DIRO
IFT 2425

EXAMEN INTRA

Max Mignotte

DIRO, Département d'Informatique et de Recherche Opérationnelle, local 2377
Http : //www.iro.umontreal.ca/∼mignotte/ift2425/
E-mail : mignotte@iro.umontreal.ca

Date : *24/02/2022*

I	Amplification d'erreur (30 pts)
II	Erreur en arithmétique flottante (24 pts)
III	Recherche des racines d'une équation (59 pts)
Total	113 points.

TOUS DOCUMENTS PERSONNELS, CALCULATRICES ET CALCULATEURS AUTORISÉS

I. Amplification d'erreur (30 **pts**)

Soit la formule suivante :

$$z = x^{y^w}$$

que l'on peut lire de la façon suivante ; x à la puissance y (avec y qui est lui-même à la puissance w). Dans cette formule, $x=4$, $y=3$ et $w=2$ sont des valeurs imprécises ou approximées avec (pour ces trois variables) 0.1% d'erreur relative.

1. En utilisant la méthode de propagation d'erreur (basé sur l'approximation de Taylor de la fonction $z(x,y,w)$ au premier ordre), donner l'approximation de z et son incertitude (en l'exprimant par $z = z^\star \pm \Delta z$. Servez-vous ensuite de ce calcul d'erreur pour souligner les chiffres significatifs "exacts" (cse) de la précédente approximation. Arrondir finalement cette approximation au nombre de cse adéquat et donner son incertitude relative.
   <14 pts>
2. Peut-on utiliser la méthode de la fourchette pour obtenir cette incertitude Δz ? Expliquer clairement pourquoi et si oui, utiliser cette méthode (de la fourchette) pour obtenir une estimation de Δz et retrouver l'incertitude relative de z. Rappeler aussi quelle critique peut-on faire à la méthode de la question précédente par rapport à cette méthode de la fourchette.
   <11 pts>
3. Supposons que x soit une variable imprécise avec 2% d'erreur relative (et toujours $x = 4$ comme approximation). Quelle erreur relative peut-on se permettre de faire sur la variable y et w ($y = 3$ et $w = 2$ comme valeur approximée) pour que la contribution de l'erreur de ces trois variables soit identique (*i.e.*, de même importance) dans la contribution finale (*i.e.*, totale) de l'erreur contenue dans z ?
   <5 pts>

Réponse

$\boxed{1}$

On a donc $\Delta x = 0.004$, $\Delta y = 0.003$ et $\Delta w = 0.002$.

La valeur de z approximée ($z^\star$) est la valeur de z en x, y et w approximée, *i.e.*, $z^\star = 4^{3^2} = 262144$ et pour calculer l'incertitude de $z(x,y,w)$, fonction de ses trois variables incertaines, en se souvenant du fait que $y^w = \exp(w \ln(y))$, et donc :

$$z(x,y,w) = x^{y^w} = \exp\Big((\ln x)\cdot \exp[w\,(\ln y)]\Big)$$

On doit maintenant calculer la différentielle de $z(.)$:

$$\begin{aligned}
\Delta z(x,y,w) &= \left|\frac{\exp\big(w\,(\ln y)\big)}{x}\, z\right|\cdot \Delta x \;+\; \left|(\ln x)\,\frac{w}{y}\exp\big(w\,(\ln y)\big)\, z\right|\cdot \Delta y \;+\; \left|(\ln x)\,(\ln y)\,\exp\big(w\,(\ln y)\big)\, z\right|\cdot \Delta w \\
&= \left|\frac{y^w}{x}\, z\right|\cdot \Delta x \;+\; \left|(\ln x)\,\frac{w}{y}\, y^w\, z\right|\cdot \Delta y \;+\; \left|(\ln x)\,(\ln y)\, y^w\, z\right|\cdot \Delta w
\end{aligned}$$

Numériquement, on obtient :

$$\begin{aligned}
\Delta z(x,y,w) &= \left|\frac{3^2}{4}\cdot 262144\right|\cdot 0.004 \;+\; \left|\ln(4)\cdot\frac{2}{3}\cdot 3^2\cdot 262144\right|\cdot 0.003 \;+\; \left|\ln(4)\cdot\ln(3)\cdot 3^2\cdot 262144\right|\cdot 0.002 \\
&\approx 2\,359.296 + 6541.357 + 7186.415 \approx 16087 < 0.5\times 10^5
\end{aligned}$$

