



CALCOLO DEGLI AUTOVALORI E FONDAMENTI DELLA MATEMATICA NUMERICA (Consistenza, stabilità e convergenza dei problemi numerici)

In questa lezione verranno introdotti in modo rigoroso i concetti fondamentali del calcolo numerico, con particolare attenzione alla nozione di buona posizione di un problema, alla stabilità, al condizionamento, alla convergenza e all'accuratezza. Analizzeremo dunque i concetti di stabilità, condizionamento e convergenza e inizialmente, introdurremo tali nozioni in modo astratto e generale. Successivamente, esamineremo esempi di problemi specifici per quantificare tali concetti e per renderli operativi.

Per cominciare, desideriamo riassumere il quadro generale di riferimento della matematica numerica. Il calcolo numerico si propone di risolvere in modo approssimato problemi formulati nell'ambito delle scienze matematiche o di altre discipline applicate. In termini generali, si può considerare un problema fisico formulabile mediante un modello matematico. La soluzione del problema fisico, che indicheremo genericamente con u_F , dove u rappresenta la soluzione associata al problema fisico F , avrà un significato che dipende dalla natura del problema originario. Il problema fisico sarà in generale formulabile tramite un modello matematico che in maniera molto generale indichiamo come

$$F(u, d) = 0,$$

il quale può essere rappresentato da un'equazione o da un sistema di equazioni, che possono essere di natura lineare o non lineare, differenziale, integrale, algebrica, e così via. Indicheremo con d l'insieme dei dati del problema. Si supporrà, quindi, che la soluzione del problema matematico, che denotiamo con u , dipenda dai dati d . In questo contesto, i dati rappresentano variabili note, mentre u è la variabile incognita del problema.

In generale, il problema matematico non ammette una soluzione esplicita tramite formule chiuse. Per tale motivo, è necessario sostituirlo con un problema numerico, o con un metodo numerico. Il problema numerico si formula in modo analogo, e si rappresenta come

$$F_n(u_n, d_n) = 0,$$

dove il pedice n sta a significare numerico, con u_n e d_n che indicano rispettivamente la soluzione approssimata e i dati discreti associati al problema.

Il problema F_n rappresenta un'equazione, o un sistema di equazioni, che deve in qualche modo approssimare l'equazione che definisce il problema matematico originario F . L'insieme dei dati numerici è rappresentato dalla variabile d_n . In alcuni casi, l'indice n potrà indicare anche la dimensione del problema; esamineremo di volta in volta il significato preciso attribuito a tale indice. La soluzione del problema numerico è denotata da u_n , mentre F_n costituisce una relazione funzionale che sostituisce quella del problema esatto F . Di conseguenza, F_n dovrà approssimare in modo adeguato F . Vedremo successivamente come definire alcuni criteri utili per confrontare F con F_n .

Un problema numerico deve essere risolto mediante l'utilizzo di un calcolatore, attraverso l'implementazione di algoritmi codificati in un linguaggio di programmazione. È quindi spesso necessario introdurre un cosiddetto algoritmo computazionale, che rappresenta la versione eseguibile del metodo numerico e che sarà concretamente risolto dal calcolatore. La soluzione finale di questo processo è quella restituita dal calcolatore, che indicheremo con \bar{u}_n . Utilizziamo il simbolo soprasseduto per sottolineare che, in generale, \bar{u}_n sarà diversa da u_n , a causa del fatto che il calcolatore



opera, come si vedrà, con la cosiddetta aritmetica finita. Ciò comporta l'approssimazione e l'arrotondamento dei numeri reali e delle operazioni algebriche effettuate su numeri reali.

In ultima analisi, l'obiettivo sarà quello di dimostrare che \bar{u}_n risulti sufficientemente vicina a u_F , ovvero alla soluzione del problema fisico. Se tale condizione è soddisfatta, si potrà concludere che l'intero processo di approssimazione del problema fisico - che ha comportato la formulazione di un modello matematico, la sostituzione di tale modello con un metodo numerico, la costruzione di un algoritmo computazionale e la sua successiva implementazione sul calcolatore - ha prodotto un risultato finale sufficientemente accurato da rappresentare in maniera attendibile la soluzione del problema fisico originario.

