Soluzione della prima prova parziale di Calcolo Numerico del 1/4/2016

Esercizio 1

(a) I numeri di macchina positivi hanno la forma

$$x = 2^p(.1d_2d_3\cdots d_t)$$

questo è un numero intero se $p \geq t$, oppure $d_{p+1} = d_{p+2} = \ldots = d_t = 0$. Quindi:

per $p = 1$ il solo numero intero è $2(0.1)_2 = 1$

per $p = 2$ abbiamo $2^2(.10) = 2$, e $2^2(.11) = 3$ quindi 2 configurazioni

per $p = k < t$ abbiamo 2^{k-1} poichè la prima cifra è necessariamente 1.

per ogni $p \geq t$ abbiamo 2^{t-1} interi, quindi in totale $2^{t-1}(M - t + 1)$, poiché per ognuno degli $M - t$ possibili esponenti abbiamo 2^{t-1} possibili mantisse.

In totale i numeri di macchina interi positivi sono quindi $2^{t-1}(M-t+1)+\sum_{k=1}^{t-1} 2^{k-1} = (M-t+1)2^{t-1}+2^{t-1}-1$ e quindi considerando anche zero ed i numeri interi negativi otteniamo che i numeri di macchina interi sono

$$1 + (M - t + 1)2^t + 2(2^{t-1} - 1) = (M - t + 2)2^t - 1, \quad \text{se } M \geq t.$$

Se invece $M < t$ abbiamo che i numeri di macchina sono $1+2\sum_{k=1}^{M} 2^{k-1} = 1 + 2(2^M - 1) = 2^{M+1} - 1$.

(b) Un numero in base 2 è dispari se la cifra delle unità è 1. Si osserva che dato $x = 2^p(.1d_2 \ldots d_t)_2$ se $p > t$ il numero è necessariamente pari, perchè la cifra delle unità è zero. Vanno allora considerati solo i numeri con $p \leq t$.

Sia $M \geq t$,

se $p = t$ abbiamo che i numeri dispari hanno la forma $x = 2^t(.1d_2d_3 \ldots d_{t-1}1)$ e sono in totale 2^{t-2},

se $p = k < t, k > 1$ abbiamo $2^k(.1d_2 \ldots d_{k-1}1)$ e quindi 2^{k-2} possibili mantisse.

se $p = 1$ abbiamo solo il numero $2^1(.1) = 1$.

Sommando abbiamo che i numeri positivi dispari sono $1+\sum_{k=2}^{t} 2^{k-2} = 2^{t-1}$.

(c) I numeri dispari sono in generale meno di quelli pari perchè se $M > t$ tutti i numeri con esponente $p > t$ sono pari. Ad esempio se $M = 4$ e $t = 3$, gli interi par risultano $(M-t-1)2^{t-1}+2^{t-1}-1 = 2*2^2+2^2-1 = 8+4-1 = 11$, mentre ci sono solo $2^{t-1} = 4$ interi positivi dispari.

Per $M < t$ si hanno 2^{M-1} numeri dispari.

Esercizio 2

(a) Il coefficiente di amplificazione risulta

$$c_x = \frac{x}{f(x)} f'(x) = \frac{x}{x(1+x)}(1+2x) = \frac{1+2x}{1+x}.$$

Poichè $|c_x|$ tende a $+\infty$ per $x \to -1$, abbiamo che il problema è malcondizionato in un intorno di -1. Non abbiamo invece problemi di malcondizionamento per $x \to \pm\infty$ poiché $\lim_{x\to\pm\infty} |c_x| = 2$.

(b) L'errore algoritmico del primo metodo è dato da

$$\epsilon^{(1)} = \epsilon_3 + \frac{(1+x)^2}{f(x)}\epsilon_2 + (\frac{(1+x)^2}{f(x)} + 1)\epsilon_1,$$

dove ϵ_1 è l'errore locale dovuto al calcolo di $x+1$, ϵ_2 è dovuto al calcolo di $(1+x)^2$ ed ϵ_3 è dovuto alla sottrazione. Poiché $\lim_{x\to 0} \left|\frac{(1+x)^2}{f(x)}\right| = \infty$ l'algoritmo non é stabile in un intorno di 0.

