



**UNIMORE**  
UNIVERSITÀ DEGLI STUDI DI  
MODENA E REGGIO EMILIA

UNIVERSITÀ DEGLI STUDI DI MODENA E REGGIO EMILIA  
Dipartimento di Scienze Fisiche, Informatiche e Matematiche

---

Corso di Laurea Magistrale in Matematica

# Caratterizzazione del ciclo respiratorio nei pazienti sottoposti a trattamento radioterapico alla mammella attraverso strumenti di SGRT (Surface-Guided Radiotherapy)

Relatrice:  
Prof.ssa Silvia  
Bonettini

Candidata:  
Giulia Acerbi

Correlatrice:  
Dott.ssa Valeria  
Trojani

---

Anno Accademico 2024-2025



# Abstract

*Background:* Il tumore alla mammella rimane, ad oggi, la neoplasia più comune nel sesso femminile. La radioterapia (RT) si è dimostrata efficace nel trattare questo tipo di tumore, tuttavia, il cambio significativo del tracciato respiratorio tra diverse sedute di trattamento deve essere tenuto in considerazione e gestito correttamente per assicurare la buona riuscita clinica di questi trattamenti. L'obiettivo di questo elaborato è caratterizzare il respiro di pazienti sottoposti a trattamento radioterapico alla mammella per l'arruolamento in trattamenti a respiro controllato, scegliendo come pazienti potenziali quelle che hanno, sulla base di dati retrospettivi, un tracciato di respirazione ripetibile in regime a respiro libero.

*Metodi:* Sono stati utilizzati i segnali provenienti dal monitoraggio del movimento su tre assi (verticale, VRT, longitudinale, LNG, e laterale, LAT) e un segnale complessivo (MAG) della ROI di 177 pazienti durante un trattamento suddiviso in 13/15 frazioni, acquisiti tramite strumenti di SGRT in uso nella pratica clinica.

Per la caratterizzazione del ciclo respiratorio, sono state individuate le varie componenti in frequenza media e relativa ampiezza dei segnali analizzati, tramite la trasformata di Fourier, ricavandone poi media, deviazione standard e percentili.

Al fine di evidenziare due classi di pazienti, "ripetibili" e "non ripetibili", abbiamo calcolato l'errore relativo percentuale (ER%) relativo al segnale MAG di ogni seduta, in modo da evidenziare pazienti con pattern respiratori significativamente differenti durante le varie sedute di trattamento. In seguito, abbiamo cercato una correlazione tra l'appartenenza a una delle due classi sopra definite ed età e volume del tumore dei pazienti, utilizzando un test di confronto della media (Wilcoxon-Mann-Whitney).

*Risultati:* L'ER% relativo al segnale MAG ha mediana 8.5% e IQR (5.3-12.9)%, di conseguenza, con cutoff del 5%, il 23% dei pazienti risulta avere un ciclo respiratorio "ripetibile" e i restanti "non ripetibile". L'ER% ha media  $(3.2 \pm 1.1)\%$  per la classe "ripetibile", e  $(12.8 \pm 10.4)\%$  per la classe "non ripetibile". Le ampiezze e frequenze medie relative al segnale MAG risultano avere media  $(1.5 \pm 0.7)mm$  e  $(0.27 \pm 0.07)Hz$  per la classe "ripetibile", e  $(2.7 \pm 1.5)mm$  e  $(0.28 \pm 0.08)Hz$  per la classe "non ripetibile". Infine, secondo il test di Wilcoxon, la ripetibilità del respiro risulta correlata significativamente con il volume del tumore dei pazienti, ma non con l'età.

*Conclusioni:* I risultati ottenuti evidenziano la fattibilità della metodologia proposta riguardo la caratterizzazione del movimento respiratorio delle pazienti sottoposte a trattamento RT della mammella. Il motivo dei valori di ER% ottenuti per il DB può essere imputata ai dati retrospettivi estratti da pazienti che hanno eseguito il trattamento in regime di respiro libero e senza istruzioni su come gestire il movimento respiratorio. Questi risultati evidenziano quindi l'importanza di un

protocollo di gestione del movimento che potrebbe portare beneficio al trattamento e alla prognosi del paziente.

# Indice

<b>Introduzione</b>	<b>5</b>
<b>1 Aspetti teorici</b>	<b>7</b>
1.1 Interazione radiazione-materia . . . . .	7
1.1.1 Radiazioni ionizzanti . . . . .	7
1.1.2 Interazione dei fotoni con la materia . . . . .	8
1.2 Radioterapia . . . . .	10
1.2.1 Processo Radioterapico . . . . .	10
1.3 Radioterapia a fasci esterni . . . . .	13
1.3.1 3D-CRT . . . . .	14
1.3.2 VMAT . . . . .	14
1.3.3 IGRT . . . . .	14
1.4 Radioterapia per Tumore alla mammella . . . . .	14
1.5 Surface Guided Radiation Therapy (SGRT) . . . . .	15
1.5.1 AlignRT . . . . .	15
1.5.2 Vantaggi e svantaggi della SGRT . . . . .	16
1.6 Tecniche di controllo del movimento respiratorio . . . . .	17
1.6.1 Breath-hold . . . . .	17
1.6.2 Gating respiratorio . . . . .	18
1.6.3 Metodi di tumor-tracking . . . . .	18
1.7 Elaborazione di segnali . . . . .	18
1.7.1 Classificazione dei segnali . . . . .	19
1.7.2 Serie di Fourier . . . . .	19
1.7.3 Trasformata di Fourier (FT) . . . . .	21
1.7.4 Convoluzione e Filtraggio . . . . .	28
1.7.5 Filtraggio delle frequenze . . . . .	32
1.7.6 Formula di inversione in $L^1$ . . . . .	34
1.7.7 Estensione della FT in $L^2(\mathbb{R})$ . . . . .	35
1.7.8 Elaborazione di segnali discreti . . . . .	37
1.7.9 Filtri discreti . . . . .	43
1.8 Test statistici . . . . .	44
1.8.1 Test di normalità . . . . .	44
1.8.2 Test U di Mann-Whitney . . . . .	45
1.9 Caratterizzazione del movimento respiratorio e di segnali . . . . .	46
<b>2 Materiale e metodi</b>	<b>47</b>
2.1 Dataset e pazienti . . . . .	47
2.2 Elaborazione del segnale . . . . .	47

2.2.1	Pre-processing . . . . .	47
2.2.2	Identificazione dei picchi . . . . .	50
2.2.3	Calcolo delle statistiche . . . . .	50
2.3	Test di normalità . . . . .	50
2.4	Classificazione dei pazienti . . . . .	51
2.5	Analisi statistica . . . . .	51
<b>3</b>	<b>Risultati</b>	<b>52</b>
3.1	Pazienti e elaborazione del segnale . . . . .	52
3.2	Classificazione dei pazienti . . . . .	62
3.3	Analisi statistica . . . . .	66
<b>4</b>	<b>Discussioni</b>	<b>67</b>
	<b>Conclusioni</b>	<b>68</b>
<b>A</b>	<b>Materiale di background</b>	<b>70</b>
A.1	Spazi $L^p$ . . . . .	70
A.2	Convergenza . . . . .	71
<b>B</b>	<b>Codici Python</b>	<b>72</b>
B.1	Elaborazione dei segnali e creazione di un unico DataFrame . . . . .	72
B.2	Test di normalità . . . . .	78
B.3	Esempi di pazienti . . . . .	80
B.4	Classificazione dei pazienti . . . . .	82
B.5	Analisi statistica . . . . .	85

# Introduzione

Il carcinoma della mammella è il tumore più frequentemente diagnosticato nelle donne in Italia. La maggior parte dei tumori maligni mammari è diagnosticato in fase iniziale quando il trattamento chirurgico può essere più spesso conservativo [1]. Questo è spesso associato a radioterapia (RT) postoperatoria, che si divide principalmente in due approcci: WBI (Whole Breast Irradiation - Irradiazione dell'intero seno) e PBI (Partial Breast Irradiation - Irradiazione parziale del seno). Entrambe le tecniche mirano a distruggere eventuali cellule tumorali residue, ma differiscono per volume trattato, durata e indicazioni cliniche.

La RT ai distretti toracici e addominali, di cui fanno parte i trattamenti RT alla mammella, tuttavia, presenta sfide significative legate al movimento del target indotto dalla respirazione. I tumori che si sviluppano in questi distretti possono infatti spostarsi significativamente durante il ciclo respiratorio, portando a deviazioni nella distribuzione della dose. Pertanto, un bersaglio potrebbe non ricevere una copertura della dose adeguata, mentre un tessuto normale potrebbe essere esposto a dosi più elevate del previsto. L'incertezza causata dalla respirazione potrebbe essere minimizzata tramite la gestione del movimento respiratorio [11].

I metodi di gestione del movimento respiratorio comunemente usati in RT includono metodi di **gating respiratorio**, metodi di **breath-hold**, e metodi di **tracciamento del tumore in tempo reale** [20]. I primi due, in particolare, richiedono l'utilizzo di tecniche di imaging per monitorare il movimento del tumore.

La radioterapia guidata dall'immagine (**SGRT**) è una tecnica non invasiva di surface imaging in grado di fornire il monitoraggio in tempo reale del movimento della superficie del torace/seno senza radiazioni aggiuntive [11].

Questo lavoro di tesi è stato condotto presso la Struttura Complessa di **Fisica Medica** dell'**IRCCS di Reggio Emilia**. I segnali di movimento analizzati in questo elaborato sono stati ottenuti utilizzando il sistema di **AlignRT** di VisionRT per monitorare i movimenti di una regione di interesse (ROI) posizionata in corrispondenza del seno di pazienti sottoposti a trattamento di radioterapia a fasci esterni alla mammella in regime a respiro libero.

Lo scopo di questo lavoro di tesi è caratterizzare il ciclo respiratorio dei pazienti, partendo dai loro segnali di movimento su tre assi. Cercheremo poi di classificare i pazienti in base alla ripetibilità del loro movimento respiratorio e vedere come questa ripetibilità si relaziona con altri parametri clinici (volume del tumore ed età).

La tesi è stata organizzata nel seguente modo:

- Capitolo 1: aspetti teorici;
- Capitolo 2: materiali e metodi;
- Nel Capitolo 3: risultati;

- Nel Capitolo 4: discussioni;
- Conclusioni.

# Capitolo 1

## Aspetti teorici

### 1.1 Interazione radiazione-materia

#### 1.1.1 Radiazioni ionizzanti

Con **radiazione** si intende la propagazione di energia attraverso lo spazio o un qualunque mezzo materiale, sotto forma di onde (**radiazioni elettromagnetiche**) o di energia cinetica propria di alcune particelle (**radiazioni corpuscolari**).

Le **radiazioni elettromagnetiche (r.e.)** sono caratterizzate da campi elettrici e magnetici oscillatori che si propagano contemporaneamente nello spazio e possono essere suddivise in **radiazioni ionizzanti** e **non ionizzanti**. Questa suddivisione si basa sulla capacità della radiazione di ionizzare atomi e molecole.

Le **r.e. ionizzanti** possiedono abbastanza energia da estrarre elettroni dagli atomi colpiti da tale radiazione, producendo quindi fenomeni di ionizzazione.

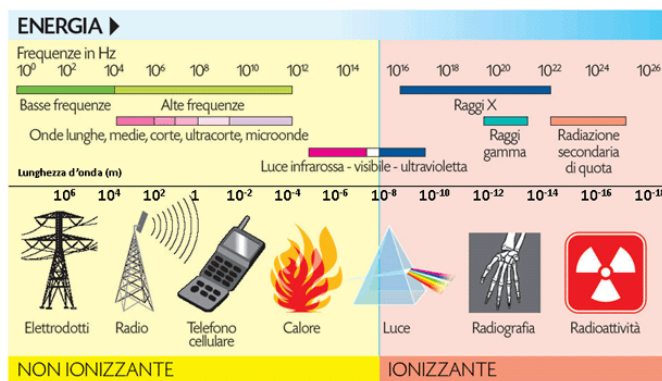


Figura 1.1: Suddivisione delle radiazioni elettromagnetiche in ionizzanti e non ionizzanti [9]

Ogni radiazione, interagendo con la materia, cede energia alla struttura atomica/molecolare del materiale attraversato. Se l'energia è sufficiente (**radiazioni ionizzanti**), si verificano nel materiale effetti distruttivi (frammentazioni, rotture di legami, ionizzazione,...). Le radiazioni ionizzanti possono essere suddivise in:

- elettromagnetiche ( $m = 0, E = hv$ )<sup>1</sup>  $\longrightarrow$  raggi X e  $\gamma$

<sup>1</sup>L'equazione  $E = hv$  è la legge di Planck che definisce l'energia ( $E$ ) di un singolo quanto di radiazione (fotone) come direttamente proporzionale alla sua frequenza ( $\nu$ ), dove  $h$  è la costante di Planck

- corpuscolari ( $m > 0, E_k$ )<sup>2</sup>  $\longrightarrow$  particelle  $\alpha, \beta^\pm, p, n$

Le radiazioni ionizzanti hanno diversi meccanismi di ionizzazione e possono essere raggruppate in **ionizzanti direttamente** o **indirettamente**.

Le **radiazioni corpuscolari cariche elettricamente** ( $\alpha, \beta^\pm, p$ ) ionizzano **direttamente** gli atomi del mezzo. A causa della loro carica, queste particelle interagiscono direttamente attraverso **forze di Coulomb** con gli elettroni degli atomi del mezzo attraversato, provocando ionizzazione o eccitazione lungo il loro percorso.

Le radiazioni elettricamente neutre, in particolare **fotoni (raggi X e gamma)** e **neutroni**, invece, poiché sono neutre, non interagiscono direttamente con gli elettroni atomici, ma interagiscono con il mezzo, rilasciando particelle cariche secondarie veloci (elettroni, protoni). Sono poi queste particelle secondarie a ionizzare il mezzo.

### 1.1.2 Interazione dei fotoni con la materia

I raggi X sono il tipo di radiazione principalmente utilizzato nelle tecniche di radioterapia a fasci esterni.

Quando i raggi X passano attraverso un mezzo, i fotoni trasferiscono la loro energia alla materia per mezzo di complesse interazioni con i nuclei e gli elettroni atomici. Alcune di queste interazioni provocano la fuoriuscita di un elettrone orbitale da un atomo, con conseguente ionizzazione, o la creazione di una coppia elettrone-positrone. A loro volta, questi elettroni producono ionizzazione del mezzo [7].

Le tre forme principali di interazione dei fotoni con la materia, che sono di importanza clinica nella radioterapia, sono l'**effetto fotoelettrico**, l'**effetto Compton** e la **produzione di coppie** [13].

#### Effetto fotoelettrico

L'**effetto fotoelettrico** si verifica quando un fotone, interagisce con un elettrone delle orbite più interne, cedendo tutta la sua energia. Il fotone scompare e l'elettrone acquista energia cinetica pari alla differenza dell'energia del fotone incidente ( $E_f$ ) con quella di legame dell'elettrone. La ionizzazione provoca riassetto degli altri elettroni con emissione di **radiazioni X caratteristiche** o con l'emissione di un **elettrone**. L'effetto fotoelettrico è più probabile per mezzi ad alto Z e per fotoni a bassa energia [7].

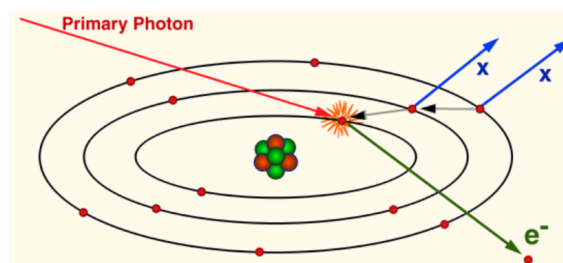


Figura 1.2: Effetto fotoelettrico

---

<sup>2</sup> $E_k$  =energia cinetica

## Effetto Compton

Detto anche **scattering incoerente**, si verifica quando un fotone interagisce con un elettrone libero o degli orbitali più esterni (debolmente legato al nucleo) cedendo parte della sua energia. Come risultato si ha l'emissione di un **elettrone** con una sua energia cinetica ( $E_c$ ) e di un fotone secondario (**gamma Compton**) di energia ( $E_fC$ ) che si propaga in direzione diversa rispetto a quella del fotone originario secondo un **angolo di scatter** che dipende dall'energia ceduta all'elettrone. L'elettrone e il fotone di scattering possono a loro volta interagire con la materia fino ad esaurire la loro energia.

L'energia che viene dissipata è uguale all'energia necessaria per ionizzare l'atomo (corrispondente all'energia di legame,  $E_l$ , dell'elettrone espulso) più l'energia cinetica ( $E_c$ ) che acquista l'elettrone (proporzionale alla velocità che gli viene impressa). L'**energia del fotone Compton** è uguale alla differenza tra l'energia del fotone incidente ( $E_{fi}$ ) e l'energia dissipata:

$$E_{fC} = E_{fi} - (E_l + E_c)$$

Il fotone Compton può essere deviato in qualsiasi direzione, anche retrodiffuso; maggiore è l'energia ceduta all'elettrone, maggiore è l'angolo di deflessione (formato dalla traiettoria del fotone primario con quella del fotone secondario). Inoltre, maggiore è l'energia del fotone incidente, maggiore è l'energia ceduta all'elettrone.

Poiché il fotone non viene assorbito del tutto, ma solo deviato, l'effetto Compton genera radiazione diffusa [7].

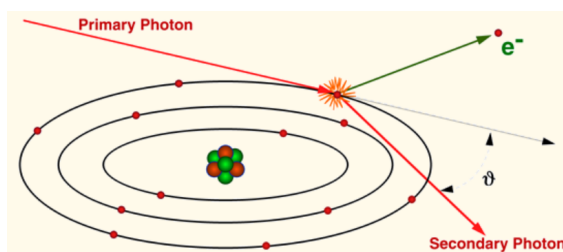


Figura 1.3: Effetto Compton [7]

## Produzione di coppie

Detto anche **effetto fotonucleare**, accade per fotoni di **energia superiore a 1.02 MeV**, corrispondente alla massa delle due particelle che vengono generate dal fenomeno.

Il fotone, interagendo col campo di forza del nucleo, scompare con la contemporanea creazione di 2 particelle: un **elettrone** e un **positrone**; tutta l'energia oltre la soglia di 1.02 MeV è distribuita in ugual misura tra le due particelle sotto forma di energia cinetica. L'elettrone così prodotto può provocare ionizzazioni, mentre il positrone va incontro ad annichilazione, con la conseguente produzione di 2 radiazioni gamma di 0.511 MeV dirette in direzioni diametralmente opposte [7].

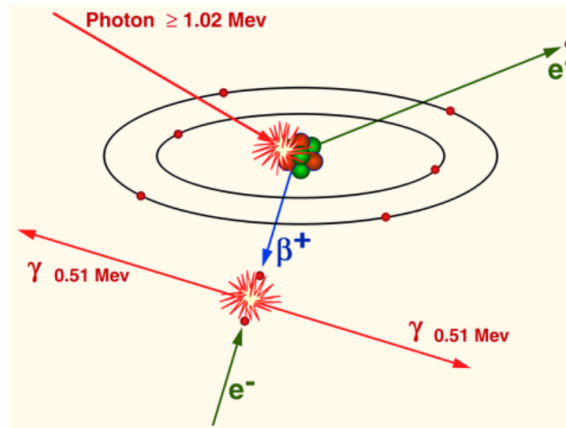


Figura 1.4: Produzione di coppie

## 1.2 Radioterapia

La **radioterapia (RT)** è una pratica medica che utilizza le radiazioni ionizzanti ad alta energia per uccidere le cellule tumorali e arrestarne la proliferazione.

In radioterapia, il successo nell'eradicare un tumore dipende sia dalla **radiosensibilità** del tumore, che dalla **tolleranza del tessuto normale (NTT)** circostante [13]. Infatti, quando si irradiano dei tessuti per eliminare le cellule tumorali è inevitabile che parte della dose depositata danneggi anche le cellule sane dei tessuti adiacenti al tumore. L'obiettivo della radioterapia è quindi quello di massimizzare il danno alle cellule tumorali minimizzando quello ai tessuti sani. Questo approccio si basa sulla diversa risposta delle cellule tumorali e normali alla radioterapia.

Per sfruttare questa differenza, la radioterapia si avvale della tecnica di **frazionamento della dose**, che consiste nel suddividere la dose totale da somministrare al paziente in sedute distanti nel tempo, in modo da permettere ai tessuti sani di riparare i danni prima della somministrazione successiva [4].

### 1.2.1 Processo Radioterapico

Il processo radioterapico è un percorso complesso dove vengono coinvolte diverse figure professionali: il medico radioterapista, il fisico medico, il tecnico sanitario di radiologia medica (TSRM) e l'infermiere. Le fasi più importanti del processo radioterapico sono:

- **valutazione iniziale** tramite visita con il medico radioterapista che analizza la storia clinica del paziente, valuta l'indicazione al trattamento e stabilisce l'obiettivo del trattamento;
- **simulazione del trattamento** tramite acquisizione TC con eventuali sistemi di immobilizzazione;
- **definizione del target** e degli organi a rischio;
- **pianificazione dosimetrica** del trattamento e verifica dosimetrica;
- **esecuzione del trattamento**, con la verifica del posizionamento e la successiva erogazione del trattamento [10].

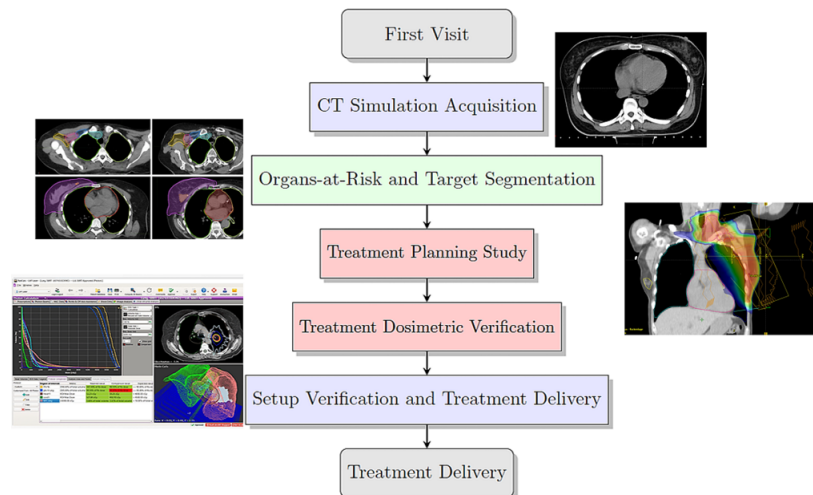


Figura 1.5: Schema del processo di radioterapia [19]

### Simulazione del trattamento

Per la realizzazione di un piano di trattamento, è necessario effettuare una tomografia computerizzata (TC), detta **TC di centraggio** o di simulazione, del paziente posizionato esattamente come sarà durante le sedute di trattamento, per cui anche il lettino della TC deve essere uguale a quello di trattamento.

In base al tipo di trattamento, sulla pelle del paziente vengono posizionati uno o più reperi puntiformi metallici nei punti in cui si incrociano i laser di centraggio in modo da essere visibili sulla TC e calcolare le coordinate spaziali dell'isocentro in fase di pianificazione; nei punti in cui sono posizionati i reperi sul paziente vengono poi effettuati dei **tatuaggi** permanenti puntiformi per riprodurre il corretto posizionamento anche durante le sedute di trattamento.

L'immagine TC è fondamentale perché da essa possono essere estratte le informazioni sulle densità elettroniche dei tessuti, necessarie per il calcolo della dose durante la fase di pianificazione del trattamento [10].

### Definizione dei volumi di irradiazione

Dopo l'acquisizione delle immagini, la fase successiva prevede la contornazione e la definizione del volume target.

I volumi di interesse vengono definiti come segue:

- **GTV (Gross Target Volume)**: volume macroscopico di tumore, ossia quello palpabile e visibile mediante le varie modalità di imaging; comprende il tumore primitivo e le eventuali metastasi
- **CTV (Clinical Target Volume)**: volume individuato dal radioterapista che contiene il GTV e tutte le zone che possono essere sede di diffusione di malattia subclinica;
- **PTV (Planning Target Volume)**: volume per la specifica di dose; è legato all'indeterminazione di tutti gli elementi che concorrono al trattamento, come spostamenti, volumi, ecc;

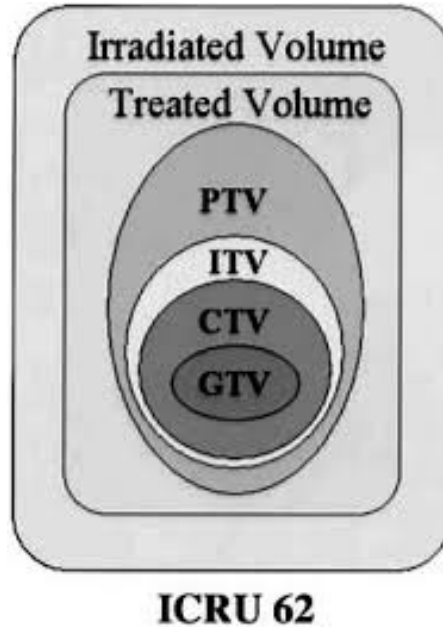


Figura 1.6: Definizione dei volumi di interesse secondo l'ICRU Report 62

- **OAR (Organ At Risk)**: tessuti sani che si trovano in prossimità della zona da irradiare e che potrebbero subire danni in seguito al trattamento; gli OAR possono influenzare significativamente la pianificazione del trattamento;
- **PRV (Planning organ at Risk Volume)**: organo a rischio a cui è stato aggiunto un margine adeguato per tenere in considerazione le variazioni di forma e volume e l'eventuale movimento dell'organo.

### Pianificazione dosimetrica del trattamento

Una volta completata la segmentazione, le immagini TC vengono trasmesse al **sistema di pianificazione del trattamento (TPS)**, un software avanzato utilizzato in radioterapia per pianificare e ottimizzare il trattamento del paziente. Il TPS integra dati anatomici del paziente, informazioni tecniche sul fascio di radiazione e algoritmi fisici e matematici per calcolare la distribuzione di dose di radiazioni all'interno del paziente.

L'elaborazione del piano di trattamento è un processo gestito e controllato dal fisico medico, il quale utilizza il TPS per ottimizzare la distribuzione di dose, bilanciando la massima irradiazione al PTV e la protezione degli OAR.

Prima di effettuare il piano di trattamento sul paziente, nel caso di tecniche più complesse come Radioterapia a Intensità Modulata (IMRT) e Terapia ad Arco Modulato Volumetricamente (VMAT), è necessario effettuare un controllo dosimetrico del trattamento, per verificare la coerenza tra la distribuzione di dose pianificata e la distribuzione di dose effettivamente erogata [10].

### Esecuzione del trattamento

Il paziente viene fatto accomodare sul lettino di trattamento e posizionato correttamente dal personale TSRM, in modo che le croci individuate dai laser corrispondano

ai tatuaggi eseguiti durante la fase di simulazione. Una volta posizionato correttamente, vengono inviati al lettino gli spostamenti previsti dal piano di trattamento per far coincidere l'isocentro del piano con l'isocentro del LINAC.

Per verificare la correttezza del posizionamento, possono essere acquisite immagini da confrontare con la TC di centraggio. Il medico radioterapista valuta, se necessario, gli spostamenti per la corrispondenza tra simulazione e trattamento, che vengono nuovamente inviati al lettino. Se disponibile, è possibile utilizzare un sistema SGRT per posizionare correttamente il paziente [10].

### 1.3 Radioterapia a fasci esterni

Nella **radioterapia a fasci esterni** la radiazione viene diretta verso la zona tumorale da un macchinario chiamato acceleratore lineare (LINAC dall'inglese LINear ACcelerator) [4].

I **LINAC** utilizzati in ambito clinico sono macchinari con design compatto ed efficiente in grado di generare sia terapie a elettroni che a raggi X.

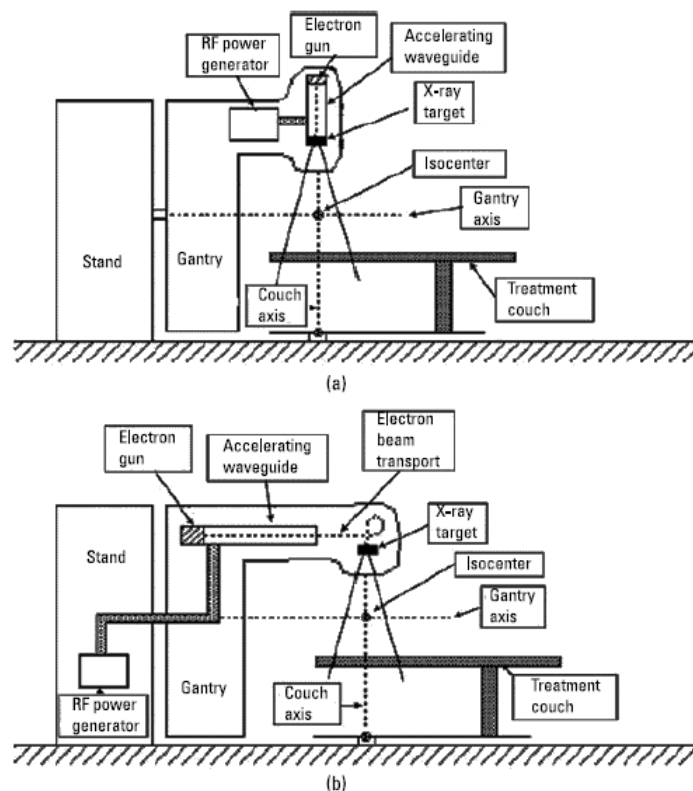


Figura 1.7: Le due configurazioni base che può assumere un LINAC montato su gantry. a) la guida d'onda è posizionata perpendicolarmente all'asse del gantry b) la guida d'onda è posizionata parallelamente all'asse del gantry [4].

In stanza di trattamento il LINAC è montato su un *gantry*: la rotazione di  $360^\circ$  del *gantry* attorno al paziente sdraiato permette all'acceleratore di trattare il tumore da più angolazioni, in modo tale da diminuire la dose assorbita dai tessuti sani. Il paziente viene posizionato in modo che il centro del tumore coincida con l'**isocentro**

della macchina, ovvero il punto di intersezione tra l'asse di rotazione del *gantry* e l'asse del fascio terapeutico [4].

I LINAC utilizzati per la radioterapia a fasci esterni supportano tecniche di trattamento avanzate, tra cui la Radioterapia Conformazionale 3D (**3D-CRT**), la Radioterapia a Intensità Modulata (**IMRT**), la Terapia ad Arco Modulato Volumetricamente (**VMAT**) e la Radioterapia Guidata da Immagini (**IGRT**).

