sottospazi concatenati di uno spazio topologico daranno origine a moduli graduati o multi-graduati.

1.2.1 Moduli graduati

La definizione di modulo graduato si basa su quella di anello graduato.

Sia A un anello.

Definizione 1.2.1. *Diciamo che A è un **anello graduato** se come gruppo abeliano può essere scritto come somma diretta $A = \bigoplus_{u \in \mathbb{N}} A_u$, dove gli A_u sono sottogruppi di A tali che $A_u A_v \subseteq A_{u+v}$.*

In particolare, segue che A_0 è un sottoanello di A e che ogni A_u è un A_0 -modulo.

Sia ora M un A -modulo.

Definizione 1.2.2. *Diciamo che M è un **A -modulo graduato** se come gruppo abeliano può essere scritto come $M = \bigoplus_{u \in \mathbb{N}} M_u$, dove gli M_u sono gruppi abeliani tali che $A_u M_v \subseteq M_{u+v}$.*

Segue dalla definizione che ogni M_u è un A_0 -modulo. Gli elementi di ciascun E_u sono chiamati **elementi omogenei di grado u** di M .

Noi lavoreremo con moduli graduati su $\mathbb{K}[x]$, dove \mathbb{K} è un campo. Questo perché in tal caso, vale il seguente teorema, che risulterà fondamentale nel Capitolo 3.

Teorema 1.2.3 (Teorema di Struttura per moduli graduati su un PID graduato). *Ogni modulo M graduato su $\mathbb{K}[x]^2$ si decompone con unicità nel modo seguente:*

$$M \cong \left(\bigoplus_{i=1}^N \mathbb{K}[x](-k_i) \right) \oplus \left(\bigoplus_{j=1}^M \mathbb{K}[x]/(x^{l_j-m_j}) \right) \quad (1.3)$$

dove i k_i sono elementi omogenei tali che $d_i \mid d_{i+1}$ e gli $l_j, m_j \in \mathbb{Z}$ e con $\mathbb{K}[x](-k_i)$ indichiamo la k_i -traslazione nei gradi di $K[x]$.

Come in molti Teoremi di Struttura, anche in questo caso quello che si ottiene è una parte libera ed una parte di torsione.

Concludiamo la sezione con un esempio e, per ulteriori approfondimenti, rimandiamo a [19].

Esempio 1.2.4. *Sia \mathbb{K} un campo. L'anello dei polinomi $\mathbb{K}[x_1, \dots, x_r]$ è un \mathbb{K} -modulo graduato, che per semplicità indichiamo con M . Abbiamo $M_0 = \mathbb{K}$. Inoltre, gli elementi omogenei di grado u di M sono i monomi in x_1, \dots, x_r di grado u , cioè i monomi $x_1^{n_1} \cdot \dots \cdot x_r^{n_r}$ con $\sum_{i=1}^r n_i = u$.*

1.2.2 Moduli multi-graduati

Nel seguito della trattazione, avremo a che fare con funzioni a valori in \mathbb{R} , che daranno origine a moduli graduati come quelli appena definiti, ma anche con funzioni a valori in \mathbb{R}^d , le quali invece daranno origine a moduli multi-graduati, che pertanto ci apprestiamo a definire.

Come viene descritto in [9], a noi interesseranno moduli multi-graduati definiti su $A := \mathbb{K}[x_1, \dots, x_d]$, dove \mathbb{K} è un campo e $d \in \mathbb{N}$, $d > 1$. Sia M un A -modulo.

Definizione 1.2.5. *Diciamo che M è \mathbb{N}^d -graduato se $M = \bigoplus_{u \in \mathbb{N}^d} M_u$ dove gli M_u sono \mathbb{K} -spazi vettoriali tali che*

$$\forall i = 1, \dots, d, u \in \mathbb{N}^d \quad x_i \cdot M_u \subseteq M_{u+e_i}.$$

I multi-indici $u \in \mathbb{N}^d$ sono chiamati **multi-gradi** e gli elementi di ciascun M_u sono detti **elementi omogenei di multi-grado u** di M .

Dalla definizione, come vedremo meglio in seguito, segue che possiamo pensare ad un modulo \mathbb{N}^d -graduato come ad una collezione di \mathbb{K} -spazi vettoriali M_u e mappe \mathbb{K} -lineari $x_i \cdot : M_u \rightarrow M_{u+e_i}$ tali che $(x_j \cdot) \circ (x_i \cdot) = (x_i \cdot) \circ (x_j \cdot) = x_{i+j} \cdot$. Ad esempio, per $d = 2$ abbiamo:

²è un PID perché \mathbb{K} è un campo

$$\begin{array}{ccccc}
\vdots & & \vdots & & \vdots \\
x_2 \cdot \uparrow & & x_2 \cdot \uparrow & & x_2 \cdot \uparrow \\
M_{(0,2)} & \xrightarrow{x_1 \cdot} & M_{(1,2)} & \xrightarrow{x_1 \cdot} & M_{(2,2)} \xrightarrow{x_1 \cdot} \dots \\
x_2 \cdot \uparrow & & x_2 \cdot \uparrow & & x_2 \cdot \uparrow \\
M_{(0,1)} & \xrightarrow{x_1 \cdot} & M_{(1,1)} & \xrightarrow{x_1 \cdot} & M_{(2,1)} \xrightarrow{x_1 \cdot} \dots \\
x_2 \cdot \uparrow & & x_2 \cdot \uparrow & & x_2 \cdot \uparrow \\
M_{(0,0)} & \xrightarrow{x_1 \cdot} & M_{(1,0)} & \xrightarrow{x_1 \cdot} & M_{(2,0)} \xrightarrow{x_1 \cdot} \dots
\end{array}$$

dove ogni quadrato è commutativo. Inoltre, per passare ad esempio da $M_{(0,0)}$ ad $M_{(0,2)}$ avremmo la mappa $x_2^2 \cdot$, che in realtà si può ottenere dalla composizione di $x_2 \cdot$ con se stessa.

Concludiamo la sezione mostrando un semplice esempio.

Esempio 1.2.6. Per ogni $u = (u_1, \dots, u_d) \in \mathbb{N}^d$ poniamo $M_u := \langle x^u \rangle_{\mathbb{K}} = \langle x_1^{u_1} \cdot x_2^{u_2} \cdot \dots \cdot x_d^{u_d} \rangle_{\mathbb{K}}$. Allora $\bigoplus_{u \in \mathbb{N}^d} M_u$ è un $\mathbb{K}[x_1, \dots, x_d]$ -modulo \mathbb{N}^d -graduato.

Capitolo 2

Moduli di persistenza

Uno degli scopi di questo lavoro di Tesi è quello di confrontare funzioni, o meglio definire strumenti matematici che permettano di stabilire quanto due funzioni siano “simili” tra loro e soprattutto che permettano di distinguere funzioni differenti. Un modo per mettere a confronto funzioni è fornito dai *moduli di persistenza*. I moduli di persistenza possono essere definiti in due modi: uno algebrico e uno categoriale.

In questo capitolo studieremo le due definizioni e le confronteremo. Vedremo in seguito come costruire moduli di persistenza a partire da funzioni, nel Capitolo 3.

2.1 Definizione categoriale

Siano \mathbb{K} un campo, (\mathcal{P}, \leq) un poset e d un intero strettamente positivo. Indichiamo con \mathbf{Vec} la categoria dei \mathbb{K} -spazi vettoriali di dimensione finita.

Definizione 2.1.1. *Un \mathcal{P} -modulo di persistenza è un funtore*

$$M : \mathcal{P} \rightarrow \mathbf{Vec}.$$

Come anticipato in precedenza, in seguito vedremo come associare moduli di persistenza a funzioni. In particolare, tali funzioni assumeranno valori in \mathbb{R}^d e in \mathbb{N}^d , ed è per questo motivo che vorremmo considerare come poset di base \mathbb{R}^d ed \mathbb{N}^d . Pertanto, risulta necessario dotare tali insiemi di una relazione d'ordine parziale. Procediamo sfruttando il confronto coordinata per coordinata. Per ogni $u = (u_1, \dots, u_d), v = (v_1, \dots, v_d)$ in \mathbb{N}^d o in \mathbb{R}^d , diciamo che

$$u \leq v \quad \text{se e solo se} \quad \forall i = 1, \dots, d \quad u_i \leq v_i. \quad (2.1)$$

Chiaramente, nel caso in cui $d = 1$ ritroviamo la relazione d'ordine naturale su \mathbb{N} e su \mathbb{R} .

Per i nostri scopi, sarà necessario lavorare sotto la seguente assunzione.

Assunzione 2.1.2. *Considereremo sempre \mathcal{P} -moduli di persistenza di tipo finito.*

2.2 Definizione algebrica

Siano \mathbb{K} un campo e d un intero strettamente positivo.

Definizione 2.2.1. *Un modulo di persistenza algebrico è un $\mathbb{K}[x_1, \dots, x_d]$ -modulo \mathbb{N}^d -graduato (con gradazione standard).*

Anche in questo caso, ci occorre assumere la seguente condizione.

Assunzione 2.2.2. *Considereremo sempre $\mathbb{K}[x_1, \dots, x_d]$ -moduli di persistenza finitamente generati.*

2.3 Relazione tra le definizioni categoriale ed algebrica

Abbiamo visto due possibili definizioni per i moduli di persistenza. Le due definizioni sono legate? Nel caso lo siano, come si può passare da una definizione all'altra?

Inoltre, abbiamo preannunciato che in seguito ci servirà poter lavorare con \mathbb{R}^d - e \mathbb{N}^d -moduli di persistenza. Che relazione intercorre tra di essi? E come si può passare da uno all'altro, se possibile? In questa sezione risponderemo a queste domande grazie ai seguenti fatti, che ora ci limitiamo ad elencare per averne una visione d'insieme, ma subito dopo dimostreremo:

- ogni \mathbb{N}^d -modulo di persistenza M determina un \mathbb{R}^d -modulo di persistenza M' ,
- ogni \mathbb{R}^d -modulo di persistenza M' determina un \mathbb{N}^d -modulo di persistenza M
- ogni \mathbb{N}^d -modulo di persistenza M determina un modulo di persistenza algebrico M'' su $\mathbb{K}[x_1, \dots, x_d]$,
- ogni modulo di persistenza algebrico M'' su $\mathbb{K}[x_1, \dots, x_d]$ definisce un \mathbb{N}^d -modulo di persistenza M .

Indicando con α, β, γ e δ rispettivamente le quattro assegnazioni appena elencate, si dimostra poi che:

- $\beta(\alpha(M)) = M$,
- $\delta(\gamma(M)) = M$,
- $\gamma(\delta(M'')) = M''$.

2.3. RELAZIONE TRA LE DEFINIZIONI CATEGORIALE ED ALGEBRICA 29

Invece, non è in generale vero che $\alpha(\beta(M')) = M'$. Inoltre, in realtà, si può dimostrare di più: abbiamo un'equivalenza di categorie tra i \mathcal{P} -moduli di persistenza di tipo finito e i $\mathbb{K}[x_1, \dots, x_d]$ -moduli di persistenza finitamente generati.

Per garantire i risultati appena elencati, occorre lavorare sotto le assunzioni Assunzione 2.1.2 e Assunzione 2.2.2.

Nelle dimostrazioni utilizzeremo la seguente notazione.

Notazione 2.3.1. *Sia M un \mathcal{P} -modulo di persistenza. Allora:*

- per ogni $p \in \mathcal{P}$ indichiamo con $M_p := M(p)$ il \mathbb{K} -spazio vettoriale associato a p tramite M ;
- per ogni $p, q \in \mathcal{P}$ tali che $p \leq q$ indichiamo con $\mu_{p,q}^M := M(p \leq q)$ l'applicazione lineare $M_p \rightarrow M_q$ associata al morfismo $p \leq q$ tramite M .

Osservazione 2.3.2. *Sia M un \mathcal{P} -modulo di persistenza. Per definizione di funtore, abbiamo:*

- per ogni $p \in \mathcal{P}$, $\mu_{p,p}^M = id_{M_p}$;
- per ogni $p, q, r \in \mathcal{P}$ tali che $p \leq r \leq q$, $\mu_{p,r}^M = \mu_{r,q}^M \circ \mu_{p,r}^M$.

Proposizione 2.3.3. *Ogni \mathbb{N}^d -modulo di persistenza M determina un \mathbb{R}^d -modulo $\alpha(M) := M' : \mathbb{R}^d \rightarrow \mathbf{Vec}$ definito da:*

- per ogni $p \in \mathbb{R}^d$, $M'_p := M_{\lfloor p \rfloor}$,
- per ogni $p, q \in \mathbb{R}^d$ con $p \leq q$, $\mu_{p,q}^{M'} := \mu_{\lfloor p \rfloor, \lfloor q \rfloor}^M$,

dove per ogni $p = (p_1, \dots, p_d) \in \mathbb{R}^d$ indichiamo con

$$\lfloor p \rfloor := \begin{cases} (\lfloor p_1 \rfloor, \dots, \lfloor p_d \rfloor) & \text{se } (p_1, \dots, p_d) \geq (0, \dots, 0), \\ (0, \dots, 0) & \text{altrimenti.} \end{cases}$$

Dimostrazione. Basta osservare che le proprietà funtoriali di M vengono trasferite ad M' .

Per ogni $p \in \mathbb{R}^d$ abbiamo

$$\mu_{p,p}^{M'} = \mu_{\lfloor p \rfloor, \lfloor p \rfloor}^M = id_{M_{\lfloor p \rfloor}} = id_{M'_p}.$$

Per ogni $p, q, r \in \mathbb{R}^d$ abbiamo

$$\mu_{p,q}^{M'} = \mu_{\lfloor p \rfloor, \lfloor q \rfloor}^M = \mu_{\lfloor r \rfloor, \lfloor q \rfloor}^M \circ \mu_{\lfloor p \rfloor, \lfloor r \rfloor}^M = \mu_{r,q}^{M'} \circ \mu_{p,r}^{M'}.$$

□

Proposizione 2.3.4. *Ogni \mathbb{R}^d -modulo di persistenza M' determina, sfruttando l'immersione canonica $\mathbb{N}^d \hookrightarrow \mathbb{R}^d$, un \mathbb{N}^d -modulo di persistenza $\beta(M') := M$ definito come segue:*

- per ogni $p \in \mathbb{N}^d$, $M_p := M'_p$,
- per ogni $p, q \in \mathbb{N}^d$ con $p \leq q$, $\mu_{p,q}^M := \mu_{p,q}^{M'}$.

Dimostrazione. Analogamente alla Proposizione 2.3.3, nuovamente le proprietà functoriali di M seguono da quelle di M' . \square

Proposizione 2.3.5. *Per ogni \mathbb{N}^d -modulo di persistenza M abbiamo $\beta(\alpha(M)) = M$.*

Dimostrazione. Per definizione,

$$\beta'(\alpha'(M)) : \mathbb{N}^d \hookrightarrow \mathbb{R}^d \xrightarrow{\lfloor \cdot \rfloor} \mathbb{N}^d \xrightarrow{M} \mathbf{Vec},$$

ma $\mathbb{N}^d \hookrightarrow \mathbb{R}^d \xrightarrow{\lfloor \cdot \rfloor} \mathbb{N}^d$ è l'identità su \mathbb{N}^d e dunque concludiamo che $\beta(\alpha(M)) = M$. \square

Osservazione 2.3.6. *Invece, in generale non è vero che per ogni \mathbb{R}^d -modulo di persistenza M' si abbia $\alpha(\beta(M')) = M'$. Ad esempio, se consideriamo l' \mathbb{R}^d -modulo di persistenza definito da:*

- per ogni $p \in \mathbb{R}^d$,

$$M'_p := \begin{cases} \mathbb{K} & \text{se } p = (\frac{1}{2}, \dots, \frac{1}{2}) \\ 0 & \text{altrimenti;} \end{cases}$$

- per ogni $p, q \in \mathbb{R}^d$ con $p \leq q$,

$$\mu_{p,q}^{M'} := \begin{cases} id_{M'_p} & \text{se } p = q = (\frac{1}{2}, \dots, \frac{1}{2}) \\ 0 & \text{altrimenti,} \end{cases}$$

allora per costruzione $\alpha(\beta(M')) = \alpha(0) = 0 \neq M'$.

Proposizione 2.3.7. *Ogni \mathbb{N}^d -modulo di persistenza definisce un $\mathbb{K}[x_1, \dots, x_d]$ -modulo di persistenza algebrico $\gamma(M) := M''$ definito come segue:*

$$M'' := \bigoplus_{p \in \mathbb{N}^d} M_p,$$

dove per ogni $p \in \mathbb{N}^d$

- l'insieme M''_p degli elementi omogenei di grado p di M'' coinciderà con M_p ;

2.3. RELAZIONE TRA LE DEFINIZIONI CATEGORIALE ED ALGEBRICA 31

- per ogni $m \in M_p''$ e per ogni $k \in \mathbb{N}^d$ l'azione $m \cdot x^k$ di x^k su m coincide con $\mu_{p,p+k}^M$.

Dimostrazione. Per dimostrare che un tale M'' è un $\mathbb{K}[x_1, \dots, x_d]$ -modulo \mathbb{N}^d -graduato basta dimostrare che per ogni $p, k \in \mathbb{N}^d$ e per ogni $m \in M_p''$ l'azione $m \cdot x^k$ restituisce un elemento di grado $p+k$, il resto è immediato. Abbiamo

$$m \cdot x^k = \mu_{p,p+k}^M(m) \in M_{p+k} = M_{p+k}''$$

e questo conclude la dimostrazione. \square

Proposizione 2.3.8. *Ogni modulo di persistenza algebrico M'' definito su $\mathbb{K}[x_1, \dots, x_d]$ determina un \mathbb{N}^d -modulo di persistenza $\delta(M'') := M : \mathbb{N}^d \rightarrow \mathbf{Vec}$ definito da:*

- per ogni $p \in \mathbb{N}^d$ $M_p := M_p''$;
- per ogni $p, q \in \mathbb{N}^d$ con $p \leq q$ $\mu_{p,q}^M : M_p \rightarrow M_q \mid m \mapsto m \cdot x^{q-p}$

Dimostrazione. Sicuramente si tratta di un'assegnazione $\mathbb{N}^d \rightarrow \mathbf{Vec}$ ben posta. Verifichiamo le proprietà functoriali.

Siano $p \in \mathbb{N}^d$ e $m \in M_p$. Allora $\mu_{p,p}^M(m) = m \cdot x^{p-p} = m$. Segue che per ogni $p \in \mathbb{N}^d$ $\mu_{p,p}^M = id_{M_p}$.

Siano $p, q, r \in \mathbb{N}^d$ con $p \leq r \leq q$ e $m \in M_p$. Allora

$$\begin{aligned} \mu_{p,q}^M(m) &= m \cdot x^{q-p} = m \cdot x^{(q-r)+(r-p)} = \\ &= (m \cdot x^{r-p}) \cdot x^{q-r} = \mu_{p,r}^M(m) \cdot x^{q-r} = \\ &= \mu_{r,q}^M(\mu_{p,r}^M(m)) = (\mu_{r,q}^M \circ \mu_{p,r}^M)(m). \end{aligned}$$

Segue che per ogni $p, q, r \in \mathbb{N}^d$ con $p \leq r \leq q$ abbiamo $\mu_{p,q}^M = \mu_{r,q}^M \circ \mu_{p,r}^M$. \square

Proposizione 2.3.9. *Per ogni \mathbb{N}^d -modulo di persistenza M e per ogni modulo di persistenza algebrico M'' definito su $\mathbb{K}[x_1, \dots, x_d]$ abbiamo $\delta(\gamma(M)) = M$ e $\gamma(\delta(M'')) = M''$.*

Dimostrazione. La tesi segue per costruzione di $\gamma(M)$ e di $\delta(M'')$. \square

In questo capitolo abbiamo visto le definizioni di modulo di persistenza e le reciproche relazioni. Si tratta di due definizioni molto astratte e che non lasciano trasparire il motivo per cui si chiamino così: perché parliamo di *persistenza*? Per rispondere a questa domanda, ci occorre parlare di omologia persistente, che sarà l'oggetto centrale del prossimo capitolo.

Capitolo 3

Omologia persistente

La geometria, con cui almeno ad un livello elementare ci confrontiamo fin da bambini e di cui perciò abbiamo una visione intuitiva e istintiva, è incentrata sulle caratteristiche degli oggetti che cambiano nel momento in cui li deformiamo (si pensi ad esempio alle aree e ai volumi); al contrario, la topologia è incentrata sulle caratteristiche degli oggetti che *non* cambiano nel momento in cui li deformiamo (a patto di farlo in modo continuo). Per studiare tali caratteristiche spesso risulta utile servirsi di strumenti algebrici: è così che nasce la *topologia algebrica*.

Uno degli strumenti fondamentali e più studiati della topologia algebrica è costituito dall'**omologia**: i gruppi di omologia, che talvolta sono in realtà spazi vettoriali, sono collezioni di gruppi associate ad uno spazio topologico ed indicizzate da un parametro intero p . Intuitivamente, il p -esimo gruppo di omologia di uno spazio topologico permette di contare il numero di cavità di dimensione p che quello spazio presenta.

L'**omologia persistente** permette di compiere un ulteriore passo: essa consente di tenere traccia delle variazioni omologiche che avvengono in uno spazio topologico nel momento in cui lo “sezioniamo”, permette cioè di individuare quanto *persistono* le caratteristiche omologiche dello spazio muovendosi tra le sue “sezioni”.

D'altra parte, l'omologia persistente ci consente, come preannunciato nel Capitolo 2, di associare moduli di persistenza a funzioni.

In particolare, distingueremo due casi:

- caso monoparametrico: costruiamo moduli di persistenza a partire da funzioni a valori in \mathbb{R} ;
- caso multiparametrico: costruiamo moduli di persistenza a partire da funzioni a valori in \mathbb{R}^d con $d > 1$.

Per ciascun caso, vedremo invarianti e descrittori; sarà grazie ad essi che potremo confrontare tra loro moduli di persistenza.

3.1 Definizione e prime proprietà

In questa sezione completiamo lo schema riportato in Figura 3.4: nel Capitolo 2 abbiamo visto la definizione algebrica e funtoriale, mentre ora studieremo la definizione topologica.

Siano X uno spazio topologico e $\Phi : X \rightarrow \mathbb{R}^d$ una funzione continua. Per ogni $u = (u_1, \dots, u_d) \in \mathbb{R}^d$, denotiamo con X_Φ^u l'insieme di sottolivello di f rispetto ad u , cioè il sottoinsieme

$$X_\Phi^u := \{x \in X \mid \Phi(x) \leq u\} \subseteq X, \quad (3.1)$$

dotato della topologia di sottospazio, dove con \leq indichiamo la relazione d'ordine definita da (2.1), data dal confronto coordinata per coordinata. Chiaramente, comunque scelti $u, v \in \mathbb{R}^d$ tali che $u \leq v$, si ha che $X_\Phi^u \subseteq X_\Phi^v$ e questa osservazione ci conduce alla definizione seguente.

Definizione 3.1.1. *La filtrazione di X tramite Φ è la sequenza degli spazi topologici X_Φ^u al variare di u in \mathbb{R}^d , dove i morfismi sono dati dalle mappe di inclusione $X_u \hookrightarrow X_v$.*

Chiamiamo Φ una **funzione filtrante** per X .

Assunzione 3.1.2. *D'ora in avanti lavoreremo assumendo che:*

- X sia compatto;
- i valori critici di X rispetto a Φ , ovvero i valori di \mathbb{R}^d in corrispondenza dei quali si hanno dei cambiamenti nei gruppi di omologia dei rispettivi insiemi di sottolivello¹, siano in numero finito.
- l' \mathbb{R}^d -modulo a cui darà origine sia di tipo finito e, equivalentemente, il corrispondente $\mathbb{F}[x_1, \dots, x_d]$ -modulo di persistenza sia finitamente generato.

Tali condizioni non sono limitanti: si tratta di caratteristiche proprie dei problemi provenienti dai contesti applicativi.

Per poter in qualche modo visualizzare una filtrazione, occorre scegliere un numero finito di valori del parametro u . Solitamente, la scelta ricade sui *valori critici* appena menzionati $a_1, \dots, a_n \in \mathbb{R}^d$, ordinati in modo crescente rispetto alla relazione d'ordine 2.1. Tuttavia, al momento non abbiamo ancora gli strumenti per determinarli. Inoltre, alla luce di quanto visto nella Sezione

¹Ne daremo una definizione precisa più avanti.

2.3 e dell'Assunzione 3.1.2, possiamo lavorare senza perdita di generalità con l'insieme $\{1, \dots, n\}^d$ — oppure $\{1, \dots, n_1\} \times \{1, \dots, n_2\} \times \dots \times \{1, \dots, n_d\}$ — ponendo² inoltre $X_{(0, \dots, 0)} := \emptyset$.

Nei capitoli successivi ci concentreremo nei casi in cui $d = 1$ e $d = 2$, perciò vediamo come si configurano le filtrazioni in questi due casi³.

Caso $d = 1$. Abbiamo

$$\emptyset = X_{\Phi}^0 \hookrightarrow X_{\Phi}^1 \hookrightarrow \dots \hookrightarrow X_{\Phi}^n.$$

Osservazione 3.1.3. Osserviamo che la sequenza è “limitata” e “discreta” grazie all'Assunzione 3.1.2; a priori, si tratta di una sequenza “illimitata” e “continua”, ma:

- essendo X compatto e Φ continua, prima di X_{Φ}^1 avremmo una sequenza sì infinita, ma di insiemi vuoti e mappe vuote, mentre dopo $X_{\Phi}^n = X$ avremo una sequenza infinita di mappe identità $X \rightarrow X$, perciò possiamo limitarci a considerare la sequenza tra X_{Φ}^0 e X_{Φ}^n ,
- essendo i valori critici in numero finito, tra X_{Φ}^0 e X_{Φ}^n abbiamo una quantità discreta di sottospazi e relative inclusioni.

Questo vale anche in generale per ogni $d \in \mathbb{R}^d$.

Caso $d = 2$. Abbiamo

$$\begin{array}{ccccccc}
 X_{\Phi}^{(0,n)} & \longrightarrow & X_{\Phi}^{(1,n)} & \longrightarrow & \dots & \longrightarrow & X_{\Phi}^{(n,n)} \\
 \uparrow & & \uparrow & & & & \uparrow \\
 \vdots & & \vdots & & \ddots & & \vdots \\
 \uparrow & & \uparrow & & & & \uparrow \\
 X_{\Phi}^{(0,1)} & \longrightarrow & X_{\Phi}^{(1,1)} & \longrightarrow & \dots & \longrightarrow & X_{\Phi}^{(n,1)} \\
 \uparrow & & \uparrow & & & & \uparrow \\
 \emptyset = X_{\Phi}^{(0,0)} & \longrightarrow & X_{\Phi}^{(1,0)} & \longrightarrow & \dots & \longrightarrow & X_{\Phi}^{(n,0)}.
 \end{array}$$

²Per l'Assunzione 3.1.2, essendo Φ continua essa ammetterà un valore minimo, perciò in ogni caso esiste sempre un valore del parametro per cui il corrispondente insieme di sottolivello sia vuoto.

³I diagrammi sono raffigurati con il minor numero possibile di mappe di inclusione, ma ricordiamo che in realtà abbiamo un'inclusione ogni volta che consideriamo $a_i \leq a_j$. Ognuna si ottiene come composizione di appropriate mappe tra quelle rappresentate.

Vediamo subito un semplice esempio, che sarà fondamentale nel seguito della trattazione, dove sarà più volte ripreso e portato avanti; pertanto, ci riferiremo ad esso come Esempio 1.

Esempio 3.1.4 (Esempio 1). Sia $X := S^1$ la circonferenza unitaria centrata nell'origine e consideriamo le seguenti funzioni continue definite su di essa:

$$\begin{aligned} f : X &\rightarrow \mathbb{R} \mid (x, y) \mapsto x, \\ g : X &\rightarrow \mathbb{R} \mid (x, y) \mapsto y, \\ \Phi &:= (f, g) : X \rightarrow \mathbb{R}^2. \end{aligned}$$

Per rendere visualizzabili tali filtrazioni, è necessario discretizzarle, perciò fissiamo i valori $-1, -\frac{\sqrt{2}}{2}, 0, \frac{\sqrt{2}}{2}, 1$ per f e g e tutte le possibili coppie di tali valori per Φ .

