Zeri di funzioni e teorema di Sturm

Enrico Bertolazzi

Dipartimento di Ingegneria Meccanica e Strutturale
Università degli Studi di Trento
via Mesiano 77, I – 38050 Trento, Italia
`Enrico.Bertolazzi@ing.unitn.it`

1 Eliminazione delle radici multiple

Il metodo di Newton-Raphson è un metodo generalmente del secondo ordine, ma se vogliamo approssimare una radice multipla di un polinomio la convergenza degrada al primo ordine. Un modo per evitare questo degrado delle prestazioni è quello di sostituire al polinomio che contiene radici multiple un polinomio con le stesse radici ma semplici. Questo problema apparentemente complicato ha una soluzione molto semplice: Sia $p(x)$ un polinomio monico fattorizzato come segue[1]

$$p(x) = \prod_{i=1}^{k} (x - x_i)^{m_i},$$

allora il polinomio derivata prima si scriverà come

$$\begin{aligned} p'(x) &= q(x) \prod_{i=1}^{k} (x - x_i)^{m_i - 1}, \\ q(x) &= \sum_{i=1}^{k} m_i \prod_{j=1}^{k}{}^{(i)} (x - x_i), \end{aligned}$$

osserviamo che il polinomio $q(x)$ non ha radici in comune con il polinomio $p(x)$ infatti

$$q(x_\ell) = \sum_{i=1}^{k} m_i \prod_{j=1}^{k}{}^{(i)} (x_\ell - x_i) = m_\ell \prod_{j=1}^{k}{}^{(\ell)} (x_\ell - x_i) \neq 0, \qquad \ell = 1, 2, \ldots, k$$

Quindi il polinomio $p(x)$ e $p'(x)$ hanno in comune solo le radici di $p(x)$ con molteplicità maggiore di 1. Se x_i è una radice comune con molteplicità m_i allora la sua molteplicità in $p'(x)$ sarà $m_i - 1$.

[1] nella formula $\prod_{j=1}^{k}{}^{(i)}$ significa che vengono fatti tutti i prodotti con indice j che va da 1 a k escluso i

Ricordiamo che un polinomio $M(x)$ è il massimo comun divisore tra due polinomi $P(x)$ e $Q(x)$ se vale

1. $M(x)$ divide $P(x)$ e $Q(x)$, cioè

$$P(x) = M(x)S(x), \qquad Q(x) = M(x)T(x), \tag{1}$$

dove $S(x)$ e $T(x)$ sono a loro volta polinomi (anche di grado 0)

2. Se $N(x)$ è un altro polinomio che divide $P(x)$ e $Q(x)$ allora $N(x)$ divide $M(x)$ cioè

$$M(x) = W(x)N(x),$$

per $W(x)$ opportuno polinomio.

Teorema 1. *Se $M(x)$ è il massimo comun divisore tra due polinomi $P(x)$ e $Q(x)$ e se α è una radice comune a $P(x)$ e $Q(x)$ cioè:*

$$P(\alpha) = 0, \qquad Q(\alpha) = 0,$$

allora necessariamente α è anche una radice di $M(x)$.

Dimostrazione. Se così non fosse dalla (1) avremo che il polinomio $\mathbf{x} - \alpha$ divide sia $S(x)$ che $T(x)$ cioè

$$P(x) = M(x)(\mathbf{x} - \alpha)S^1(x), \qquad Q(x) = M(x)(\mathbf{x} - \alpha)T^1(x),$$

e $M(x)(\mathbf{x} - \alpha)$ sarebbe un divisore comune a $S(x)$ e $T(x)$ che non divide $M(x)$. ■

Teorema 2. *Se α è una radice di $M(x)$ massimo comun divisore tra i polinomi $P(x)$ e $Q(x)$ allora α è una radice comune a $P(x)$ e $Q(x)$ cioè:*

$$P(\alpha) = 0, \qquad Q(\alpha) = 0,$$

Dimostrazione. dalla (1) segue immediatamente

$$\begin{aligned} P(\alpha) &= M(\alpha)S(\alpha) = 0, \\ Q(\alpha) &= M(\alpha)T(\alpha) = 0, \end{aligned}$$

Conseguenza di questi due teoremi è che il massimo comun divisore tra due polinomi è costituito dal prodotto delle radici in comune con molteplicità il minimo tra le due.