### 1.3.1 3D-CRT

La radioterapia conformazionale 3D (3D-CRT) si basa sulla combinazione di diversi campi di radiazione statici, i quali incidono sul volume bersaglio da diverse direzioni con intensità costante. Aumentando il numero di campi, migliora la conformazione della dose e si riduce l'irraggiamento ai tessuti sani mantenendo costante quella erogata al tumore. L'aumento del numero dei campi porta però a un incremento delle basse dosi diffuse agli organi sani [10].

### 1.3.2 VMAT

La VMAT è una tecnica a intensità modulata non statica: durante il trattamento radioterapico, il *dose rate*, la velocità del gantry e la forma del campo di radiazioni sono variabili. Il gantry del LINAC ruota continuamente durante l'irradiazione, favorendo la massima focalizzazione delle radiazioni sui tessuti tumorali,

Il bersaglio, in pochi minuti, viene raggiunto con precisione, riducendo ulteriormente il rischio di danni collaterali agli organi sani.

Potenzialmente, la VMAT si applica a tutte le neoplasie, ma è particolarmente utile per quelle patologie in cui la dose deve essere focalizzata al massimo per preservare organi molto vicini, come i tumori della testa, del collo e del polmone [10].

### 1.3.3 IGRT

I LINAC sono dotati di un **On-Board Imager** (OBI), un sistema automatizzato che utilizza raggi X per la corretta identificazione del tumore da parte del clinico a seguito di variazioni anatomiche giornaliere durante il trattamento. Questa tecnica è chiamata **radioterapia guidata dall'immagine** (IGRT).

## 1.4 Radioterapia per Tumore alla mammella

Il cancro al seno è il tumore più frequentemente diagnosticato nelle donne a livello globale: quasi il 40% delle pazienti trattate presso l'Unità di Radioterapia dell'AUSL-IRCCS di Reggio Emilia sono pazienti con tumore al seno. La radioterapia è stata tradizionalmente integrata nella terapia conservativa del seno (BCT), per eliminare eventuali residui di malattia microscopica e ridurre il rischio di recidiva locale.

Il regime di radioterapia alla mammella convenzionale mirava a ottimizzare il controllo locale del tumore riducendo al minimo le tossicità tardive, somministrando 50 Gy in 25 frazioni (2 Gy x 25), cioè in cinque settimane di trattamento. Le sfide logistiche affrontate dai pazienti e dai sistemi sanitari a causa della durata prolungata

del trattamento hanno stimolato studi sulla radioterapia ipofrazionata, che prevede una dose più elevata per frazione in un arco temporale moderatamente ridotto. È stato dimostrato che la RT moderatamente ipofrazionata di 40 Gy in 15 frazioni (2.67 Gy x 15) in tre settimane per il cancro al seno è efficace, conveniente e sicura per tutte le indicazioni, i volumi target e le tecniche, anche negli stadi avanzati come evidenziato nello studio prospettico randomizzato di Pechino pubblicato nel 2019. [18].

Il regime di dose di base è, quindi, di 40–42,5 Gy (2,67 Gy/frazione, 5 volte a settimana, in 15–16 frazioni) oppure 45–50,4 Gy (1,8–2 Gy/frazione, 5 volte a settimana, in 25–28 frazioni) [8]. Tuttavia, nel 2023, lo studio FAST-Forward ha dimostrato che la radioterapia ultra-ipofrazionata di 26 Gy in 5 frazioni distribuite in una settimana è un'alternativa sicura ed efficace alla radioterapia locale di 40 Gy in 15 frazioni per pazienti sottoposti a radioterapia adiuvante dopo chirurgia primaria per il cancro al seno in fase iniziale [6]. Questo approccio offre numerosi vantaggi, tra cui maggiore comodità per i pazienti, costi sanitari ridotti e un uso ottimale delle risorse. [19].

## 1.5 Surface Guided Radiation Therapy (SGRT)

L'introduzione della radioterapia guidata dalle immagini (IGRT) nei primi anni 2000 ha offerto importanti miglioramenti in termini di accuratezza dei trattamenti radioterapici. Nonostante questi progressi, le capacità di imaging intra-frazione dei sistemi IGRT basati sul LINAC erano ancora limitate, dando origine a una nuova tecnica per il monitoraggio intra-frazione dei pazienti: la **radioterapia guidata dalle superfici (SGRT)**.

Tra i principali sistemi di SGRT, attualmente disponibili per l'utilizzo clinico, vi sono **AlignRT** (Vision RT, London, United Kingdom) e **Catalyst™** (C-Rad, Upsalla, Sweden). Nel seguito ci concentreremo sul sistema AlignRT di VisionRT, tramite il quale sono stati acquisiti i segnali oggetto di analisi in questa tesi.

### 1.5.1 AlignRT

**AlignRT** è un sistema di Surface-Guided Radiotherapy (SGRT) per il posizionamento e il monitoraggio dei pazienti in radioterapia. Utilizza 6 telecamere stereoscopiche a infrarossi per creare una mappa 3D della superficie del paziente. Questa mappa viene confrontata con quella costruita dalla TAC di centraggio (su cui è stato costruito il piano di trattamento) e il sistema propone gli spostamenti da effettuare per ottenere una corrispondenza tra le due mappe. Poiché la superficie di riferimento è definita rispetto all'isocentro del trattamento, gli algoritmi calcolano le traslazioni e le rotazioni in sei gradi di libertà (*lateral, longitudinal, vertical, yaw, roll, pitch*) (Figura 1.8) necessarie per correggere la posizione del paziente in tempo reale [15]. Inoltre, se si verifica un superamento delle tolleranze imposte nei sei gradi di libertà, il sistema garantisce l'immediata sospensione del fascio di radiazioni.

Nelle aree anatomiche in cui la superficie del paziente è adatta per il bersaglio della radioterapia, come nel caso del seno e del cervello, l'imaging superficiale è stato rapidamente integrato nella pratica clinica grazie alla sua elevata precisione temporale e spaziale [15].

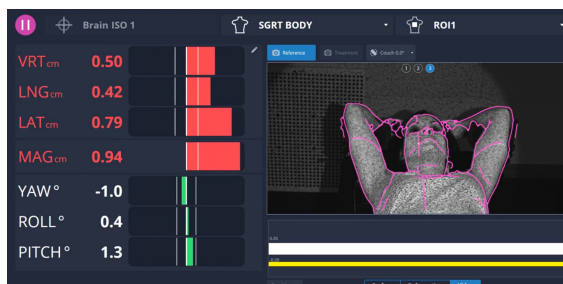


Figura 1.8: Sistema a 6 gradi di libertà in AlignRT Advance

## 1.5.2 Vantaggi e svantaggi della SGRT

Tra i **vantaggi** della SGRT, abbiamo sicuramente l'**alta precisione e accuratezza**. Inoltre, come già citato in precedenza, questa tecnologia prevede un monitoraggio in tempo reale del paziente durante il trattamento, riducendo drasticamente il rischio di errori di posizionamento.

Un altro vantaggio offerto dalla grande precisione di questa tecnologia è la possibilità di **minimizzare l'esposizione dei tessuti sani alle radiazioni**, riducendo il rischio di effetti collaterali. Utilizzando **gating respiratorio** è possibile trattare tumori in distretti respiratori complicati, come, ad esempio, quelli presenti nel torace o nell'addome superiore. La SGRT non richiede l'uso di marcatori o tatuaggi sulla pelle, riducendo il disagio per il paziente ed evitando segni permanenti. Inoltre, non essendo necessarie procedure invasive o fastidiose, il comfort del paziente è significativamente migliorato grazie all'utilizzo della SGRT. La SGRT ha anche il vantaggio di essere **compatibile con altri sistemi di imaging**. Infatti, può essere utilizzata in combinazione con altre tecnologie come CT, MRI e PET [20].

Inoltre, è **sensibile ai cambiamenti della superficie corporea del paziente** (ad esempio, variazioni del peso corporeo) e ciò può influire negativamente sulla precisione.

In alcuni sistemi SGRT, la zona di acquisizione è limitata a una regione di interesse (ROI, dall'inglese Region Of Interest) che rappresenta l'area di trattamento (vedi Figura 1.9), come nel caso di tumore al seno.

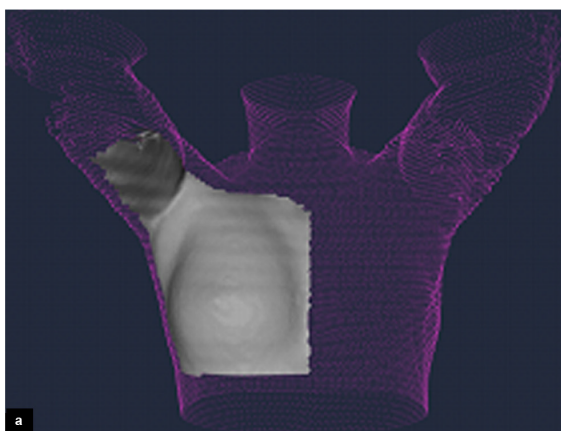


Figura 1.9: Esempio di regione di interesse nella SGRT

## 1.6 Tecniche di controllo del movimento respiratorio

Il **movimento respiratorio** rappresenta la **fonte di incertezza maggiore** sulla posizione del target nel caso di trattamenti a tumori toracici o addominali, risultando in una deviazione della distribuzione di dose rispetto a quella pianificata a sfavore del target stesso e dei tessuti sani [10].

In particolare, nella RT per il cancro al seno, il movimento del target varia solitamente da 1 mm a oltre 20 mm di spostamento in alcuni casi. La maggior parte dei tumori della mammella (circa il 78%) si muove con uno spostamento massimo inferiore a 10 mm *peak-to-peak* [20]. Inoltre, il movimento del seno durante la respirazione varia da 0,8 a 10 mm nella direzione antero-posteriore (AP) (VRT per AlignRT) [20]. È stato dimostrato che l'entità del movimento del target è **più significativa nella direzione AP** rispetto alle direzioni destra-sinistra (LAT per AlignRT) e craniocaudale (LNG per AlignRT) [20]. A causa del movimento respiratorio, si creano degli **artefatti da movimento** anche durante l'acquisizione dell'immagine di centraggio.

Non è sempre possibile osservare direttamente il movimento tumorale durante il trattamento, ma è possibile osservare delle strutture surrogate, ad esempio la parete toracica o dei marker posizionati sulla superficie del paziente, oppure utilizzare il segnale fornito da uno spirometro, il cui segnale è correlato al movimento del tumore.

Sono stati introdotti **diversi sistemi** in grado di **monitorare il movimento degli organi** dovuto alla respirazione; questo è particolarmente importante non solo nei trattamenti della mammella sinistra, ma anche nei trattamenti in cui il target è situato vicino a organi o tessuti che si muovono durante la respirazione e che devono essere contemporaneamente salvaguardati in maniera particolare dalle radiazioni [10].

I metodi più comunemente utilizzati insieme alla SGRT per il controllo del movimento intra-frazione indotto dalla respirazione nel trattamento del cancro al seno sono i **metodi di apnea** (*breath-hold*), i metodi di **gating respiratorio** e i metodi di **tracciamento del tumore in tempo reale** [20].

### 1.6.1 Breath-hold

Le tecniche di *breath-hold* si basano sulla gestione del movimento del target da parte del paziente.

Una delle tecniche di *breath-hold* più diffuse è la *deep-inspiration breath-hold* (**DIBH**), nella quale il paziente viene istruito a trattenere il fiato alla fine dell'inspirazione. Il paziente deve poter familiarizzare con il sistema di DIBH tramite sessioni apposite di allenamento, necessarie anche per studiare l'abilità del paziente a effettuare apnee riproducibili. Il paziente viene sottoposto a due acquisizioni TC: una a respiro libero e una in apnea alla fine dell'inspirazione. Durante le sedute di trattamento, il fascio viene acceso solo quando il livello di apnea, se sufficientemente stabile, si trova in un intervallo prestabilito. Il trattamento è, quindi, suddiviso in segmenti, con pause per permettere al paziente di respirare normalmente.

La DIBH viene utilizzata in radioterapia, in particolare nei tumori della mammella sinistra, a causa della vicinanza dei campi di radiazioni al cuore [10]. Durante l'inspirazione, il diaframma sposta il cuore posteriormente e inferiormente, allonta-

nandolo dal seno, portando a una potenziale riduzione delle tossicità cardiache e polmonari.

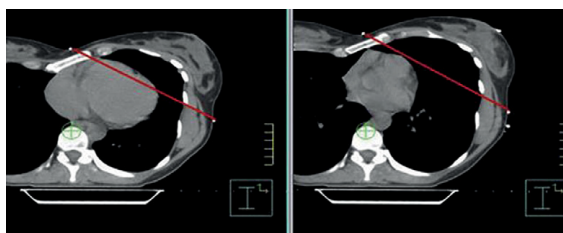


Figura 1.10: Posizione del cuore nelle sezioni assiali di TC dello stesso paziente con carcinoma mammario in condizioni di respiro libero (sinistra) e in DIBH (destra). La linea rossa indica il confine del campo di trattamento tangenziale per l'irradiazione dell'intera mammella (WBI) [20].

Durante un trattamento in DIBH, si utilizza un sistema di monitoraggio per garantire che il paziente mantenga una posizione stabile. Tra le tecnologie di monitoraggio più comunemente usate vi è la **SGRT** [10].

### 1.6.2 Gating respiratorio

Nei metodi di gating, l'irraggiamento viene sincronizzato con il ciclo respiratorio del paziente, il quale respira liberamente, e l'erogazione viene consentita solo in una particolare porzione del ciclo respiratorio, detta **gate** [10].

Il gating respiratorio viene solitamente implementato in due modalità: **gating basato sulla fase** e **gating basato sull'ampiezza**. Il primo si ottiene definendo un insieme di fasi (**gates** o finestre) durante un ciclo respiratorio completo. Il fascio di irradiazione è attivo solo in una o poche finestre. La seconda modalità, invece, si realizza impostando un **valore di soglia** sull'ampiezza del segnale respiratorio. Quando il segnale respiratorio scende sotto la soglia predefinita, il fascio di irradiazione si attiva [20].

### 1.6.3 Metodi di tumor-tracking

In questi metodi, il fascio di radiazioni si muove dinamicamente in modo da seguire la posizione del tumore in movimento; questi sistemi possono essere implementati usando i collimatori multi-leaf o un acceleratore lineare attaccato a un braccio robotico o facendo muovere il lettino in automatico. Idealmente, questi metodi possono eliminare la necessità di introdurre dei margini al CTV per la definizione del PTV [10].

## 1.7 Elaborazione di segnali

Finora abbiamo introdotto brevemente i principali aspetti di teoria delle radiazioni e di radioterapia. Le prossime due sezioni saranno, invece, dedicate all'esposizione dei concetti matematici necessari per comprendere il lavoro svolto in questo elaborato di tesi e descritto nel prossimo capitolo. Al fine di eliminare il più possibile il rumore ad alte frequenze dai segnali, preservando l'integrità del contenuto respiratorio, i

segnali sono stati filtrati prima di essere analizzati. Siccome le tecniche di filtraggio dei segnali sono basate sulla trasformata di Fourier, in questa sezione ne riportiamo la definizione e i principali risultati. L'esposizione che segue, gli esempi e i grafici riportati sono tratti da [5].

### 1.7.1 Classificazione dei segnali

In questo paragrafo riportiamo alcune definizioni di base.

- **Funzioni pari:**  $f(x) = f(-x)$ 
  - Es:  $\cos(x)$ ,  $x^{2n}$
  - $\int_{-a}^a f(x)dx = 2 \int_0^a f(x)dx$
- **Funzioni dispari:**  $f(x) = -f(-x)$ 
  - Es:  $\sin(x)$ ,  $x^{2n+1}$
  - $\int_{-a}^a f(x)dx = 0$
- **Funzioni a supporto compatto** con supporto  $X \subset \mathbb{R}$ :  $f(x) = 0$  se  $x \notin X$ .
  - Es.  $\chi_{[-T,T]} = \begin{cases} 1 & \text{se } x \in [-T, T] \\ 0 & \text{altrimenti} \end{cases}$
  - $\mathbf{C}_0^n(\mathbb{R})$ ,  $0 \leq n \leq \infty$ , è l'insieme di tutte le funzioni  $f : \mathbb{R} \rightarrow \mathbb{C}$  a supporto compatto che sono continuamente differenziabili n-volte.
- **Funzioni periodiche** di periodo T:  $f(x) = f(x + T)$ 
  - Indichiamo con  $L_T^p(\mathbb{R})$ , con  $1 \leq p < +\infty$ , lo spazio di tutte le funzioni periodiche  $f : \mathbb{R} \rightarrow \mathbb{C}$  di periodo T tali che  $\|f\|_{L_T^p(\mathbb{R})} = \left(\frac{1}{T} \int_0^T |f(x)|^p dx\right)^{\frac{1}{p}} < \infty$
  - $L_T^2(\mathbb{R})$  è uno **spazio di Hilbert** dotato del prodotto scalare

$$\langle f, g \rangle = \frac{1}{T} \int_0^T f(x)g^*(x)dx$$

### 1.7.2 Serie di Fourier

Le serie di Fourier sono uno strumento fondamentale per analizzare e sintetizzare funzioni periodiche. Introduciamo prima la seguente famiglia di funzioni T-periodiche:

$$\phi_n(x) = e^{i\frac{2\pi}{T}nx} = \cos\left(\frac{2\pi}{T}nx\right) + i\sin\left(\frac{2\pi}{T}nx\right), \quad \forall n \in \mathbb{Z}$$

detta **base di Fourier**.

**1.7.1 Definizione (Serie di Fourier):** La serie di Fourier di una funzione  $T$ -periodica  $f$ , è formalmente definita come

$$\sum_{n \in \mathbb{Z}} \langle f, \phi_n \rangle \phi_n(x) = \sum_{n \in \mathbb{Z}} \hat{f}_n e^{i \frac{2\pi}{T} n x} \quad (1.1)$$

dove l' $n$ -esimo coefficiente di Fourier,  $\hat{f}_n$ , è definito come

$$\hat{f}_n = \langle f, \phi_n \rangle = \frac{1}{T} \int_0^T f(x) e^{-i \frac{2\pi}{T} n x} dx \quad (1.2)$$

La serie di Fourier di  $f$  può essere vista come il limite del polinomio trigonometrico

$$S_N^f(x) = \sum_{|n| \leq N} \hat{f}_n e^{i \frac{2\pi}{T} n x} \quad (1.3)$$

$$\sum_{n \in \mathbb{Z}} \hat{f}_n e^{i \frac{2\pi}{T} n x} = \lim_{N \rightarrow +\infty} S_N^f(x)$$

Sotto alcune ipotesi, la serie di Fourier **converge** (in qualche senso) e il suo limite coincide con  $f$ . In questi casi possiamo scrivere

$$f(x) = \sum_{n \in \mathbb{Z}} \hat{f}_n e^{i \frac{2\pi}{T} n x} \quad (1.4)$$

dove  $\hat{f}_n = \frac{1}{T} \int_0^T f(x) e^{-i \frac{2\pi}{T} n x} dx$ .

**Osservazione 1.** La serie di Fourier (quando converge) consente di esprimere una funzione  $T$ -periodica  $f$  come somma (infinita) di onde sinusoidali  $\phi_n(x)$  pesate a diverse frequenze. Inoltre, il termine  $\hat{f}_n e^{i \frac{2\pi}{T} n x}$  è detto  **$n$ -esima armonica** di  $f$ .

### Serie di Fourier per funzioni a valori reali

Se  $f : \mathbb{R} \rightarrow \mathbb{R}$ , la sua serie di Fourier può essere scritta come

$$f(x) = \frac{a_0}{2} + \sum_{n=1}^{\infty} a_n \cos\left(\frac{2\pi}{T} n x\right) + b_n \sin\left(\frac{2\pi}{T} n x\right) \quad (1.5)$$

dove  $a_0 = 2\hat{f}_0$ ,  $a_n = \frac{2}{T} \int_0^T f(x) \cos\left(\frac{2\pi}{T} n x\right) dx$  e  $b_n = \frac{2}{T} \int_0^T f(x) \sin\left(\frac{2\pi}{T} n x\right) dx$

*Nota 1.* Ci aspettiamo che una funzione altamente oscillante avrà coefficienti di Fourier grandi in corrispondenza di armoniche ad alta frequenza.

**1.7.1 Esempio:** Consideriamo il seguente polinomio trigonometrico

$$\begin{aligned} f(x) = c_1 \cos(2\pi x) + c_n \cos(2\pi n x) &\implies \hat{f}_{\pm 1} = \frac{c_1}{2}, \\ &\hat{f}_{\pm n} = \frac{c_n}{2}, \\ &\hat{f}_k = 0 \quad \forall k \neq 1, n \end{aligned}$$

Come possiamo vedere dalla Figura 1.11, la forma di  $f(x)$  è più simile alla sua prima armonica (rispetto all' $n$ -esima) se  $c_1$  è significativamente più grande di  $c_n$ .

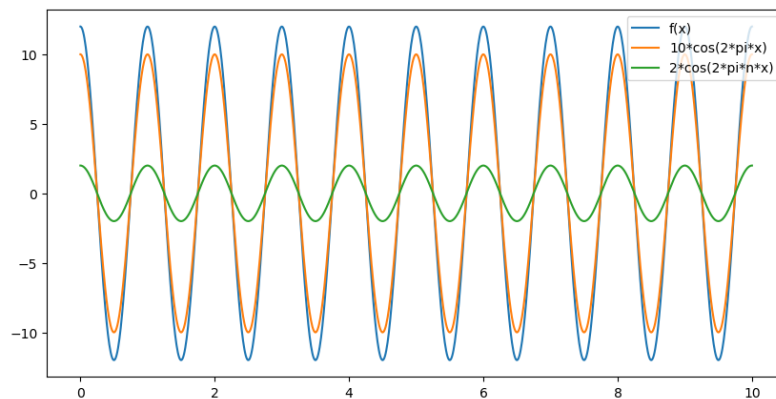


Figura 1.11: Grafico di  $f(x) = 10\cos(2\pi x) + 2\cos(2\pi n x)$  (in blu) e delle sue armoniche  $10\cos(2\pi x)$  (arancione) e  $2\cos(2\pi n x)$  (verde) con  $n = 1000$ .

## Convergenza della serie di Fourier

**1.7.1 Teorema (Carleson, 1966):** Se  $f \in L^2_{2\pi}(\mathbb{R})$ , allora la sua serie di Fourier converge quasi ovunque a  $f$ .

**1.7.2 Corollario:** Se la serie di Fourier **converge** nella norma  $L^1_{2\pi}(\mathbb{R})$ , allora  $f$  è uguale quasi ovunque alla sua serie di Fourier. Inoltre, se  $f$  è **continua** in un qualche punto  $x \in \mathbb{R}$ , allora la sua serie di Fourier converge puntualmente in  $x$ .

## 1.7.3 Trasformata di Fourier (FT)

Uno degli strumenti più importanti sia per l'analisi che per la sintesi di segnali è la trasformata di Fourier e l'analisi armonica a essa legata. Di seguito faremo un riassunto dei risultati più importanti sulla FT per spiegare le sue implicazioni nel processo di analisi/sintesi.

Per semplicità, restringiamo la nostra analisi al caso **unidimensionale**, cioè consideriamo una funzione  $f : \mathbb{R} \rightarrow \mathbb{C}$ .

**1.7.2 Definizione (Trasformata di Fourier):** La Trasformata di Fourier di  $f$  è la funzione  $\hat{f} : \mathbb{R} \rightarrow \mathbb{C}$  definita dal seguente integrale, quando **converge**:

$$\hat{f}(\omega) = \int_{-\infty}^{+\infty} f(x)e^{-i\omega x} dx \quad (1.6)$$

Sotto alcune ipotesi, vale la seguente **formula di inversione**

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \hat{f}(\omega)e^{i\omega x} d\omega \quad (1.7)$$

L'integrale a destra nella formula precedente è detto **Trasformata di Fourier Inversa (IFT)**.

### Relazione con la serie di Fourier

Sia  $f^{2T}(x)$  una funzione  $2T$ -periodica e definiamo la funzione a supporto compatto,  $f$ , con supporto  $[-T, T]$  come

$$f(x) = f^{2T}(x)\chi_{[-T, T]}(x)$$

allora

$$\begin{aligned} \hat{f}_n^{2T} &= \frac{1}{2T} \int_{-T}^T f^{2T}(x)e^{-i\frac{2\pi}{2T}nx} dx \\ &= \frac{1}{2T} \int_{-\infty}^{\infty} f(x)e^{-i\frac{\pi}{T}nx} dx \\ &= \frac{1}{2T} \hat{f}\left(\frac{\pi}{T}n\right) \end{aligned}$$

Questo ci dice che c'è una corrispondenza tra l' $n$ -esimo coefficiente della serie di Fourier della funzione  $2T$ -periodica  $f^{2T}$  e il valore della Trasformata di Fourier della funzione a supporto compatto  $f$  calcolata in  $\frac{\pi}{T}n$ .

L'esempio nella Figura 1.12 mostra che il significato dell'argomento della FT è legato alla frequenza, quindi, anche se trattiamo funzioni non periodiche, ci riferiremo all'argomento della FT come **frequenza**.

**Osservazione 2.** La FT consente di rappresentare una funzione/segnale  $f$  come **somma (integrale) infinita** di componenti sinusoidali  $e^{i\omega x}$  a diverse frequenze e  $\{\hat{f}(\omega) | \omega \in \mathbb{R}\}$  fanno da coefficienti in questa decomposizione, misurando "quanta oscillazione" è presente in  $f$  alla frequenza  $\omega$ . Al contrario, se si conosce lo spettro completo di  $f$ , cioè si conosce  $\hat{f}(\omega)$  per tutte le frequenze  $\omega \in \mathbb{R}$ , allora si può ricostruire (**sintetizzare**) l'intero segnale  $f(x)$  sommando/integrando tutte le componenti  $e^{i\omega x}\hat{f}(\omega)$ , se vale (1.7)

Poichè  $e^{-i\omega x} = \cos(\omega x) - i\sin(\omega x)$ , possiamo scrivere la FT come

$$\hat{f}(\omega) = \int_{-\infty}^{+\infty} f(x)\cos(\omega x)dx - i \int_{-\infty}^{+\infty} f(x)\sin(\omega x)dx \quad (1.8)$$

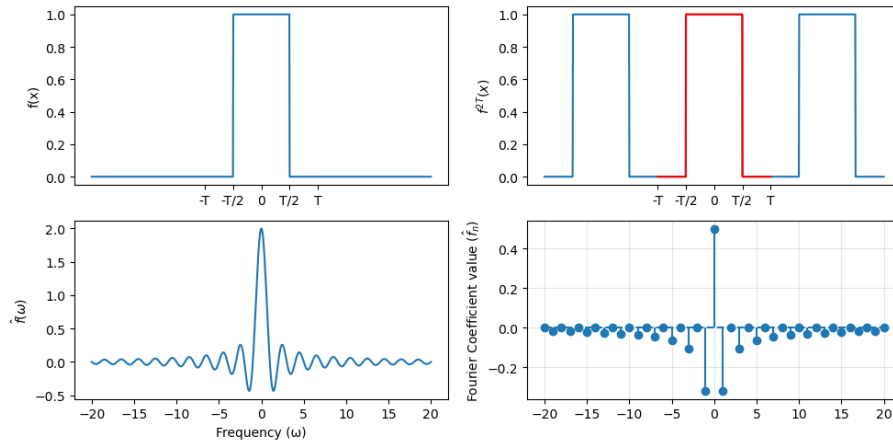


Figura 1.12: Confronto tra i coefficienti della serie di Fourier dell'onda quadra,  $f^{2T}(x)$ , e la trasformata di Fourier,  $\hat{f}(\omega)$ , della funzione a supporto compatto  $f(x)$ .

**1.7.1 Proposizione (Proprietà di simmetria della FT):** *Supponendo che gli integrali in (1.8) convergano, valgono le seguenti proprietà:*

1. se  $f$  è **pari**, allora  $\hat{f}(\omega) = \int f(x)\cos(\omega x)dx = \hat{f}(-\omega)$ , cioè  $\hat{f}$  è **pari**;
2. se  $f$  è **dispari**, allora  $\hat{f}(\omega) = -i \int f(x)\sin(\omega x)dx = -\hat{f}(-\omega)$ , cioè  $\hat{f}$  è **dispari**;
3. se  $f$  è **reale e pari**, allora  $\hat{f}$  è **reale e pari**;
4. se  $f$  è **reale e dispari**, allora  $\hat{f}$  è **immaginaria e dispari**;
5. se  $f$  è **reale**, allora  $\hat{f}(\omega) = (\hat{f}(-\omega))^*$

### Interpretazione della FT

Lo scopo di questa sezione è illustrare quale tipo di informazione qualitativa si può dedurre dallo spettro di una funzione, cioè dalla conoscenza della sua FT su tutto il piano complesso. Per fare questo, consideriamo la seguente funzione

$$f(x) = \cos(3x)e^{-\frac{x^2}{4\pi}}$$

la quale è una senoide di frequenza 3 moltiplicata per una Gaussiana che forza un decadimento rapido su di essa. La sua FT è data da

$$\hat{f}(\omega) = \pi e^{-\pi(\omega-3)^2} + \pi e^{-\pi(\omega+3)^2}$$

Nella Figura 1.13 sono riportati i grafici della funzione e della sua FT. Osserviamo che sia la funzione che la sua FT non hanno supporto compatto, anche se decadono rapidamente a 0 verso  $\pm\infty$ . Da questi grafici possiamo osservare che i picchi della FT corrispondono alle frequenze  $\omega = \pm 3$ , che è esattamente la frequenza della senoide in  $f$ . Pertanto, la FT di  $f$  è stata in grado di determinare il contenuto in frequenza di  $f$ .

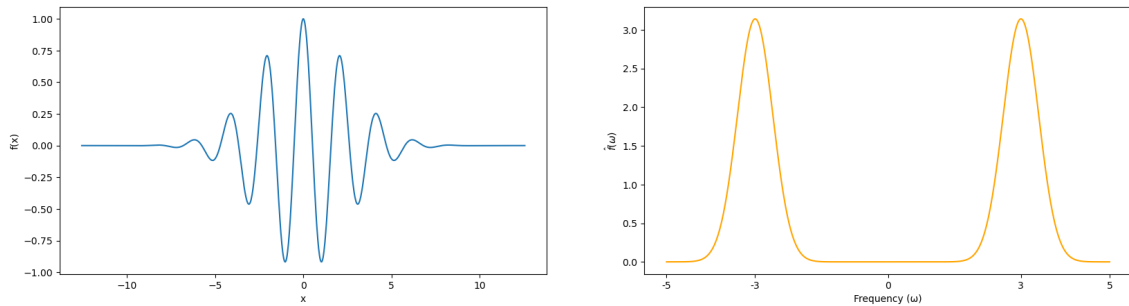


Figura 1.13: Plot della funzione  $f(x) = \cos(3x)e^{-\frac{x^2}{4\pi}}$  (sinistra) e della sua FT (destra).