Esplicitiamo alcuni insiemi di sottolivello a titolo di esempio, tutti gli altri si trovano in modo analogo tramite (3.1) e saranno direttamente raffigurati in Figura 3.1. Posto $O := (-1, 0)$, come rappresentato in figura, otteniamo:

$$\begin{aligned} X_f^{-1} &= \{(x, y) \in S^1 \mid f(x, y) \leq -1\} = \{(x, y) \in S^1 \mid x \leq -1\} = \{O\}; \\ X_g^{\frac{\sqrt{2}}{2}} &= \{(x, y) \in S^1 \mid g(x, y) \leq \frac{\sqrt{2}}{2}\} = \{(x, y) \in S^1 \mid y \leq \frac{\sqrt{2}}{2}\}; \\ X_\Phi^{(-1, \frac{\sqrt{2}}{2})} &= \{(x, y) \in S^1 \mid f(x, y) \leq -1 \text{ e } g(x, y) \leq \frac{\sqrt{2}}{2}\} = X_f^{-1} \cap X_g^{\frac{\sqrt{2}}{2}} = \{O\}. \end{aligned}$$

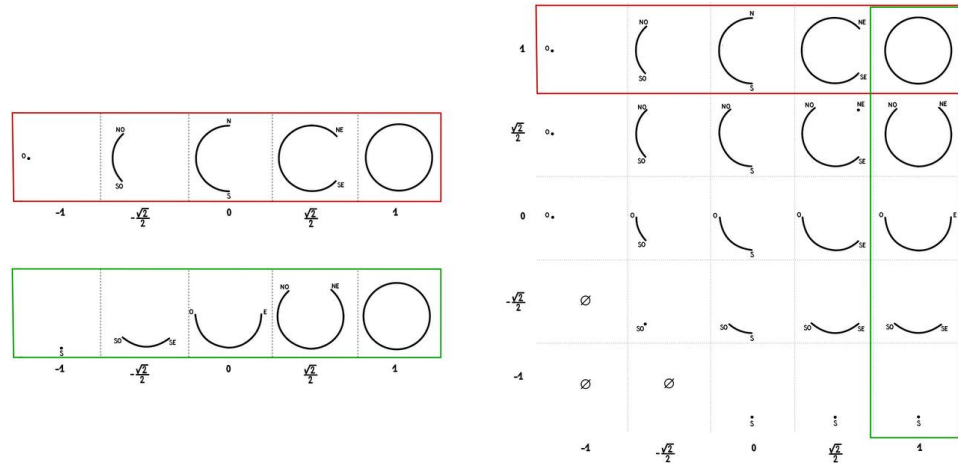


Figura 3.1: A sinistra: filtrazioni indotte da f (in rosso) e da g (in verde). A destra: filtrazione indotta da Φ

Osservazione 3.1.5. Data una filtrazione, le inclusioni coinvolte sono funzioni continue, e possono perciò indurre morfismi in omologia (singolare).

Per ogni $u, v \in \mathbb{R}^d$ con $u \leq v$, denoteremo con $h_p^{u,v}$ il morfismo indotto in omologia dalla mappa di inclusione $X_\Phi^u \hookrightarrow X_\Phi^v$.

Pertanto, scelto un anello per i coefficienti — che non indicheremo per non appesantire la notazione —, è naturale passare da una filtrazione ad una sequenza in omologia, e questo ci porta alla seguente definizione.

Definizione 3.1.6. Per ogni grado $p \in \mathbb{Z}$, il **modulo di omologia persistente** di grado p è la sequenza in omologia

$$PH(\Phi) := (\{H_p(X_\Phi^u)\}_{u \in \mathbb{R}^d}, \{h_p^{u,v}\}_{u \leq v}).$$

Nuovamente, ci soffermiamo sui casi $d = 1$ e $d = 2$ ⁴:

Caso $d = 1$ Abbiamo

$$PH(\Phi) : \mathbf{0} = H_p(X_\Phi^0) \longrightarrow H_p(X_\Phi^1) \longrightarrow \dots \longrightarrow H_p(X_\Phi^n).$$

Caso $d = 2$ Abbiamo

$$\begin{array}{ccccccc} H_p(X_\Phi^{(0,n)}) & \longrightarrow & H_p(X_\Phi^{(1,n)}) & \longrightarrow & \dots & \longrightarrow & H_p(X_\Phi^{(n,n)}) \\ \uparrow & & \uparrow & & & & \uparrow \\ \vdots & & \vdots & & \ddots & & \vdots \\ \uparrow & & \uparrow & & & & \uparrow \\ H_p(X_\Phi^{(0,1)}) & \longrightarrow & H_p(X_\Phi^{(1,1)}) & \longrightarrow & \dots & \longrightarrow & H_p(X_\Phi^{(n,1)}) \\ \uparrow & & \uparrow & & & & \uparrow \\ \mathbf{0} = H_p(X_\Phi^{(0,0)}) & \longrightarrow & H_p(X_\Phi^{(1,0)}) & \longrightarrow & \dots & \longrightarrow & H_p(X_\Phi^{(n,0)}) \end{array}$$

Nel caso in cui $d = 1$, parliamo di *modulo di omologia persistente mono-parametrico*, altrimenti di *modulo di omologia persistente multiparametrico*.

Come anticipato nell'Osservazione 1.1.15, noi lavoreremo con l'omologia a coefficienti in un campo (tipicamente $\mathbb{Z}/2\mathbb{Z}$), che omettiamo nella notazione utilizzata; tale scelta è giustificata dal Teorema 1.1.13. Ricordiamo che, in tal caso, i gruppi di omologia coinvolti risultano in realtà degli spazi vettoriali sul campo fissato. In ogni caso, vedremo che in generale la scelta dei coefficienti in un campo, anziché in un semplice anello, porta ulteriori vantaggi, tra cui ad esempio la possibilità di ottenere, in alcuni casi, un teorema di struttura per i moduli di persistenza.

⁴dove, per lo stesso discorso portato avanti in precedenza, non riportiamo tutte le mappe esistenti tra gli oggetti coinvolti nella sequenza.

Osservazione 3.1.7. *Il modulo di omologia persistente appena definito non è altro che un \mathbb{R}^d -modulo di persistenza come da Definizione 2.1.1 (di tipo finito): si verifica che l'assegnazione che ad ogni $u \in \mathbb{R}^d$ associa $H_p(X_{\Phi}^u)$ e ad ogni $u \leq v$ associa $h_p^{u,v} : H_p(X_{\Phi}^u) \rightarrow H_p(X_{\Phi}^v)$ è un funtore $\mathbb{R}^d \rightarrow \mathbf{Vec}$.*

Esempio 3.1.8 (Esempio 1). *Portiamo avanti l'Esempio 3.1.4. Per calcolare l'omologia, lavoriamo a coefficienti in $\mathbb{Z}/2\mathbb{Z}$; otterremo dei $(\mathbb{Z}/2\mathbb{Z})$ -spazi vettoriali di cui ci limiteremo ad indicarne solo le dimensioni.*

Gli unici gradi in cui l'omologia di S^1 non è nulla sono $p = 0$ e $p = 1$; pertanto anche l'omologia delle filtrazioni considerate è non nulla al più solo per $p = 0$ e $p = 1$.

L'omologia di grado 0 “conta” il numero di componenti connesse; più precisamente la dimensione di un $(\mathbb{Z}/2\mathbb{Z})$ -spazio vettoriale di omologia di grado 0 è uguale al numero di componenti connesse (gli 0-cicli che non sono 0-bordi) dello spazio considerato. In modo analogo, l'omologia in grado 1 “conta” il numero di 1-cicli che non sono 1-bordi. Per una spiegazione più dettagliata, rimandiamo agli argomenti affrontati nel Capitolo 1. In conclusione, arriviamo alle tabelle riportate in Figura 3.2.

Abbiamo visto come sia possibile parlare di moduli di persistenza sia adottando un punto di vista astratto-algebrico che concreto-topologico e ne abbiamo studiato, in entrambi i casi, le prime proprietà. Tuttavia, non è ancora chiaro come mai parliamo proprio di *persistenza*. La definizione di modulo di omologia persistente ci può già fornire un'intuizione del motivo dell'utilizzo di tale termine: intuitivamente, guardando i moduli di omologia riusciamo a tenere traccia di quando hanno origine e cessano di esistere classi di omologia nel corso della filtrazione considerata, e dunque quanto una classe di omologia *persista* nella filtrazione.

Idea. Per esempio, consideriamo la filtrazione su $X := S^1$ riportata in Figura 3.3. Osservando la filtrazione, possiamo individuare una componente connessa — cioè una classe di omologia di grado 0 — che compare per la prima volta in $X^{-\frac{\sqrt{2}}{2}}$ e persiste fino alla fine della filtrazione e una componente connessa che compare in $X^{\frac{\sqrt{2}}{2}}$ e scompare entrando in X^1 . In altre parole, abbiamo:

- una componente connessa che “nasce” all'istante $-\frac{\sqrt{2}}{2}$ e non “muore” mai, cioè di persistenza infinita;
- una componente connessa che “nasce” all'istante $\frac{\sqrt{2}}{2}$ e “muore” all'istante 1.

Renderemo il tutto più formale in seguito, ma già a questo livello di analisi possiamo apprezzare la natura della persistenza ed iniziare a renderci conto delle problematiche che dovremo affrontare nella sua formalizzazione. Inoltre, vedremo che esistono vari descrittori ed invarianti della persistenza

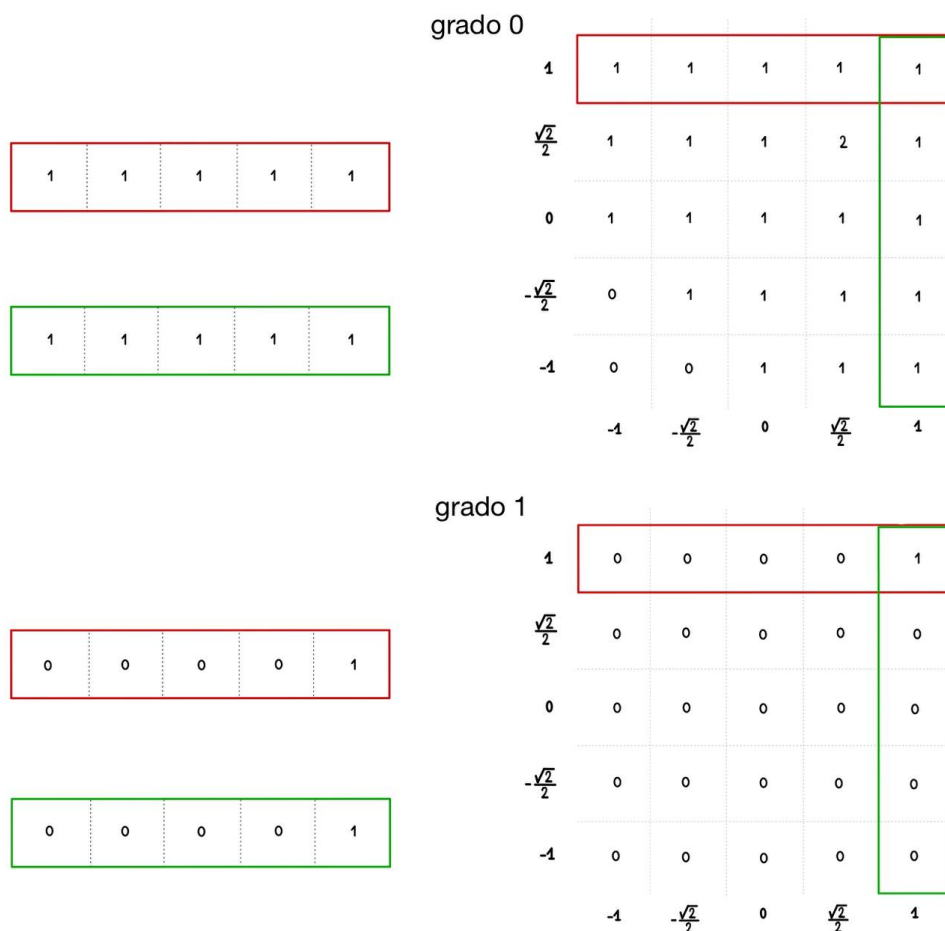


Figura 3.2: Dimensioni dei gruppi di omologia (spazi vettoriali su $\mathbb{Z}/2\mathbb{Z}$) che costituiscono il modulo di omologia persistente delle funzioni dell'Esempio 1: a sinistra f (in rosso) e g (in verde), a destra Φ

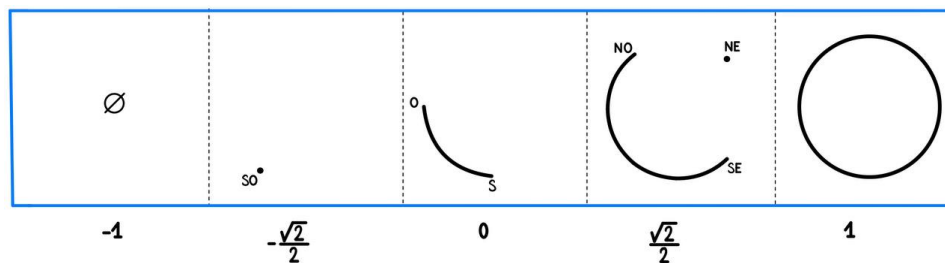


Figura 3.3: Una filtrazione su S^1 .

che presentano tuttavia differenze sostanziali tra caso monoparametrico e multiparametrico. Per questi motivi, analizzeremo i due casi separatamente.

Prima di passare oltre, per concludere questa prima parte del capitolo, riassumiamo tutte le definizioni viste finora e le relative corrispondenze, presentate nei Capitoli 2 e 3, nello schema riportato in Figura 3.4.

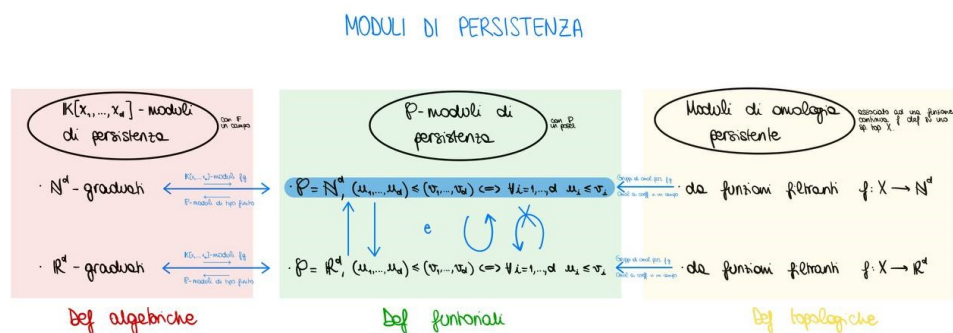


Figura 3.4: Definizioni di modulo di persistenza e loro relazioni.

3.2 Caso monoparametrico

In questa sezione ci occuperemo di persistenza monoparametrica, cioè lavoreremo con $d = 1$. In particolare, lavoreremo con funzioni filtranti f a valori in \mathbb{R} .

Innanzitutto, formalizzeremo l’Idea presentata nella Sezione 3.1, e questo mostrerà che i moduli di omologia persistente descrivono classi di omologia che “nascono” e “muoiono” all’interno della filtrazione considerata; tali informazioni sono codificate nelle cosiddette *coppie di persistenza*. Inoltre mostreremo che, viceversa, i moduli di omologia persistente monoparametrici sono determinati proprio dalle coppie di persistenza, grazie al *teorema di struttura*. Avendo a questo punto una caratterizzazione completa dei moduli di omologia persistente, ci occuperemo dei seguenti aspetti:

- Come visualizzare moduli di persistenza?
- Come confrontare moduli di persistenza?
- Gli oggetti introdotti sono stabili al variare della funzione filtrante?

3.2.1 Invarianti e descrittori

Alla base dei numerosi invarianti e descrittori dell'omologia persistente monoparametrica ci sono due nozioni: i *numeri di Betti persistenti* e il concetto di *molteplicità*. A partire da questi definiremo:

- le *coppie di persistenza*, coinvolti nel *teorema di struttura per moduli di persistenza monoparametrici* e alla base di vari descrittori;
- i *diagrammi di persistenza*, un tipo particolare di descrittori, grazie ai quali potremo definire una (pseudo)metrica tra moduli di persistenza monoparametrici: la *distanza di bottleneck*.

Inoltre, grazie al teorema di struttura si dimostra che in realtà tutti gli invarianti per i moduli di omologia persistente monoparametrici sono tra loro equivalenti.

Teorema di struttura

Dato che stiamo lavorando a coefficienti in un campo \mathbb{K} , sappiamo che $\mathbb{K}[x]$ è un PID e pertanto ricadiamo nelle ipotesi del Teorema 1.2.3. Quello che si vede è che se M è un modulo di omologia persistente allora la sua decomposizione della forma 1.3 è totalmente determinata dalle coppie di persistenza, cioè abbiamo il seguente teorema.

Teorema 3.2.1 (Teorema di Struttura per moduli di omologia persistente (di tipo finito)). *Ogni modulo di omologia persistente su $\mathbb{K}[x]$ si decompone con unicità nel modo seguente:*

$$M \cong \left(\bigoplus_{i=1}^N \mathbb{K}[x](-a_i) \right) \oplus \left(\bigoplus_{j=1}^M \mathbb{K}[x]/(x^{(b_j-c_j)}) \right)$$

con $a_i, b_j, c_j \in \mathbb{R}$.

Una dimostrazione si può trovare in [8].

Come anticipato, il Teorema 3.2.1 è alla base della completezza di tutti gli invarianti del caso monoparametrico e li rende tutti equivalenti. Questo perché, grazie alla corrispondenza γ stabilita nel Capitolo 2 tra \mathcal{P} -moduli di persistenza e moduli di persistenza algebrici, come vedremo, i valori che permettono di determinare la decomposizione descritta nel Teorema 3.2.1 sono le coppie di persistenza, che sono alla base delle definizioni dei vari invarianti e descrittori dei moduli di persistenza monoparametrici. Le coppie di persistenza e gli altri invarianti si definiscono a partire dai *numeri di Betti persistenti* e il concetto di *molteplicità*, che descriveremo nella prossima sezione.

Numeri di Betti persistenti e molteplicità

Siano X uno spazio topologico che soddisfi l'Assunzione 3.1.2, $\Phi : X \rightarrow \mathbb{R}^d$ una funzione continua, con $d > 1$, e per ogni $p \in \mathbb{Z}$ e $u, v \in \mathbb{R}^d$ denotiamo con $h_p^{u,v}$ il morfismo indotto in omologia dall'inclusione $X_\Phi^u \hookrightarrow X_\Phi^v$. Sappiamo che se i coefficienti dell'omologia sono in un campo, allora i gruppi di omologia che emergono da una filtrazione sono in realtà spazi vettoriali su quel campo e le mappe tra di essi sono applicazioni lineari. Dato che noi stiamo lavorando con i coefficienti in un campo, segue che la seguente definizione è ben posta.

Definizione 3.2.2. *Definiamo $H_p^{u,v} = \text{im } h_p^{u,v}$ e lo chiamiamo un p -esimo gruppo di omologia persistente.*

Definiamo inoltre $\beta_p^f(u, v) := \dim H_p^{u,v}$ un p -esimo numero di Betti persistente.

*Ponendo $\Delta^+ := \{(u, v) \in \mathbb{R} \times \mathbb{R} \mid u < v\}$, otteniamo una collezione di funzioni $\beta_p^f : \Delta^+ \rightarrow \mathbb{N}$ associate a f , chiamate **funzioni dei numeri di Betti persistenti (PBNF)**.*

Osservazione 3.2.3. *Dati $u, v \in \mathbb{R}$, si ha che $H_p^{u,v}$ è il quoziente tra $Z_p(X_f^u)$ e $B_p(X_f^v) \cap Z_p(X_f^u)$.*

Per garantire alcuni risultati che avremo modo di trattare nel corso del capitolo, abbiamo bisogno della seguente Assunzione.

Assunzione 3.2.4. *Assumiamo che le PBNF siano continue a destra.*

Alla base di alcuni descrittori dell'omologia persistente, che avremo modo di vedere nel corso di questo capitolo, risiede il concetto di molteplicità dei punti in Δ^+ e in $\{(u, \infty) \mid u \in \mathbb{R}\}$, che si basa sui numeri di Betti persistenti appena definiti. Essa inoltre permette di individuare i valori critici.

Definizione 3.2.5. *La molteplicità di $(u, v) \in \Delta^+$ è il numero finito non negativo $\mu_f(u, v)$ definito da*

$$\min_{\substack{\varepsilon > 0 \\ v - \varepsilon > u + \varepsilon}} \{(\beta_f^p(u + \varepsilon, v - \varepsilon) - \beta_f^p(u - \varepsilon, v - \varepsilon)) + \\ -(\beta_f^p(u + \varepsilon, v + \varepsilon) - \beta_f^p(u - \varepsilon, v + \varepsilon))\}.$$

La molteplicità di (u, ∞) con $u \in \mathbb{R}$ è il numero finito non negativo $\mu_f(u, \infty)$ definito da

$$\min_{\substack{\varepsilon > 0 \\ v > u + \varepsilon}} \beta_f^p(u + \varepsilon, v) - \beta_f^p(u - \varepsilon, v).$$

Intuitivamente, tale quantità permette di contare il numero di classi di omologia di grado p che “nascono” in X_f^u e “muoiono” in X_f^v nel primo caso, mentre non muoiono mai nel secondo caso. I *valori critici* sono i valori $a_1, \dots, a_n \in \mathbb{R}$ in corrispondenza dei quali la molteplicità passa da essere nulla ad essere positiva, ovvero sono i valori che compaiono in almeno una coppia (u, v) con molteplicità positiva. Più precisamente, abbiamo la seguente definizione.

Definizione 3.2.6. *Diciamo che $u \in \mathbb{R}$ è un **valore critico** di f se per ogni $\varepsilon > 0$ esistono $u', u'' \in \mathbb{R}$ tali che $u' < u < u''$, $\|u' - u\|_\infty, \|u'' - u\|_\infty \leq \varepsilon$ e l'omomorfismo $h_p^{u', u''}$ indotto in omologia dall'inclusione $X_f^{u'} \hookrightarrow X_f^{u''}$ non sia un isomorfismo.*

Inoltre, si dimostra il seguente teorema.

Teorema 3.2.7. *Sia $p \in \Delta^+ \cup \{(u, \infty) \mid u \in \mathbb{R}\}$. Se $\mu_f(p) > 0$ allora:*

- *se $p = (u, v) \in \Delta^+$ allora u e v sono valori critici per f ;*
- *se $p = (u, \infty)$ allora u è un valore critico per f .*

Questo ci conduce direttamente alla prossima sezione: le *coppie di persistenza* non sono altro che coppie di punti critici.

Coppie di persistenza

Procediamo ora nella formalizzazione dell'Idea nella Sezione 3.1, rendendo formalmente più rigorosi i concetti di “nascita” e “morte” di una classe di omologia.

Nella prossima definizione, lavoreremo con i valori critici $a_1, \dots, a_n \in \mathbb{N}$, supponendo senza perdita di generalità che siano $1, \dots, n$; le definizioni che vedremo si potranno generalizzare a valori qualunque grazie alla Definizione 3.2.5, passando da considerare coppie di valori critici (i, j) a considerare coppie di valori $(u, v) \in \Delta^+ \cup \{(u, \infty) \mid u \in \mathbb{R}\}$ con $\mu_f(u, v) > 0$. Inoltre aggiungiamo ai valori considerati il valore $-\infty$, che ci permette di includere nella filtrazione anche l'insieme vuoto, e $n + 1$ che invece “rappresenta” tutti gli $u \in \mathbb{R}$ tali che $u > a_n$, assumendo che $H_p(X_f^{n+1}) = 0$.

Definizione 3.2.8. *Siano $l \in \{1, \dots, n\}$ e $\xi \in H_p(X_f^l)$. Posto*

$$i := \min\{k \leq l \mid \xi \in H_p^{k,l} \text{ e } \xi \notin H_p^{k-1,l}\},$$

diciamo che ξ nasce in X_f^i .

Analogamente, fissato $j \in \{l + 1, \dots, n\}$, diciamo che ξ muore entrando in X_f^j se $h_p^{l,j+1}(\xi) \neq 0$ e $h_p^{l,j}(\xi) = 0$.

*Chiamiamo **coppie di persistenza** le coppie (a_i, a_j) che indicano nascita e morte di una classe di omologia all'interno della filtrazione considerata.*

Osservazione 3.2.9. *Lavorando con i valori critici e le assunzioni descritte poco sopra, la molteplicità assume la forma seguente. Per ogni $i, j = 1, \dots, n+1$ tali che $i < j \leq n+1$ abbiamo*

$$\mu_f(i, j) = (\beta_f^p(i, j-1) - \beta_f^p(i, j)) - (\beta_f^p(i-1, j-1) - \beta_f^p(i-1, j)) \in \mathbb{N}.$$

Risulta ora chiaro che essa esprima il numero di classi indipendenti che nascono in X_f^i e muoiono entrando in X_f^j ; infatti la prima delle differenze esprime il numero di classi indipendenti che nascono al più tardi in X_f^i e muoiono in X_f^j , mentre la terza differenza esprime il numero di classi indipendenti che sono nate al più tardi in X_f^{i-1} e muoiono in X_f^j .

In particolare, $\mu_f(i, n+1)$ conta il numero delle classi indipendenti che nascono in X_f^i e “muoiono entrando in X_f^{n+1} ”, cioè che in realtà non muoiono mai nel corso della filtrazione ed è per enfatizzare tale “longevità” che si utilizza la notazione $a_{n+1} = n+1 = \infty$.

Osservazione 3.2.10. *Anche se stiamo lavorando a coefficienti in un campo e dunque i gruppi di omologia risultano spazi vettoriali, in generale le pesime classi di omologia in $H_p(X_f^{j-1})$ che nascono in X_f^i e muoiono entrando in X_f^{j-1} potrebbero non costituire uno spazio vettoriale. Perciò non si può parlare di dimensione in questo contesto. In effetti, l'introduzione di $\mu_f(i, j)$ sopperisce a questa mancanza e in un certo senso la spiega: $\mu_f(i, j)$ coinvolge somme alternate di dimensioni di spazi vettoriali.*

Una prima caratteristica utile delle coppie di persistenza è che esse consentono di dare una definizione *quantitativa* di persistenza di una classe di omologia.

Definizione 3.2.11. *Per ogni $i, j = 1, \dots, n$ con $i < j$ tale che $\mu_p^{i,j} \neq 0$, per ogni classe di omologia ξ che nasce in X_f^i e muore entrando in X_f^j , definiamo la **persistenza** di ξ la quantità $\text{pers}(\xi) := a_j - a_i$.*

Invece, quando $j = \infty$, allora per ogni $i = 1, \dots, n$ e per ogni classe di omologia ξ che nasce in $X_f^{a_i}$ e muore entrando in $X_f^{a_j}$ — cioè che non muore mai — poniamo $\text{pers}(\xi) := \infty = (n+1) - i$.

Esempio 3.2.12 (Esempio 1). *Riprendiamo l'Esempio 3.1.8. Ci concentriamo sulla filtrazione indotta da f e lavoriamo in grado 0.*

Sappiamo che $\dim(H_0(X_f^1)) = \dim(H_0(S^1)) = 1$, perciò abbiamo un'unica classe in omologia ξ . Quando nasce? Osservando la rappresentazione in Figura 3.1, a livello intuitivo ci accorgiamo subito che nasce in X_f^{-1} . Abbiamo perciò trovato una coppia di persistenza $(-1, \infty)$ in grado 0 per f , e la persistenza della corrispondente classe è ∞ . Osserviamo che per ogni $\varepsilon > 0$ e per ogni $v \in \mathbb{R}$ tale che $v > -1 + \varepsilon$ abbiamo

$$\beta_f^0(-1 + \varepsilon, v) = 1$$

poiché dopo X_f^{-1} nella filtrazione abbiamo sempre sottospazi costituiti da una componente connessa, e abbiamo

$$\beta_f^0(-1 - \varepsilon, v) = 0$$

poiché prima di X_f^{-1} tutti gli insiemi della filtrazione sono vuoti. Pertanto,

$$\mu_f(-1, \infty) = \min_{\substack{\varepsilon > 0 \\ v > -1 + \varepsilon}} \beta_f^p(-1 + \varepsilon, v) - \beta_f(-1 - \varepsilon, v) = 1.$$

Le coppie di persistenza permettono di riassumere sia le informazioni algebriche che quelle topologiche di un modulo di omologia persistente. Esse infatti consentono di ottenere la decomposizione dei moduli di omologia persistente e di tenere traccia dei cambiamenti in omologia all'interno di una filtrazione. Più precisamente:

- dal punto di vista algebrico, esse entrano in gioco nel teorema di struttura, in quanto sono proprio i valori che determinano la decomposizione dei moduli di persistenza come descritto nel Teorema 3.2.1: le coppie del tipo (a_i, ∞) determinano la componente libera $\bigoplus_{i=1}^N \mathbb{K}[x](-a_i)$, mentre le coppie del tipo (a_i, a_j) determinano la componente di torsione $\bigoplus_{l=1}^M \mathbb{K}[x]/(x^{(b_l - c_l)})$;
- dal punto di vista topologico, esse permettono di definire numerosi descrittori topologici, tra cui i *diagrammi di persistenza*, che studieremo nella prossima sezione.

