

# LA MATEMATICA DEI SUONI

Settimana Matematica, Dipartimento di Matematica, Pisa

5 – 7 Febbraio 2020

## 1 Segnali sonori

### 1.1 Fisica dei suoni

Per poter analizzare matematicamente un segnale sonoro abbiamo bisogno di creare un modello che traduca in termini matematici cosa intendiamo come suono.

Per provare a capire cosa è, e come si propaga, un suono, proviamo ad immaginare un esempio concreto. Supponiamo di avere un'ascoltatrice, Alice, ed un percussionista, Bob. Alice si trova ad una certa distanza da Bob, che ha in mano un tamburo. Ad un certo punto, Bob colpisce energicamente la membrana del tamburo, che inizia a vibrare.

L'esperienza ci insegna che Alice udirà il suono del tamburo; ma perché questo succede? Qual è la ragione fisica di questo fenomeno? Possiamo riassumerlo nei seguenti passi:

1. Prima che il tamburo venga percosso, l'aria nell'ambiente ha una pressione costante (almeno in assenza di altri stimoli). Questa sarà completamente determinata dalla pressione atmosferica.
2. Il tamburo viene percosso e comincia a vibrare, ed in particolare la sua membrana oscilla alzandosi ed abbassandosi un certo numero di volte al secondo<sup>1</sup>. I cambiamenti di posizione spaziale della membrana causano delle alterazioni locali nella pressione dell'aria circostante, che viene compressa o decompressa.
3. I cambiamenti di pressione locale si propagano nello spazio. Intuitivamente possiamo immaginarci l'aria che è stata compressa che cerca di espandersi a sua volta comprimendo quella attorno, e quella decompressa che invece si restringe.
4. Dopo un certo lasso di tempo, questi cambiamenti di pressione arrivano vicino al timpano di Alice, e questo traduce questi cambiamenti di pressione in segnali elettrici che arrivano al cervello, dove noi sentiamo il suono del tamburo.

Probabilmente, la maggior parte delle persone non riflette su questi passaggi quando sente un tamburo suonare. A noi serve averne almeno una comprensione ad alto livello

---

<sup>1</sup>Questo numero è quello che viene chiamato *frequenza*, e misurato in Hertz (Hz). Questo concetto riemergerà molto spesso in questi giorni!

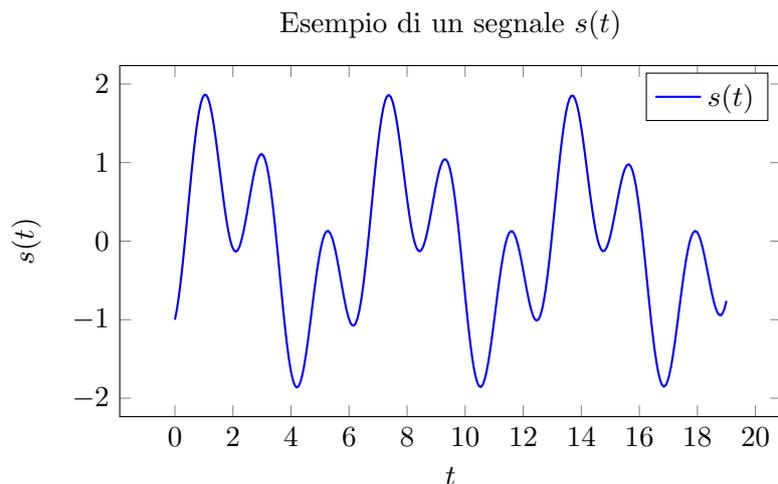


Figura 1: Esempio di un segnale acustico rappresentato come funzione del tempo.

per capire come rappresentare un segnale sonoro (ad esempio quello del tamburo). La trasmissione del suono attraverso lo spazio è quella che fisicamente si chiama *onda*. Non c'è una vera trasmissione di materia: l'aria cambia pressione momentaneamente ma non si sposta — avviene solo una trasmissione di *informazione*.

Questa trasmissione ha però bisogno della materia per avvenire! Se Alice e Bob fossero nel vuoto, allora Alice non avrebbe nessuna possibilità di sentire il rumore del tamburo.

## 1.2 Rappresentazione matematica di un'onda sonora

Ora che abbiamo capito, ad alto livello, come si propagano i suoni, possiamo proporre un modello matematico per studiarli. Come rappresentare il suono udito da Alice?

Quello che Alice sente si può pensare come una variazione di pressione sul suo timpano, all'interno dell'orecchio<sup>2</sup>. Supponiamo di fissare un istante iniziale  $t = 0$ , e rappresentiamo queste variazioni di pressione come un segnale  $s(t)$ , che potrebbe avere un grafico simile a quello in Figura 1.

Ricordiamoci che il suono ha un significato solo come variazione di pressione, e dunque non ha particolare rilevanza conoscere il valore della funzione  $s(t)$  in un dato istante  $t_0$ . Per essere in grado di apprezzare un suono, dobbiamo “ascoltare” la funzione su un intervallo di tempo di ampiezza maggiore di zero.

---

<sup>2</sup>Per essere precisi, Alice avrà quasi certamente due orecchi, ed in effetti sentirà due suoni contemporaneamente, dando origine ad un segnale “stereo”. Ma per ora ignoriamo questa complicazione.