Per evidenziare meglio le informazioni che si possono ricavare dalla FT di una funzione, consideriamo una leggera variazione dell'esempio precedente:

$$f(x) = (\sin(x) + \cos(10x)/10)e^{-\frac{x^2}{4\pi}} \quad (1.9)$$

Anche in questo caso la FT può essere calcolata analiticamente ed è data da

$$\hat{f}(\omega) = \frac{\pi}{10}(e^{-\pi(\omega-10)^2} + e^{-\pi(\omega+10)^2}) + i\pi(e^{-\pi(\omega-1)^2} - e^{-\pi(\omega+1)^2})$$

I grafici di  $f$  e di  $\hat{f}$  sono riportati nella Figura 1.14. In questo caso la FT è a valori complessi e le informazioni principali sullo spettro devono essere ricavate dal suo modulo. Anche in questo caso possiamo osservare che  $|\hat{f}(\omega)|$  presenta dei picchi corrispondenti alle frequenze delle sue componenti sinusoidali. Inoltre, l'altezza dei picchi è proporzionale all'ampiezza della sinusoide corrispondente. In generale, se  $|\hat{f}(\omega)|$  ha un picco a una certa frequenza  $\omega$ , possiamo aspettarci che  $f$  esibisca  $\omega$  oscillazioni in un qualche intervallo di ampiezza  $2\pi$  contenuto nel suo dominio. Inoltre, l'ampiezza di queste oscillazioni è legata al valore assunto da  $|\hat{f}(\omega)|$  in corrispondenza di quel picco. Per questo motivo,  $|\hat{f}(\omega)|$  è detto anche **risposta in ampiezza** di  $f$ .

*Nota 2.* La FT **non** ci dice la **posizione** sulla retta reale in cui avvengono le oscillazioni. Indica solamente che in un qualche intervallo del suo dominio la funzione si comporta in modo simile ad una sinusoide a una data frequenza.

### Assunzioni e Proprietà

La Trasformata di Fourier viene solitamente introdotta in  $L_1(\mathbb{R})$ , dove l'analisi è più semplice, e poi viene estesa a spazi più generali, come  $L^2(\mathbb{R})$ . Infatti, se  $f \in L_1(\mathbb{R})$ , ovvero

$$\int |f(x)|dx < +\infty,$$

l'integrale in (1.6) converge e, quindi, la FT (1.6) è ben definita per ogni  $\omega \in \mathbb{R}$ , come affermato nel prossimo teorema.

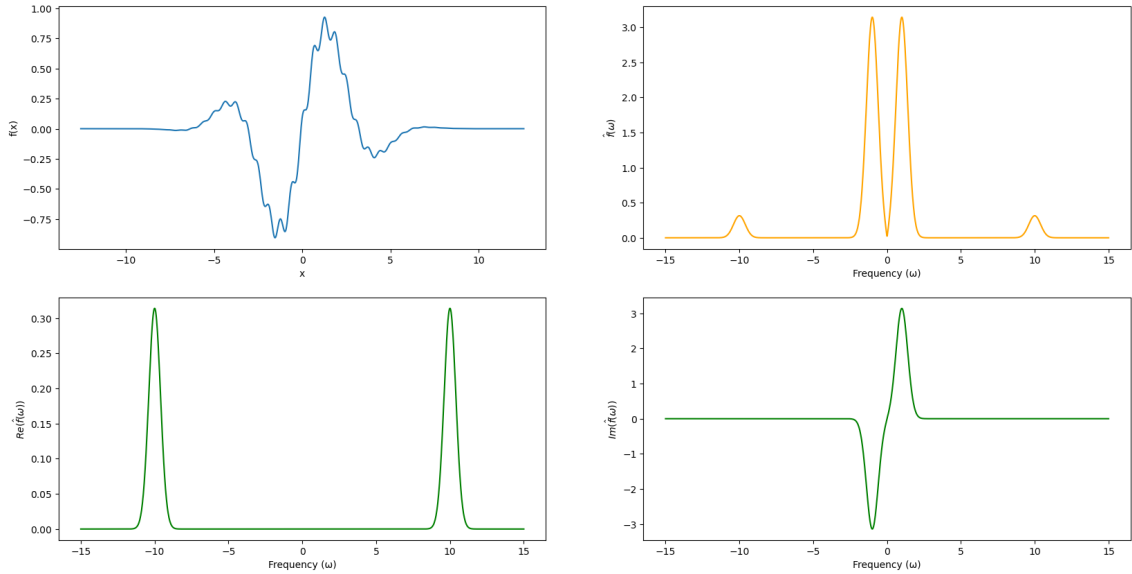


Figura 1.14: Plot della funzione  $f(x) = (\sin(x) + \cos(10x))/10 e^{-\frac{x^2}{4\pi}}$  (in alto a sinistra) e della sua FT; modulo (in alto a destra), parte reale (in basso a sinistra) e parte immaginaria (in basso a destra).

**1.7.3 Teorema (Continuità e limitatezza della FT):** Se  $f \in L_1(\mathbb{R})$ , allora  $\hat{f} : \mathbb{R} \rightarrow \mathbb{C}$  è una funzione **continua** e **limitata** e, in particolare, abbiamo che

$$\|\hat{f}\|_\infty \leq \|f\|_1 \quad (1.10)$$

*Dimostrazione.* Sfruttando le proprietà degli integrali abbiamo che

$$|\hat{f}(\omega)| \leq \int |f(x)| \cdot |e^{-i\omega x}| dx = \int |f(x)| dx < +\infty \quad (1.11)$$

dove, l'ultima disuguaglianza segue dal fatto che  $f \in L_1(\mathbb{R})$ . Inoltre, per un qualunque  $h \in \mathbb{R}$  abbiamo che

$$|\hat{f}(\omega + h) - \hat{f}(\omega)| = \left| \int f(x)(e^{-i(\omega+h)x} - e^{-i\omega x}) dx \right| \quad (1.12)$$

$$\leq \int |f(x)| |e^{-i(\omega+h)x} - e^{-i\omega x}| dx \quad (1.13)$$

$$= \int |f(x)| |e^{-ihx} - 1| \cdot |e^{-i\omega x}| dx \quad (1.14)$$

$$= \int |f(x)| |e^{-ihx} - 1| dx \quad (1.15)$$

In particolare, abbiamo che  $|e^{-ihx} - 1| = |\cos(hx) - i\sin(hx) - 1| = (\cos(hx) - 1)^2 + \sin^2(hx) = 2 - 2\cos(hx) \leq 2$ . Di conseguenza,  $|f(x)| |e^{-ihx} - 1| \leq 2|f(x)| \quad \forall x \in \mathbb{R}$ . Ricordando l'assunzione  $f \in L_1(\mathbb{R})$ , possiamo applicare il Teorema della Convergenza Dominata di Lebesgue, il quale assicura che l'ultimo integrale sopra è finito e

che

$$\lim_{h \rightarrow 0} |\hat{f}(\omega + h) - \hat{f}(\omega)| \leq \int |f(x)| \underbrace{(\lim_{h \rightarrow 0} |e^{-ihx} - 1|)}_{=0} dx = 0$$

che dimostra la continuità di  $f$  □

Il seguente teorema riassume alcune proprietà utili della FT.

**1.7.4 Teorema (Proprietà della FT):** Supponiamo che  $f, g \in L^1(\mathbb{R})$ ,  $a, b \in \mathbb{C}$ ,  $s \in \mathbb{R}$ ,  $s > 0$ ,  $x_0, \omega_0 \in \mathbb{R}$ . Allora

1. se  $g(x) = af(x) + bh(x)$ , allora  $\hat{g}(\omega) = a\hat{f}(\omega) + b\hat{h}(\omega)$ ;
2. se  $g(x) = f(x - x_0)$ , allora  $\hat{g}(\omega) = e^{-ix_0\omega} \hat{f}(\omega)$ ;
3. se  $g(x) = f(\frac{x}{s})$ , allora  $\hat{g}(\omega) = s\hat{f}(s\omega)$ ;
4. se  $g(x) = e^{i\omega_0 x} f(x)$ , allora  $\hat{g}(\omega) = \hat{f}(\omega - \omega_0)$ ;
5. se  $f \in C^n(\mathbb{R})$  e  $f^{(j)} \in L^1(\mathbb{R})$ ,  $\forall j = 0, \dots, n$ , allora  $f^{(n)}(\omega) = (i\omega)^n \hat{f}(\omega)$ ;
6. se  $x \mapsto x^n f(x) \in L^1(\mathbb{R})$ , allora  $\hat{f}$  è differenziabile  $n$  volte e  $(\hat{f})^{(n)}(\omega) = \hat{g}(\omega)$ , dove  $g(x) = (-ix)^n f(x)$ ;
7. **Formula di moltiplicazione:**  $\int f(x)\hat{h}(x)dx = \int \hat{f}(x)h(x)dx$ ;

**1.7.5 Teorema (Riemann-Lebesgue):** Se  $f \in L^1(\mathbb{R})$ , allora

$$\lim_{\omega \rightarrow \pm\infty} \hat{f}(\omega) = 0$$

*Dimostrazione.* Prima di tutto, ricordiamo che  $C_0^\infty(\mathbb{R})$  è denso in  $L^1(\mathbb{R})$ . Di conseguenza  $\exists$  una famiglia di funzioni  $\{f_k\}_{k \in \mathbb{N}} \subset C_0^\infty(\mathbb{R})$  tale che  $f_k \xrightarrow{L^1(\mathbb{R})} f$  al divergere di  $k$ , cioè,  $\lim_{k \rightarrow +\infty} \|f_k - f\|_1 = 0$ .

Passo 1) Osserviamo che, per la (1.10),

$$\lim_{k \rightarrow \infty} \|\hat{f}_k - \hat{f}\|_\infty \leq \lim \|f_k - f\|_{L^1} = 0,$$

cioè,  $\{\hat{f}_k\}_{k \in \mathbb{N}}$  **converge uniformemente** ad  $\hat{f}$ .

Passo 2) Osserviamo che  $g \in C_0^\infty(\mathbb{R})$  implica che  $g^{(n)} \in C_0^\infty(\mathbb{R}) \subset L^1(\mathbb{R})$ , cosa che ci consente di applicare la Proprietà (5) del Teorema precedente:

$$g^{(n)}(\omega) = (i\omega)^n \hat{g}(\omega).$$

Quando  $\omega \neq 0$ , possiamo riscrivere l'uguaglianza precedente come

$$|\hat{g}(\omega)| = \frac{|g^{(n)}(\omega)|}{|(i\omega)^n|} = \frac{|g^{(n)}(\omega)|}{|\omega|^n} \leq \frac{\|g^{(n)}\|_\infty}{|\omega|^n},$$

Ora, osservando che  $\|g^{(n)}\|_\infty$  è finito e indipendente da  $\omega$ , prendendo il limite da entrambe le parti otteniamo che  $\lim_{\omega \rightarrow \pm\infty} |\hat{g}(\omega)| = 0$ .

Posta  $g = f_k$  e usando il fatto che  $\{f_k\}_{k \in \mathbb{N}}$  converge uniformemente ad  $\hat{f}$ , otteniamo il risultato.  $\square$

Un'implicazione interessante del Teorema di Riemann-Lebesgue (R-L) è la seguente:

**1.7.6 Corollario:** *Supposto che  $f \in L^1(\mathbb{R})$ . Se  $f \in C_0^n(\mathbb{R})$ , allora*

$$\hat{f}(\omega) = O\left(\frac{1}{|\omega|^n}\right) \quad \text{per } \omega \rightarrow \pm\infty$$

Questo risultato ci dice che più una funzione è regolare più velocemente la sua FT decade a 0.

Vediamo ora alcuni esempi significativi di funzioni e della loro FT.

**1.7.2 Esempio (Funzione caratteristica):** *Consideriamo ora la funzione caratteristica dell'intervallo  $[-T, T]$ , con  $T > 0$ :*

$$f(x) = \chi_{[-T, T]}(x)$$

La sua FT è data da

$$\hat{f}(\omega) = 2T \operatorname{sinc}(\omega T),$$

dove

$$\operatorname{sinc}(x) = \begin{cases} \frac{\sin(x)}{x} & x \neq 0 \\ 1 & x = 0 \end{cases}$$

è la **funzione seno cardinale**.

In questo caso  $\operatorname{sinc} \notin L^1(\mathbb{R}) \implies \hat{f} \notin L^1(\mathbb{R})$ . Di conseguenza, come vedremo in seguito, la IFT (1.7) non può essere applicata alla funzione caratteristica.

**1.7.3 Esempio (Funzione gaussiana):** *Partiamo dal caso più semplice*

$$f(x) = e^{-x^2} \implies \hat{f}(\omega) = \sqrt{\pi} e^{-\frac{\omega^2}{4}}$$

Osserviamo che la FT di una gaussiana è sempre una gaussiana.

- Dalla (3) del Teorema 1.7.4, con  $s = \frac{1}{\sqrt{a}}$ ,  $e^{-ax^2} \xrightarrow{FT} \sqrt{\frac{\pi}{a}} e^{-\frac{\omega^2}{4a}}$
- Dalla (2)  $e^{-a(x-x_0)^2} \xrightarrow{FT} \sqrt{\frac{\pi}{a}} e^{-ix_0\omega} e^{-\frac{\omega^2}{4a}}$
- Dalla (3) con  $s = \sqrt{2}\sigma$ ,  $e^{-\frac{x^2}{2\sigma^2}} \xrightarrow{FT} \sqrt{2\pi}\sigma e^{-\frac{\sigma^2\omega^2}{2}}$
- Dalla (2)  $e^{-\frac{(x-x_0)^2}{2\sigma^2}} \xrightarrow{FT} \sqrt{2\pi}\sigma e^{-ix_0\omega} e^{-\frac{\sigma^2\omega^2}{2}}$

**1.7.4 Esempio (Funzione Lorenziana):** Consideriamo la funzione

$$f(x) = \frac{1}{(ax)^2 + 1},$$

La sua FT è data da

$$\hat{f}(\omega) = \frac{\pi}{a} e^{-\frac{|\omega|}{a}}$$

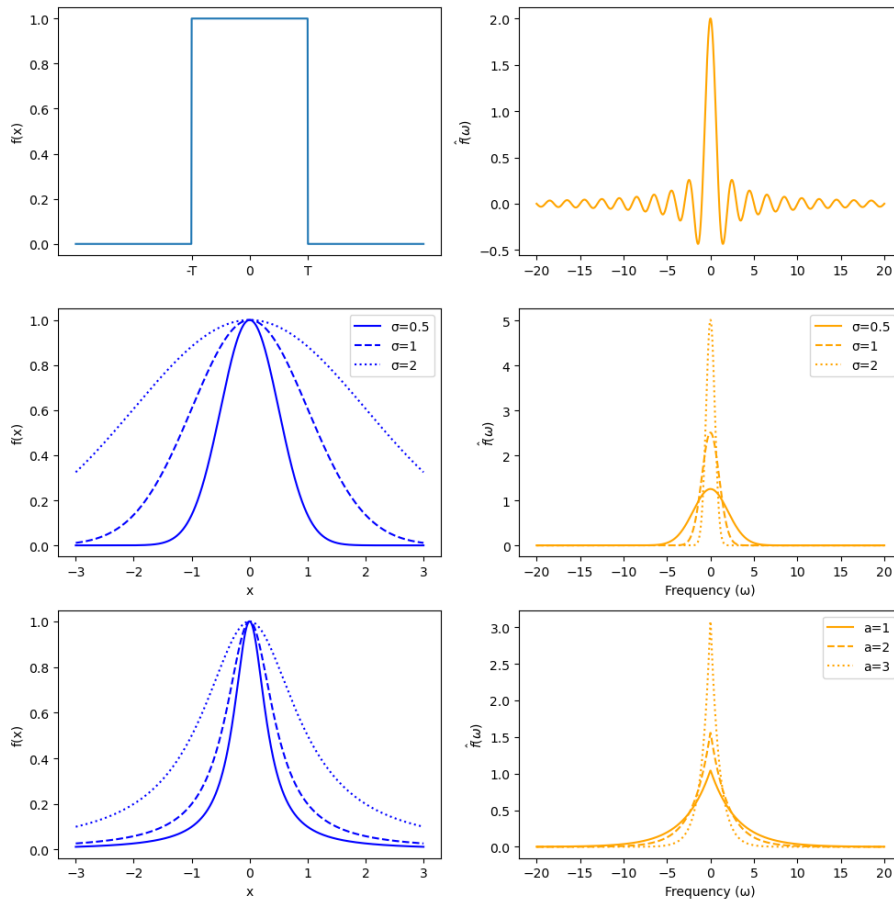


Figura 1.15: Esempi di coppie funzione e FT per la funzione  $f(x) = \chi_T(x)$  (prima riga), la Gaussiana con  $f(x) = e^{-\frac{x^2}{2\sigma^2}}$  per diversi valori di  $\sigma$  (seconda riga) e la Lorenziana  $f(x) = \frac{1}{(ax)^2+1}$  per diversi valori di  $a$  (terza riga).

**1.7.4 Convoluzione e Filtraggio**

Abbiamo visto come analizzare un segnale usando la FT. Ora introduciamo una delle operazioni più interessanti tra funzioni, alla base dell'elaborazione di segnali, che consente di modificare il loro comportamento oscillatorio, attenuando o amplificando le oscillazioni. Questa operazione è detta **convoluzione**, e la sua definizione formale è riportata di seguito.

**1.7.3 Definizione (Convoluzione):** La convoluzione di due funzioni  $f$  e  $h$  è definita come

$$(f * h)(x) = \int f(x - y)h(y)dy = \int f(y)h(x - y)dy. \quad (1.16)$$

**Osservazione 3.** La definizione è **ben posta** quando l'integrale che definisce la convoluzione converge, ovvero, quando  $f \in L^1(\mathbb{R})$  e  $h \in L^\infty(\mathbb{R})$ . In questo caso abbiamo che  $\|f * h\|_\infty \leq \|f\|_1 \|h\|_\infty$ . Più in generale, vale la **disuguaglianza di Young**:

$$\|f * h\|_r \leq \|f\|_p \|h\|_q, \quad \text{con} \quad \frac{1}{r} = \frac{1}{p} + \frac{1}{q} - 1, \quad 1 \leq p, q, r \leq +\infty \quad (1.17)$$

La quale implica che, se  $f \in L^p(\mathbb{R})$  e  $h \in L^q(\mathbb{R})$ , allora  $f * h$  è ben definita e  $f * h \in L^r(\mathbb{R})$ .

Osserviamo inoltre che la convoluzione gode della **proprietà di simmetria**:

$$(f * h)(x) = (h * f)(x)$$

Prima di dare dei risultati teorici sulla convoluzione, è importante sottolineare il significato di questa operazione. In pratica, fare la convoluzione di  $h$  con  $f$  produce una nuova funzione  $g = f * h$ , i cui valori sono una sorta di **media (integrale)** dei valori di  $f$ , dove i valori di  $h$  fanno da pesi. In particolare, per ogni  $x$ , dobbiamo traslare  $f$  di  $x$  e moltiplicarla per  $h$ , per poi integrare il prodotto sulla retta reale. La funzione risultante sarà una versione trasformata di  $f$ , la cui forma dipende dalle caratteristiche di  $h$ . Per sottolineare questo concetto, consideriamo un semplice esempio in cui  $h$  è una funzione a campana e  $f$  è una funzione costante a tratti con supporto compatto.

**1.7.5 Esempio:** Prendiamo  $f$  come una combinazione lineare di funzioni caratteristiche modulate e traslate e ha come una Lorenziana:

$$f(x) = \sum_{i=1}^n \mu_i \underbrace{\chi_{T_i}(x - \tau_i)}_{f_i(x)}, \quad h_a(x) = \frac{1}{1 + (ax)^2} \quad (1.18)$$

In questo caso, sfruttando la linearità dell'integrale, la convoluzione  $f * h_a$  può essere calcolata esplicitamente e ha forma

$$(f * h_a)(x) = \sum_{i=1}^p \mu_i u_i(x)$$

dove

$$u_i(x) = (f_i * h_a)(x) = \int f_i(x - y) h_a(y) dy = \int \chi_{T_i}(x - \tau_i - y) \frac{1}{1 + (ay)^2} dy \quad (1.19)$$

$$= \int_{x-\tau_i-T_i}^{x-\tau_i+T_i} \frac{1}{1 + (ay)^2} dy = \frac{1}{a} [\arctan(ay)]_{x-\tau_i-T_i}^{x-\tau_i+T_i} \quad (1.20)$$

$$= \frac{1}{a} [\arctan(a(x - \tau_i + T_i)) - \arctan(a(x - \tau_i - T_i))]. \quad (1.21)$$

I grafici della funzione  $f$  (con  $T = (\frac{1}{2}, 7, 2, \frac{1}{2}, 1)$ ,  $\tau = (-\frac{2}{3}, 0, 2, 1, 3)$ ),  $h_a$  e  $(f * h_a)$  per  $a \in \{10, 2\}$  sono presenti nella Figura 1.16. Da questo esempio possiamo dedurre un paio di fatti interessanti riguardo alla convoluzione.

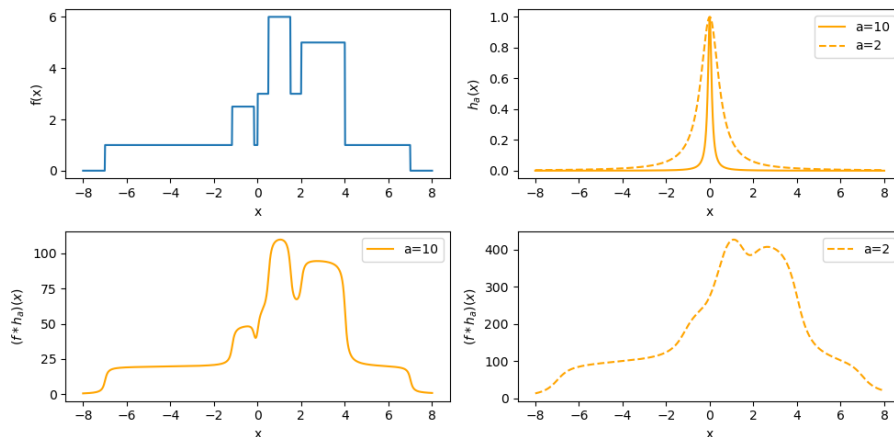


Figura 1.16: In alto: grafici di una funzione costante a tratti (a sinistra) e di una Lorenziana con diverse ampiezze. In basso: grafici della convoluzione tra le funzioni sopra.

**Osservazione 4. (Regolarità)** Prima di tutto, possiamo osservare che la convoluzione  $(f * h_a)$  è una versione lisciata di  $f$ . Tuttavia, poiché l'operazione di convoluzione è simmetrica, possiamo anche affermare che  $(f * h_a)$  è una versione skyline di  $h_a$ . Grosso modo, le caratteristiche di  $(f * h_a)$  sono quelle di  $f$  e  $h_a$  mischiate

insieme. Questo fatto è cruciale quando si vuole elaborare un segnale, ovvero, dato un segnale, costruirne uno nuovo rinforzando o rimuovendo alcune sue caratteristiche. Ad esempio, se vogliamo che il segnale risultante sia liscio, dovremo farne la convoluzione con una funzione liscia.

Il comportamento della convoluzione rispetto alla regolarità è formalmente descritto nel seguente teorema, che, grosso modo, ci dice che la convoluzione  $f * h$  è liscia quanto la più liscia tra  $f$  ed  $h$ .

**1.7.7 Teorema:** Siano  $f \in L^1(\mathbb{R})$  e  $h \in C_b^n(\mathbb{R}) = C^n(\mathbb{R}) \cap L^\infty(\mathbb{R})$  con  $n \in \mathbb{N}$ . Allora

$$g = f * h \in C_b^n(\mathbb{R}) \quad e \quad g^{(j)} = h^{(j)} * f, \quad \forall j = 0, \dots, n$$

**Osservazione 5. (Oscillazioni)** Dalla Figura 1.18, possiamo osservare anche che più è larga  $h_a$ , cioè, più piccolo è  $a$ , più piatta è  $(f * h_a)$ . In particolare, se  $h_a$  è sufficientemente larga, alcune oscillazioni in  $f$  vengono sopresse, come se fossero state assorbite da  $h_a$ . Questo fatto, come spiegato nel teorema seguente, ha una controparte nel dominio delle frequenze.

**1.7.8 Teorema (di Convoluzione):** Siano  $f, h \in L^1(\mathbb{R})$ . Allora  $g = f * h \in L^1(\mathbb{R})$ ,  $\|g\|_1 = \|f * h\|_1 \leq \|f\|_1 \|h\|_1$  e

$$\hat{g}(\omega) = \widehat{f * h}(\omega) = \hat{f}(\omega) \hat{h}(\omega) \tag{1.22}$$

*Dimostrazione.* La norma  $L^1$  del prodotto della convoluzione si può scrivere come

$$\int \left| \int f(x-y)h(y)dy \right| dx \leq \int \int |f(x-y)||h(y)|dydx \tag{1.23}$$

Facciamo vedere prima che il seguente integrale converge:

$$\begin{aligned} \int \left( \int |f(x-y)||h(y)|dx \right) dy &= \int |h(y)| \left( \int |f(x-y)|dx \right) dy \\ &= \int |h(y)| \|f\|_1 dy = \|h\|_1 \|f\|_1 \end{aligned}$$

Di conseguenza, dal Teorema di Fubini, la parte destra della disuguaglianza (1.23) è uguale a  $\|h\|_1 \|f\|_1$ .

$$\implies f * h \in L^1(\mathbb{R}) \quad e \quad \|f * h\|_1 \leq \|f\|_1 \|h\|_1$$

Infine, invocando nuovamente il Teorema di Fubini, segue direttamente dalla definizione di FT che

$$\begin{aligned}
 \widehat{f * h}(\omega) &= \int \left( \int f(x-y)h(y)dy \right) e^{-i\omega x} dx \\
 &= \int \int f(x-y)e^{-i\omega(x-y)}h(y)e^{-i\omega y} dx dy \\
 &= \int h(y)e^{-i\omega y} \left( \int f(x-y)e^{-i\omega(x-y)} dx \right) dy \\
 &= \hat{f}(\omega)\hat{h}(\omega)
 \end{aligned}$$

□

Il teorema precedente ci dice che una convoluzione nel dominio dello spazio corrisponde ad un prodotto nel dominio delle frequenze. In particolare, quando applicato all'esempio (1.18), otteniamo che la FT della funzione  $g(x) = (f * h_a)(x)$  è data da

$$\hat{g}(\omega) = 2a\pi \sum_{k=1}^p \mu_k e^{-i\tau_k \omega} \frac{\sin(\omega T_k)}{T_k} e^{-\frac{|\omega|}{a}}$$

Il cui grafico è riportato nella Figura 1.17. Si può osservare che le oscillazioni della FT della funzione costante a tratti sono ridotte nella FT della convoluzione dalla moltiplicazione con la FT della funzione Lorenziana la quale, specialmente per valori piccoli di  $a$ , decade molto rapidamente.

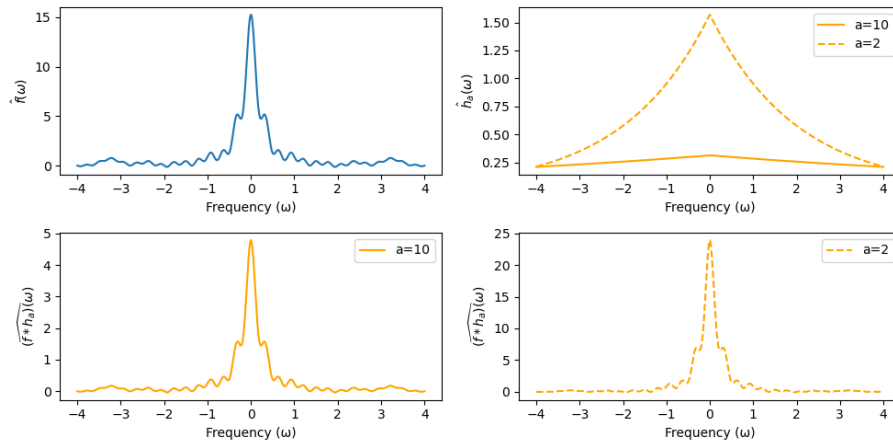


Figura 1.17: In alto: FT della funzione costante a tratti (sinistra) e della Lorenziana (destra) in (1.18). In basso: FT della loro convoluzione.

### 1.7.5 Filtraggio delle frequenze

Il Teorema di convoluzione suggerisce che, se vogliamo sopprimere (o enfatizzare) le oscillazioni di un segnale  $f$  ad una certa frequenza, preservando le altre, dobbiamo farne la convoluzione con una funzione  $h$  la cui FT ha valori piccoli (o grandi) in modulo in corrispondenza di quelle frequenze. Tale funzione,  $h$ , è detta **filtro**, e l'intera

procedura, al centro dell'elaborazione di segnali, è detta **filtraggio**. In particolare,  $h$  è detta **risposta all'impulso** del filtro, a causa della seguente uguaglianza

$$h = h * \delta,$$

dove  $\delta$  è una funzione detta **impulso unitario** o **impulso di Dirac** tale che  $\delta(x) = 0$  ovunque tranne che in  $x = 0$  e  $\int_{\mathbb{R}} \delta(x) = 1$ .

La FT,  $\hat{h}$ , è detta, invece, **risposta in frequenza** di  $h$ , in quanto, presa  $f(x) = e^{i\omega x}$ , abbiamo che

$$(f * h)(x) = \int f(x-y)h(y)dy = \int h(y)e^{i\omega(x-y)}dy = e^{i\omega x} \int h(y)e^{-i\omega y}dy = e^{i\omega x}\hat{h}(\omega).$$

Diamo ora alcune definizioni.

**1.7.4 Definizione (Funzione spazio limitata):** Una funzione a supporto compatto è detta anche **spazio limitata**.

**1.7.5 Definizione (Funzione a banda limitata):** Una funzione  $h$  la cui FT,  $\hat{h}$ , è a supporto compatto è detta **a banda limitata**.

