Gruppi di matrici

Paolo Caressa

Roma, 1996

1 Coordinate canoniche su un gruppo di matrici

Consideriamo il gruppo lineare generale $GL(n, \mathbb{R})$ formato dalle matrici invertibili, cioè da quelle che hanno il determinante diverso da zero. Osserviamo che, mentre l'insieme $M_n(\mathbb{R})$ di tutte le matrici quadrate $n \times n$ è uno spazio vettoriale (e quindi in modo ovvio una varietà differenziabile) identificabile con $\mathbb{R}^{n^2}$ e con coordinate fornite dalle entrate delle matrici, l'insieme delle matrici invertibili non lo è, perché, ad esempio, non contiene la matrice 0. Tuttavia $GL(n, \mathbb{R})$ è definito come

$$GL(n, \mathbb{R}) = M_n(\mathbb{R}) \setminus \{X \mid \det(x) = 0\}$$

e quindi, dato che la funzione $X \mapsto \det(x)$ è un polinomio nelle entrate della matrice X, è di certo continua sicché $GL(n, \mathbb{R})$ è un aperto in M_n, cioè è una varietà differenziabile di dimensione n^2.

Osserviamo inoltre che $GL(n, \mathbb{R})$ è un gruppo rispetto al prodotto di matrici e che (esercizio!) questa operazione è C^∞ fra le varietà

$$GL(n, \mathbb{R}) \times GL(n, \mathbb{R}) \to GL(n, \mathbb{R})$$

Questo fatto si esprime dicendo che $GL(n, \mathbb{R})$ è un *gruppo di Lie.*

Introduciamo nello spazio delle matrici una norma ponendo

$$||A|| := \sum_{i,j}^{1\ldots n} |a_{ij}|^2$$

ove $((a_{ij}))$ sono le coordinate della matrice A, i.e. le sue entrate.

Allora è ovvio che $||\cdot||$ è una norma sullo spazio delle matrici, in particolare si ha che

$$||A+B|| \leq ||A|| + ||B||$$

Questa norma è compatibile col prodotto di matrici, nel senso che vale il seguente

Lemma 1.1 *Se A e B sono matrici in $M_n(\mathbb{R})$ allora*

$$||AB|| \leq ||A|| \cdot ||B||$$

La dimostrazione è un esercizio che poggia sulla diseguaglianza di Cauchy–Schwartz. Il lemma precedente si può esprimere dicendo che M_n è un'algebra di Banach.

Ora sappiamo che $GL(n,\mathbb{R})$ è una varietà differenziabile rispetto alla struttura differenziabile indotta da quella di M_n, come avviene per ogni aperto in una varietà. Per capire meglio come è fatta questa struttura, proviamo a scrivere delle coordinate locali in $GL(n,\mathbb{R})$.

Osserviamo che, avendo introdotto una norma su M_n (che ovviamente lo rende uno spazio di Banach!) possiamo parlare di convergenza di serie: allora dimostriamo il

Lemma 1.2 *Se $||X|| < 1$, la matrice $A = 1+X$ è invertibile, cioè appartiene a $GL(n,\mathbb{R})$.*

DIMOSTRAZIONE: Consideriamo la serie di matrici

$$B = 1 - X + X^2 - X^3 + ...$$

Questa serie converge se $||X|| < 1$. Infatti, per la diseguaglianza triangolare della norma di matrici e per il Lemma 1:

$$\begin{aligned} ||X^m - X^{m+1} + X^{m+2} - \ldots \pm X^{m+k-1}|| \leq & \\ \leq ||X^m|| \cdot (1 + ||X|| + \cdots + ||X||^{k-1}) &= \\ = ||X||^m \frac{1-||X||^k}{1-||X||} & \end{aligned}$$

Dunque la serie delle somme parziali della serie B è una successione di Cauchy se $||X|| < 1$ e quindi la serie converge. Ma si ha

$$AB = (1+X)(1-X+X^2-X^3+\ldots) = 1$$

e quindi $B = A^{-1}$ cioè A è invertibile.

QED

Ora consideriamo la matrice $1 \in GL(n, \mathbb{R})$ e definiamo un intorno di essa con la condizione

$$||A - 1|| < 1$$

Cioè le matrici A che verificano la condizione precedente, per il Lemma 2, formano un intorno U della matrice 1 in $GL(n, \mathbb{R})$. A questo punto, per avere un intorno di una qualsiasi altra matrice $B \in GL(n, \mathbb{R})$ basta considerare $B \cdot U$ che è un intorno di B in quanto la moltiplicazione di matrici è C^∞. In questo modo abbiamo le coordinate locali sul gruppo: scriviamole in concreto. Sia B la matrice intorno alla quale vogliamo le coordinate. Allora, se $C = B^{-1} = ((c_{ij}))$ si pone:

$$\begin{aligned} x_{ij}(A) &= \sum_{k=1}^{n} c_{ik} a_{kj} - \delta_{ij} \\ x_{ij}(B) &= 0 \end{aligned}$$

Le coordinate $\{x_{ij}\}$ sono valide per ogni matrice A tale che

$$||A - B|| < ||B||$$

Osserviamo che in un gruppo di Lie qualsiasi basta definire le coordinate in un intorno dell'origine per averle su tutto il gruppo semplicemente moltiplicando.