Diagrammi di persistenza

Finora abbiamo visto che un modulo di omologia persistente descrive classi di omologia che nascono e muoiono nel corso della filtrazione e che esso è totalmente determinato da tali informazioni. In particolare, le coppie di persistenza ci hanno permesso di trovare una caratterizzazione algebrica dei moduli di omologia persistente. In realtà esse ci consentono anche di definire descrittori topologici di tali moduli, grazie ai quali riusciamo anche a “visualizzarli”. I principali descrittori sono i *diagrammi di persistenza*, i *barcode*, i *landscape* e le *persistence images*. Noi analizzeremo i *diagrammi di persistenza*.

Definizione 3.2.13. Sia $p \in \mathbb{Z}$ un grado di omologia. Il **diagramma di persistenza** di grado p di f è il multiset dato dall'unione $Dgm_p(f)$ tra:

- l'insieme dei punti $(a_i, a_j) \in \Delta^+$, cioè con $i < j$, di molteplicità $\mu_p^{i,j}$ non nulla, contati con la propria molteplicità;
- l'insieme dei punti della diagonale $\{(u, u) \mid u \in \mathbb{R}\}$ contati con molteplicità infinita⁵.

⁵Questa condizione sarà utile in seguito quando parleremo di matching tra diagrammi di persistenza.

In particolare, i punti (a_i, ∞) rappresentanti classi che non muoiono mai vengono chiamati *punti essenziali di persistenza* e le corrispondenti classi *classi di omologia essenziali*.

A meno che non si renda necessario indicare esplicitamente a quale grado p ci stiamo riferendo, scriveremo semplicemente $Dgm(f)$.

Esempio 3.2.14 (Esempio 1). *Portiamo avanti l'Esempio 3.2.12. Abbiamo individuato una coppia di persistenza di grado 0 per $f: (-1, \infty)$, di molteplicità 1. In realtà, essa è l'unica coppia $(u, v) \in \Delta^+ \cup \{(u, \infty)\}$ con molteplicità non nulla, perciò il diagramma di persistenza in grado 0 associato a f è semplicemente quello rappresentato in Figura 3.5.*

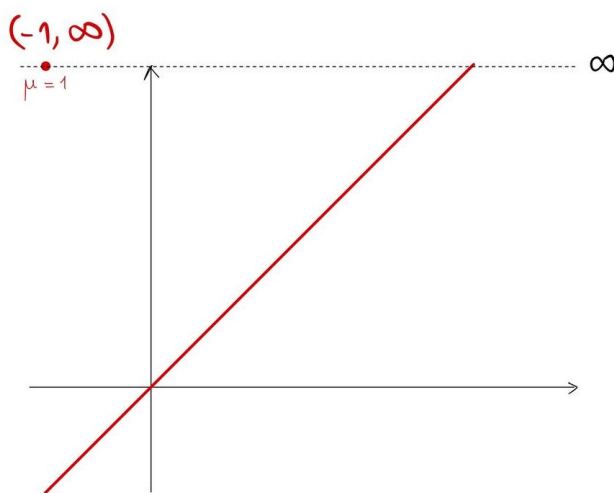


Figura 3.5: Diagramma di persistenza di grado 0 associato a f nell'Esempio 1.

Osservazione 3.2.15. *Osserviamo che*

- se una classe di omologia ha persistenza s allora il punto che la rappresenta in $Dgm(f)$ avrà distanza $\frac{s}{\sqrt{2}}$ dalla diagonale;
- i punti di $Dgm(f)$ si trovano per costruzione sopra o al di sopra della diagonale, poiché $a_i \leq a_j$ per ogni $i < j$.

3.2.2 Distanza di bottleneck e stabilità

Ora che abbiamo un modo per rappresentare i moduli di persistenza, i diagrammi di persistenza presentati nella sezione precedente, viene naturale domandarsi come confrontarli tra loro. Esistono varie possibili (pseudo)metriche tra moduli di persistenza, tra cui la *distanza interleaving* e di *Wasserstein*. Noi invece vedremo la *distanza di bottleneck*, che è una distanza

tra diagrammi di persistenza, mentre una pseudometrica tra moduli di persistenza. Essa si basa sul concetto di *matching* tra diagrammi di persistenza. Siano f e g due funzioni filtranti su X a valori reali.

Definizione 3.2.16. Per ogni $p \in \mathbb{Z}$ fissato, un **matching** tra $Dgm_p(f)$ e $Dgm_p(g)$ è una qualsiasi biezione tra di essi, tenendo in considerazione le molteplicità dei punti dei due diagrammi.

Osservazione 3.2.17. È sempre possibile stabilire biezioni tra i diagrammi di persistenza per definizione: i punti della diagonale vengono contati con molteplicità infinita.

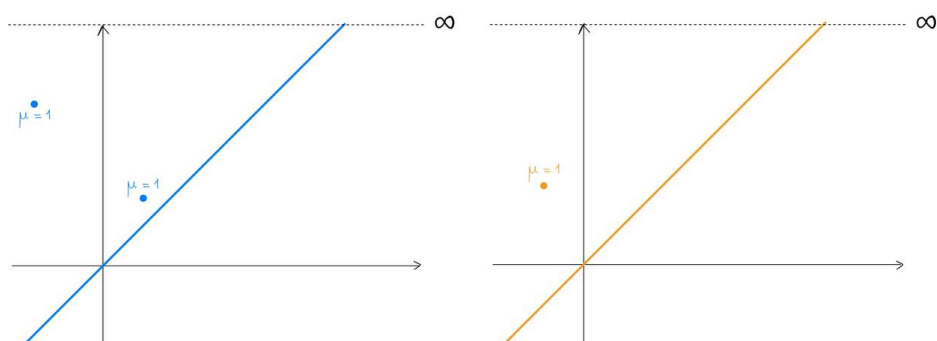


Figura 3.6: Due possibili diagrammi di persistenza.

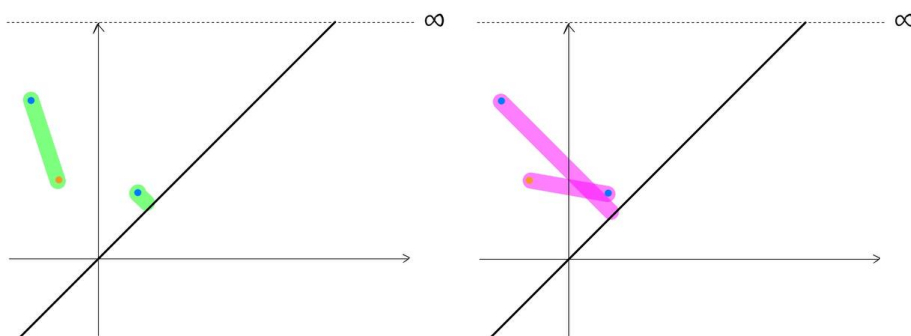


Figura 3.7: Possibili matching tra i due diagrammi di persistenza rappresentati in Figura 3.6.

Esempio 3.2.18. Consideriamo i diagrammi di persistenza rappresentati in Figura 3.6. I possibili matching tra i due sono quelli rappresentati in Figura 3.7. In particolare, in questo caso si annuncia necessario l'utilizzo della diagonale.

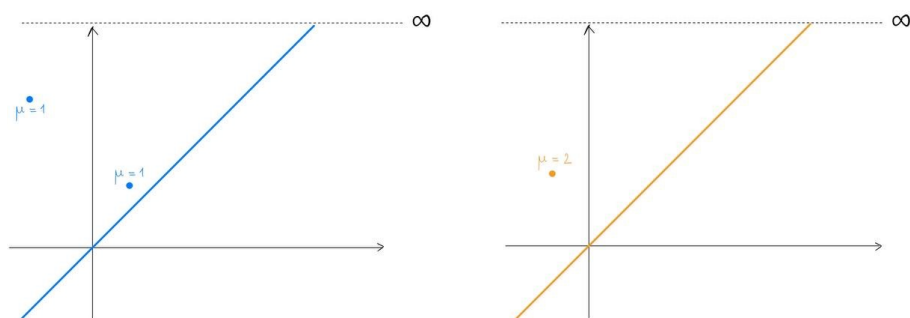


Figura 3.8: Due possibili diagrammi di persistenza.

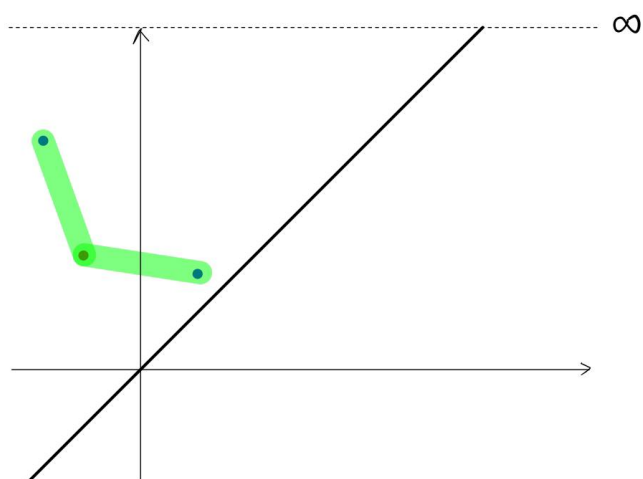


Figura 3.9: Possibili matching tra i due diagrammi di persistenza rappresentati in Figura 3.8.

Se invece ad esempio il punto non diagonale del secondo diagramma (quello giallo) avesse avuto molteplicità 2, come mostrato in Figura 3.8, sarebbe stato possibile individuare anche un matching che non coinvolgesse alcun punto della diagonale, rappresentato in Figura 3.9.

A questo punto l'idea che ci condurrà alla definizione della distanza di bottleneck è molto semplice e naturale: fissato un matching tra $Dgm(f)$ e $Dgm(g)$ si definisce il suo costo come il massimo delle distanze ($\|\cdot\|_\infty$) tra le coppie di punti individuate dal matching — con qualche accorgimento per i punti essenziali, e poi si considera il minimo costo al variare dei possibili matching. Da questa idea di “imbuto” viene il nome “bottleneck”.

Definizione 3.2.19. Sia $\varphi : Dgm(f) \rightarrow Dgm(g)$ un matching. Definiamo il **costo** di φ come segue

$$cost(\varphi) := \max\{d_\infty(\alpha, \varphi(\alpha)) \mid \alpha \in Dgm(f)\},$$

dove per ogni $x = (x_1, x_2), y = (y_1, y_2) \in \overline{\mathbb{R}^2}$ si ha

$$d_\infty(x, y) := \begin{cases} \|x - y\|_\infty & \text{se } x_1, x_2, y_1, y_2 \in \mathbb{R}^2 \\ |x_1 - y_1| & \text{se } x_2 = \infty \text{ e } y_2 = \infty \\ \infty & \text{se } x_2 = \infty \text{ e } y_2 \neq \infty \text{ oppure } x_2 \neq \infty \text{ e } y_2 = \infty. \end{cases}$$

Indichiamo con \mathcal{M} l'insieme di tutti i possibili matching tra $Dgm(f)$ e $Dgm(g)$.

Definizione 3.2.20. La **distanza di bottleneck** tra $Dgm(f)$ e $Dgm(g)$ è definita come segue

$$d_B(Dgm(f), Dgm(g)) := \min\{cost(\varphi) \mid \varphi \in \mathcal{M}\}.$$

È chiaro che d_B risulti una distanza. Ciò non rimane vero nel momento in cui si definisce $d_B(PH(f), PH(g)) := d_B(Dgm(f), Dgm(g))$; in particolare esistono moduli di omologia persistente diversi con distanza di bottleneck nulla. La distanza di bottleneck tra moduli di persistenza non è una metrica ma una pseudometrica.

Osservazione 3.2.21. Nelle due definizioni precedenti, il minimo e il massimo considerati sono ben posti grazie all'Assunzione 3.1.2. Lavorando invece in generale, vengono definiti rispettivamente come \inf e \sup .

Osservazione 3.2.22. Osserviamo che, per definizione di distanza di bottleneck, includere la diagonale nei diagrammi di persistenza anche quando non sarebbe necessario non solo ci permette di trovare bigezioni anche quando non è possibile mettere in corrispondenza le coppie di persistenza dei due diagrammi, ma ci consente di minimizzare ulteriormente i costi dei matching. Ad esempio, consideriamo i diagrammi di persistenza in Figura 3.10. I pos-

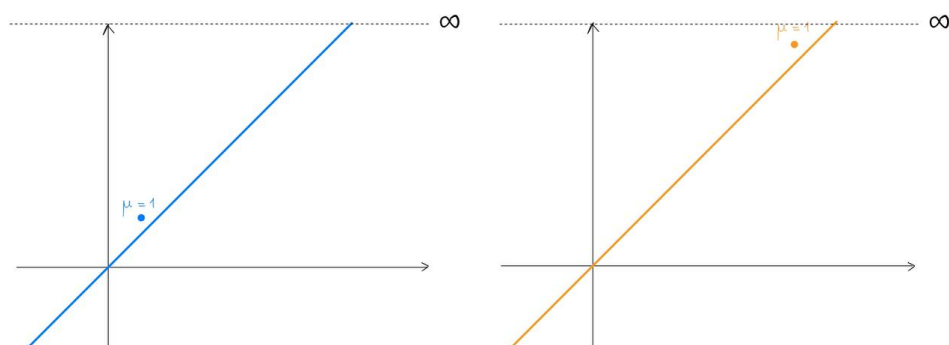


Figura 3.10: Due possibili diagrammi di persistenza.

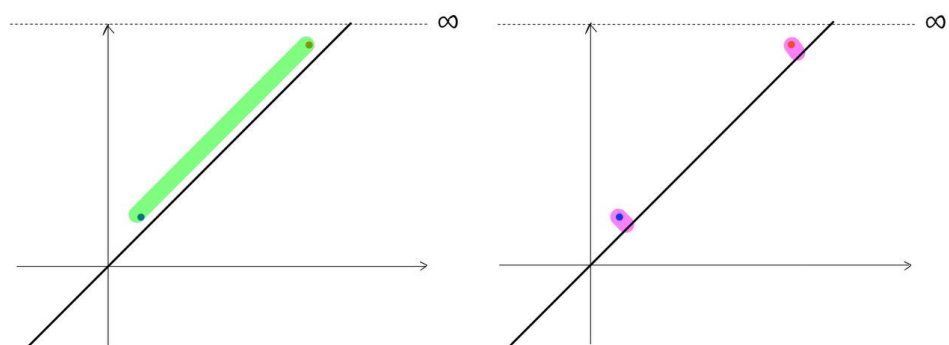


Figura 3.11: Possibili matching tra i due diagrammi di persistenza rappresentati in Figura 3.10.

sibili matching sono quelli rappresentati in Figura 3.11. Se considerassimo solamente il primo dei due matching, quello che non coinvolge la diagonale, avremmo un costo molto più alto rispetto al costo che otteniamo dal secondo matching, che è possibile costruire solamente se si considerano anche due punti sulla diagonale. Dato che la distanza di bottleneck minimizza il costo, in generale ci conviene perciò considerare tutti i possibili matching, anche quelli che coinvolgono punti della diagonale.

Esempio 3.2.23 (Esempio 1). Nell'Esempio 3.2.14 abbiamo trovato il diagramma di persistenza di grado 0 associato ad f . In modo analogo si trova quello di grado 0 associato a g . I possibili matching sono pertanto quelli rappresentati in Figura 3.12 e la distanza di bottleneck è realizzata dal primo. Pertanto essa risulta nulla.

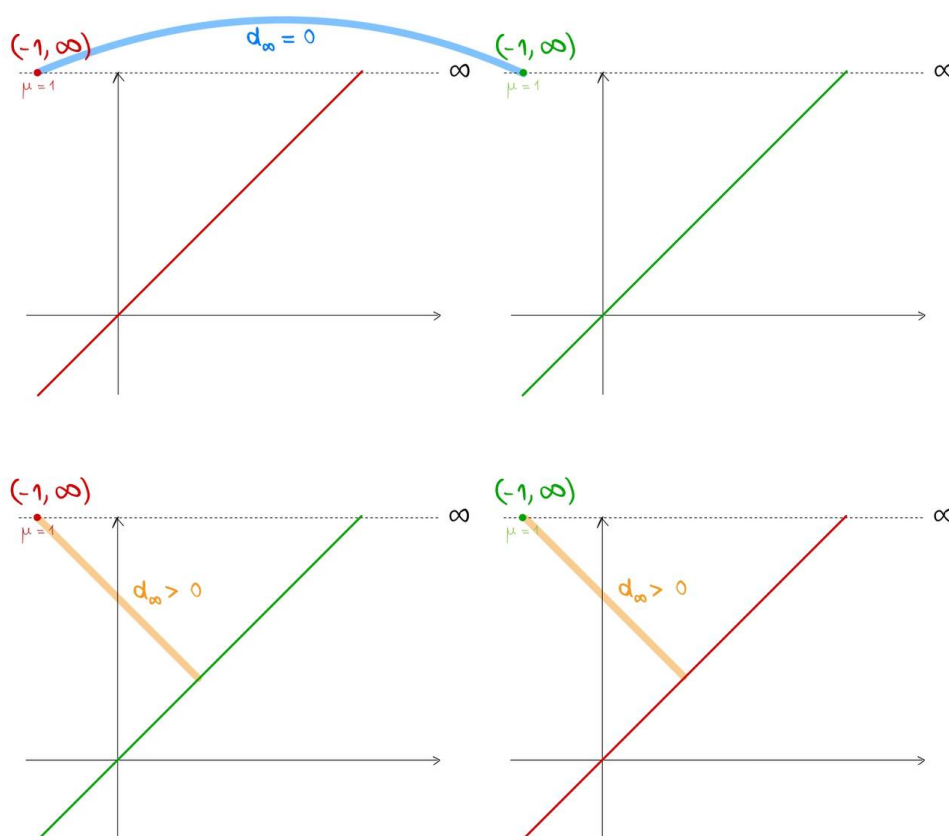


Figura 3.12: Possibili matching tra i diagrammi di persistenza di f e g dell'Esempio 1.

Stabilità della distanza di bottleneck

Per far sì che la distanza di bottleneck sia utilizzabile nella pratica, è necessario che sia stabile al variare delle funzioni filtranti, cioè vogliamo assicurarci che funzioni “simili” determinino diagrammi “simili”. Effettivamente si dimostra che, sotto certe ipotesi, la distanza di bottleneck è controllata dalla norma infinito tra le funzioni filtranti, come stabilisce il seguente teorema.

Teorema 3.2.24. *Sia X uno spazio topologico triangolabile e siano $f, g : X \rightarrow \mathbb{R}$ regolari⁶. Allora per ogni $p \geq 0$*

$$d_B(Dgm_p(f), Dgm_p(g)) \leq \|f - g\|_\infty.$$

Grazie all’Assunzione 3.1.2, nel nostro caso è sufficiente assumere che X sia triangolabile per garantire la tesi. Per una dimostrazione si può fare riferimento a [7, 6].

Notazione 3.2.25. *In seguito indicheremo $d_B(f, g) := d_B(Dgm(f), Dgm(g))$.*

Grazie al Teorema di Struttura e alla stabilità della distanza di bottleneck, si dimostra che i diagrammi di persistenza sono invarianti completi per moduli di omologia persistente monoparametrica. Inoltre, i diagrammi di persistenza sono invarianti per omotopia per lo spazio topologico sottostante. Si possono trovare dimostrazioni e motivazioni di questi fatti in [13, 24, 12, 21].

3.3 Caso multiparametrico

La teoria della persistenza monoparametrica che abbiamo appena illustrato risulta molto completa e rigorosa. Perché allora estendere le nozioni viste a funzioni a valore in \mathbb{R}^d ? A prescindere dal fatto che la teoria multiparametrica è molto interessante, come avremo modo di vedere, in realtà essa risulta il modo più naturale per formalizzare alcuni problemi di natura applicativa. Ad esempio, se abbiamo un dataset costituito da immagini a colori su scala RGB, risulta poco sensato e poco rappresentativo lavorare con un approccio monoparametrico, mentre risulta molto più naturale lavorare utilizzando tre parametri: quello per quantificare il rosso (R), quello per quantificare il verde (G) e quello per quantificare i blu (B). Pertanto è importante portare avanti ed analizzare anche la persistenza multiparametrica.

D’ora in avanti lavoreremo perciò con $d > 1$.

⁶Per *funzione filtrante regolare* qui intendiamo una funzione $X \rightarrow \mathbb{R}$ continua t.c. i gruppi di omologia degli insiemi di sottolivello hanno rango finito e i valori critici sono in numero finito.

3.3.1 Invarianti e descrittori

Anche nel caso multiparametrico abbiamo invarianti e descrittori che si costruiscono a partire dai *numeri di Betti persistenti multiparametrici* e la *molteplicità*. A partire da tali nozioni potremo definire un concetto analogo a quello di diagramma di persistenza: lo *spazio di persistenza*. Dopodiché ci concentreremo sul caso biparametrico ($d = 2$), dove definiremo una (pseudo)distanza tra moduli di persistenza biparametrici: la *distanza di matching*.

Quello che invece non possiamo generalizzare è il teorema di struttura: nel caso multiparametrico non è possibile avere una decomposizione analoga e univocamente determinata come avviene invece nel caso monoparametrico. Di conseguenza, non esiste nemmeno il concetto analogo a quello di coppie di persistenza.

Nonostante questo possa scoraggiare, avremo modo di vedere che, nonostante questo, la persistenza multiparametrica rimane comunque un valido strumento, che in alcuni casi in realtà è molto più efficace rispetto alla persistenza monoparametrica.

Numeri di Betti di persistenza multiparametrici e molteplicità

A differenza di quelli monoparametrici, i moduli di omologia persistente multiparametrici non ammettono una rappresentazione concisa analoga a quella delle coppie di persistenza; per questo motivo, non ammettono un teorema di struttura. Tuttavia, è possibile comunque generalizzare le idee alla base di tali strutture tramite i *numeri di Betti persistenti multiparametrici* e questo ci permetterà di costruire lo *spazio di persistenza* associato ad una funzione filtrante a valori in \mathbb{R}^d , che costituisce il concetto analogo del diagramma di persistenza associato ad una funzione filtrante a valori in \mathbb{R} .

Siano $\Phi : X \rightarrow \mathbb{R}^2$ una funzione filtrante e $p \in \mathbb{Z}$. Analogamente al caso monoparametrico, possiamo definire i *gruppi di omologia persistente di grado p* .

Definizione 3.3.1. *Siano $u, v \in \mathbb{R}^d$ tali che $u \leq v$ rispetto alla relazione d'ordine definita da 2.1. Chiamiamo l'immagine $H_p^{u,v}$ dell'omomorfismo $H_p(X_\Phi^u) \rightarrow H_p(X_\Phi^v)$ un p -esimo **gruppo di omologia persistente**. Il suo rango $\beta_\Phi^p(u, v)$ è chiamato un p -esimo **numero di Betti persistente multiparametrico**.*

*Ponendo $\Delta_d^+ := \{(u, v) \in \mathbb{R}^d \times \mathbb{R}^d \mid u < v\}$, otteniamo una collezione di funzioni $\beta_\Phi^p : \Delta_d^+ \rightarrow \mathbb{N} \mid (u, v) \mapsto \beta_\Phi^p(u, v)$ associate a Φ , chiamate **funzioni dei numeri di Betti persistenti multiparametrici**, che nuovamente denotiamo con *PBNF*.*

Fino a qui, non ci sono stati problemi nel generalizzare le analoghe definizioni date nel caso monoparametrico. Ripercorrendo quanto visto per quel caso, una volta definiti i numeri di Betti persistenti multiparametrici vorremmo sfruttare tale nozione per definire la molteplicità di una coppia $(u, v) \in \Delta_d^+$, ma già qui sorgono i primi problemi e non ci si riesce in modo così immediato. Questo è dovuto al fatto che, a differenza del caso monoparametrico, in questo caso le PBNF non sono necessariamente continue⁷.

Per ogni $\bar{u} \in \mathbb{R}^d$ poniamo $\mathbb{R}_\pm^d(\bar{u}) := \{u \in \mathbb{R}^d \mid u < \bar{u} \text{ oppure } \bar{u} < u\}$. Si possono dimostrare le seguenti proposizioni.

Proposizione 3.3.2. *Sia $U := (\bar{u}, \bar{v}) \in \Delta_d^+$. Esiste $\varepsilon > 0$ tale che l'insieme*

$$W_\varepsilon(U) := \{(u, v) \in \mathbb{R}_\pm^d(\bar{u}) \times \mathbb{R}_\pm^d(\bar{v}) \mid \|u - \bar{u}\|_\infty, \|v - \bar{v}\|_\infty < \varepsilon\}$$

è un insieme aperto di Δ_d^+ e non contiene alcun punto di discontinuità di β_Φ^p .

Proposizione 3.3.3. *Sia $\bar{u} \in \mathbb{R}^d$. Esiste $\varepsilon > 0$ tale che l'insieme*

$$V_\varepsilon(U) := \left\{ (u, v) \in \mathbb{R}_\pm^d(\bar{u}) \times \mathbb{R}^d \mid \|u - \bar{u}\|_\infty < \varepsilon \text{ e } \forall_{i=1, \dots, d} v_i > \frac{1}{\varepsilon} \right\}$$

è un insieme aperto di Δ_d^+ e non contiene alcun punto di discontinuità di β_Φ^p .

Ora, per ogni $(u, v) \in \Delta_d^+$, $e \in \mathbb{R}^d$ con $e > 0$ e $u + e < v - e$ consideriamo la quantità

$$\begin{aligned} \mu_\Phi^e(u, v) := & \beta_\Phi^p(u + e, v - e) - \beta_\Phi^p(u - e, v - e) + \\ & - \beta_\Phi^p(u + e, v + e) + \beta_\Phi^p(u - e, v + e). \end{aligned}$$

Grazie all'Assunzione 3.1.2, siamo nell'ipotesi di gruppi di omologia finitamente generati, perciò sicuramente $\mu_\Phi^e(u, v) \in \mathbb{Z}$; inoltre, si verifica facilmente che in realtà si tratta di un intero non negativo e che è non decrescente in e . Per la Proposizione 3.3.2, ogni termine nella somma che definisce $\mu_\Phi^e(u, v)$ è costante per ogni $e > 0$ con $\|e\|_\infty$.

Similmente, la quantità

$$\beta_\Phi^p(u + e, v) - \beta_\Phi^p(u - e, v)$$

è un intero non negativo. Inoltre, è non decrescente in e ed è non decrescente in v e per la Proposizione 3.3.3 si ha che ogni termine che la definisce è costante per ogni $e > 0$ con $\|e\|_\infty$ sufficientemente piccolo e ogni $v \in \mathbb{R}^d$ con $v_i > \frac{1}{\|e\|_\infty}$.

Tutto questo giustifica e motiva la seguente definizione.

⁷Anche nel caso monoparametrico non è detto che lo siano, ma sotto le nostre ipotesi di lavoro lo sono.

Definizione 3.3.4. La **molteplicità** di $(u, v) \in \Delta_d^+$ è il seguente numero intero e non negativo:

$$\mu_\Phi(u, v) := \min_{\substack{e > 0 \\ v - e > u + e}} \mu_\Phi^e(u, v).$$

La **molteplicità** di (u, ∞) con $u \in \mathbb{R}^d$ è il seguente numero intero non negativo:

$$\mu_\Phi(u, \infty) := \min_{\substack{e > 0 \\ v > v + e}} \{\beta_\Phi^p(u + e, v) - \beta_\Phi^p(u - e, v)\}.$$

Come nel caso monoparametrico, anche in questo caso i valori critici sono i valori che compaiono in coppie di valori con molteplicità positiva, come mostrato in [4]. Pertanto, nella prossima sezione definiremo lo *spazio di persistenza* in modo analogo a quanto visto per i diagrammi di persistenza.