Il supporto di  $\hat{h}$  è la **banda** del segnale  $h$ .

L'ampiezza del supporto è l'**ampiezza di banda** del segnale  $h$ .

Sappiamo dal Teorema di Riemann-Lebesgue che  $\hat{h}$  è limitata e che  $\lim_{|\omega| \rightarrow \infty} \hat{h}(\omega) = 0$ . Di conseguenza, possiamo estendere la definizione di ampiezza di banda anche al caso in cui la FT non è a supporto compatto, come l'ampiezza dell'intervallo nel dominio delle frequenze dove i valori di  $\hat{h}$  sono maggiori o uguali di una frazione del suo valore massimo (ad esempio la metà).

### Classificazione dei filtri

I filtri possono essere classificati in base alla forma della loro FT in:

- **Filtri passa basso:**

- Si usano quando l'obiettivo è attenuare le oscillazioni ad alta frequenza, preservando le basse frequenze.
- **Banda:** un intervallo intorno all'origine.
- **Proprietà tipica:**  $\hat{h}(0) = \int h(x)dx \neq 0$
- Esempi: Gaussiana, Lorenziana.

- **Filtri passa alto:**

- Si usano quando l'obiettivo è ridurre le basse frequenze, preservando le alte frequenze con solo una piccola attenuazione.
- **Banda:** non include un intervallo intorno all'origine.
- **Proprietà tipica:**  $\int h(x)dx \simeq 0$

- Esempi: modulazione di un filtro passa basso, derivata di un filtro passa basso ( $\hat{h}'(\omega) = i\omega\hat{h}(\omega)$ )

- **Filtri passa banda:** si usano quando l'obiettivo è attenuare tutte le frequenze al di fuori di un dato intervallo (banda).

I **filtri ideali** passa basso, passa alto e passa banda sarebbero la funzione  $h$  tale che  $\hat{h}(\omega) = \chi_{\Omega}(\omega)$ , dove  $\Omega = [-\omega_0, \omega_0]$ ,  $\Omega = \{\omega \in \mathbb{R}^n : \|\omega\| \geq \omega_0\}$ ,  $\Omega = [\omega_0 - s, \omega_0 + s]$ , rispettivamente, per qualche  $\omega_0, s \in \mathbb{R}$ ,  $s > 0$ . Sfortunatamente, questo tipo di funzione non esiste in  $L^1(\mathbb{R})$ , in quanto tutte le funzioni in questo spazio hanno una Trasformata di Fourier continua (dal Teorema di Continuità e Limitatezza della FT), ma esiste in  $L^2(\mathbb{R})$ . Infatti, la definizione di FT, il Teorema di Convulsione e la formula di inversione (IFT) possono essere estese in  $L^1(\mathbb{R}) \cap L^2(\mathbb{R})$  (cosa che vedremo più avanti).

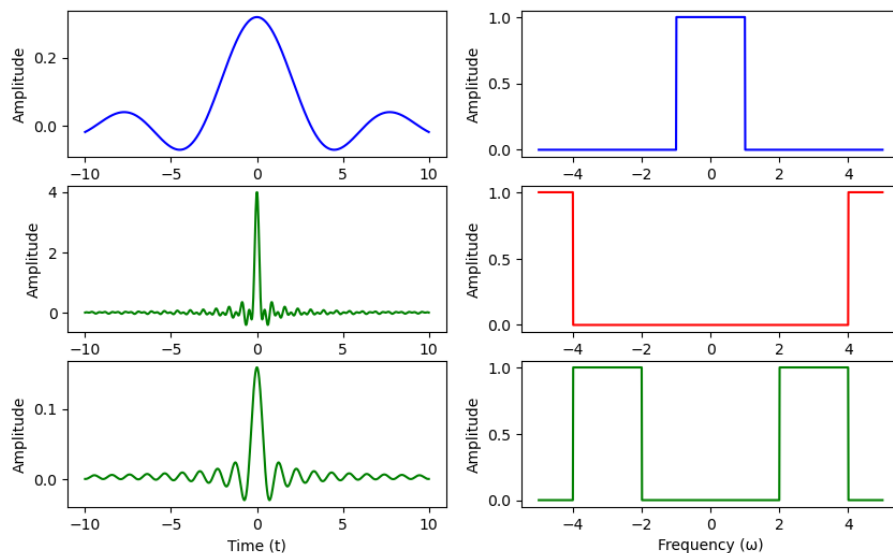


Figura 1.18: Esempi di filtri ideali. Prima riga: filtro ideale passa basso (sinistra) con la sua FT (destra). Seconda riga: filtro ideale passa alto (sinistra) con la sua FT (destra). Terza riga: filtro ideale passa banda (sinistra) con la sua FT (destra).

### 1.7.6 Formula di inversione in $L^1$

Per filtrare le frequenze di un segnale di input  $f$ , possiamo seguire il seguente schema:

- calcolare  $\hat{f}$ ;
- calcolare  $\hat{f} \cdot \hat{h}$ , dove  $h$  è un **filtro**;
- calcolare il nuovo segnale  $g(x)$  come

$$g(x) = \frac{1}{2\pi} \int \hat{f}(\omega)\hat{h}(\omega)e^{i\omega x} d\omega$$

A questo punto, però, la domanda è: " $g(x)$  è **equivalente** a  $f * h$ ?".

La formula di inversione nel prossimo teorema è fondamentale per completare lo schema dell'elaborazione di un segnale basata sulla Trasformata di Fourier, in quanto ci fornisce la fase di sintesi.

**1.7.9 Teorema (Trasformata di Fourier Inversa):** Se  $f, \hat{f} \in L^1(\mathbb{R})$ , allora la seguente uguaglianza vale q.o. in  $\mathbb{R}$

$$f(x) = \frac{1}{2\pi} \int \hat{f}(\omega) e^{i\omega x} d\omega \quad (1.24)$$

In particolare, vale in ogni punto  $x$  tale che  $f$  è **continua** in  $x$ .

Il seguente corollario è una conseguenza diretta della Teorema di inversione.

**1.7.10 Corollario:** Se  $f, \hat{f} \in L^1(\mathbb{R})$ , allora

$$\hat{\hat{f}}(x) = 2\pi f(-x) \quad q.o.$$

e l'uguaglianza vale ovunque  $f$  è continua in  $x$ .

Il Teorema di inversione, quindi, ci dice che lo schema del filtraggio delle frequenze visto sopra è ben definito solo quando  $f, h \in L^1(\mathbb{R})$  e  $\hat{f} \cdot \hat{h} \in L^1(\mathbb{R})$ .

**1.7.11 Corollario:** Se  $f, h \in L^1(\mathbb{R})$  e  $\hat{f} \cdot \hat{h} \in L^1(\mathbb{R})$ , allora

$$(f * h)(x) = \frac{1}{2\pi} \int \hat{f}(\omega) \hat{h}(\omega) e^{i\omega x} d\omega \quad q.o.$$

Queste restrizioni sulla scelta del filtro possono essere rilassate usando una definizione più generale di FT (FT in  $L^2$ ).

### 1.7.7 Estensione della FT in $L^2(\mathbb{R})$

Se  $f \in L^2(\mathbb{R})$  ma  $f \notin L^1(\mathbb{R})$ , la sua FT non può essere calcolata con l'integrale di Fourier (1.6), perchè  $f(x)e^{-i\omega x}$  non è integrabile. In questo caso la FT viene definita come un limite usando la FT di funzioni in  $L^1(\mathbb{R}) \cap L^2(\mathbb{R})$ .

**1.7.6 Definizione (FT in  $L^2(\mathbb{R})$ ):** Sia  $f \in L^2(\mathbb{R})$  e  $\{f_n\}_{n \in \mathbb{N}} \subset L^1(\mathbb{R}) \cap L^2(\mathbb{R})$  una famiglia di funzioni tale che

$$\lim_{n \rightarrow \infty} f_n = f \quad \text{in } L^2.$$

Definiamo la **Trasformata di Fourier** di  $f$ , denotata con  $\mathcal{F}f$ , come il seguente limite

$$\mathcal{F}f = \lim_{n \rightarrow \infty} \hat{f}_n$$

dove il limite è in  $L^2$ , ovvero  $\|\hat{f}_n - \mathcal{F}f\|_{L^2} \rightarrow 0$ .

Il fatto che  $L^1(\mathbb{R}) \cap L^2(\mathbb{R})$  è un sottoinsieme **denso** di  $L^1(\mathbb{R})$  assicura che la precedente definizione sia ben posta, in quanto, per ogni funzione  $f \in L^2(\mathbb{R})$ , possiamo

sempre trovare una famiglia di funzioni  $\{f_n\}_{n \in \mathbb{N}} \subset L^1(\mathbb{R}) \cap L^2(\mathbb{R})$  che converge ad  $f$  in  $L^2$ .

Il seguente teorema ci dice che  $\mathcal{F}$  è un mappa **invertibile** su  $L^2(\mathbb{R})$ .

**1.7.12 Teorema (Teorema di Plancherel):** Siano  $f, h \in L^2(\mathbb{R})$ . Allora, abbiamo

1.  $\|f\|_2 = \frac{1}{\sqrt{2\pi}} \|\mathcal{F}f\|_2$
2.  $\langle f, h \rangle = \frac{1}{2\pi} \langle \mathcal{F}f, \mathcal{F}h \rangle$
3. L'uguaglianza  $\mathcal{F}\mathcal{F}f(x) = 2\pi f(-x)$  vale q.o.  $\implies \mathcal{F}$  è una mappa **invertibile** su  $L^2(\mathbb{R})$  e la sua **inversa** è
4.  $\mathcal{F}^{-1}f(x) = \frac{1}{2\pi} \mathcal{F}f(-x)$  q.o.

Le prime due uguaglianze sono dette **identità di Parseval**.

**Teorema di Convoluzione in  $L^2(\mathbb{R})$**

Osserviamo che, se  $f \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})$ , allora  $\mathcal{F}f = \hat{f}$ . Inoltre tutte le proprietà principali della FT rispetto alla traslazione, scalatura, modulazione e differenziazione, enunciate nel Teorema 1.7.4, valgono anche in  $L^2(\mathbb{R})$ .

**1.7.2 Proposizione:** Se  $f, h \in L^2(\mathbb{R})$ , allora  $(f * h)(x)$  è **ben definita**, in quanto, per la disuguaglianza di Young,

$$\|f * h\|_\infty \leq \|f\|_2 \|h\|_2$$

Si può dimostrare anche che

$$\lim_{|x| \rightarrow \infty} (f * h)(x) = 0$$

Tuttavia, non è garantito che  $f * h$  appartenga a  $L^1(\mathbb{R})$  o a  $L^2(\mathbb{R})$ .

**1.7.13 Teorema (Teorema di convoluzione in  $L^1 - L^2$ ):** Supponiamo che  $f \in L^1(\mathbb{R})$  e  $h \in L^2(\mathbb{R})$ . Allora  $g = f * h \in L^2(\mathbb{R})$  e

$$\mathcal{F}g = \hat{f} \cdot \mathcal{F}h \quad \text{q.o.}$$

**1.7.14 Teorema (Formula di inversione in  $L^1 - L^2$ ):** Se  $f \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})$ , allora  $\mathcal{F}f(\omega) = \hat{f}(\omega)$  e

$$\mathcal{F}\hat{f}(x) = 2\pi f(-x) \quad \text{q.o.}$$

A questo punto, abbiamo un **algoritmo per il filtraggio di frequenze in  $L^1 - L^2$  ben posto**: dato un segnale di input  $f \in L^1(\mathbb{R})$

- **Analisi:** calcolare la FT  $\hat{f}$ ;

- **Filtraggio** : dato un filtro  $h \in L^2(\mathbb{R})$ , calcolare  $\hat{f} \cdot \mathcal{F}h$ ;
- **Sintesi**: calcolare il nuovo segnale  $\mathcal{F}^{-1}(\hat{f} \cdot \mathcal{F}h)(x)$

Output:  $g = f * h \in L^2(\mathbb{R})$

### 1.7.8 Elaborazione di segnali discreti

Nelle sezioni precedenti abbiamo sviluppato gli strumenti matematici necessari a descrivere, analizzare e manipolare funzioni di variabili continue. Questa sezione punta a trasporre tutto questo materiale da un contesto continuo e infinito a un contesto discreto e finito, ovvero il contesto in cui vivono le applicazioni e i modelli del mondo reale. Infatti, nella maggior parte delle applicazioni, come nel caso di questo lavoro di tesi, i dati a disposizione sono un insieme di misurazioni di una certa quantità fisica che varia nel tempo e/o nello spazio. Queste misurazioni corrispondono alle valutazioni di una funzione (possibilmente sconosciuta) in un insieme di punti. Anche se nelle applicazioni questo insieme di punti è finito, iniziamo supponendo di avere a disposizione una quantità numerabile di valori di  $f$ , come meglio formalizzato nella seguente definizione.

**1.7.7 Definizione:** Sia  $f$  una funzione definita su  $\mathbb{R}$  e  $\{x_j\}_{j \in \mathbb{Z}} \subset \mathbb{R}$  una sequenza senza punti di accumulazione.

I valori  $\{f_j = f(x_j)\}_{j \in \mathbb{Z}}$  vengono chiamati **campioni** di  $f$  in  $\{x_j\}_{j \in \mathbb{Z}}$  e i punti  $x_j$  vengono chiamati **punti di campionamento**.

Se  $x_j = x_0 + j\delta_x$ ,  $j \in \mathbb{Z}$ , per qualche  $x_0 \in \mathbb{R}$ , dove  $\delta_x$  è un numero positivo fissato, questi sono detti **punti di campionamento equispaziati**; il numero  $\delta_x$  è detto **spaziatura di campionamento**. Il reciproco  $f_s = \frac{1}{\delta_x}$  è detto **tasso o frequenza di campionamento**

Come già detto, nelle applicazioni reali i dati non sono solo discreti, ma anche finiti. Questo tipo di segnali viene rappresentato da vettori e matrici di numeri reali o complessi. In questa sezione introdurremo gli strumenti per processare segnali in questa forma, che sono le controparti degli operatori funzionali descritti precedentemente.

### Trasformata di Fourier Discreta (DFT)

Introduciamo ora una trasformata per vettori di dimensione finita, analoga alla FT definita per le funzioni, che è connessa in modo significativo alla FT quando le componenti del vettore sono campioni di una funzione.

**1.7.8 Definizione (DFT):** Sia  $\mathbf{f} = (f_0, f_1, \dots, f_{N-1})^T \in \mathbb{C}^N$  un vettore  $N$ -dimensionale. Allora la **Trasformata di Fourier Discreta (DFT)** di  $\mathbf{f}$  è il vettore  $N$ -dimensionale  $\hat{\mathbf{f}} = (\hat{f}_0, \hat{f}_1, \dots, \hat{f}_{N-1})^T \in \mathbb{C}^N$  definito da

$$\hat{f}_m = \sum_{n=0}^{N-1} f_n e^{-i \frac{2\pi}{N} mn} \quad m = 0, \dots, N-1 \quad (1.25)$$

Chiaramente, la DFT definita sopra è una **mappa lineare** da  $\mathbb{C}^N$  a  $\mathbb{C}^N$ . Di conseguenza, può essere espressa in termini di un prodotto vettoriale di matrici. Infatti, introducendo la matrice  $\mathbf{F} \in \mathbb{C}^{N \times N}$ , detta **matrice DFT** (o di **Fourier**) definita come

$$F_{mn} = e^{-i\frac{2\pi}{N}mn}, \quad m, n = 0, \dots, N-1 \quad (1.26)$$

abbiamo che la DFT può essere scritta in forma vettoriale anche come

$$\hat{\mathbf{f}} = \mathbf{F}\mathbf{f} \quad (1.27)$$

**1.7.15 Teorema:** La matrice di Fourier  $\mathbf{F}$  è **non singolare** e la sua inversa è data da

$$\mathbf{F}^{-1} = \frac{1}{N}\mathbf{F}^H \quad (1.28)$$

dove  $\mathbf{F}^H$  è l'**Hermitiana** di  $\mathbf{F}$ :

$$(\mathbf{F}^H)_{nm} = F_{mn}^* = e^{i\frac{2\pi}{N}nm}, \quad n, m = 0, \dots, N-1.$$

Inoltre, indicando con  $\mathbf{w}^{(k)}$  la  $k$ -esima colonna di  $\mathbf{F}^H$  e  $\mathbf{v}^{(k)} = \frac{1}{\sqrt{N}}\mathbf{w}^{(k)}$ , l'insieme  $\{\mathbf{v}^{(0)}, \dots, \mathbf{v}^{(N-1)}\}$  è una **base ortonormale** di  $\mathbb{C}^N$ .

*Dimostrazione.* Osserviamo che un elemento della matrice  $\mathbf{F}$  può essere scritto anche come  $F_{mn} = \xi^{mn}$ , dove  $\xi = e^{-i\frac{2\pi}{N}}$  è una **radice primitiva N-esima dell'unità**, in quanto  $\xi^N = e^{-i2\pi} = 1$  e  $\xi^k \neq 1 \quad \forall k < N$ . Inoltre, usando la formula per la somma parziale di una serie geometrica

$$\sum_{k=0}^{N-1} q^k = \frac{1 - q^N}{1 - q}$$

con  $q = \xi^\ell$ , per  $\ell \in \{-N+1, \dots, 0, \dots, N-1\}$ , abbiamo che

$$\sum_{k=0}^{N-1} \xi^{\ell k} = \begin{cases} N & \text{se } \ell = 0 \\ \frac{1 - \xi^{\ell N}}{1 - \xi^\ell} = \frac{1 - (\xi^N)^\ell}{1 - \xi^\ell} = 0 & \text{se } \ell \neq 0 \end{cases} \quad (1.29)$$

Consideriamo ora l'elemento di indice  $(n, j)$  della matrice  $\mathbf{F}^H\mathbf{F}$ :

$$(\mathbf{F}^H\mathbf{F})_{nj} = \sum_{k=0}^{N-1} e^{i\frac{2\pi}{N}nk} e^{-i\frac{2\pi}{N}kj} = \sum_{k=0}^{N-1} e^{-i\frac{2\pi}{N}k(j-n)} = \sum_{k=0}^{N-1} \xi^{k(j-n)} = N\delta_{nj}$$

dove  $\delta_{nj}$  è il delta di Kronecker. Di conseguenza, abbiamo che  $\mathbf{F}^H\mathbf{F} = \mathbf{N}\mathbf{I}$  e, quindi la (1.28) è dimostrata.

Posta  $\mathbf{V} = \frac{1}{\sqrt{N}}\mathbf{F}^H$ , segue che  $\mathbf{V}\mathbf{V}^H = \mathbf{I}$ , ovvero le colonne di  $\mathbf{V}$  sono ortonormali tra di loro.  $\square$

Una diretta conseguenza della (1.28) è la seguente formula di inversione.

**1.7.16 Corollario (Trasformata di Fourier Discreta Inversa (IDFT)):** Sia  $\mathbf{f} \in \mathbb{C}^N$  e sia  $\hat{\mathbf{f}}$  la sua DFT. Allora

$$\mathbf{f} = \frac{1}{N} \mathbf{F}^H \hat{\mathbf{f}}$$

la cui  $n$ -esima componente è definita come

$$f_n = \frac{1}{N} \sum_{m=0}^{N-1} \hat{f}_m e^{i\frac{2\pi}{N}mn} \quad (1.30)$$

Indicando con  $\mathbf{v}^{(n)}$ ,  $n = 0, \dots, N-1$ , le colonne di  $\mathbf{V}$ , la (1.30) si scrive anche come

$$\mathbf{f} = \frac{1}{\sqrt{N}} \sum_{m=0}^{N-1} \hat{f}_m \mathbf{v}^{(m)}$$

Osserviamo, quindi, che le componenti del vettore DFT, a meno di un fattore  $\frac{1}{N}$ , sono i coefficienti di  $\mathbf{f}$  rispetto alla base ortonormale  $\{\mathbf{v}^{(0)}, \dots, \mathbf{v}^{(N-1)}\}$ . La DFT è l'analisi del vettore  $\mathbf{f}$  rispetto a questa base. L'espressione esplicita del  $k$ -esimo elemento di questa base è la seguente

$$\mathbf{v}^{(k)} = \frac{1}{\sqrt{N}} \begin{bmatrix} 1 \\ e^{i\frac{2\pi}{N}k} \\ e^{i\frac{2\pi}{N}2k} \\ \vdots \\ e^{i\frac{2\pi}{N}(N-1)k} \end{bmatrix}$$

Tale base ha le seguenti proprietà:

- $\mathbf{v}^{(0)}$  ha componenti costanti uguali a  $\frac{1}{\sqrt{N}}$ ;
- Supponendo che  $N$  sia pari,  $\mathbf{v}^{(k)} = \mathbf{v}^{(N-k)*}$ , per  $k = 1, \dots, \frac{N}{2} - 1$ . Infatti, abbiamo che

$$(\mathbf{v}^{(N-k)})_m = e^{i\frac{2\pi}{N}(N-k)m} = \underbrace{e^{i2\pi m}}_{=1} e^{-i\frac{2\pi}{N}km} = (\mathbf{v}^{(k)})_m^*$$

- $(\mathbf{v}^{(\frac{N}{2})})_m = (-1)^m$ ,  $m = 0, \dots, N-1$

Le colonne della matrice di Fourier inversa (non normalizzata),  $\mathbf{w}^{(k)}$ , possono essere viste come campioni delle sinusoidi  $e^{i\omega x}$ , alla frequenza  $\omega = \frac{2\pi}{N}k$  e nei punti di campionamento  $\{0, \dots, N-1\}$ .

Ulteriori proprietà della DFT sono raccolte nella seguente proposizione.

**1.7.3 Proposizione (Identità di Parseval):** Siano  $\mathbf{f}, \mathbf{h} \in \mathbb{C}^N$ . Allora,

$$\sum_{n=0}^{N-1} f_n h_n^* = \frac{1}{N} \sum_{m=0}^{N-1} \hat{f}_m \hat{h}_m^*$$

$$\sum_{n=0}^{N-1} |f_n|^2 = \frac{1}{N} \sum_{m=0}^{N-1} |\hat{f}_m|^2$$

È spesso utile considerare la DFT come trasformata di sequenze periodiche, non solo di vettori. Più precisamente,

**1.7.9 Definizione (Estensione periodica di un vettore  $\mathbb{C}^N$ ):** Dato un vettore  $\mathbf{f} \in \mathbb{C}^N$ , possiamo definire la sequenza periodica  $\{f_n\}_{n \in \mathbb{Z}}$  associata ad  $\mathbf{f}$  tale che

$$f_{n+N} = f_n, \quad \forall n \in \mathbb{Z}.$$

Allora, la definizione della DFT in (1.25) può essere estesa a tutti gli indici  $m \in \mathbb{Z}$ , generando la sequenza periodica  $\{\hat{f}_m\}_{m \in \mathbb{Z}}$  tale che  $\hat{f}_{m+N} = \hat{f}_m$ .

Le seguenti proprietà di queste sequenze periodiche possono essere derivate direttamente dalla definizione di DFT.

**1.7.4 Proposizione:** Sia  $\mathbf{f} \in \mathbb{C}^N$  e  $\{f_n\}_{n \in \mathbb{Z}}$  la sequenza periodica ad essa associata. Allora,

1. se  $\mathbf{f}$  è a valori reali, allora  $\hat{f}_m^* = \hat{f}_{N-m}$ ;
2. se  $\mathbf{f}$  è pari, ovvero  $f_{N-n} = f_n$ , allora  $\hat{\mathbf{f}}$  è pari;
3. se  $\mathbf{f}$  è dispari, ovvero  $f_{N-n} = -f_n$ , allora  $\hat{\mathbf{f}}$  è dispari;
4.  $\forall p \in \mathbb{Z}$ ,

$$\sum_{n=0}^{N-1} f_{n+p} e^{-i\frac{2\pi}{N}mn} = e^{i\frac{2\pi}{N}pm} \hat{f}_m$$

Calcolare la DFT tramite applicazione diretta della definizione porterebbe ad una complessità di  $O(N^2)$ , dovuta alla moltiplicazione matrice vettore ( $\mathbf{F}\mathbf{f}$ ). Nel 1965, Cooley e Tukey proposero un algoritmo molto efficiente per il calcolo della DFT ad un costo di  $O(N \log N)$ , noto come **Fast Fourier Transform (FFT)**, molto importante nel campo dell'elaborazione di segnali.

### Convoluzione discreta

**1.7.10 Definizione:** Siano  $\mathbf{f}, \mathbf{h} \in \mathbb{C}^N$ , con le corrispondenti sequenze periodiche di periodo  $N$ . Allora, il loro **prodotto di convoluzione ciclica** è il vettore  $\mathbf{g} \in \mathbb{C}^N$  la cui  $m$ -esima componente è definita come

$$g_m = \sum_{n=0}^{N-1} h_{m-n} f_n, \quad m = 0, \dots, N-1 \quad (1.31)$$

È facile verificare che l'equazione (1.31) considerata per un qualunque  $m \in \mathbb{Z}$ , definisce una sequenza periodica di periodo  $N$ . La notazione comune per il prodotto di convoluzione ciclica è

$$\mathbf{g} = \mathbf{h} * \mathbf{f}.$$

Usando la corrispondenza di  $\mathbf{f}$  e  $\mathbf{h}$  con le loro sequenze periodiche, è facile vedere che il prodotto di convoluzione ciclica è **simmetrico**:

$$\mathbf{h} * \mathbf{f} = \mathbf{f} * \mathbf{h}.$$

Vediamo ora due proprietà del prodotto di convoluzione ciclica che ci serviranno nel seguito:

- se  $f_n = \delta_{n0} \implies \mathbf{f} = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \implies \mathbf{h} * \mathbf{f} = \mathbf{h}$
- se  $f_n = \delta_{n\frac{N}{2}} \implies \mathbf{h} * \mathbf{f} = \mathbf{h}^s$

dove il vettore  $\mathbf{h}^s$  è il vettore ottenuto da  $\mathbf{h}$  scambiando le sue due metà e  $\delta_{nm}$  è il delta di Kronecker.

*Nota 3.* Il vettore la cui  $n$ -esima componente è  $\delta_{n\frac{N}{2}}$  viene spesso chiamato **vettore impulso** e, a meno di una traslazione delle componenti, può essere considerato l'**elemento unitario** del prodotto di convoluzione ciclica.

Un'altra analogia molto interessante con la convoluzione è il seguente teorema, che è la controparte discreta del Teorema di convoluzione.

**1.7.17 Teorema (Teorema di convoluzione ciclica):** Siano  $\mathbf{f}, \mathbf{h}, \mathbf{g} \in \mathbb{C}^N$ , con le corrispondenti sequenze periodiche di periodo  $N$ . Se  $\mathbf{g} = \mathbf{h} * \mathbf{f}$ , allora

$$\hat{g}_m = \hat{h}_m \cdot \hat{f}_m, \quad m = 0, \dots, N-1 \quad (1.32)$$

### Matrici circolanti

Il prodotto di convoluzione ciclica è strettamente legato ad una classe speciale di matrici  $N \times N$ , dette **matrici circolanti** (o **cicliche**) definite sotto.

**1.7.11 Definizione (Matrice circolante):** Una **matrice circolante**  $N \times N$  è una matrice le cui righe (o colonne) sono una permutazione ciclica degli elementi di un vettore  $a \in \mathbb{C}^N$ , cioè

$$\mathbf{A} = \begin{pmatrix} a_0 & a_1 & a_2 & \dots & a_{N-2} & a_{N-1} \\ a_{N-1} & a_0 & a_1 & \dots & a_{N-3} & a_{N-2} \\ & & & \vdots & & \\ a_1 & a_2 & a_3 & \dots & a_{N-1} & a_0 \end{pmatrix}$$

Le matrici circolanti sono definite completamente dalla loro prima riga (o colonna), quindi richiedono solo  $N$  posti in memoria, invece di  $N^2$ .

L'operatore in (1.31) può essere espresso anche in termini di un prodotto matrice-vettore

$$\mathbf{h} * \mathbf{f} = \mathbf{H}\mathbf{f} \quad (1.33)$$

dove  $\mathbf{H}$  è la matrice circolante la cui prima colonna è uguale al vettore  $\mathbf{h}$  stesso:

$$\mathbf{H} = \begin{pmatrix} h_0 & h_{N-1} & h_{N-2} & \dots & h_1 \\ h_1 & h_0 & h_{N-1} & \dots & h_2 \\ h_2 & h_1 & h_0 & \dots & h_3 \\ \dots & \dots & \dots & \dots & \dots \\ h_{N-1} & h_{N-2} & h_{N-3} & \dots & h_0 \end{pmatrix}$$

Ogni matrice circolante ha gli stessi autovettori, come affermato dal prossimo teorema.

**1.7.18 Teorema:** Sia  $\mathbf{H} \in \mathbb{C}^{N \times N}$  una matrice circolante e indichiamo con  $\mathbf{h}$  la sua prima colonna. Allora, gli **autovettori** di  $\mathbf{H}$  sono le colonne della matrice  $\mathbf{V} = \frac{1}{\sqrt{N}}\mathbf{F}^H$ , mentre gli **autovalori** corrispondenti sono le componenti della DFT di  $\mathbf{h}$ .

*Dimostrazione.* Dalla (1.28) sappiamo che  $\mathbf{V}$  ha colonne ortogonali con norma unitaria. Usando le identità (1.33) e (1.27) nella (1.32), otteniamo

$$(\mathbf{F}\mathbf{H}\mathbf{f})_m = \hat{h}_m(\mathbf{F}\mathbf{f})_m$$

che può essere scritta anche come

$$\mathbf{F}\mathbf{H} = \mathbf{\Lambda}_h\mathbf{F}$$

dove  $\mathbf{\Lambda}_h \in \mathbb{C}^{N \times N}$  indica la matrice diagonale i cui elementi diagonali sono solo gli  $\hat{h}_m$ . Allora segue che

$$\mathbf{H} = \mathbf{V}\mathbf{\Lambda}_h\mathbf{V}^H \quad (1.34)$$

il che significa che le colonne ortonormali di  $\mathbf{V}$  sono gli autovettori di  $\mathbf{H}$ , mentre gli autovalori sono gli elementi diagonali di  $\mathbf{\Lambda}_h$ .

□

In pratica, per ottenere l'intero spettro di una matrice circolante, è sufficiente calcolare la DFT della sua prima riga (o colonna). Questa proprietà è molto utile anche per determinare se una matrice circolante è non singolare e, in questo caso calcolare la sua inversa.