L'errore algoritmico del secondo metodo è dato da

$$\epsilon^{(2)} = \epsilon_2 + \epsilon_1,$$

dove ϵ_1 è l'errore locale dovuto al calcolo di $x+1$, ϵ_2 è dovuto al calcolo di $x(1+x)$. Poichè $|\epsilon_i| < u$, abbiamo che $|\epsilon^{(2)}| < 2u$ quindi l'algoritmo è sempre stabile, ed è quindi da preferire al primo algoritmo per ogni valore di x.

Esercizio 3

(a) I punti fissi di $g(x)$ sono tali che

$$(1-k)x = \log(x+1)$$

Attraverso il metodo della separazione grafica si vede subito che

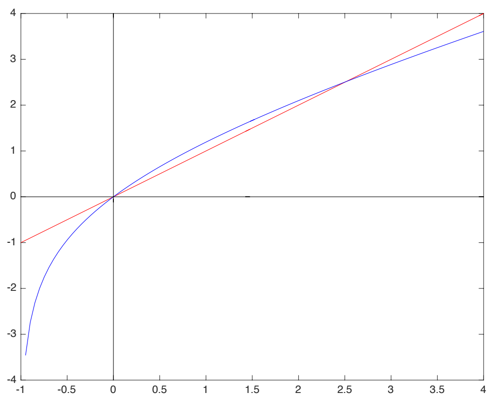
per $k = 0$ abbiamo la sola soluzione $\alpha = 0$,

per $1-k > 1$ cioè $k < 0$, 2 soluzioni, $\alpha = 0$ e $\beta < 0$.

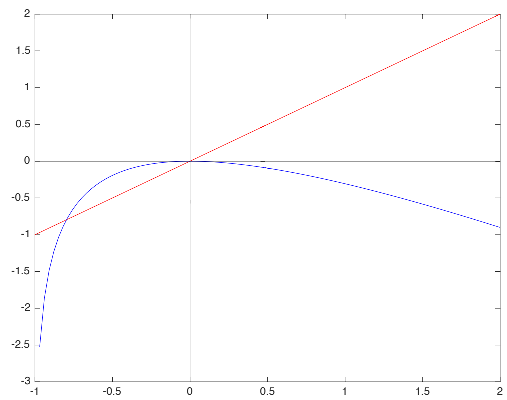
per $0 < k < 1$ due soluzioni $\alpha = 0$ e $\beta > 0$.

per $k \geq 1$ la sola soluzione $\alpha = 0$.

(b) Per $k = 1/2$ abbiamo che $g(x) = 1/2x + \log(x+1)$ ha 2 soluzioni $\alpha = 0$ e $\beta > 0$. Poiché $g(2) = 1 + \log(3) > 2$ e $g(3) = 3/2 + \log(4) < 3$, abbiamo che $2 < \beta < 3$. Abbiamo che $g'(x) = 1/2 + \frac{1}{x+1}$, quindi $g'(0) = 3/2 > 1$. Non abbiamo quindi convergenza alla soluzione $\alpha = 0$. Si osserva che poiché $g"(x) = -\frac{1}{(1+x)^2} < 0$ $g'(x)$ risulta una funzione decrescente. Possiamo usare questa informazione per stimare $g'(\beta)$ che sarà tale che $g'(2) > g'(\beta) > 0$. Abbiamo che $g'(2) = 5/6 < 1$ e quindi anche $0 < g'(\beta) < 1$ che ci assicura la convergenza locale con ordine 1. Osservando il grafico



abbiamo convergenza monotona crescente a partire da ogni $x_0 \in (0, \beta]$ e decrescente per ogni $x_0 > \beta$. Per $k = -1$ abbiamo oltre alla soluzione $\alpha = 0$ anche la soluzione $-0.9 < \beta < -1/2$. $g(x)$ risulta $g(x) = -x + \log(1+x)$. Il grafico risulta



Dall'analisi di $g'(x) = -1 + \frac{1}{(1+x)}$ abbiamo che $g'(0) = 0$ e quindi convergenza alla soluzione $\alpha = 0$, inoltre il metodo ha ordine 2. Abbiamo convergenza monotona ad α per ogni$x_0 > \beta$. Inoltre se indichiamo con $\bar{x}$, $\bar{x} > 0$ il punto tale che $g(\bar{x}) = \beta$, abbiamo convergenza a α anche per ogni $0 < x_0 < \bar{x}$.

Poiché $g"(x) < 0$, $g'(x)$ risulta una funzione decrescente, quindi $g'(\beta) > g'(-1/2) = 1$. Non abbiamo quindi convergenza a β poiché $g'(\beta) > 1$.