### 1.3 Range udibile, e altezza di un suono

Il nostro orecchio è in grado di riconoscere alcune particolari funzioni  $s(t)$ , ovvero funzioni *periodiche*. Una funzione  $s(t)$  si dice di periodo<sup>3</sup>  $T$  se per ogni  $t \in \mathbb{R}$  vale  $s(t) = s(t+T)$ . In particolare, è chiaro che una funzione con questa proprietà è completamente determinata dal suo valore sull'intervallo  $[0, T)$  perché da lì in poi comincia a “ripetersi”.

Per semplicità, consideriamo come esempio concreto la funzione  $s(t) = A \sin(2\pi\omega t)$ , per qualche valore di  $\omega, A > 0$ , che ha periodo  $T = \frac{1}{\omega}$ . Definiamo anche una quantità detta frequenza, che si definisce come  $T^{-1}$ ; in questo caso specifico, la frequenza non è altro che  $\omega$ . Dal punto di vista fisico, se  $s(t)$  è una funzione del tempo, il periodo si misura in secondi, mentre la frequenza in Hertz, con simbolo  $\text{Hz} := \frac{1}{s}$ .

Dato un tale segnale, il nostro orecchio è in grado di riconoscere distintamente l'ampiezza  $A$  (che noi percepiamo come intensità sonora), e la frequenza  $\omega$ , che noi percepiamo come “altezza” del suono.

In effetti, ad ogni frequenza corrisponde una certa nota musicale. Per fare un esempio, prima di un concerto sinfonico è usanza che l'oboe suoni un La, su cui si accorderanno tutti gli altri strumenti. Se misurassimo in quel momento la pressione sonora in prossimità dello strumento, assumendo un assoluto silenzio del resto del mondo, osserveremmo esattamente il grafico di una funzione di periodo  $T = \frac{1}{440\text{Hz}} \approx 0.002273s$ ; 440Hz è la frequenza del La utilizzato come riferimento per l'accordatura nella musica moderna.

In aggiunta ad ampiezza e frequenza, il nostro orecchio è in grado di percepire la forma dell'onda. Data una funzione periodica si potrà determinare se è il suono di un oboe, di un violino, o di un pianoforte.

Sfortunatamente, non possiamo udire suoni con frequenze arbitrarie. Ad esempio, le frequenze inferiori a 20Hz non vengono percepite dalla maggior parte delle persone, così come quelle superiori ai 20000Hz. Questa soglia superiore si abbassa sempre di più con l'invecchiamento, fino ad arrivare sensibilmente sotto i 10000Hz. Con una visione un po' antropocentrica, questi suoni vengono chiamati rispettivamente *infrasuoni*, ed *ultrasuoni*.

Altri animali sono in grado di udire suoni in range molto più ampi; ad esempio, un pipistrello ha un range dell'udibile che si estende fino agli 80000Hz.

## 2 Dal continuo al discreto

Il nostro modello di un suono per un tempo  $T$  come una funzione  $s(t): [0, T] \rightarrow \mathbb{R}$  funziona molto bene dal punto di vista strettamente matematico. Tuttavia, il nostro obiettivo è elaborare e rappresentare suoni tramite un computer, o comunque in modo digitale.

Un computer è un dispositivo con una quantità *finita* di memoria, mentre è abbastanza chiaro che le funzioni da un'intervallo  $[0, T]$  nei numeri reali siano un numero infinito.

---

<sup>3</sup>Notiamo che, secondo questa definizione, una funzione di periodo  $T$  è periodica anche di periodo  $2T, 3T, \dots$ . In generale, consideriamo come periodo il più piccolo valore di  $T$  per cui  $s$  soddisfa questa proprietà.

In effetti, nonostante comparare cardinalità infinite sia una cosa da non prendere troppo alla leggera, è possibile mostrare che queste sono anche “di più” dei numeri naturali, e anche più dei reali.

## 2.1 Discretizzazione

La via più naturale per passare da un modello *continuo*  $s(t)$  di un suono ad uno discreto è *discretizzare* la variabile del tempo. In pratica, si tratta di definire una funzione  $x: I \subseteq \mathbb{N} \rightarrow \mathbb{R}$  come segue:

$$x(j) := s(j\Delta t), \quad j = 0, \dots, \left\lfloor \frac{T}{\Delta t} \right\rfloor.$$

Useremo la notazione  $x_j$  al posto di  $x(j)$ , per suggerire la naturale interpretazione di  $x$  come un vettore. L'intervallo di tempo  $\Delta t$  è la distanza fra due successive misurazioni del segnale  $s(t)$  che vengono salvate in questo vettore; in inglese, si usa il termine *sample*.

È immediatamente chiaro che, dato un segnale  $s(t)$ , per ogni scelta dell'intervallo di *sampling*  $\Delta t$  il vettore  $x$  è univocamente determinato. Apparentemente, c'è però molta perdita di informazione in questa restrizione ad istanti discreti.

