Appunti su alcuni argomenti di Analisi II per Fisica (A.A. 1998/99)

In questi appunti tratto alcuni argomenti che sul libro non sono trattati o sono trattati in modo impreciso, o che per vari motivi potrebbero non essere risultati sufficientemente chiari da quanto spiegato in aula. Ho diviso tali appunti in 4 paragrafi. Qui uso una numerazione a parte per ogni paragrafo (negli enunciati, nelle formule). Si intende che se faccio riferimento per esempio alla formula 3 si intende quella dello stesso paragrafo in cui ci troviamo (salvo avviso contrario).

1. Sistemi differenziali lineari

Studieremo qui i sistemi differenziali lineari (del primo ordine) di n equazioni in n incognite (considereremo nel seguito sottointeso che i sistemi e le equazioni differenziali che tratteremo saranno **sempre in forma normale**). Questi sono sistemi della forma

$$
\begin{aligned}
&y_1'(t) = a_{1,1}(t)y_1(t) + a_{1,2}(t)y_2(t) + ... + a_{1,n}(t)y_n(t) + b_1(t)\\
&y_2'(t) = a_{2,1}(t)y_1(t) + a_{2,2}(t)y_2(t) + ... + a_{2,n}(t)y_n(t) + b_2(t)\\
&.....\\
&.....\\
&.....\\
&y_n'(t) = a_{n,1}(t)y_1(t) + a_{n,2}(t)y_2(t) + ... + a_{n,n}(t)y_n(t) + b_n(t)
\end{aligned}
$$

ove le funzioni $a_{i,j}$ e b_i sono assunte continue da un intervallo aperto I a valori in $\mathbf{R}$. Dalla teoria generale dei sistemi di equazioni segue che le soluzioni di tale sistema sono definite in tutto I, per essere piú precisi, ogni problema di Cauchy relativo a tale sistema ha una (sola) soluzione su tutto I (poiché il sistema è lineare, cf. Giusti teoria pg. 106). Nel seguito per soluzione del sistema intenderemo sempre soluzione in tutto I. Spesso ma non necessariamente I sarà tutto $\mathbf{R}$. Possiamo riscrivere il sistema precedente in forma piú compatta come

$$y_i'(t) = \sum_{j=1}^{n} a_{i,j}(t)y_j(t) + b_i(t), \text{ per } i = 1, ..., n$$

o anche in forma matriciale

$$y'(t) = A(t)y(t) + b(t) \tag{1}$$

ove y è il vettore (colonna) $(y_1, ..., y_n)$, b è il vettore (colonna) $(b_1, ..., b_n)$, la matrice A è la matrice $n \times n$ che ha al posto i, j $a_{i,j}$, e il prodotto $A(t)y(t)$ è inteso ovviamente come prodotto di matrici. Qui se abbiamo una matrice (in particolare un vettore) di funzioni derivabili che al posto i, j ha la funzione $c_{i,j}$ chiameremo come ovvio derivata di tale matrice la matrice che al posto i, j ha la funzione $c_{i,j}'$ (in altri termini la derivata di una matrice si definisce derivando i singoli elementi). In particolare $y'(t)$ è il vettore colonna $(y_1'(t), ..., y_n'(t))$. Useremo spesso la forma matriciale perché piú agile. Il sistema si dice omogeneo se il vettore b è il vettore nullo, ossia se tutte le funzioni b_i sono nulle. In genere non si riesce a risolvere esplicitamente un sistema di equazioni tipo (1), a parte il caso di una sola equazione cioè $n = 1$, o quando le funzioni $a_{i,j}$ (non necessariamente le b_i) sono costanti, nel qual caso il sistema è detto a coefficienti costanti; però si possono dare

informazioni generali sull'insieme delle soluzioni, e questo è quello che faremo nei prossimi teoremi. Cominciamo col caso del sistema omogeneo.

Teorema 1. *Supponiamo che $b = 0$, cioè il sistema (1) sia omogeneo. Allora*
a) L'insieme V delle soluzioni di (1) costituisce uno spazio vettoriale di dimensione n.
b) Fissato $t_0 \in I$, un numero finito di soluzioni $y_{(1)}, ..., y_{(m)}$ di (1) (non necessariamente $m = n$) sono linearmente indipendenti se e solo se i vettori di $\mathbf{R}^n$ $y_{(1)}(t_0), ..., y_{(m)}(t_0)$ sono linearmente indipendenti.
Dimostrazione. Prima proviamo che V è uno spazio vettoriale. Siano y e z in V e siano $a, b \in \mathbf{R}$. Abbiamo da provare che $ay + bz \in V$. Infatti

$$(ay + bz)'(t) = ay'(t) + bz'(t) = aA(t)y(t) + bA(t)z(t) = A(t)(ay(t) + bz(t))$$

ove nella prima uguaglianza abbiamo usato la linearità della derivata, nella seconda il fatto che y e z sono soluzioni di (1), e nella terza la linearità del prodotto con una fissata matrice. Ora proviamo che fissato $t_0 \in I$ l'applicazione $\phi : V \to \mathbf{R}^n$ definita da

$$\phi(y) = y(t_0)$$

è un'applicazione lineare biiettiva, ossia un isomorfismo di spazi vettoriali. Da questo seguirà subito che V ha dimensione n in quanto isomorfo a $\mathbf{R}^n$ e anche b), poiché b) dice in sostanza che m elementi di V sono linearmente indipendenti se e solo se le loro immagini mediante ϕ lo sono. Per provare che ϕ è lineare e biiettiva, osserviamo che la linearità è ovvia in quanto

$$\phi(ay + bz) = ay(t_0) + bz(t_0) = a\phi(y) + b\phi(z)$$

se $y, z \in V$, $a, b \in \mathbf{R}$. Inoltre ϕ è biiettiva se e solo se per ogni $u \in \mathbf{R}^n$ esiste unica $y \in V$ con $\phi(y) = u$, in altri termini per ogni $u \in \mathbf{R}^n$ esiste unica soluzione y di (1) tale che $y(t_0) = u$, ma questo enunciato è vero per il teorema di esistenza e unicità. ∎

Nell'enunciare b) ho preferito usare la notazione $y_{(i)}$ invece di y_i perché quest'ultima potrebbe sembrare la componente iesima di un vettore y. Notiamo che b) dice in sostanza che m soluzioni sono linearmente indipendenti se e solo se lo sono in un punto arbitrario di I. Questo fatto sarà sottointeso nel seguito. Consideriamo ora il sistema (1) non necessariamente omogeneo. Allora il corrispondente sistema omogeneo

$$y'(t) = A(t)y(t) \tag{2}$$

cioè quello ottenuto da (1), "togliendo" il termine b è chiamato sistema omogeneo associato a (1). Proviamo il seguente

Teorema 2. *Data una soluzione $\overline{y}$ di (1), allora una funzione y da I in $\mathbf{R}^n$ è soluzione di (1) se e solo se $y - \overline{y}$ è soluzione di (2).*
Dimostrazione. Tenuto conto che $\overline{y}$ è soluzione di (1), y è soluzione di (1) se e solo se

$$y'(t) = A(t)y(t) + b(t) \iff y'(t) - \overline{y}'(t) = A(t)y(t) + b(t) - (A(t)\overline{y}(t) + b(t))$$
$$\iff (y - \overline{y})'(t) = A(t)(y(t) - \overline{y}(t))$$

e il teorema è provato. ∎

Il teorema precedente dice in sostanza che una volta conosciuta una particolare soluzione $\overline{y}$ di (1) tutte le soluzioni si ottengono sommando a $\overline{y}$ tutte le soluzioni del sistema omogeneo associato. Supponiamo ora di volere risolvere (1) e di conoscere tutte le soluzioni dell'omogeneo associato. Per quanto visto ora basta trovare *una* soluzione di (1). Questa si può trovare usando quello che si chiama metodo della variazione delle costanti. In base alle considerazione fatte finora si capisce che il limite di questo metodo consiste nel fatto che dobbiamo conoscere le soluzioni dell'omogeneo associato, cosa che succede in genere solo quando il sistema è a coefficienti costanti o in altri casi particolari. Però a priori il metodo si può usare anche in altri casi, se per qualche motivo si conoscono ugualmente le soluzioni dell'omogeneo associato. Per Teorema 1 le soluzioni di (2) sono uno spazio vettoriale di dimensione n. Possiamo perciò trovarne n linearmente indipendenti $u_{(1)}, ..., u_{(n)}$. Ovviamente ognuna di queste soluzioni sarà un vettore di n componenti.
Chiamiamo ora U la matrice che ha come colonna jesima il vettore colonna $u_{(j)}$, ossia che al posto i, j ha $U_{i,j} = (u_{(j)})_i$ (la componenente iesima del vettore $u_{(j)}$). Dal fatto che ogni funzione vettoriale $u_{(j)}$ soddisfa (2) segue subito con semplice calcolo che la matrice U verifica l'equazione analoga

$$U'(t) = A(t)U(t). \tag{3}$$

Cerchiamo una soluzione di (1) della forma

$$\tilde{y}(t) = \sum_{i=1}^{n} c_i(t) u_{(i)}(t), \tag{4}$$

cioè cerchiamo una soluzione come "combinazione a coefficienti variabili" di $u_{(1)}, ..., u_{(n)}$. La (4) si può riscrivere in forma matriciale come

$$\tilde{y}(t) = U(t)c(t). \tag{5}$$

Infatti quest'ultima espressione equivale a

$$\tilde{y}_j(t) = \sum_{i=1}^{n} u_{j,i}(t)c_i(t) = \sum_{i=1}^{n} (u_{(i)})_j(t)c_i(t) \quad j = 1, ..., n$$

che è (4). Per andare avanti notiamo che la derivata del prodotto di due matrici derivabili B e C si calcola $(BC)' = BC' + B'C$. Questo si verifica facilmente; infatti BC al posto i, j ha $\sum_k b_{i,k}c_{k,j}$ che ha come derivata $\sum_k b'_{i,k}c_{k,j} + \sum_k b_{i,k}c'_{k,j}$ (ho qui omesso i limiti tra cui varia l'indice k per agilità di notazione; inoltre ho indicato con $b_{i,j}$ (risp. $c_{i,j}$) l'elemento di posto i, j della matrice B (risp. C)). Da (5), tenuto conto di (3), si ha

$$\tilde{y}'(t) = U'(t)c(t) + U(t)c'(t) = A(t)U(t)c(t) + U(t)c'(t) = A(t)\tilde{y}(t) + U(t)c'(t).$$