Esempio 1. Siano

$$P(x) = 3(x-1)^3(x+1)^2(x-3),$$
$$Q(x) = 2(x-1)^2(x+1)(x-3)^3,$$

il loro massimo comun divisore è

$$M(x) = (x-1)^2(x+1)(x-3).$$

Sia $m(x)$ il massimo comun divisore tra i polinomi $p(x)$ e $p'(x)$ allora per i discorsi precedenti vale

$$m(x) = \prod_{i=1}^{k}(x-x_i)^{m_i-1}.$$

Allora il polinomio

$$\frac{p(x)}{m(x)} = \prod_{i=1}^{k}(x-x_i),$$

conterrà tutte le radici di $p(x)$ e solo radici semplici. Il problema è come trovare questo massimo comun divisore. Per fare questo ci viene in aiuto un algoritmo classico, l'algoritmo di Euclide (Euclide di Alessandria circa 365–300 a.C.) per il calcolo del massimo comun divisore.

Algorithm *Algoritmo di Euclide per il M.C.D. di $p(x)$ e $q(x)$*
Input: $p(x)$, $q(x)$
1. **if** $\partial p > \partial q$
2. **then** $p_0 \leftarrow p$; $p_1 \leftarrow q$
3. **else** $p_0 \leftarrow q$; $p_1 \leftarrow p$
4. $i \leftarrow 1$
5. **repeat**
6. (∗ p_{i+1} è il resto della divisione di p_{i-1} con p_i. ∗)
7. (∗ $p_{i-1}(x) = s_i(x)p_i(x) + p_{i+1}(x)$ ∗)
8. $i \leftarrow i+1$
9. **until** $p_i \equiv 0$
10. (∗ p_{i-1} è il massimo comun divisore. ∗)

Esempio 2. Sia dato

$$\begin{aligned} p(x) &= (x-1)^3(x-2)(x+1)^2, \\ &= x^6 - 3x^5 + 6x^3 - 3x^2 - 3x + 2, \\ p'(x) &= 6x^5 - 15x^4 + 18x^2 - 6x - 3, \end{aligned}$$

calcoliamo il massimo comun divisore $m(x)$ con l'algoritmo di Euclide:

1. Inizializzazione:
$$p_0(x) = x^6 - 3x^5 + 6x^3 - 3x^2 - 3x + 2,$$
$$p_1(x) = 6x^5 - 15x^4 + 18x^2 - 6x - 3,$$
2. Calcolo p_2:
$$\begin{aligned} p_0(x) &= p_1(x)s_1(x) + p_2(x), \\ x^6 - 3x^5 + 6x^3 - 3x^2 - 3x + 2 &= \left(6x^5 - 15x^4 + 18x^2 - 6x - 3\right)\left(\frac{x}{6} - \frac{1}{12}\right) \\ &+ \left(\frac{7}{4} - 3x - \frac{x^2}{2} + 3x^3 - \frac{5}{4}x^4\right). \end{aligned}$$
3. Calcolo p_3:
$$\begin{aligned} p_1(x) &= p_2(x)s_1(x) + p_3(x), \\ 6x^5 - 15x^4 + 18x^2 - 6x - 3 &= \left(-\frac{24x}{5} + \frac{12}{25}\right)\left(\frac{7}{4} - 3x - \frac{x^2}{2} + 3x^3 - \frac{5}{4}x^4\right) \\ &+ \frac{96}{25}\left(-x^3 + x^2 + x - 1\right). \end{aligned}$$
4. Calcolo p_4:
$$\begin{aligned} p_2(x) &= p_3(x)s_1(x) + p_4(x), \\ 6x^5 - 15x^4 + 18x^2 - 6x - 3 &= \left(\frac{125x}{384} - \frac{175}{384}\right)\frac{96}{25}\left(-x^3 + x^2 + x - 1\right) + 0. \end{aligned}$$
5. Poiché $p_4 \equiv 0$ il massimo comun divisore è p_3
$$\frac{96}{25}\left(-x^3 + x^2 + x - 1\right),$$
   e poiché è unico a meno di una costante moltiplicativa scegliamo
$$m(x) = -x^3 + x^2 + x - 1.$$

dividendo $p(x)$ per $m(x)$ otteniamo
$$\frac{p(x)}{m(x)} = -\frac{x^3}{2} + x^2 + \frac{x}{2} - 1.$$