Proviamo a fare un esempio pratico. Consideriamo la funzione periodica  $s_1(t) = \sin(6\pi t)$ , definita su  $[0, 1]$ , e supponiamo di scegliere come  $\Delta t = \frac{1}{12}$ . Se costruiamo la corrispondente successione  $x$  otteniamo il vettore:

$$x = [ 0 \quad 1 \quad 0 \quad -1 \quad 0 \quad 1 \quad 0 \quad -1 \quad 0 \quad 1 \quad 0 \quad -1 ]$$

Consideriamo ora la funzione  $s_2(t) := 2\sin(9\pi t)\cos(3\pi t)$ . Queste sono chiaramente funzioni diverse (e quindi suoni diversi), come visibile in Figura 2, eppure le valutazioni  $s_2(j\Delta t)$  forniscono esattamente lo stesso vettore  $x$ . Questo significa che non saremmo in grado di distinguere questi due suoni solamente misurandoli a tempi discreti. Almeno, non scegliendo  $\Delta t := \frac{1}{12}$ .

## 2.2 Teorema di Shannon, Nyquist rate

Tuttavia, la situazione non è tragica come potrebbe inizialmente sembrare. A patto di scegliere l'intervallo  $\Delta t$  sufficientemente piccolo, possiamo garantire di non perdere nessuna parte del segnale fra quelle che possiamo sentire. Ricordiamoci che il nostro orecchio distingue l'altezza delle funzioni in base alla loro frequenza, o periodo. Dunque, supponiamo ora che il nostro segnale di interesse sia composto solamente da frequenze all'interno di un intervallo  $[0, B]$ . Per esempio, potremmo supporre di avere un segnale  $s(t)$  ottenuto sovrapponendo segnali con frequenze diverse:

$$s(t) = s_1(t) + s_2(t) + \dots \tag{1}$$

dove  $s_1(t)$  è una funzione periodica con frequenza  $f_1$ ,  $s_2(t)$  con frequenza  $f_2$ , e così via, in modo che per ogni  $j$  si abbia  $0 \leq f_j \leq B$ . Le funzioni che sovrapponiamo per ottenere

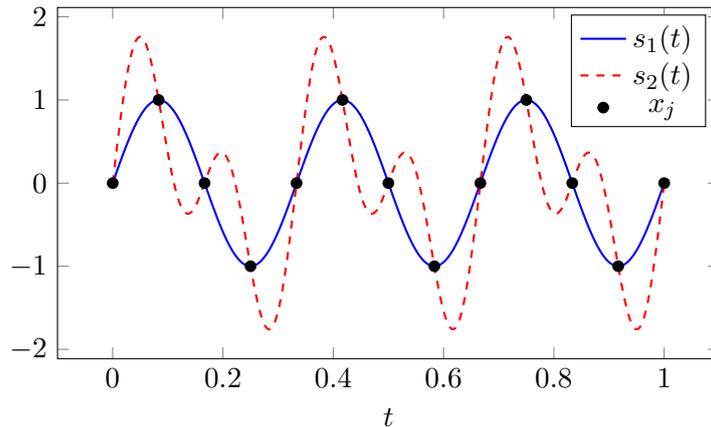


Figura 2: Esempio di due funzioni distinte, per cui il processo di discretizzazione fornisce esattamente lo stesso vettore  $x$ .

$s(t)$  possono essere molte, anche infinite. In questo contesto, possiamo enunciare il seguente risultato<sup>4</sup>.

**Teorema 1** (Shannon-Nyquist). *Sia  $s(t)$  una funzione che non contenga frequenze al di fuori di  $[0, B]$ , ad esempio ottenuta come descritto in (1). Allora, la funzione  $s(t)$  può essere ricostruita univocamente da un suo qualunque sampling con  $\Delta t < \frac{1}{2B}$ .*

L'intervallo di tempo  $\Delta t = \frac{1}{2B}$ , ovvero il più grande possibile che permette di non perdere alcuna informazione relativa alle frequenze minori di  $B$ , viene solitamente chiamato *Nyquist rate*, o *frequenza di Nyquist*.

### 2.3 Compressione audio, frequenze udibili

L'osservazione che, dato un sampling rate  $\Delta t$  sufficientemente piccolo, è possibile rappresentare senza perdita un segnale che rappresenta delle frequenze limitate è alla base dell'elaborazione audio moderna. Quando nel 1980 Philips e Sony dovettero scegliere un modo per memorizzare l'audio sull'appena nato CD Audio, si trovarono a dover selezionare un  $\Delta t$  abbastanza piccolo per rappresentare fedelmente tutti i suoni udibili, ma allo stesso tempo non troppo piccolo per permettere di memorizzare brani di una durata soddisfacente.

Ricordiamo che la massima frequenza udibile è intorno ai 20.000Hz, per cui il relativo Nyquist sampling rate, il minimo accettabile, è di circa 40.000Hz. Per questa ragione, venne scelto (ed è abbastanza in uso tutt'ora), il valore di 44100Hz. In particolare, quando un segnale continuo, ad esempio quello registrato dal microfono, viene discretizzato

<sup>4</sup>L'enunciato del seguente risultato è un pochino semplificato rispetto alla versione originale, per mancanza di tutti gli strumenti teorici necessari ad enunciarlo e dimostrarlo nella forma più completa.

con un sampling rate di 44100Hz, un essere umano non noterà nessuna differenza fra il segnale originale e quello ricostruito.

Ovviamente, un pipistrello inorridirebbe<sup>5</sup> ad ascoltare un CD, rispetto ad un concerto dal vivo.