Quindi $\tilde{y}$ risolve (1) se e solo se si ha

$$U(t)c'(t) = b(t). \tag{6}$$

Poiché le soluzioni $u_{(1)}, ..., u_{(n)}$ erano assunte linearmente indipendenti, per ogni t i vettori $u_{(1)}(t), ..., u_{(n)}(t)$ sono linearmente indipendenti, quindi la matrice $U(t)$ è invertibile e da (6) si ricava $c'(t) = U(t)^{-1}b(t)$. A questo punto, trovati $c_i'(t)$ si ricavano c_i per integrazione, e da (4) si ricava la soluzione $\tilde{y}$ di (1). Per maggiore chiarezza riscriviamo (6) in forma di sistema esplicito, cioè

$$\sum_{j=1}^{n}(u_{(j)})_i(t)c_j'(t) = b_i(t) \quad i = 1, ..., n. \tag{7}$$

Vediamo ora di applicare le considerazioni svolte al caso di equazioni differenziali lineari di ordine n. Queste sono equazioni della forma

$$y^{(n)}(t) = a_{n-1}(t)y^{(n-1)}(t) + ... + a_1(t)y'(t) + a_0(t)y(t) + b(t). \tag{8}$$

ove ancora le funzioni a_i e b sono assunte continue da un intervallo aperto I a valori in $\mathbf{R}$. L'equazione (8) è detta omogenea se $b = 0$. L'equazione

$$y^{(n)}(t) = a_{n-1}(t)y^{(n-1)}(t) + ... + a_1(t)y'(t) + a_0(t)y(t) \tag{9}$$

viene chiamata omogenea associata a (8). Ora come noto dalla teoria, ogni equazione di ordine n si "trasforma" in un sistema del primo ordine, nel senso che le soluzioni del sistema sono tutte e sole le funzioni vettoriali $(y, y', ..., y^{(n-1)})$ ove y è una soluzione dell'equazione data (cioè son tutti e soli i vettori z con $z_i = y^{(i-1)}$ con y soluzione dell'equazione), e nel nostro caso il sistema associato a (8) è dato da

$y_1'(t) = y_2(t)$
.....
.....
$y_{n-1}'(t) = y_n(t)$
$y_n'(t) = a_{n-1}(t)y_n(t) + ... + a_1(t)y_2(t) + a_0(t)y_1(t) + b(t)$

mentre ovviamente quello associato a (9) è dato da

$y_1'(t) = y_2(t)$
.....
.....
$y_{n-1}'(t) = y_n(t)$
$y_n'(t) = a_{n-1}(t)y_n(t) + ... + a_1(t)y_2(t) + a_0(t)y_1(t)$

(ho il dubbio di avere scritto male gli indici durante la spiegazione in aula) Notiamo anche che m soluzioni dell'equazione omogenea (9) (y_1,...,y_m) sono linearmente indipendenti se e solo se lo sono le corrispondenti soluzioni del sistema associato (che sono, come visto, la n-upla composta dalla funzione e dalle sue prime $n-1$ derivate). Per vedere questo, basta notare che una combinazione lineare delle m funzioni cioè

$$\sum_{i=1}^{m} c_i y_i \tag{10}$$

è nulla se e solo se la stessa combinazione lineare delle corrispondenti soluzioni del sistema associato è nulla poiché questo equivale a dire

$$\sum_{i=1}^{m} c_i y_i^{(j)} = 0 \quad \text{per } j = 0, 1, ..., n-1 \tag{11}$$

e (10) è un caso particolare di (11) ($j = 0$), mentre (11) segue da (10) derivando j volte. Noi potremo ora risolvere l'equazione (8) se sapremo risolvere (9) (e questo è possibile quando l'equazione è a coefficienti costanti, come vedremo in dettaglio nel prossimo paragrafo) in questo modo: Dalle soluzioni di (9), noi ricaviamo come detto quelle del sistema associato a (9) che come si vede subito è il sistema omogeneo associato al sistema associato a (8). Perciò col metodo della variazione delle costanti possiamo risolvere il sistema associato a (8) e quindi l'equazione (8). In concreto siano date n soluzioni linearmente indipendenti di (9) y_1,...,y_n. Consideriamo le corrispondenti n soluzioni u_1,...., u_n (linearmente indipendenti) del sistema associato a (9), e cerchiamo una soluzione del sistema associato a (8) data da $u(t) = \sum_{i=1}^{n} c_i(t)u_i(t)$. Ritornando all'equazione (8) avremo che u è della forma $(y, y', ..., y^{(n-1)})$ ove y è una soluzione di (8) e si deve avere (per $j = 0, 1, ..., n-1$)

$$y^{(j)}(t) = \sum_{i=1}^{n} c_i(t) y_i^{(j)}(t)$$

e in particolare

$$y(t) = \sum_{i=1}^{n} c_i(t) y_i(t), \tag{12}$$

determinando cosí la nostra soluzione y di (8). Per determinare le funzioni c_i bisogna risolvere il sistema (7) per il sistema differenziale associato a (8), cioè il sistema

$$\sum_{i=1}^{n} c_i'(t) y_i^{(j)}(t) = \begin{cases} 0 & \text{se } j < n-1 \\ b(t) & \text{se } \quad j = n-1 \end{cases} \tag{13}$$

al variare di $j = 0, 1, ..., n-1$ (notare che il "b" di (1) corrisponde nel sistema associato a (8) al vettore $(0, ..., 0, b)$). Voglio ora dare una spiegazione piú diretta, ossia senza passare al sistema associato a (8), di tale procedimento. Supponiamo dunque che y sia data da (12) con c_i che soddisfano (13). Da (12) derivando otteniamo

$$y' = \sum c_i y_i' + \sum c_i' y_i = \sum c_i y_i'$$

ove la seconda uguaglianza segue dalla prima equazione del sistema (13) (se $n > 1$). Derivando questa otteniamo

$$y'' = \sum c_i y_i'' + \sum c_i' y_i' = \sum c_i y_i''$$

ove la seconda uguaglianza segue dalla seconda equazione di (13) (se $n > 2$) e man mano derivando ogni riga dalla precedente e usando le equazioni di (13) otteniamo

$$y^{(j)} = \sum c_i y_i^{(j)} + \sum c_i' y_i^{(j-1)} = \sum c_i y_i^{(j)}$$

per $j < n$ e

$$y^{(n)} = \sum c_i y_i^{(n)} + \sum c_i' y_i^{(n-1)} = \sum c_i y_i^{(n)} + b.$$

(è ovvio che per formalizzare questo ragionamento bisognerebbe usare l'induzione sull'indice di derivazione). Da queste formule, tenuto conto che $y_1, ..., y_n$ sono soluzioni dell'omogenea associata (9), si deduce

$$y^{(n)} = (\sum c_i y_i^{(n)}) + b = \big(\sum c_i (a_{n-1} y_i^{(n-1)} + ... + a_1 y_i' + a_0 y_i)\big) + b$$

$$= a_{n-1}(\sum c_i y_i^{(n-1)}) + ... + a_1(\sum c_i y_i') + a_0(\sum c_i y_i) + b$$

$$a_{n-1} y^{(n-1)} + ... + a_1 y' + a_0 y + b$$

e quindi y è davvero una soluzione di (8). Per allegerire la notazione in questo discorso non ho evidenziato la dipendenza da t delle funzioni e ho sottointeso che tutte le sommatorie sono per i che varia da 1 a n.

2. Soluzione delle equazioni lineari omogenee a coefficienti costanti

Per studiare le equazioni a coefficienti costanti è conveniente ambientare le equazioni nell'ambito delle delle funzioni da $\mathbf{R}$ in $\mathbf{C}$. Per fare questo cominciamo a generalizzare al caso di funzioni a valori complessi alcune nozioni usuali per funzioni a valori reali. Nota comunque che lo spazio di partenza è sempre $\mathbf{R}$ (eventualmente privato di un punto). Noi *definiamo* il limite di una funzione $f : \mathbf{R} \setminus \{t_0\} \to \mathbf{C}$ nel punto $t_0 \in \mathbf{R}$ nel seguente modo:

$$f(t) \underset{t \to t_0}{\longrightarrow} z \iff \forall\, \varepsilon > 0\ \exists \delta > 0 |\ 0 < |t - t_0| < \delta \Rightarrow |f(t) - z| < \varepsilon$$

ove ovviamente $|\ |$ denota il modulo di un numero complesso. Nella definizione precedente si intende che z è un numero complesso (non consideriamo qui definizioni di limiti infiniti). Segue subito dalla definizione che $f(t) \underset{t \to t_0}{\longrightarrow} z$ equivale a $(\mathcal{R}f(t), \mathcal{I}f(t)) \underset{t \to t_0}{\longrightarrow} (\mathcal{R}(z), \mathcal{I}(z))$, ove $\mathcal{R}$ indica la parte reale e $\mathcal{I}$ la parte immaginaria; in altri termini, parlando in modo non del tutto preciso, si può "identificare" $\mathbf{C}$ con $\mathbf{R}^2$ e studiare il limite della corrispondente funzione a valori in $\mathbf{R}^2$, e questo semplicemente perché dalle definizioni il modulo in $\mathbf{C}$ in questo modo corrisponde esattamente alla norma in $\mathbf{R}^2$, cioè $|x + iy| = ||(x, y)|| (= \sqrt{x^2 + y^2})$ (qui e nel seguito sarà sottointeso che quando scrivo un numero complesso $x + iy$, salvo avviso contrario x e y sono numeri reali e quindi rispettivamente la parte reale e la parte immaginaria del numero). Da queste considerazioni segue subito che $f(t) \underset{t \to t_0}{\longrightarrow} z \iff \mathcal{R}f(t) \underset{t \to t_0}{\longrightarrow} \mathcal{R}z$ e $\mathcal{I}f(t) \underset{t \to t_0}{\longrightarrow} \mathcal{I}z$ (poichè una funzione a valori in $\mathbf{R}^2$ ha come limite un certo elemento di $\mathbf{R}^2$ se e solo se le componenti della funzione hanno come limite le componenti dell'elemento, in altri termini (f_1, f_2) ha come limite (l_1, l_2) se e solo

se f_1 ha come limite l_1 e f_2 ha come limite l_2; questo lo avevamo visto studiando i limiti di funzioni tra spazi metrici). Si prova allora facilmente la seguente

Proposizione 1. *Fissiamo $t_0 \in \mathbf{R}$. In questa proposizione sottointendiamo che le funzioni f e g considerate sono definite in $\mathbf{R} \setminus \{t_0\}$ a valori in $\mathbf{C}$. Allora*
a) Se $c \in \mathbf{C}$ e f è la funzione identicamente uguale a c si ha $f(t) \underset{t\to t_0}{\longrightarrow} c$.
b) Se $f(t) = t$, allora $f(t) \underset{t\to t_0}{\longrightarrow} t_0$ (notare che f è in realtà a valori in $\mathbf{R}$; ovviamente questo non è un problema).
c) Se $f(t) \underset{t\to t_0}{\longrightarrow} z_1$ e $g(t) \underset{t\to t_0}{\longrightarrow} z_2$, allora $f(t) + g(t) \underset{t\to t_0}{\longrightarrow} z_1 + z_2$ e $f(t)g(t) \underset{t\to t_0}{\longrightarrow} z_1 z_2$.
Dimostrazione. a) e b) seguono subito dalla definizione (e si dimostrano come i corrispondenti enunciati per funzioni a valori in $\mathbf{R}$). Anche la c) si dimostra esattamente come il corrispondente enunciato per funzioni reali. Vediamo ad esempio che $f(t)g(t) \underset{t\to t_0}{\longrightarrow} z_1 z_2$. Occorre premettere l'osservazione che come nel caso di funzioni a valori reali avendo f limite per $t \to t_0$, esiste $\delta_1 > 0$ tale che f è limitata nell'insieme

$$I^*(t_0, \delta_1) = \{t \in \mathbf{R} : 0 < |t - t_0| < \delta_1\},$$

per esempio, si vede subito che prendendo come δ_1 il δ corrispondente a $\varepsilon = 1$ nella definizione di limite si ha $|f(t)| < |z_1| + 1$ se $t \in I^*(t_0, \delta_1)$. Poi scriviamo

$$|f(t)g(t) - z_1 z_2| = |f(t)(g(t) - z_2) + z_2(f(t) - z_1)| \leq$$

$$|f(t)(g(t) - z_2)| + |z_2(f(t) - z_1)| = |f(t)|\,|g(t) - z_2| + |z_2|\,|f(t) - z_1|.$$