**1.7.5 Proposizione:** Sia  $\mathbf{H} \in \mathbb{C}^{N \times N}$  una matrice circolante. Indichiamo con  $\mathbf{h}$  la sua prima colonna. La matrice  $\mathbf{H}$  è **non singolare**  $\Leftrightarrow \hat{h}_m \neq 0 \quad \forall m = 0, \dots, N-1$ . In questo caso

$$\mathbf{H}^{-1} = \mathbf{V}\Lambda_h^{-1}\mathbf{V}^H$$

La proposizione precedente è una diretta conseguenza dell'equazione (1.34) e del fatto che  $\mathbf{V}$  è una matrice ortogonale.

Da questa proposizione, segue che, conoscendo  $\mathbf{g}, \mathbf{h} \in \mathbb{C}^N$  tali che  $\mathbf{g} = \mathbf{h} * \mathbf{f} = \mathbf{H}\mathbf{f}$ , possiamo ricavare  $\mathbf{f}$  come:

$$\mathbf{f} = \mathbf{H}^{-1}\mathbf{g} = \mathbf{V}\Lambda_h^{-1}\mathbf{V}^H\mathbf{g}$$

### 1.7.9 Filtri discreti

Filtrare un segnale discreto, significa farne la convoluzione con un altro segnale discreto dato, detto **filtro** (o *kernel*), ottenendone uno nuovo le cui caratteristiche dipendono dalle proprietà del filtro. Poichè il teorema di convoluzione ciclica stabilisce un'equivalenza tra la convoluzione nel dominio dello spazio e la moltiplicazione nel dominio delle frequenze, il processo di filtraggio può essere completamente caratterizzato nel dominio delle frequenze. Per questa ragione i filtri vengono classificati in base alla loro risposta in frequenza, cioè la sua DFT shiftata, in quanto mostra quali frequenze saranno attenuate o amplificate nel segnale filtrato.

#### Rappresentazione dei filtri

Dato un filtro discreto  $\mathbf{h} \in \mathbb{C}^N$ , possiamo rappresentarlo principalmente in due modi:

- Risposta all'impulso  $\mathbf{h}^{IR}$ ;
- Risposta in frequenza  $\mathbf{h}^{FR}$ ;

#### (1) Risposta all'impulso

$$\mathbf{h}^{IR} \equiv \mathbf{h} = \begin{pmatrix} h_0 \\ \vdots \\ h_{N-1} \end{pmatrix} \longleftrightarrow (\mathbf{e}_0, \dots, \mathbf{e}_{N-1})$$

Consiste nella solita rappresentazione di un vettore mediante le sue componenti rispetto alla base canonica di  $\mathbb{C}^N$ ,  $(\mathbf{e}_0, \dots, \mathbf{e}_{N-1})$ . Viene chiamata risposta all'impulso a causa dell'uguaglianza

$$\mathbf{h} = \mathbf{h} * \delta,$$

dove  $\delta = (1, 0, \dots, 0)^T \in \mathbb{C}^N$  è detto **vettore impulso**.

## (2) Risposta in frequenza

Abbiamo visto che la base  $\{\mathbf{v}^{(0)}, \dots, \mathbf{v}^{(N-1)}\}$ , formata dalle colonne della matrice di Fourier moltiplicate per un fattore  $\frac{1}{\sqrt{N}}$ , è una base ortonormale di  $\mathbb{C}^N$  e che possiamo rappresentare un segnale  $\mathbf{h} \in \mathbb{C}^N$  rispetto a questa base tramite le componenti della sua DFT come

$$\mathbf{h} = \sum_{m=0}^{N-1} \hat{h}_m \mathbf{v}^{(m)}.$$

Di conseguenza, supposto che le componenti del vettore  $\mathbf{h}$  siano campioni di un segnale  $h$  di una variabile reale, la **risposta in frequenza** di  $\mathbf{h}$  è definita come

$$\mathbf{h}^{FR} \equiv \delta_x (\hat{\mathbf{h}}^s)^s \quad (1.35)$$

dove  $\delta_x$  è la spaziatura del campionamento.

La ragione di questo doppio shift è la seguente uguaglianza approssimata

$$h_m^{FR} \simeq \hat{h}(\omega_m) \quad (1.36)$$

dove  $\omega_m = -\frac{\pi}{\delta_x} + \frac{2\pi}{N\delta_x}m$ ,  $m = 0, \dots, N-1$ .

Tramite il doppio shift, quindi, l'**m-esima componente** della risposta in frequenza di  $\mathbf{h}$  corrisponde al **valore della risposta in frequenza** del segnale  $h$  alla frequenza  $\omega_m = -\frac{\pi}{\delta_x} + \frac{2\pi}{N\delta_x}m$ .

## Classificazione dei filtri

La classificazione dei filtri discreti è basata sulla forma della loro risposta in frequenza:

- **passa basso:** la risposta in frequenza ha componenti nulle (o piccole) in corrispondenza delle alte frequenze (indici vicini a 0 e vicini a  $N-1$ );
- **passa alto:** la risposta in frequenza ha componenti nulle (o piccole) in corrispondenza delle basse frequenze (indici intorno a  $\frac{N}{2}$ );
- **passa banda:** la risposta in frequenza ha componenti nulle (o piccole) ovunque tranne che in corrispondenza di una data banda di frequenze;

## 1.8 Test statistici

In questa sezione verrà riassunto il funzionamento dei test statistici utilizzati in questo lavoro di tesi (Shapiro-Wilk e Wilcoxon-Mann-Whitney).

### 1.8.1 Test di normalità

Una valutazione della normalità dei dati è un prerequisito per molti test statistici, poiché la normalità dei dati è un'ipotesi di base nei test parametrici. Esistono due principali tipi di metodi per valutare la normalità: metodi **grafici** e metodi **numerici** (inclusi i test statistici).

I **test statistici** hanno il vantaggio di fornire un giudizio oggettivo sulla normalità, ma hanno lo svantaggio di non essere talvolta sufficientemente sensibili con

dataset di piccole dimensioni o di essere eccessivamente sensibili con dataset di grandi dimensioni. L'**interpretazione grafica** ha il vantaggio di consentire un buon giudizio per valutare la normalità in situazioni in cui i test numerici potrebbero essere troppo o troppo poco sensibili. Tuttavia, la valutazione della normalità tramite metodi grafici richiede molta esperienza per evitare interpretazioni errate. Se non si ha una buona esperienza, è meglio fare affidamento sui metodi numerici.

Esistono diversi test statistici per verificare la normalità dei dati continui; tra questi, i più noti e ampiamente utilizzati sono il test di **Shapiro–Wilk**, il test di Kolmogorov–Smirnov, il test di Anderson-Darling e il test di D'Agostino-Pearson. Per tutti questi test, l'ipotesi nulla afferma che i dati provengono da una popolazione distribuita normalmente. Quando  $p > 0.05$ , l'ipotesi nulla viene accettata e i dati vengono considerati distribuiti normalmente [14]. Il test di Shapiro–Wilk è il metodo più appropriato per dataset di piccole dimensioni ( $< 50$  campioni), sebbene possa essere utilizzato anche con un numero di campioni ( $n$ ) maggiore. Per questo motivo, in questo lavoro di tesi, si è scelto di utilizzare il test di **Shapiro–Wilk**, avendo a che fare con dataset di piccola dimensione.

### Shapiro-Wilk

Proposto nel 1965, questo test calcola una statistica  $W$  che testa se un campione aleatorio  $\mathbf{x} = (x_1, \dots, x_n)$ , proviene da una distribuzione normale. Valori piccoli di  $W$  indicano un distacco dalla normalità. La statistica  $W$  viene calcolata come segue:

$$W = \frac{(\sum_{i=1}^n a_i x_{(i)})^2}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

dove gli  $x_{(i)}$  sono i valori del campione ordinati in ordine crescente, le  $a_i$  sono costanti generate da medie, varianze e covarianze delle statistiche d'ordine di un campione di dimensione  $n$  proveniente da una distribuzione normale, e  $\bar{x}$  è la media campionaria di  $\mathbf{x}$  [17].

### 1.8.2 Test U di Mann-Whitney

Il **test U di Mann-Whitney**, noto anche come **test dei ranghi di Wilcoxon** o **test di Wilcoxon-Mann-Whitney (WMW)**, è un test statistico non parametrico, la cui ipotesi nulla è che due popolazioni provengano dalla stessa distribuzione, ovvero che non ci sia nessuna differenza significativa tra i due gruppi. L'ipotesi alternativa, invece, è che un gruppo tende ad avere valori significativamente più grandi della variabile di interesse rispetto all'altro.

Al contrario dei test parametrici, come il test t e ANOVA, il test di Mann-Whitney **non fa assunzioni di normalità**, cioè non assume che i gruppi provengano da distribuzioni normali. Di conseguenza, se le assunzioni del test t non valgono, ma le osservazioni sono **indipendenti**, il test di Mann-Whitney può essere un metodo appropriato per confrontare due gruppi, se i dati sono almeno di **natura ordinale**. Una misurazione ordinale non deve necessariamente avere intervalli uguali tra le unità, ma deve essere composta da numeri ordinati (ad es. primo, secondo, terzo,...).

Il test di Mann-Whitney può essere utilizzato per confrontare due gruppi senza fare alcuna assunzione sulle distribuzioni. Viene utilizzato, solitamente, quando non

c'è ragione di credere che i dati siano distribuiti normalmente, quando si hanno pochi campioni a disposizione o quando le varianze sono eterogenee [12].

I dati di ciascun gruppo vengono innanzitutto ordinati in ordine crescente in base al valore della variabile dipendente; dopodiché, ai valori dell'intero dataset viene assegnato un rango: al soggetto con valore minore viene assegnato rango 1, a quello con il secondo valore più piccolo viene assegnato rango 2, e così via, finché al soggetto con valore più alto viene assegnato rango  $N$ , dove  $N = n_1 + n_2$  è il numero totale di campioni presenti nel dataset.

I ranghi vengono poi sommati per ciascun gruppo e il valore di  $U$  è determinato come il più piccolo tra i seguenti due valori:

$$U_1 = n_1 n_2 + \frac{n_1(n_1 + 1)}{2} - R_1$$

$$U_2 = n_1 n_2 + \frac{n_2(n_2 + 1)}{2} - R_2$$

dove  $n_1$  e  $n_2$  sono rispettivamente il numero di campioni presenti nel gruppo 1 e 2;  $R_1$  e  $R_2$  sono le somme dei ranghi per il gruppo 1 e 2.

La statistica  $U$  viene poi confrontata con il valore critico riportato nelle tabelle appropriate al livello di significatività,  $\alpha$ , scelto (solitamente 0.05) e per i valori di  $n_1, n_2$ . A questo punto, se  $U \leq U_{critico}$ , l'ipotesi nulla viene rifiutata, altrimenti viene accettata. Se l'ipotesi nulla viene accettata, significa che non c'è nessuna differenza significativa tra i due gruppi.

## 1.9 Caratterizzazione del movimento respiratorio e di segnali

La modellizzazione del movimento respiratorio interpretato come segnale è un argomento che è stato affrontato nella letteratura scientifica degli ultimi anni. Di seguito, vengono presentati gli articoli che sono stati utilizzati come base del metodo sviluppato in questo elaborato.

Ashe et al.[3] hanno sviluppato un metodo per rilevare i respiri utilizzando sensori di movimento, applicando un filtro sulla base dei valori di frequenza ragionevolmente attesi in questo tipo di movimenti. Utilizzando i punti di riferimento di fase istantanei dalla rappresentazione analitica dei segnali cinematici e di flusso, hanno, poi, identificato i singoli respiri e derivato segnali di frequenza respiratoria (FR) a 1 Hz.

Renner et al.[21], hanno dimostrato che codificare esplicitamente la natura ciclica del segnale respiratorio nei dati di training consente una significativa riduzione dei dataset di training ottenibili da volontari sani. La tecnica di filtraggio descritta in questo articolo è stata adattata al nostro dataset e utilizzata nello sviluppo di questo elaborato di tesi.

Infine, Richter et al.[22], hanno valutato l'influenza del movimento respiratorio nella radioterapia postoperatoria dell'intera mammella (WBI). I valori di ampiezza riportati in questo articolo sono stati presi come riferimento letterario per quelli ottenuti in questa tesi.

# Capitolo 2

## Materiale e metodi

### 2.1 Dataset e pazienti

È stato condotto uno studio prospettico su 177 pazienti con cancro al seno, trattati presso l'AUSL di Reggio Emilia e sottoposti a trattamento di WBI con tecnica 3D-CRT a respiro libero suddiviso in 13/15 frazioni. Durante ogni frazione, lo spostamento della superficie della ROI, posizionata in corrispondenza del seno, è stato monitorato su 6 gradi di libertà (VRT, LNG, LAT, YAW, ROLL, PITCH) dal sistema di AlignRT, il quale restituisce, di conseguenza, 7 segnali di movimento: un segnale per ogni grado di libertà più un segnale MAG, che rappresenta lo spostamento complessivo lungo i tre assi (VRT, LNG e LAT), dove il valore del segnale MAG all'istante  $t$  si ottiene come norma euclidea del vettore  $(VRT(t), LNG(t), LAT(t))$  in  $\mathbb{R}^3$ :

$$MAG(t) = \sqrt{VRT(t)^2 + LNG(t)^2 + LAT(t)^2}$$

Le colonne del dataset restituito da AlignRT per ogni frazione sono 22. Di queste, quelle che ci interessano sono:

- **Elapsed Time(sec)**: tempo di acquisizioni in secondi
- **D.MAG(mm)**: valore del segnale MAG in mm
- **D.VRT(mm)**, **D.LNG(mm)**, **D.LAT(mm)**: spostamenti in mm su 3 direzioni (verticale, longitudinale e laterale, rispettivamente)
- **Reported Beam State**: variabile binaria che vale 1 quando il LINAC emette il fascio di radiazioni e 0 altrimenti.

### 2.2 Elaborazione del segnale

Vediamo ora come sono stati elaborati i segnali delle singole frazioni.

#### 2.2.1 Pre-processing

Ogni segnale è stato ristretto all'intervallo di *beam on* per eliminare eventuali artefatti dovuti al setup del paziente all'inizio e alla fine del trattamento (Figura 3.1). La frequenza di acquisizione ( $fs$ ) media dei segnali da parte di AlignRT è  $(22.0 \pm 4.6)$  Hz. A causa di queste fluttuazioni, tutti i segnali sono stati interpolati (se

$fs < 20$ ) (Figura 3.2) o ricampionati (se  $fs > 20$ ) (Figura 3.3) a una frequenza di campionamento costante di 20 Hz. Per l'interpolazione è stata utilizzata una spline quadratica (funzione `interp1d` di `scipy.interpolate`), mentre per il ricampionamento è stato utilizzato il metodo `resample` di `pandas`. Per ridurre il rumore, abbiamo applicato ai segnali un filtro Butterworth lowpass usando la funzione `butter` di `scipy.signal`. Per un design del filtro più piatto possibile, sono stati selezionati un numeratore e denominatore del filtro di ordine 8 e una frequenza di taglio di 1 Hz, assicurando così l'attenuazione del rumore ad alta frequenza e, contemporaneamente, la preservazione dell'integrità dei segnali di movimento. La scelta della frequenza di taglio è stata basata sul presupposto che il movimento del seno durante il trattamento è dovuto principalmente alla respirazione e che possiamo ragionevolmente aspettarci che la frequenza della respirazione umana si trovi al di sotto di 60 respiri al minuto nella maggior parte delle circostanze.

### Filtro Butterworth

Il filtro di Butterworth rappresenta il miglior compromesso tra attenuazione e risposta in fase ed è uno dei filtri nel dominio delle frequenze più comunemente usati. Non presenta ondulazioni nella *passband* né nella *stopband*, e per questo motivo viene talvolta chiamato **filtro a risposta massimamente piatta**. Il filtro di Butterworth raggiunge questa *flatness* a scapito di una regione di transizione relativamente ampia tra *passband* e *stopband*, con caratteristiche transitorie medie.

La risposta in frequenza di un filtro di Butterworth lowpass di ordine  $n$  e frequenza di taglio  $\omega_c$  (rad/s) è data da:

$$H(\omega) = \hat{h}(\omega) = \sqrt{\frac{1}{1 + (\frac{\omega}{\omega_c})^{2n}}} \quad (2.1)$$

Con il filtro Butterworth, nelle basse frequenze, possiamo ottenere un guadagno più vicino a 1, e man mano che la frequenza aumenta, il guadagno diminuisce. Il modo in cui avviene questa transizione dipende molto dall'ordine del filtro; con un ordine di filtro basso si ottiene una pendenza morbida, mentre con un ordine di filtro più alto la curva può sembrare più simile ad una funzione gradino (*step function*), con transizioni molto nette e a frequenze più alte avrà un guadagno molto basso [2] (vedi Figure 2.1 e 2.2).

### Zero-phase filtering

Il **filtraggio a fase zero (zero-phase filtering)** è una metodologia per ottenere un filtraggio senza alcuna distorsione di fase [16].

Si dice funzione di trasferimento,  $X(z)$ , di un segnale discreto di valori  $x_n$ , la **trasformata z** della sua risposta all'impulso, definita come segue:

$$X(z) = \sum_{n=-\infty}^{+\infty} x_n z^{-n}$$

Siano  $X(z)$ ,  $Y_1(z)$  e  $H(z)$  le funzioni di trasferimento dell'input, dell'output e del filtro rispettivamente. L'output,  $Y(z)$  è dato da:

$$Y_1(z) = H(z)X(z) \quad (2.2)$$

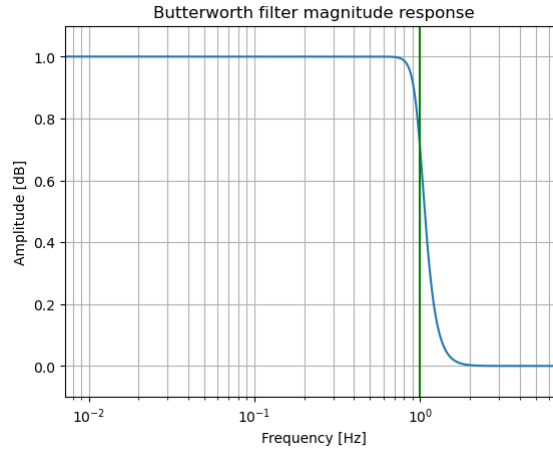


Figura 2.1: Risposta in ampiezza,  $|H(\omega)|$ , di un filtro di Butterworth lowpass di ordine 8 e frequenza di taglio 1 Hz (linea verde) utilizzato in questo lavoro di tesi

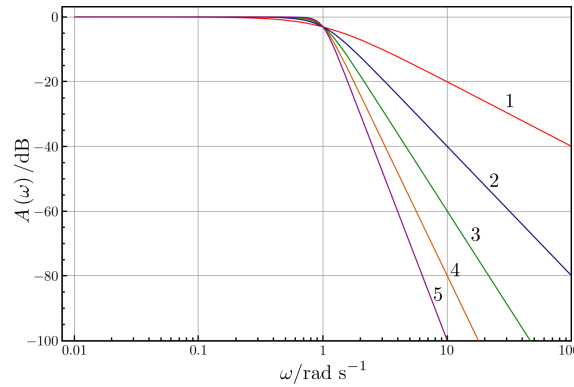


Figura 2.2: Confronto tra la risposta in ampiezza,  $|H(\omega)|$ , del filtro di Butterworth al variare dell'ordine dall'1 al 5 con frequenza di taglio  $\omega_c = 1$  rad/s

Questo è il filtraggio convenzionale, dove in  $Y_1(z)$  viene introdotta dal filtraggio una distorsione di fase non-lineare.

Se il filtro viene applicato a  $Y_1(z)$  nella direzione opposta, l'output  $Y_2(z)$  è dato dalla (2.3). Ora, se  $Y_2(z)$  viene nuovamente invertito temporalmente, l'output risultante  $Y(z)$  è dato dalla (2.4).

$$Y_2(z) = Y_1(z^{-1})H(z) \quad (2.3)$$

$$Y(z) = Y_2(z^{-1}) \quad (2.4)$$

Sostituendo la (2.2) e (2.3) nella (2.4), si ottiene la (2.5) e sostituendo  $z = e^{i\omega}$  (andando a calcolare le funzioni di trasferimento nel cerchio unitario) nella (2.5) si ottiene la (2.6), dove  $X(e^{i\omega}) = X(\omega)$ ,  $Y(e^{i\omega}) = Y(\omega)$  e  $H(e^{i\omega}) = H(\omega)$  sono la risposta in frequenza del segnale di input, di output e del filtro, rispettivamente.

$$Y(z) = Y_1(z)H(z^{-1}) = X(z)H(z)H(z^{-1}) \quad (2.5)$$

$$Y(e^{i\omega}) = X(e^{i\omega})|H(e^{i\omega})|^2 \quad (2.6)$$

L'output finale dato dalla (2.6) è influenzato solo dall'ampiezza del filtro applicato e non presenta alcuna distorsione di fase [16].

Nel nostro caso, è stato effettuato filtraggio a fase zero, utilizzando la funzione `sosfiltfilt` di `scipy.signal`, che consente, quindi, di nullificare la distorsione di fase introdotta dal filtro di Butterworth (Figura 3.4 e 3.5)

### 2.2.2 Identificazione dei picchi

Per individuare i picchi di inspirazione ed espirazione nei segnali, è stata utilizzata la funzione `find_peaks` di `scipy.signal` con dei parametri specifici. Vediamo ora come sono stati calcolati questi parametri.

Abbiamo calcolato la **Trasformata di Fourier discreta** (DFT) unidimensionale dei segnali tramite la funzione `fft` di `scipy.fft`, la quale utilizza a sua volta l'algoritmo di **Fast Fourier Transform** (FFT) (vedi Figura 3.6). Da questa abbiamo ricavato le frequenze presenti nel segnale filtrato e calcolato una loro **media pesata**,  $f_m$ . Infine, abbiamo calcolato il periodo medio dei segnali  $T_m$  come reciproco di  $f_m$ :

$$T_m = \frac{1}{f_m}$$

Tutto questo è servito per calcolare il parametro `dist`, che indica la distanza minima tra picchi successivi, da passare alla funzione `find_peaks`, come  $dist = f_s \frac{2}{3} T_m$ , dove  $f_s = 20Hz$  è la frequenza di campionamento. Come parametro `prominence`, invece, è stato preso 0.1, affinché i picchi siano sufficientemente staccati dalla baseline (con valori superiori si rischia di non individuare dei picchi bassi ma significativi).

Con questi parametri, è stata applicata la funzione `find_peaks` al segnale per trovare i picchi di inspirazione (in rosso nella Figura 3.7 e indicati con R), dopodiché è stata applicata nuovamente al segnale capovolto rispetto all'asse y per trovare i picchi di espirazione (in nero nella Figura 3.7 e indicati con B).

Dopo aver creato un DataFrame per ogni segnale e inserito al suo interno tutti i picchi trovati con tipo (R o B) e rispettive coordinate (Figura 3.8), abbiamo eliminato eventuali picchi successivi dello stesso tipo, e calcolato le ampiezze dei singoli respiri come escursione tra picchi successivi e, solo per il segnale MAG, moltiplicato per 2 questi valori.

Per eliminare eventuali outlier nelle ampiezze, ci siamo ristretti tra il 1° e il 99° percentile (Figura 3.9).

### 2.2.3 Calcolo delle statistiche

Infine, per ogni segnale sono state calcolate **media**, **deviazione standard**, 10°, 25°, 50°, 75° e 90° percentile delle ampiezze e **frequenza media**. Quest'ultima è stata calcolata come rapporto tra numero di picchi di inspirazione (R) e ampiezza dell'intervallo di beam on (in secondi).

L'analisi svolta nella sezione (2.2) è stata ripetuta per ogni frazione di ciascun paziente e le statistiche calcolate sono state raccolte in un unico dataset.

## 2.3 Test di normalità

Il dataset ottenuto include una colonna "PtID" con l'ID del paziente, una colonna "date" con la data della singola frazione e una colonna per ciascuna statistica calcolata in precedenza per ciascun segnale (MAG, VRT, LNG, LAT). Di queste, abbiamo

considerato soltanto l'ampiezza media ( $A_{media}$ ) dei 4 tipi di segnale, raggruppato i valori per ID del paziente ed eseguito un test di normalità di **Shapiro-Wilk** sui 13/15 valori di ampiezza media di ciascun paziente, ritenuti distribuiti normalmente se  $p > 0.05$ .

La scelta del test di normalità di Shapiro-Wilk è stata ritenuta ragionevole, in quanto i valori su cui viene eseguito non sono numerosi in questo caso.

Per valori di  $A_{media}$  che sono stati ritenuti distribuiti normalmente, è stato calcolato un **errore relativo percentuale** ( $ER\%$ ) come segue:

$$ER\%(\mathbf{x}) = \frac{1}{\bar{x}} \frac{S^2}{\sqrt{k}} \cdot 100 \quad (2.7)$$

dove  $\mathbf{x} = [x_1, \dots, x_k]$  è il vettore dei valori di  $A_{media}$  di uno dei segnali di un paziente,  $k$  è il numero di frazioni del paziente,  $\bar{x}$  è la media campionaria di  $\mathbf{x}$  e  $S^2$  è la **varianza campionaria** di  $\mathbf{x}$ :

$$S^2 = \frac{\sum_{i=1}^k (x_i - \bar{x})^2}{k - 1} \quad (2.8)$$

Se, invece,  $\mathbf{x}$  non risulta distribuito normalmente, al posto della media campionaria è stata utilizzata la **mediana**,  $\tilde{x}$ , in (2.7) per calcolare  $ER\%(\mathbf{x})$ .

I valori di  $ER\%$ , così ottenuti, sono stati poi uniti a informazioni cliniche aggiuntive sui pazienti raccolte in precedenza:

- **volume del tumore (CTV)(cc)**: Clinical Target Volume del tumore in cc;
- **età**: età della paziente alla prima frazione;
- **sx/dx**: posizione del tumore (seno sx o dx);
- **Fr**: numero di frazioni (13 o 15);

## 2.4 Classificazione dei pazienti

Al fine di evidenziare due classi di pazienti, li abbiamo suddivisi in base ai loro valori di  $ER\%$  relativo al segnale MAG ( $ER\%_{MAG}$ ) nel seguente modo:

- se  $ER\%_{MAG} \leq cutoff \implies$  il paziente viene classificato come "ripetibile", nel senso che riesce a mantenere un pattern respiratorio sufficientemente costante durante le sedute del trattamento;
- altrimenti  $\implies$  viene classificato come "non ripetibile".

Abbiamo provato a prendere due diversi valori di cutoff, 5% e 10%, in quanto sono entrambi sufficientemente bassi, ma il primo è più selettivo e il secondo meno.

## 2.5 Analisi statistica

Infine, abbiamo cercato una **correlazione** tra l'appartenenza ad una delle due classi sopra definite e le variabili cliniche "volume del tumore (CTV)(cc)", "età" e "sx/-dx", con entrambi i cutoff, utilizzando il **test di Wilcoxon-Mann-Whitney** e le differenze sono state considerate significative per  $p < 0.05$ .

# Capitolo 3

## Risultati

### 3.1 Pazienti e elaborazione del segnale

Le 177 pazienti considerate in questo studio hanno un'età compresa tra i 35 e gli 85 anni (mediana (IQR) = 64(56-71) anni) e volume del tumore (CTV) medio di 708.8 cc (mediana (IQR) = 662.5(461.3-917.9)cc). 108 pazienti sono state sottoposte ad un trattamento suddiviso in 13 frazioni con dose di 3 Gy a frazione e 69 pazienti ad un trattamento suddiviso in 15 frazioni con dose di 2.7 Gy a frazione. 88 pazienti hanno ricevuto radioterapia al seno destro e 89 al seno sinistro.

Di seguito, nelle Figure dalla 3.1 alla 3.9, è riportato un esempio delle varie fasi di elaborazione dei segnali descritte nel capitolo precedente.

I segnali YAW, ROLL e PITCH, non sono stati considerati, in quanto rimangono quasi costanti per tutti i pazienti: la media per paziente dello spostamento medio di YAW, ROLL e PITCH è rispettivamente  $(0.46 \pm 0.17)^\circ$ ,  $(0.38 \pm 0.13)^\circ$ ,  $(0.42 \pm 0.15)^\circ$

Vediamo ora due pazienti esempio: una paziente con  $ER\%_{MAG} < 5\%$  e una con  $ER_{MAG}\% > 10\%$ . La prima paziente (Pz 1) ha 51 anni, volume del tumore (CTV) = 896.2 cc, ha subito un trattamento di 13 frazioni e ha  $ER\%_{MAG} = 4.2\%$ , mentre, la seconda (Pz 2) ha 54 anni, volume del tumore (CTV) = 859.6 cc, ha subito un trattamento in 15 frazioni e ha  $ER\%_{MAG} = 16.3\%$ . Entrambe sono state trattate al seno destro.

Nelle Figure 3.10 e 3.11 sono riportati i grafici dei loro segnali di movimento durante una frazione esempio, nella Tabella 3.1 i valori di ampiezza e frequenza media dei segnali di movimento su tutte le loro frazioni e nelle Figure 3.12 e 3.13 l'andamento delle loro ampiezze e frequenze medie nelle varie frazioni.