### 3 Serie di Fourier

Riassumiamo le questioni che abbiamo chiarito fino a questo punto:

- Un suono si può rappresentare matematicamente come una funzione  $s(t)$ , definita sull'intervallo di tempo  $[0, t_{\max}]$  di interesse.
- Se questo suono corrisponde ad un'unica nota prodotta da uno strumento musicale, la funzione  $s(t)$  è periodica di periodo  $T = \frac{1}{f}$ , con  $f$  uguale alla frequenza della nota scelta.
- La *forma* di questa funzione periodica determina il timbro dello strumento che ascoltiamo.

Se vogliamo provare a riprodurre uno strumento utilizzando un computer già sappiamo che dobbiamo restringerci ad un sampling discreto, invece che ad una funzione definita su un insieme continuo, ma abbiamo bisogno di capire come generare una funzione di una data forma.

Per raggiungere questo obiettivo, facciamo un breve detour attraverso questioni apparentemente scorrelate, ma che invece ci torneranno molto utili a breve.

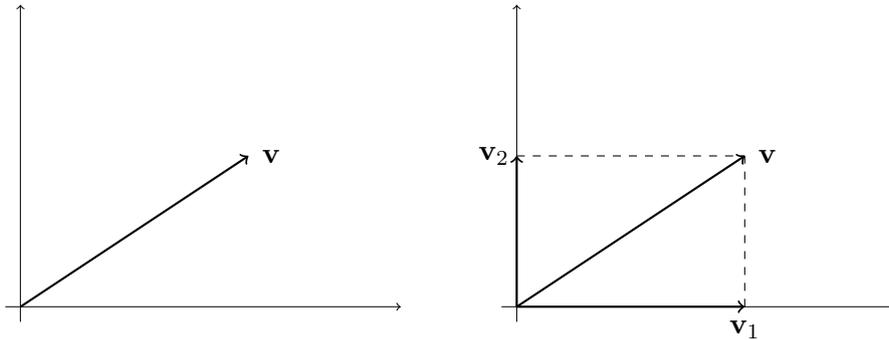
#### 3.1 Spazi vettoriali, prodotti scalari, e coordinate

Cos'è un vettore? Per il momento non ci preoccupiamo di dare una definizione precisa, ma partiamo dall'esempio con cui è più facile avere dimestichezza: i vettori all'interno del piano cartesiano  $\mathbb{R}^2$ .

Un vettore  $\mathbf{v}$  viene rappresentato graficamente sul piano cartesiano, il che permette di individuare la sue "componenti" lungo l'asse delle  $x$  e delle  $y$ , come evidenziato nel piano a destra.

---

<sup>5</sup>In realtà, anche questo non è particolarmente vero, dato che la maggior parte dei suoni di interesse in un CD hanno le frequenze principali intorno ai 500Hz o 1000Hz; probabilmente, un pipistrello potrebbe notare giusto una minima differenza.



Le componenti  $\mathbf{v}_1$  e  $\mathbf{v}_2$  sono semplicemente una delle tante decomposizioni additive di  $\mathbf{v}$ , ovvero una scrittura di  $\mathbf{v}$  come  $\mathbf{v} = \mathbf{v}_1 + \mathbf{v}_2$ . Tuttavia, questa scelta ha dei vantaggi notevoli:

- Le componenti  $\mathbf{v}_1$  e  $\mathbf{v}_2$  vengono scelte come multipli di due vettori di lunghezza unitaria lungo l'asse  $x$  ed  $y$ , che possiamo chiamare  $\mathbf{e}_1$  ed  $\mathbf{e}_2$ . In particolare, possiamo scrivere  $\mathbf{v}_1 = v_1\mathbf{e}_1$  e  $\mathbf{v}_2 = v_2\mathbf{e}_2$ , dove  $v_1, v_2$  sono ora dei numeri reali.
- Come vedremo, è relativamente semplice calcolare le componenti di  $\mathbf{v}$  utilizzando un *prodotto scalare*.

Cos'è un prodotto scalare? In maniera astratta possiamo pensarlo come una particolare funzione  $f(v, w)$  che, dati in input due vettori, ci restituisce un numero (ovvero uno *scalare*, da cui il nome). La notazione usuale per questa funzione  $f(v, w)$  è  $\langle v, w \rangle$ . Questa funzione per meritare il nome di *prodotto scalare*, deve avere alcune proprietà aggiuntive, ad esempio:

$$\begin{aligned}
 \langle \mathbf{v}, \mathbf{v} \rangle &\geq 0, & \langle \mathbf{v}, \mathbf{v} \rangle = 0 &\iff \mathbf{v} \equiv \mathbf{0} \\
 \langle \alpha \mathbf{v}, \mathbf{w} \rangle &= \alpha \langle \mathbf{v}, \mathbf{w} \rangle, & \langle \mathbf{v}, \mathbf{w} \rangle &= \langle \mathbf{w}, \mathbf{v} \rangle, \\
 \langle \mathbf{v} + \mathbf{u}, \mathbf{w} \rangle &= \langle \mathbf{v}, \mathbf{w} \rangle + \langle \mathbf{u}, \mathbf{w} \rangle.
 \end{aligned} \tag{2}$$

per ogni  $\alpha \in \mathbb{R}$ . Quando si studia fisica, è usuale definire questo prodotto scalare come  $\langle v, w \rangle := \|v\| \cdot \|w\| \cdot \cos(\theta)$ , dove  $\|\cdot\|$  indica la lunghezza Euclidea, e  $\theta$  è l'angolo "interno" fra due vettori. Dal punto di vista matematico, in generale si preferisce seguire il percorso inverso, ovvero definire cos'è un prodotto scalare, ed utilizzarlo per introdurre l'angolo interno e la lunghezza.