Lo spazio tangente al gruppo $GL(n, \mathbb{R})$ nel suo punto 1 si può identificare con lo spazio delle matrici M_n nel modo seguente: sia $A(t)$ una curva in $GL(n, \mathbb{R})$, e supponiamo che questa curva passi per 1 al tempo $t = 0$, i.e. $A(0) = 1$. Allora il vettore tangente ad $A(t)$ in 1 è

$$\left(\frac{dA(t)}{dt} \right)_{t=0}$$

Viceversa, se X è una matrice qualsiasi allora la curva $A(t) = 1 + tX$ appartiene a $GL(n, \mathbb{R})$ per i valori di t sufficientemente piccoli (in virtù del Lemma 2), e si ha ovviamente:

$$A(0) = 1 \qquad \left(\frac{dA(t)}{dt} \right)_{t=0} = X$$

e quindi ogni matrice fa parte dello spazio tangente a $GL(n, \mathbb{R})$ in 1, che dunque si identifica con $M_n(\mathbb{R})$.

Un altro esempio notevole di gruppo di matrici è il *gruppo lineare speciale*

$$SL(n, \mathbb{R}) := \{X \in M_n(\mathbb{R}) | \det(X) = 1\}$$

Ovviamente si tratta di un sottogruppo di $GL(n,\mathbb{R})$, chiuso. Per vedere che è una varietà differenziabile, dato che è una ipersuperficie, basta verificare che ogni suo punto sia regolare, cioè che gli jacobiani delle coordinate abbiano rango massimo.

Questo è un esercizio che può svolgersi come segue: dapprima si fa vedere che la dimensione dello spazio tangente nel punto 1 è n^2-1 cioè la stessa della varietà, e questo si può verificare derivando la relazione $\det(A(t))=1$ (ove $A(t)$ è una curva che passa per 1 in $SL(n,\mathbb{R})$) per ottenere $\mathrm{tr}(X)=0$, e concludere quindi che lo spazio tangente in 1 si identifica a quello delle matrici a traccia nulla, che ha dimensione n^2-1. Infine si conclude che ogni punto B di $SL(n,\mathbb{R})$ è regolare spostandolo, con la moltiplicazione per B^{-1}, che è un diffeomorfismo, in 1.

Un gruppo importante di matrici è il *gruppo ortogonale*

$$O(n) := \{X\in M_n(\mathbb{R})|\ X\cdot X^T = 1\}$$

Questo è il gruppo delle matrici ortogonali, cioè quelle che preservano il prodotto scalare standard in $\mathbb{R}^n$: se $x=(x_1,\dots,x_n)$ e $y=(y_1,\dots,y_n)$ sono due vettori in $\mathbb{R}^n$ scritti in una fissata base, allora il loro prodotto scalare è

$$\langle x,y\rangle = \sum_k x_k y_k$$

Allora una matrice A è ortogonale se

$$\langle Ax,Ay\rangle = \langle x,y\rangle$$

cioè se

$$A\cdot A^T = 1$$

Quest'ultima equazione, corrisponde ad un sistema di $n(n-1)/2$ equazioni scalari, che quindi definiscono una superficie di questa dimensione in M_n. Di nuovo vediamo che non è singolare in 1, e quindi in nessun altro punto. Anche questo è un esercizio che consiste nel derivare la relazione precedente per ottenere $X+X^T=0$ e quindi concludere che lo spazio tangente in 1 si identifica allo spazio delle matrici antisimmetriche, che ha dimensione proprio $n(n-1)/2$.

Osserviamo che il gruppo $O(n)$ non solo è chiuso in M_n, ma è anche limitato, nella norma delle matrici. Infatti si ha

$$||AA^T|| = 1$$

cioè

$$n = \mathrm{tr}(1) = \mathrm{tr}(AA^T)) = \sum_k\sum_h a_{kh}a_{kh} = \sum_{kh} a_{kh}^2$$

e quindi una matrice ortogonale si trova nella sfera di centro l'origine e raggio $\sqrt{n}$: ne segue che $O(n)$ è limitato, e, essendo anche chiuso, compatto.