Per questioni computazionali (ad esempio nella costruzione di successioni di Sturm) a volte è più conveniente usare una versione modificata dell'algoritmo tenendo cento del fatto che se $m(x)$ è il massimo comun divisore tra due polinomi anche $cm(x)$ lo è dove c è un qualunque scalare non nullo. La versione è la seguente

Algorithm *Algoritmo di Euclide per il M.C.D. di $p(x)$ e $q(x)$ (versione modificata)*

Input: $p(x)$, $q(x)$

if $\partial p > \partial q$
  then $p_0 \leftarrow p$; $p_1 \leftarrow q$
  else $p_0 \leftarrow q$; $p_1 \leftarrow p$
$i \leftarrow 1$
repeat
    (* p_{i+1} è il resto della divisione di p_{i-1} con p_i. *)
    (* $p_{i-1}(x) = s_i(x)p_i(x) - c_i p_{i+1}(x)$ *)
    $i \leftarrow i + 1$
until $p_i \equiv 0$
(* p_{i-1} è il massimo comun divisore *)

Osserviamo che l'algoritmo *Algoritmo di Euclide per il M.C.D. di $p(x)$ e $q(x)$* produce effettivamente il massimo comun divisore. Vale infatti il seguente teorema:

Teorema 3. *L'algoritmo Algoritmo di Euclide per il M.C.D. di $p(x)$ e $q(x)$ termina in un numero finito m di passi e l'ultimo resto non nullo è il massimo comun divisore dei polinomi.*

Dimostrazione. Innanzitutto osserviamo che l'algoritmo termina in un numero finito di passi. Infatti il resto della divisione ha grado strettamente minore del divisore e quindi ad ogni divisione si riduce il grado dei polinomi coinvolti. Quando il grado è 0 il polinomio è uno scalare. La divisione di un polinomio per uno scalare ha resto nullo e quindi anche in questo caso estremo l' algoritmo termina. L'algoritmo può essere quindi scritto come segue

$$
\begin{aligned}
p_0(x) &= s_1(x)p_1(x) - c_1 p_2(x), && (1.0)\\
p_1(x) &= s_2(x)p_2(x) - c_2 p_3(x), && (1.1)\\
&\vdots\\
p_{k-1}(x) &= s_k(x)p_k(x) - c_k p_{k+1}(x), && (1.k)\\
&\vdots\\
p_{m-2}(x) &= s_{m-1}(x)p_{m-1}(x) - c_{m-1} p_m(x), && (1.m-1)\\
p_{m-1}(x) &= s_m(x)p_m(x), && (1.m)
\end{aligned}
$$

e segue subito che $p_m(x)$ l'ultimo resto non nullo divide $p_{m-1}(x)$, $p_{m-2}(x)$,..., $p_0(x)$ e quindi è un divisore comune. Viceversa sia $q(x)$ un divisore di $p_0(x)$ e $p_1(x)$, allora dalla (1.0) segue se $q(x)$ divide $p_2(x)$ e dalla (1.1) segue che $q(x)$ divide $p_3(x)$ e così via fino ad arrivare alla (1.m) da cui segue che $q(x)$ divide $p_m(x)$ e quindi $p_m(x)$ è il massimo comun divisore. ∎

1.1 Separazione delle radici

E‘ desiderabile avere il modo per conoscere il numero di *radici reali* in un dato intervallo. Se questo è possibile è possibile tramite un metodo di bisezione separare tutte le radici reali e approssimarle fino alla approssimazione voluta. Questo problema può essere risolto grazie alle successioni che prendono il nome dal suo scopritore Jacques Charles Francois Sturm 1803–1855:

Definizione 1. La successione di funzioni continue definite su $[a, b]$:

$$\mathcal{F}(x) = \{f_0(x), f_1(x), \dots, f_m(x)\},$$

forma una *successione di Sturm su* $[a, b]$ se vale:

1. $f_0(x)$ ha al più zero semplici in $[a, b]$;
2. $f_m(x)$ non ha zeri su $[a, b]$;
3. per ogni $\alpha \in [a, b]$ tale che $f_k(\alpha) = 0$, allora $f_{k-1}(\alpha)f_{k+1}(\alpha) < 0$;
4. per ogni $\alpha \in [a, b]$ tale che $f_0(\alpha) = 0$, allora $f_0'(\alpha)f_1(\alpha) < 0$;

Definizione 2. Data una successione di Sturm $\mathcal{F}(x) = \{f_0(x), f_1(x), \dots, f_m(x)\}$ per ogni punti x associamo un vettore di $m + 1$ numeri reali. A questo vettore possiamo associare un numero intero detto variazione di segno. Questo numero rappresenta il numero di volte che scorrendo la successione di numeri c'è un cambio di segno. Ad esempio la successione

$$\{\underbrace{1, 0, -2,}_{*} \underbrace{-3, 4,}_{*} 3, 0, 1\},$$

ha 2 variazioni di segno (marcate con $*$). Osserviamo che $3, 0, 1$ ha variazione di segno nulla. Infatti lo 0 va considerato come elemento neutro e va rimosso dal computo.

Per ogni successione di Sturm vale il seguente teorema:

Teorema 4 (Sturm). *Data una successione di Sturm* $\mathcal{F}(x) = \{f_0(x), f_1(x), \dots, f_m(x)\}$ *su* $[a, b]$ *il numero di zeri di* $f_0(x)$ *in* (a, b) *è dato dalla differenza tra il numero di variazioni di segno in* $\mathcal{F}(b)$ *e* $\mathcal{F}(a)$.

Dimostrazione. Il numero di variazioni di segno cambia man mano che x passa da a a b solo a causa del cambio di segno di qualche funzione della successione di Sturm. Assumiamo che per un qualche $\hat{x} \in (a, b)$ valga $f_k(\hat{x}) = 0$ per $0 < k < m$. In un intorno di $\hat{x}$ dalla condizione 3 sarà

x	$f_{k-1}(x)$	$f_k(x)$	$f_{k+1}(x)$
$\hat{x} + \epsilon$	$+$	$\pm$	$-$
$\hat{x}$	$+$	0	$-$
$\hat{x} - \epsilon$	$+$	$\pm$	$-$

oppure

x	$f_{k-1}(x)$	$f_k(x)$	$f_{k+1}(x)$
$\hat{x} + \epsilon$	$-$	$\pm$	$+$
$\hat{x}$	$-$	0	$+$
$\hat{x} - \epsilon$	$-$	$\pm$	$+$

in ogni caso il passaggio da $\hat{x}$ non cambia il numero di variazioni di segno. Sia ora $\hat{x}$ uno zero di $f_0(x)$ e vediamo la variazione di segno nell'intorno di $\hat{x}$, osserviamo che dalla condizione 4 della definizione 1 avremo

x	$f_0(x)$	$f_1(x)$
$\hat{x}+\epsilon$	$+$	$+$
$\hat{x}$	0	$+$
$\hat{x}-\epsilon$	$-$	$+$

oppure

x	$f_0(x)$	$f_1(x)$
$\hat{x}+\epsilon$	$-$	$-$
$\hat{x}$	0	$-$
$\hat{x}-\epsilon$	$+$	$-$

in ogni caso in numero delle variazioni di segno cresce al passare di x per uno zero di $f_0(x)$. Combinando queste osservazioni otteniamo il teorema. ■

1.1.1 Costruzione della successione di Sturm per un polinomio

E‘ relativamente facile costruire una successione di Sturm per un polinomio. Sia $f_0 \equiv p_n$ un polinomio di grado n, definiamo $f_1 \equiv -p'_n$. Dividiamo $f_0(x)$ per $f_1(x)$ e chiamiamo il resto $-f_2(x)$. Poi dividiamo $f_1(x)$ per $f_2(x)$ e chiamiamo il resto $-f_3(x)$. Continuiamo così finché il procedimento termina. Abbiamo così la successione:

$$\begin{aligned} f_0(x) &= q_1(x)f_1(x) - f_2(x), \\ f_1(x) &= q_2(x)f_2(x) - f_3(x), \\ &\vdots \\ f_{k-1}(x) &= q_k(x)f_k(x) - f_{k+1}(x), \\ &\vdots \\ f_{m-2}(x) &= q_{m-1}(x)f_{m-1}(x) - f_m(x), \\ f_{m-1}(x) &= q_m(x)f_m(x). \end{aligned} \tag{2}$$