Ora, fissato $\varepsilon > 0$ basta prendere $\delta_2 > 0$ tale che se $0 < |t - t_0| < \delta_2$ si ha $|g(t) - z_2| < \frac{\varepsilon}{2(|z_1|+1)}$ e $|f(t) - z_1| < \frac{\varepsilon}{2(|z_2|+1)}$ (e questo è possibile per le ipotesi $f(t) \underset{t\to t_0}{\longrightarrow} z_1$ e $g(t) \underset{t\to t_0}{\longrightarrow} z_2$) perché se $0 < |t - t_0| < \delta (= \min\{\delta_1, \delta_2\})$ si abbia $|f(t)g(t) - z_1 z_2| < \varepsilon$, e questo chiude la dimostrazione. ∎

Diciamo ora che $f : \mathbf{R} \to \mathbf{C}$ è continua in $t_0 \in \mathbf{R}$ se $f(t) \underset{t\to t_0}{\longrightarrow} f(t_0)$ e che f è continua se è continua se è continua in tutti i $t_0 \in \mathbf{R}$. Dalla Prop. 1 segue che le costanti sono continue, che la funzione identica (f definita da $f(t) = t$) è continua, che somma e prodotto di funzioni continue in un certo punto sono continue in quel punto. Perciò i polinomi (a coefficienti complessi), che si ottengono dalle funzioni costanti e dalla funzione identità mediante un numero finito di operazioni di somma e prodotto sono, sono funzioni continue (a rigore si dovrebbe dimostrare l'enunciato per induzione sul grado del polinomio).
Data una funzione $f : \mathbf{R} \to \mathbf{C}$ diciamo che f è derivabile in $t_0 \in \mathbf{R}$ se esiste

$$\lim_{t\to t_0} \frac{f(t) - f(t_0)}{t - t_0}$$

(o equivalentemente $\lim_{h\to 0} \frac{f(t_0+h)-f(t_0)}{h}$). In tal caso tale limite si chiama derivata di f in t_0 (indicato con $f'(t_0)$). Ovviamente diremo che f è derivabile se è derivabile in tutti i punti di $\mathbf{R}$. Notiamo anche che se spezziamo f nella parte reale e in quella immaginaria, cioè

$f = u + iv$ con u e v a valori reali, allora, dati $t_0 \in \mathbf{R}$ e $z = x + iy \in \mathbf{C}$, $\exists\, f'(t_0) = z$ se e solo se $\frac{u(t)+iv(t)-(u(t_0)+iv(t_0))}{t-t_0} \underset{t\to t_0}{\longrightarrow} x + iy$ se e solo se $\frac{u(t)-u(t_0)}{t-t_0} + i\frac{v(t)-v(t_0)}{t-t_0} \underset{t\to t_0}{\longrightarrow} (x + iy)$ se e solo se $\frac{u(t)-u(t_0)}{t-t_0} \underset{t\to t_0}{\longrightarrow} x$ e $\frac{v(t)-v(t_0)}{t-t_0} \underset{t\to t_0}{\longrightarrow} y$, quindi

$$\exists\, f'(t_0) = z \iff \exists\, u'(t_0) = x \text{ e } \exists\, v'(t_0) = y,$$

in altre parole la determinazione della derivata di una funzione si riduce a quella delle due funzioni reali date da parte reale e parte immaginaria di quella funzione (nella precedente catena di equivalenze l'ultima segue dalle considerazioni iniziali in base alle quali il limite di una funzione a valori complessi si "riconduce al limite" della parte reale e della parte immaginaria della funzione). Noto ancora che avremmo potuto dare ovviamente definizioni di limiti continuità e derivata anche per funzioni non necessariamente definite in tutto $\mathbf{R}$ o $\mathbf{R} \setminus \{t_0\}$, ma per semplicità ho preferito considerare solo questo caso perché sufficiente per i nostri scopi. Si dimostra esattamente come nel caso di funzioni reali che una funzione derivabile in un punto è continua in quel punto. Inoltre si dimostra come nel caso reale che valgono le usuali regole di derivazione, cioè ad esempio: la derivata di una costante è 0, la derivata della funzione identità è $= 1$, se f e g sono derivabili in un punto $t_0 \in \mathbf{R}$ allora $f+g$ e fg sono deribabili in t_0 e $(f+g)'(t_0) = f'(t_0)+g'(t_0)$, $(fg)'(t_0) = f(t_0)g'(t_0)+f'(t_0)g(t_0)$. Per esempio accenno a come si dimostra l'ultima formula. Nelle nostre ipotesi si ha

$$\frac{f(t)g(t) - f(t_0)g(t_0)}{t - t_0} = f(t)\frac{g(t) - g(t_0)}{t - t_0} + g(t_0)\frac{f(t) - f(t_0)}{t - t_0}$$

e si conclude che l'espressione tende a $f(t_0)g'(t_0) + g(t_0)f'(t_0)$ per $t \to t_0$ poiché $f(t) \underset{t\to t_0}{\longrightarrow} f(t_0)$ in quanto f, essendo derivabile in t_0, è continua in t_0 come detto prima, e poi si usa Prop. 1. Usando queste considerazioni si vede che i polinomi a coefficienti complessi sono derivabili e di derivano come nel caso reale (si segue esattamente la stessa dimostrazione), cioè la derivata della funzione

$$f(t) = \sum_{k=0}^{n} a_k t^k$$

vale

$$f'(t) = \sum_{k=1}^{n} k a_k t^{k-1}.$$

Si definisce in ovvia analogia col caso reale la derivata di ordine n di una funzione f (denotata con $f^{(n)}$). Segue facilmente dalle considerazioni precedenti (per induzione su n) che la derivata di ordine n è un operatore lineare nel senso che se f e g hanno derivata di ordine n in $t_0 \in \mathbf{R}$ allora, dati $a, b \in \mathbf{C}$ la funzione $af + bg$ ha derivata di ordine n in t_0 data da $af^{(n)}(t_0) + bg^{(n)}(t_0)$. Per proseguire diamo ora la definizione dell'esponenziale complesso, e cioè poniamo

$$e^{x+iy} = e^x(\cos(y) + i\sin(y)).$$

Cerco ora di spiegare perché si usa proprio questa definizione. Ricordiamo che in Analisi I era definito e^x per x reale. Naturalmente uno è libero di dare la definizione che gli piace di piú dell'esponenziale di un numero complesso che non era in precedenza definito e quindi a priori non sappiamo che cosa è. Però in genere in casi come questo sembra opportuno dare una definizione che estenda in un modo naturale quella dell'esponenziale reale. Ora notiamo che se noi consideriamo la serie di Taylor di e^x otteniamo $e^x = \sum_{k=0}^{+\infty} \frac{x^k}{k!}$, che avevamo dimostrato valida per tutti gli x reali, e se al posto di x mettiamo iy nella serie precedente e separiamo parte reale e parte immaginaria otteniamo con semplici calcoli $\cos(y)+i\sin(y)$ (almeno formalmente perchè non abbiamo trattato in modo rigoroso le serie complesse quindi non sappiamo bene che cosa significa tale serie). Questa considerazione rende naturale dare la definizione $e^{iy} = \cos(y) + i\sin(y)$. Sottolineo ancora una volta che questa non è una *dimostrazione* di tale formula perché non possiamo dimostrare qualcosa su un oggetto come e^{iy} che non era stato in precedenza definito, ma spiega perché tale denizione appare ragionevole. Ora, definito e^{iy}, sembra naturale pretendere che anche in campo complesso l'esponenziale della somma sia uguale al prodotto degli esponenziali quindi sembra naturale porre $e^{x+iy} = e^x e^{iy} = e^x(\cos(y) + i\sin(y))$, ottenendo la formula data. Notiamo che, come nel caso reale, l'esponenziale di un numero complesso è sempre $\neq 0$ (questo fatto sarà sottointeso nel seguito). Infatti e^{x+iy} è il prodotto di e^x che, in quanto esponenziale reale è $\neq 0$, e del numero $\cos(y) + i\sin(y)$, che non può essere 0, in quanto il coseno e il seno di un numero reale non si annullano mai simultaneamente. Si verifica anche con semplice calcolo che come nel caso di numeri reali se z_1, z_2 sono numeri complessi allora $e^{z_1+z_2} = e^{z_1}e^{z_2}$. Proviamo ora che dato un numero complesso $z = x + iy$, la funzione

$$f(t) = e^{zt}$$

si deriva secondo la regola dell'esponeziale reale ossia ha derivata uguale a ze^{zt}. Per provare questa formula, basta scrivere $e^{zt} = e^{xt}(\cos(yt) + i\sin(yt))$, e notare che questa funzione è derivabile in quanto sono derivabili sia la parte reale sia la parte immaginaria e la derivata sia ottiene derivando queste ossia $f'(t) = \frac{d}{dt}(e^{xt}\cos(yt)) + i\frac{d}{dt}(e^{xt}\sin(yt)) = xe^{xt}\cos(yt) - ye^{xt}\sin(yt) + i(xe^{xt}\sin(yt) + ye^{xt}\cos(yt))$, e con semplici calcoli si ha

$$f'(t) = (x + iy)e^{xt}(\cos(yt) + i\sin(yt)) = ze^{zt}.$$

Finora abbiamo fatto delle considerazioni introduttive sulle funzioni definite sui reali a valori complessi. Ora passiamo alle equazioni a coefficienti costanti (in generale complessi) omogenee. Per studiare le soluzioni reali di equazioni a coefficienti reali ci conviene introdurre come concetto ausiliario le equazioni a coefficienti complessi e considerarne le soluzioni non solo reali ma anche complesse. Un'equazione lineare omogenea a coefficienti costanti complessi di ordine n è un'espressione della forma

$$\sum_{k=0}^{n} a_k y^{(k)}(t) = 0 \tag{1}$$

ove a_k (i coefficienti) sono numeri complessi e $a_n \neq 0$, (se $a_n = 0$ viene ovviamente un'equazione di ordine minore). Le soluzioni di tale equazione sono le funzioni $y : \mathbf{R} \to \mathbf{C}$ soddisfacenti (1).