Spazio di persistenza

Lo *spazio di persistenza* costituisce l'analogo multiparametrico dei diagrammi di persistenza.

Definizione 3.3.5. Lo **spazio di persistenza** di grado p associato a Φ è il multiset $\text{Spc}(\Phi)$ dato dall'unione tra :

- l'insieme di tutti i punti in $\Delta_d^+ \cup \{(u, \infty) \mid u \in \mathbb{R}^d\}$ con molteplicità μ_Φ strettamente positiva, contati con la loro molteplicità
- l'insieme dei punti della diagonale $\{(u, u) \mid u \in \mathbb{R}^d\}$ contati con molteplicità infinita.

Esso può essere pensato come all'analogo multiparametrico dei diagrammi di persistenza del caso monoparametrico non solo perché è costruito in modo analogo, ma anche perché ne condivide alcune proprietà. Per approfondire questi aspetti rimandiamo a [4].

Tuttavia, a differenza dei diagrammi di persistenza, gli spazi di persistenza non sono invarianti per moduli di omologia persistente.

3.3.2 Distanza di matching e stabilità

D'ora in avanti nel corso di questo capitolo lavoreremo con funzioni filtranti $X \rightarrow \mathbb{R}^2$, per studiare nel dettaglio il caso biparametrico. Questo perché nel caso biparametrico è possibile "codificare" tutte le informazioni di una funzione $\Phi : X \rightarrow \mathbb{R}^2$ tramite una collezione $\{\Phi_{(a,b)}\}_{(a,b) \in [0,1] \times \mathbb{R}}$ di funzioni a valori reali e dunque è possibile, in particolare, definire una (pseudo)distanza tra moduli di persistenza biparametrici riconducendosi alla distanza di bottleneck. Questa procedura prende il nome di *metodo di foliazione* ed ora la

illustriamo nel dettaglio.

Consideriamo Λ^+ l'insieme di tutte le rette contenute in \mathbb{R}^2 a pendenza (strettamente) positiva. Possiamo parametrizzare tale spazio su $(0, 1) \times \mathbb{R}$ come segue: ad ogni retta r in Λ^+ è associata un'unica coppia $(a, b) \in (0, 1) \times \mathbb{R}$ con

- $a \in (0, 1)$ tale che il vettore $(a, 1 - a)$ sia un vettore direzionale per r ,
- $b \in \mathbb{R}$ tale che $(b, -b)$ sia un punto appartenente a r ,

e scriveremo $r = r_{(a,b)}$.

Osserviamo che, data $\Phi := (f, g) : X \rightarrow \mathbb{R}$ una funzione filtrante, per ogni $r_{(a,b)} \in \Lambda^+$ possiamo associare ad ogni suo punto di coordinate $(u, v) = (u(t), v(t)) = t(a, 1 - a) + (b, -b)$ un sottoinsieme $X_t^{a,b} := X_{\Phi}^{(u(t), v(t))}$. Ciò suggerisce che ogni $r_{(a,b)} \in \Lambda^+$ definisce una filtrazione⁸ $\{X_t^{a,b}\}_{t \in \mathbb{R}}$ di cui in particolare possiamo considerare il diagramma di persistenza. Questo ci permette di dare la seguente definizione.

Definizione 3.3.6. *Per ogni $\Phi : X \rightarrow \mathbb{R}^2$ abbiamo una famiglia di diagrammi di persistenza, uno per ciascuna retta $r_{(a,b)}$ al variare di $(a, b) \in (0, 1) \times \mathbb{R}$. Per semplicità, chiamiamo tale collezione un **diagramma di persistenza 2D** di Φ .*

Osservazione 3.3.7. *Si verifica facilmente che la filtrazione $\{X_t^{a,b}\}_{t \in \mathbb{R}}$ è indotta dalla seguente funzione filtrante a valori reali:*

$$\Phi_{(a,b)} : X \rightarrow \mathbb{R} \mid x \mapsto \max\left\{\frac{f(x) - b}{a}, \frac{g(x) + b}{1 - a}\right\}.$$

Noi la utilizzeremo nella sua versione normalizzata

$$\Phi_{(a,b)}^* := \min\{a, 1 - a\}\Phi_{(a,b)}.$$

Esempio 3.3.8 (Esempio 1). *Continuiamo ad analizzare la situazione descritta negli esempi precedenti, ma questa volta ci concentriamo sulla funzione a valori in \mathbb{R}^2 , cioè Φ . Per $(x, y) \in S^1$, poniamo*

$$\Phi^*(x, y) := \Phi_{(\frac{1}{2}, 0)}^*(x, y) = \frac{1}{2} \max\{2x, 2y\} = \max\{x, y\}.$$

La filtrazione indotta da Φ^ è riportata in Figura 3.13. Il lettore attento si accorgerà che è la stessa filtrazione che si può leggere lungo l'antidiagonale della bifiltrazione di Φ riportata in Figura 3.1.*

⁸Le inclusioni dei sottoinsiemi così ottenuti sono garantite dal fatto che le rette considerate sono a pendenza positiva, perciò abbiamo effettivamente una filtrazione.

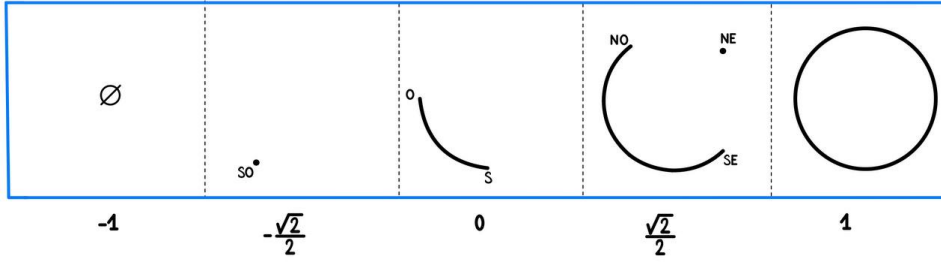


Figura 3.13: Filtrazione indotta dalla funzione Φ^* dell'Esempio 3.3.8

Grazie all'Osservazione 3.3.7 e al Teorema di Riduzione⁹, possiamo ricostruire interamente le funzioni dei numeri di Betti persistenti β_Φ considerando tutte le sole $\beta_{\Phi_{(a,b)^*}}$, che sono a loro volta totalmente descritte dai diagrammi di persistenza $Dgm(\Phi_{(a,b)^*}^*)$. È pertanto possibile “spacchettare” lo studio della persistenza biparametrica in uno studio collettivo di carattere monoparametrico. In particolare, possiamo dare la seguente definizione.

Definizione 3.3.9. *Siano $\Phi, \Psi : X \rightarrow \mathbb{R}^2$ due funzioni filtranti. Definiamo la distanza di matching tra Φ e Ψ la quantità*

$$d_{match}(\Phi, \Psi) := \sup_{(a,b) \in (0,1) \times \mathbb{R}} d_B(\Phi_{(a,b)^*}^*, \Psi_{(a,b)^*}^*),$$

dove abbiamo usato la Notazione 3.2.25.

Esempio 3.3.10 (Esempio 1). *In generale, calcolare a mano esplicitamente la distanza di matching è un'operazione dispendiosa, anche in situazioni semplici come quella degli esempi portati avanti finora. Tuttavia possiamo darne una stima.*

Consideriamo Φ^* definita nell'Esempio 3.3.8 e per $(x, y) \in S^1$, poniamo

$$F^*(x, y) := F_{(\frac{1}{2}, 0)}^*(x, y) = \frac{1}{2} \max\{2x, 2y\} = x = f(x).$$

La filtrazione indotta da Φ^* è quella in Figura 3.13, mentre quella indotta da F^* è la stessa indotta da f e dunque otteniamo un'unica coppia di persistenza di grado 0, $(-1, \infty)$ di molteplicità 1, come visto nell'Esempio 3.2.12. Possiamo osservare i due diagrammi di persistenza e i possibili matching tra loro in Figura 3.14. Otteniamo così la seguente distanza di bottleneck:

$$d_B(\Phi^*, F^*) = 1 - \frac{\sqrt{2}}{4}.$$

Possiamo pertanto concludere che $d_{match}(\Phi, F) \geq 1 - \frac{\sqrt{2}}{4}$.

⁹riferimento

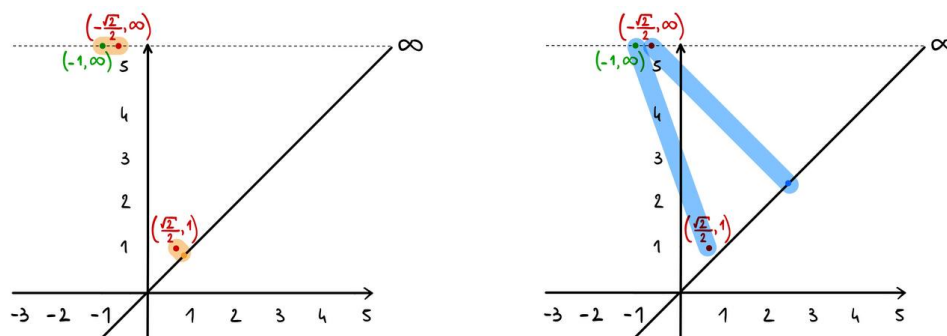


Figura 3.14: Possibili matching tra le funzioni Φ^* ed F^* dell'Esempio 3.3.10.

Infine, si può estendere lo spazio dei parametri (a, b) a $[0, 1] \times \mathbb{R}$, cioè considerare anche linee orizzontali e verticali, grazie a [14]. In particolare, è possibile definire:

$$\Phi_{(0,b)}^*(x) := \lim_{(a',b') \rightarrow (0,b)} \Phi_{(a',b')}^*$$

$$\Phi_{(1,b)}^*(x) := \lim_{(a',b') \rightarrow (1,b)} \Phi_{(a',b')}^*$$

Si dimostra poi che

$$\Phi_{(0,b)}^*(x) = \max\{f(x) - b, 0\}$$

$$\Phi_{(1,b)}^*(x) = \max\{0, g(x) + b\}$$

Questo ci tornerà utile nel Capitolo 6. In particolare, ci servirà la seguente osservazione.

Osservazione 3.3.11. *Supponiamo f e g a valori non negativi; se anche non lo fossero di per sé, basterebbe considerare una loro opportuna traslazione rigida. Allora, per $a = 0, b = 0$ otteniamo $\Phi_{(0,0)}^* = f$ e $\Phi_{(1,0)}^* = g$. In particolare,*

$$d_{\text{match}}(f, g) \geq d_B(f, g).$$

Stabilità della distanza di matching

Anche in questo caso abbiamo un teorema di stabilità, analogo a quello che coinvolge la distanza di bottleneck.

Teorema 3.3.12. *Sia X uno spazio topologico triangolabile e siano $\Phi, \Psi : X \rightarrow \mathbb{R}^2$ regolari¹⁰. Allora per ogni $p \geq 0$*

$$d_{\text{match}}(\Phi, \Psi) \leq \|f - g\|_\infty.$$

¹⁰Come prima, per *funzione filtrante regolare* qui intendiamo una funzione $X \rightarrow \mathbb{R}^2$ continua t.c. i gruppi di omologia degli insiemi di sottolivello hanno rango finito e i valori critici sono in numero finito.

Nuovamente, grazie alla nostra Assunzione 3.1.2, basta assumere X triangolare per soddisfare le ipotesi del teorema. Per una dimostrazione, si può consultare [5].

Capitolo 4

Problema 1: differenze nell'efficacia discriminante tra approccio monoparametrico e multiparametrico

Nei capitoli precedenti abbiamo descritto la persistenza monoparametrica e multiparametrica, individuandone le caratteristiche salienti e descrivendone alcuni invarianti. Ora le metteremo a confronto e questo ci porterà a individuare e formulare un problema, di cui al momento non esistono formalizzazioni o studi specifici in letteratura.

4.1 Formulazione del Problema 1

Alla luce di quanto visto nei capitoli precedenti, date due funzioni $\Phi_1 := (f_1, g_1), \Phi_2 := (f_2, g_2) : X \rightarrow \mathbb{R}^2$, definite su uno spazio topologico X compatto, l'omologia persistente permette di confrontarle in due modi. Ad ognuna delle funzioni coinvolte possiamo associare un modulo di omologia persistente, che denotiamo rispettivamente con M_{f_i}, M_{g_i} (monoparametrici) e M_{Φ_i} (multiparametrici), per $i = 1, 2$, e a questo punto possiamo mettere a confronto Φ_1 e Φ_2 adottando due punti di vista.

Approccio monoparametrico : confrontiamo (M_{f_1}, M_{g_1}) con (M_{f_2}, M_{g_2}) ;

Approccio multiparametrico : confrontiamo M_{Φ_1} e M_{Φ_2} .

Tali approcci sono equivalenti? In questo capitolo, cercheremo di rispondere a questa domanda basandoci sugli strumenti finora a nostra disposizione. Vedremo che, da un lato, da un punto di vista teorico e formale l'approccio monoparametrico è più robusto, in quanto presenta un numero nettamente maggiore di invarianti rispetto al secondo — i quali sono per giunta

completi, a differenza di quelli presenti nel secondo caso — e una minor complessità computazionale¹; dall'altro lato mostreremo che nonostante questo è fin troppo facile costruire esempi in cui Φ_1 e Φ_2 sono indistinguibili per il primo approccio ma non per il secondo. È in questo senso che la teoria multiparametrica garantisce una maggior capacità discriminante rispetto a quella monoparametrica e il mio lavoro di Tesi è volto proprio a comprendere i motivi teorici di tale differenza, nonché spiegare da un punto di vista più intuitivo cosa essa rappresenti.

4.1.1 Invarianti e descrittori: differenze tra caso monoparametrico e multiparametrico

Come abbiamo in parte visto nel Capitolo 3, i moduli di persistenza monoparametrici presentano una varietà di descrittori topologici e invarianti completi, che ci permettono di caratterizzare in modo esaustivo la loro struttura. Abbiamo già illustrato i numeri di Betti persistenti e i diagrammi di persistenza, entrambi invarianti completi; altri invarianti sono ad esempio i barcode, il rank invariant e i P -intervalli, anch'essi completi. Possibili riferimenti per le definizioni e le proprietà di tali invarianti sono [9, 13, 24]. Il fatto che la teoria monoparametrica risulti così ricca di invarianti e che essi siano completi, risiede nel fatto che, nel contesto monoparametrico, possiamo fare affidamento sul Teorema di Struttura 3.2.1, il quale garantisce una decomposizione univoca dei moduli di persistenza tramite le coppie di persistenza, ed esse sono il concetto tramite il quale si definiscono gli altri invarianti del caso monoparametrico. In questo modo, *la teoria monoparametrica risulta molto salda e ordinata dal punto di vista teorico.*

Nel caso dei moduli di persistenza multiparametrici, invece, la situazione è più complessa e meno soddisfacente dal punto di vista della completezza degli invarianti. Innanzitutto, non tutti gli invarianti del caso monoparametrico trovano il loro esatto analogo nel caso multiparametrico; inoltre, gli invarianti che è possibile costruire in questo caso non sono completi.

Gli invarianti principali del caso multiparametrico sono i numeri di Betti persistenti multiparametrici e il persistence space, che abbiamo già descritto nel Capitolo 3, il rank invariant, il multirank invariant e i blockcodes (possibili riferimenti: [3, 9]), nessuno dei quali risulta completo. In particolare, il multirank invariant è completo solo per moduli interval-decomposable, perciò possiamo affermare che per ottenere la completezza bisogna comunque ricondursi al contesto monoparametrico. La situazione può diventare sconcertante affrontando la lettura di [3]: lì viene dimostrato che *non può* esistere un invariante completo nel caso multiparametrico.

Aggiungendo a queste considerazioni il fatto che, almeno inizialmente, la

¹Questo è vero in linea di principio. In realtà nel corso degli anni sono stati compiuti passi avanti notevoli, e al momento non ci sono differenze sostanziali tra i due approcci.

teoria multiparametrica era molto più complessa dal punto di vista computazionale, una domanda che può sorgere spontanea è:

Perché non accontentarsi della persistenza monoparametrica?

Il punto è che *la persistenza multiparametrica è più **pratica** [18] di quella monoparametrica*. Infatti, le filtrazioni sorgono naturalmente quando si cerca di calcolare nella pratica gli invarianti topologici di uno spazio topologico e un approccio multiparametrico garantisce una maggior precisione. In altre parole, nonostante gli invarianti del caso multiparametrico siano incompleti, essi sono più efficaci nel distinguere spazi topologici differenti, come avremo modo di vedere nella prossima sezione.

4.1.2 Quale approccio garantisce una maggior efficacia discriminante?

Nella sezione precedente, abbiamo illustrato i numerosi invarianti e descrittori della persistenza monoparametrica e quelli, non altrettanto numerosi e non completi, della persistenza multiparametrica. Possiamo inoltre ragionevolmente aspettarci una maggior complessità computazionale del caso multiparametrico rispetto a quello monoparametrico². Pertanto, sembrerebbe che non ci siano vantaggi nell'utilizzo dell'approccio multiparametrico al posto di quello monoparametrico.

In realtà, sorprendentemente l'approccio multiparametrico risulta molto più efficace di quello monoparametrico nel confrontare funzioni, in particolare nel distinguere funzioni diverse. Vediamo portando avanti l'Esempio 1 sviluppato nei capitoli precedenti.

Esempio 4.1.1 (Esempio 1). *Sia $X := S^1$ la circonferenza unitaria centrata nell'origine e consideriamo le seguenti funzioni continue:*

$$\begin{aligned} f : X &\rightarrow \mathbb{R} \mid (x, y) \mapsto x, \\ g : X &\rightarrow \mathbb{R} \mid (x, y) \mapsto y, \\ \Phi &:= (f, g) : X \rightarrow \mathbb{R}^2, \\ F &:= (f, f) : X \rightarrow \mathbb{R}^2, \\ G &:= (g, g) : X \rightarrow \mathbb{R}^2. \end{aligned}$$

Le filtrazioni che possiamo considerare sono quelle rappresentate³ in Figura 4.1, e le tabelle contenenti i corrispondenti numeri di Betti persistenti di grado 0 sono riportate in Figura 4.1. Essendo Φ , F e G evidentemente fun-

²Questo è vero in linea di principio. Come anticipato, in realtà nel corso degli anni sono stati compiuti passi avanti notevoli, e al momento non ci sono differenze sostanziali tra i due casi.

³Ricordiamo che per rendere una filtrazione visualizzabile, è necessario discretizzarla. A tal scopo, in questo esempio abbiamo fissato i valori $-1, -\frac{\sqrt{2}}{2}, 0, \frac{\sqrt{2}}{2}, 1$.

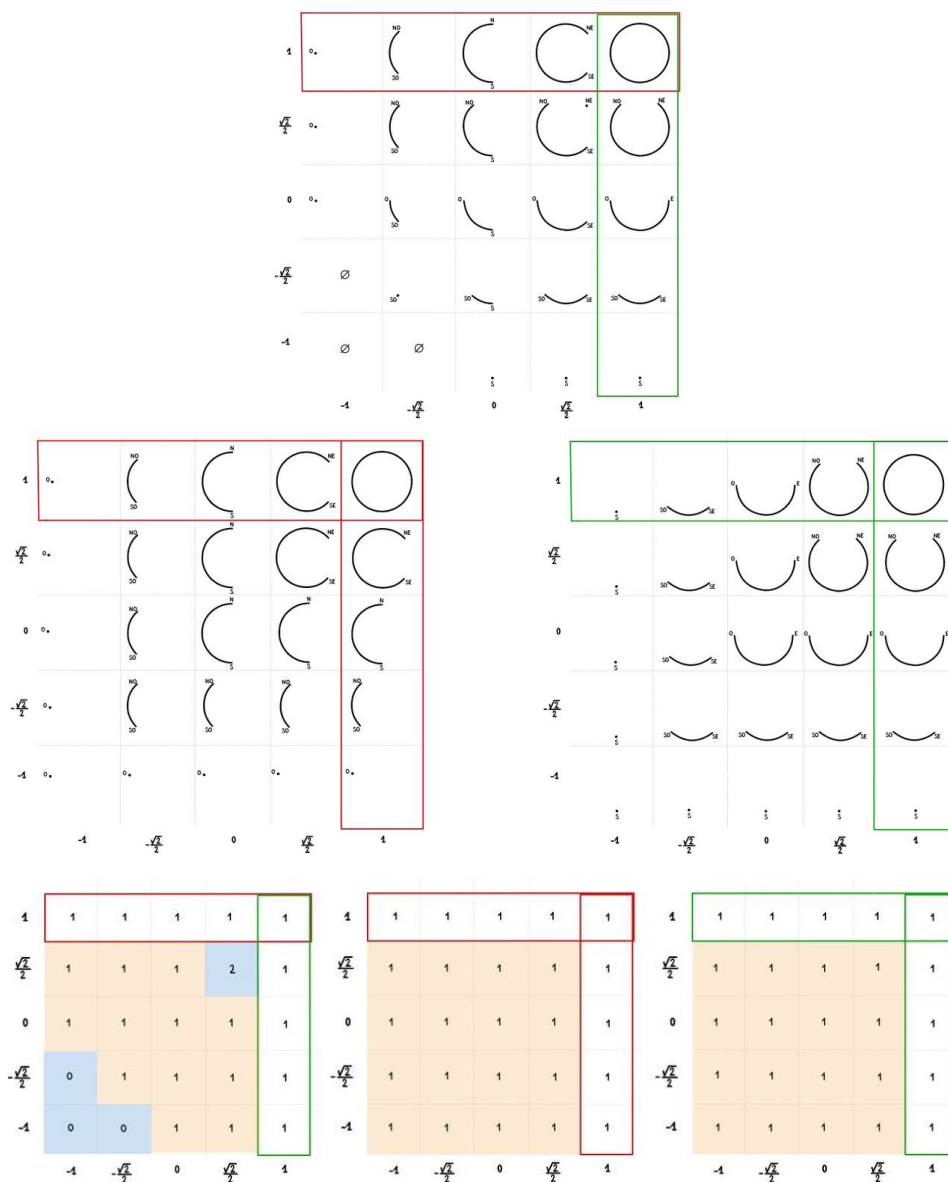


Figura 4.1: Sopra: filtrazioni indotte da f, g, Φ, F e G dell'Esempio 1. Sotto: Dimensioni dei corrispondenti gruppi di omologia in grado zero.

zioni diverse — poiché sono f e g ad essere funzioni distinte —, vorremmo poterle distinguere. Procediamo seguendo i due approcci separatamente per controllare se siano equivalenti o no.

Approccio monoparametrico. Per procedere con questo approccio dobbiamo considerare le coppie di moduli (M_f, M_g) , (M_f, M_f) e (M_g, M_g) , rappresentate dai bordi incorniciati delle tabelle in Figura 4.1, che sono indistinguibili. Pertanto, lavorando in modo monoparametrico non è possibile distinguere Φ da F né Φ da G — cioè non riusciamo a distinguere f e g .

Approccio multiparametrico Viceversa, in questo caso dobbiamo confrontare i moduli M_Φ, M_F e M_G , raffigurati dalle intere tabelle in Figura 4.1. In particolare, osservando i riquadri gialli, ovvero le parti “puramente biparametriche” dei moduli di persistenza considerati, riusciamo a distinguere M_Φ da M_F e da M_G .⁴

Osserviamo che, dato che l’omologia di S^1 è sempre nulla per $p > 1$, l’unico altro grado in omologia in cui potremmo sperare di trovare differenze già considerando la persistenza monoparametrica è il grado 1. Nemmeno in questo caso si riesce nell’intento.

In questo esempio risulta pertanto evidente che la persistenza multiparametrica abbia un maggior potere discriminante rispetto a quella monoparametrica, nonostante non abbiamo ancora a disposizione un modo per misurare formalmente la differenza tra i due casi.

4.1.3 Formulazione del problema

Tirando le somme, ci troviamo di fronte alle seguenti questioni aperte.

Problema 1: Come mai, nonostante le sue proprietà facciano presumere il contrario, l’omologia persistente multiparametrica risulta più efficace dal punto di vista discriminativo rispetto a quella monoparametrica? E che significato hanno le informazioni aggiuntive ricavate tramite l’approccio multiparametrico rispetto a quello monoparametrico? Come si può formalizzare e misurare il gap informativo tra i due?

4.2 Esempi

L’Esempio 1 che abbiamo appena considerato non è l’unico esempio possibile di situazioni in cui due funzioni diverse non sono distinguibili con

⁴Invece, comunque M_F e M_G continuano ad essere indistinguibili; sarà questa l’osservazione chiave che ci condurrà alla definizione di correlazione topologica nel capitolo 6.

l'approccio monparametrico ma lo sono con quello multiparametrico. Al contrario, è molto facile trovare esempi analoghi. Ne presentiamo alcuni, che continueremo a studiare anche successivamente.

Notazione 4.2.1. *D'ora in avanti, tutti gli spazi topologici e le funzioni considerate nell'Esempio i presenteranno un pedice i . Ad esempio, nell'Esempio 1 d'ora in avanti scriveremo: $X_1 := S^1$, $f_1 := f$, $g_1 := g$, $\Phi_1 := \Phi$, $F_1 := F$ e $G_1 := G$.*

Esempio 4.2.2 (Esempio 2). *Consideriamo un toro $X_2 := T$ e le funzioni $f_2 : X \rightarrow \mathbb{R} \mid (x, y, z) \mapsto x$ e $g_2 : X \rightarrow \mathbb{R} \mid (x, y, z) \mapsto y$, e poniamo $\Phi_2 := (f_2, g_2)$. Nuovamente rispetto ad opportuni valori fissati, la filtrazione biparametrica indotta da Φ_2 è rappresentata in Figura 4.2. Lavorando come*

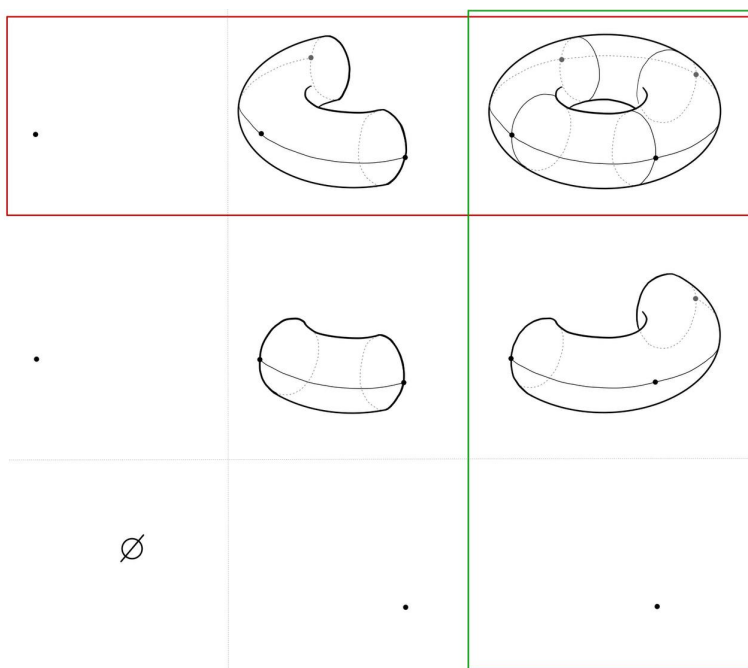


Figura 4.2: Bifiltrazione indotta da Φ_2 nell'Esempio 2.

nell'esempio precedente considerando anche $F_2 := (f_2, f_2)$ e $G_2 := (g_2, g_2)$ e ragionando con l'omologia in grado 0, si conclude che in questo caso le sole informazioni monparametriche non sono sufficienti per discriminare tra f e g .