Tuttavia non è un gruppo connesso, ma ha due componenti: le matrici ortogonali a determinante 1 e quelle a determinante -1. La componente dell'identità è un sottogruppo:

$$SO(n, \mathbb{R}) := O(n) \cap SL(n, \mathbb{R})$$

normale (perché?) costituito dalle rotazioni proprie della sfera in $\mathbb{R}^n$: è ovviamente compatto.

Per vedere che $O(n)$ ha due componenti connesse si può procedere come segue: intanto si dimostra che $SO(n)$ è connesso, dato che per ogni matrice $A \in SO(n)$ esiste una matrice $B \in O(n)$ tale che

$$BAB^{-1} = \begin{pmatrix} \cos\varphi_1 & \operatorname{sen}\varphi_1 & 0\dots 0 \\ -\operatorname{sen}\varphi_1 & \cos\varphi_1 & 0\dots 0 \\ 0\dots & \dots & \dots 0 \\ 0\dots 0 & 0\dots 0 & \cos\varphi_k \operatorname{sen}\varphi_k \\ 0\dots 0 & 0\dots 0 & -\operatorname{sen}\varphi_k \cos\varphi_k \end{pmatrix}$$

(se $n = 2k$) oppure

$$BAB^{-1} = \begin{pmatrix} \cos\varphi_1 & \operatorname{sen}\varphi_1 & 0\dots 0 & 0 \\ -\operatorname{sen}\varphi_1 & \cos\varphi_1 & 0\dots 0 & 0 \\ 0\dots & \dots & \dots 0 & 0 \\ 0\dots 0 & 0\dots 0 & \cos\varphi_k \operatorname{sen}\varphi_k & 0 \\ 0\dots 0 & 0\dots 0 & -\operatorname{sen}\varphi_k \cos\varphi_k & 0 \\ 0\dots 0 & 0\dots 0 & 0 \quad 0 & 1 \end{pmatrix}$$

(se $n = 2k + 1$). Queste forme canoniche sono facilmente ottenute per induzione a partire dal caso $n = 2$, nel quale evidentemente

$$SO(2) = \left\{ \begin{pmatrix} \cos\varphi & \operatorname{sen}\varphi \\ -\operatorname{sen}\varphi & \cos\varphi \end{pmatrix} \mid \varphi \in [0, 2\pi) \right\} \equiv S^1$$

Allora considerando le matrici $C(t)$ che hanno la forma delle precedenti ma con al posto delle costanti $\{\varphi_i\}$ le variabili $\{t\varphi_i\}$ otteniamo una curva continua $t \mapsto B^{-1}C(t)B$ che connette ogni punto di $SO(n)$ con l'identità. Quindi $SO(n)$ è connesso e, dato che le matrici ortogonali a determinante -1 sono diffeomorfe a $SO(n)$ attraverso la moltiplicazine per la matrice -1, vediamo che $O(n)$ è unione di due connessi e quindi ha due componenti connesse.

Il fatto che $O(n)$ abbia due componenti connesse si riflette su $GL(n, \mathbb{R})$. Infatti le matrici invertibili e quelle ortogonali sono legate dalla cosiddetta *decomposizione polare*.

Teorema 1.3 *Se $X \in GL(n, \mathbb{R})$ allora esiste un'unica matrice $A \in O(n)$ ed un'unica matrice simmetrica P definita positiva tali che $X = AP$. In altri termini ha luogo la decomposizione*

$$GL(n, \mathbb{R}) = O(n) \times \{P \in M_n | \ P = P^T \ , \ P \textit{ definita positiva}\}$$

DIMOSTRAZIONE: Ricordiamo intanto che una matrice P è simmetrica se è uguale alla sua trasposta, ed è definita positiva se soddisfa una delle sue condizioni seguenti (la cui equivalenza è un facile esercizio):

(1) Ogni autovalore di P è positivo.

(2) Per ogni vettore $x \in \mathbb{R}^n$ $x^T P x > 0$.

Ora se X è invertibile, allora $Q := X^T X$ è una matrice simmetrica definita positiva. Esiste allora un'unica matrice simmetrica positiva P tale che $P^2 = Q$ (questo è un esercizio che richiede solo il teorema spettrale, cioè il fatto che una matrice simmetrica è diagonalizzabile). Ora poniamo

$$A := XP^{-1}$$

Questa matrice è ortogonale, infatti

$$\begin{aligned} AA^T &= XP^{-1}(P^{-1})^T X^T = XP^{-2}X^T \\ &= XQ^{-1}X^T = X(X^T X)^{-1}X^T \\ &= XX^{-1}X^{T-1}X^T = 1 \end{aligned}$$

Abbiamo cosìla nostra decomposizione. L'unicità è un esercizio (si tenga conto dell'unicità della radice quadrata P di una matrice simmetrica definita positiva Q).