Secondo la definizione sopra abbiamo immediatamente che

$$\langle \mathbf{e}_1, \mathbf{e}_1 \rangle = 1, \quad \langle \mathbf{e}_2, \mathbf{e}_2 \rangle = 1, \quad \langle \mathbf{e}_1, \mathbf{e}_2 \rangle = \langle \mathbf{e}_2, \mathbf{e}_1 \rangle = 0$$

da cui segue, utilizzando le proprietà introdotte in (2):

$$\begin{aligned}
 \langle \mathbf{v}, \mathbf{e}_1 \rangle &= \langle v_1\mathbf{e}_1 + v_2\mathbf{e}_2, \mathbf{e}_1 \rangle = v_1\langle \mathbf{e}_1, \mathbf{e}_1 \rangle + v_2\langle \mathbf{e}_2, \mathbf{e}_1 \rangle = v_1, \\
 \langle \mathbf{v}, \mathbf{e}_2 \rangle &= \langle v_1\mathbf{e}_1 + v_2\mathbf{e}_2, \mathbf{e}_2 \rangle = v_1\langle \mathbf{e}_1, \mathbf{e}_2 \rangle + v_2\langle \mathbf{e}_2, \mathbf{e}_2 \rangle = v_2.
 \end{aligned}$$

Se fissiamo la nostra scelta di vettori  $\mathbf{e}_1$  ed  $\mathbf{e}_2$ , allora possiamo identificare  $\mathbf{v}$  tramite le sue *coordinate*, ovvero i numeri  $v_1, v_2$ , e scriviamo

$$\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}.$$

Utilizzando questa notazione, è anche più semplice esprimere il prodotto scalare fra  $\mathbf{v}$  e  $\mathbf{w}$  in termini delle coordinate:

$$\langle \mathbf{v}, \mathbf{w} \rangle = v_1 w_1 + v_2 w_2.$$

### 3.2 Basi di seni e coseni

Ora che ci siamo un po' rinfrescati le idee su come funziona il prodotto scalare canonico su  $\mathbb{R}^2$ , possiamo fare un paio di osservazioni preliminari:

- La stessa idea si potrebbe estendere in più dimensioni. Per esempio, in uno spazio tridimensionale avremo bisogno di 3 vettori  $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$  per “costruire” tutti gli altri, che potrebbero essere orientati secondo l’asse  $x, y, z$ .
- Allo stesso tempo, la scelta che abbiamo fatto di  $\mathbf{e}_1, \mathbf{e}_2$  non è necessariamente l’unica possibile: ad esempio, cosa succede se invece che  $\mathbf{e}_2$  scegliamo  $-\mathbf{e}_2$ ? La proprietà che fa funzionare veramente le cose è che  $\langle \mathbf{e}_1, \mathbf{e}_2 \rangle = 0$ , ovvero che i vettori siano ortogonali, e che  $\langle \mathbf{e}_i, \mathbf{e}_i \rangle = 1$  per ogni  $i$ . In più dimensioni, chiederemo sempre che

$$\langle \mathbf{e}_i, \mathbf{e}_j \rangle = \begin{cases} 1 & \text{se } i = j \\ 0 & \text{altrimenti} \end{cases}.$$

Torniamo al nostro caso delle funzioni periodiche, e supponiamo di avere un segnale  $s(t)$  periodico su  $[0, 2\pi]$ . Vogliamo usare la stessa idea di decomposizione che abbiamo usato per i vettori in  $\mathbb{R}^2$ , ovvero provare a decomporre il nostro segnale  $s(t)$  come combinazione di funzioni periodiche più semplici, ad esempio seni e coseni.

È chiaro che, per ogni scelta di coefficienti  $a_i$  per  $i = 0, \dots, k$  e  $b_i$  per  $i = 1, \dots, k$ , la funzione

$$s(t) = \frac{a_0}{2} + a_1 \cos(t) + b_1 \sin(t) + a_2 \cos(2t) + b_2 \sin(2t) + \dots + a_k \cos(kt) + b_k \sin(kt) \quad (3)$$

è una funzione periodica su  $[0, 2\pi]$ . Ma è vera anche l’implicazione inversa? Ovvero, è possibile scrivere ogni funzione periodica come combinazioni di seni e coseni? La risposta è negativa in tutta questa generalità, ma vedremo che ci possiamo avvicinare molto.

Immaginiamo che  $s(t)$  sia il nostro vettore  $\mathbf{v}$ , e  $\sin(jt)$  e  $\cos(jt)$  svolgano il ruolo dei nostri  $\mathbf{e}_j$ . Per determinare i coefficienti  $a_i, b_i$ , che invece svolgeranno il ruolo di  $v_i$ , avremmo bisogno di un prodotto scalare. Dato che i nostri vettori ora sono delle funzioni, abbiamo bisogno di qualcosa che, date in input due generiche funzioni  $s_1(t)$  ed  $s_2(t)$ , restituisca un numero.