Esempio 4.2.3 (Esempio 3). *Consideriamo $X_3 := I^2 = [0, 1] \times [0, 1]$ e le funzioni $f_3 : X_3 \rightarrow \mathbb{R} \mid (x, y) \mapsto \min\{x, y\}$ e $g_3 : X_3 \rightarrow \mathbb{R} \mid (x, y) \mapsto 1 - \max\{x, y\}$, e poniamo $\Phi_3 := (f_3, g_3)$. Nuovamente rispetto ad opportuni valori fissati, la filtrazione biparametrica indotta da Φ_3 è rappresentata in Figura 4.3. Lavorando come negli esempi precedenti considerando an-*

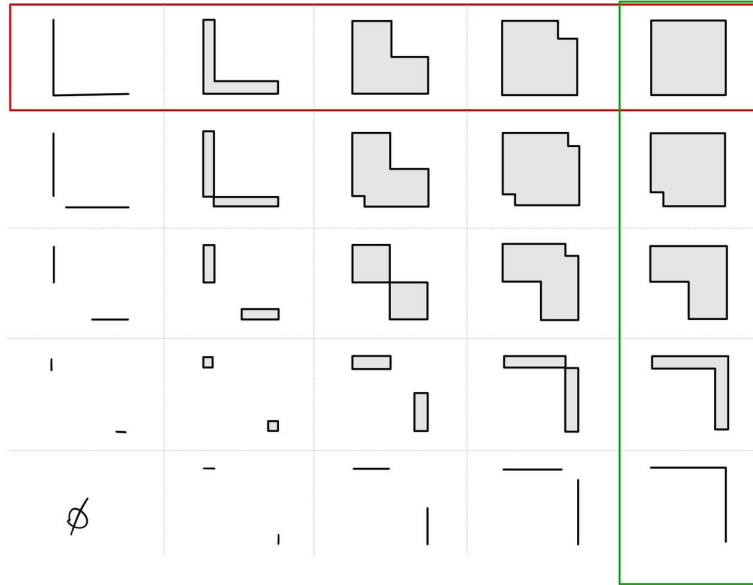


Figura 4.3: Bifiltrazione indotta da Φ_3 nell'Esempio 3.

che $F_3 := (f_3, f_3)$ e $G_3 := (g_3, g_3)$ e ragionando con l'omologia in grado 0, si conclude che in questo caso le sole informazioni monoparametriche non sono sufficienti per discriminare tra f_3 e g_3 .

Negli esempi precedenti e nell'Esempio 4.1.1 abbiamo lavorato con l'omologia in grado 0; si possono ottenere risultati simili anche lavorando in grado 1, come vedremo nel prossimo esempio.

Esempio 4.2.4 (Esempio 4). Consideriamo ora una sfera $X_4 := S^2$ e le funzioni $f_4 : X_4 \rightarrow \mathbb{R} \mid (x, y, z) \mapsto x$, $g_4 : X \rightarrow \mathbb{R} \mid (x, y, z) \mapsto z$ e $\Phi_4 := (f_4, g_4)$. La bifiltrazione indotta da Φ_4 è rappresentata in Figura 4.4, dove nuovamente la rappresentiamo rispetto ad alcuni valori discreti fissati e non tutti i possibili valori reali.

In particolare, $X_{4, \Phi}^{(\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2})}$ è lo spazio rappresentato in Figura 4.5 e possiede omologia non nulla in grado 1, mentre si vede facilmente che le bifiltrazione indotte da $F_4 := (f_4, f_4)$ e $G_4 := (g_4, g_4)$ non presentano un tale sottospazio. Perciò, anche in questo caso si conclude che le sole informazioni monoparametriche non sono sufficienti per distinguere f da g .

È molto facile costruire esempi simili anche nel caso dei complessi simpliciali, come descritto in Appendice A. Senza necessariamente entrare nei dettagli di tali costruzioni, potete osservare le tabelle in Figura A.9, Figura A.10 e Figura A.11 e apprezzare le differenze tra la parte multiparametrica e quella monoparametrica.

La letteratura fornisce molte conferme sul maggior potere discriminante della multipersistenza rispetto alla monopersistenza, ma d'altro canto non

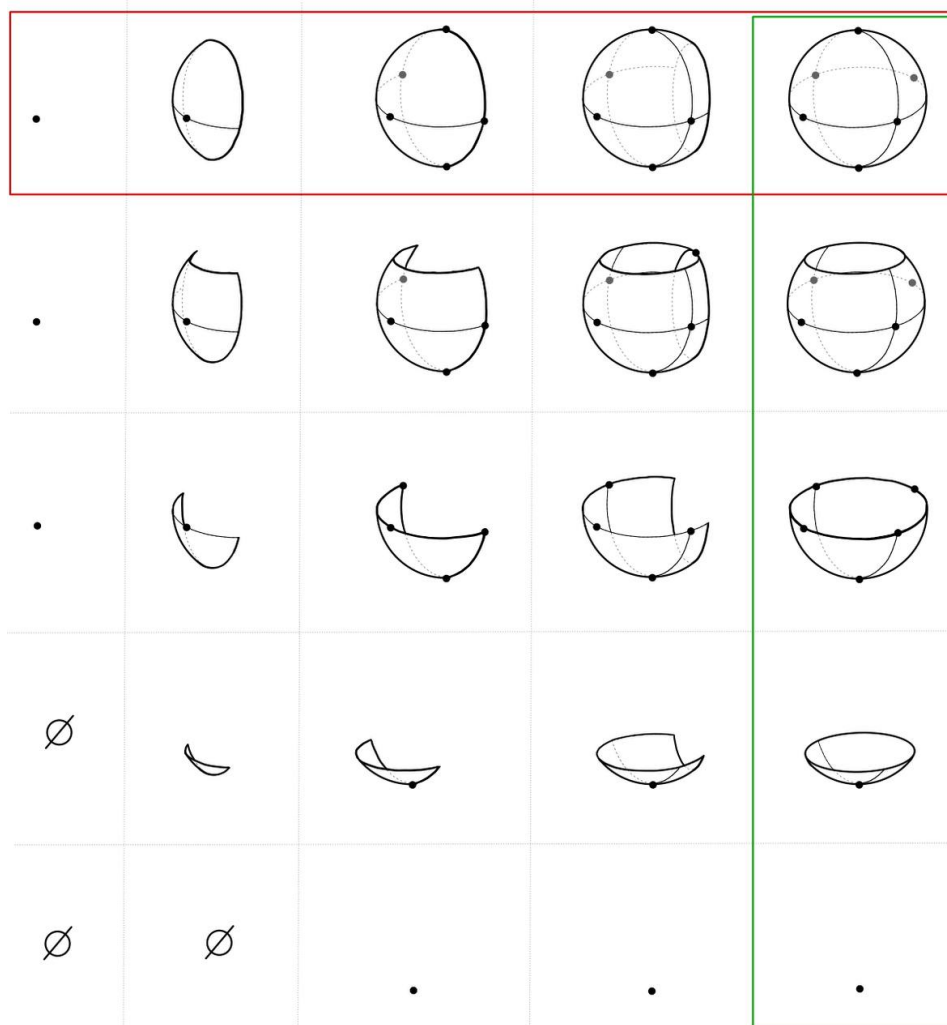


Figura 4.4: Bifiltrazione indotta dalla funzione Φ descritta nell'Esempio 4.2.4.

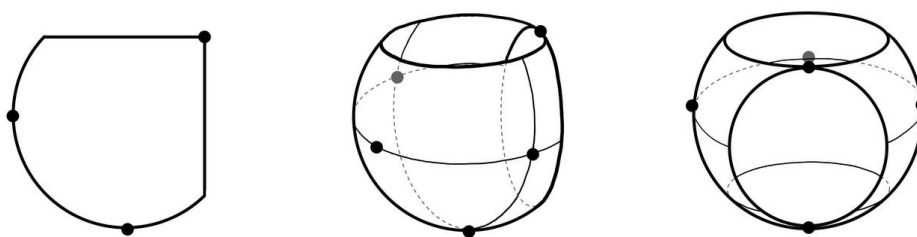


Figura 4.5: Sottospazio, osservato da diversi punti di vista, nella filtrazione indotta da Φ_4 nell'Esempio 4.

ha mai affrontato la questione in modo diretto. Lo stesso sviluppo e studio di invarianti può in realtà fornire delle prospettive interessanti a riguardo, tuttavia non esistono studi dedicati principalmente a tale tematica.

Nella mia Tesi invece, queste problematiche costituiscono uno dei due poli attorno cui orbita l'intero lavoro. Prima di addentrarci nella questione, nel prossimo capitolo ci dedicheremo alla descrizione del secondo problema "polo" della mia Tesi, il quale a prima vista sembrerà del tutto scollegato a quello appena formulato, ma che in realtà si rivelerà la sua chiave di lettura.

Capitolo 5

Problema 2: correlazione tra funzioni

In questo capitolo ci addentreremo in un contesto diverso da quello considerato finora e ci occuperemo di questioni apparentemente scollegate dai capitoli precedenti. Il motivo di questa deviazione rispetto al percorso delineato finora sarà chiaro nel prossimo, nonché ultimo, capitolo, dove troveremo un collegamento tra quanto stiamo per affrontare e quanto visto in precedenza.

5.1 Formulazione del Problema 2

Nelle indagini statistiche e nelle analisi dati, tra gli strumenti più utili spiccano quelli che consentono di effettuare riduzione dimensionale e clusterizzazione; le motivazioni e i vantaggi sono molteplici, dalla possibilità di poter visualizzare i dati più facilmente a quella di poter ridurre la complessità computazionale del problema considerato, e così via. In questo capitolo ci occuperemo di presentare uno degli strumenti base coinvolti in tali tecniche, ovvero la **correlazione**, e di individuare un contesto in cui, almeno per il momento, le nozioni analoghe presentano notevoli limitazioni e in cui, in realtà, non esiste un vero e proprio analogo. Più precisamente, ci poniamo il seguente problema: esiste un concetto di correlazione tra funzioni? Con quali proprietà?

5.1.1 Correlazione tra variabili

La correlazione è uno strumento statistico che rileva e misura eventuali relazioni (lineari o monotone) tra due o più variabili. Quando due variabili sono correlate, non è detto che tra di esse intercorra una relazione di tipo causa-effetto; piuttosto significa che esiste una tendenza nei cambiamenti di una delle due rispetto ai cambiamenti dell'altra; ad esempio all'aumentare

di una anche l'altra aumenta.

Siano A e B due variabili (quantitative) rilevate su una stessa popolazione campione $\Omega := \{\omega_i\}_{i=1,\dots,n}$, cioè

$$\begin{aligned} A : \Omega &\rightarrow \mathbb{R} \mid \omega_i \mapsto a_i, \\ B : \Omega &\rightarrow \mathbb{R} \mid \omega_i \mapsto b_i, \end{aligned}$$

Con abuso di notazione, indichiamo con $A := (a_1, \dots, a_n)$ e $B := (b_1, \dots, b_n)$, vettori in \mathbb{R}^n i quali rappresentano le immagini rispettivamente di A e B .

Definizione 5.1.1. *La correlazione (lineare) tra A e B è definita come*

$$\rho(A, B) := \frac{(A - \bar{a})^t (B - \bar{b})}{\|A - \bar{a}\| \|B - \bar{b}\|},$$

dove con \bar{a} e \bar{b} indichiamo la media rispettivamente di A e di B e $\|\cdot\|$ indica la norma euclidea in \mathbb{R}^n .

Vediamo un semplice esempio.

Esempio 5.1.2. *Consideriamo una popolazione Ω di n studentesse e studenti universitari. Per ciascuno di essi, cioè per ciascun $\omega_i \in \Omega$, registriamo il suo peso a_i e la sua altezza b_i . Otteniamo così due variabili*

$$\begin{aligned} A : \Omega &\rightarrow \mathbb{R} \mid \omega_i \mapsto a_i, \\ B : \Omega &\rightarrow \mathbb{R} \mid \omega_i \mapsto b_i. \end{aligned}$$

Possiamo rappresentare i dati raccolti tramite un grafico in \mathbb{R}^2 , detto scatter-plot: per ogni $i = 1, \dots, n$ consideriamo il punto di \mathbb{R}^2 di coordinate (a_i, b_i) . Otteniamo così il grafico mostrato in Figura 5.1. È evidente dal grafico che ci sia la seguente tendenza nei dati raccolti: all'aumentare dell'altezza, in generale, aumenta anche il peso. La correlazione tra A e B esprime proprio tale tendenza; in questo caso, essa vale $\rho(A, B) = 0.61$ [23].

Solitamente un'indagine statistica coinvolge un numero superiore a 2, anche di gran lunga, di variabili, e in questi casi conoscere quali tra queste siano correlate può risultare decisivo; si può decidere di considerare solo una fra due variabili correlate e questo permette di ridurre la dimensione del problema studiato, con tutti i vantaggi che ne conseguono. Riprendendo l'Esempio 5.1.2, se per esempio consideriamo anche l'età, le pulsazioni al minuto e il numero di sigarette consumate in media al giorno, non possiamo più visualizzare tali dati in modo immediato come visto in Figura 5.1, poiché in questo caso a ciascun individuo è associato un vettore in \mathbb{R}^5 e non in \mathbb{R}^2 (o in \mathbb{R}^3); invece, cercando quali variabili sono correlate, ci saremmo accorti che peso e altezza sono correlate, così come taglia di scarpe ed altezza, e pertanto avremmo potuto decidere di considerare solo le tre variabili date da altezza, pulsazioni e numero di sigarette, riuscendo a visualizzare tali dati in \mathbb{R}^3 .

Abbiamo ora gli strumenti per formalizzare il Problema 2.

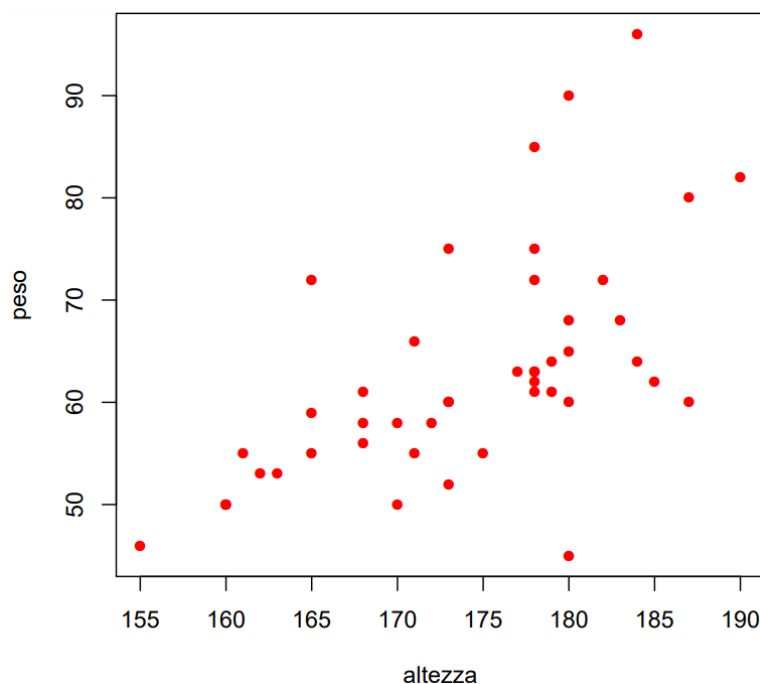


Figura 5.1: Scatterplot dei pesi e delle altezze di un gruppo di studentesse e studenti universitari. Immagine tratta da [23].

5.1.2 Formulazione del problema

Abbiamo visto che le variabili sono funzioni che ad ogni individuo della popolazione associano un valore. *Cosa succede se invece a ciascun individuo associamo una funzione?* Ad esempio, riprendendo l'Esempio 5.1.2, per ogni studente coinvolto nell'indagine potremmo considerare due funzioni che ad ogni punto della superficie corporea considerata associano rispettivamente temperatura corporea e pressione sanguigna rilevate in quel punto. Più precisamente, vorremmo poter considerare due “variabili” del tipo

$$\begin{aligned} f : \Omega &\rightarrow \mathcal{F} \mid \omega_i \mapsto f_i, \\ g : \Omega &\rightarrow \mathcal{F} \mid \omega_i \mapsto g_i, \end{aligned}$$

con \mathcal{F} un opportuno insieme di funzioni $\omega_i \rightarrow \mathbb{R}$ al variare di i , dove con abuso di notazione identifichiamo ogni individuo con la propria superficie corporea, per poi definire un concetto di correlazione tra f e g analogo a quello appena visto per variabili “classiche”. In tal caso, non è chiaro come procedere.

Questo esempio ci pone pertanto di fronte alle seguenti domande.

Problema 2: Come facciamo a determinare se due funzioni sono correlate?
E cosa intendiamo esattamente per correlazione tra funzioni?

Nelle prossime sezioni, indaghiamo quali risposte possiamo ottenere dalla letteratura attuale e vedremo che al momento non esiste una generalizzazione vera e propria del concetto di correlazione nel contesto delle funzioni, ma solo risposte parziali al problema.

5.2 Correlazione tra funzioni

5.2.1 Stato dell'Arte

In questa sezione esploreremo la letteratura esistente riguardo al concetto di correlazione tra funzioni. Ai fini della trattazione, non affronteremo gli argomenti seguenti con la massima generalità possibile, ma cercheremo di individuarne i concetti salienti e di acquisirne anche una visione concreta ed intuitiva tramite esempi. Per uno studio più approfondito, rimandiamo a [2, 15].

Marking point

Nel cercare di affrontare il problema, ci si rende subito conto di quanto si complichino la questione nel passare da variabili a funzioni. Tra i due casi, in realtà, ne esiste uno intermedio: il caso delle variabili del tipo

$$A : \Omega \mapsto \mathbb{R}^d \mid \omega_i \mapsto a_i,$$

con $a_i = (\alpha^1, \dots, \alpha^d)$. In questa situazione, al contrario del contesto delle funzioni, esiste un concetto di correlazione. Una possibile strategia è perciò quella di passare in qualche modo dal considerare funzioni a al considerare vettori, ed è qui che entrano in gioco i cosiddetti *marking point* o *marker*. Vediamo come nel prossimo esempio.

Esempio 5.2.1. *Consideriamo sempre la stessa popolazione di studentesse e studenti universitari Ω , di cui questa volta siamo interessati a conoscere temperatura corporea e pressione sanguigna in ciascun punto delle rispettive superfici corporee. Come accennato in precedenza, quello che stiamo facendo è associare a ciascun individuo ω_i due funzioni $f_i : \omega_i \rightarrow \mathbb{R}$ e $g_i : \omega_i \rightarrow \mathbb{R}$, dove identifichiamo ciascun individuo con la propria superficie corporea. Per poter associare a ciascun ω_i due vettori anziché due funzioni, si può procedere selezionando dei punti salienti e rilevare le grandezze di interesse solamente in quei punti. Tali punti sono chiamati **marker**. Ad esempio, potremmo selezionare il centro della fronte, il centro dell'addome, e i punti estremali di ciascun arto, per un totale di sei valori per ciascun individuo. Otteniamo così due variabili*

$$A : \Omega \rightarrow \mathbb{R}^6 \mid \omega_i \mapsto a_i$$

$$B : \Omega \rightarrow \mathbb{R}^6 \mid \omega_i \mapsto b_i,$$

di cui riusciamo a calcolare la correlazione.

Il metodo appena illustrato nell'Esempio 5.2.1 risulta sempre efficace? Ragionandoci un minimo, ci si rende subito conto che non può essere così. Nel prossimo esempio vediamo un problema illustriamo una situazione in cui questa strategia non può funzionare.

Esempio 5.2.2. *Questa volta, al posto di una popolazione di studenti, consideriamo una popolazione $\{X_i\}_{i=1,\dots,N}$ di superfici molecolari, e supponiamo di voler studiare la compatibilità reciproca delle corrispondenti molecole. Tale proprietà dipende non soltanto, come ci si può aspettare, dalla forma delle molecole, ma anche da alcune loro caratteristiche chimico-fisiche, descritte tramite funzioni definite su ciascuna superficie molecolare X_i e a valori in \mathbb{R} . Per esempio, due di queste funzioni sono la distribuzione di carica elettrostatica e l'indice di idrofobicità, di cui possiamo vedere una rappresentazione, su due delle superfici molecolari considerate, in Figura 5.2.*

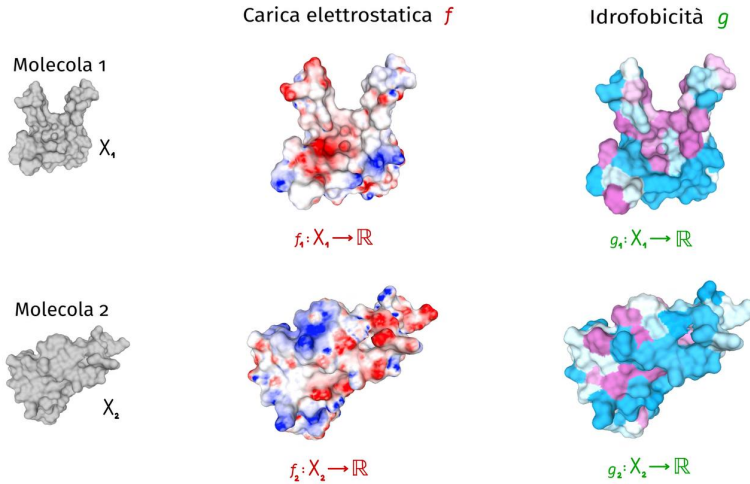


Figura 5.2: Rappresentazione delle funzioni di distribuzione di carica elettrostatica e indice di idrofobicità di due molecole campione. Immagini tratte da [22].

Più precisamente, abbiamo quindi

$$\begin{aligned} f : \Omega &\rightarrow \mathcal{F} \mid \omega_i \mapsto f_i, \\ g : \Omega &\rightarrow \mathcal{F} \mid \omega_i \mapsto g_i. \end{aligned}$$

A questo punto vorremmo procedere come nell'Esempio 5.2.1, ma non è così chiaro come fare: il mondo molecolare non è così omogeneo e uniforme come quello umano, e risulta molto complicato individuare dei marker point che risultino sensati e rilevanti per ciascuna molecola (talvolta non è possibile farlo nemmeno per una singola molecola). In questo caso, risulta perciò impossibile seguire la strada dei marker point.

È pertanto evidente che questa strategia presenti in generale dei limiti nelle possibilità di utilizzo.

Cross-correlation

Allontanandoci per un momento dal mondo della statistica e immergendoci in quello dell'analisi (di Fourier, armonica e dei segnali), riusciamo a trovare un altro possibile approccio. Dati due segnali $\mathbb{R} \rightarrow \mathbb{C}$ si può definire un'operazione tra di essi, molto simile al più noto prodotto di convoluzione, che prende il nome di **cross-correlation** e che permette di stabilire quanto i due segnali siano sovrapponibili.

Definizione 5.2.3. *Siano $f, g : \mathbb{R} \rightarrow \mathbb{C}$. Per $x \in \mathbb{R}$ poniamo*

$$h(x) := f \star g(x) := \int_{-\infty}^{\infty} f(t-x)g(t)dt. \quad (5.1)$$

Sotto certe ipotesi, si ottiene così una funzione ben definita $h : \mathbb{R} \rightarrow \mathbb{C}$ e, in tal caso, h prende il nome di funzione cross-correlation tra f e g .

La cross-correlation tra due funzioni è profondamente legata alla più nota convoluzione, infatti:

$$f \star g = (\overline{f \circ \Upsilon}) * g, \quad (5.2)$$

Dove Υ denota la riflessione $\mathbb{C} \rightarrow \mathbb{C} \mid x \mapsto -x$. Per questo motivo, possiamo dedurre in quali contesti sia ben definita la cross-correlation. Innanzitutto, tramite l'espressione (5.1) otteniamo una funzione ben definita quando sono ben definite la convoluzione e la riflessione; perciò sicuramente è possibile considerare f e g non solo definite su \mathbb{R} , ma anche su \mathbb{R}^d e in generale su un gruppo topologico localmente compatto ([15]). Più precisamente, possiamo partire dalle classi di funzioni per cui sia ben definito l'operatore di convoluzione e poi utilizzare il legame descritto da (5.2) per stabilire per quali classi di funzioni sia ben definito l'operatore di cross-correlation. Chiaramente quest'ultimo risulta ben definito su tutti gli spazi di funzioni su cui sia ben definito l'operatore di convoluzione e che siano invarianti per riflessione e coniugazione complessa; ad esempio, tutti gli spazi $L^p(\mathbb{R}^d)$ soddisfano tali condizioni. Con i dovuti accorgimenti, tutto questo può essere adattato alle funzioni periodiche ed inoltre esiste anche un analogo discreto della cross-correlation, utilizzato ad esempio in statistica nell'analisi delle serie storiche [2].

Vediamo in un semplice esempio come mai la cross-correlation costituisca un indice di "sovrapponibilità" fra due segnali.

Esempio 5.2.4. Consideriamo la funzione $f := \sin$. Essendo una funzione 2π periodica, la cross-correlation $f \star f$ è definita dall'espressione seguente:

$$f \star f(x) = \int_0^{2\pi} f(t-x)f(t)dt = \int_0^{2\pi} \sin(t-x)\sin(t)dt, \quad (5.3)$$

cioè si calcola allo stesso modo illustrato in precedenza, ma come dominio di integrazione si considera un intervallo che abbia lunghezza pari al periodo della funzione. Osseviamo che il calcolo di $f \star f$ prevede per ogni x di integrare f moltiplicata per una sua traslata di ampiezza x . In particolare osserviamo che:

- se $x = 0$ allora la traslazione di f di ampiezza x coincide con f ; essendo perciò le due funzioni totalmente sovrapponibili, ci aspettiamo un alto valore per $f \star f(0)$;
- se $x = \frac{\pi}{2}$ allora la traslazione di f di ampiezza x coincide con la funzione $-\cos$; essendo perciò le due funzioni sfasate, ci aspettiamo $f \star f(\frac{\pi}{2}) = 0$;
- se $x = \pi$ allora la traslazione di f di ampiezza x coincide con $-\sin$; essendo perciò le due funzioni a fasi opposte, ci aspettiamo un basso valore per $f \star f(\pi)$.

Portiamo avanti il calcolo di 5.3 e vediamo se le nostre ipotesi sono verificate. Ad esempio usando il fatto che $\sin(A)\sin(B) = \frac{1}{2}(\cos(A-B) - \cos(A+B))$, si ricava facilmente che

$$\int_0^{2\pi} \sin(t-x)\sin(t)dt = \pi \cos(x),$$

perciò

$$f \star f(x) = \pi \cos(x).$$

Concludiamo che in effetti i valori di cross-correlation ottenuti rispecchiano l'intuizione e soddisfano le nostre aspettative iniziali. Le tre situazioni sono rappresentate in Figura 5.3, dove: per $x = 0, \frac{\pi}{2}, \pi$, a sinistra sono rappresentate f e la sua traslata $f(\cdot - x)$, in centro è raffigurata la funzione integranda in (5.3) e a destra viene individuato il valore $f \star f(x)$ sul grafico di $f \star f$.

Osservazione 5.2.5. L'esempio appena descritto non è altro che una riscrittura dell'Esempio 1: considerare le funzioni “componente x ” e “componente y ” sulla circonferenza unitaria S^1 è equivalente a considerare le funzioni \sin e \cos su un qualsiasi intervallo reale di ampiezza 2π . Questa osservazione ci tornerà utile nel Capitolo 6.

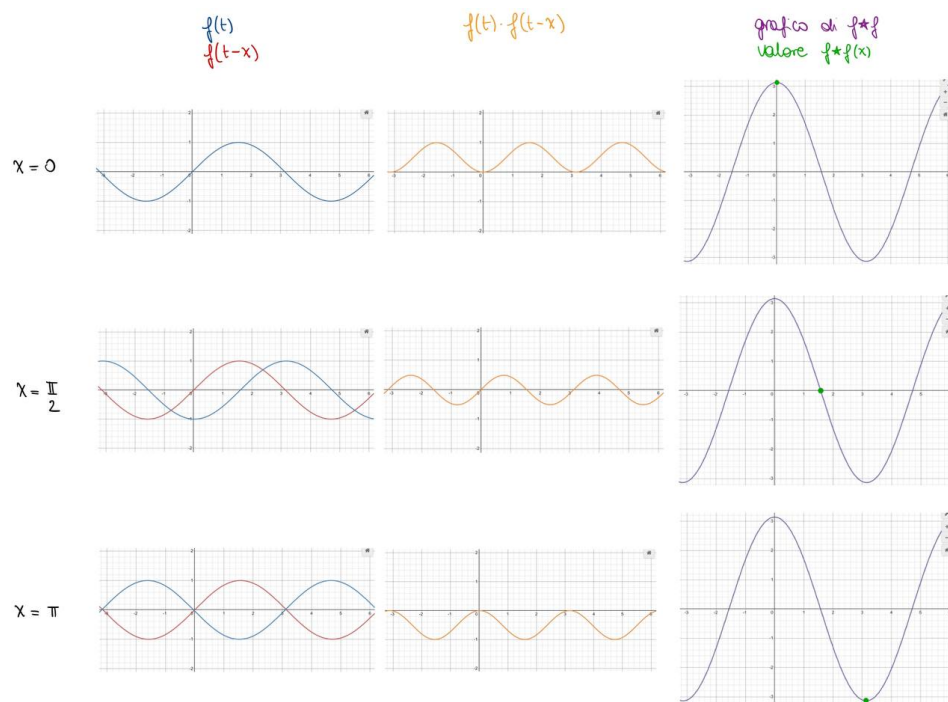


Figura 5.3: A sinistra: due funzioni senoide con tre diversi sfasamenti. In centro, le due funzioni rappresentate tramite una curva di Lissajous. A destra: la funzione di cross-correlation risultante.