Questa successione a parte il segno di f_i è esattamente la successione per il calcolo del massimo comun divisore di un polinomio, e f_m è il massimo comun divisore tra f_0 e f_1. La successione di polinomi

$$p_i(x) = \frac{f_i(x)}{f_m(x)}, \qquad i = 0, 1, \ldots, m$$

è una successione di Sturm, infatti

1. $p_0(x)$ ha al più zero semplici in $[a, b]$; essendo il rapporto tra il polinomio originario ed il massimo comun divisore tra il polinomio originario e la sua derivata prima.

2. $p_m(x)$ non ha zeri su $[a, b]$; infatti è una costante.

3. per ogni $\alpha \in [a,b]$ tale che $p_k(\alpha) = 0$, allora $p_{k-1}(\alpha)p_{k+1}(\alpha) < 0$; infatti dalla (2) abbiamo

$$p_{k-1}(\alpha) = -p_{k+1}(\alpha).$$

Osserviamo che se $p_{k-1}(\alpha) = p_k(\alpha) = p_{k+1}(\alpha) = 0$ allora dalla (2) seguirebbe

4. per ogni $\alpha \in [a,b]$ tale che $p_0(\alpha) = 0$, allora $p_0'(\alpha)p_1(\alpha) = -p_0'(\alpha)^2 < 0$;

1.2 Limitazione delle radici

Per poter usare il metodo di bisezione e separare le radici con Sturm è necessario conoscere una prima stima anche se grossolana dell' intervallo in cui stanno tutte le radici di un polinomio. Per fare questo occorre osservare che la seguente matrice in forma di Frobenius (Ferdinand Georg Frobenius 1849–1917)

$$\mathbf{A} = \begin{bmatrix} 0 & 1 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 1 & & \vdots \\ \vdots & & & \ddots & \ddots & 0 \\ 0 & 0 & 0 & \cdots & 0 & 1 \\ \dfrac{a_0}{a_n} & -\dfrac{a_1}{a_n} & \dfrac{a_2}{a_n} & \cdots & (-1)^{n-1}\dfrac{a_{n-2}}{a_n} & (-1)^n\dfrac{a_{n-1}}{a_n} \end{bmatrix},$$

ha come polinomio caratteristico

$$p(\lambda) = |\,\mathbf{A} - \lambda\mathbf{I}\,| = \frac{(-1)^n}{a_n}\left(a_0 + a_1\lambda + a_2\lambda^2 + \cdots + a_{n-2}\lambda^{n-2} + a_{n-1}\lambda^{n-1} + a_n\lambda^n\right), \tag{3}$$

(basta sviluppare lungo l' ultima riga). Applicando il teorema di Gershgorin alla matrice otteniamo che le radici del polinomio (3) soddisfano

$$\left|\lambda - \frac{a_{n-1}}{a_n}\right| \leq \left|\frac{a_0}{a_n}\right| + |a_1| + \cdots + \left|\frac{a_{n-2}}{a_n}\right|, \qquad |\lambda| \leq 1.$$

Applicando il teorema di Gershgorin alla matrice trasposta otteniamo che le radici del polinomio (3) soddisfano

$$|\lambda| \leq \left|\frac{a_0}{a_n}\right|, \qquad |\lambda| \leq \left|\frac{a_i}{a_n}\right| + 1, \qquad \left|\lambda - \frac{a_{n-1}}{a_n}\right| \leq 1.$$

Possiamo applicare le disegualianze appena viste (indebolendole un poco) ad un polinomio qualunque,

$$p(x) = a_0 + a_1x + a_2x^2 + \cdots + a_{n-1}x^{n-1} + a_nx^n,$$

ottenendo le seguenti limitazioni

$$|x| \leq \max\left\{1, \sum_{i=0}^{n-1}\left|\frac{a_i}{a_n}\right|\right\},$$

$$|x| \leq \max\left\{\left|\frac{a_0}{a_n}\right|, 1+\max_{i=1}^{n-1}\left|\frac{a_i}{a_n}\right|\right\}.$$