Osservazione 2. Osserviamo che *se l'equazione è a coefficienti reali*, allora la funzione complessa $y(t) = u(t) + iv(t)$ ove u e v sono la parte reale e quella immaginaria di y, é soluzione dell'equazione se e solo se u e v lo sono. Infatti

$$\sum_{k=0}^{n} a_k y^{(k)}(t) = 0$$

se e solo se

$$\sum_{k=0}^{n} a_k (u^{(k)}(t) + iv^{(k)}(t)) = 0$$

se e solo se

$$\sum_{k=0}^{n} a_k u^{(k)}(t) + i\big(\sum_{k=0}^{n} a_k v^{(k)}(t)\big) = 0.$$

Poiché le due espressioni in sommatoria sono numeri reali essendo gli a_k reali per ipotesi, l'ultima uguaglianza equivale a dire che le due sommatorie sono entrambe 0, ossia u e v soddisfano (1). ∎

Definiamo ora C^∞ l'insieme delle funzioni da $\mathbf{R}$ in $\mathbf{C}$ che hanno derivate di tutti gli ordini. Si vede subito che C^∞ è uno spazio vettoriale. Dato un polinomio a coefficienti complessi P con

$$P(\lambda) = \sum_{k=0}^{n} a_k \lambda^k \tag{2}$$

ad esso possiamo associare quello che chiamiamo l'operatore differenziale

$$P(D) = \sum_{k=0}^{n} a_k D^k.$$

Per definizione $P(D)$ è l'applicazione da C^∞ in C^∞ data da

$$(P(D)(y))(t) = \sum_{k=0}^{n} a_k y^{(k)}(t) \tag{3}$$

cosí l'equazione (1) si può riscrivere come

$$P(D)(y) = 0. \tag{4}$$

(si intende in questo contesto che la derivata di ordine 0 di una funzione è la funzione stessa, quindi $D^0(y) = y^{(0)} = y$; l'applicazione D^0 coincide quindi con l'identità e la indichiamo con I). Si vede facilmente che in effetti la funzione $P(D)(y)$ definita in (3) è in C^∞, avendo supposto che y è in C^∞, quindi è corretto dire che $P(D)$ è un'applicazione da C^∞ in C^∞,

ed è facile verificare che $P(D)$ è lineare (essendo ogni derivata di ordine n un operatore lineare). In questo modo possiamo scrivere ogni equazione della forma (1) nella forma (4) ove P è il polinomio dato da (2). Tale polinomio, associato all'equazione, è anche chiamato *polinomio caratteristico* dell'equazione.

Osservazione 3. Si vede facilmente che se P_1 e P_2 sono polinomi allora
a) $(P_1 + P_2)(D) = P_1(D) + P_2(D)$.
b) $(P_1P_2)(D) = (P_1(D)) \circ (P_2(D)) = (P_2(D)) \circ (P_1(D))$.

Per verificare le due formule scriviamo i polinomi P_1 e P_2 nella forma $P_1(\lambda) = \sum_{k=0}^{n} a_k\lambda^k$ e $P_2(\lambda) = \sum_{k=0}^{m} b_k\lambda^k$. Notiamo anche (e questo renderà piú agile la dimostrazione di a)) che possiamo supporre $n = m$ ponendo eventualmente $= 0$ i coefficienti dei gradi piú alti di uno dei due polinomi (per esempio se P_1 ha grado 4 e P_2 ha grado 2, scriviamo P_2 come $0 \cdot \lambda^4 + 0 \cdot \lambda^3 + b_2\lambda^2 + b_1\lambda^1 + b_0$. La formula in a) significa per definizione che per ogni $y \in C^\infty$ si ha $((P_1 + P_2)(D))(y) = P_1(D)(y) + P_2(D)(y)$. Ma questo si verifica facilmente poiché supponendo $n = m$ come detto sopra si ha $(P_1 + P_2)(\lambda) = \sum_{k=0}^{n} (a_k + b_k)\lambda^k$. Perciò

$$((P_1+P_2)(D))(y) = \sum_{k=0}^{n}(a_k+b_k)y^{(k)} = \sum_{k=0}^{n} a_k y^{(k)} \ + \sum_{k=0}^{n} b_k y^{(k)} = (P_1(D))(y) + (P_2(D))(y)$$

e questo prova a). Per provare b) intanto cerchiamo di capire che cosa significa l'espressione $(P_1(D)) \circ (P_2(D))$. Questa è semplicemente da interpretare come la composizione dei due operatori differenziali $P_1(D)$ e $P_2(D)$. Questi sono applicazioni da C^∞ in C^∞, e quindi la loro composizione è ancora un' applicazione da C^∞ in C^∞. Osserviamo anche che per provare b) basta provare la prima uguaglianza in b) poiché l'uguaglianza tra il primo e il terzo membro di b) segue da quella tra il primo e il secondo scambiando P_1 e P_2 e tenendo conto che ovviamente $P_1P_2 = P_2P_1$. Proviamo dunque tale prima uguaglianza. Prima osserviamo che

$$(P_1P_2)(\lambda) = P_1(\lambda)P_2(\lambda) = \big(\sum_{k=0}^{n} a_k\lambda^k\big)\big(\sum_{h=0}^{m} b_h\lambda^h\big) =$$

$$\sum_{h=0}^{m}\big(b_h\lambda^h \sum_{k=0}^{n} a_k\lambda^k\big) = \sum_{h=0}^{m}\sum_{k=0}^{n} a_k b_h \lambda^{k+h}.$$

Nell'uguaglianza tra la prima e la seconda riga si è portata dentro la sommatoria in h la sommatoria in k che è costante in quanto non dipende da h (cioè si è usata la formula $c\sum d_h = \sum(cd_h)$ con $c =$ la sommatoria in k), nella successiva uguaglianza si è portato il termine $b_h\lambda^h$ dentro la sommatoria in k. Usando questa formula si ha

$$\big((P_1P_2)(D)\big)(y) = \sum_{h=0}^{m}\sum_{k=0}^{n} a_k b_h y^{(k+h)}.$$

D'altra parte si ha

$$\big(P_1(D)\big)\circ\big(P_2(D)\big)(y) = P_1(D)(P_2(D)(y)) = P_1(D)\big(\sum_{h=0}^{m} b_h y^{(h)}\big)$$

$$= \sum_{h=0}^{m} b_h\big(P_1(D)(y^{(h)})\big) = \sum_{h=0}^{m} b_h\big(\sum_{k=0}^{n} a_k y^{k+h}\big) = \sum_{h=0}^{m}\sum_{k=0}^{n} a_k b_h y^{(k+h)}$$

(nell'uguaglianza tra la prima e la seconda riga si è usato il fatto che $P_1(D)$ è un operatore lineare) e quindi b) è provata. ∎

Notiamo che b) della precedente osservazione dice in sostanza che se ho un polinomio P prodotto di due polinomi, il corrispondente operatore differenziale $P(D)$ è la composizione (in un ordine arbitrario) dei corrispondenti operatore differenziali. Con un ovvio passaggio induttivo se ho un polinomio P prodotto di un numero finito di polinomi, il corrispondente operatore differenziale $P(D)$ è la composizione (in un ordine arbitrario) dei corrispondenti operatore differenziali. Supponiamo ora nel seguito che P sia un polinomio di grado n scritto nella forma (2) con $a_n = 1$ (per i nostri scopi supporre $a_n = 1$ non è una restrizione reale poiché le equazioni che consideriamo possono ricondursi a equazioni con $a_n = 1$ dividendo per a_n), e supponiamo d'ora in poi che i *coefficienti* di P siano *reali*. Per il teorema fondamentale dell'algebra si ha

$$P(\lambda) = (\lambda - \lambda_1)^{\mu_1} \cdot ... \cdot (\lambda - \lambda_m)^{\mu_m}$$

ove $\lambda_k, k = 1, ..., m$ sono le radici (complesse) di P distinte tra di loro e μ_k è la molteplicità di λ_k. Ricordo anche che essendo P a coefficienti reali, se λ è una radice di P, anche il suo coniugato $\overline{\lambda}$ è radice di P e con la stessa molteplicità di λ. Questo fatto sarà sottointeso nel seguito. Da b) dell'Osservazione 3 si deduce che

$$P(D) = (D - \lambda_1 I)^{\mu_1} \circ ... \circ (D - \lambda_m I)^{\mu_m} \tag{5}$$

ove $(D - \lambda I)^{\mu}$ con μ intero positivo significa $(D - \lambda I)$ composto con se stesso μ volte. In questa formula si vede l'utilità di considerare le equazioni in campo complesso poiché in campo reale non vale il teorema fondamentale dell'algebra, quindi non si riesce a ottenere una formula tipo la (5). Comunque, come accennato in precedenza, da (b) di Osserv.3 si vede che nella formula (5) in realtà possiamo cambiare a piacimento l'ordine della composizione e mettere come ultimo termine un qualunque $(D - \lambda_k I)^{\mu_k}$, e non necessariamente $(D - \lambda_m I)^{\mu_m}$. Perciò se y soddisfa

$$(D - \lambda I)^{\mu}(y) = 0 \tag{6}$$

per $\lambda = \lambda_k, \mu = \mu_k$, allora y soddisfa l'equazione (4), in quanto se ad esempio $k = m$, da $(D - \lambda_m I)^{\mu_m}(y) = 0$, usando (5) segue $P(D)(y) = 0$, dato che tutti gli operatori $(D - \lambda_1 I)^{\mu_1}$, $(D - \lambda_2 I)^{\mu_2}$, ecc. essendo lineari mandano la funzione 0 in se stessa. Noi ora cerchiamo delle soluzioni di (6) per ogni $k = 1, ..., m$ e poi proveremo che tutte queste funzioni che per quanto detto sono soluzioni di (4), sono in realtà una base dello spazio delle soluzioni di (4), e quindi le soluzioni di (4) saranno esattamente le combinazioni lineari delle funzioni trovate. Piú precisamente per ogni k troveremo μ_k soluzioni complesse di

(6) corrisponsenti a $\lambda = \lambda_k$, $\mu = \mu_k$. Mettendo insieme tutte queste soluzioni al variare di k, otterremo n soluzioni complesse di (4). Prendendo parte reale e parte immaginaria otterremo n soluzioni reali di (4). Proveremo che queste sono linearmente indipendenti e quindi formano una base dello spazio delle soluzioni di (4) che, come noto, ha dimensione n.

Lemma 4. *Le funzioni $t^h e^{\lambda t}$ per $h = 0, 1, ..., \mu - 1$, sono soluzioni (complesse) di (6).*

Dimostrazione. Noi potremmo verificare direttamente che le funzioni date sono soluzioni di (6), ma preferisco far vedere come si arriva a trovare queste funzioni. In altri termini *cerco* le soluzioni di (6) e ottengo le funzioni dette e non mi limito a fare una verifica che non spiega come sono arrivato a pensare proprio quelle funzioni. Se $\mu = 1$ (6) diventa $(D - \lambda I)(y) = 0$, ossia

$$y' = \lambda y \tag{7}$$

e una soluzione di questa è $y(t) = e^{\lambda t}$. Infatti avevamo provato poco dopo aver introdotto l'esponenziale complesso che la funzione data soddisfa (7) e abbiamo la tesi se $\mu = 1$. Supponiamo ora $\mu = 2$. Quindi l'equazione (6) diventa

$$(D - \lambda I)^2(y)\big(= (D - \lambda I)((D - \lambda I)(y))\big) = 0. \tag{8}$$