Proponiamo il seguente candidato prodotto scalare:

$$\langle s_1(t), s_2(t) \rangle := \frac{1}{\pi} \int_0^{2\pi} s_1(t) s_2(t) dt.$$

In effetti, questa definizione soddisfa le proprietà richieste in (2), anche se non lo verificiamo. Perché questo prodotto scalare facciamo veramente al caso nostro, è necessario che  $\langle \cos(it), \cos(jt) \rangle = 0 = \langle \sin(it), \sin(jt) \rangle$  ogni qual volta  $i \neq j$ , e così che  $\langle \cos(it), \sin(jt) \rangle = 0$  per ogni  $i, j$ . È veramente così? Proviamo a verificarlo in qualche caso semplice, lasciando tutte le prove nel caso più generale per esercizio. Osserviamo che

$$\frac{\cos((i+j)t) + \cos((i-j)t)}{2} = \cos(it) \cos(jt),$$

come si può stabilire con le usuali formule per  $\cos(\alpha \pm \beta)$ . Nel caso non avessimo eccessiva familiarità con gli integrali, per ora ci basta ricordare che

$$\int_a^b f(t) dt = \text{“Area sottesa al grafico di } f(t) \text{ fra } a \text{ e } b\text{”},$$

il che in particolare permette di mostrare abbastanza facilmente che:

$$\int_0^{2\pi} f_1(t) + f_2(t) dt = \int_0^{2\pi} f_1(t) dt + \int_0^{2\pi} f_2(t) dt.$$

Consideriamo ora  $\langle \cos(it), \cos(jt) \rangle$ ; possiamo scrivere

$$\langle \cos(it), \cos(jt) \rangle = \frac{1}{\pi} \int_0^{2\pi} \cos(it) \cos(jt) dt \quad (4)$$

$$= \frac{1}{2\pi} \int_0^{2\pi} \cos((i+j)t) dt + \frac{1}{2\pi} \int_0^{2\pi} \cos((i-j)t) dt. \quad (5)$$

Prima di continuare, osserviamo che,

$$\int_0^{2\pi} \cos(nt) dt = 0 \quad \forall n \in \mathbb{N} \setminus \{0\}. \quad (6)$$

Possiamo darne una “dimostrazione grafica”. Se  $n$  è positivo, il grafico di  $\cos(nt)$  è una funzione periodica di periodo  $\frac{2\pi}{n}$ . All’interno di ogni periodo, è positiva per un intervallo di ampiezza  $\frac{\pi}{2n}$ , poi negativa per un altro intervallo di ampiezza  $\frac{\pi}{n}$ , e ancora negativa per un altro intervallo di ampiezza  $\frac{\pi}{2n}$ .

In questi due intervalli la funzione è esattamente uguale, ma specchiata rispetto all’asse delle  $x$ , e quindi la somma delle due aree<sup>6</sup> (considerando il segno) è nulla. Ad esempio, in Figura 3 è riportata una rappresentazione grafica delle parte positiva e negativa dell’integrale della funzione  $\cos(5x)$ .

<sup>6</sup>Nella definizione di integrale l’area di una funzione negativa ha segno negativo.

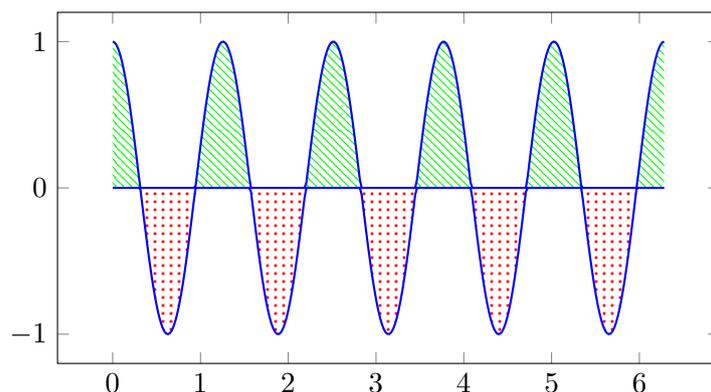


Figura 3: Rappresentazione grafica dell'integrale fra 0 e  $2\pi$  della funzione  $\cos(5x)$ .

Grazie a (6), combinata con l'espressione in (5), è un semplice esercizio<sup>7</sup> concludere che

$$\langle \cos(it), \cos(jt) \rangle = \begin{cases} 0 & \text{se } i \neq j \\ 1 & \text{se } i = j \neq 0 \\ 2 & \text{se } i = j = 0 \end{cases} \quad (7)$$

Ora che abbiamo verificato che i coseni soddisfano la proprietà di *ortogonalità*, potremmo verificare anche le altre relazioni, ad esempio che  $\langle \sin(it), \cos(jt) \rangle = 0$ , e quella che coinvolge due funzioni seno. Sarebbe però un conto tedioso, e dunque lo lasciamo al lettore.

Mettendo insieme tutti questi risultati possiamo finalmente determinare una formula per i coefficienti  $a_i$  e  $b_i$  per una generica funzione  $s(t)$ :

$$\begin{aligned} a_i &= \langle s(t), \cos(it) \rangle = \frac{1}{2\pi} \int_0^{2\pi} \cos(it) s(t) dt, \\ b_i &= \langle s(t), \sin(it) \rangle = \frac{1}{2\pi} \int_0^{2\pi} \sin(it) s(t) dt. \end{aligned} \quad (8)$$

La presenza di un 2 per  $i = j = 0$  in (7) chiarisce anche perché abbiamo scelto di mettere  $\frac{a_0}{2}$  invece che  $a_0$  nell'espressione (3).