Questo costituisce il quadro generale di riferimento all'interno del quale si colloca l'analisi numerica, detta anche, in modo equivalente, matematica numerica/discreta o calcolo numerico. Si considereranno dunque, in sequenza, un problema fisico, un problema matematico, un metodo numerico e un algoritmo computazionale. Tuttavia, nello specifico, l'analisi numerica non si occupa dell'intero processo. Essa si concentra, in prima istanza, sulla sostituzione del modello matematico con un metodo numerico e, successivamente, sulla sua realizzazione sul calcolatore.

L'obiettivo primario dell'analisi numerica è quello di ottenere una soluzione u_n che sia convergente alla soluzione esatta u del problema matematico. In altri termini, si richiede che

$$\lim_{n \rightarrow \infty} u_n = u,$$

ovvero che, al crescere dell'indice n , la soluzione approssimata u_n tenda alla soluzione esatta u .

Ciò implicherà che il metodo numerico costituisce una buona approssimazione del modello matematico. Questa condizione deve verificarsi quando n tende all'infinito, dove n rappresenta, in generale, la dimensione o la complessità del problema numerico. Pertanto, sarà di nostro interesse dimostrare che l'errore

$$e_n = u_n - u,$$

tenda a zero al tendere di n all'infinito, ovvero

$$\lim_{n \rightarrow \infty} e_n = 0.$$

Abbiamo osservato, nei precedenti nuclei, che esistono due modalità principali per misurare la distanza tra due entità matematiche: la misura assoluta e la misura relativa percentuale. In particolare, definiamo l'errore assoluto e relativo come

$$\begin{aligned} \text{errore assoluto: } & \|u_n - u\|, \\ \text{errore relativo (percentuale): } & \frac{\|u_n - u\|}{\|u\|}, \end{aligned}$$

dove u rappresenta la soluzione del modello matematico. Tale soluzione può essere, a seconda del contesto, un numero reale, un vettore o una matrice. Per questa ragione, si ricorre all'uso della norma per quantificare sia la grandezza delle entità considerate, sia la distanza tra due entità omologhe. Pertanto, la norma $\|u_n - u\|$ fornisce una misura assoluta dell'errore, mentre il rapporto $\frac{\|u_n - u\|}{\|u\|}$ fornisce una misura relativa, o percentuale, dell'errore. Sarà quindi di nostro interesse ottenere una stima dell'errore percentuale commesso sulla soluzione, ovvero della distanza, misurata in norma, tra u_n e la soluzione esatta u , rapportata alla norma della soluzione stessa $\|u\|$.



Nel menzionare la convergenza della soluzione numerica u_n verso la soluzione esatta u , abbiamo introdotto il concetto di convergenza. È ora opportuno caratterizzare le condizioni necessarie affinché tale convergenza si verifichi. Per garantire la convergenza, è necessario che siano soddisfatte due proprietà fondamentali: la consistenza e la stabilità del metodo numerico.

La proprietà di consistenza assicura che, nel passaggio dal problema matematico al problema numerico, non venga persa alcuna informazione rilevante. In altre parole, il problema numerico deve essere una rappresentazione fedele del problema matematico dal punto di vista strutturale. Vedremo in seguito come questa “somiglianza” possa essere definita e verificata in modo rigoroso.

La seconda proprietà necessaria è la stabilità del problema numerico. Come accennato e visto per alcuni problemi specifici nelle lezioni precedenti, la stabilità consiste essenzialmente nella capacità di controllare la soluzione numerica in funzione dei dati del problema numerico. Più precisamente, si richiede che tale controllo sia uniforme rispetto alla dimensione del problema. Con questa espressione si intende che, se ad esempio si affronta un sistema costituito da dieci equazioni in dieci incognite, si desidera che la soluzione numerica possa essere controllata in funzione dei dati tramite una costante che non dipenda dalla dimensione del sistema (in questo caso, dieci). Analogamente, se si considera un sistema con mille equazioni in mille incognite, la costante di controllo dovrà rimanere indipendente da questa dimensione. In tal modo, si evita che la costante cresca indefinitamente con l'aumentare della dimensione del problema, mantenendo il controllo stabile anche per problemi di grande scala.