Ovviamente ogni soluzione di (7), (in particolare $y(t) = e^{\lambda t}$), è anche soluzione di (8), in quanto se y risolve (7) allora $(D - \lambda I)((D - \lambda I)y) = (D - \lambda I)(0) = 0$. D'altra parte, per il caso precedente, ogni y che risolve

$$(D - \lambda I)(y)(t)(= y'(t) - \lambda y(t)) = e^{\lambda t} \tag{9}$$

risolve (8). Per risolvere (9) usiamo il metodo della varizione delle costanti. Si potrebbe porre l'obiezione che tale metodo l'avevamo visto per equazioni reali e questa è un'equazione complessa. Comunque il metodo si dimostra valido nel campo complesso esattamente nello stesso modo poiché tutti i procedimenti usati nel caso reale (essenzialmente la regola di derivazione del prodotto di funzioni) sappiamo che valgono ancora in campo complesso. Per usare il metodo della variazione delle costanti notiamo che l'equazione omogenea associata a (9) è (7). Una soluzione di (7) è data da $e^{\lambda t}$. Quindi cerco una soluzione di (9) della forma $c(t)e^{\lambda t}$ con $c(t)$ da determinarsi, mediante l'equazione $c'(t)e^{\lambda t} = e^{\lambda t}$, che equivale a $c'(t) = 1$, quindi la funzione $te^{\lambda t}$ è soluzione di (9). In conclusione la funzione $t^h e^{\lambda t}$ è soluzione di (8) per $h = 0, 1$, che è quello che si doveva provare. Noi abbiamo provato la tesi nel caso $\mu = 2$ usando il fatto che la tesi valeva se $\mu = 1$. Usando questo stesso metodo possiamo trovare le soluzioni cercate per μ via via crescenti, verificando che, note le soluzioni di (6) per un certo μ, si trovano quelle per $\mu + 1$, aggiungendo a quelle trovate per μ la funzione $t^\mu e^{\lambda t}$. Formalizziamo questo procedimento procedendo per induzione. Supponiamo dunque che la tesi valga per μ e consideriamo l'equazione

$$(D - \lambda I)^{\mu+1}(y) = 0. \tag{10}$$

Proviamo che le funzioni $t^h e^{\lambda t}$ per $h = 0, 1, ..., \mu$ sono soluzioni di (10). Questo completerà il passo induttivo. Notiamo che se $h = 0, 1, ..., \mu - 1$ allora la funzione data è soluzione di (6) per l'ipotesi induttiva e quindi anche di (10) in quanto

$$(D-\lambda I)^{\mu+1}(y) = (D-\lambda I)\big((D-\lambda I)^{\mu}(y)\big) = (D-\lambda I)(0) = 0.$$

Rimane da provare che la funzione $t^{\mu}e^{\lambda t}$ è soluzione di (10). Riscriviamo (10) come

$$(D-\lambda I)^{\mu}((D-\lambda I)(y)) = (D-\lambda I)^{\mu}(y' - \lambda y) = 0.$$

Per l'ipotesi induttiva quindi (10) è soddisfatta se (non solo se)

$$y'(t) - \lambda y(t) = t^{\mu-1}e^{\lambda t}. \tag{11}$$

Cerco col metodo della variazione delle costanti una soluzione di (11) della forma $c(t)e^{\lambda t}$, e procedendo come nel caso precedente ottengo $c'(t) = t^{\mu-1}$ quindi ho una soluzione con $c(t) = \frac{1}{\mu}t^{\mu}$. Tenuto conto che per la linearità i multipli di soluzioni sono soluzioni si avrà che $t^{\mu}e^{\lambda t}$ è soluzione di (10), e questo conclude la dimostrazione. In realtà usando lo stesso procedimento in modo solo leggermente piú raffinato si vede che tutte le soluzioni di (6) sono le combinazioni lineari delle funzioni trovate, ma tralasciamo questo fatto perché non necessario nel seguito. ∎

Per concludere la dimostrazione ci servono ancora due lemmi.

Lemma 5. *Sia f la funzione da $\mathbf{R}$ in $\mathbf{C}$ definita da $f(t) = e^{\lambda t}P(t)$ ove $\lambda \in \mathbf{C} \setminus \{0\}$, è P e un polinomio. Allora si ha*

$$f^{(n)}(t) = P_n(t)e^{\lambda t}$$

ove P_n è un opportuno polinomio avente lo stesso grado di P. In particolare se P non è il polinomio 0 anche P_n non è il polinomio 0.

Dimostrazione. L'ultimo enunciato segue dai precedenti in quanto 0 è l'unico polinomio avente grado -1. Per provare la tesi basterà (per un ovvio procedimento induttivo) verificarla per $n = 1$. Si ha

$$f'(t) = e^{\lambda t}(\lambda P(t) + P'(t)).$$

Dobbiamo provare che il polinomio $\lambda P + P'$ ha lo stesso grado di P. Questo è ovvio se il grado di P è -1. In caso contrario il grado di P' è sempre $<$ di quello di P, perciò essendo per ipotesi $\lambda \neq 0$ e quindi il grado di λP uguale a quello di P, si deduce la tesi (il grado della somma di due polinomi di gradi diversi è uguale al massimo dei gradi dei due polinomi). ∎

Lemma 6. *Siano dati m numeri complessi $\mu_1, ..., \mu_m$ diversi tra di loro, e m polinomi a coefficienti complessi $P_1, ..., P_m$. Allora l'uguaglianza*

$$\sum_{k=1}^{m} P_k(t)e^{\mu_k t} = 0 \tag{12}$$

vale per tutti i t reali solo nel caso che tutti i polinomi P_k siano nulli.

Dimostrazione. Procediamo per induzione su m. Se $m = 1$ (12) si riduce a $P_1(t)e^{\mu_1 t} = 0$, e poiché l'esponenziale è sempre $\neq 0$, segue $P_1 = 0$. Supponiamo ora che la tesi sia vera per m e proviamola per $m+1$. Supponiamo quindi di avere

$$\sum_{k=1}^{m+1} P_k(t)e^{\mu_k t} = 0.$$

Riscriviamo questa uguaglianza come

$$-P_{m+1}(t) = \sum_{k=1}^{m} P_k(t)e^{(\mu_k - \mu_{m+1})t}, \tag{13}$$

e deriviamo un numero sufficiente di volte perché a sinistra si ottenga 0 (precisamente bisogna derivare una volta piú del grado di P_{m+1}). Ora dall'ipotesi che i μ_k sono tutti diversi tra loro si ha che i numeri $\mu_k - \mu_{m+1}$ sono tutti diversi tra loro e diversi da 0. Perciò per il Lemma 5 si otterrà

$$0 = \sum_{k=1}^{m} \tilde{P}_k(t)e^{(\mu_k - \mu_{m+1})t}$$

ove $\tilde{P}_k = 0$ se e solo se $P_k = 0$. D'altra parte per l'ipotesi induttiva si deve avere $\tilde{P}_k = 0$, e quindi $P_k = 0$ per ogni $k = 1, ..., m$. Rimane da provare che $P_{m+1} = 0$ ma questo segue subito da (13). ∎

Scriviamo ora le radici dell'equazione (2), che è l'equazione caratteristica della nostra equazione differenziale (4), separando quelle reali da quelle non reali e accoppiando quelle complesse coniugate.

$$\lambda_k \in \mathbf{R}, k = 1, ..., l, \quad \lambda_k = \alpha_k + i\beta_k, \lambda_{k+s} = \alpha_k - i\beta_k, k = l+1, ..., l+s$$

ove ovviamente α_k e β_k sono numeri reali e $\beta_k \neq 0$, e inoltre $l + 2s = m$. Ricordo che μ_k è la molteplicità di λ_k e chiaramente $\mu_k = \mu_{k+s}$ per $k = l+1, ..., l+s$, poiché le radici λ_k e λ_{k+s} sono coniugate. Grazie alle considerazioni svolte finora le funzioni

$$t^h e^{\lambda_k t}, \ h = 0, ..., \mu_k - 1, \ k = 1, ..., m \tag{14}$$

sono soluzioni complesse di (4), e sono chiaramente n in quanto il loro numero è dato dalla somma di tutte le molteplicità delle radici, che è uguale al grado del polinomio caratteristico, cioè appunto n. Inoltre associamo a queste soluzioni delle soluzioni reali ottenute lasciando le stesse funzioni in corrispondenza dei λ_k reali, ossia per $k = 1, ..., l$, e sostituendo alle soluzioni

$$t^h e^{(\alpha_k + i\beta_k)t}, t^h e^{(\alpha_k - i\beta_k)t} \quad h = 0, ..., \mu_k - 1, \ k = l+1, ..., l+s$$

le rispettive parti reali e immaginarie, che vengono precisamente

$$t^h e^{\alpha_k t}\cos(\beta_k t) \quad \text{e} \quad t^h e^{\alpha_k t}\sin(\beta_k t).$$

Si vede facilmente che in questo modo abbiamo n funzioni reali poiché abbiamo sostituito alle due funzioni complesse $t^h e^{(\alpha_k + i\beta_k)t}$ e $t^h e^{(\alpha_k - i\beta_k)t}$ le due funzioni reali $t^h e^{\alpha_k t}\cos(\beta_k t)$

e $t^h e^{\alpha_k t}\sin(\beta_k t)$. Sia E l'insieme delle funzioni cosí ottenute. Arriviamo dunque a provare il teorema conclusivo.

Teorema 7. *Le soluzioni reali di (4) sono tutte e sole le combinazioni lineari delle funzioni in E.*

Dimostrazione. Come già accennato in precedenza, poiché l'equazione è di ordine n e quindi l'insieme delle soluzioni di (4) è uno spazio vettoriale di dimensione n, e E è un insieme di n soluzioni, basta provare che queste sono linearmente indipendenti. Per ottenere questo, prima proviamo che le funzioni date da (14) sono linearmente indipendenti su $\mathbf{C}$ cioè l'unica loro combinazione lineare a coefficienti complessi è quella con tutti i coefficienti nulli. Supponiamo dunque che si abbia

$$\sum_{k=1}^{m}\sum_{h=0}^{\mu_k-1} c_{h,k}t^h e^{\lambda_k t} = 0$$

con $c_{h,k} \in \mathbf{C}$ e proviamo che si deve avere che $c_{h,k} = 0$ per ogni h, k. Ora riscriviamo l'espressione precedente come

$$\sum_{k=1}^{m}\Big(\sum_{h=0}^{\mu_k-1} c_{h,k}t^h\Big)e^{\lambda_k t} = 0.$$

Quindi per il lemma 6 i polinomi $P_k(t) = \sum_{h=0}^{\mu_k-1} c_{h,k}t^h$ devono essere tutti nulli e quindi tutti i $c_{h,k}$ devono essere tutti nulli, come volevasi. Ora proviamo che le funzioni in E sono linearmente indipendenti. Per ottenere questo l'idea è rimpiazzare una combinazione lineare di funzioni di E con una di funzioni di (14) in particolare notando che dati numeri u e v ogni loro combinazione lineare si può esprimere come una combinazione lineare di $u+iv$ e $u-iv$ e fatti i calcoli, $au+bv = \frac{a-ib}{2}(u+iv) + \frac{a+ib}{2}(u-iv)$, nel nostro caso sarà $u = t^h e^{\alpha_k t}\cos(\beta_k t)$, $v = t^h e^{\alpha_k t}\sin(\beta_k t)$. Quindi se

$$\sum_{k=1}^{l}\sum_{h=1}^{\mu_k-1} c_{h,k}t^h e^{\lambda_k t} + \sum_{k=l+1}^{l+s}\sum_{h=0}^{\mu_k-1} a_{h,k}t^h e^{\alpha_k t}\cos(\beta_k t) + \sum_{k=l+1}^{l+s}\sum_{h=0}^{\mu_k-1} b_{h,k}t^h e^{\alpha_k t}\sin(\beta_k t) = 0 \tag{15}$$