QED

Ora, le matrici simmetriche definite positive sono uno spazio convesso (ogni matrice siffatta è diagonalizzabile con elementi diagonali positivi e quindi congiungibile alla matrice 1 con un segmento formato di matrici simmetriche positive) e quindi contraibile, sicché $GL(n, \mathbb{R})$ ha le stesse componenti di $O(n)$, cioè due: le matrici a determinante positivo, e le matrici a determinante negativo.

Questi risultati si possono interpretare osservando che le matrici invertibili possono caratterizzarsi dal fatto che i vettori colonna (o i vettori riga) che le compongono sono esattamente le basi di $\mathbb{R}^n$, mentre le matrici ortogonali sono caratterizzate dal fatto che i vettori colonna (o i vettori riga) che le compongono sono esattamente le basi ortonormali di $\mathbb{R}^n$ rispetto al prodotto scalare standard. Le due componenti connesse di $GL(n,\mathbb{R})$ e $=(n)$ corrispondono all'esistenza di due orientazioni in $\mathbb{R}^n$ e quindi alla distinzione in due classi delle basi e delle basi ortonormali.

Osserviamo che la decomposizione polare delle matrici ha un analogo nel caso complesso. Si possono difatti introdurre il gruppo lineare generale complesso $GL(n,\mathbb{C})$ delle matrici complesse di determinante non nullo, il gruppo lineare speciale complesso $SL(n,\mathbb{C})$ delle matrici complesse di determinante 1 ed il *gruppo unitario*

$$U(n) := \{X \in M_n(\mathbb{C}) |\ X \cdot X^* = 1\}$$

ove X^* è la matrice aggiunta a X cioè la coniugata della trsposta. Allora si dimostra, come nel caso reale, che tutte queste sono varietà differenziabili e che $U(n)$ è compatta, e formata dalle matrici che preservano il prodotto scalare hermitiano standard in $\mathbb{C}^n$. Si lascia come esercizio la dimostrazione del teorema di decomposizione polare nel caso complesso, che afferma essere ogni matrice complessa invertibile prodotto in modo unico di una matrice unitaria per una matrice simmetrica definita positiva (complessa).

Esercizio 1.4 *Dimostrare che $GL(n,\mathbb{C})$ è connesso, dimostrando prima che se X è una matrice complessa invertibile, allora esiste una matrice Y invertibile tale che YXY^{-1} è triangolare superiore (cioè ha tutti zeri sotto la diagonale: si usi l'esistenza di autovalori in $\mathbb{C}$ per una matrice qualsiasi) e quindi mostrando che ogni matrice invertibile può connettersi alla matrice identità con una curva continua contenuta in $GL(n,\mathbb{C})$.*

Il fatto che $GL(n,\mathbb{C})$ sia connesso mentre $GL(n,\mathbb{R})$ non lo è può intuitivamente vedersi come dovuto al fatto che nel caso reale $GL(n,\mathbb{R})$ è ottenuto togliendo da uno spazio vettoriale di dimensione n^2 una curva di dimensione 1, $\{\det(X)=0\}$, mentre $GL(n,\mathbb{C})$ è ottenuto togliendo ad uno spazio di dimensione reale $2n^2$, cioè $\mathbb{C}^n$, una curva complessa, sempre $\{\det(X)=0\}$, che però ha dimensione reale 2, ed è chiaro che non si può sconnettere uno spazio togliendogliene uno di codimensione 2, mentre è ben possibile sconnettere uno spazio togliendogliene uno di codimensione 1. Si pensi a $GL(1,\mathbb{R})$, che è $\mathbb{R}\setminus\{0\}$, ed a $GL(1,\mathbb{C})$, che è $\mathbb{C}\setminus\{0\}$: nell'un caso si ha una retta meno un punto, che quindi la sconnette, nell'altro un piano meno un punto, che non può certo sconnetterlo...

Osserviamo che, se $n=1$, allora $GL(1,\mathbb{C})=\mathbb{C}\setminus\{0\}$ e $U(1)=S^1$ (circonferenza unitaria, cioè i numeri complessi di modulo 1), sicché la decomposizione polare si riduce a quella ben nota per i numeri complessi: $z=\rho e^{i\vartheta}$ ove $\rho=|z|$ e $\vartheta = Arg(z)$. Osserviamo che il gruppo $U(1)$ è quindi connesso, e tale risulta in generale il gruppo unitario di dimensione n:

Esercizio 1.5 *Dimostrare la seguente forma normale delle matrici unitarie: per ogni matrice $A \in SU(n)$ esiste una matrice $B \in U(n)$ tale che*

$$BAB^{-1} = \begin{pmatrix} e^{i\varphi_1} & 0 & 0\dots 0 \\ 0 & e^{i\varphi_2} & 0\dots 0 \\ \dots & & \\ 0 & 0\dots 0 & e^{i\varphi_n} \end{pmatrix}$$

ove $\varphi_1 + \dots + \varphi_n = 2l\pi$ per $l \in \mathbb{Z}$ e gli $\{e^{i\varphi_j}\}$ sono gli autovalori della matrice A. Utilizzare questo fatto per dimostrare che $SU(n)$ e $U(n)$ sono connessi.