ID	MAG		VRT		
	$A_{media}(mm)$	$f_{media}(Hz)$	$A_{media}(mm)$	$f_{media}(Hz)$	
Pz 1	$0.8 \pm 0.2$	$0.26 \pm 0.04$	$0.5 \pm 0.1$	$0.24 \pm 0.04$	
Pz 2	$2.2 \pm 1.1$	$0.34 \pm 0.07$	$1.3 \pm 0.2$	$0.26 \pm 0.04$	
		LNG		LAT	
Pz 1	$0.4 \pm 0.1$	$0.30 \pm 0.05$	$0.3 \pm 0.1$	$0.25 \pm 0.04$	
Pz 2	$1.9 \pm 0.3$	$0.28 \pm 0.02$	$0.4 \pm 0.1$	$0.26 \pm 0.02$	

Tabella 3.1: Ampiezza e frequenza media dei segnali delle due pazienti esempio su tutte le loro frazioni

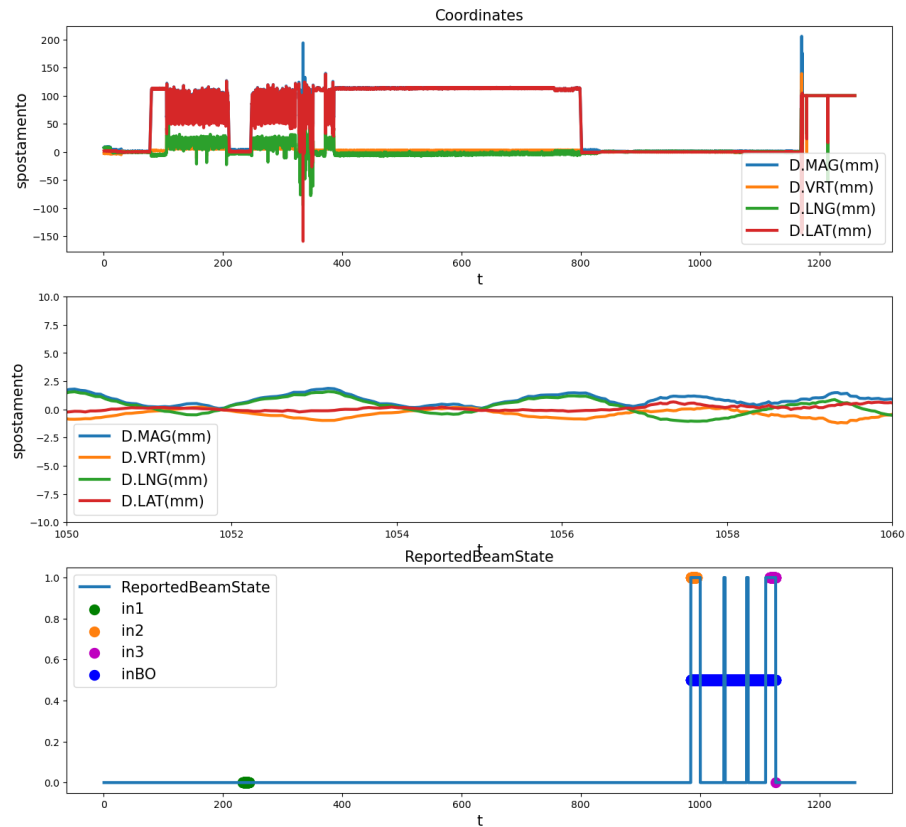


Figura 3.1: Esempio di segnali di una frazione prima del pre-processamento. La prima figura mostra i 4 segnali di una frazione su tutto l'intervallo di trattamento, la seconda è uno zoom degli stessi segnali su un intervallo di 10 secondi, infine, nella terza sono rappresentati i valori della variabile ReportedBeamState e i vari intervalli del trattamento, compreso intBO (intervallo di *beam on*)

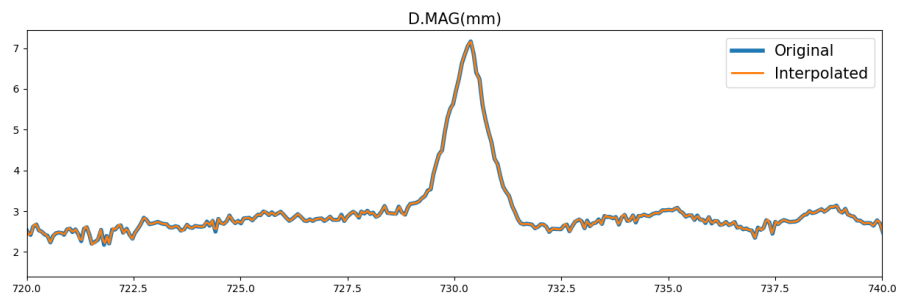


Figura 3.2: Esempio di parte di un segnale MAG con frequenza di campionamento 14 Hz che è stato interpolato con spline quadratica a 20 Hz

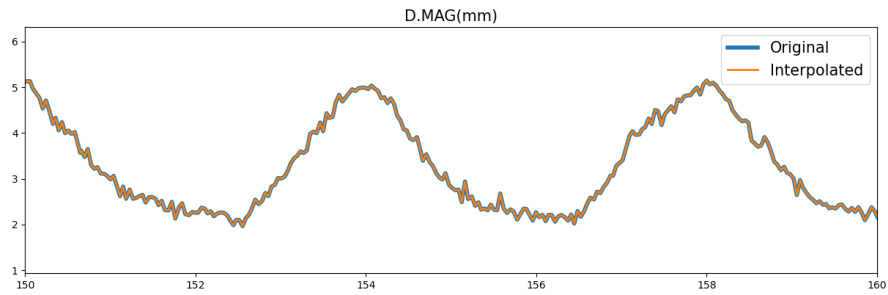


Figura 3.3: Esempio di parte di un segnale MAG con frequenza di campionamento 50 Hz che è stato ricampionato a 20 Hz

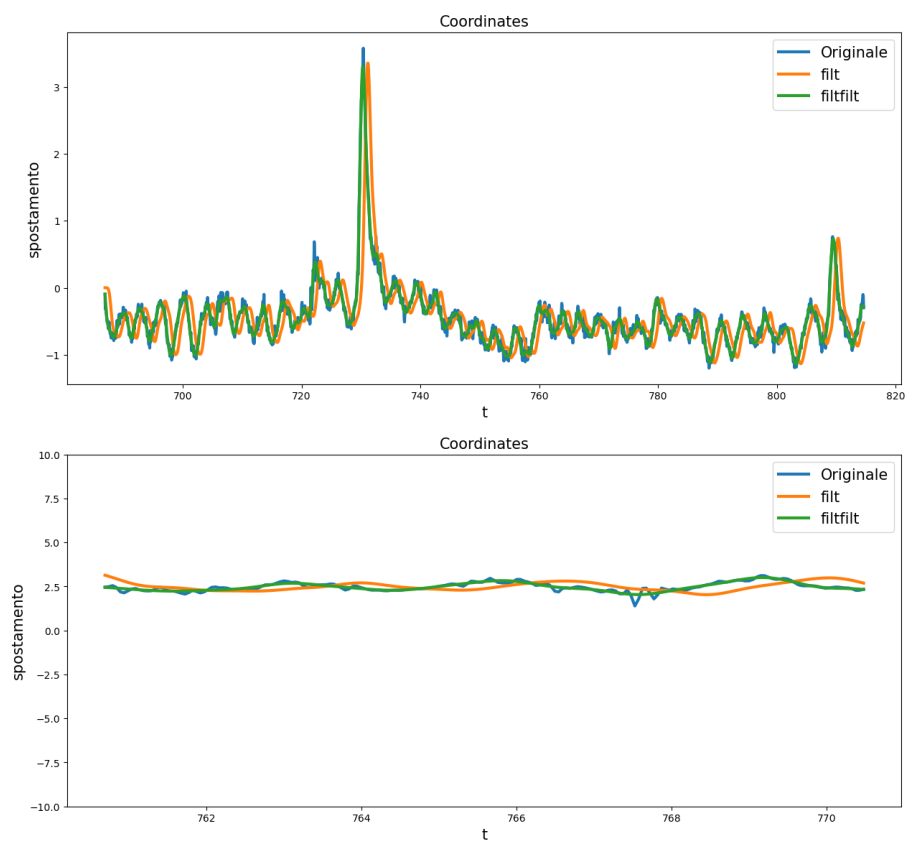


Figura 3.4: Confronto tra segnale VRT originale, VRT filtrato una volta (filt) e VRT con zero-phase filtering (filtfilt)

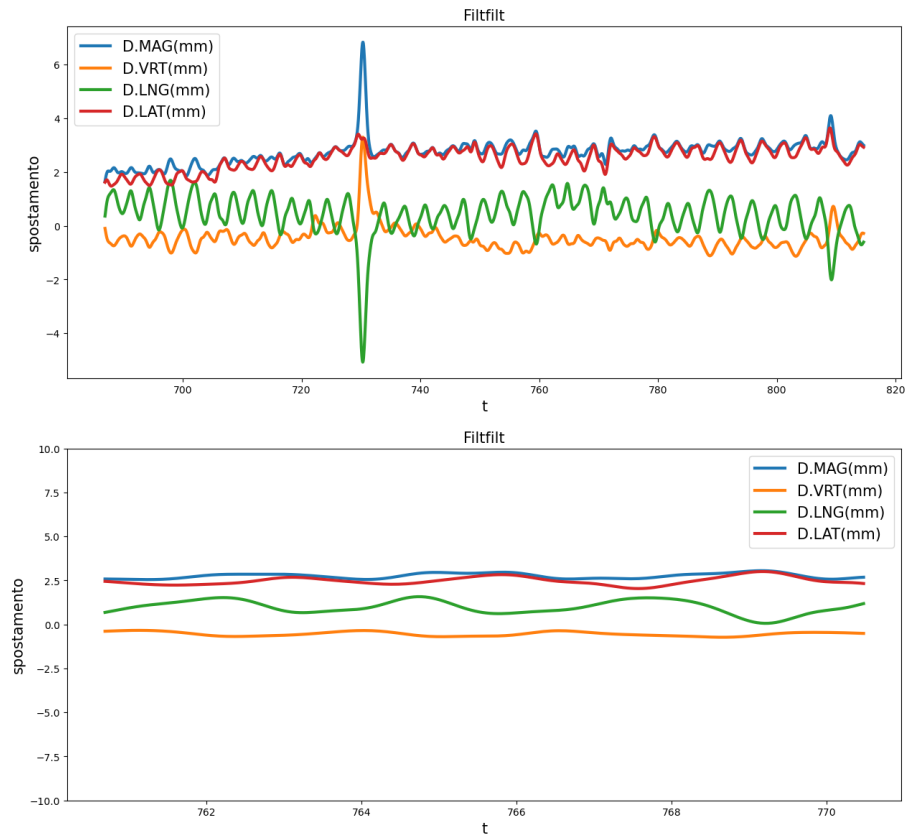


Figura 3.5: Segnali dopo il filtraggio a fase zero. Sopra sono riportati i segnali MAG, VRT, LNG e LAT durante tutto l'intervallo di *beam on*, mentre sotto è riportato uno zoom dei sei segnali su un intervallo di 10 secondi.

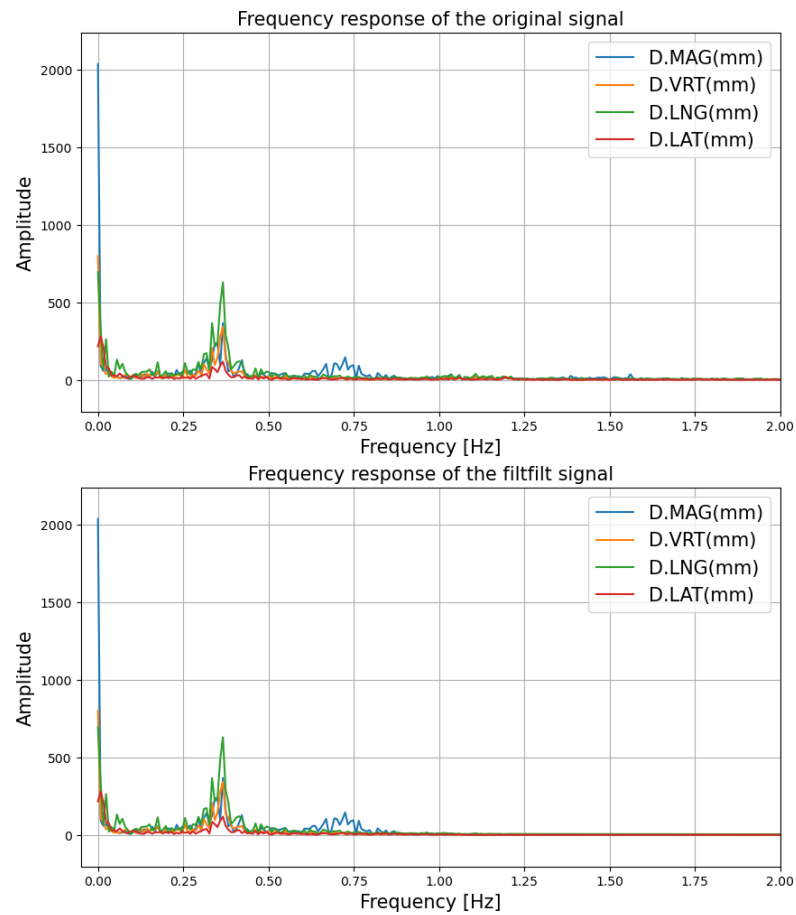


Figura 3.6: Confronto tra la DFT dei segnali originali (sopra) e la DFT dei segnali filtrati con funzione `sosfiltfilt` (sotto). Tramite il filtraggio, sono state rimosse alcune ondulazioni presenti nella DFT del segnale originale in corrispondenza di frequenze maggiori di 1 Hz.

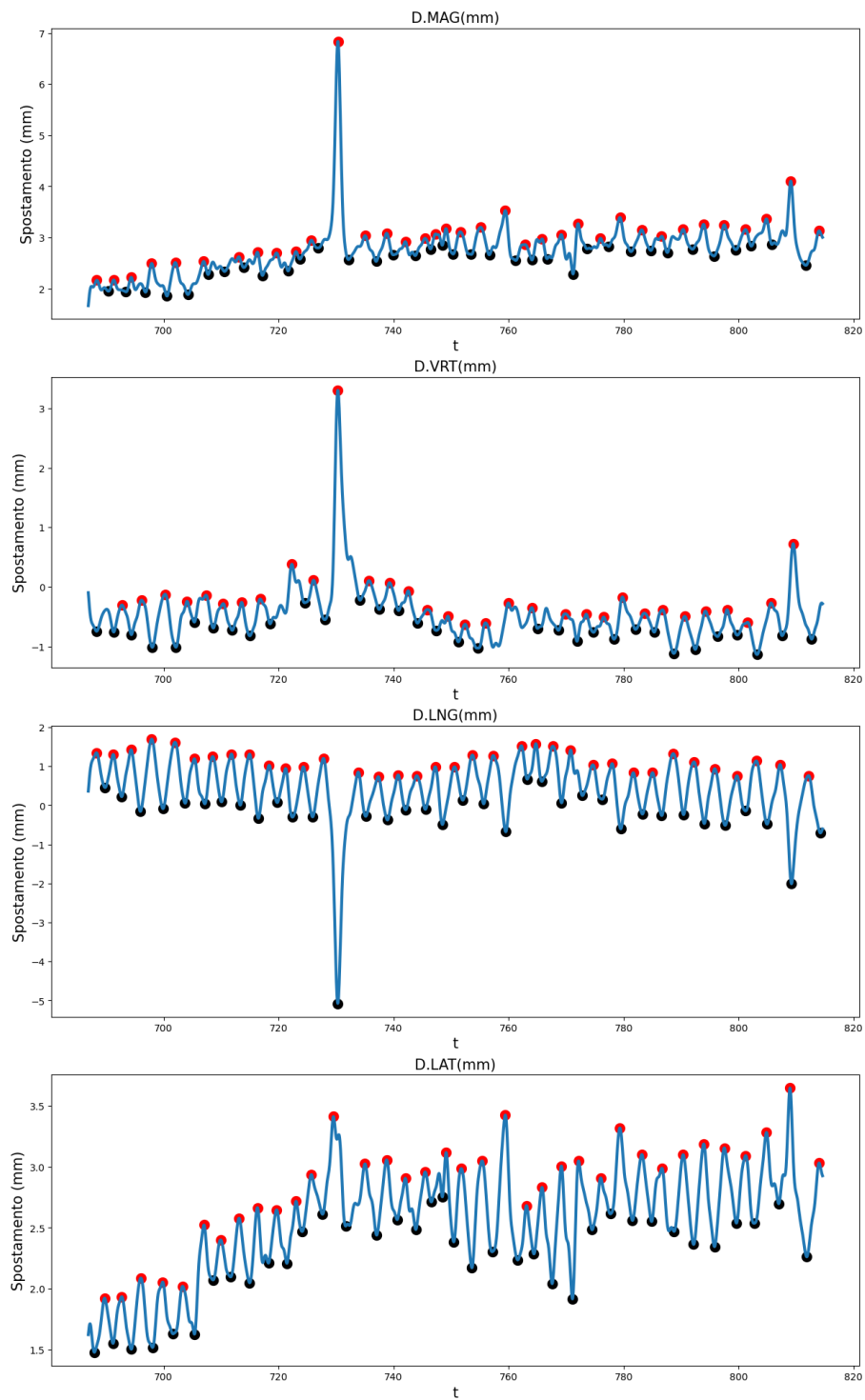


Figura 3.7: Esempio di individuazione dei picchi dei segnali con funzione *find\_peaks*

	△ Peak type	# Elapsed Time(sec)	# D.MAG(mm)
0	R	688.361	2.1683277413119897
1	R	691.291	2.1681977453131087
2	R	694.369	2.2207523771433393
3	R	695.978	2.094772894685768
4	R	697.922	2.493496277106618
5	R	702.045	2.510257414765551
6	R	706.963	2.531363379019287
7	R	708.593	2.5179244387474897
8	R	713.043	2.623187166381962

Figura 3.8: Esempio di DataFrame per i picchi del segnale MAG

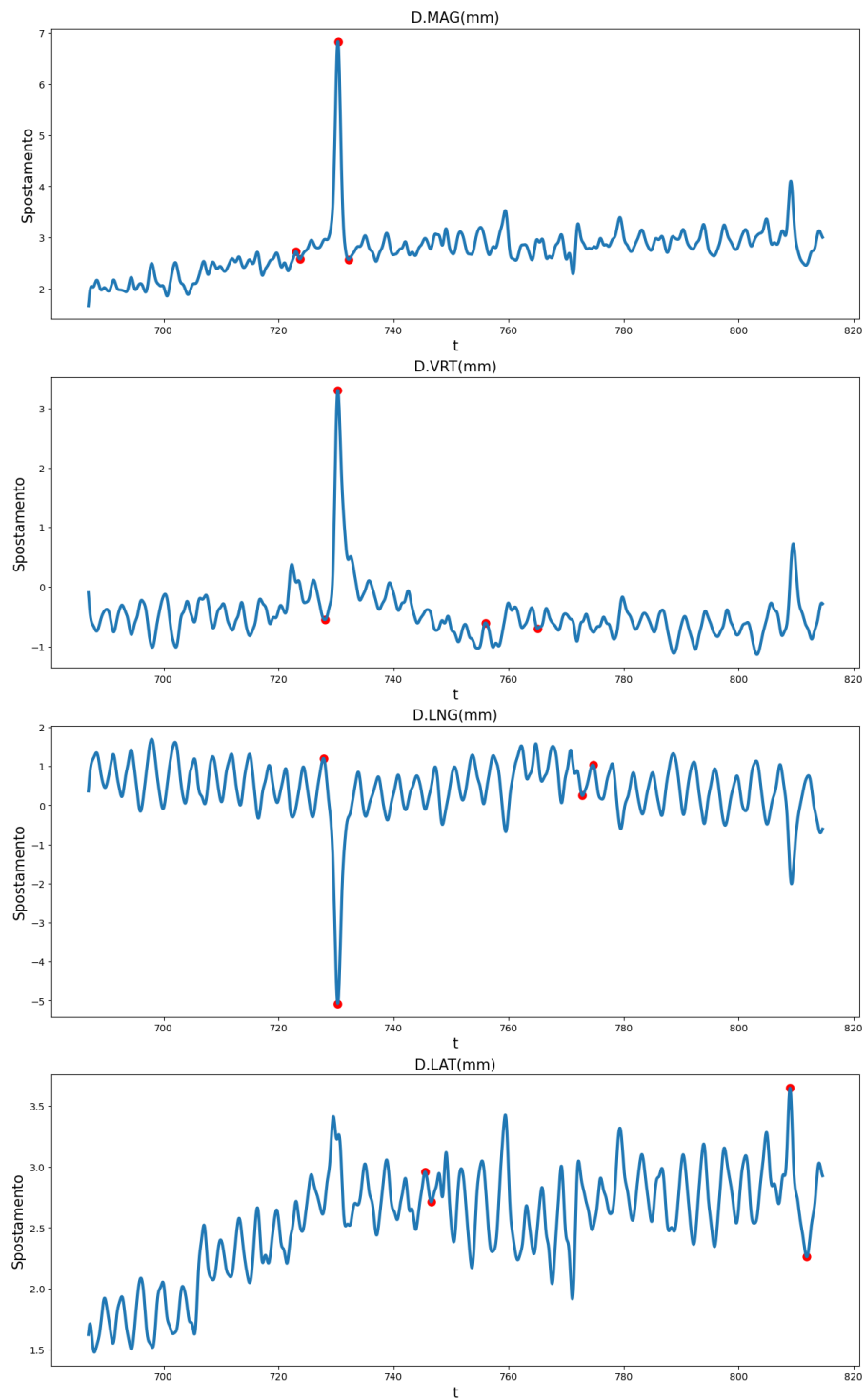


Figura 3.9: Coppie di picchi eliminate dopo il restringimento tra il 1° e 99° percentile. Vengono eliminate solo le ampiezze più grandi e più piccole.

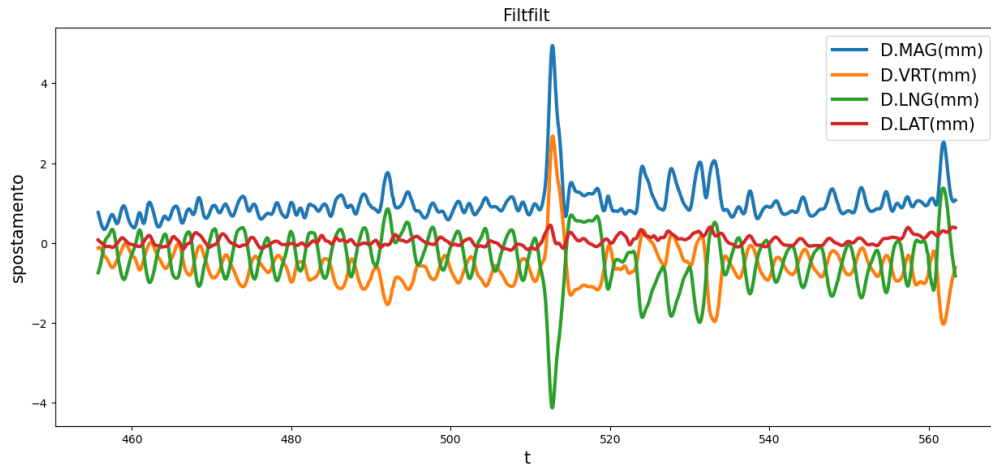


Figura 3.10: Frazione esempio della prima paziente con  $ER\% = 4.2\%$ . Andamento dei segnali filtrati nell'intervallo di beam on.

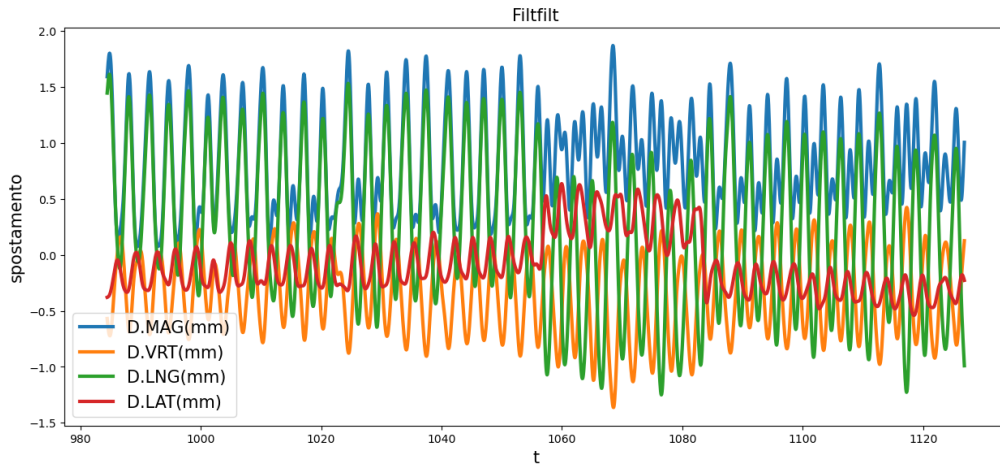


Figura 3.11: Frazione esempio della seconda paziente con  $ER\% = 16.3\%$ . Andamento dei segnali filtrati nell'intervallo di beam on.

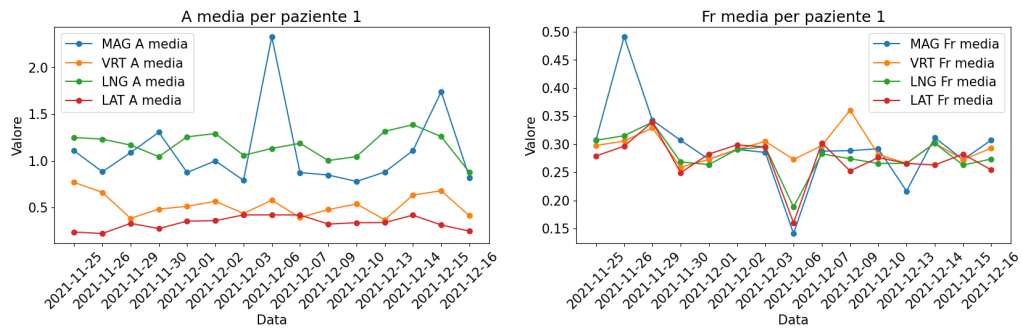


Figura 3.12: Andamento delle ampiezze (sinistra) e frequenze medie (destra) della prima paziente sulle varie frazioni

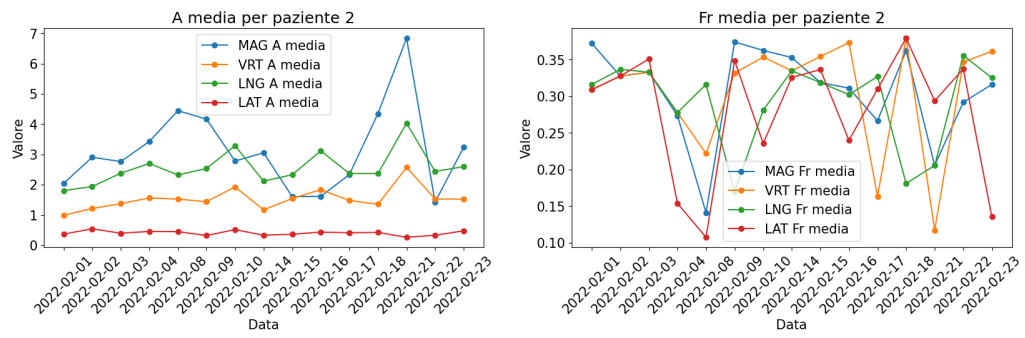


Figura 3.13: Andamento delle ampiezze (sinistra) e frequenze medie (destra) della seconda paziente sulle varie frazioni

## 3.2 Classificazione dei pazienti

L'ER% relativo al segnale MAG ha mediana 8.5% e IQR (5.3-12.9)%, di conseguenza, con **cutoff del 5%**, il 23% dei pazienti risulta avere un ciclo respiratorio "ripetibile" e i restanti "non ripetibile". L'ER% ha media  $(3.2 \pm 1.1)\%$  per la classe "ripetibile", e  $(12.8 \pm 10.4)\%$  per la classe "non ripetibile".

Con cutoff del 10%, invece, i pazienti con ciclo respiratorio "ripetibile" diventano la maggioranza (59%). Con questo cutoff, l'ER% ha media  $(5.6 \pm 2.3)\%$  per la classe "ripetibile", e  $(17.7 \pm 12.2)\%$  per la classe "non ripetibile".

### Distribuzioni delle variabili in base alla ripetibilità

Consideriamo per le analisi seguenti il cutoff del 5% descritto sopra.

Le ampiezze medie relative al segnale MAG risultano avere media  $(1.5 \pm 0.7)mm$  per la classe "ripetibile" e  $(2.7 \pm 1.5)mm$  per la classe "non ripetibile". Le frequenze medie, invece sono molto simili:  $(0.26 \pm 0.07)Hz$  e  $(0.27 \pm 0.08)Hz$  per i "ripetibili" e "non ripetibili", rispettivamente.

La mediana (IQR) del volume del tumore (CTV) per le classi "ripetibile" e "non ripetibile" è, rispettivamente, 763.0(632.3 – 979.5)cc e 614.6(429.7 – 859.6)cc. L'età media delle classi "ripetibile" e "non ripetibile" è, rispettivamente,  $65 \pm 9$  e  $63 \pm 10$  anni. Dalle figure 3.14 e 3.15, in cui sono riportati gli istogrammi e boxplot di queste distribuzioni, possiamo vedere un leggero distacco tra le medie delle distribuzioni "ripetibile" e "non ripetibile", sia per età che per volume del tumore, ma c'è maggiore distinzione tra le distribuzioni delle due classi per valori di volume del tumore.

Il 55% e 49% delle pazienti "ripetibili" e "non ripetibili", rispettivamente, è stato trattato al lato sinistro (Figura 3.16), il che suggerisce che non ci sia correlazione tra la posizione del tumore e la ripetibilità del ciclo respiratorio.

Andando a suddividere le pazienti in 3 fasce di età (18-55, 56-65 e 66+), possiamo osservare che le fasce con la percentuale più alta di pazienti "ripetibili" sono le ultime due, con il 25% di pazienti "ripetibili" (Figura 3.17).

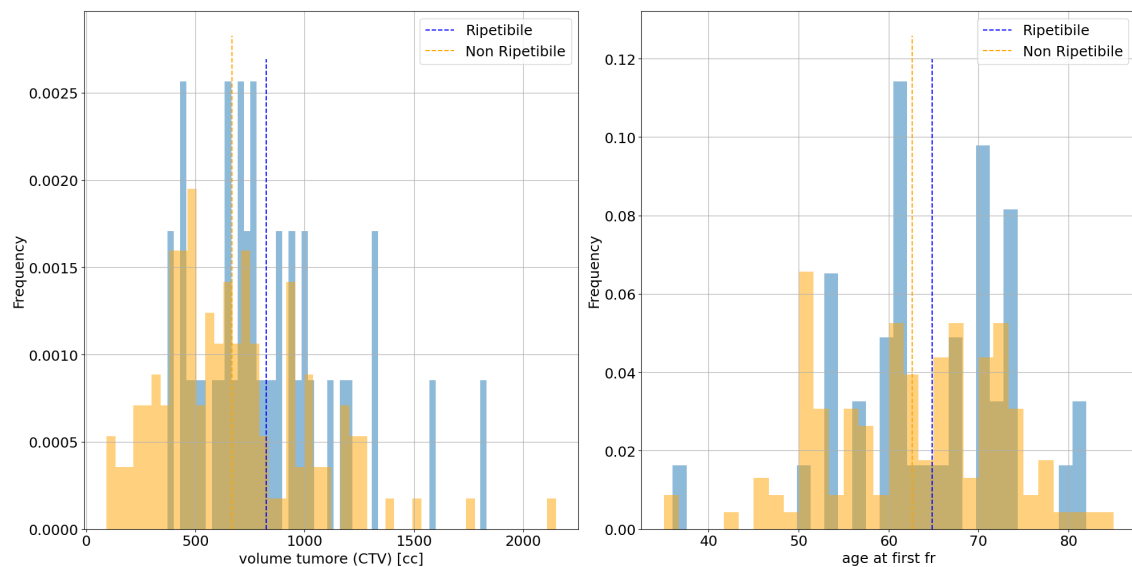


Figura 3.14: Istogrammi delle distribuzioni di volume del tumore (sinistra) ed età (destra) in base alla ripetibilità. Le linee tratteggiate rappresentano le medie.