### 3.3 Risultati ed esempi di convergenza

Ora che abbiamo un risultato per determinare i coefficienti  $a_i, b_i$ , li possiamo calcolare per ogni funzione del tipo riportato in (3). In effetti, nessuno ci obbliga a fermarci a sommare un numero finito di funzioni seno e coseno, e possiamo estendere l'interesse a

<sup>7</sup>L'unico caso che rimane scoperto è l'integrale in (6) per il caso  $n = 0$ , ma  $\cos(0 \cdot t) \equiv 1$ , e dunque si tratta semplicemente di calcolare l'area di un rettangolo.

funzioni del tipo:

$$s(t) = \frac{a_0}{2} + \sum_{i=1}^{\infty} a_i \cos(it) + b_i \sin(it). \quad (9)$$

L'espressione qui sopra, con i coefficienti  $a_i, b_i$  che possono essere determinati tramite (8), è detta *serie di Fourier di  $s(t)$* .

È naturale chiedersi se ogni funzione periodica su  $[0, 2\pi]$  debba per forza essere del tipo sopra, per qualche opportuna scelta di  $a_i, b_i$ . Più precisamente, ricordiamoci sempre che noi dovremo lavorare su un computer, e sommare un numero infinito di addendi mal si combina con la memoria finita di cui disponiamo, e anche con la quantità finita di tempo che abbiamo a disposizione.

Ci chiediamo dunque: data una funzione  $s(t)$  che rispetti determinate ipotesi, posso concludere che, per un valore di  $k$  sufficientemente grande, la somma in (9) è una buona approssimazione di  $s(t)$ ? In effetti è così, e determinare esattamente quando questo succede (e come succede) ha impegnato molti matematici alla fine del XIX secolo.

La risposta è: dipende; come regola generale, più la funzione è “regolare”, più è veloce la sua convergenza. Ad esempio, nel caso in cui la funzione sia derivabile con continuità, la convergenza è automaticamente garantita.

Facciamo qualche esempio per chiarirci le idee. Consideriamo la funzione  $s(t) = e^{\sin t + \cos t}$ , definita sull'intervallo  $[-\pi, \pi]$ , e osserviamo che:

- La funzione  $s(t)$  è periodica di periodo  $2\pi$ , i.e., se la estendessimo su tutto  $\mathbb{R}$  avremmo  $s(t + 2k\pi) = s(t)$  per ogni  $k \in \mathbb{Z}$ .
- La funzione è  $C^m$  per ogni  $m \geq 0$ , ovvero è  $C^\infty$ : possiamo derivarla un numero arbitrario di volte, e otterremo sempre una funzione continua (e periodica!).
- La funzione non è data da una somma finita di seni e coseni, i.e.,

$$s(t) \neq \sum_{i=0}^k a_i \cos(it) + b_j \sin(it)$$

per ogni scelta di  $a_i, b_i$  e  $k$  finito.

Come accennato prima, data la regolarità di questa funzione, ci aspettiamo invece che una somma *infinita* di seni e coseni sia in grado di rappresentare  $s(t)$ , e dunque che al crescere degli elementi che sommiamo, con  $a_i, b_i$  scelti come dall'equazione (8), la differenza fra  $s(t)$  e la somma converga a zero.

Chiamiamo  $s_k(t)$  per  $k = 0, 1, \dots$  la somma parziale dei coefficienti di Fourier ottenuta sommando fino all'indice  $k$ . Per cui:

$$s_0(t) = \frac{a_0}{2}, \quad s_1(t) = \frac{a_0}{2} + a_1 \cos(t) + b_1 \sin(t), \quad \dots$$

In Figura 4 possiamo osservare come queste funzioni si avvicinano ad  $s(t)$  mano a mano che l'indice  $k$  cresce. È immediatamente chiaro che già  $s_3(t)$  è difficilmente

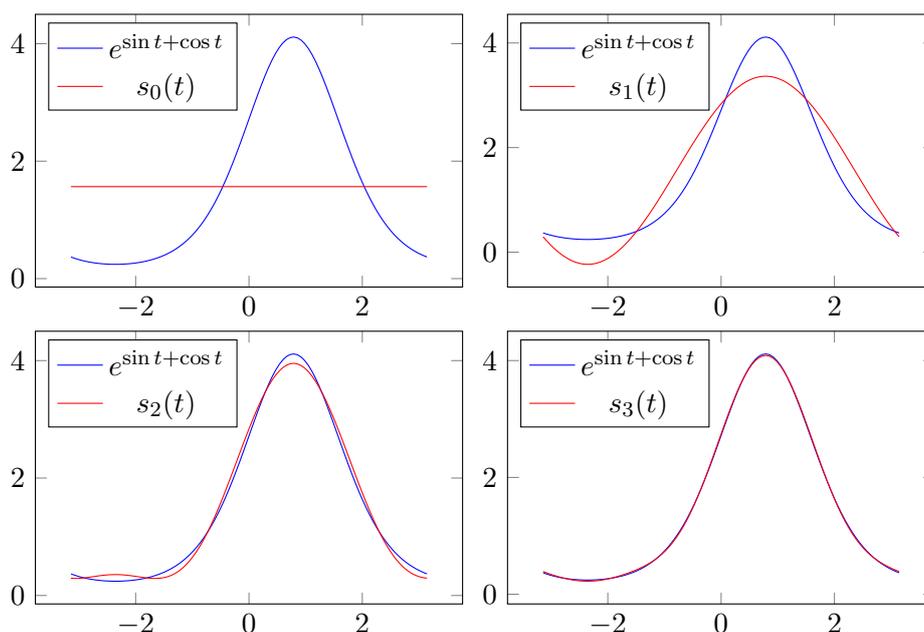


Figura 4: Convergenza delle somme parziali della serie di Fourier di  $s(t) = e^{\sin t + \cos t}$ .