Procediamo ora a una caratterizzazione più precisa del concetto di consistenza, che in precedenza abbiamo introdotto in modo qualitativo ed euristico. La consistenza implica che il limite di $F_n(u, d)$ sia pari a zero, ossia

$$\lim_{n \rightarrow \infty} F_n(u, d) = 0.$$

Ricordiamo che il problema numerico è

$$F_n(u_n, d_n) = 0,$$

Pertanto, richiedere che $F_n(u, d)$ sia esattamente pari a zero equivale a pretendere che la soluzione esatta, corrispondente ai dati indicati con d , soddisfi il problema numerico. Tuttavia, ciò non è generalmente vero, ovvero non si ha in generale $F_n(u, d) = 0$. Si richiede invece che il limite di $F_n(u, d)$ sia zero al tendere di n all'infinito. In altre parole, si chiede che

$$\lim_{n \rightarrow \infty} (F_n(u, d) - F(u, d)) = 0,$$

considerando che $F(u, d) = 0$ per definizione, in quanto essa rappresenta la soluzione esatta del problema. Qualitativamente, si richiede che la legge numerica F_n , utilizzata per approssimare la legge matematica esatta F , si avvicini a quest'ultima. Più precisamente, si richiede la convergenza di F_n a F al tendere di n all'infinito, ovvero

$$F_n \rightarrow F, \quad \text{per } n \rightarrow \infty.$$

Si tratta indubbiamente di un concetto generale, espresso in termini astratti. Procediamo ora con un esempio specifico, nel quale tenderemo di tradurre in termini euristici quanto espresso in maniera astratta.

Qualora si desideri approssimare la derivata di una funzione u rispetto a x in un determinato punto

$$\frac{du}{dx}$$



si può considerare che la derivata, ove esistente, è definita come il limite dei rapporti incrementali di u intorno a quel punto

$$\frac{du}{dx} = \lim_{h \rightarrow 0} \frac{u(x+h) - u(x)}{h}.$$

Pertanto, è possibile sostituire la derivata con un rapporto incrementale in avanti, all'indietro o centrato. Il rapporto incrementale in avanti consiste nel prendere il punto di interesse, in cui si vuole calcolare la derivata, e un punto vicino, sostituendo la curva in corrispondenza del punto considerato con la retta passante per questi due punti. La derivata prima della funzione in quel punto viene così approssimata dal coefficiente angolare della retta così definita. Se l'ampiezza dell'intervallo tra i due punti tende a zero, tale procedimento risulta consistente, in quanto al limite si riconduce esattamente alla definizione di derivata. Un ragionamento analogo può essere applicato nel caso dei rapporti incrementali calcolati prendendo un punto precedente a quello considerato. Analogamente, nel caso dei rapporti incrementali centrati, la derivata prima in un punto viene sostituita dal coefficiente angolare della retta che passa per un punto antecedente e un punto successivo rispetto al punto di interesse.

Abbiamo dunque esaminato il significato di consistenza e ne abbiamo fornito una definizione rigorosa. Inoltre, è stato presentato un esempio semplice che illustra in termini euristici tale concetto.

Introduciamo ora un concetto leggermente più articolato, quello di stabilità. Prima di affrontare direttamente la stabilità, è opportuno richiamare brevemente il problema matematico di partenza e rievocare il concetto di problema ben posto secondo la definizione di Hadamard. Questa digressione risulta fondamentale, in quanto apre la strada alla definizione rigorosa del concetto di stabilità numerica.

Analizziamo innanzitutto gli aspetti euristici del concetto di stabilità. La stabilità implica che, in presenza di piccole variazioni nei dati di un problema, tali variazioni determinino soltanto corrispondenti modifiche contenute nella soluzione.