Le rang du dernier cse est le rang 5 (celui des dizaines de milliers), donc une façon bien rigoureuse d'écrire cette variable incertaine z en arrondissant correctement au nombre de cse adéquat est d'écrire finalement :

$$z = \underline{2}60\,000 \pm 17000$$

et le résultat est exprimé avec $(16087/262144) \approx$ 6.15% d'erreur relative.
<14 pts>

$\boxed{\mathbf{2}}$

La fonction $z(x,y,w) = x^{y^w}$ est continue et monotone (croissante) lorsque les trois variables imprécises x, y et w vont croitre dans leur intervalle de confiance. Du coup, il sera facile d'obtenir, sans ambiguïté, l'intervalle de confiance (ou de variation) de la variable imprécise z, ce qui veut dire que l'on peut donc appliquer la méthode de la fourchette (et vérifier si on ne s'est pas trompé à la question précédente).

Les variables x, y et w sont contenues dans les intervalles de confiance suivants $x \in [3.996 \quad 4.004]$, $y \in [2.997 \quad 3.003]$ et $w \in [1.998 \quad 2.002]$ et :

$$\begin{aligned} z(x,y,w) &\in [x_{\min}^{y_{\min}^{w_{\min}}} \; , \; x_{\max}^{y_{\max}^{w_{\max}}}] \\ &\in [246\,578.9062 \quad , \quad 278\,778.3983] \end{aligned}$$

La connaissance de cet intervalle de confiance nous permet de trouver facilement $\Delta z \approx 16100 < 0.5 \times 10^5$, soit aussi une estimation de $z = \underline{2}60\,000 \pm 17000$ avec un cse comme dans la première question et une erreur relative de $(17000/262144 \approx 6.15\%$ d'erreur relative.

Cette méthode (de la fourchette) est la plus précise car la méthode de propagation d'erreur basée sur l'approximation de Taylor reste une approximation. Plus précisément, la méthode de propagation d'erreur est basée sur une approximation de Taylor du premier ordre (et donc sur une hypothèse de linéarité de la fonction z dans les voisinages ou intervalles de confiance définis précédemment). Cependant, même dans le cas d'une fonction très fortement non linéaire comme notre fonction double puissance $z = x^{y^w}$, son approximation par une linéarité reste très bonne et on observe des résultats quasi équivalents !
<11 pts>

Nota : Cet exemple nous montre encore toute la puissance de la méthode de propagation d'erreur basée sur l'approximation de Taylor qui peut s'appliquer, je le rappelle dans tous les cas et pas seulement lorsqu'on peut obtenir, comme dans ce cas particulier, l'intervalle de confiance (ou de variation) de la variable finale imprécise z, c-a-d, la borne min ou max de z à partir des bornes min et max des différentes variables de la fonction $z(.)$.

$\boxed{\mathbf{3}}$

Le plus simple est d'utiliser la différentielle de $z(.)$ obtenu à la question 1. qui exprime la contribution des trois erreurs dans l'erreur totale faite sur $z(.)$:

$$\begin{aligned} \Delta z(x,y,w) &= \left|\frac{3^2}{4} \cdot 262144\right| \cdot \Delta x \; + \; \left|\ln(4) \cdot \frac{2}{3} \cdot 3^2 \cdot 262144\right| \cdot \Delta y \; + \; \left|\ln(4) \cdot \ln(3) \cdot 3^2 \cdot 262144\right| \cdot \Delta w \\ &= \underbrace{589\,824 \cdot \Delta x}_{=47\,185.92 \text{ dans le cas où } \Delta x = 0.08} \; + \; 2\,180\,452.494 \cdot \Delta y \; + \; 3\,593\,207.857 \cdot \Delta w \end{aligned}$$

Pour une même contribution entre ces trois termes/erreurs, il faut donc que :

$$\begin{aligned} 2\,180\,452.494 \cdot \Delta y = 47\,185.92 \quad &\text{et} \quad 3\,593\,207.857 \cdot \Delta w = 47\,185.92 \\ \text{c-a-d,} \qquad \Delta y = 0.021640425 \quad &\text{et} \quad \Delta w = 0.013131976 \end{aligned}$$

Donc pour une erreur relative $(\Delta x/x)$ de 2%, il y aura une contribution de même poids ou de même importance pour les trois erreurs dans l'erreur totale faite sur $z(.)$ lorsque l'erreur relative $\Delta y/y = 0.021640425/3 \approx 0.72\,\%$ pour y et une erreur relative de $\Delta w/w = 0.013131976/2 \approx 0.66\,\%$ pour w.
<5 pts>