distinguibile, ad occhio nudo, dalla funzione originale. In effetti, questo è un caso in cui la funzione è molto regolare, e dunque la convergenza è piuttosto veloce.

Consideriamo invece un esempio di funzione discontinua, come  $s(t) = \frac{t}{|t|}$ , definita sempre su  $[-\pi, \pi]$ , e posta uguale a 0 in 0. Questa funzione vale  $-1$  sui numeri negativi, e  $1$  su quelli positivi. Dunque, il suo grafico fa un “salto” di altezza 2 a  $t = 0$ . Seni e coseni sono funzioni continue, e lo stesso vale per le loro somme. Dunque riesce difficile immaginare come possano convergere ad una funzione con un salto! In Figura 5 vediamo come se la cavano.

Ricordiamo che per la funzione precedente, che era molto regolare,  $s_3(t)$  era già un’ottima approssimazione della funzione. Ora, invece  $s_3(t)$  è ancora molto distante; si comincia ad intuire anche un fenomeno interessante: vicino al punto di discontinuità la convergenza si comincia a perdere, e mentre la funzione approssimante converge nei punti dove la funzione è regolare, in prossimità del salto di discontinuità ci sono dei “rimbalzi”. Questo fenomeno è ben noto, ed è detto *fenomeno di Gibbs*. Si verifica spesso quanto tentiamo di approssimare con funzioni regolari altre funzioni che, invece, non lo sono.

Lo stesso fenomeno si riscontra, anche se in maniera più debole, con un rallentamento della convergenza per funzioni la cui derivata (oppure la derivata  $m$ -esima, per qualche  $m$ ) non è continua. Non possiamo addentrarci nei dettagli di questa teoria, ma riportiamo in maniera qualitativa quello che ci si può aspettare; data una funzione con  $m$  derivate continue, le somme di Fourier convergeranno a questa funzione in modo

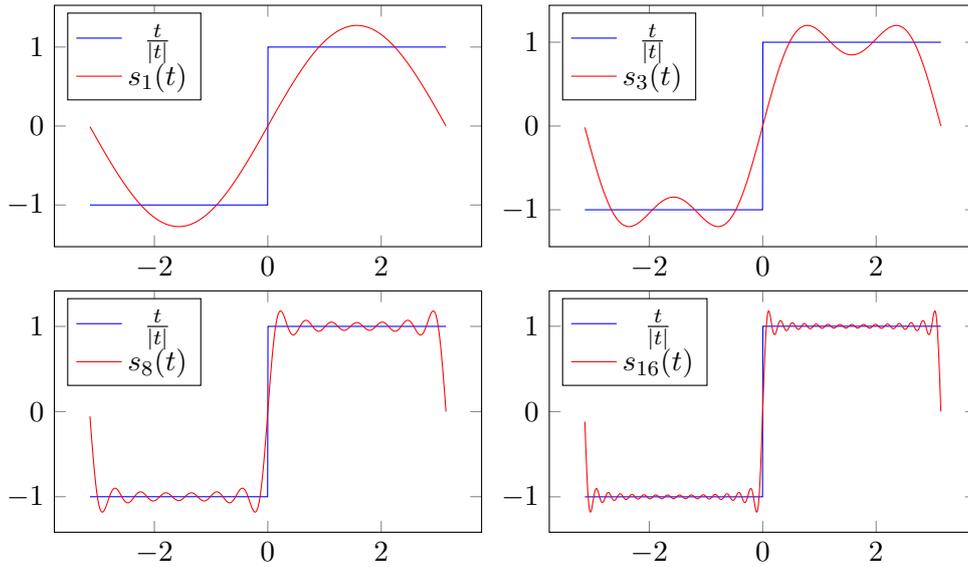


Figura 5: Convergenza delle somme parziali della serie di Fourier di  $s(t) = \frac{t}{|t|}$ , con  $s(0) := 0$ . Nelle ultime funzioni riportate si comincia a vedere il fenomeno di Gibbs, con dei “rimbalzi” intorno al punto di discontinuità di  $s(t)$ .

polinomiale, ovvero:

$$|s(t) - s_j(t)| \leq Cj^{-m},$$

dove  $C$  è una qualche costante opportuna che dipende dalla funzione in considerazione. Se invece la funzione ha un numero infinito di derivate continue, come ad esempio  $s(t) = e^{\sin t + \cos t}$  che abbiamo considerato in precedenza, allora la convergenza sarà più veloce di ogni polinomio, e avremo  $|s(t) - s_j(t)| \leq C\rho^j$  per un’opportuna costante  $C$  ed un numero reale  $0 < \rho < 1$ . Potrebbe non essere immediato, ma questo secondo tipo di convergenza è molto più veloce della prima.