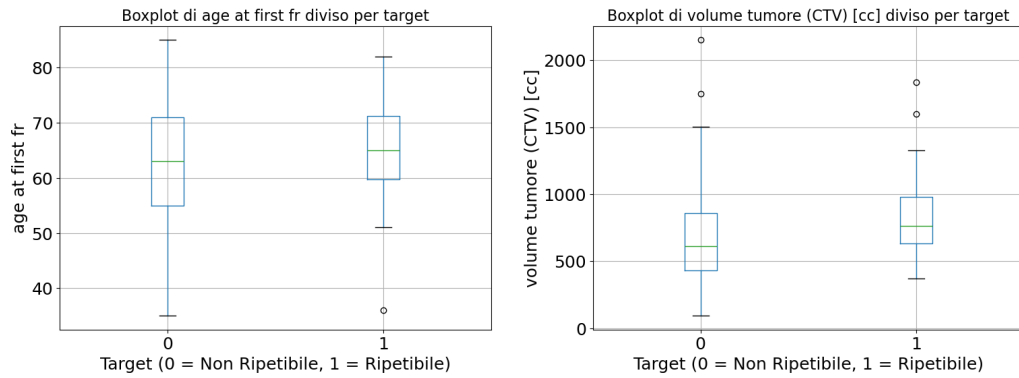


Figura 3.15: Box plot delle distribuzioni di età (sinistra) e volume del tumore(destra) in base alla ripetibilità

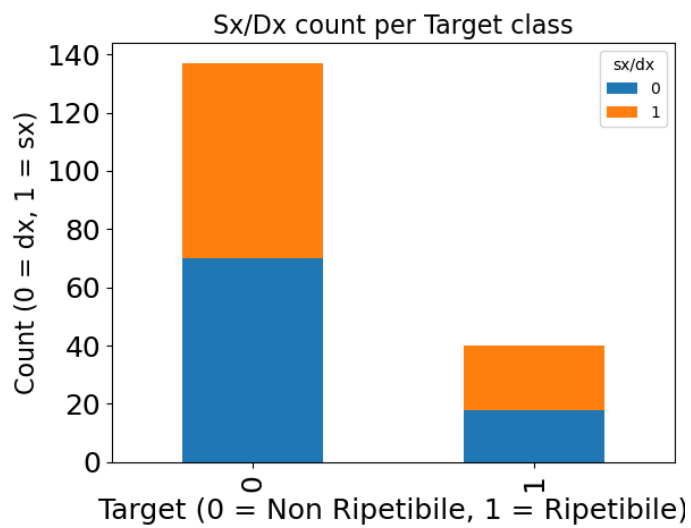


Figura 3.16: Bar plot delle distribuzioni della posizione del tumore in base alla ripetibilità

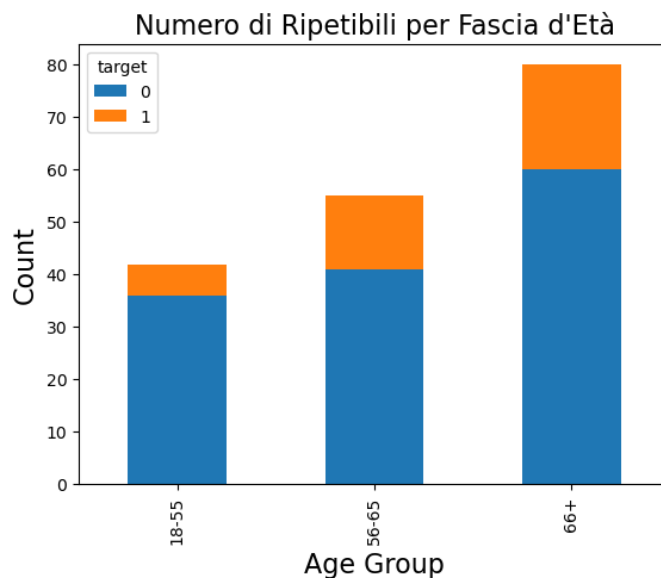


Figura 3.17: Bar plot del numero di pazienti "ripetibili" e "non ripetibili" presenti suddivisi in 3 fasce di età

### Distribuzione dell'ER% per fasce di età e volume

Andando a suddividere le pazienti in 5 fasce di età di età, l'ER% è distribuito come segue:

	18-45	46-55	56-65	66-75	76+
ER%	$8.3 \pm 3.1$	11.8(7.1 – 16.4)	8.4(5.1 – 11.9)	$8.8 \pm 6.0$	$11.9 \pm 8.4$

Tabella 3.2: Statistiche (media $\pm$  dev.std o mediana(IQR)) dell'ER% relativo al segnale MAG per le varie fasce di età

Andandole, invece, a dividere in fasce di volume del tumore, l'ER% è distribuito come segue:

	0-400	401-800	801-1200	1200+
ER%	$13.9 \pm 7.7$	7.5(5.2 – 11.3)	$9.4 \pm 6.5$	$9.2 \pm 6, 7$

Tabella 3.3: Statistiche (media $\pm$  dev.std o mediana(IQR)) dell'ER% relativo al segnale MAG per le varie classi di volume del tumore

Nelle Figure 3.18 e 3.19 sono riportati i box plot dell'ER% suddiviso per classi di età e volume, rispettivamente.

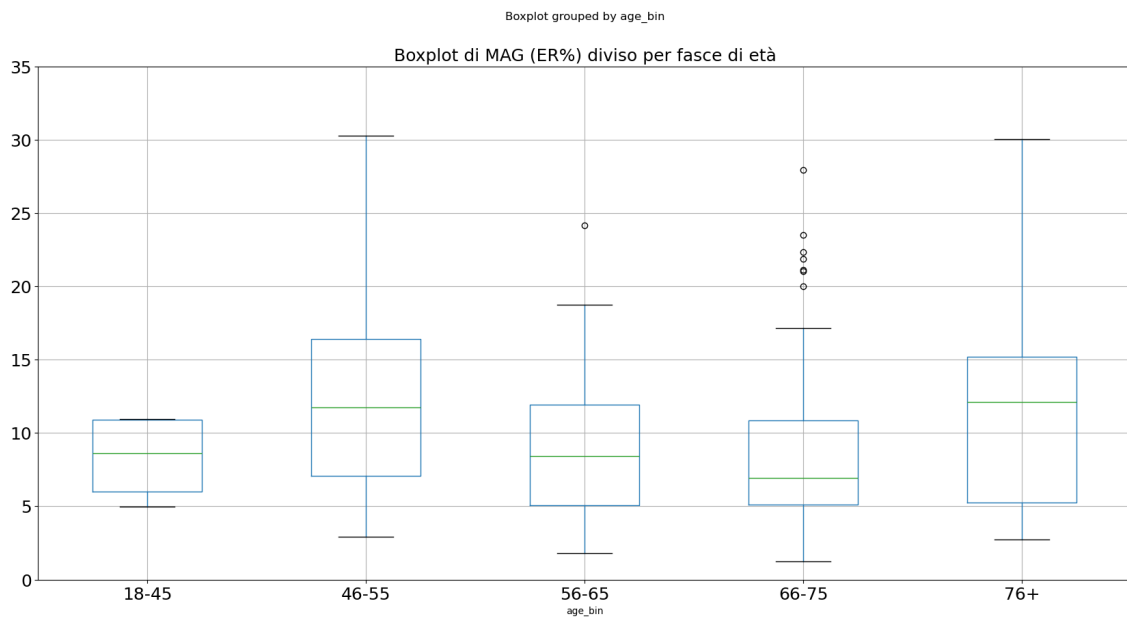


Figura 3.18: Distribuzione dell'ER% per classi di età

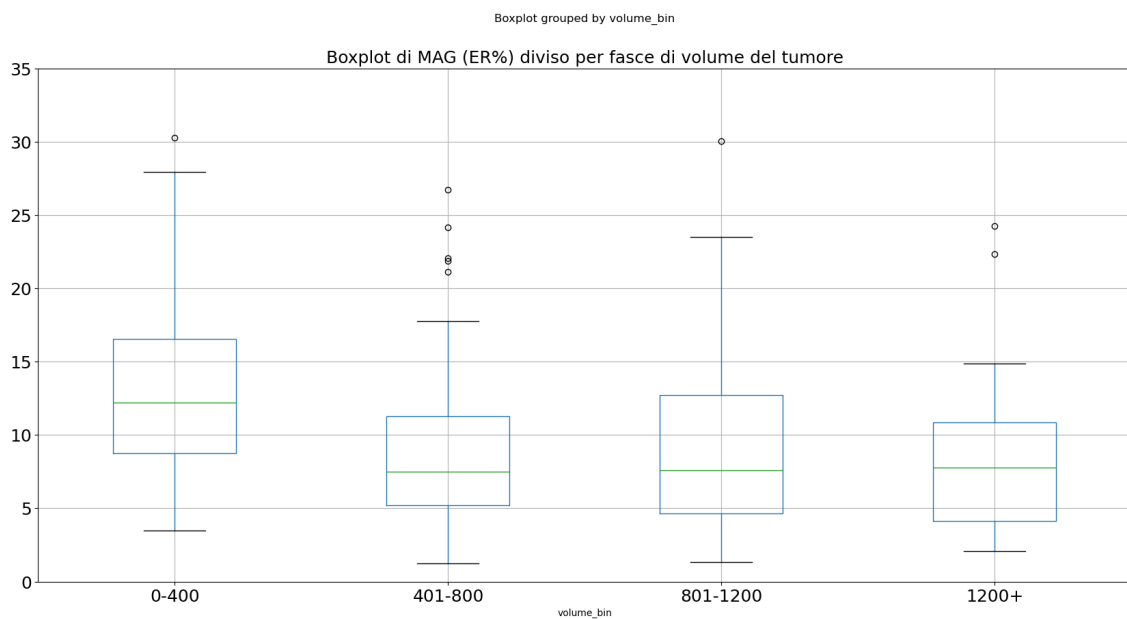


Figura 3.19: Distribuzione dell'ER% per classi di volume del tumore

Per quanto riguarda la distribuzione dell'ER% nelle varie fasce di età, possiamo osservare un leggero distacco tra le medie delle distribuzioni, ma nessuna si distacca nettamente dalle altre.

Riguardo il volume del tumore, invece, la fascia 0-400 si distingue dalle altre per maggiori valori di ER% in media.

### 3.3 Analisi statistica

Infine, secondo il test di confronto della media di Wilcoxon-Mann-Whitney, con cutoff del 5%, la ripetibilità del respiro risulta correlata significativamente solo con volume del tumore, ma non con età e posizione del tumore (sx/dx). Infatti, possiamo vedere dai boxplot nella Figura (3.15) che il gruppo di pazienti "ripetibili" tende ad avere tumori di volume maggiore rispetto ai "non ripetibili", ma non ci sono differenze significative di età o nella posiazione del tumore tra i due gruppi.

Se invece consideriamo un cutoff del 10%, la ripetibilità del respiro risulta correlata significativamente sia con volume del tumore che con età delle pazienti, in quanto il distacco tra le distribuzioni aumenta, e le pazienti "ripetibili" tendono ad avere volume del tumore ed età maggiore rispetto alle "non ripetibili". La posizione del tumore, invece, continua a non essere correlata con la ripetibilità.

Nella Tabella 3.4 sono riportate le statistiche del test di Wilcoxon per entrambi i cutoff.

		Statistiche U	p value
5%	Volume tumore	3549.0	0.0045*
	Età	3111.0	0.1934
	sx/dx	2907.0	0.5001
10%	Volume tumore	4839.0	0.0019*
	Età	4494.5	0.0374*
	sx/dx	3858.5	0.8311

Tabella 3.4: Statistiche U e p-value del test di Wilcoxon per cutoff del 5 e 10%. \* valori del p value < 0.05

# Capitolo 4

## Discussioni

In questa tesi, ci siamo posti l'obiettivo di caratterizzare il movimento respiratorio delle pazienti, partendo dai loro segnali di movimento, acquisiti con AlignRT. Le abbiamo poi classificate in base alla ripetibilità del loro movimento respiratorio e cercato eventuali correlazioni con i parametri clinici a disposizione (volume del tumore ed età). Come argomento scientifico non è ancora stato affrontato approfonditamente per via della *novelty* delle tecniche e della radioterapia adattativa, di conseguenza siamo limitati nel numero di studi da poter prendere come riferimento.

I risultati ottenuti evidenziano la fattibilità della metodologia proposta riguardo la caratterizzazione del movimento respiratorio delle pazienti sottoposte a trattamento RT alla mammella. La suddivisione delle pazienti in classi di ripetibilità, ha, infatti, evidenziato un divario significativo nelle ampiezze medie. L'ampiezza media relativa al segnale MAG è risultata essere  $(1.5 \pm 0.7)mm$  per la classe "ripetibile", in linea con i valori trovati da Richer et al. [22], mentre, la classe "non ripetibile" presenta valori di ampiezza media significativamente maggiori,  $(2.7 \pm 1.5)mm$ . Questa differenza rispetto alla letteratura, può essere attribuita al diverso sistema con cui sono stati acquisiti i segnali di movimento.

L' $ER_{MAG}\%$  risulta essere abbastanza alto, soprattutto per un cutoff del 5%, con il quale solo il 23% di pazienti risulta "ripetibile". Il motivo dei valori di ER% ottenuti per il dataset può essere attribuito ai dati retrospettivi estratti da pazienti che hanno eseguito il trattamento in regime di respiro libero e senza istruzioni su come gestire il movimento respiratorio. Questi risultati evidenziano quindi l'importanza di un protocollo di gestione del movimento che potrebbe portare beneficio al trattamento e alla prognosi del paziente.

Il valore del cutoff può essere regolato in base al livello di ripetibilità del movimento respiratorio richiesto. È sufficiente passare a una soglia del 10%, valore ancora ragionevolmente basso, per ottenere un 59% di pazienti ripetibili. I cutoff del 5% e 10% sono stati scelti in quanto ragionevolmente bassi, ma per sceglierlo in modo più rigoroso andrebbe testato su dati clinici prospettici, che noi non avevamo a disposizione, o su pazienti a cui sono state fornite istruzioni specifiche per il controllo del respiro.

Dalla suddivisione dei pazienti in base alla ripetibilità del movimento respiratorio, emergono deboli correlazioni con i parametri clinici, in particolare con il volume del tumore, che restano basse anche al variare del cutoff. Inoltre, le pazienti "ripetibili" risultano avere un volume del tumore ed età leggermente maggiori. Una limitazione di questo studio è la scarsità di dati clinici completi a disposizione.

In questo studio, abbiamo considerato solamente le pazienti sottoposte ad un trattamento ipofrazionato in 13/15 frazioni. Un confronto interessante che si può fare riguarda come varia la ripetibilità in un regime di trattamento minore (5 frazioni). In questo caso, ci aspettiamo valori minori di errore relativo e, di conseguenza, un maggior numero di "ripetibili", a parità di cutoff, in quanto il trattamento si svolge su un intervallo temporale minore.

# Conclusioni

Abbiamo intrapreso questo studio con l'obiettivo di caratterizzare il respiro di pazienti sottoposte a trattamento radioterapico alla mammella per l'arruolamento in trattamenti a respiro controllato, scegliendo come potenziali pazienti quelle che hanno, sulla base di dati retrospettivi, un tracciato di respirazione ripetibile in regime a respiro libero.

Siamo partiti dal monitoraggio del movimento di una ROI posizionata in corrispondenza della superficie del seno delle pazienti su tre assi, utilizzando strumenti di SGRT. Per la caratterizzazione del ciclo respiratorio è stato fatto uso della Trasformata di Fourier per ricavare informazioni sulle principali frequenze presenti nei segnali e individuare le varie componenti in frequenza e ampiezza media. Abbiamo poi usufruito dei dati ricavati da questa analisi per evidenziare pazienti con pattern respiratori significativamente differenti durante le varie sedute del trattamento, suddividendoli in due classi, "ripetibili" e "non ripetibili", in base a un certo cutoff. Infine, abbiamo cercato una correlazione tra l'appartenenza a una delle due classi sopra definite ed età e volume del tumore dei pazienti.

I risultati ottenuti evidenziano la fattibilità della metodologia proposta riguardo la caratterizzazione del movimento respiratorio delle pazienti sottoposte a trattamento RT della mammella. Infatti, le pazienti che risultano "ripetibili" (con cutoff del 5%) riescono a mantenere pattern respiratori sufficientemente costanti durante le varie sedute, con valori di ampiezza media del segnale MAG a bassa variabilità e in accordo con la letteratura. Queste pazienti saranno, di conseguenza, potenziali candidate per trattamenti a respiro controllato.

Il motivo dei valori di ER% ottenuti per il DB può essere imputata ai dati retrospettivi estratti da pazienti che hanno eseguito il trattamento in regime di respiro libero e senza istruzioni su come gestire il movimento respiratorio. Questi risultati evidenziano quindi l'importanza di un protocollo di gestione del movimento che potrebbe portare beneficio al trattamento e alla prognosi del paziente.

# Appendice A

## Materiale di background

### A.1 Spazi $L^p$

**A.1.1 Definizione (Norme  $p$ ):** Data una funzione  $f : X \rightarrow \mathbb{C}$ , la **norma  $p$**  di  $f$  è definita com segue:

- Se  $1 \leq p < \infty$

$$\|f\|_{L^p(X)} \equiv \|f\|_p = \left( \int_X |f(x)|^p dx \right)^{\frac{1}{p}}$$

- Se  $p = \infty$

$$\|f\|_{L^\infty(X)} \equiv \|f\|_\infty = \sup_{x \in X} |f(x)|$$

Uno spazio  $L^p$ , con  $1 \leq p \leq \infty$ , è definito come segue:

$$L^p(X) = \{f : X \rightarrow \mathbb{C} : \|f\|_p < \infty\}$$

In generale, gli spazi  $L^p$  sono spazi di Banach (spazi vettoriali normati completi). Inoltre  $L^2(X)$  è uno spazio di Hilbert dotato del prodotto scalare

$$\langle f, g \rangle = \int_X f(x)g^*(x)dx$$

dove  $g^*(x)$  è il complesso coniugato di  $g(x)$ .

**A.1.1 Teorema (Inclusione degli spazi  $L^p$ ):** Se  $|X| < \infty$ , allora

$$L^q(X) \subset L^p, \quad \forall 1 \leq p < q \leq \infty$$

Osserviamo che questo non vale se  $X = \mathbb{R}^n$ .

## A.2 Convergenza

**A.2.1 Teorema (Teorema della Convergenza Dominata di Lebesgue):** Sia  $X$  uno spazio misurabile e  $(f_n)_{n \in \mathbb{N}}$  una famiglia di funzioni  $f_n : X \rightarrow \mathbb{C}$ ,  $\forall n \in \mathbb{N}$  tali che  $f_n$  converge quasi ovunque a  $f$  in  $X$ .  
Se  $|f_n(x)| \leq g(x)$  q.o.  $\forall n \in \mathbb{N}$  e  $\int_X g(x) dx < \infty$ , allora  $f$  è **integrabile** e

$$\int_X \lim_{n \rightarrow \infty} f_n(x) dx = \lim_{n \rightarrow \infty} \int_X f_n(x) dx$$

**A.2.2 Teorema (Teorema di Fubini):** Se

$$\int_{-\infty}^{+\infty} \left( \int_{-\infty}^{+\infty} |f(x, y)| dx \right) dy < \infty \quad (f \in L^1(\mathbb{R} \times \mathbb{R}))$$

allora

$$\begin{aligned} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y) dx dy &= \int_{-\infty}^{+\infty} \left( \int_{-\infty}^{+\infty} f(x, y) dx \right) dy \\ &= \int_{-\infty}^{+\infty} \left( \int_{-\infty}^{+\infty} f(x, y) dy \right) dx \end{aligned}$$

# Appendice B

## Codici Python

In questa Appendice sono riportati i codici Python eseguiti per ottenere i risultati riportati nel Capitolo 3.

### B.1 Elaborazione dei segnali e creazione di un unico DataFrame

```
1 import numpy as np
2 import pandas as pd
3 import matplotlib.pyplot as plt
4
5 from scipy import signal
6 from scipy.signal import butter, freqz
7 from scipy.signal import hilbert, find_peaks
8 from scipy.fft import fft, fftfreq
9 from scipy.interpolate import interp1d
10 from scipy.signal import resample
11
12 import os
13 import shutil
14 import random
15 import glob
16
17 from collections import Counter
18
19 from os.path import isfile, join
20 from sklearn.preprocessing import LabelEncoder
21 # import cv2
22 from datetime import datetime
23
24 import warnings
25 warnings.filterwarnings('ignore')
26
27 seed_value= 1
28 os.environ['PYTHONHASHSEED']=str(seed_value)
29 random.seed(seed_value)
30 np.random.seed(seed_value)
31
32 from IPython.display import display, HTML
33 display(HTML("<style>.container { width:85% !important; }</style>"))
    )
```

```

34
35 #Parametri filtro
36 nb = 8 # ordine del filtro
37 Wn = 1 # frequenze di taglio (Hz)
38 filter_type = 'lowpass' # tipo di filtro
39
40 col_to_plot = [
41     'D.MAG(mm)',
42     'D.VRT(mm)',
43     'D.LNG(mm)',
44     'D.LAT(mm)'
45 ]
46 new_col = ['num_fr', 'PtID', 'date', 'fs',
47     'D.MAG - A media(mm)',
48     'D.MAG - A Range(mm)',
49     'D.MAG - A Mediana(mm)',
50     'D.MAG - A Std(mm)',
51     'D.MAG - A 10° percentile(mm)',
52     'D.MAG - A 25° percentile(mm)',
53     'D.MAG - A 75° percentile(mm)',
54     'D.MAG - A 90° percentile(mm)',
55     'D.MAG - Fr media(Hz)',
56     'D.VRT - A media(mm)',
57     'D.VRT - A Range(mm)',
58     'D.VRT - A Mediana(mm)',
59     'D.VRT - A Std(mm)',
60     'D.VRT - A 10° percentile(mm)',
61     'D.VRT - A 25° percentile(mm)',
62     'D.VRT - A 75° percentile(mm)',
63     'D.VRT - A 90° percentile(mm)',
64     'D.VRT - Fr media(Hz)',
65     'D.LNG - A media(mm)',
66     'D.LNG - A Range(mm)',
67     'D.LNG - A Mediana(mm)',
68     'D.LNG - A Std(mm)',
69     'D.LNG - A 10° percentile(mm)',
70     'D.LNG - A 25° percentile(mm)',
71     'D.LNG - A 75° percentile(mm)',
72     'D.LNG - A 90° percentile(mm)',
73     'D.LNG - Fr media(Hz)',
74     'D.LAT - A media(mm)',
75     'D.LAT - A Range(mm)',
76     'D.LAT - A Mediana(mm)',
77     'D.LAT - A Std(mm)',
78     'D.LAT - A 10° percentile(mm)',
79     'D.LAT - A 25° percentile(mm)',
80     'D.LAT - A 75° percentile(mm)',
81     'D.LAT - A 90° percentile(mm)',
82     'D.LAT - Fr media(Hz)',
83 ]

1 # Lettura dei file CSV e creazione del dataframe
2 data_type = ['5', '15']
3 df = pd.DataFrame(columns=new_col)
4 for type in data_type:
5     for (subdir, folder, files) in os.walk(f'D:/Tirocinio/{type}
6         fr_toelab'):

```

```

7
8     for file in files:
9         if file.endswith('.csv'):
10            date = file.split('_')[1]
11            PtID = file.split('_')[0]
12            date = date.split('.')[0]
13            # Rimozione di caratteri extra dopo la data
14            date = date.split('(')[0]
15            # Tieni solo la parte prima della '('
16            date = datetime.strptime(date, '%y%m%d').date()
17
18            df_plt = pd.read_csv(os.path.join(subdir, file),
19                                sep=';', decimal='.', encoding='latin1')
20            df_plt.drop(['Unnamed: 0'], axis=1, inplace=True)
21
22            # Restrizione all'intervallo di BeamON
23            #-----BeamInd-----
24            beam_state_index = df_plt['ReportedBeamState'][df_plt
25 ['ReportedBeamState'].diff() != 0].index.tolist()
26            start_first_beam = beam_state_index[1]
27            end_last_beam = beam_state_index[-1]
28            #-----INTERVAL BEAMON-----
29            df_intB0 = df_plt.loc[start_first_beam:end_last_beam]
30            df_intB0.reset_index(inplace=True)
31            df_intB0.drop(['index'], axis=1, inplace=True)
32
33            N = df_intB0.shape[0] # numero di campioni
34            deltax = df_intB0['Elapsed Time(sec)'][1]-df_intB0['
35 Elapsed Time(sec)'][0] # intervallo di campionamento (s)
36            fs = 1 / deltax # frequenza di campionamento (Hz)
37
38            # Interpolazione del segnale per avere un
39            # campionamento uniforme di 20 Hz
40            new_fs = 20.0 # nuova frequenza di campionamento (Hz)
41
42            if fs < 20.0:
43                new_N = int(N * (new_fs / fs)) # nuovo numero di
44                # campioni (N : new_N = fs : new_fs)
45                new_time = np.linspace(df_intB0['Elapsed Time(sec)']
46                ].iloc[0], df_intB0['Elapsed Time(sec)'].iloc[-1], new_N)
47                interpolated_signals = {}
48                for col in col_to_plot:
49                    f = interp1d(df_intB0['Elapsed Time(sec)'],
50                                df_intB0[col], kind='quadratic')
51                    interpolated_signals[col] = f(new_time)
52
53                df_intB0_new = pd.DataFrame(interpolated_signals)
54                df_intB0_new['Elapsed Time(sec)'] = new_time
55                df_intB0_new = df_intB0_new[['Elapsed Time(sec)'] +
56                col_to_plot]
57                N = new_N
58
59            else :
60                if fs > 20.0:
61                    # ricampionamento del segnale a 20 Hz
62                    from scipy.signal import resample
63                    new_N = int(N * (new_fs / fs)) # nuovo numero

```

```

57     di campioni (N : new_N = fs : new_fs)
58         df_intB0_new = df_intB0.copy()
59         df_intB0_new['Elapsed Time(sec)'] = pd.
to_timedelta(df_intB0_new['Elapsed Time(sec)'], unit='s')
60         df_intB0_new.resample(f'{new_N}s', on='Elapsed
Time(sec)')
61         # Converti 'Elapsed Time(sec)' in secondi (
float)
62         sec = df_intB0_new['Elapsed Time(sec)'].dt.
total_seconds()
63         df_intB0_new['Elapsed Time(sec)'] = sec
64         N = new_N
65         # Filtraggio con filtro Butterworth
66         sos = butter(nb, Wn, fs=new_fs, btype= filter_type,
output='sos')
67         b, a = butter(nb, Wn, fs=new_fs, btype= filter_type,
output='ba')
68         w ,h = freqz(b, a, fs=new_fs, worN=1000)
69
70         # Zero phase filtering
71         df_filtfilt = df_intB0_new.copy()
72         for i, col in enumerate(col_to_plot[:4]):
73             df_filtfilt[col] = signal.sosfiltfilt(sos,
df_intB0_new[col])
74
75         # Calcolo della FFT del segnale filtrato e media
pesata delle frequenze
76         df_filtfilt_FT = np.zeros((df_filtfilt.shape[0], 5),
dtype=complex)
77         weighted_mean = np.zeros((5,1), dtype=float)
78         for i, col in enumerate(col_to_plot):
79             y= df_filtfilt[col]
80             freq = fftfreq(len(y), deltax)
81             df_filtfilt_FT[:,i]= fft(y)
82             # Calcolo la media pesata delle frequenze
83             weighted_mean[i] = np.sum(freq[:N//2] * abs(
df_filtfilt_FT[:N//2,i])) / np.sum(abs(df_filtfilt_FT[:N//2,i]))
84
85         # Calcolo il periodo corrispondente alla media pesata
delle frequenze
86         periods = 1 / weighted_mean
87
88         # Identificazione dei picchi
89         peaks_pos_list = []
90         peaks_neg_list = []
91         comp_sec = new_fs # numero di campioni in 1 secondo (
frequenza di campionamento)
92         for i, col in enumerate(col_to_plot):
93             dist = comp_sec *(2/3*periods[i]) # distanza minima
tra picchi in numero di campioni
94             prominence = 0.1 # distacco minimo dei picchi
dalla baseline
95             peaks_pos, _ = find_peaks(df_filtfilt[col],
distance = dist, prominence = prominence) # 1 sec = 20 samples
96             peaks_neg, _ = find_peaks(-df_filtfilt[col],
distance = dist, prominence = prominence)
97             peaks_pos_list.append(peaks_pos)

```

```

98         peaks_neg_list.append(peaks_neg)
99
100         # Creiamo dei dataframe per i picchi
101         df_picchi = {'D.MAG(mm)': pd.DataFrame() ,
102                     'D.VRT(mm)' : pd.DataFrame(),
103                     'D.LNG(mm)':pd.DataFrame(),
104                     'D.LAT(mm)':pd.DataFrame()}
105         for i, col in enumerate(col_to_plot):
106             peak_type_pos = pd.Series(['R']*len(
107 peaks_pos_list[i]), name='Peak type')
108             peak_type_neg = pd.Series(['B']*len(
109 peaks_neg_list[i]), name='Peak type')
110             peak_type = pd.DataFrame(pd.concat([peak_type_pos
111 , peak_type_neg], ignore_index=True, axis=0))
112             peak_list = np.concatenate((peaks_pos_list[i],
113 peaks_neg_list[i]), axis=0)
114             time = df_filtfilt['Elapsed Time(sec)'][peak_list
115 ].reset_index(drop=True)
116             coord = df_filtfilt[col][peak_list].reset_index(
117 drop=True)
118             df_picchi[col] = peak_type.join([time, coord])
119
120         # Ordiniamo i picchi in base al tempo e rimuoviamo i
121 picchi successivi dello stesso tipo
122
123         for col in col_to_plot:
124             df_picchi[col].sort_values(by=['Elapsed Time(sec)
125 '], ascending=True, inplace=True)
126             df_picchi[col].reset_index(drop=True, inplace=
127 True)
128             drop_indices = []
129             for i in range(df_picchi[col].shape[0] - 1):
130                 if df_picchi[col]['Peak type'][i] ==
131 df_picchi[col]['Peak type'][i+1]:
132                     drop_indices.append(i+1)
133             df_picchi[col].drop(drop_indices, inplace=True)
134             df_picchi[col].reset_index(drop=True, inplace=
135 True)
136
137         df_amplitude = {}
138         for i,col in enumerate(col_to_plot):
139             if col in ['D.MAG(mm)']:
140                 df_amplitude[col] = 2*np.abs(df_picchi[col][col
141 ].diff())
142             else:
143                 df_amplitude[col] = np.abs(df_picchi[col][col].
144 diff())
145             df_amplitude[col] = df_amplitude[col].dropna()
146             df_amplitude[col].reset_index(drop=True, inplace=
147 True)
148
149         # Restringimento tra 1° e 99° percentile
150
151         amplitude = [df_amplitude[col] for col in col_to_plot
152 ]
153         percentile_1 = [np.percentile(amplitude[i],1) for i
154 in range(len(amplitude))]
155         percentile_99 = [np.percentile(amplitude[i],99) for i