Esercizio 1.6 *Dimostrare che, come varietà, $U(n)$ è diffeomorfo al prodotto diretto di S^1 per $SU(n)$.*

Per concludere lasciamo l'esercizio di verificare che lo spazio tangente alla matrice identità della varietà $U(n)$ è lo spazio vettoriale delle matrici antihermitiane, cioè delle X complesse tali che $X + X^* = 0$.

2 Mappa esponenziale

Le matrici, per molti versi, sono una generalizzazione dei numeri, nel senso che, non solo si possono effettuare su di esse le usuali operazioni algebriche, ma, come abbiamo visto, anche il calcolo infinitesimale (limiti, serie, ...) per mezzo della norma definita su M_n. In realtà si può, per alcune classi di matrici, fare molto di più grazie al teorema spettrale ad esempio, che consente di diagonalizzare le matrici simmetriche. Allora potremo applicare qualsiasi funzione alle matrici, semplicemente applicandola agli elementi diagonali, cioè agli autovalori. In realtà il calcolo di funzioni continue o addirittura boreliane può farsi per operatori definiti in spazi di Hilbert, attraverso la cosiddetta Teoria Spettrale.

A noi interessa comunque, definire l'esponenziale di matrici, che è lo strumento fondamentale col quale, ad esempio, viene formulata la teoria dei sistemi di equazioni differenziali ordinarie. Se consideriamo una matrice A qualsiasi, vogliamo cioè dare senso all'espressione $\exp(A)$, in modo da, nel caso di matrici 1×1, ottenere il familiare esponenziale di numeri reali o complessi. Lavoreremo sui $\mathbb{R}$, ma tutto ciò che diremo si potrà facilmente estendere a matrici complesse.

Prima di parlare di matrici, dobbiamo però premettere un lemma fondamentale che riguarda l'esponenziale dei numeri reali, che dovremo utilizzare in alcuni punti cruciali: probabilmente si tratta di un risultato noto, ma è troppo importante, specie la tecnica della sua dimostrazione, per non cogliere l'occasione di richiamarlo qui:

Lemma Fondamentale 2.1 *Se $f : \mathbb{R} \to \mathbb{R}$ è una funzione continua che soddisfa alle seguenti proprietà:*

$$f(0) = 1$$
$$f(t+s) = f(t)f(s)$$

allora esiste una costante k tale che per ogni $t \in \mathbb{R}$ $f(t) = e^{kt}$.

DIMOSTRAZIONE: Intanto osserviamo che le ipotesi su f implicano che

$$f(t)f(-t) = f(t-t) = 1$$

e quindi, per ogni $t \in \mathbb{R}$ $f(t) \neq 0$. Ora dimostriamo che la nostra funzione f è necessariamente derivabile infinite volte.

Consideriamo una funzione $w(t) \in C^\infty(\mathbb{R})$, nulla al di fuori di un certo intorno di $0 \in \mathbb{R}$ e tale che

$$\int_{-\infty}^{\infty} f(t)w(t)\, dt \neq 0$$

Una tale funzione esiste certamente dato che f non è mai nulla. Ora consideriamo la relazione $f(t+s) = f(t)f(s)$, moltiplichiamone ambo i membri per $w(s)$ ed integriamo rispetto a s ottenendo:

$$\int_{-\infty}^{\infty} f(t+s)w(s)ds = f(t)\int_{-\infty}^{\infty} f(s)w(s)ds$$

Ora poniamo

$$c := \int_{-\infty}^{\infty} f(s)w(s)ds$$

Allora è $c \neq 0$ e quindi dalla relazione precedente otteniamo:

$$f(t) = \frac{1}{c}\int_{-\infty}^{\infty} f(t+s)w(s)ds = \frac{1}{c}\int_{-\infty}^{\infty} f(u)w(u-t)du$$

Il secondo membro di quest'equazione è infinitamente derivabile rispetto a t, in quanto lo è $w(u-t)$ e dato che l'integrale si calcola sul supporto di w che può supporsi limitato. Ne segue che anche $f(t) \in C^\infty(\mathbb{R})$.