con i coefficienti $c_{h,k}, a_{h,k}, b_{h,k}$ reali, riscriviamo l'espressione precedente come

$$\sum_{k=1}^{l}\sum_{h=1}^{\mu_k-1} c_{h,k}t^h e^{\lambda_k t} + \sum_{k=l+1}^{l+s}\sum_{h=0}^{\mu_k-1} \frac{a_{h,k}-ib_{h,k}}{2}t^h e^{\lambda_k t} + \sum_{k=l+1}^{l+s}\sum_{h=0}^{\mu_k-1} \frac{a_{h,k}+ib_{h,k}}{2}t^h e^{\lambda_{k+s} t} = 0$$

tenuto conto che

$$t^h e^{\alpha_k t}\cos(\beta_k t) + it^h e^{\alpha_k t}\sin(\beta_k t) = t^h e^{(\alpha_k+i\beta_k)t} = t^h e^{\lambda_k t}$$

e

$$t^h e^{\alpha_k t}\cos(\beta_k t) - it^h e^{\alpha_k t}\sin(\beta_k t) = t^h e^{(\alpha_k - i\beta_k)t} = t^h e^{\lambda_{k+s} t}$$

se $k = l+1, ..., l+s$. Dal fatto che abbiamo appena visto che le funzioni in (14) sono linearmente indipendenti su $\mathbf{C}$ segue che per $h = 0, ..., \mu_k - 1$, $c_{h,k} = 0$ per $k = 1, ..., l$, e $\frac{a_{h,k}+ib_{h,k}}{2} = \frac{a_{h,k}-ib_{h,k}}{2} = 0$ per $k = l+1, ..., l+s$, ma questo implica facilmente che tutti gli $a_{h,k}$ e tutti i $b_{h,k}$ sono pure nulli e abbiamo provato che tutti i coefficienti della nostra combinazione lineare (formula (15)) sono nulli. ∎

3. Precisazioni su argomenti vari

In questo paragrafo precisiamo alcuni punti che sul libro o a lezione non erano detti in modo preciso. Nel seguito G.T. vorrà dire Giusti Teoria.

Nel teorema 2.1 del cap. 3 del G.T. (pg. 86), si dice che il problema di Cauchy ha un'unica soluzione in $I(t_0, r_0)$, ma in realtà analizzando la dimostrazione si vede che si prova che esiste una e una sola funzione in X che risolve il problema dato. Questo prova ovviamente l'esistenza della soluzione ma non l'unicità perché potrebbe rimanere il dubbio che esista una soluzione del problema di Cauchy non appartenente ad X. Comunque si può completare la dimostrazione provando che ogni soluzione y del problema dato è in realtà in X, ossia per la definizione data di X, che per ogni $t \in I(t_0, r_0)$ si ha

$$|y(t) - u_0| \leq r_2. \tag{1}$$

Proviamo addirittura che la disuguaglianza in (1) è stretta. Per assurdo supponiamo che ciò non accada. Allora almeno uno dei due insiemi

$$E_+ = \{t \in]t_0, t_0 + r_0[: |y(t) - u_0| \geq r_2\}$$

$$E_- = \{t \in]t_0 - r_0, t_0[: |y(t) - u_0| \geq r_2\}$$

è non vuoto. Supponiamo per esempio che $E_+ \neq \emptyset$ e sia $\overline{t} = \inf E_+$. Chiaramente $\overline{t} > t_0$ poiché y vale u_0 in t_0 e quindi per t in tutto un opportuno intorno di t_0 (per continuità) deve valere $|y(t) - u_0| < r_2$. Quindi $t_0 < \overline{t} < t_0 + r_0$. Dalle proprietà dell'inf segue che se $t \in]t_0, \overline{t}[$ si ha $t \notin E_+$ e quindi $|y(t) - u_0| < r_2$ e quindi $(t, y(t)) \in I \times J$ per definizione di I, J, r_0. Segue che per tali t si ha $|f(t, y(t))| \leq M$. Perciò

$$|y(\overline{t}) - u_0| = |\int_{t_0}^{\overline{t}} f(s, y(s))\, ds| \leq \int_{t_0}^{\overline{t}} |f(s, y(s))|\, ds \leq \int_{t_0}^{\overline{t}} M\, ds = M(\overline{t} - t_0) < r_2$$

poiché $M(\overline{t} - t_0) \leq M r_0 < r_2$, ma dalla definizione di $\overline{t}$ segue per continuità che $\overline{t} \in E_+$, ottenendo una contraddizione. Se $E_- \neq \emptyset$ si procede analogamente. ∎

Nel teorema 5.1 cap. 5, pg.189 (G.T.) non è precisato che nella formula (5.1) si intende che $0 \times \infty = 0$. Inoltre la dimostrazione non è completa. Questa funziona bene se gli insiemi E e F considerati sono *limitati* (o piú generalmente di misura finita), ma non nel caso generale. Questo prima di tutto perché la conclusione che si prova lí è che se E e F

sono misurabili allora $E \times F$ ha misura esterna uguale alla misura interna, ma se questo valore non è finito non segue automaticamente che $E \times F$ è misurabile. Inoltre nel passare al limite non si capisce cosa succede se uno dei due insiemi ha misura 0 e l'altro $+\infty$, poiché non posso dire niente sull'eventuale limite del prodotto di misure che tendono a 0 per misure che tendono a $+\infty$. Si rimedia comunque nel seguente modo: Definendo $I_m(0,r)$ come la palla aperta di centro l'origine e raggio r in $\mathbf{R}^m$ si ha che $E \cap I_n(0,h)$ e $F \cap I_k(0,h)$ sono ovviamente *limitati* e misurabili. Quindi possiamo applicare il teorema avendo che il loro prodotto è un sottoinsieme misurabile di $\mathbf{R}^{n+k}$ e si ha

$$m_{n+k}\big((E \cap I_n(0,h)) \times (F \cap I_k(0,h))\big) = m_n(E \cap I_n(0,h))\ m_k(F \cap I_k(0,h)). \qquad (2)$$

Poiché

$$E \times F = \bigcup_{h=1}^{+\infty} \Big((E \cap I_n(0,h)) \times (F \cap I_k(0,h)) \Big)$$

e tale unione è *crescente* in h (che qui faccio variare solo sugli interi positivi), segue che $E \times F$ è misurabile, in quanto unione numerabile di insiemi misurabili e da (2)

$$m_{n+k}(E \times F) = \lim_{h \to +\infty} \Big(m_n(E \cap I_n(0,h)) m_k(F \cap I_k(0,h)) \Big) \qquad (3)$$

e nel secondo membro di quest'ultima uguaglianza il primo fattore tende a $m_n(E)$ e il secondo a $m_k(F)$ proprio per la definizione di misura. Quindi il limite nella formula (3) vale $m_n(E)m_k(F)$ se questo prodotto non è una forma indeterminata cioè del tipo $0 \times \infty$. Se invece per esempio $m_n(E) = 0$ e $m_k(F) = +\infty$ allora $m_n(E \cap I_n(0,h)) \leq m_n(E) = 0$, e quindi il prodotto di cui devo fare il limite nel secondo membro di (3) vale 0 per ogni h. Perciò tale limite vale 0 e la dimostrazione è completata. ∎

La formula

$$\mathcal{F} = \mathcal{F}_0 \cup (\mathbf{R}^n \times]-\infty, 0]) \qquad (4)$$

pg. 205, circa metà pagina G.T., non è corretta poiché se $f(t) = 0$ il punto $(t, 0)$ è nel secondo membro dell'uguaglianza (4) ma non nel primo. Si rimedia facilmente sostituendola con

$$\mathcal{F} = \Big(\mathcal{F}_0 \cup (\mathbf{R}^n \times]-\infty, 0]) \Big) \setminus \Big((f^{-1}(0)) \times \{0\} \Big)$$

che come è facile verificare è valida. L'insieme sottratto, in quanto sottoinsieme dell'insieme $\mathbf{R}^n \times \{0\}$, che ha misura 0 per il teorema sulla misura di un prodotto cartesiano, è misurabile (e ha misura 0). Perciò $\mathcal{F}$ è lo stesso misurabile e si continua come nel testo. ∎

La formula 5.15 cap. 6 (G.T. pg.230) non è del tutto corretta nel senso che vale solo se si suppone $\alpha(x) < \beta(x)$ in $]a, b[$ (inoltre bisogna precisare che le funzioni α e β devono essere continue, o piú generalmente misurabili, su $]a, b[$). Se tale assunzione non vale l'integrale in dx deve essere fatto non sull'intervallo $]a, b[$ ma sul suo sottoinsieme

$$A = \{x \in]a, b[: \alpha(x) < \beta(x)\}.$$

Per esempio l'integrale di una funzione f (integrabile in E) sull'insieme E dato da

$$E = \{(x, y) \in \mathbf{R}^2 : 0 < x < 2, x < y < x^2\}$$

deve essere calcolato come

$$\int_1^2 (\int_x^{x^2} f(x, y)\, dy) dx$$

cioè in dx tra 1 e 2 e non tra 0 e 2. Per spiegare la formula osserviamo intanto che nelle ipotesi date E è un insieme misurabile poiché vale $E = G_-(\beta) \cap G_+(\alpha)$ con α e β misurabili (v. paragrafo seguente per maggiori delucidazioni sul sottografico G_- e sul sopragrafico G_+). Poi nella spiegazione del libro bisogna solo osservare che in dy si integra sull'intervallo $]\alpha(x), \beta(x)[$ e questo equivale a integrare tra $\alpha(x)$ e $\beta(x)$ solo nel caso che $\alpha(x) < \beta(x)$. In caso contrario l'intervallo considerato è l'insieme vuoto, per cui per quegli x l'integrale in dy viene 0. Quindi l'integrale in dx su $]a, b[$ si spezza nella somma dell'integrale su A e dell'integrale su $]a, b[\setminus A$, e quest'ultimo integrale viene 0. Analogamente nella formula per gli integrali tripli di G.T. pg. 232 (dopo la parola "risulterà") l'integrale in $dx\, dy$ deve essere fatto nel caso generale, non su F ma su $\{(x, y) \in F : \alpha(x, y) < \beta(x, y)\}$. Inoltre le funzioni α e β devono essere misurabili su F, che a sua volta deve essere misurabile. ∎

Voglio qui spiegare un esempio di problema di Cauchy per equazioni a variabili separabili, trovando la soluzione massimale. L'esempio è semplice e l'avevo spiegata in aula, ma non sono sicuro che sia risultato chiaro. Qui voglio spiegare tutti i passaggi col massimo rigore. Sia dato il problema

$$\begin{cases} y'(t) = y^3(t) \\ y(0) = u \end{cases} \tag{5}$$

per $u \in \mathbf{R}$. Tra le soluzioni dell'equazione cerchiamo intanto le funzioni costanti c che annullano la funzione $x \mapsto x^3$ ossia le c tali che $c^3 = 0$, ossia $c = 0$. Questa è ovviamente la soluzione di (5) per $u = 0$, su tutto $\mathbf{R}$. Le altre soluzioni si ottengono come è noto dividendo per $y^3(t)$ poiché nell'intervallo in cui sono soluzioni non si annullano. Fissiamo d'ora in poi $u \neq 0$ e allora una funzione derivabile y sull'intervallo $I =]a, b[$ con $a < 0 < b$ (ovviamente l'intervallo deve contenere il punto "iniziale" 0) risolve (5) su I se e solo se vale

$$\begin{cases} \frac{y'(t)}{y^3(t)} = 1 \quad \forall\, t \in I \\ y(0) = u \end{cases} \tag{6}$$