Ad esempio, consideriamo la risoluzione di un'equazione differenziale ordinaria (argomento che tratteremo nel dettaglio nei nuclei successivi del corso) su un determinato intervallo. Supponiamo che il dato iniziale esatto sia u_0 e indichiamo con $u_n(t)$ la soluzione numerica corrispondente a questo dato. Se invece si considerasse un dato iniziale differente v_0 , la soluzione numerica associata sarà indicata con $v_n(t)$. Si ottengono dunque due soluzioni numeriche riferite a due dati iniziali distinti u_0 e v_0 . La distanza tra tali soluzioni viene valutata tramite la differenza, per ogni valore della variabile t , tra $u_n(t)$ e $v_n(t)$. Ci si attende che una piccola differenza tra u_0 e v_0 corrisponda a differenze contenute tra $u_n(t)$ e $v_n(t)$. Qualora questo comportamento si verifichi, il metodo numerico che ha prodotto tali soluzioni sarà definito stabile, per ogni possibile scelta dei dati iniziali u_0 e v_0 .

Vediamo ora la definizione rigorosa di stabilità, richiamando il concetto di ben positura di un problema secondo Hadamard. Un problema si dice ben posto nel senso di Hadamard se la sua soluzione esiste ed è unica ed inoltre dipende in modo continuo dai dati per qualunque scelta possibile degli stessi. In termini formali, un problema matematico del tipo $F(u, d) = 0$ è ben posto (nel senso di Hadamard) se soddisfa le seguenti condizioni:

- esiste una ed una sola soluzione;
- la soluzione dipende in maniera continua dai dati.



Di conseguenza, un problema formulabile in termini astratti come $F(u, d) = 0$ deve ammettere una soluzione unica, la quale deve inoltre dipendere in modo continuo dai dati. La dipendenza continua dai dati significa che, se si considera un dato d la cui soluzione corrispondente è u , quindi tale che $F(u, d) = 0$, e successivamente un altro dato d^* , la cui soluzione associata è u^* , ovvero $F(u^*, d^*) = 0$, allora, se la differenza $d - d^*$ è “piccola”, anche la differenza $u - u^*$ deve essere “piccola”, ossia controllabile in funzione della differenza tra i dati $d - d^*$.

In termini rigorosi, utilizzando il linguaggio dell'analisi matematica, tale condizione si esprime come segue

$$\forall d, \forall \epsilon > 0 \exists K = K(\epsilon, d) > 0: \|d - d^*\| < \epsilon \Rightarrow \|u - u^*\| \leq K,$$

Ciò significa che, per ogni dato d e per ogni $\epsilon > 0$, arbitrariamente piccolo, esiste una costante positiva K , che dipende a priori da ϵ e da d , tale che, se la differenza tra i dati $\|d - d^*\|$ (in una norma opportuna) è inferiore a ϵ , allora anche la differenza tra le soluzioni $\|u - u^*\|$ risulterà inferiore a K . Tale costante K rappresenta il fattore di controllo della dipendenza continua della soluzione dai dati del problema.

Non è sufficiente esprimere la dipendenza continua dai dati mediante la sola formulazione matematica vista. Desideriamo, infatti, introdurre un indicatore quantitativo di tale dipendenza. Sebbene questo aspetto risulti secondario nel contesto del problema puramente analitico, esso assume importanza cruciale nel contesto numerico. Nel problema numerico, alla stabilità delle soluzioni (intesa come dipendenza continua dai dati) deve corrispondere il concetto di convergenza del metodo. Per questo motivo, vedremo pertanto che le costanti K introdotte nella descrizione della dipendenza continua entreranno direttamente nella definizione di convergenza.

Introduciamo ora quello che viene definito numero di condizionamento del problema matematico. Tale grandezza rappresenta un indicatore quantitativo della sensibilità della soluzione rispetto a variazioni nei dati. La sua definizione è data dalla seguente formula

$$K = \sup_{\delta d} \left(\frac{\|\delta u\|}{\|u\|} / \frac{\|\delta d\|}{\|d\|} \right).$$

Essa misura la variazione relativa della soluzione u rispetto a una variazione relativa dei dati d , nel limite in cui la perturbazione δd tende a zero.