La cross-correlation è pertanto un valido strumento per studiare la correlazione tra due funzioni nell'ambito dell'analisi armonica. Tuttavia, tale definizione non soddisfa pienamente le caratteristiche che vorremmo ottenere dalla nozione che incarna la nostra idea di correlazione tra funzioni:

- f e g devono necessariamente essere definite su \mathbb{R}^d o su gruppi topologici localmente compatti, noi vorremmo invece una definizione di correlazione per funzioni definite almeno su spazi topologici generici, possibilmente anche su complessi simpliciali e altre strutture topologiche;
- f e g devono necessariamente essere definite su uno stesso dominio, mentre noi vorremmo misurare la correlazione tra funzioni potenzialmente definite su domini diversi tra loro;
- gli spazi di funzioni che consentono di ottenere un operatore ben definito non garantiscono un'ampia generalità, perlomeno in relazione ai contesti applicativi in cui vorremmo utilizzare la definizione di correlazione topologica, in cui è possibile dover trattare funzioni "brutte a piacere".

5.2.2 Esistono definizioni più generali?

In generale non sembra esistere in letteratura un concetto che sia il perfetto analogo della correlazione nel contesto delle funzioni.

Vista la natura degli oggetti in gioco, è possibile che ci si possa servire di strumenti propri dell'analisi funzionale per sopperire a questa mancanza, in aggiunta al prodotto di convoluzione appena illustrato. Tuttavia, oltre al fatto che non è detto che si possa procedere in tal modo (e, per quanto riguarda la convoluzione, abbiamo appena visto che permette di considerare solamente una piccola parte delle possibili casistiche), questo lavoro di Tesi mira a definire un concetto di correlazione servendosi di strumenti che apparentemente non hanno nulla a che spartire con gli argomenti trattati in questo capitolo. È proprio a questo punto che torna in gioco ciò che abbiamo per il momento messo da parte: l'omologia persistente. Infatti, in modo forse sorprendente, l'omologia persistente sembra essere lo strumento sul quale basarsi per definire un concetto di **correlazione topologica** tra funzioni, come vedremo nel prossimo capitolo.

Capitolo 6

Due problemi, due soluzioni connesse: differenza e correlazione topologiche

Nei due capitoli precedenti abbiamo individuato ed esaminato dettagliatamente due questioni aperte: da un lato, la maggior capacità discriminante, per il momento ancora priva di spiegazione e formalizzazione, della multi-persistenza rispetto a quella della monopersistenza; dall'altro l'esigenza di definire un concetto di correlazione tra funzioni. Apparentemente, questi problemi sembrano slegati tra loro. Al contrario, le due questioni sono in realtà due facce di una stessa medaglia, come vedremo in questo capitolo, dove sfrutteremo proprio il legame tra i due problemi per dare una nuova definizione; in particolare, l'idea cruciale sarà quella di formalizzare e quantificare il gap informativo tra caso multiparametrico e monoperametrico per definire una correlazione tra due funzioni a valori reali che sono le componenti di una funzione filtrante a valori in \mathbb{R}^2 . Infine sfrutteremo questa definizione per confrontare funzioni definite su domini diversi.

6.1 Soluzione 1: differenza topologica

6.1.1 Definizione

Per prima cosa, abbiamo bisogno di un modo per far comunicare tra loro il mondo multiparametrico con quello monoperametrico. Più precisamente, vorremmo definire una quantità che, date due funzioni a valori in \mathbb{R}^2 , permetta di misurare e formalizzare la differenza tra ciò che rileviamo nelle due bifiltrazioni indotte nella loro totalità e ciò che invece rileviamo limitandoci a considerarne i bordi, ovvero limitandoci a considerare le sole filtrazioni monoperametriche indotte dalle componenti delle due funzioni. Formalmente, procediamo come segue.

Siano X uno spazio topologico compatto e siano $\Phi_1 := (f_1, g_1) : X \rightarrow \mathbb{R}^2$ e $\Phi_2 := (f_2, g_2) : X \rightarrow \mathbb{R}^2$ due funzioni continue. Per confrontarle possiamo seguire due strade: fissato un grado di omologia p , possiamo

- calcolare la distanza di matching $d_{match}(\Phi_1, \Phi_2)$ tra Φ_1 e Φ_2 ;
- calcolare le distanze di bottleneck tra le loro componenti, cioè calcolare $d_B(f_1, f_2)$ e $d_B(g_1, g_2)$. in grado p , cioè utilizzando i diagrammi di persistenza di grado p .

Osserviamo che la prima si colloca nell'ambito biparametrico, mentre la seconda nell'ambito monoparametrico; un'idea per poter rispondere al Problema 1 potrebbe perciò essere quella di combinare queste grandezze e confrontarle. Questo ci porta alla seguente definizione.

Definizione 6.1.1. *Chiamiamo*

$$\Delta(\Phi_1, \Phi_2) := d_{match}(\Phi_1, \Phi_2) - \max\{d_B(f_1, f_2), d_B(g_1, g_2)\}$$

la *differenza topologica tra Φ_1 e Φ_2* .

Anche tale quantità è riferita al grado p . Non lo indichiamo esplicitamente nella definizione poiché in seguito sarà sempre chiaro dal contesto a quale grado ci staremo riferendo.

Osservazione 6.1.2. *Osserviamo che, essendo la distanza di matching il massimo tra le distanze di bottleneck delle foliazioni di Φ_1 e Φ_2 , per l'Osservazione 3.3.7, la loro differenza topologica sarà sempre non negativa,*

$$\Delta(\Phi_1, \Phi_2) \geq 0.$$

6.1.2 Prime proprietà

La prima proprietà da verificare è che la differenza topologica sia nulla quando le due funzioni considerate sono uguali.

Proposizione 6.1.3. *Siano $\Phi_1 := (f_1, g_1), \Phi_2 := (f_2, g_2) : X \rightarrow \mathbb{R}^2$ due funzioni continue. Se $\Phi_1 = \Phi_2$ allora $\Delta(\Phi_1, \Phi_2) = 0$.*

Dimostrazione. Essendo per ipotesi $\Phi_1 = \Phi_2 = 0$, abbiamo che $d_{match}(\Phi_1, \Phi_2) = 0 = d_B(f_1, f_2) = d_B(g_1, g_2)$, da cui segue la tesi. \square

Ricordiamo che la distanza di bottleneck e la distanza di matching sono pseudometriche; in particolare, esistono funzioni filtranti diverse che inducono gli stessi diagrammi di persistenza. Pertanto, la seguente proposizione non è ridondante rispetto a quella appena dimostrata.

Proposizione 6.1.4. *Siano $\Phi_1 := (f_1, g_1), \Phi_2 := (f_2, g_2) : X \rightarrow \mathbb{R}^2$ due funzioni continue. Se $d_{match}(\Phi_1, \Phi_2) = 0$ allora $\Delta(\Phi_1, \Phi_2) = 0$.*

Dimostrazione. Per ipotesi, $d_{match}(\Phi_1, \Phi_2) = 0$. Da cui segue che anche $d_B(f_1, f_2) = 0 = d_B(g_1, g_2)$ e dunque abbiamo la tesi. \square

6.1.3 Esempi

Esempio 6.1.5 (Esempio 1). *Riprendiamo l'Esempio 4.1.1. Abbiamo dato una stima della distanza di matching tra Φ_1 ed F_1 nell'Esempio 3.3.10. Pertanto, $\Delta(\Phi_1, F_1) \geq 1 - \frac{\sqrt{2}}{4}$.*

È possibile dare una stima della differenza topologica negli esempi Esempio 2, Esempio 3 ed Esempio 4 in modo totalmente analogo.

6.2 Soluzione 2: correlazione topologica tra le componenti di una funzione

Nella sezione precedente, abbiamo illustrato uno strumento, la differenza topologica, che costituisce una prima risposta ad una delle tematiche del Problema 1, cioè come formalizzare e misurare il gap informativo tra omologia persistente biparametrica e monoparametrica. Ora vediamo come poter ulteriormente sviluppare tale strumento in modo da completare il quadro del Problema 1 e fornire anche dei primi contributi per il Problema 2.

6.2.1 Definizione

Il passo naturale da compiere successivamente è quello di sfruttare la definizione di differenza topologica appena data per definire il concetto di correlazione topologica tra le due componenti di una funzione continua a valori in \mathbb{R}^2 .

Siano X uno spazio topologico compatto e $\Phi := (f, g) : X \rightarrow \mathbb{R}^2$ una funzione continua. Possiamo considerare le seguenti funzioni, che risulteranno per costruzione continue a loro volta: $F := (f, f) : X \rightarrow \mathbb{R}^2$ e $G := (g, g) : X \rightarrow \mathbb{R}^2$. A questo punto, abbiamo tre funzioni biparametriche da poter confrontare: una è la funzione di partenza, mentre le altre due “rendono biparametriche” le due componenti f e g di Φ . Pertanto, risulta sensato calcolare le due differenze topologiche $\Delta(\Phi, F)$ e $\Delta(\Phi, G)$ e considerarne il massimo. Osserviamo che

$$\begin{aligned} \Delta(\Phi, F) &= d_{\text{match}}(\Phi, F) - \max\{d_B(f, f), d_B(g, f)\} = \\ &= d_{\text{match}}(\Phi, F) - d_B(g, f), \end{aligned}$$

e, analogamente,

$$\Delta(\Phi, G) = d_{\text{match}}(\Phi, G) - d_B(f, g).$$

La quantità $\max\{\Delta(\Phi, F), \Delta(\Phi, G)\}$ permette, per costruzione, di formalizzare e misurare il gap informativo tra multipersistenza e monopersistenza; per renderla meno grossolana e più facilmente interpretabile, rimane solo da

normalizzarla e fare in modo che risulti vicina a 0 quando non c'è correlazione e, viceversa, vicina a 1 quando c'è correlazione. Questo ci porta alla seguente definizione.

Definizione 6.2.1. *Definiamo la correlazione topologica tra le componenti f e g di Φ la quantità*

$$\Delta(\Phi) := \begin{cases} 1 - \frac{\max\{\Delta(\Phi, F), \Delta(\Phi, G)\}}{\Delta(\Phi, F) + \Delta(\Phi, G)} & \text{se } \Delta(\Phi, F) + \Delta(\Phi, G) \neq 0, \\ 1 & \text{altrimenti.} \end{cases}$$

Come vedremo nella prossima sezione, tanto più non c'è differenza tra caso monoparametrico e multiparametrico e tanto più le funzioni in gioco sono correlate tra loro.

6.2.2 Prime proprietà

Iniziamo a studiare la nozione appena definita verificando quando raggiunge i valori limite.

Per definizione, per avere correlazione uguale ad 1 occorre che sia $\Delta(\Phi, F)$ che $\Delta(\Phi, G)$ siano nulle; questo significa che non troveremo informazioni aggiuntive dalla persistenza multiparametrica di Φ rispetto alle due monoparametriche di F e G , come formalizzato nella seguente proposizione.

Proposizione 6.2.2.

$$\Delta(\Phi) = 1 \iff \begin{cases} d_{\text{match}}(\Phi, F) = d_B(f, g) \\ d_{\text{match}}(\Phi, G) = d_B(f, g) \end{cases}$$

Dimostrazione. Per definizione

$$\begin{aligned} \Delta(\Phi) = 1 &\iff \Delta(\Phi, F) + \Delta(\Phi, G) = 0 \\ &\iff \Delta(\Phi, F) = 0 = \Delta(\Phi, G) \\ &\iff \begin{cases} d_{\text{match}}(\Phi, F) = d_B(f, g) \\ d_{\text{match}}(\Phi, G) = d_B(f, g) \end{cases} \end{aligned}$$

□

Più precisamente, questo è equivalente a richiedere:

$$\begin{cases} \forall (a, b) \in [0, 1] \times \mathbb{R} & d_B(\Phi_{(a,b)}^*, F_{(a,b)}^*) = d_B(f, g) \\ \forall (a, b) \in [0, 1] \times \mathbb{R} & d_B(\Phi_{(a,b)}^*, G_{(a,b)}^*) = d_B(f, g). \end{cases}$$

È in questo senso che in tal caso la multipersistenza non ci dà ulteriori informazioni rispetto alla monopersistenza.

Inoltre, un caso particolare è il seguente.

Corollario 6.2.3. *Siano $f : X \rightarrow \mathbb{R}$ una funzione continua e $\Phi := (f, f)$. Allora $\Delta(\Phi) = 1$.*

Dimostrazione. Per costruzione, $F = G = \Phi = (f, f)$, perciò, per la Proposizione 6.1.3, abbiamo $\Delta(\Phi, F) = 0 = \Delta(\Phi, G)$, da cui segue la tesi \square

D'altra parte, per avere invece correlazione uguale ad 0 occorre che solo una tra $\Delta(\Phi, F)$ e $\Delta(\Phi, G)$ sia nulla; questo significa che troveremo informazioni aggiuntive dalla persistenza multiparametrica di Φ rispetto ad una delle due monoparametriche di F e G , ovvero quella che porta a differenza topologica non nulla, come formalizzato nella seguente proposizione.

Proposizione 6.2.4.

$$\Delta(\Phi) = 0 \iff \begin{cases} d_{match}(\Phi, G) = d_B(f, g) \\ d_{match}(\Phi, F) > d_B(f, g) \end{cases} \vee \begin{cases} d_{match}(\Phi, G) > d_B(f, g) \\ d_{match}(\Phi, F) = d_B(f, g) \end{cases}$$

Dimostrazione. Per definizione,

$$\begin{aligned} \Delta(\Phi) = 0 &\iff 1 - \frac{\max\{\Delta(\Phi, F), \Delta(\Phi, G)\}}{\Delta(\Phi, F) + \Delta(\Phi, G)} = 0 \\ &\iff \frac{\max\{\Delta(\Phi, F), \Delta(\Phi, G)\}}{\Delta(\Phi, F) + \Delta(\Phi, G)} = 1 \\ &\iff \frac{\Delta(\Phi, F)}{\Delta(\Phi, F) + \Delta(\Phi, G)} = 1 \text{ oppure } \frac{\Delta(\Phi, G)}{\Delta(\Phi, F) + \Delta(\Phi, G)} = 1 \\ &\iff \Delta(\Phi, G) = 0, \Delta(\Phi, F) \neq 0 \text{ oppure } \Delta(\Phi, F) = 0, \Delta(\Phi, G) \neq 0. \end{aligned}$$

Per semplicità, portiamo avanti il primo caso, cioè $\Delta(\Phi, G) = 0, \Delta(\Phi, F) \neq 0$ (l'altro caso risulta analogo):

$$\Delta(\Phi, G) = 0, \Delta(\Phi, F) \neq 0 \iff \begin{cases} d_{match}(\Phi, G) = d_B(f, g) \\ d_{match}(\Phi, F) > d_B(f, g) \end{cases}$$

\square

Più precisamente, questo è equivalente a richiedere:

$$\begin{cases} \forall (a, b) \in [0, 1] \times \mathbb{R} & d_B(\Phi_{(a,b)}^*, G_{(a,b)}^*) = d_B(f, g) \\ \exists (a, b) \in [0, 1] \times \mathbb{R} \text{ t.c. } & d_B(\Phi_{(a,b)}^*, F_{(a,b)}^*) > d_B(f, g) \end{cases}$$

ed è in questo senso che in tal caso la persistenza multiparametrica risulta più informativa di quella monoparametrica.

6.2.3 Esempi

Vediamo ora alcuni esempi di correlazione topologica tra le componenti di una funzione.

Esempio 6.2.5 (Esempio 1). *Riprendiamo l'Esempio 6.1.5. Abbiamo visto che $d_B(f_1, g_1) = d_B(f_1, f_1) = d_B(g_1, g_1) = 0$, pertanto:*

$$\begin{aligned}\Delta(\Phi_1, F_1) &= d_{\text{match}}(\Phi_1, F_1) - \max\{d_B(f_1, f_1), d_B(g_1, f_1)\} = \\ &= d_{\text{match}}(\Phi_1, F_1) = \\ &= \max_{a \in [0,1], b \in \mathbb{R}} \{d_B(\Phi_{1,(a,b)}^*, F_{1,(a,b)}^*)\},\end{aligned}$$

e, analogamente,

$$\Delta(\Phi_1, G_1) = \max_{a \in [0,1], b \in \mathbb{R}} \{d_B(\Phi_{1,(a,b)}^*, G_{1,(a,b)}^*)\}.$$

In questo caso, per ogni $a \in [0, 1]$ e $b \in \mathbb{R}$ si ha che $F_{1,(a,b)}^*$ e $G_{1,(a,b)}^*$ hanno gli stessi diagrammi di persistenza (per lo stesso motivo per cui li hanno f e g , come visto nell'Esempio 3.2.23). Pertanto, risulta:

$$\Delta(\Phi_1, F_1) = \Delta(\Phi_1, G_1),$$

da cui segue che

$$\Delta(\Phi_1) = 1 - \frac{\max\{\Delta(\Phi_1, F_1), \Delta(\Phi_1, G_1)\}}{\Delta(\Phi_1, F_1) + \Delta(\Phi_1, G_1)} = 1 - \frac{\Delta(\Phi_1, F_1)}{2\Delta(\Phi_1, F_1)} = \frac{1}{2}.$$

Osservazione 6.2.6. *Come anticipato nell'Osservazione 5.2.5, l'Esempio 1 e l'Esempio 5.2.4 sono due riscritture di una stessa situazione.*

Nel linguaggio dell'Esempio 5.2.4 è stato possibile confrontare le funzioni sin e cos tramite la cross correlation. In particolare, abbiamo visto che: sin e sin sono totalmente sovrapponibili e hanno infatti cross-correlation massima (pari a 1), mentre sin e cos sono sfasate e infatti hanno cross-correlation nulla. In modo analogo, sappiamo che, per il Corollario 6.2.3, $\Delta(F_1) = 1 = \Delta(G_1)$, mentre abbiamo visto nell'Esempio 6.2.5 che $\Delta(\Phi_1) = \frac{1}{2}$; possiamo perciò affermare che il concetto di correlazione topologica sia compatibile con il concetto di cross-correlation. Il fatto che $\Delta(\Phi_1)$ non sia nullo, ma solo un valore intermedio rispetto all'intervallo dei possibili valori, può essere dovuto al fatto che al momento la definizione di Δ è "grossolana", essendo qui presentata nella sua prima formulazione; sicuramente, raffinandola¹ si potrà ottenere valore 0 anche per $\Delta(\Phi_1)$. Quello che importa è che si abbia valore massimo per funzioni sovrapponibili e un valore "a metà" dell'intervallo di valori considerato per funzioni sfasate, così come avviene per la cross-correlation: nel primo caso, il range di valori possibili è $[0, 1]$, mentre nel secondo è $[-1, 1]$.

¹Rimandiamo per questo al Capitolo 7.

Esempio 6.2.7 (Esempio 2). *Procedendo come nell'Esempio 6.2.5 si calcola nuovamente $\Delta(\Phi_2) = \frac{1}{2}$.*

Esempio 6.2.8 (Esempio 3). *Procedendo come nell'Esempio 6.2.5 si calcola nuovamente $\Delta(\Phi_3) = \frac{1}{2}$.*

6.3 Soluzione 2: correlazione topologica tra collezioni di funzioni

Siamo finalmente giunti alla definizione cruciale di questo lavoro di Tesi di Laurea. Abbiamo appena dato una definizione di correlazione tra le componenti di una funzione a valori in \mathbb{R}^2 ; in particolare, tali componenti risultano definite sullo stesso dominio. Noi vorremmo invece riuscire a confrontare funzioni definite su domini diversi, come abbiamo visto nel Capitolo 5. A tal fine, entra in gioco la prossima definizione.

6.3.1 Definizione

Consideriamo un numero finito N di spazi topologici compatti X_1, \dots, X_N e due collezioni di funzioni $f := \{f_i\}_{i=1, \dots, N}$ e $g := \{g_i\}_{i=1, \dots, N}$, dove per ogni $i = 1, \dots, N$ abbiamo $f_i : X_i \rightarrow \mathbb{R}$ e $g_i : X_i \rightarrow \mathbb{R}$ continue. Per poter utilizzare le definizioni precedenti, consideriamo le funzioni $\Phi_i := (f_i, g_i) : X_i \rightarrow \mathbb{R}^2$; in questo modo, possiamo ottenere una collezione di valori $\{\Delta(\Phi_i)\}_{i=1, \dots, N}$ e, considerandone la media, una prima definizione di correlazione tra f e g .

Definizione 6.3.1. *Definiamo la **correlazione tra f e g** come la media aritmetica di $\{\Delta(\Phi_i)\}$, cioè $\text{corr}(f, g) := \frac{1}{N} \sum_{i=1}^N \Delta(\Phi_i)$.*

Questa definizione giustifica anche la scelta della Notazione 4.2.1, infatti nella prossima sezione considereremo collezioni di funzioni a partire dagli esempi Esempio 1, Esempio 2 ed Esempio 3.

6.3.2 Esempi

Vediamo subito un esempio riguardante la definizione appena introdotta.

Esempio 6.3.2 (Esempio 1 + Esempio 3). *Siano $X_1 := S^1$ e $X_3 := I^2$ e consideriamo le funzioni “coordinata x ” e “coordinata y ” (o loro combinazioni), rappresentate dalle collezioni $f := \{f_1, f_2\}$ e $g := \{g_1, g_3\}$. Poniamo inoltre $\Phi_i := (f_i, g_i)$. Abbiamo visto negli esempi Esempio 6.2.5 ed Esempio 6.2.7 che $\Delta(\Phi_1) = \frac{1}{2} = \Delta(\Phi_3)$, perciò $\text{corr}(f, g) = \frac{1}{2}$.*

Già da questo semplice esempio è pertanto evidente il potenziale della definizione di correlazione topologica: essa permette di confrontare funzioni definite su domini diversi. In realtà è possibile generalizzare ulteriormente: nell'esempio appena descritto, comunque gli spazi considerati si immergono

in modo minimale \mathbb{R}^2 . Invece nel prossimo esempio considereremo la circonferenza e il toro unitari, che si immergono in modo minimale rispettivamente in \mathbb{R}^2 ed \mathbb{R}^3 .

Esempio 6.3.3 (Esempio 1 + Esempio 2). *Siano $X_1 := S^1$ e $X_2 := T$ e consideriamo le funzioni “coordinata x ” e “coordinata y ”, rappresentate dalle collezioni $f := \{f_1, f_2\}$ e $g := \{g_1, g_2\}$ dove $f_i : X_i \rightarrow \mathbb{R} \mid P \mapsto P_x$. Poniamo inoltre $\Phi_i := (f_i, g_i)$. Abbiamo visto negli esempi Esempio 6.2.5 ed Esempio 6.2.7 che $\Delta(\Phi_1) = \frac{1}{2} = \Delta(\Phi_2)$, perciò $\text{corr}(f, g) = \frac{1}{2}$.*

L’idea di utilizzo di tale definizione è la seguente. Ispirandoci al problema presentato nell’Esempio 5.2.2, consideriamo:

- al posto di S^1 , I^2 e T , tre superfici molecolari;
- al posto delle collezioni di funzioni “coordinata x ” e “coordinata y ”, le collezioni di funzioni “densità di carica” e “idrofobicità”.

Allora, calcolando $\text{corr}(f, g)$, riusciamo a capire quanto influisca una proprietà chimico-fisica sull’altra.

Capitolo 7

Conclusioni e sviluppi futuri

In questo lavoro di Tesi, abbiamo analizzato lo Stato dell'Arte della teoria dell'omologia persistente, per arrivare a presentare e studiare due problemi apparentemente scollegati tra loro, i quali invece sono strettamente legati. Da un lato abbiamo messo in luce come ci sia un'evidente differenza tra la capacità discriminante dell'omologia persistente multiparametrica e monoparametrica, ma non esistano ancora in letteratura né formalizzazioni né motivazioni di tale comportamento (Problema 1). Dall'altro abbiamo mostrato che, soprattutto nelle applicazioni, potrebbe essere utile definire un concetto di correlazione tra funzioni, ma che al momento tale concetto non esiste, almeno nella massima generalità possibile (Problema 2). Abbiamo poi visto come l'omologia persistente possa fornirci gli strumenti per rispondere, a un tempo, ad entrambe le domande. L'idea alla base di tutto il lavoro è stata la seguente: più due funzioni f e g a valori reali sono correlate e meno gap informativo è presente tra la filtrazione biparametrica indotta da $\Phi := (f, g)$ e le due filtrazioni monoparametriche indotte da f e g . Questo ci ha portato a dare due definizioni: la differenza topologica tra due funzioni a valori reali, che costituisce una prima risposta al Problema 1, e la correlazione topologica tra due funzioni a valori reali e tra due collezioni di funzioni reali, che costituisce una prima risposta al Problema 2 e che formalizza proprio l'idea appena descritta. Infine, abbiamo iniziato ad indagare le proprietà delle definizioni date e le abbiamo applicate in alcuni esempi, che danno prime e significative conferme del fatto che esse realizzino le idee intuitive che ci hanno portato alla loro formulazione.

In futuro sarebbe sicuramente interessante portare avanti lo studio di ulteriori proprietà e applicare tali concetti a più esempi, aumentandone via via la complessità e la natura pratica, in modo da adattare le definizioni per renderle il più possibile efficaci. Le definizioni attuali di differenza topologica e di correlazione topologica tra due funzioni a valori reali sono basate su strumenti topologici propri della topologia algebrica e dell'Analisi Dati Topologica, in particolar modo sulla *distanza di bottleneck* e la *distanza*

di matching, mentre l'attuale definizione di correlazione topologica tra due collezioni di funzioni prevede il calcolo di una media di varie differenze topologiche, e dunque coinvolge anche uno strumento statistico. Probabilmente è possibile rendere più precise, mirate ed efficaci tutte le definizioni, verificando sia l'impiego di differenti strumenti topologici che quello di differenti strumenti statistici. Ad esempio, una delle prime operazioni da fare sarebbe provare a sostituire, nella definizione di correlazione topologica tra due collezioni di funzioni, la media con diversi strumenti statistici più raffinati e confrontare i risultati ottenuti per stabilire quale sia la miglior definizione possibile di correlazione topologica. Inoltre, la definizione attuale di correlazione topologica prevede valori in $[0, 1]$; per avere una totale generalizzazione, bisognerebbe estendere l'intervallo di possibili valori a $[-1, 1]$. Infine, al momento la definizione di correlazione topologica tra due funzioni non tiene conto dell'estensione del dominio e del codominio: ad esempio la correlazione tra $f : S^1 \rightarrow \mathbb{R} \mid (x, y) \mapsto x$ e $g : S^1 \rightarrow \mathbb{R} \mid (x, y) \mapsto 2x$ secondo la definizione attuale non è massima, mentre ci aspetteremmo che lo fosse.

Durante lo studio e la scrittura necessari per questo lavoro di Tesi, abbiamo inoltre individuato ulteriori strade percorribili per portare avanti la ricerca. Le presentiamo divise per dominio di interesse.

7.1 Sviluppi topologici

Durante il lavoro di ricerca, abbiamo intrapreso anche uno studio dei problemi basato sulla sequenza di Mayer-Vietoris. Per chiare ragioni di tempo (ad un certo punto, purtroppo o per fortuna, bisogna laurearsi) non siamo riusciti ad indagare a fondo questo approccio, ma sicuramente porterebbe a sviluppi interessanti. In topologia algebrica, la sequenza di Mayer-Vietoris permette di studiare l'omologia di uno spazio topologico X dividendolo in sottospazi la cui omologia è più facile da calcolare; perciò in aggiunta consente anche di studiare la relazione tra omologia globale e locale di X . Per questo, in questi anni sono stati cercati risultati analoghi per l'omologia persistente ([10, 1, 20]). L'idea che abbiamo solamente iniziato ad esplorare per il momento si basava solamente sulla classica versione di Mayer-Vietoris e sicuramente l'utilizzo della sequenza di Mayer-Vietoris per l'omologia persistente potrebbe fornire ulteriori interessanti spunti.

Osserviamo inoltre che questa trattazione ha preso in considerazione solamente funzioni continue definite su spazi topologici compatti. Pertanto, sarebbe rilevante generalizzare lo studio a spazi e funzioni più generiche.