(ove in (6) è sottointeso che $y(t) \neq 0$ se $t \in I$ altrimenti l'espressione in (6) non avrebbe senso) se e solo se (integrando) esiste $c \in \mathbf{R}$ tale che

$$\begin{cases} -\frac{1}{2y^2(t)} = t + c \quad \forall\, t \in I \\ y(0) = u \end{cases} \tag{7}$$

se e solo se

$$-\frac{1}{2y^2(t)} = t - \frac{1}{2u^2} \quad \forall\, t \in I \tag{8}$$

(poiché sostituendo $t = 0$ si ottiene che per y che verifica la prima "riga" di (7) y soddisfa la seconda riga di (7), se e solo se $c = -\frac{1}{2u^2}$). D'altra parte (8) implica che $t - \frac{1}{2u^2} < 0$ per tutti i $t \in I$ ossia $I \subseteq]-\infty, \frac{1}{2u^2}[$ (questa considerazione è comoda perché solo su questi intervalli si potrà proseguire il procedimento arrivando a (11)), quindi (8) equivale a

$$-\frac{1}{2y^2(t)} = t - \frac{1}{2u^2} \quad \forall\, t \in I \subseteq]-\infty, \frac{1}{2u^2}[\tag{9}$$

e (9) è equivalente a

$$y^2(t) = \frac{1}{\frac{1}{u^2} - 2t} \quad \forall\, t \in I \subseteq]-\infty, \frac{1}{2u^2}[\tag{10}$$

e (10) equivale a: esiste $\alpha : I \to \{-1, 1\}$ tale che

$$y(t) = \alpha(t)\sqrt{\frac{1}{\frac{1}{u^2} - 2t}} \quad \forall\, t \in I \subseteq]-\infty, \frac{1}{2u^2}[\tag{11}$$

in altri termini per ogni $t \in I$, $y(t)$ vale + o − la radice in (11). Questo non equivale a dire che

$$y(t) = \pm\sqrt{\frac{1}{\frac{1}{u^2} - 2t}} \quad \forall\, t \in I \subseteq]-\infty, \frac{1}{2u^2}[$$

perché quest'ultima formula implica che il segno + o − sia lo stesso per tutti i t, e chiaramente la deduzione algebrica da (10) non permette di dire questo. Però dato che (e questo segue anche da (11)) y non si annulla mai in I, e avendo assunto dall'inizio che y è derivabile e quindi continua in I, per il teorema dell'esistenza degli zeri y non può cambiare segno in I. Perciò in tutto I ha il segno che ha in 0 cioè $\mathrm{sign}(u)$. Perciò (10) equivale a

$$y(t) = \mathrm{sign}u\sqrt{\frac{1}{\frac{1}{u^2} - 2t}}\,\forall \quad t \in I \subseteq]-\infty, \frac{1}{2u^2}[\tag{12}$$

In conclusione il problema (5) ha soluzione su I se e solo se $I \subseteq]-\infty, \frac{1}{2u^2}[$, e nel caso y è dato dall'espressione (12), in altri termini la soluzione massimale di (5) è data da $y(t) = \mathrm{sign}u\sqrt{\frac{1}{\frac{1}{u^2}-2t}}$ sull'intervallo $]-\infty, \frac{1}{2u^2}[$.

4. Riassunto della teoria della misura e dell'integrazione

Riassumo in questo paragrafo alcune delle principali proprietà della teoria della misura e dell'integrazione. Richiamo senza giustificazione i fatti noti dalla teoria e giustifico, magari brevemente, gli altri. Si sottointende, che salvo avviso contrario gli insiemi considerati sono sottoinsiemi di $\mathbf{R}^n$. Cominciamo a vedere come è fatta la classe degli insiemi misurabili (in $\mathbf{R}^n$).

1. Ogni insieme aperto (in particolare $\emptyset$ e $\mathbf{R}^n$) è misurabile.

2. L'unione di una famiglia finita o numerabile di insiemi misurabile è un insieme misurabile.

3. L'intersezione di una famiglia finita o numerabile di insiemi misurabili è un insieme misurabile.

4. Se A e B sono insiemi misurabili l'insieme differenza $A \setminus B$ è misurabile. In particolare il complementare di un insieme B misurabile è un insieme misurabile (basta prendere $A = \mathbf{R}^n$). Segue da 1 che i chiusi sono misurabili in quanto complementari degli aperti.

5. Non tutti gli insiemi sono misurabili. In $\mathbf{R}$ un esempio è l'insieme V di Vitali visto a un'esercitazione. In $\mathbf{R}^n$ si può provare in modo analogo che $V \times [0,1]^{n-1}$ non è misurabile. Comunque gli insiemi che "si incontrano in pratica" sono misurabili.

Vediamo ora qualche proprietà della misura, che, ricordo, è una funzione m_n definita dalla classe degli insiemi misurabili in $\mathbf{R}^n$ a valori in $[0, +\infty]$.

6. Se A e B sono misurabili e $A \subseteq B$ allora $m_n(A) \leq m_n(B)$.

7. Se A è misurabile e $m_n(A) = 0$, e $B \subseteq A$, allora B è misurabile (e allora per 6 $m_n(B) = 0$).

8. Se abbiamo una famiglia finita di insiemi misurabili $A_1, ..., A_k$ si ha

$$m_n(\bigcup_{i=1}^{k} A_i) \leq \sum_{i=1}^{k} m_n(A_i)$$

e vale l'uguaglianza se (non solo se) gli insiemi A_i sono a due a due disgiunti.

9. Vale una proprietà analoga per l'unione numerabile, cioè, se abbiamo una famiglia numerabile di insiemi misurabili $A_1, ..., A_k, ...$ si ha

$$m_n(\bigcup_{i=1}^{\infty} A_i) \leq \sum_{i=1}^{\infty} m_n(A_i)$$

e vale l'uguaglianza se gli insiemi A_i sono a due a due disgiunti. Se invece la successione A_i è crescente nel senso che si ha $A_i \subseteq A_{i+1}$ per $i = 1, 2, ...$ allora

$$m_n(\bigcup_{i=1}^{\infty} A_i) = \lim_{i \to +\infty} m_n(A_i).$$

10. Se A e B sono misurabili, $B \subseteq A$ e $m_n(B) < +\infty$, allora

$$m_n(A \setminus B) = m_n(A) - m_n(B)$$

(nota che questa formula non avrebbe senso se fosse $m_n(B) = +\infty$ poiché allora si avrebbe $m_n(A) - m_n(B) = +\infty - (+\infty)$). Tale formula segue dalla semplice osservazione che $A = B \cup (A \setminus B)$, e B e $A \setminus B$ sono insiemi disgiunti, quindi per 8, $m_n(A) = m_n(B) + m_n(A \setminus B)$.

11. Se A_i è una successione di insiemi misurabili "decrescente" nel senso che $A_i \supseteq A_{i+1}$ per ogni $i = 1, 2, ...$ allora

$$m_n(\bigcap_{i=1}^{\infty} A_i) = \lim_{i \to +\infty} m_n(A_i)$$

a patto che $m_n(A_1) < +\infty$ (senza questa condizione la formula non vale poiché basta considerare $A_i = [i, +\infty[$, la cui intersezione è vuota e ha quindi misura 0, mentre ogni A_i ha misura $+\infty$). Per provare la formula basta osservare che, posto $B_i = A_1 \setminus A_i$, da 10 $m_n(B_i) = m_n(A_1) - m_n(A_i)$ e la successione B_i è chiaramente crescente e per le regole di De Morgan l'unione dei B_i vale

$$A_1 \setminus (\bigcap_{i=1}^{+\infty} A_i),$$

quindi applicando l'ultima formula in 9 si ottiene

$$m_n(A_1) - m_n(\bigcap_{i=1}^{+\infty} A_i) = m_n\big(A_1 \setminus (\bigcap_{i=1}^{+\infty} A_i)\big)$$

$$\lim_{i \to +\infty} m_n(B_i) = \lim_{i \to +\infty} (m_n(A_1) - m_n(A_i))$$

e quindi la tesi.

12. Ogni insieme costituito da un solo punto $y \in \mathbf{R}^n$ ha misura 0. Infatti $\{y\}$ è misurabile in quanto chiuso. Inoltre $\{y\} \subseteq \prod_{i=1}^{n} [y_i - \varepsilon, y_i + \varepsilon]$, quindi per ogni $\varepsilon > 0$ si ha

$$m_n(\{y\}) \leq m_n\big(\prod_{i=1}^{n} [y_i - \varepsilon, y_i + \varepsilon]\big) = (2\varepsilon)^n.$$

13. Se A è misurabile e B ha misura 0 allora $m_n(A \cup B) = m_n(A)$. Infatti $m_n(A) \leq m_n(A \cup B) \leq m_n(A) + m_n(B) = m_n(A)$.

14. Segue da 13 che se A e B sono misurabili e

$$m_n(A) = m_n(B) < +\infty, \ A \subseteq C \subseteq B$$

allora C è misurabile, e $m_n(C) = m_n(A)$. Infatti segue da 10 che $m_n(B \setminus A) = 0$, quindi, essendo $C \setminus A \subseteq B \setminus A$, segue (da 7) che $C \setminus A$ è misurabile e $m_n(C \setminus A) = 0$, e per finire basta notare che $C = A \cup (C \setminus A)$.

15. Se a e b sono numeri reali e $a < b$, si ha $m_1([a, b[) = m_1(]a, b[) = m_1(]a, b]) = m_1([a, b]) = b - a$. Infatti l'ultima uguaglianza segue dalla definizione di misura e gli altri intervalli si ottengono dall'intervallo chiuso togliendo un insieme fatto da uno o due punti e quindi avente misura nulla.

16. Se $E \subseteq \mathbf{R}^n$ e $F \subseteq \mathbf{R}^k$ sono insiemi misurabili, allora $E \times F$ è misurabile in $\mathbf{R}^{n+k}$ e $m_{n+k}(E \times F) = m_n(E) m_k(F)$ con la convenzione $0 \times (+\infty) = 0$.

17. Gli insiemi finiti o numerabili hanno misura 0 (poiché unione finita o numerabile di punti, e i punti hanno misura 0).