La formula, sebbene possa apparire complessa, ha in realtà un'interpretazione piuttosto semplice. Si considera il rapporto tra la variazione della soluzione, indicata con δu , e la variazione dei dati, indicata con δd . In particolare, si ha:

$$\delta u = u - u^*, \quad \delta d = d - d^*,$$

dove u^* e u sono le soluzioni corrispondenti rispettivamente ai dati d e d^* . L'obiettivo è valutare gli errori in termini relativi, poiché questi forniscono informazioni più significative rispetto agli errori assoluti, in quanto tengono conto degli ordini di grandezza in gioco. A tal fine, si considera il rapporto tra l'errore relativo sulla soluzione, $\frac{\|\delta u\|}{\|u\|}$ e l'errore relativo sui dati $\frac{\|\delta d\|}{\|d\|}$. Si prende infine il massimo valore (estremo superiore) di tale rapporto al variare di tutte le possibili variazioni δd .

Procediamo dunque a caratterizzare il rapporto tra le variazioni delle soluzioni e le variazioni dei dati, considerando tutte le possibili variazioni nei dati. Tra questi rapporti, si individua quello massimo, ossia il più grande valore possibile. Tale valore è indicato con K e rappresenta il numero di condizionamento del modello matematico.



Un valore di K contenuto indica che il modello è ben condizionato e, pertanto, le soluzioni risulteranno stabili rispetto alle variazioni dei dati. Al contrario, un valore molto elevato di K implica un potenziale effetto di amplificazione degli errori nei dati, o della loro distanza, sulle soluzioni. Questo aspetto avrà, come si vedrà, importanti ripercussioni dal punto di vista numerico.

Qualitativamente, si può affermare che, se il numero di condizionamento K è molto maggiore di uno, il problema è mal condizionato, mentre se K è approssimativamente pari a uno, il problema è ben condizionato. In sintesi:

K = numero di condizionamento del problema,
 $K \gg 1$ problema mal condizionato,
 $K \simeq 1$: problema ben condizionato.

Non si stanno definendo in questo caso i concetti in modo quantitativo. Infatti, un valore di K approssimativamente uguale a uno può indicare un numero ragionevolmente piccolo. Il numero K fornisce una misura dell'amplificazione potenziale degli errori nei dati sulle soluzioni.

Si intende ora fornire una caratterizzazione numerica operativa del numero di condizionamento K . La definizione precedentemente data, espressa in termini di estremo superiore di un insieme numerico, risulta poco pratica, in quanto si riferisce a un estremo superiore di insiemi. Pertanto, si desidera disporre di una formula semplice che consenta di calcolare esplicitamente K . A tal fine, si introduce una ipotesi: si suppone che il problema matematico iniziale, definito dall'equazione

$$F(u, d) = 0,$$

permetta di esprimere la soluzione u direttamente in funzione del dato d . In altre parole, si assume l'esistenza di una funzione f - da non confondere con F - tale che la soluzione del problema possa essere rappresentata come

$$F(u, d) = 0 \Rightarrow u = f(d).$$

Esaminiamo alcuni esempi che consentano di comprendere come questa situazione sia, in effetti, piuttosto comune nei problemi matematici. Presentiamo alcuni casi particolarmente semplici. A scopo esclusivamente didattico, consideriamo il problema di risolvere l'equazione lineare

$$au + b = 0,$$

dove u è l'incognita ed a e b sono due coefficienti assegnati. La soluzione di questo problema è banalmente

$$u = -\frac{b}{a}, \quad (a \neq 0),$$

dove, in questo caso l'insieme dei dati d è

$$d = \{a, b\},$$

ed $F(u, d)$ è

$$F(u, d) = au + b,$$

da cui, il problema matematico è

$$F(u, d) = 0.$$

La funzione $f(d)$ è quella che, applicata al dato d , restituisce la soluzione u , ovvero $u = f(d)$. Nel caso specifico, si ha quindi



$$u = -\frac{b}{a} \Rightarrow f(d) = -\frac{b}{a}.$$

Un ulteriore esempio, ancora molto semplice, è rappresentato dalle equazioni algebriche di secondo grado

$$u^2 + 2pu + q = 0,$$

dove p e q sono coefficienti noti (assegnati) e u è l'incognita. L'insieme dei dati, in questo caso, è