7.2 Sviluppi algebrici

Nel Capitolo 2, abbiamo presentato i moduli di persistenza oggetti di natura categoriale e algebrica. In particolare, concentrandoci sulla natura algebrica

dei moduli di persistenza come moduli graduati su anelli di polinomi è possibile studiarli tramite i classici strumenti dell'algebra commutativa, come le funzioni di Hilbert e le risoluzioni minimali.

Un ulteriore passo potrebbe consistere nella definizione di operazioni algebriche tra moduli di persistenza, a partire dallo studio di quelle già esistenti tra moduli graduati. Data $\Phi := (f, g) : X \rightarrow \mathbb{R}^2$, abbiamo tre moduli di persistenza:

$$PH(f) \quad PH(g) \quad PH(\Phi).$$

Se potessimo considerare un prodotto di moduli di persistenza \star , potremmo capire quando $PH(f) \star PH(g) = PH(\Phi)$, ovvero avremmo un modo algebrico per stabilire quando la persistenza multiparametrica indotta da Φ risulta più informativa rispetto a quella monoparametrica indotta da f e g , e dunque anche un nuovo modo per studiare la correlazione tra f e g .

7.3 Sviluppi geometrici

Un'altra possibilità potrebbe essere quella di esplorare come strumenti provenienti dalla geometria algebrica, differenziale e complessa possano essere adattati a questo contesto. Ricollegandoci al discorso sulla possibilità di definire un prodotto tra moduli di persistenza, si potrebbe trarre ispirazione dalle classi caratteristiche per costruire uno strumento simile nell'ambito della persistenza; questo perché esse, tra le altre cose, permettono di stabilire se un fascio (o un fibrato) sia banale, cioè sia un prodotto di fasci (o fibrati).

Inoltre potrebbe essere interessante, soprattutto nell'ottica di confrontare funzioni definite su domini diversi, indagare se sia possibile definire morfismi di filtrazioni. Ad esempio, se abbiamo due spazi topologici compatti X_1, X_2 e due funzioni filtranti definite su di essi $\Phi_1 : X_1 \rightarrow \mathbb{R}^d, \Phi_2 : X_2 \rightarrow \mathbb{R}^d$, sotto quali ipotesi esiste una funzione $\alpha : X_1 \rightarrow X_2$ continua tale che $\forall p \in \mathbb{Z}, \forall u \in \mathbb{R}^d \quad H_p(X_{\Phi_2 \circ \alpha}) \cong H_p(X_{\Phi_1}^u)$ e vengano preservate le inclusioni tra gli insiemi di sottolivello? In questo modo si possono considerare le filtrazioni indotte da Φ_1 e $\Phi_2 \circ \alpha$ che risultano entrambe definite su X_1 .

$$\begin{array}{ccc} X_1 & \searrow \Phi_1 & \\ \downarrow \alpha & & \mathbb{R}^d \\ X_2 & \nearrow \Phi_2 & \end{array}$$

Si tratta di richieste troppo forti per ottenere risultati utili e significativi nella pratica? Le condizioni possono essere indebolite? Tale α è unica? Esiste una sua definizione costruttiva? Secondo me si tratta di riflessioni molto

naturali, sebbene non sembrano essere mai state affrontate in letteratura, e rispondere a tali domande può rivelarsi molto proficuo.

7.4 Sviluppi computazionali e implementativi

Infine, ci sono numerosi aspetti computazionali ed implementativi che sarebbe significativo studiare, nonché necessario in vista di applicazioni future. Ne elenchiamo alcuni, divisi per argomenti:

- Persistenza multiparametrica:
 - Confronto tra omologia multipersistente e monopersistente: eseguire dei benchmark su vari tipi di dataset per valutare se la maggior capacità discriminante dell’approccio multiparametrico si realizzi solo in contesti “teorici” o anche nella pratica.
 - Condurre esperimenti per valutare in quali contesti la capacità discriminante e caratterizzante dell’omologia multipersistente risultino migliori di quelle dei più comuni metodi di learning.
- Calcolo della matching distance e di Δ :
 - Rendere efficiente il calcolo della distanza di matching, tenendo conto che dalla teoria sappiamo che essa si realizza in un numero limitato di casistiche per i valori (a, b)
 - Sviluppo di un calcolo efficiente di Δ sulla base delle implementazioni portate avanti nel punto precedente.
 - Estendere il calcolo della distanza di matching e di Δ al caso 3D.
- Calcolo della correlazione $corr(f, g)$:
 - Sulla base dei punti precedenti, sviluppare e implementare un algoritmo per il calcolo di $corr(f, g)$.
 - Portare avanti molti test per valutare la correttezza della definizione di $corr(f, g)$, ed eventualmente migliorarla utilizzando ulteriori strumenti statistici.
 - valutare l’efficacia di $corr(f, g)$ nella riduzione dimensionale - cioè nella riduzione del numero di funzioni che descrivono un problema/fenomeno - confrontando i risultati ottenuti con altri approcci probabilistico-statistici e di learning.
 - Validare i nostri strumenti in svariati contesti applicativi (ad esempio, studi di compatibilità tra molecole).

7.5 Possibili applicazioni

L'applicazione più importante dei concetti introdotti con questo lavoro di Tesi è quella della riduzione dimensionale di problemi. Come presentato nel Capitolo 5, date alcune collezioni di funzioni, ciascuna delle quali rappresentante una proprietà di interesse per il problema studiato (ad esempio, nell'Esempio 5.2.2 le collezioni di funzioni rappresentano proprietà fisico-chimiche delle molecole considerate), tramite la definizione di correlazione topologica introdotta poc'anzi è possibile stabilire quali tra tali collezioni non aggiungono informazioni rilevanti per il problema. La finalità più ambiziosa sarebbe proprio quella di sfruttare le nozioni presentate in questo lavoro di Tesi per sviluppare metodi e algoritmi di ricerca di "basi" di funzioni per i problemi considerati.

Sarebbe inoltre sicuramente interessante e utile integrare tali teorie in algoritmi di machine learning e confrontarli con le classiche tecniche di machine learning; in particolare, sarebbe significativo condurre questo confronto nell'ambito della shape detection e recognition.

Probabilmente portando avanti la ricerca e migliorando le attuali definizioni sarà possibile individuare anche ulteriori contesti applicativi adatti all'utilizzo di metodi basati su queste teorie.

Appendice A

Omologia persistente di complessi simpliciali

Specialmente lavorando nel contesto dell'Analisi Dati Topologica, è molto probabile imbattersi in dataset descritti tramite complessi simpliciali. La teoria della persistenza presentata in questo lavoro di Tesi, fortunatamente, trova il suo corrispondente in questo contesto. Pertanto, in questa appendice generalizzeremo i concetti affrontati nel caso continuo al contesto dei complessi simpliciali.

A.1 Complessi simpliciali

Per prima cosa, vediamo cosa intendiamo per *complesso simpliciale*. In generale, con la parola “complesso” si intende una collezione di oggetti della stessa tipologia e che soddisfino certe proprietà; nel caso dei complessi *simpliciali*, gli oggetti coinvolti saranno per l'appunto i semplici. Abbiamo già visto nel Capitolo 1 una possibile definizione del concetto di semplice (standard), nella Definizione 1.1.1. Ora lavoreremo con una definizione leggermente più generica e ci addenteremo maggiormente nei dettagli.

Definizione A.1.1. *Per $m, n \in \mathbb{N}$ con $n \leq m$, un n -simpleso in \mathbb{R}^m è l'involuppo convesso di un insieme P di $n + 1$ punti di \mathbb{R}^m linearmente indipendenti.*

La differenza con i semplici standard della Definizione 1.1.1 è solamente nella scelta dei punti a partire dai quali si costruisce l'involuppo convesso: in quel caso erano gli elementi della base canonica di \mathbb{R}^n , ed è a questo che devono l'etichetta di semplici *standard*.

Diciamo che la dimensione di un n -simpleso è proprio n . Inoltre, dato σ ¹

¹Attenzione: nel Capitolo 1 indicavamo con Δ_n gli n -semplici standard e con σ i semplici singolari. Invece qui indichiamo con σ generici n -semplici di \mathbb{R}^m .

un n -simpleso, per ogni $n' \in \mathbb{N}$ con $0 \leq n' \leq n$, una **faccia** di σ di dimensione n' è un n' -simpleso dato dall'involuppo convesso di un sottoinsieme non vuoto di P . Una faccia è detta *propria* se è l'involuppo convesso di un sottoinsieme *proprio* di P . Le facce di dimensione $n - 1$ di σ sono chiamate **facce** di σ e sono $n + 1$ facce.

Definizione A.1.2. Un **complesso simpliciale** (geometrico) è un insieme, al più numerabile, K di semplici che soddisfi le seguenti proprietà:

1. K contiene ogni faccia di ciascun suo semplice;
2. per ogni coppia di semplici $\sigma, \tau \in K$, la loro intersezione $\sigma \cap \tau$ può essere o vuota oppure una faccia sia di σ che di τ .

Esempio A.1.3. In Figura A.1 e in Figura A.2 sono rappresentati due complessi simpliciali. Invece in Figura A.3 è rappresentata una collezione di semplici che non costituiscono un complesso simpliciale.

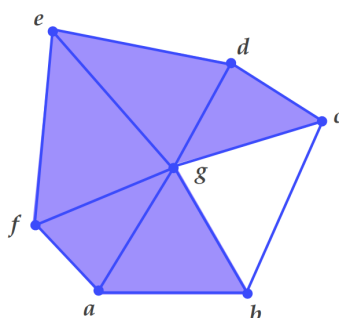


Figura A.1: Esempio di complesso simpliciale. Immagine tratta da [16].

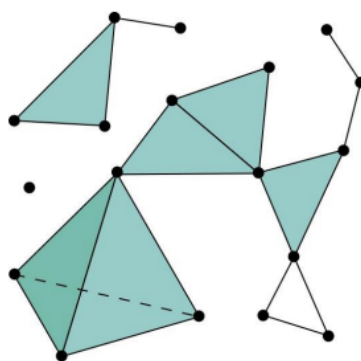


Figura A.2: Esempio di complesso simpliciale. Immagine tratta da [16].

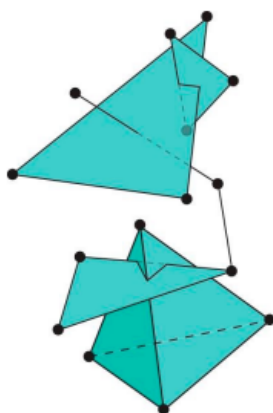


Figura A.3: Collezione di semplici che non costituisce un complesso simpliciale. Immagine tratta da [16].

Nella proprietà 2. della Definizione A.1.2, stiamo chiedendo che i semplici che costituiscono K si possano intersecare solo tramite facce. Se K è finito, la **dimensione** di K è la massima tra le dimensioni dei suoi semplici.

I complessi simpliciali geometrici sono realizzazioni geometriche di un concetto più astratto di complesso simpliciale — i *complessi simpliciali astratti*, definiti tramite generici insiemi e non a partire da involucri convessi di punti in \mathbb{R}^m [9] — ma si dimostra che quelli geometrici e quelli astratti sono in corrispondenza; perciò parleremo semplicemente di complessi simpliciali.

A.2 Omologia simpliciale

In questa sezione ripercorreremo a grandi linee quanto visto nel Capitolo 1 per la definizione dell'omologia singolare, per arrivare a definire l'*omologia simpliciale*.

Sia K un complesso simpliciale finito di dimensione n . Per ogni $p \in \mathbb{N}$ con $0 \leq p \leq n$, denotiamo con m_p il numero dei p -simplessi di K . In modo totalmente analogo rispetto a quanto visto per i simplessi singolari, definiamo un complesso di catene per K .

Definizione A.2.1. Una *p -catena* in K è una somma formale $\sum_{i=1}^{m_p} k_i \sigma_i$ con $k_i \in \mathbb{Z}$ e σ_i un p -simpleso di K . In altre parole, le p -catene di K sono gli elementi del gruppo abeliano libero generato dai p -simplessi di K , che denotiamo con $C_p(K)$.

Inoltre, se $p \geq 1$, per ogni p -simpleso σ di vertici $\{v_0, \dots, v_p\}$ possiamo definire

$$\partial_p^K \sigma := \sum_{i=0}^p [v_0, \dots, \hat{v}_i, \dots, v_p],$$

dove con $[v_0, \dots, \hat{v}_i, \dots, v_p]$ denotiamo il $(p-1)$ -simpleso generato dai vertici di σ ad eccezione di v_i . Estendendo in modo \mathbb{Z} -lineare ∂_p alle p -catene di K , otteniamo un omomorfismo $\partial_p^K : C_p(K) \rightarrow C_{p-1}(K)$ detto **operatore di bordo**.

Invece, per $p = 0$, poniamo $C_{-1}(K) := \mathbf{0}$ e $\partial_0^K := \mathbf{0}$.

Nuovamente, come per il Lemma 1.1.6, si dimostra che la composizione dell'operatore di bordo con se stesso è nulla e pertanto si ottiene un complesso di catene $C.(X) := \{(C_p(K), \partial_p)\}$.

Come visto nel caso singolare, a questo punto si definiscono il **gruppo dei p -cicli** $Z_p(K) := \ker \partial_p$ e **dei p -bordi** $B_p(K) := \text{im } \partial_{p+1}$ e si definisce il **p -esimo gruppo di omologia** di K come il loro quoziente $H_p(K) := Z_p(K)/B_p(K)$, ben definito grazie al fatto che la composizione $\partial_{p+1} \circ \partial_p = 0$.

Osservazione A.2.2. *Si può dimostrare che se X è uno spazio topologico triangolabile e la sua triangolazione è data da un complesso singolare K , allora K e X sono omeomorfi. In particolare, hanno stessa omologia, cioè per ogni $p \geq 0$ $H_p(X) \cong H_p(K)$, dove a sinistra abbiamo il p -esimo gruppo di omologia singolare di X , mentre a destra abbiamo il p -esimo gruppo di omologia simpliciale di K .*

Inoltre, due differenti triangolazioni di X danno origine a omologie simpliciali isomorfe.

A.3 Filtrazioni di complessi simpliciali e moduli di omologia simpliciale persistente

Anche nel contesto dei complessi simpliciali è possibile parlare di filtrazioni e, conseguentemente, di omologia persistente. Per semplicità, tratteremo solo il caso monoparametrico, ma anche il caso multiparametrico si adatta in modo analogo.

Sia K un complesso simpliciale.

Assunzione A.3.1. *Anche in questo contesto, assumiamo di lavorare nelle condizioni descritte dalle assunzioni Assunzione 2.1.2, Assunzione 2.2.2 e Assunzione 3.1.2, dove la richiesta di compattezza per gli spazi topologici considerati viene sostituita dalla richiesta di finitezza dei complessi simpliciali considerati.*

Definizione A.3.2. Una **filtrazione** di K è una sequenza concatenata di suoi sottocomplessi, cioè una sequenza del tipo

$$\emptyset = K_0 \hookrightarrow K_1 \hookrightarrow \dots \hookrightarrow K_n = K,$$

dove n è un intero positivo e $K_1, \dots, K_n = K$ sono sottocomplessi di K tali che $K_1 \subseteq \dots \subseteq K_n = K$.

Esempio A.3.3. Possiamo osservare un esempio di filtrazione di un complesso simpliciale in Figura A.4.

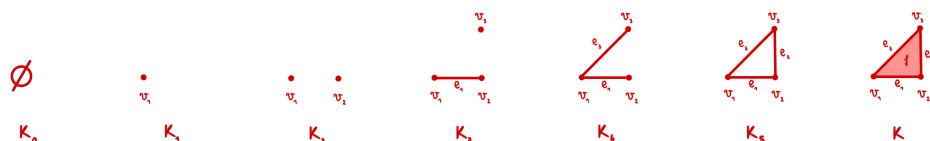


Figura A.4: Filtrazione di un complesso simpliciale costituito da un 2-simplesso e tutte le sue facce.

Anche in questo contesto è possibile indurre filtrazioni a partire da funzioni, che questa volta saranno definite simplex-wise (cioè simplessi per simplessi) ed inoltre dovranno garantire che gli insiemi di sottolivello siano sottocomplessi contenuti l'uno nell'altro, proprietà che non risulta immediata in questo contesto.

Le funzioni che garantiscono tale condizione sono le *funzioni monotone simplex-wise*, definite come segue.

Definizione A.3.4. Sia $f : K \rightarrow \mathbb{R}$ una funzione definita simplex-wise. Diciamo che f è **monotona simplex-wise** se e solo se comunque scelti σ e σ' due simplessi contenuti in K tali che $\sigma \subseteq \sigma'$ allora $f(\sigma) \leq f(\sigma')$.

Tale proprietà assicura che, per ogni $u \in \mathbb{R}$ fissato, $f^{-1}((-\infty, u])$ sia un sottocomplesso di K . In particolare, considerando ad esempio i valori critici $a_1, \dots, a_n \in \mathbb{R}$ e ponendo per ogni $i \in \{1, \dots, n\}$ $K_i := f^{-1}((-\infty, a_i])$ e $K_0 = \emptyset$, otteniamo una filtrazione simpliciale

$$\emptyset = K_0 \hookrightarrow K_1 \hookrightarrow \dots \hookrightarrow K_n = K.$$

In tal caso, chiamiamo f una **funzione filtrante** per K .

Esempio A.3.5. Esempio funzione monotona simplex-wise e corrispondente filtrazione indotta.

Un modo pratico per definire funzioni monotone simplex-wise consiste nel costruirle a partire dalle cosiddette *vertex function*.

Definizione A.3.6. Una **vertex function** su K è una funzione a valori reali definita sui suoi vertici, cioè $f : V(K) \rightarrow \mathbb{R}$.

Vediamo ora come possiamo definire una funzione monotona simplex-wise a partire da una vertex function f . L'idea è quella di estendere tale funzione dapprima agli archi, successivamente alle facce e così via, salendo ogni volta di dimensione. Per ogni arco σ poniamo $f(\sigma)$ il massimo valore che f assume sui suoi vertici; dopodiché, per ogni faccia σ poniamo $f(\sigma)$ il massimo valore che f assume sui suoi archi. In generale, per ogni k -simpleso σ in K , poniamo $f(\sigma)$ il massimo valore che f assume sui $(k-1)$ -simplessi contenuti in σ .

Esempio A.3.7. Dato il complesso simpliciale K dell'esempio A.3.3, definiamo la seguente vertex function $f : V(K) \rightarrow \mathbb{R}$:

$$f(v_1) = -1, \quad f(v_2) = 0, \quad f(v_3) = 1.$$

La funzione monotona simplex-wise che si ottiene a partire da f è quella rappresentata in figura A.5.

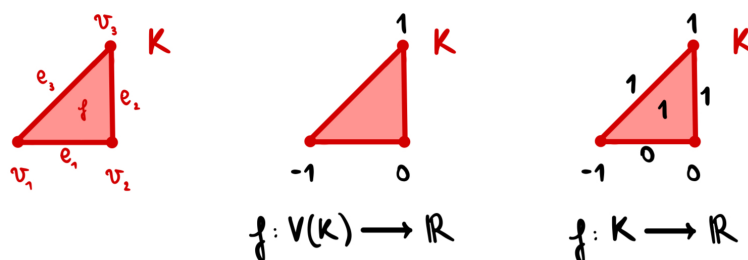


Figura A.5: Da sinistra a destra: complesso simpliciale K , vertex function f dell'esempio A.3.7, funzione monotona simplex-wise indotta dalla vertex function f .

È possibile definire filtrazioni a partire direttamente da vertex function, senza passare tramite le loro estensioni monotone simplex-wise.

Sia $f : V(K) \rightarrow \mathbb{R}$ una vertex function su K . Esplicitiamo l'insieme dei vertici di K : $V(K) = \{v_1, v_2, \dots, v_N\}$ per qualche $N \in \mathbb{N}_{\geq 1}$, ordinati in modo tale da avere $f(v_1) \leq f(v_2) \leq \dots \leq f(v_N)$.

Definizione A.3.8. Sia $v \in V(K)$. Il **lower star** di v in K è l'insieme $Lst(v)$ dei simplessi in $St(v)$ i cui vertici, ad esclusione di v stesso, siano minori di v rispetto all'ordine fissato su $V(K)$. In modo simmetrico, l'**upper star** di v in K è l'insieme $Ust(v)$ dei simplessi in $St(v)$ i cui vertici, ad esclusione di v stesso, siano maggiori di v rispetto all'ordine fissato su $V(K)$.

È ora possibile definire due filtrazioni a partire da f .

Definizione A.3.9. Per ogni $i = 1, \dots, N$, denotiamo con $K_{f(v_i)}$ il sotto-simpleso di K generato dai vertici v_1, \dots, v_i e aggiungiamo un vertice v_0 tale che $f(v_0) = -\infty$. Otteniamo così la seguente filtrazione

$$\emptyset = K_{f(v_0)} \hookrightarrow K_{f(v_1)} \hookrightarrow \dots \hookrightarrow K_{f(v_n)} = K,$$

dove per ogni $i = 1, \dots, n$ si ha $K_{f(v_i)} \setminus K_{f(v_{i-1})} = Lst(v_i)$ e per questo viene chiamata **filtrazione lower star** indotta da f .

Analogamente, considerando i vertici in ordine invertito e aggiungendo un vertice v_{N+1} tale che $f(v_{N+1}) = -\infty$, otteniamo la **filtrazione upper star** indotta da f , cioè la filtrazione

$$\emptyset = K_{f(v_{N+1})} \hookrightarrow K_{f(v_N)} \hookrightarrow \dots \hookrightarrow K_{f(v_1)} = K,$$

dove per ogni $i = 1, \dots, N$ si ha $K_{f(v_i)} \setminus K_{f(v_{i+1})} = Ust(v_i)$.

Anche in questo caso, f prende il nome di **funzione filtrante**.

Esempio A.3.10. In Figura A.6 e in Figura A.7 sono rappresentati alcuni esempi dei concetti appena introdotti.

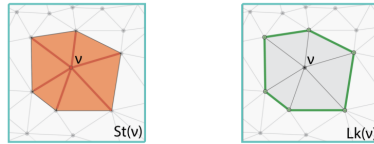


Figura A.6: Esempio del lower star e il lower link di un 2-simplesso e tutte le sue facce. Immagine tratta da [16]

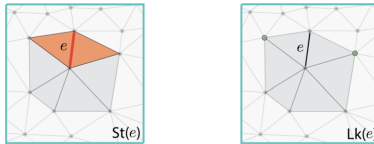


Figura A.7: Esempio del lower star e il lower link di un arco, cioè un 1-simplesso. Immagine tratta da [16]

Se una filtrazione simpliciale viene presentata senza esplicitare una funzione che la induca, possiamo comunque costruire una funzione filtrante monotona simplex-wise per tale filtrazione.

Definizione A.3.11. Sia $\emptyset = K_0 \hookrightarrow K_1 \hookrightarrow \dots \hookrightarrow K_n = K$ una filtrazione. Definiamo $f : K \rightarrow \mathbb{R}$ monotona simplex-wise come segue: per ogni $i = 1, \dots, n$ tale che $K_i \neq K_{i-1}$ e per ogni simplesso σ in $K_i \setminus K_{i-1}$ poniamo $f(\sigma) = i$.

In questo caso la **funzione filtrante** della filtrazione è f .

Esempio A.3.12. Consideriamo la filtrazione dell'esempio A.3.3, che non è indotta da alcuna funzione in modo esplicito. Costruiamo una funzione filtrante per tale filtrazione. Poniamo: $f(v_1) = 1, f(v_2) = 2, f(v_3) = f(e_1) = 3, f(e_3) = 4, f(e_2) = 5, f(K) = 6$.

Definizione A.3.13. Anche nel contesto simpliciale, una filtrazione $\emptyset = K_0 \hookrightarrow K_1 \hookrightarrow \dots \hookrightarrow K_n = K$ induce una sequenza in omologia (simpliciale):

$$\mathbf{0} = H_p(K_0) \longrightarrow H_p(K_1) \longrightarrow \dots \longrightarrow H_p(K_n), \quad (\text{A.1})$$

che prende il nome di **modulo di omologia persistente simpliciale**.

Come nel caso continuo, noi lavoreremo con l'omologia a coefficienti in $\mathbb{Z}/2\mathbb{Z}$ e dunque i gruppi di omologia saranno dei campi vettoriali su $\mathbb{Z}/2\mathbb{Z}$.

In questo contesto, le definizioni riguardanti i gruppi di omologia persistente, la nascita e la morte di classi di omologia e la persistenza sono totalmente analoghe a quelle descritte nel Capitolo 3. Vediamo un esempio.

Esempio A.3.14. Riprendiamo l'esempio A.3.3. La componente connessa contenente v_2 , cioè la sua classe di omologia in grado zero, nasce in K_2 e muore entrando in K_3 ; in K_5 nasce un ciclo che poi muore entrando in K_6 , e così via. Invece, la componente connessa contenente v_1 nasce in K_1 e "non muore mai". Più in dettaglio, abbiamo:

- in grado zero, $\text{pers}([v_1]) = \infty$, $\text{pers}([v_2]) = 3 - 2 = 1$, $\text{pers}([v_3]) = 4 - 3 = 1$;
- in grado uno, $\text{pers}([v_1, v_2, v_3]) = 6 - 5 = 1$

Si ottengono pertanto tutti gli invarianti e descrittori del caso, a partire dai diagrammi di persistenza e la distanza di bottleneck.

Inoltre, in modo analogo si possono adattare i concetti del caso multipa-

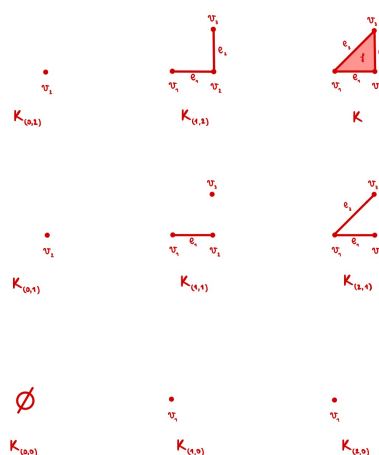


Figura A.8: Rappresentazione di una filtrazione biparametrica di un complesso simpliciale costituito da un 2-simplesso

rametrico descritti nel Capitolo 3 all'ambito dei complessi simpliciali. Un esempio di filtrazione biparametrica è riportato in Figura A.8.

A.3.1 Problema 1 e Problema 2 nel contesto simpliciale

Anche nel contesto dei complessi simpliciali è molto facile costruire esempi in cui l'informazione ottenuta con un approccio biparametrico risulta maggiore rispetto a quella ottenuta con un approccio monoparametrico e si riesce ad apprezzare la relazione tra questo gap informativo e la correlazione tra le funzioni filtranti coinvolte. Vediamone alcuni.

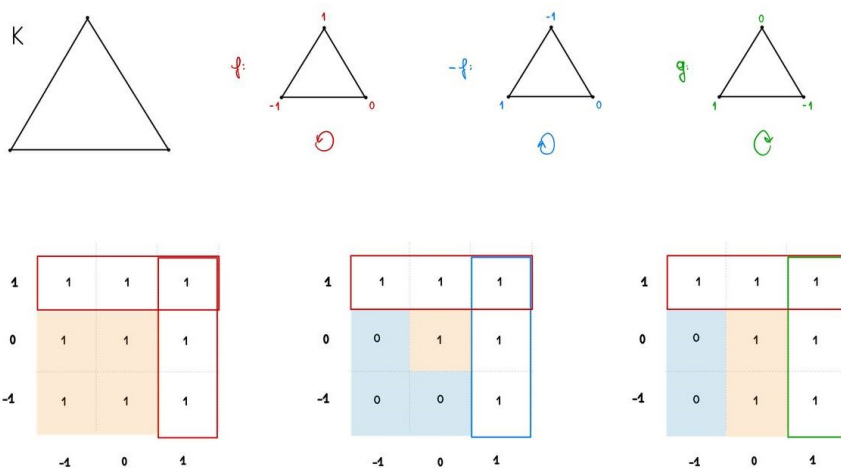


Figura A.9: Un complesso simpliciale K , con tre vertex function definite su di esso e le rispettive tabelle rappresentanti i corrispondenti numeri di betti persistenti di grado 0. Sotto le tre funzioni è raffigurato il loro “andamento”, per facilitare il lettore a farsi un’idea di quali di esse potrebbero essere correlate oppure no.