Ora osserviamo che la nostra funzione f soddisfa all'equazione differenziale

$$\frac{df}{dt} = kf$$

ove k è una costante. Infatti, se deriviamo ambo i membri della $f(t+s) = f(t)f(s)$ rispetto a t e poniamo $t = 0$ otteniamo:

$$f'(s) = f'(0)f(s)$$

e quindi, per $k = f'(0)$, f soddisfa all'equazione precedente.

Ma è un fatto elementare che ogni soluzione dell'equazione differenziale ordinaria omogenea $f' = kf$ è della forma $f(t) = e^{kt}$, e questo conclude la dimosrazione del lemma.

QED

Il fatto più interessante del risultato precedente è che la sola continuità implica l'infinita differenziabilità di una funzione, se questa soddisfa ad una condizione "gruppale" come quella di trasformare somme in prodotti: ciò si riscontra in molti altri casi.

Consideriamo ora una matrice qualsiasi $X \in M_n$, e poniamo

$$\exp(X) := 1 + \frac{X}{1!} + \frac{X^2}{2!} + \frac{X^3}{3!} + \dots$$

cioè la serie esponenziale classica. Questa è una scrittura che ha formalmente senso perché coinvolge solo somme e prodotti di matrici, e, perché definisca una matrice, dobbiamo trovarne un dominio di convergenza.

Lemma 1. *Per ogni matrice X la serie* $\exp(X)$ *converge.*

DIMOSTRAZIONE: Usiamo il criterio di Cauchy tramite la diseguaglianza triangolare della norma di matrici ed il Lemma 1.1:

$$\Big|\Big|\frac{X^m}{m!} + \frac{X^{m+1}}{(m+1)!} + \dots + \frac{X^{m+k-1}}{(m+k-1)!}\Big|\Big| < \\ < \frac{||X||^m}{m!} + \dots + \frac{||X||^{m+k-1}}{(m+k-1)!}$$

Ma la serie numerica $\exp||X||$ converge per ogni X e quindi le somme parziali di $\exp(X)$ costituiscono una successione di Cauchy nella norma delle matrici il che dimostra la convergenza della serie.

QED

Alcune proprietà formali dell'esponenziale valgono ancora per l'esponenziale di matrici:

Lemma 2. *Se X e Y sono matrici che commutano, cioè $XY = YX$ allora*

$$\exp(X + Y) = \exp(X)\exp(Y)$$

DIMOSTRAZIONE: Per l'ipotesi fatta su X e Y otteniamo:

$$\begin{aligned}\exp(X)\exp(Y) &= \left(\sum_{k=0}^{\infty}\frac{X^k}{k!}\right)\left(\sum_{h=0}^{\infty}\frac{Y^h}{h!}\right)\\ &= \sum_{m=0}^{\infty}\frac{1}{m!}\left(\sum_{h+k=m}\frac{m!}{k!h!}X^kY^h\right)\\ &= \sum_{m=0}^{\infty}\frac{1}{m!}(X+Y)^m\\ &= \exp(X+Y)\end{aligned}$$

Infatti se X e Y commutano, vale la formula del binomiale per calcolare $(X+Y)^n$ esattamente come per gli scalari.

QED

Come esercizio si fornisca un esempio di due matrici che non commutano e non verificano la tesi del lemma precedente, e si dimostri il

Lemma 3. *Per ogni matrice X, la matrice $\exp(X)$ è invertibile, e risulta*

$$(\exp(X))^{-1} = (-X)$$

Inoltre

$$\exp(X^T) = \exp(X)^T$$

e, se C è invertibile:

$$C^{-1}\exp(X)C = \exp(C^{-1}XC)$$

Possiamo cioè dire che exp è una mappa

$$\exp : M_n(\mathbb{R}) \to GL(n,\mathbb{R})$$

e quindi va dallo spazio tangente di $GL(n,\mathbb{R})$ in 1 in $GL(n,\mathbb{R})$.

Questo fatto vale per ognuno dei gruppi di matrici che abbiamo presentato, cioè:

Proposizione 4.