```

```

140     in range(len(amplitude))]
141
142         ind_outliers = [(amplitude[i] < percentile_1[i]) | (
143 amplitude[i] > percentile_99[i]) for i in range(len(amplitude))]
144         amplitude_outliers = [amplitude[i][ind_outliers[i]]
145 for i in range(len(amplitude))]
146
147         df_amplitude_no = {}
148         for i, col in enumerate(col_to_plot):
149             df_amplitude_no[col] = df_amplitude[col][~
150 ind_outliers[i]]
151             df_amplitude_no[col].reset_index(drop=True,
152 inplace=True)
153
154         # Calcolo delle statistiche
155         amplitude_mean = np.zeros((5,1))
156         amplitude_mean_no = np.zeros((5,1))
157         amplitude_median = np.zeros((5,1))
158         amplitude_std = np.zeros((5,1))
159         amplitude_range = np.zeros((5,1))
160         a_percentile_10 = np.zeros((5,1))
161         a_percentile_90 = np.zeros((5,1))
162         a_percentile_25 = np.zeros((5,1))
163         a_percentile_75 = np.zeros((5,1))
164         freq_mean = np.zeros((5,1))
165         ampiezza_intB0_new = df_intB0_new['Elapsed Time(sec)']
166 ].iloc[-1] - df_intB0_new['Elapsed Time(sec)'].iloc[0]
167         for i, col in enumerate(col_to_plot):
168             amplitude_mean[i] = df_amplitude[col].mean()
169             amplitude_mean_no[i] = df_amplitude_no[col].mean
170 ()
171             amplitude_median[i] = df_amplitude_no[col].median
172 ()
173             amplitude_std[i] = df_amplitude_no[col].std()
174             amplitude_range[i] = df_amplitude_no[col].max() -
175 df_amplitude_no[col].min()
176             a_percentile_10[i] = np.percentile(
177 df_amplitude_no[col], 10)
178             a_percentile_90[i] = np.percentile(
179 df_amplitude_no[col], 90)
180             a_percentile_25[i] = np.percentile(
181 df_amplitude_no[col], 25)
182             a_percentile_75[i] = np.percentile(
183 df_amplitude_no[col], 75)
184             peak_R = df_picchi[col][df_picchi[col]['Peak type
185 ']] == 'R']
186             freq_mean[i] = peak_R.shape[0] /
187 ampiezza_intB0_new
188
189         # Creazione del DataFrame riga per riga
190         df_riga = pd.DataFrame([[type, PtID, date, fs,
191                                 amplitude_mean_no[0],
192                                 amplitude_range[0],
193                                 amplitude_median[0],
194                                 amplitude_std[0],
195                                 a_percentile_10[0],
196                                 a_percentile_25[0],

```

```

181     a_percentile_75 [0] ,
182     a_percentile_90 [0] ,
183     freq_mean [0] ,
184     amplitude_mean_no [1] ,
185     amplitude_range [1] ,
186     amplitude_median [1] ,
187     amplitude_std [1] ,
188     a_percentile_10 [1] ,
189     a_percentile_25 [1] ,
190     a_percentile_75 [1] ,
191     a_percentile_90 [1] ,
192     freq_mean [1] ,
193     amplitude_mean_no [2] ,
194     amplitude_range [2] ,
195     amplitude_median [2] ,
196     amplitude_std [2] ,
197     a_percentile_10 [2] ,
198     a_percentile_25 [2] ,
199     a_percentile_75 [2] ,
200     a_percentile_90 [2] ,
201     freq_mean [2] ,
202     amplitude_mean_no [3] ,
203     amplitude_range [3] ,
204     amplitude_median [3] ,
205     amplitude_std [3] ,
206     a_percentile_10 [3] ,
207     a_percentile_25 [3] ,
208     a_percentile_75 [3] ,
209     a_percentile_90 [3] ,
210     freq_mean [3] ,],
211     columns=new_col)
212     # Aggiungo la riga appena creata al DataFrame
213     df = pd.concat([df, df_riga], ignore_index=True)

1 # Conto il numero di frazioni per paziente e tengo solo quelli con
   almeno 13 frazioni
2 fr_per_patient = df['PtID'].value_counts()
3 less_than_13fr = fr_per_patient[fr_per_patient < 13].index
4 df_13= df.loc[(1-df['PtID'].isin(less_than_13fr)).astype(bool)]
5 df_13 = df_13.reset_index(drop=True)
6
7 # Ci sono dei pazienti con troppe frazioni, quindi rimuovo quelli
   con una numero di frazioni diverso da 13 o 15
8 df_1315 = df_13[df_13['PtID'].isin(fr_per_patient[fr_per_patient.
   isin([13, 15]).index])]
9 df_1315 = df_1315.reset_index(drop=True)

```

## B.2 Test di normalità

```

1 # Per pazienti con un numero di frazioni n = 13-15, facciamo un
   test di normalità sulle ampiezze medie della MAG, VRT, LNG, LAT
2
3 alpha = 0.05
4
5 PtID = df_1315['PtID'].unique()
6 print(f'Number of patients : {len(df["PtID"].unique())}')

```

```

7 print(f'Number of patients with 13-15 fractions: {len(PtID)}\n')
8
9 # Test di normalit  di Shapiro-Wilk (pi  adatto per piccoli dataset
  )
10 from scipy.stats import shapiro
11 alpha = 0.05
12 normally_dist_feature_1= {}
13 df_rel_error_1 = pd.DataFrame(columns=['PtID', 'Feature', 'Relative
  Error (%)'])
14 for i in PtID:
15     df_patient = df_1315[df_1315['PtID']==i]
16     #print(f'Patient ID: {i}\n')
17     for col in ['D.MAG - A media(mm)', 'D.VRT - A media(mm)', 'D.LNG
  - A media(mm)', 'D.LAT - A media(mm)', ]:
18         data = df_patient[col]
19         stat, p = shapiro(data)
20         #print(f'Normality test for {col}: stat={stat:.4f}, p-value
  ={p:.4f}')
21         if p > alpha:
22             #print(f'Fail to reject the null hypothesis: {col} is
  normally distributed (p > {alpha})\n')
23             if i not in normally_dist_feature_1.keys():
24                 normally_dist_feature_1[i] = [col]
25             else:
26                 normally_dist_feature_1[i].append(col)
27
28         # Calcolo varianza e errore relativo
29         variance = np.var(df_patient[col], ddof=1)
30         mean = np.mean(df_patient[col])
31         abs_error = variance / np.sqrt(len(df_patient[col]))
32         rel_error = abs_error / mean * 100
33
34         df_rel_error_1 = pd.concat([df_rel_error_1, pd.
  DataFrame({'PtID': [i], 'Feature': [col], 'Relative Error (%)':
  [rel_error])}], ignore_index=True)
35
36     else:
37         #print(f'Reject the null hypothesis: {col} is not
  normally distributed (p <= {alpha})\n')
38         # Calcolo varianza e errore relativo
39         variance = np.var(df_patient[col], ddof=1)
40         median = np.median(df_patient[col])
41         abs_error = variance / np.sqrt(len(df_patient[col]))
42         rel_error = abs_error / median * 100
43
44         df_rel_error_1 = pd.concat([df_rel_error_1, pd.
  DataFrame({'PtID': [i], 'Feature': [col], 'Relative Error (%)':
  [rel_error])}], ignore_index=True)
45
46 df_rel_error2 = pd.DataFrame(columns=['PtId', 'MAG (ER%)', 'VRT (
  ER%)', 'LNG (ER%)', 'LAT (ER%)'])
47 df_rel_error2['PtId'] = df_rel_error_1['PtID'].unique()
48 df_rel_error2 = df_rel_error2.set_index('PtId')
49
50 for i in df_rel_error2.index:
51     df_temp = df_rel_error_1[df_rel_error_1['PtID']==i]
52     for col in df_temp['Feature']:
53         if col == 'D.VRT - A media(mm)':

```

```

54         df_rel_error2.loc[i, 'VRT (ER%)'] = df_temp[df_temp['
Feature'] == col]['Relative Error (%)'].values[0]
55         elif col == 'D.LNG - A media(mm)':
56             df_rel_error2.loc[i, 'LNG (ER%)'] = df_temp[df_temp['
Feature'] == col]['Relative Error (%)'].values[0]
57             elif col == 'D.LAT - A media(mm)':
58                 df_rel_error2.loc[i, 'LAT (ER%)'] = df_temp[df_temp['
Feature'] == col]['Relative Error (%)'].values[0]
59                 elif col == 'D.MAG - A media(mm)':
60                     df_rel_error2.loc[i, 'MAG (ER%)'] = df_temp[df_temp['
Feature'] == col]['Relative Error (%)'].values[0]
61 df_rel_error2.reset_index( inplace=True)
62 columns = ['MAG (ER%)', 'VRT (ER%)', 'LNG (ER%)', 'LAT (ER%)']
63 for col in columns:
64     df_rel_error2[col] = df_rel_error2[col].astype(float)

```

```

1 # Importo il file excel con le informazioni sui pazienti con 15
   frazioni
2 df_infoPZ_15fr = pd.read_excel("percorso file con informazioni sui
   pazienti")
3 df_infoPZ_15fr.drop(columns=['Unnamed: 0'], inplace=True)
4
5 # Unisco le informazioni ai dati degli ER% calcolati con il test di
   Shapiro-Wilk
6 df_prima_analisi = pd.merge(df_rel_error2_1, df_infoPZ_15fr, how='
   left', left_on='PtId', right_on='PtID')
7 df_prima_analisi = df_15fr_final_1.drop(columns=['PtId'])
8
9 # Rimuovo i pazienti con 'Fr' diverso da 13 o 15
10 df_prima_analisi = df_prima_analisi[df_prima_analisi['Fr'].isin
   ([13, 15])]
11 df_prima_analisi = df_prima_analisi.reset_index(drop=True)

```

### B.3 Esempi di pazienti

```

1 df_new = df_prima_analisi.copy()
2 df2 = df_1315.copy()
3 columns = ['D.MAG - A media(mm)', 'D.MAG - Fr media(Hz)',
4           'D.VRT - A media(mm)',
5           'D.VRT - Fr media(Hz)',
6           'D.LNG - A media(mm)', 'D.LNG - Fr media(Hz)',
7           'D.LAT - A media(mm)',
8 df2 = df2[['PtID', 'date'] + columns]

1 # Seleziono una paziente con ER% < 5 tra quelle presenti nel
   DataFrame df_new
2 np.random.seed(6)
3 pz1 = df_new[(df_new['target'] == 1) & (df_new['MAG (ER%)'] < 5)][
   'PtID'].sample(n=1).iloc[0]
4 print(f'ID di un paziente "Ripetibile": {pz1}')
5 print(f"ER% della paziente selezionata: {df_new[df_new['PtID'] == pz1
   ]['MAG (ER%)'].iloc[0]}%")
6 print(f"Volume del tumore: {df_new[df_new['PtID'] == pz1]['volume
   tumore (CTV) [cc]'].iloc[0]} cc")
7 print(f"Etá: {df_new[df_new['PtID'] == pz1]['age at first fr'].iloc
   [0]} anni")

```

```

8 print(f"Sx/Dx: {'sx' if df_new[df_new['PtID']==pz1]['sx/dx'].iloc
    [0] == 1 else 'dx'}")
9
10 df_pz1 = df2[df2['PtID']==pz1]

```

```

1 # Grafici A media e Fr media per le 4 coordinate in funzione del
    tempo per la paziente selezionata
2 plt.figure(figsize=(20, 10))
3 plt.subplot(2, 2, 1)
4 for i, coord in enumerate(['MAG', 'VRT', 'LNG', 'LAT']):
5     plt.plot(df_pz1['date'], df_pz1[f'D.{coord} - A media(mm)'],
6             marker='o', label=f'{coord} A media')
7     plt.title(f'A media per paziente {pz_ripetibile}')
8     plt.xlabel('Data')
9     plt.ylabel('Valore')
10    plt.xticks(rotation=45)
11    plt.legend()
12    plt.grid()
13 plt.subplot(2, 2, 2)
14 for i, coord in enumerate(['MAG', 'VRT', 'LNG', 'LAT']):
15     plt.plot(df_pz1['date'], df_pz1[f'D.{coord} - Fr media(Hz)'],
16             marker='o', label=f'{coord} Fr media')
17     plt.title(f'Fr media per paziente {pz_ripetibile}')
18     plt.xlabel('Data')
19     plt.ylabel('Valore')
20     plt.xticks(rotation=45)
21     plt.legend()
22     plt.grid()

```

```

1 # Statistiche della paziente con ER% < 5%
2 df_pz1.describe()

```

```

1 # Seleziono una paziente con ER% > 10 tra quelle presenti nel
    DataFrame df_new
2 pz2 = df_new[(df_new['target']==0) & (df_new['MAG (ER%)'] > 10)][
    'PtID'].sample(n=1).iloc[0]
3 print(f'ID di un paziente "Non Ripetibile": {pz2}')
4 print(f"ER% della paziente selezionata: {df_new[df_new['PtID']==pz2
    ]['MAG (ER%)'].iloc[0]}%")
5 print(f"Volume del tumore: {df_new[df_new['PtID']==pz2]['volume
    tumore (CTV) [cc]'].iloc[0]} cc")
6 print(f"Et a: {df_new[df_new['PtID']==pz2]['age at first fr'].iloc
    [0]} anni")
7 print(f"Sx/Dx: {'sx' if df_new[df_new['PtID']==pz2]['sx/dx'].iloc
    [0] == 1 else 'dx'}")
8 df_pz2 = df2[df2['PtID']==pz2]

```

```

1 # Grafici A media e Fr media per le 4 coordinate in funzione del
    tempo per la paziente selezionata
2 plt.figure(figsize=(20, 10))
3 plt.subplot(2, 2, 1)
4 for i, coord in enumerate(['MAG', 'VRT', 'LNG', 'LAT']):
5     plt.plot(df_pz2['date'], df_pz2[f'D.{coord} - A media(mm)'],
6             marker='o', label=f'{coord} A media')
7     plt.title(f'A media per paziente {pz2}')
8     plt.xlabel('Data')
9     plt.ylabel('Valore')
10    plt.xticks(rotation=45)

```

```

10     plt.legend()
11     plt.grid()
12
13 plt.subplot(2, 2, 2)
14 for i, coord in enumerate(['MAG', 'VRT', 'LNG', 'LAT']):
15     plt.plot(df_pz2['date'], df_pz2[f'D.{coord} - Fr media(Hz)'],
16             marker='o', label=f'{coord} Fr media')
17     plt.title(f'Fr media per paziente {pz_non_ripetibile}')
18     plt.xlabel('Data')
19     plt.ylabel('Valore')
20     plt.xticks(rotation=45)
21     plt.legend()
22     plt.grid()

1 # Statistiche della paziente con ER% > 10%
2 df_pz2.describe()

```

## B.4 Classificazione dei pazienti

```

1 # Classifico le pazienti in base al loro valore di MAG (ER%)
2 coord = 'MAG (ER%)'
3 other_coord = ['VRT (ER%)', 'LNG (ER%)', 'LAT (ER%)']
4 df_new.drop(columns=other_coord, inplace=True)
5
6 # Classificazione pazienti in respiro Ripetibile o Non ripetibile a
7   seconda che MAG (ER%) sia <= cuttuff o > cuttuff
8 cuttuff = 5 #10
9 df_new['target'] = np.where(df_new[coord] <= cuttuff, 'Ripetibile',
10                            'Non ripetibile')
11
12 # Plot delle distribuzioni
13 cat_cols = ['sx/dx', 'Fr', 'target']
14 num_cols = df_new.select_dtypes(include=['int64', 'float64']).
15             columns
16
17 plt.figure(figsize=(20,10))
18 for (i,col) in enumerate(df_new.columns[2:]):
19     plt.subplot(2,4,i+1)
20     if col in cat_cols:
21         df_new[col].value_counts().plot(kind='bar')
22     else:
23         df_new[col].hist(bins=30)
24     plt.xlabel(col)
25     plt.ylabel('Frequency')
26
27 plt.tight_layout()
28 plt.show()
29
30 # Stampo le misure di centralità e dispersione delle variabili
31 for col in df_new.columns[2:]:
32     print(col.upper())
33     if col in cat_cols:
34         print('Moda:', df_new[col].mode()[0])
35         print()
36     elif col=='age at first fr':
37         print('Media:', df_new[col].mean())

```

```

35     print('Deviazione standard:', df_new[col].std())
36     print()
37     else:
38     print('Mediana:', df_new[col].median())
39     print()
40     print('Range:', df_new[col].max()-df_new[col].min())
41     print('Q1:', df_new[col].quantile(q=0.25))
42     print('Q3:', df_new[col].quantile(q=0.75))
43     print()

```

## Distribuzioni delle variabili in base alla ripetibilità

```

1 # Suddivido df2 in due DataFrame separati per le due classi '
   Ripetibile' e 'Non Ripetibile' e guardo come sono distribuite le
   Ampiezze e Frequenze medie in base alla ripetibilità
2 df2_Ripetibile = df2[df2['PtID'].isin(df_Ripetibile['PtID'])]
3 df2_Non_Rip = df2[df2['PtID'].isin(df_Non_Rip['PtID'])]
4 df2_Ripetibile[columns].describe(), df2_Non_Rip[columns].describe()

1 # Suddivido il df_new in due DataFrame separati per le due classi '
   Ripetibile' e 'Non Ripetibile'
2 df_Ripetibile = df_new[df_new['target']==1]
3 df_Non_Rip = df_new[df_new['target']==0]
4
5 # plot delle distribuzioni di "age at first fr" e "volume tumore (
   CTV) [cc]" in base alla ripetibilità (istogrammi)
6 n_bin = [50, 30]
7
8 plt.figure(figsize=(20,10))
9 for (i,col) in enumerate(['volume tumore (CTV) [cc]', 'age at first
   fr']):
10     plt.subplot(1,2,i+1)
11     df_Ripetibile[col].hist(bins=n_bin[i], alpha = 0.5, density=True)
12     df_Non_Rip[col].hist(bins=n_bin[i], color='orange', alpha=0.5,
   density=True)
13     plt.vlines(df_Ripetibile[col].mean(), 0,plt.ylim()[1], color='
   blue', linestyle='dashed', label='Media Ripetibile')
14     plt.vlines(df_Non_Rip[col].mean(), 0, plt.ylim()[1], color='
   orange', linestyle='dashed', label='Media Non Ripetibile')
15     plt.xlabel(col)
16     plt.ylabel('Frequency')
17     plt.legend(['Ripetibile', 'Non Ripetibile'])
18
19 plt.tight_layout()
20 plt.show()

1 # Misure di centralità e di dispersione per le due classi
2 for col in ['volume tumore (CTV) [cc]', 'age at first fr']:
3     print(col.upper())
4     if col in cat_cols:
5         print('Ripetibile:  Moda:', df_Ripetibile[col].mode()[0])
6         print('Non Ripetibile:  Moda:', df_Non_Rip[col].mode()[0])
7         print()
8     elif col=='age at first fr':
9         print('Ripetibile:  Media:', df_Ripetibile[col].mean())
10        print('\t Deviazione standard:', df_Ripetibile[col].std())
11        print('Non Ripetibile:  Media:', df_Non_Rip[col].mean())

```

```

12     print('\t Deviazione standard:', df_Non_Rip[col].std())
13     print()
14     else:
15         print('Ripetibile:  Mediana:', df_Ripetibile[col].median())
16         print('\t \tRange:', df_Ripetibile[col].max()-df_Ripetibile[col]
17         ].min())
18         print('\t \tQ1:', df_Ripetibile[col].quantile(q=0.25))
19         print('\t \tQ3:', df_Ripetibile[col].quantile(q=0.75))
20         print('Non Ripetibile:  Mediana:', df_Non_Rip[col].median())
21         print('\t \tRange:', df_Non_Rip[col].max()-df_Non_Rip[col].min
22         ())
23         print('\t \tQ1:', df_Non_Rip[col].quantile(q=0.25))
24         print('\t \tQ3:', df_Non_Rip[col].quantile(q=0.75))
25         print()

```

```

1 # Boxplot delle 2 variabili numeriche divise per classi di
   ripetibilità
2 for col in ['volume tumore (CTV) [cc]', 'age at first fr']:
3     df_new.boxplot(column=col, by='target')
4     plt.title(f'Boxplot di {col} diviso per target')
5     plt.suptitle('')
6     plt.xlabel('Target (0 = Non Ripetibile, 1 = Ripetibile)')
7     plt.ylabel(col)
8     plt.show()

```

```

1 # Distribuzione dell'ER% relativo al segnale MAG in base alla
   ripetibilità
2 df_Ripetibile['MAG (ER%)'].describe(), df_Non_Rip['MAG (ER%)'].
   describe()

```

```

1 # Confronto distribuzioni di sx/dx e Fr divise in base alla
   ripetibilità
2 plt.figure(figsize=(20,10))
3 grouped_data = df_new.groupby(["target", "sx/dx"]).size().unstack(
   fill_value=0)
4
5 # Plot the count of each value of 'sx/dx' for each 'target' class
6 grouped_data.plot(kind="bar", stacked=True)
7
8 # Add labels and title
9 plt.xlabel("Target (0 = Non Ripetibile, 1 = Ripetibile)")
10 plt.ylabel("Count (0 = dx, 1 = sx)")
11 plt.title("Sx/Dx count per Target class")
12
13
14 grouped_data = df_new.groupby(["target", "Fr"]).size().unstack(
   fill_value=0)
15 grouped_data.plot(kind="bar", stacked=True)
16 plt.xlabel("Target (0 = Non Ripetibile, 1 = Ripetibile)")
17 plt.ylabel("Count (0 = 15, 1 = 13)")
18 plt.title("Fr count per Target class")
19
20 plt.show()

```

Distribuzione dell'ER% per classi di età e volume

```

1 # Suddivido il DataFrame in base all'età in bin di 5 anni
2 bins = [0, 45, 55, 65, 75, np.inf]
3 labels = ['18-45', '46-55', '56-65', '66-75', '76+']
4 df_new['age_bin'] = pd.cut(df_new['age at first fr'], bins=bins,
5     labels=labels, right=False)
6
7 #Boxplot
8 df_new.boxplot(column='MAG (ER%)', by='age_bin', figsize=(20,10))
9 plt.ylim(0, 35)
10 # Aumentare il carattere dei numeri sull'asse x
11 plt.xticks(fontsize=15)
12 plt.yticks(fontsize=15)
13 plt.title('Boxplot di MAG (ER%) diviso per fasce di età')
14
15 # Statistiche di ER% per le diverse classi di età
16 df_new.groupby('age_bin')['MAG (ER%)'].describe()
17
18 # Suddivido il volume del tumore in bin di 400 cc
19 bins = [0, 400, 800, 1200, np.inf]
20 labels = ['0-400', '401-800', '801-1200', '1200+']
21 df_new['volume_bin'] = pd.cut(df_new['volume tumore (CTV) [cc]'],
22     bins=bins, labels=labels, right=False)
23
24 #Boxplot
25 df_new.boxplot(column='MAG (ER%)', by='volume_bin', figsize=(20,10)
26     )
27 plt.ylim(0, 35)
28 plt.xticks(fontsize=15)
29 plt.yticks(fontsize=15)
30 plt.title('Boxplot di MAG (ER%) diviso per fasce di volume del
31     tumore')
32
33 # Statistiche di ER% per le diverse classi di volume tumore
34 df_new.groupby('volume_bin')['MAG (ER%)'].describe()

```

## B.5 Analisi statistica

```

1 # test di Wilcoxon-Mann-Whitney per le variabili numeriche divise
2   per le due classi
3 from scipy.stats import mannwhitneyu
4 for col in ['volume tumore (CTV) [cc]', 'age at first fr', 'sx/dx']:
5     stat, p = mannwhitneyu(df_Ripetibile[col], df_Non_Rip[col])
6     print(f"Test di Mann-Whitney U per la variabile '{col}':")
7     print(f"Statistiche U: {stat}, P-value: {p}")
8     if p < 0.05:
9         print("C'è una differenza significativa tra le due classi."
10             )
11     else:
12         print("Non c'è una differenza significativa tra le due
13             classi.")

```

# Bibliografia

- [1] AIOM-AIRTUM-SIAPEC-IAP. «I numeri del cancro in Italia». In: *Intermedia Editore* (2020).
- [2] Noura AlHinai. *Biomedical Signal Processing and Artificial Intelligence in Healthcare*. Elsevier, 2020, p. 28.
- [3] William B Ashe1 et al. «Analysis of respiratory kinematics: a method to characterize breaths from motion signals». In: *Physiological Measurement* 43.015007 (2022). DOI: <https://doi.org/10.1088/1361-6579/ac4d1a>.
- [4] Gaia Elisa Bernazzani. «TECNICHE 4DCT NELLA GESTIONE DEL MOTO RESPIRATORIO IN SGRT». Tesi di Laurea Triennale. Università di Bologna, 2025.
- [5] Prof. Silvia Bonettini. «Numerical Algorithms for Signal and Image Processing - Lecture notes».
- [6] A. M. Brunt et al. «One versus three weeks hypofractionated whole breast radiotherapy for early breast cancer treatment: the fast-forward phase III RCT». In: *Health Technology Assessment* 27 (2023), pp. 1–176. DOI: 10.3310/WWBF1044.
- [7] Prof. Franco Bui e Prof. Diego Cecchin. *Fisica delle Radiazioni*. URL: <https://www.unipd.it/nucmed/TF/TF.physics1.ita.html>. (ultimo aggiornamento: 18 settembre 2019).
- [8] Polgár C et al. «Radiotherapy of Breast Cancer-Professional Guideline 1st Central-Eastern European Professional Consensus Statement on Breast Cancer». In: *Pathol Oncol Res.* 28.1610378 (2022). DOI: 10.3389/pore.2022.1610378.
- [9] ASMN-IRCCS Reggio Emilia Dott. Elisa Grassi S.C. Fisica Medica. *Interazione radiazione-materia*. URL: <https://www.infermierimo.unimore.it/site/home/avvisi/documento210038955.html>.
- [10] Puppo Emanuela. «Ottimizzazione di piani di trattamento con tecniche di controllo respiratorio in pazienti con tumore alla mammella sinistra». Tesi di Laurea Magistrale. Università degli studi di Genova, 2023.
- [11] Chen L et al. «Accuracy of real-time respiratory motion tracking and time delay of gating radiotherapy based on optical surface imaging technique». In: *Radiation Oncology* 15.1 (2020). DOI: 10.1186/s13014-020-01611-6.
- [12] Andrew C. Leon. *Comprehensive Clinical Psychology*. Elsevier, 1998. Cap. 3.12 Descriptive and Inferential Statistics.

- [13] Lt Gen SR Mehta, M Semwal Maj V Suhag e Maj N Sharma. «Radiotherapy : Basic Concepts and Recent Advances». In: *Med J Armed Forces India* 66.2 (2010). DOI: 10.1016/S0377-1237(10)80132-7.
- [14] Prabhaker Mishra et al. «Descriptive Statistics and Normality Tests for Statistical Data». In: *Annals of Cardiac Anaesthesia* (2019). DOI: 10.4103/aca.ACA\_157\_18.
- [15] Tommaso Mughetti. «SURFACE GUIDED RADIATION THERAPY: IL RUOLO DEL TSRM NEI CONTROLLI DI QUALITÀ CON FANTOCCHIO DINAMICO». Tesi di Laurea Triennale. Università di Bologna, CORSO DI LAUREA IN TECNICHE DI RADIOLOGIA MEDICA, PER IMMAGINI E RADIOTERAPIA, 2024.
- [16] Balaji Narayanasamy et al. «Zero-Phase-Filtering based Digital Active EMI Filter». In: *IEEE* (2020). DOI: 10.1109/IPEMC-ECCEAsia48364.2020.9367786.
- [17] NIST/SEMATECH. *e-Handbook of Statistical Methods*. Cap. 7.2.1.3. URL: <https://www.itl.nist.gov/div898/handbook/prc/section2/prc213.htm>.
- [18] Birgitte V. Offersen et al. «Hypofractionated Versus Standard Fractionated Radiotherapy in Patients With Early Breast Cancer or Ductal Carcinoma In Situ in a Randomized Phase III Trial: The DBCG HYPO Trial». In: *Journal of Clinical Oncology* 38.31 (2020), pp. 3615–3625. DOI: 10.1200/JCO.20.01363.
- [19] Marco Pio Petrillo. «EPID-Based In Vivo Dosimetry for Patient-Specific QA: A Study on Independent Dose Calculation in Breast Cancer Radiotherapy». Tesi di Laurea Magistrale. Università degli Studi di Modena e Reggio Emilia, 2024.
- [20] Elham Piruzan et al. «Target motion management in breast cancer radiation therapy». In: *Radiology and Oncology* 55.4 (2021), pp. 393–408. DOI: 10.2478/raon-2021-0040.
- [21] Andreas Renner et al. «Explicitly encoding the cyclic nature of breathing signal allows for accurate breathing motion prediction in radiotherapy with minimal training data». In: *Physics and Imaging in Radiation Oncology* 30.100594 (2024). DOI: <https://doi.org/10.1016/j.phro.2024.100594>.
- [22] Anne Richter et al. «Effect of Breathing Motion in Radiotherapy of Breast Cancer. 4D Dose Calculation and Motion Tracking via EPID». In: *Strahlentherapie und Onkologie* 185.7 (2009), pp. 425–30. DOI: 10.1007/s00066-009-1980-1.