Esempio A.3.15. Consideriamo il complesso simpliciale K rappresentato in Figura A.10 e le funzioni $f, -f, g : X \rightarrow \mathbb{R}$ indotte dalle vertex function anch’esse rappresentate in Figura A.9. Consideriamo inoltre $\Phi := (f, f)$, $\Psi := (f, -f)$ e $\xi := (f, g)$. Otteniamo i numeri di Betti persistenti di grado 0 riportati nelle tabelle in Figura A.9. In particolare, le sole informazioni monoparametriche non sono sufficienti a distinguere tra $f, -f$ e g , mentre quelle multiparametriche sì.

Esempio A.3.16. Consideriamo il complesso simpliciale K rappresentato in Figura A.10 e le funzioni $f, -f, g : X \rightarrow \mathbb{R}$ indotte dalle vertex function anch’esse rappresentate in Figura A.10. Consideriamo inoltre $\Phi := (f, f)$, $\Psi := (f, -f)$ e $\xi := (f, g)$. Otteniamo i numeri di Betti persistenti di grado 0 riportati nelle tabelle in Figura A.10. In particolare, le sole informazioni monoparametriche non sono sufficienti a distinguere tra $f, -f$ e g , mentre quelle multiparametriche sì.

Nei due esempi precedenti abbiamo lavorato in grado 0. Segue un esempio in cui si considera l’omologia di grado 1.

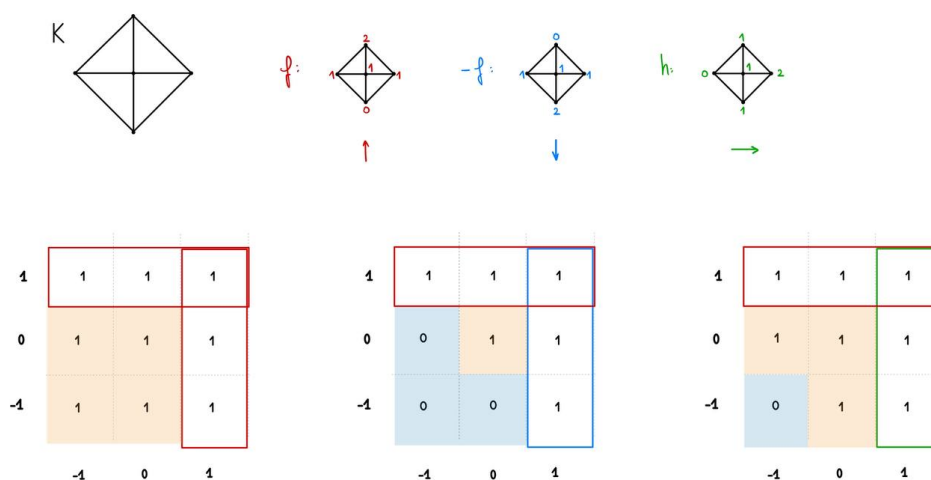


Figura A.10: Un complesso simpliciale K , con tre vertex function definite su di esso e le rispettive tabelle rappresentanti i corrispondenti numeri di betti persistenti di grado 0. Sotto le tre funzioni è raffigurato il loro “andamento”, per facilitare il lettore a farsi un’idea di quali di esse potrebbero essere correlate oppure no.

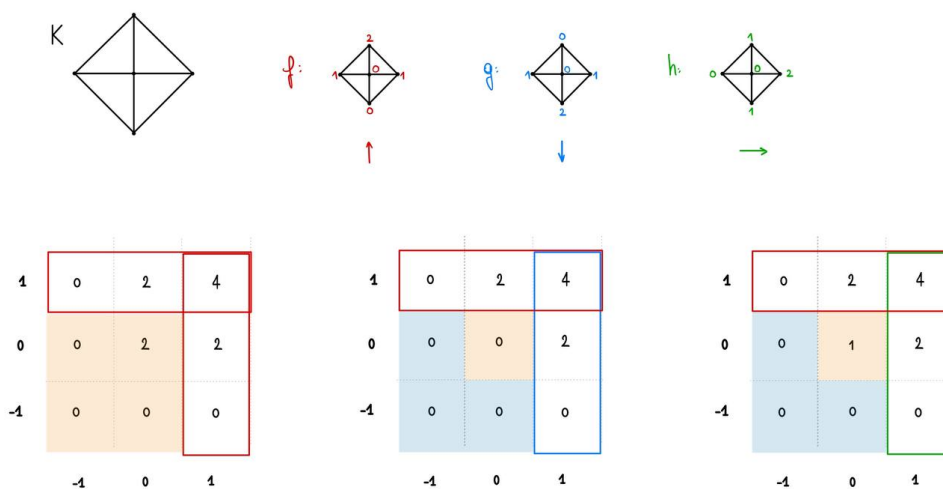


Figura A.11: Un complesso simpliciale K , con tre vertex function definite su di esso e le rispettive tabelle rappresentanti i corrispondenti numeri di betti persistenti di grado 1. Sotto le tre funzioni è raffigurato il loro “andamento”, per facilitare il lettore a farsi un’idea di quali di esse potrebbero essere correlate oppure no.

Esempio A.3.17. Consideriamo il complesso simpliciale K rappresentato in Figura A.10 e le funzioni $f, -f, g : X \rightarrow \mathbb{R}$ indotte dalle vertex function anch’esse rappresentate in Figura A.11. Consideriamo inoltre $\Phi := (f, f)$,

$\Psi := (f, -f)$ e $\xi := (f, g)$. Otteniamo i numeri di Betti persistenti di grado 1 riportati nelle tabelle in Figura A.11. In particolare, le sole informazioni monoparametriche non sono sufficienti a distinguere tra $f, -f$ e g , mentre quelle multiparametriche sì.

Appendice B

Analisi Dati Topologica

Come accennato diverse volte nel corso della trattazione, l'argomento di questa Tesi ha avuto origine nell'ambito dell'Analisi Dati Topologica (TDA). Pertanto, in quest'ultima appendice, vediamo brevemente di cosa si tratta.

L'Analisi Dati è una disciplina permette di ricavare informazioni a partire da dati. I dati vengono raccolti e successivamente trasformati (ad esempio, centralizzati, normalizzati, ecc.), con l'obiettivo di renderli accessibili e comprensibili il più possibile, in modo da ricavarne informazioni a prima vista "nascoste", come correlazioni, anomalie, tendenze e così via. Questo consente ad esempio di fare previsioni o ottimizzare processi. Si tratta di un ambito estremamente interdisciplinare, poiché basato su metodi statistici, matematici e computazionali ed è di interesse sia teorico che applicativo; ad esempio molte applicazioni provengono dall'ambito dell'economia, del marketing, della finanza, dell'ingegneria o delle scienze naturali.

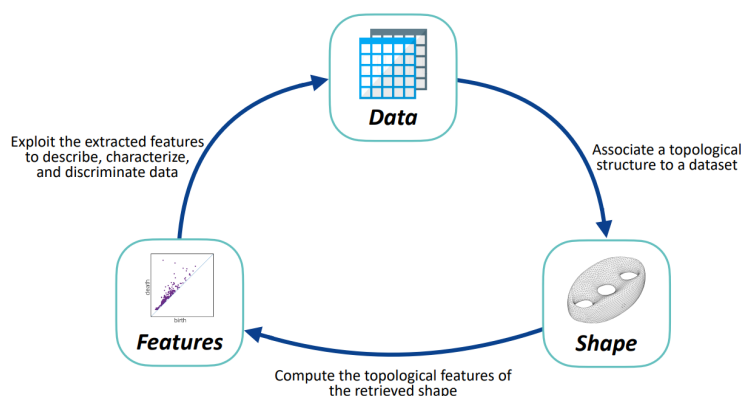


Figura B.1: Schema raffigurante il processo dei dati nel contesto dell'Analisi Dati Topologica. Immagine tratta da [16].

L'Analisi Dati Topologica studia dataset descritti tramite strutture topo-

logiche. L'idea alla base di tale disciplina è la seguente: dato un dataset, lo dotiamo di una struttura topologica e ne studiamo le caratteristiche con tutti gli strumenti topologici che abbiamo a disposizione, con il fine di descrivere, caratterizzare e discriminare i dati iniziali.

Le strutture maggiormente utilizzate per studiare dataset tramite omologia persistente, ovvero tramite filtrazioni, sono le seguenti.

Complessi simpliciali. Sono le strutture descritte in Appendice B, di cui è possibile costruire filtrazioni. Esse consentono di lavorare con dataset non organizzati, in alta dimensione e con un elevato numero di campioni.

Consideriamo un dataset rappresentato da una nuvola di punti V in \mathbb{R}^n . Studiare V lasciandola inalterata non fornisce alcuna informazione topologica rilevante, la “vera” forma del dataset può invece essere individuata costruendo un complesso simpliciale dove i punti vicini di V sono collegati tramite un opportuno simpleso. Le costruzioni standard sono:

- triangolazioni di Delaunay e diagrammi di Voronoi;

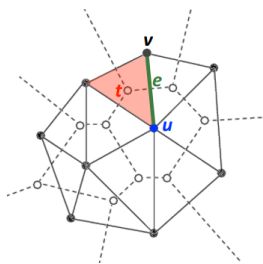


Figura B.2: Una triangolazione di Delaunay con il corrispondente diagramma di Voronoi. Le triangolazioni di Delaunay sono caratterizzate da semplici “equilibrati”, poco distorti, come quelli rappresentati in figura. Immagine tratta da [16].

- complessi di Čech;
- complessi di Vietoris-Rips;
- Alpha-shapes;
- Complessi witness.

La maggior parte di questi si basa sul concetto di *nerve complex*.

Griglie regolari Si tratta di strutture simili ai complessi simpliciali, ma in questo caso l'unità di base è costituita non più dai semplici, bensì dagli ipercubi.

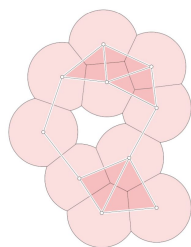


Figura B.3: Rappresentazione di una alpha-shape. Immagine tratta da [16].

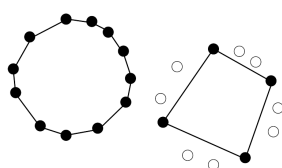


Figura B.4: Rappresentazione di un complesso witness (a destra), costruito a partire dall'insieme di sinistra. L'idea è che è possibile che non sia necessario considerare tutti i punti del dataset grezzo, ma solo alcuni. Immagine tratta da [16].

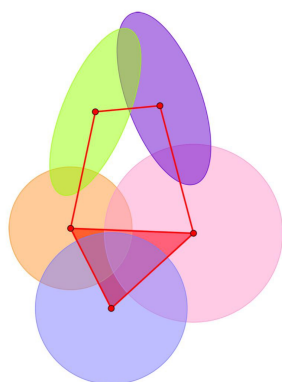


Figura B.5: Rappresentazione di un nerve complex. A partire da una collezione di insiemi si costruisce un complesso simpliciale assegnando ad ogni intersezione tra di essi un semplice, la cui dimensione è pari al numero di insiemi dell'intersezione. Immagine tratta da [16].

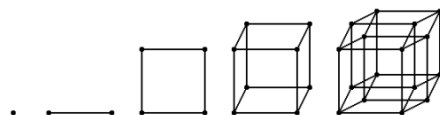


Figura B.6: Esempi di ipercubi di varie dimensioni. Immagine tratta da [16].

Complessi di celle. Anche in questo caso, l'idea è la stessa dei precedenti e l'unità di base è costituita dalle cosiddette celle, dove con n -cella indichiamo uno spazio topologico omeomorfo al disco $D^n \setminus \partial D^n$ aperto.

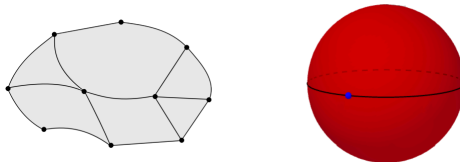


Figura B.7: Esempi di complessi di celle. Immagine tratta da [16].

Insiemi di sottolivello È la struttura utilizzata in questo lavoro di Tesi.

Grafi e reti complesse Anche a partire da strutture come i grafi pesati, è possibile costruire delle filtrazioni.

Ad esempio, dato un grafo $G := (V, E, w)$ dove V è l'insieme dei vertici di G , $E \subseteq V \times V$ è l'insieme degli archi di G e $w : E \rightarrow \mathbb{R}$ è la funzione dei pesi, è possibile definire un *Flag complex*. Fissato $\varepsilon > 0$ il flag complex di soglia ε di G si costruisce a partire dal grafo $G_\varepsilon := (V, E_\varepsilon)$, dove E_ε è l'insieme degli archi in E con peso al più ε .

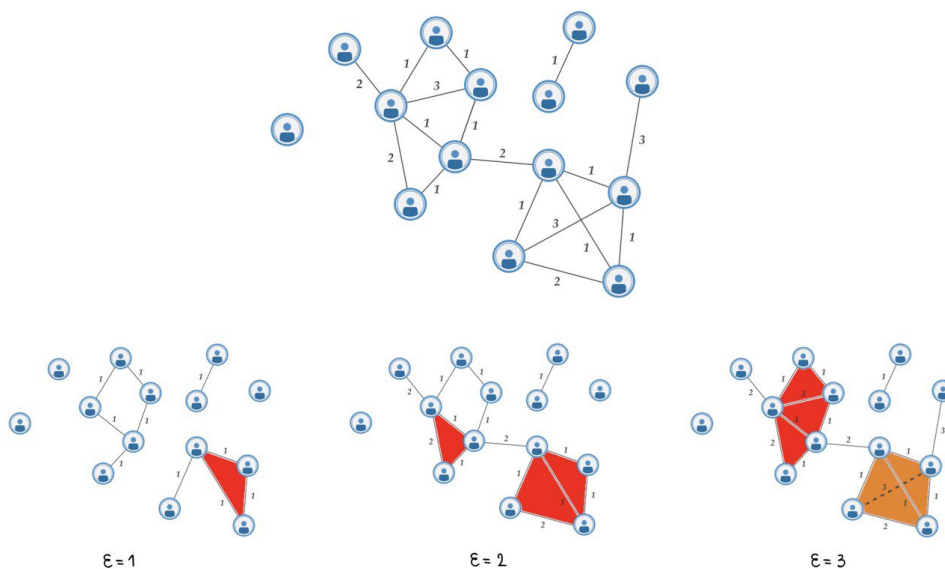


Figura B.8: Rappresentazione di alcuni flag complex. In alto abbiamo il grafo di partenza e in basso tre flag complex ad esso associati, con tre valori di soglia differenti. Immagine tratta da [16].

Per descrizioni dettagliate di tali strutture, si può fare riferimento a [9, 16].

Concludo questo lavoro segnalando un sito interessante [11], contenente un dataset completo di lavori riguardanti applicazioni di vario genere dell'Analisi Dati Topologica ad altri contesti.

Bibliografia

- [1] Dobrina Boltcheva et al. *Constructive Mayer-Vietoris Algorithm: Computing the Homology of Unions of Simplicial Complexes*. Research Report RR-7471. INRIA, dic. 2010. URL: <https://inria.hal.science/inria-00542717>.
- [2] George EP Box et al. *Time series analysis: forecasting and control*. John Wiley & Sons, 2015.
- [3] Gunnar Carlsson e Afra Zomorodian. «The theory of multidimensional persistence». In: *Proceedings of the twenty-third annual symposium on Computational geometry*. 2007, pp. 184–193.
- [4] Andrea Cerri e Claudia Landi. «The persistence space in multidimensional persistent homology». In: *Discrete Geometry for Computer Imagery: 17th IAPR International Conference, DGCI 2013, Seville, Spain, March 20-22, 2013. Proceedings 17*. Springer. 2013, pp. 180–191.
- [5] Andrea Cerri et al. «Betti numbers in multidimensional persistent homology are stable functions». In: *Mathematical Methods in the Applied Sciences* 36.12 (2013), pp. 1543–1557.
- [6] David Cohen-Steiner, Herbert Edelsbrunner e John Harer. «Extending persistence using Poincaré and Lefschetz duality». In: *Foundations of Computational Mathematics* 9.1 (2009), pp. 79–103.
- [7] David Cohen-Steiner, Herbert Edelsbrunner e John Harer. «Stability of persistence diagrams». In: *Proceedings of the twenty-first annual symposium on Computational geometry*. 2005, pp. 263–271.
- [8] William Crawley-Boevey. «Decomposition of pointwise finite-dimensional persistence modules». In: *Journal of Algebra and its Applications* 14.05 (2015), p. 1550066.
- [9] Tamal Krishna Dey e Yusu Wang. *Computational topology for data analysis*. Cambridge University Press, 2022.
- [10] Barbara Di Fabio e Claudia Landi. «A Mayer–Vietoris formula for persistent homology with an application to shape recognition in the presence of occlusions». In: *Foundations of Computational Mathematics* 11 (2011), pp. 499–527.

- [11] Barbara Giunti, Jānis Lazovskis e Bastian Rieck. *DONUT: Database of Original & Non-Theoretical Uses of Topology*. <https://donut.topology.rocks>. 2022.
- [12] Herbert Edelsbrunner e John L Harer. *Computational topology: an introduction*. American Mathematical Society, 2022.
- [13] Edelsbrunner, Letscher e Zomorodian. «Topological persistence and simplification». In: *Discrete & computational geometry* 28 (2002), pp. 511–533.
- [14] Marc Ethier et al. «Geometry of the matching distance for 2D filtering functions». In: *Journal of Applied and Computational Topology* 7.4 (2023), pp. 815–830.
- [15] Gerald B Folland. *A course in abstract harmonic analysis*. CRC press, 2016.
- [16] Ulderico Fugacci. *Topological Data Analysis*. Slide sulla parte dedicata all'Analisi Dati Topologica del corso di Matematica Discreta e Applicazioni presso Università degli Studi di Genova, Dipartimento di Matematica. URL: <https://fugacci.github.io/home/PDF%20TDA/Matematica%20Discreta%20e%20Applicazioni%20-%20TDA%20-%20v6.pdf>.
- [17] Allen Hatcher. *Algebraic topology*. 2005.
- [18] Michael Kerber. «Multi-parameter persistent homology is practical». In: *TDA & Beyond*. 2020.
- [19] Serge Lang. *Algebra*. Vol. 211. Springer Science & Business Media, 2012.
- [20] Ryan H Lewis e Afra Zomorodian. «Multicore homology via mayer vietoris». In: *arXiv preprint arXiv:1407.2275* (2014).
- [21] Steve Y Oudot. *Persistence theory: from quiver representations to data analysis*. Vol. 209. American Mathematical Soc., 2017.
- [22] Andrea Raffo et al. «SHREC 2021: Retrieval and classification of protein surfaces equipped with physical and chemical properties». In: *Computers & Graphics* 99 (2021), pp. 1–21.
- [23] Maria Piera Rogantin. *Statistica Descrittiva*. Dispense del corso di Statistica Descrittiva presso Università degli Studi di Genova, Dipartimento di Matematica.
- [24] Afra Zomorodian e Gunnar Carlsson. «Computing persistent homology». In: *Proceedings of the twentieth annual symposium on Computational geometry*. 2004, pp. 347–356.

Ringraziamenti

Eccoci giunti alla fine di questo lavoro. Se state leggendo queste righe, significa che avete avuto la pazienza (la *persistenza!*) di seguirmi fino a qui e, per questo, meritate il mio primo ringraziamento.

Ogni pagina di questa Tesi è il risultato di un cammino impregnato di studio e impegno personale, ma soprattutto è il frutto di uno percorso durato sei anni e condiviso con molte persone, che desidero pertanto ringraziare.

La mia più autentica riconoscenza va al mio Relatore, Ulderico Fugacci, per diverse ragioni: per aver definitivamente catturato il mio interesse, già emerso grazie a letture personali, per l'omologia persistente durante le sue eccentriche lezioni di Matematica Discreta e Applicazioni, tanto da convincermi a rivolgermi a lui per la Tesi; per avermi subito catapultata nel vivo dell'argomento e della ricerca attuale, riponendo molta fiducia nelle mie capacità; per la dedizione con cui ha seguito il mio lavoro; per l'entusiasmo travolgente con cui mi ha trascinato verso nuove strade e ha accolto le mie idee; per la meticolosità con cui ha controllato tutto quanto — leggendo, rileggendo, correggendo, immedesimandosi in chi potesse leggere questo lavoro, aggiungendo, rimuovendo, stravolgendo, e di nuovo leggendo, rileggendo ecc. —; per la leggerezza con cui, al tempo stesso, mi ha guidato attraverso i problemi incontrati durante il percorso, scandito da continui “Ce la faremo!” e “Siamo in un cul de sac”. Mi sono davvero sentita parte di una squadra.

Un altro tassello fondamentale nei miei anni da studentessa, e che mi ha condotta al mondo di cui fa parte il mio lavoro di Tesi, è stato il primo approccio con la topologia algebrica, avvenuto grazie al mio Correlatore, nonché Relatore della Tesi triennale, Matteo Penegini. Alle sue lezioni, dense e significative, scandite a suon di “Può succedere di tutto”, “Vediamo un esempio. Proposizione...” e “Beh... questo è facile, no?”, devo davvero molto dei miei attuali interessi e di quella che è la mia “visione” matematica.

Una delle caratteristiche della mia Tesi, ed è forse ciò che più mi ha stimolato durante questi mesi di lavoro, è che abbraccia, o almeno sfiora, moltissimi ambiti. Per questo motivo, ho avuto modo di fare affidamento

— e la fortuna di poterlo fare — anche sulle spiegazioni, i consigli e l'ascolto attivo di altre persone. In particolare, ringrazio di cuore Emanuela De Negri, Francesca Tombari, Tommaso Bruno e tutti i miei amici dottorandi importunati tra una pausa caffè e l'altra.

Ringrazio inoltre tutto il corpo docenti del DIMA e Tony e Guido: grazie a voi ho trovato la mia dimensione in un contesto di grande valore e, al tempo stesso, accogliente e familiare.

In questi mesi di scrittura della Tesi, ho avuto, per forza di cose, molto a che fare con la parola *persistenza*. Ho iniziato così a stilare un elenco di dettagli e situazioni *persistenti* nella mia vita: il gesso sulle mani e sui vestiti, le partite a carte dopo pranzo, i miei formicolii — che avrei tanto voluto non dover includere nell'elenco e invece sono davvero persistenti, nel vero senso della parola —, i viaggi in treno per andare e tornare da casa o da Ceriale, le mie passeggiate e corse fino a Boccadasse, la musica costantemente nelle orecchie, l'imbevivibile caffè delle macchinette e potrei continuare per pagine e pagine. Ma soprattutto ho individuato *persone* persistenti nella mia vita, che hanno più di tutto e tutti contribuito a rendermi chi sono oggi e che, per questo, vorrei ringraziare. Scusate se mi dilungherò, ma questa mi sembra un'ottima occasione per ringraziare, cosa che invece nella vita di tutti i giorni, per un motivo o per un altro, non si fa così facilmente.

Grazie alle mie coinquiline, a partire da Ile e Michi, per arrivare a Drilla e Alty ed infine tornare ad Ile, con Franci e mia cugina Lety. Grazie a tutte per la compagnia e le vere avventure da Coinquiline con la *C* maiuscola affrontate insieme. Con voi ho vissuto la vita universitaria che sempre mi sono immaginata e, per questo, vi ringrazio.

Grazie agli amici di una vita, chi da Alassio e chi da Piacenza, che rimangono tali anche se non ci si vede spesso, e con cui riesco ogni volta a provare due sensazioni opposte: quando ci vediamo, da un lato mi sembra che il tempo non sia mai trascorso, dall'altro, al contrario, mi rendo conto proprio di quanto tempo sia trascorso da quando ci siamo conosciuti. E questo è bellissimo.

Grazie alle Gazzelle, per avermi trascinata — no, un momento, sono stata più io a trascinare voi — in un'attività per me totalmente nuova e che mai avrei pensato avrebbe fatto parte delle mie giornate, e soprattutto per il bellissimo gruppo che si è così creato, sia tra noi gazzelcompagne che con la numerosa tifoseria gazzellica. Che dire, come sempre #staytuned, #staygazzelle.

Grazie ai miei compagni e amici di Università per la compagnia e il supporto costante, per le infinite partite a carte, lo studio matto e disperato, le bevute, le feste, le suonate e i bagni in spiaggia per alleggerire le sessioni estive, le barzellette e le battute per le quali rido solo io, le risate generali, i mugugni, i passaggi di appunti, le spiegazioni, le ore passate a ripetere alla lavagna, il fantastico coro DIMAtale, il Festival della Scienza, la storica faida “mangiamo fuori” vs “mangiamo dentro”, i karaoke, e tutto il resto. Avete aggiunto un sacco di pepe al condimento di questi anni.

Grazie agli amici dello swing e soprattutto grazie a chi mi ha fatto conoscere questo mondo, il Team YourSwing. Grazie a voi, ho avuto la possibilità di entrare in contatto e poi tuffarmi a capofitto in qualcosa che mi sono sentita, fin dal primo momento, cucito addosso come nient'altro nella mia vita. Esco sempre dalle lezioni e dalle serate sentendomi fluttuare un metro sopra il pavimento e credo che questo mondo abbia ancora moltissimo da offrirmi. Per questo, vi ringrazio di cuore anche per avermi scelta come nuovo membro del Team, cosa di cui sono molto entusiasta e orgogliosa. Un ringraziamento particolare va a Matte, in primis per gli innumerevoli passaggi che mi ha offerto in questi anni, senza i quali mi sarei persa la metà degli eventi, ma anche per essere diventato un ottimo amico; a Davide “Lucifer” per essere stato il mio primo amico sinceramente incuriosito dai miei racconti e che si è deciso a provare a ballare — e ora è completamente in fissa —; alla Fra per aver condiviso così tanto con me; a Frankie per coinvolgermi nei suoi pazzi progetti; e a Marco, a Louis, Nicole, Luca e tutti tutti quanti, per regalarmi momenti di leggerezza, non solo in pista da ballo. Ah, e poi grazie a chi ha sopportato le mie innumerevoli storie a tema swing su Instagram e soprattutto il mio propagandismo senza limiti.

Grazie ai miei *bff solo a Disney Channel* Simo, per gli innumerevoli esami preparati insieme, gli scleri — decisamente più miei che suoi — e per offrire il proprio sostegno sempre in punta di piedi; Fabri, per le camminate, la sua infinita e sincera curiosità nei confronti di ciò che e di chi lo circonda e che trasmette agli altri, la semplicità e la genuinità con cui si rapporta con le situazioni e le persone, e — soprattutto! — per averci fatto scoprire il Tiki Bar; Anna per essere stata l'amica che tutti vorrebbero avere, sempre attenta, sempre in grado di leggere le situazioni e sempre in grado di dire e fare la cosa giusta, con naturalezza e senza troppi giri di parole.

Grazie alla Ludo, chiaramente ¹ la mia vera anima gemella, con cui quest'anno ho condiviso TUTTO — dagli stessi traumi e scleri sentimentali, alla matematica, allo swing, alle Gazzelle, alla musica, alle suonate a casa con i muratori, alla birra, le suonate e i calci balilla dell'Orso, e molto molto

¹Secondo Spotify, siamo compatibili al 96%. Direi che non c'è altro da aggiungere.

altro. D'altronde, se siamo l'una il Ryan dell'altra, una ragione c'è.

Infine, non posso che ringraziare la "cosa" più persistente che esista nella mia vita: la mia famiglia. Per tutti i "San Rocco" e le altre feste trascorse insieme; per i picnic e le camminate in montagna; per i bagni in spiaggia; per le passeggiate con Camilla, Eolo, Miele e Dante; per avermi sempre supportata con lo studio e con la danza, nei miei successi e nei miei fallimenti; per avermi sempre fatto fare ciò che mi piace e anche di più; per le risate, i pianti, le litigate e gli abbracci; per avermi sempre stimolata, da tutti i punti di vista; per aver affrontato insieme i momenti difficili; per avermi sempre permesso di scegliere. E potrei continuare all'infinito, ma tutto si può riassumere dicendovi grazie per l'affetto incondizionato, i consigli, il supporto, ma soprattutto per avermi permesso di diventare quella che sono oggi. E un abbraccio speciale va alla mia nonna.

Per concludere, grazie a Genova, la città che mi ha accolto per gli studi e subito conquistata, con la sua aria di mare e i suoi bricchi, i suoi vicoli tortuosi e i suoi palazzi maestosi, la sua vita lenta e frenetica, e la sua gente essenziale ma gentile. Qui mi sono sentita "tornata a casa", nella mia Liguria, vivendo anni davvero intensi ed emozionanti, in una splendida cornice.

Mi porterò sempre nel cuore tutto questo.