$$d = \{p, q\}.$$

La funzione $F(u, d)$ è

$$F(u, d) = u^2 + 2pu + q = 0 \Rightarrow F(u, d) = 0,$$

le cui soluzioni sono

$$u_{\pm} = -p \pm \sqrt{p^2 - q},$$

di conseguenza

$$f(d) = -p \pm \sqrt{p^2 - q}, \quad u_{\pm} = f(d).$$

La funzione f , applicata all'insieme dei dati d , restituisce due soluzioni, u_+ e u_- . Pertanto, f può essere considerata una funzione a valori vettoriali, e in questo caso la soluzione u sarà rappresentata da un vettore le cui componenti sono u_+ e u_- , ovvero

$$\mathbf{u} = [u_+, u_-].$$

Consideriamo un terzo esempio, leggermente meno banale, relativo alla risoluzione di un sistema lineare della forma

$$A\mathbf{u} = \mathbf{b},$$

in cui A è la matrice dei coefficienti, \mathbf{b} il vettore termine noto ed \mathbf{u} il vettore incognita. In questo caso

$$d = \{\mathbf{b}, A\}, \quad F(\mathbf{u}, d) = A\mathbf{u} - \mathbf{b} = \mathbf{0},$$

dove d è l'insieme dei dati composto dal termine noto \mathbf{b} e la matrice dei coefficienti A . La soluzione del problema (se A è una matrice non singolare) è

$$\mathbf{u} = A^{-1}\mathbf{b} \Rightarrow \mathbf{u} = f(d),$$

dove A^{-1} è la matrice inversa di A , ovvero quella matrice che moltiplicata per A risulta nella matrice identità (ovvero, la matrice che ha tutti uno sulla diagonale principale e zero altrove).

Abbiamo quindi osservato come, in generale, i problemi matematici definiti dall'equazione

$$F(\mathbf{u}, d) = 0,$$

possano ammettere una rappresentazione esplicita della soluzione \mathbf{u} nella forma

$$\mathbf{u} = f(d).$$

Per questi problemi matematici che possono essere espressi nella forma indicata, si intende caratterizzare il numero di condizionamento mediante una formula di facile applicazione. A tal proposito, ricordiamo lo sviluppo di Taylor, che è valido per funzioni reali di variabile reale, ma si estende in generale anche a funzioni vettoriali.

Supponiamo di considerare due punti: un punto d e un punto $d + \delta d$. Il punto d può essere un numero reale, un vettore o una matrice; pertanto, d e $d + \delta d$ sono elementi di uno spazio, in generale, di dimensione arbitraria n , con $n \geq 1$.

La forma di Taylor, espressa della seguente formula

$$f(d + \delta d) = f(d) + f'(d)\delta d + o(\|\delta d\|)$$

ci assicura che il valore della funzione f nel punto $d + \delta d$ può essere espresso come il valore della funzione nel punto d , più un termine lineare dato da $f'(d)\delta d$, ovvero la derivata di f in d , nel senso appropriato, moltiplicata per l'incremento δd , e un termine residuo $o(\|\delta d\|)$, che è un infinitesimo rispetto alla norma di δd . In particolare, tale termine residuo soddisfa la proprietà



$$o(\|\delta d\|) \rightarrow 0, \quad \text{per } \delta d \rightarrow 0.$$

Un termine o piccolo di una certa quantità indica che il rapporto $\frac{o(\|\delta d\|)}{\|\delta d\|}$ tende a zero al tendere a zero della quantità δd . Pertanto, trascurando tale termine $o(\|\delta d\|)$, si può approssimare la funzione come

$$f(d + \delta d) \simeq f(d) + f'(d)\delta d.$$

Al dato d , corrisponde una soluzione u ed in maniera analoga a $d + \delta d$ corrisponde la soluzione $u + \delta u$
 $u = f(d), \quad u + \delta u = f(d + \delta d),$

da cui

$$\delta u = f(d + \delta d) - u \Rightarrow \delta u = f(d + \delta d) - f(d).$$

Ricordando adesso la definizione del numero di condizionamento K

$$K = \sup_{\delta d} \left(\frac{\|\delta u\|}{\|u\|} / \frac{\|\delta d\|}{\|d\|} \right),$$

da cui

$$K = \sup_{\delta d} \left(\frac{\|f(d + \delta d) - f(d)\|}{\|f(d)\|} / \frac{\|\delta d\|}{\|d\|} \right),$$

ovvero

$$K = \sup_{\delta d} \left(\frac{\|f(d + \delta d) - f(d)\|}{\|\delta d\|} \frac{\|d\|}{\|f(d)\|} \right),$$

in cui abbiamo sostituito $\delta u = f(d + \delta d) - f(d)$ e $u = f(d)$. Naturalmente, si considerano delle norme al fine di poter trattare casi di natura arbitrariamente generale, includendo numeri reali, vettori e matrici.