18. Ogni insieme limitato e misurabile ha misura finita.
19. Esistono insiemi illimitati di misura finita ad esempio $\mathbf{Q} \subseteq \mathbf{R}$ ha misura 0, in quanto numerabile.
20. Ogni insieme aperto non vuoto A ha misura > 0. Infatti A per definizione contiene una palla (di raggio opportunamente piccolo) ed è facile vedere che questa deve contenere un rettangolo T, che per definizione di misura ha misura > 0. Quindi $m_n(A) \geq m_n(T) > 0$.
21. Si vede facilmente dalla definizione che $m_n(\mathbf{R}^n) = +\infty$, $m_1([a, -\infty[) = m_1(]a, +\infty[) = m_1(]\infty, a]) = m_1(]-\infty, a[)$ per ogni numero reale a.
22. Sia E un insieme misurabile e siano $f_i, g_j : E \to \mathbf{R}$ funzioni continue ove i varia su un insieme finito o numerabile I, e j varia su un insieme finito o numerabile J (lo stesso enunciato vale per funzioni misurabili). Allora l'insieme

$$A = \{x \in E : f_i(x) < 0 \ \forall\, i \in I,\ g_j(x) \leq 0 \ \forall\, j \in J\}$$

è misurabile in quanto si ha

$$A = \big(\bigcap_{i \in I} B_i\big) \cap \big(\bigcap_{j \in J} C_j\big)$$

ove

$$B_i = \{x \in E : f_i(x) < 0\}, \quad C_j = \{x \in E : g_j(x) \leq 0\}$$

e gli insiemi B_i sono aperti in E poiché la f_i è continua, quindi per fatti noti dagli spazi metrici sono esprimibili come l'intersezione di un aperto di $\mathbf{R}^n$ (quindi misurabile) con l'insieme misurabile E, e quindi sono misurabili. Con un discorso analogo si vede che i C_j sono misurabili, quindi A è misurabile in quanto intersezione di una famiglia finita o numerabile di insiemi misurabili. Come esempio sia

$$A = \{(x, y) \in \mathbf{R}^2 : xy < 4, x^2 + y^2 \geq 2\}.$$

Allora A rientra nel caso detto prendendo $f_1(x, y) = xy - 4$, $g_1(x, y) = 2 - x^2 - y^2$.

Ora passiamo a funzioni misurabili e relazioni tra funzioni misurabili e insiemi misurabili
23. Ricordo che quando si parla di funzioni misurabili, si intendono nel caso piú generale funzioni a valori in $\overline{\mathbf{R}}$ (definite su un insieme *misurabile* E). Le funzioni continue sono misurabili. Inoltre somma, differenza, prodotto di funzioni misurabili sono funzioni misurabili (se possono essere definite, cioè ad esempio la somma di funzioni è definita se in nessun punto una delle due funzioni vale $+\infty$ e l'altra $-\infty$, ovviamente questi problemi non sussistono se le funzioni sono a valori reali). sup, inf, max lim, min lim di una successione di funzioni misurabili è una funzione misurabile. In particolare se esiste il limite puntuale di una successione di funzioni misurabile questo è una funzione misurabile. D'ora innanzi considereremo *sempre* (e sarà sottointeso), funzioni a valori reali
24. Data $f : E \to \mathbf{R}$ con E misurabile, allora f è misurabile se e solo se la funzione $\tilde{f} : \mathbf{R}^n \to \mathbf{R}$ definita da

$$\tilde{f}(x) = \begin{cases} f(x) & \text{se } x \in E \\ 0 & \text{se } x \notin E \end{cases}$$

è misurabile. Cioè la misurabilità di una funzione definita in E viene riportata a quella di una funzione definita su tutto $\mathbf{R}^n$. Nota che se f fosse definita a priori su tutto $\mathbf{R}^n$ (ma questo non accade sempre) si potrebbe definire semplicemente $\tilde{f} = f\varphi_E$. Per provare quanto affermato notare che per ogni $t \in \mathbf{R}$

$$\{x \in E : f(x) > t\} = \{x \in \mathbf{R}^n : \tilde{f}(x) > t\} \cap E$$

e da questo e dalla definizione di funzione misurabile si deduce che se $\tilde{f}$ è misurabile anche f è misurabile (tenuto conto che E è misurabile). Per vedere il viceversa basta ragionare analogamente usando la formula

$$\{x \in \mathbf{R}^n : \tilde{f}(x) > t\} = \begin{cases} \{x \in E : f(x) > t\} & \text{se } t \geq 0 \\ \{x \in E : f(x) > t\} \cup (\mathbf{R}^n \setminus E) & \text{se } t < 0. \end{cases}$$

25. La funzione caratteristica φ_E di un insieme E è misurabile se e solo se E è misurabile. Questo si vede semplicemente osservando che $\{x \in \mathbf{R}^n : \varphi_E(x) > t\}$ vale $\emptyset$, $\mathbf{R}^n$, o E a seconda di t e vale E se $0 \leq t < 1$, e usando la definizione di funzione misurabile.
26. Se sono dati m insiemi misurabili $E_1, ..., E_m$ *a due a due disgiunti* e $f_i : E_i \to \mathbf{R}$ funzioni misurabili su E_i, allora la funzione $f : \mathbf{R}^n \to \mathbf{R}$ definita da

$$f(x) = \begin{cases} f_i(x) & \text{se } x \in E_i \\ 0 & \text{se } x \notin \bigcup_{i=1}^{m} E_i \end{cases}$$

è misurabile. Infatti si vede facilmente, usando le notazioni di 24, che $f = \sum_{i=1}^{m} \tilde{f}_i$. Ovviamente può accadere che $\bigcup_{i=1}^{m} E_i = \mathbf{R}^n$, nel qual caso nella forma precedente si usa solo la prima definizione. Come esempio sia data la funzione definita in $\mathbf{R}^2$ da

$$f(x, y) = \begin{cases} \frac{1}{x} \text{ se } x > 0 \\ xy \text{ se } x \leq 0. \end{cases}$$

f è misurabile in quanto la funzione $\frac{1}{x}$ è misurabile, in quanto continua, sull'insieme misurabile $\{(x, y) : x > 0\}$, e analogamente la funzione xy è misurabile in quanto continua sull'insieme misurabile $\{(x, y) : x \leq 0\}$. In questo modo si vede che "praticamente tutte" le funzioni che si incontrano in pratica sono misurabile. Esistono comunque anche funzioni non misurabili come la funzione caratteristica dell'insieme di Vitali (v. 25).
27. Se E_i e f_i sono come in 26, ma l'indice i varia in un insieme numerabile (che a meno di un cambiamento di indice possiamo supporre che sia $\mathbf{N} \setminus \{0\}$), e non tra 1 e m, cioè abbiamo una famiglia numerabile di insiemi (e quindi di funzioni) e definiamo f analogamente come

$$f(x) = \begin{cases} f_i(x) & \text{se } x \in E_i \\ 0 & \text{se } x \notin \bigcup_{i=1}^{\infty} E_i \end{cases}$$

allora f è misurabile, e questo perché

$$f = \lim_{m \to +\infty} \sum_{i=1}^{m} \tilde{f}_i.$$

Come esempio la funzione parte intera è misurabile in quanto si definisce uguale alla funzione costante (e quindi misurabile) n sugli insiemi misurabili $[n, n+1[$, al variare di $n \in \mathbf{Z}$, che sono a due a due disgiunti.

28. Data una funzione f definita da un insieme misurabile E a valori reali definiamo

$$G_-(f) = \{(x, y) : x \in E : y < f(x)\}$$

$$G_+(f) = \{(x, y) : x \in E : y > f(x)\}$$

$$G(f) = \{(x, y) : x \in E : y = f(x)\}$$

in altre parole $G_-(f)$ è il sottografico di f (in G.T. chiamato $\mathcal{F}$), $G_+(f)$ è il "sopragrafico" di f, e $G(f)$ è il grafico di f. Dalla teoria sappiamo che f è misurabile se e solo se l'insieme $G_-(f)$ è misurabile e si potrebbe dedurre allo stesso modo che f è misurabile se e solo se l'insieme $G_+(f)$ è misurabile. Segue che *se f è misurabile* allora $G(f) = (E \times \mathbf{R}) \setminus (G_-(f) \cup G_+(f))$ è misurabile. Inoltre applicando il teorema di Fubini (o meglio di Tonelli) si vede facilmente che $G(f)$ ha misura 0, poichè la misura di $G(f)$ è uguale all'integrale su $\mathbf{R}^n$ della funzione caratteristica di $G(f)$ che vale

$$\int_{\mathbf{R}}^{n} \Big(\int_{\mathbf{R}} \varphi_{G(f)}(x, y)\, dy \Big) dx$$

ove $x \in \mathbf{R}^n$, $y \in \mathbf{R}$, e e per ogni $x \in \mathbf{R}^n$ l'integrale in dy vale 0 poiché per quel valore di x c'è al piú un y (e cioè $y = f(x)$ se $x \in E$) dove l'integranda è diversa da 0.

29. Come applicazione delle considerazioni di 28 vediamo che ogni iperpiano A in $\mathbf{R}^n$ $(n > 1)$ ha misura 0. Infatti si può scrivere A come

$$\{x \in \mathbf{R}^n : \sum_{i=1}^{n} a_i x_i = b\}$$

ove gli a_i sono numeri reali non tutti nulli, e b è un numero reale. Supponendo ad esempio $a_n \neq 0$ si vede subito che A è il grafico della funzione f definita su $\mathbf{R}^{n-1}$ da

$$f(x) = \frac{b - \sum_{i=1}^{n-1} a_i x_i}{a_n}.$$

Un altro esempio è l'insieme $C = \{(x, y) \in \mathbf{R}^2 : x^2 + y^2 = 1\}$, cioè la circonferenza di centro l'origine e raggio 1. Si vede facilmente che C non è un grafico, ma è l'unione dei grafici delle due funzioni (di x) $\sqrt{1 - x^2}$ e $-\sqrt{1 - x^2}$ definite sull'insieme misurabile $E = [-1, 1]$, che sono misurabili in quanto continue. Quindi C ha misura 0 in quanto unione di due

insiemi di misura 0. Ovviamente si vede analogamente che qualunque circonferenza ha misura 0, e similmente in $\mathbf{R}^n$ ha misura 0 la sfera di centro P e raggio r, che è definita come l'insieme degli $z \in \mathbf{R}^n$ che hanno distanza r da P.

30. Come ultima cosa ricordo che se una funzione f da $\mathbf{R}$ in $\mathbf{R}$ è limitata e nulla fuori da un compatto, allora se è Riemann integrabile è anche Lebesgue integrabile e i due integrali coincidono. Però f può essere Lebesgue integrabile, ma non Riemann integrabile; un esempio di questo tipo è la funzione caratteristica dell'insieme dei razionali in $[0,1]$ (che sarebbe poi la funzione di Dirichlet). Per gli integrali impropri la situazione è diversa. Supponiamo per semplicità che f sia *continua* da un intervallo aperto I a valori in $\mathbf{R}$. Allora se f è Lebesgue integrabile su I (questo è garantito in particolare se $f \geq 0$), segue che esiste l'integrale improprio di f (eventualmente infinito) e coincide con l'integrale di Lebesgue, però f potrebbe essere integrabile in senso improprio, ma non nel senso di Lebesgue. Un esempio di questo tipo è la funzione $\frac{\sin x}{x}$ sull'intervallo $]0, +\infty[$. Un modo per vedere se f è Lebesgue integrabile consiste nel calcolare gli integrali delle funzioni f^+ e f^-; essendo queste funzioni ≥ 0 (e continue come facile verificare) il loro integrale di Lebesgue coincide con quello improprio (finito o infinito). Dalla teoria risulta che f è Lebesgue integrabile se e solo se questi integrali non valgono ambedue $+\infty$. La classe delle funzioni Lebesgue integrabili è molto ampia. Per esempio si può dire che tutte le funzioni misurabili ≥ 0 sono Lebesgue integrabili, mentre se f è misurabile ma non ≥ 0, allora f è Lebesgue integrabile se e solo se gli integrali delle funzioni f^+ e f^- non sono entrambi infiniti.