(1) *L'esponenziale di una matrice a traccia nulla appartiene a $SL(n,\mathbb{R})$.*

(2) *L'esponenziale di una matrice antisimmetrica appartiene a $O(n)$.*

(3) *L'esponenziale di una matrice antihermitiana appartiene a $U(n)$.*

DIMOSTRAZIONE: Se $\mathrm{tr}(X) = 0$, consideriamo la curva

$$A(t) = \exp(tX)$$

Allora per il Lemma 2

$$A(t+s) = A(t)A(s)$$

(il che esprime che A è un gruppo ad un parametro: un classico teorema di Stone afferma che ogni gruppo ad un parametro in uno spazio di Hilbert ha questa forma esponenziale) e quindi, se $d(t) := \det(A(t))$ si ha

$$d(t+s) = d(t)d(s)$$

e quindi la funzione d soddisfa le ipotesi del Lemma Fondamentale, per cui deve esistere una costante c tale che $d(t) = e^{ct}$. Dimostriamo che $c = 0$. In effetti è

$$d(t) = \det(\exp(tX)) = \det(1 + tX + o(t)) = t\ \mathrm{tr}(X) + o(t)$$

ove $o(t)$ è una matrice tale che $lim_{t\to 0}\frac{o(t)}{t} = 0$. Allora l'ipotesi che $\mathrm{tr}(X) = 0$ implica che

$$c = \left(\frac{df}{dt}\right)_{t=0}$$

il che dimostra il primo enunciato.

Per dimostrare il secondo osserviamo che se X è antisimmetrica, i.e. $X + X^T = 0$ allora X e X^T commutano, e quindi, per il Lemma 2:

$$(\exp(X))^T \exp(X) = \exp(X^T + X) = 1$$

cioè $\exp(X) \in O(n)$.

Lasciamo per esercizio il terzo enunciato.

QED

Osserviamo che per i gruppi speciali $SO(n)$ e $SU(n) = U(n) \cap SL(n, \mathbb{C})$ gli spazi tangenti all'identità sono l'intersezione degli spazi tangenti di $O(n)$ e $U(n)$ con le matrici a traccia nulla, e quindi valgono enunciati analoghi ai precedenti anche per questi gruppi.

Prima di proseguire, formuliamo due esercizi istruttivi:

Esercizio 5. *Se X è una matrice, dimostrare che* $\det(\exp(X)) = e^{tr(x)}$.

Una proprietà cruciale della mappa esponenziale è la seguente:

Proposizione 6. *La mappa* $\exp : M_n(\mathbb{R}) \to GL(n, \mathbb{R})$ *è un diffeomorfismo locale.*

Dimostrazione: Ovviamente exp è una mappa analitica fra varietà, visto che è determinata da una serie di potenze. Per dimostrare il nostro enunciato basterà dimostrare che, nel punto $0 \in M_n$, lo jacobiano di exp è diverso da zero, ed il teorema della funzione inversa farà il resto.

Ora, se $X = ((x_{ij}))$ e $\exp(X) = A = ((a_{ij}))$, allora

$$\frac{\partial a_{kl}}{\partial x_{ij}} = \delta_{ki}\delta_{lj} + \dots$$

dove i puntini indicano i termini che si annulano in $X = 0$. Quindi la matrice jacobiana della mappa esponenziale ha la forma

$$\left(\left(\frac{\partial a_{kl}}{\partial x_{ij}}\right)\right) = ((\delta_{ki}\delta_{lj}))$$

che è una matrice unitaria di rango massimo, il che dimostra la tesi.

QED

Il significato della proposizione precedente è che esiste un intorno dell'origine nello spazio vettoriale M_n tale che, ristretta a questo intorno, exp è un diffeomorfismo su un intorno dell'identità di $GL(n, \mathbb{R})$. Quindi il suo inverso costituisce una carta intorno all'identità della varietà $GL(n, \mathbb{R})$, e questo introduce delle coordinate privilegiate su $GL(n, \mathbb{R})$ che si dicono canoniche.

Per capire esattamente come sono fatte queste coordinate, calcoliamo esplicitamente la mappa che localmente inverte exp.

Intanto possiamo considerare il caso $n = 1$, nel quale, come è noto, la mappa esponenziale realizza un isomorfismo fra il gruppo additivo dei numeri reali ed il gruppo moltiplicativo dei numeri reali positivi. In questo caso l'inversa è data dal logaritmo, e quindi, visto che quello che vogliamo fare sulle matrici deve particolarizzarsi ai numeri, prendiamo spunto da questo e definiamo una serie di matrici:

$$\ln(A) = (A-1) - \frac{(A-1)^2}{2} + \frac{(A-1)^3}{3} - \dots$$

che è formalmente la familiare serie logaritmica che, se A fosse un numero a, convergerebbe assolutamente al logaritmo naturale per i valori di a tali che $|a-1| < 1$. Per questo motivo, la serie di matrici $\ln(A)$ converge nella norma delle matrici per le matrici A tali che $||A-1|| < 1$.