Sostituendo, mediante l'approssimazione di Taylor

$$f(d + \delta d) \simeq f(d) + f'(d)\delta d,$$

il termine

$$f'(d) \simeq \frac{f(d + \delta d) - f(d)}{\delta d},$$

nell'equazione $K = \sup_{\delta d} \left(\frac{\|f(d + \delta d) - f(d)\|}{\|\delta d\|} \frac{\|d\|}{\|f(d)\|} \right)$, si ottiene la seguente stima per il numero di condizionamento K

$$K \simeq \|f'(d)\| \frac{\|d\|}{\|f(d)\|},$$

dove il segno di approssimazione \simeq è dovuto al fatto che sono stati trascurati i termini infinitesimali di ordine superiore rispetto alla norma $\|\delta d\|$.

Questa formula consente di calcolare il numero di condizionamento K di un problema utilizzando la derivata prima della funzione f , la norma del dato $\|d\|$ e la norma della soluzione $\|f(d)\|$.

Il condizionamento di un problema numerico viene definito in modo analogo a quello del problema matematico continuo. Si consideri, quindi, la sostituzione del problema continuo $F(u, d) = 0$ con il corrispondente problema numerico



$$F_n(u_n, d_n) = 0.$$

In presenza di una perturbazione sui dati, si ha

$$F_n(u_n + \delta u_n, d_n + \delta d_n) = 0.$$

Definiamo allora il numero di condizionamento numerico K_n come

$$K_n = \sup_{\delta d_n} \left(\frac{\|\delta u_n\|}{\|u_n\|} / \frac{\|\delta d_n\|}{\|d_n\|} \right).$$

dove K_n rappresenta l'estremo superiore del rapporto tra l'errore relativo sulla soluzione numerica e l'errore relativo sul dato numerico. Esso fornisce una misura della sensibilità delle soluzioni numeriche rispetto a perturbazioni nei dati.

Infine, si definisce il numero di condizionamento numerico asintotico K^{num} come il limite superiore di K^n al crescere dell'indice n , ovvero

$$K^{num} = \lim_{k \rightarrow \infty} \sup_{n \geq k} K_n.$$

Pertanto, si considerano tutti i numeri di condizionamento numerici K_n , se ne calcola l'estremo superiore per $n \geq k$, e infine si fa tendere k all'infinito. L'obiettivo è individuare una quantità che, in un certo senso, sia indipendente da n . Tale quantità, indicata con K^{num} , viene denominata numero di condizionamento numerico asintotico. Se risulta che

$$K^{num} \approx K,$$

cioè, se il numero di condizionamento numerico asintotico è comparabile al numero di condizionamento del problema matematico, allora si dirà che il metodo numerico è stabile.

Possiamo dunque concludere questa lezione affermando che il numero di condizionamento di un problema numerico è rappresentato dal valore K^{num} , come definito in precedenza. Se questo numero risulta approssimativamente uguale al numero di condizionamento K del problema matematico di partenza, si può affermare che il metodo numerico è stabile. In termini qualitativi, tale risultato può essere interpretato dicendo che il potenziale di amplificazione degli errori introdotto dal metodo numerico è confrontabile con quello intrinseco al problema matematico. In altri termini, a una variazione nei dati corrispondono variazioni nelle soluzioni che risultano, in una certa misura, proporzionali o confrontabili con le variazioni iniziali.

In sintesi, nel corso di questa lezione sono stati introdotti e definiti i concetti di buona posizione di un problema matematico e numerico, consistenza di un problema numerico e stabilità di un metodo numerico.