Osserviamo che l'identità scalare $e^{\ln\ a} = a$ può dimostrarsi considerando le serie dell'esponenziale e del logaritmo e sostituendole l'una nell'altra, e lo stesso procedimento può farsi nel caso delle matrici (lo si ripeta per esercizio) ottenendo:

$$\exp(\ln(A)) = A$$

per ogni matrice A tale che $||A - 1|| < 1$. In modo del tutto analogo si trova

$$\ln(\exp(A)) = A$$

per ogni matrice A tale che $||A|| < \ln\ 2$ (perché?). Si noti che se $||A|| < \ln\ 2$ allora $||\exp(A) - 1|| \leq e^{||A||} - 1 < 1$.

Esercizio 7. *La mappa*

$$\exp : M_2(\mathbb{R}) \to GL(2, \mathbb{R})$$

non è suriettiva in quanto lg *non è connesso mentre l'immagine di un connesso (M_n) tramite una mappa continua (exp è analitica!) lo è; si dimostri tuttavia che exp non è suriettiva nemmeno sulla componente connessa delle matrici a determinante positivo (suggerimento: provare con la matrice* $\begin{pmatrix} -2 & 0 \\ 0 & -1 \end{pmatrix}$*).*

La mappa esponenziale dalle matrici complesse in $GL(n, \mathbb{C})$ è invece suriettiva: questo può dimostrarsi considerando la forma canonica di Jordan delle matrici complesse, e dimostrando che ogni blocco di Jordan è l'immagine di una matrice tramite exp nel modo seguente: se A è un blocco di Jordan, scriviamolo come $A = (\lambda I) \cdot N$ ove le entrate diagonali di N sono tutte 1. Allora $\lambda I = \exp(A_1)$ e $N = \exp(A_2)$ ove $A_1 A_2 = A_2 A_1$, e per trovare A_2 si usa il logaritmo (esercizio).

Il calcolo delle matrici esponenziali è fondamentale nella teoria delle equazioni differenziali ordinarie: infatti un sistema di equazioni differenziali ordinarie può scriversi come:

$$\dot{\to}\, y = Ay + b$$

ove y è una funzione vettoriale e A una matrice. Allora la soluzione del sistema omogeneo associato $\dot{\to}\, y = Ay$ si ottiene per $y(t) = \exp(tA)y_0$ ove y_0 rappresenta il vettore che contiene i dati iniziali. Infatti

$$\frac{dy(t)}{dt} = \frac{d}{dt}\exp(tA)y_0 = A\exp(tA)y_0 = Ay(t)$$

Osserviamo che, per ogni matrice A, la mappa

$$t \mapsto \exp(tA)$$

è un morfismo fra il gruppo additivo dei numeri reali ed il gruppo moltiplicativo $GL(n, \mathbb{R})$.

Definizione 8. *Un morfismo di gruppi*

$$\rho : \mathbb{R} \to GL(n, \mathbb{R})$$

continuo, cioè continuo come mappa fra gli spazi topologici $\mathbb{R}$ *e* $GL(n, \mathbb{R})$ *si dice gruppo ad un parametro.*

Vogliamo dimostrare ora che ogni gruppo ad un parametro si ottiene per mezzo di un esponenziale:

Proposizione 9. *Se* $\rho : \mathbb{R} \to GL(n, \mathbb{R})$ *è un gruppo ad un parametro, esiste un'unica matrice* A *tale che per ogni* $t \in \mathbb{R}$ $\rho(t) = \exp(tA)$*.*

DIMOSTRAZIONE: Consideriamo un intorno V di 0 in M_n ed un intorno U di 1 in $GL(n, \mathbb{R})$ in modo che la mappa esponenziale ristretta a V sia un diffeomorfismo fra V ed U, come asserito dalla proposizione 6. Definiamo poi:

$$\psi(t) := \ln(\rho(t))$$

per quei $t \in \mathbb{R}$ tali che $\rho(t) \in U$ (ovviamente l'insieme di tali t sta in un certo intorno T dello zero in $\mathbb{R}$). Allora, se $t \in T$ e $s \in T$ sono tali che $\rho(t+s) \in U$ si ha

$$\begin{aligned}
\psi(t+s) &= \ln(\rho(t+s)) \\
&= \ln(\rho(t)\rho(s)) \\
&= \ln(\rho(t)) + \ln(\rho(s)) \\
&= \psi(t) + \psi(s)
\end{aligned}$$

Dunque ψ è una funzione lineare in t e quindi è della forma

$$\psi(t) = tA$$

per una determinata matrice A, sicché

$$\exp(\psi(t) = \rho(t) = \exp(tA)$$

per ogni $t \in T$.

Per ottenere il risultato per qualsiasi $t \in \mathbb{R}$ scriviamo un qualsiasi numero $x \in \mathbb{R}$ come $x = nt$ ove $t \in T$ e $n \in \mathbb{Z}$. Allora

$$\rho(x) = \rho(t)^n = \exp(tA)^n = \exp(xA)$$

QED