II. Erreur/Amplification d'Erreur en Arithmétique Flottant (24 pts)

1. Parmi les nombres suivants, dire ceux qui sont plus imprécis ou moins imprécis ou aussi imprécis en terme d'erreur relative d'affectation en notation flottante simple précision que 0.1 et donner une justification qui soit claire :

   $$0.4 \qquad 0.025 \qquad 0.5 \qquad 0.0625$$

   <4 pts>
2. On veut estimer le terme suivant en notation flottante pour des valeurs élevées de n :

   $$0.1^n - 0.1^{n+1}$$

   Identifier et citer tous les problèmes numériques associés au calcul de cette expression en notation flottante (pour des valeurs élevées de n). Proposer une expression mathématiquement équivalente qui permettrait d'augmenter la précision du calcul de cette expression en notation flottante.
   <5 pts>
3. Expliquer pourquoi le calcul numérique de ces trois expressions :
   (a) $\frac{1}{1-\exp(-x)} - \frac{1}{1+\exp(-x)} \qquad x \neq 0$
   (b) $\sin(2x) - 2\sin(x)\cos(x)$
   (c) $\frac{1}{(1-\sin x)} - \frac{1}{(1+\sin x)} \qquad x \neq \pi/2$

   peuvent conduire à des problèmes d'erreurs numériques (identifier aussi pour quelle valeur de x). Identifier et citer (tous) le(s) problème(s) numérique(s) associé(s) à chacune de ces expressions et proposer une expression mathématiquement équivalente (ou dans le pire des cas, lorsqu'on ne peut absolument pas, une expression approximativement équivalente) qui permettrait d'éviter ces problèmes numériques et/ou augmenter la précision des calculs de ces trois expressions.
   <15 pts>

Réponse

$\boxed{1}$

▷ 0.4 est aussi imprécis en terme d'erreur relative d'affectation en notation flottante simple précision que 0.1 car sa valeur s'exprime par un décalage de 2 bits vers la gauche (son erreur absolue en terme d'affectation en notation flottante est multipliée par 4 et son erreur absolue est aussi multipliée par 4 et même imprécision en terme d'erreur relative que 0.4).

▷ Même chose pour 0.025 mais pour un décalage de 2 bits vers la droite et une valeur et une erreur absolue divisée cette fois par 4 (mais une même erreur relative que 0.1).

▷ 0.5 à une représentation exacte et 0.0625 aussi (qui correspond au codage de 0.5 divisée par 8, ou un décalage de 3 ($8 = 2^3$) bits vers la droite.

<4 pts>

$\boxed{2}$

En notation flottante, le nombre 0.1 n'a pas de représentation exacte (il faudrait un nombre infini de bits pour le représenter parfaitement). Du coup 0.1^n ou 0.1^{n+1} n'auront pas non plus de représentation exacte et seront approximés ou entachés d'erreur numérique. Pour de fortes valeurs de n, on se retrouve de plus, avec le problème de la soustraction de deux nombres approximés quasi égaux (dans notre cas proche de 0) qui amplifie les erreurs numériques et de plus pour de très fortes valeurs de n on aura le problème de l'UNDERFLOW, la valeur 0.1^n ou 0.1^{n+1} seront approximés à 0.

On peut cependant réécrire facilement cette expression de telle façon que tous ces problèmes s'atténuent ou disparaissent de la façon suivante (noté que 9 et 10 ont des représentations binaires exactes) :

$$0.1^n - 0.1^{n+1} = 0.1^n(1 - 0.1) = \frac{0.9}{10^n} = \frac{9}{10^{n+1}}$$

<5 pts>

Nota : Sur mon ordinateur, cette notation flottante, avec $n=10$, on obtient pour l'expression non corrigée $9.00000143855E\!-11$ et pour l'expression corrigée $9.00000005077E\!-11$ (expression la plus précise que l'on puisse avoir).

$\boxed{\textbf{3}}$

[-a-] Pour la première expression, on aura un problème d'annulation de cse (chiffres significatifs exacts) due à la soustraction de deux nombres approximés quasi égaux. Cette annulation de cse se passe lorsque $x \to +\infty$. Dans ce cas aussi, on a le problème de l'UNDERFLOW car $\exp(-x)$ risque de s'approximer à zéro. Dans le cas où $x \to +\infty$, il vaudrait mieux donc réécrire cette expression de la façon suivante :

$$\frac{1}{1-\exp(-x)} - \frac{1}{1+\exp(-x)} = \frac{2\exp(-x)}{(1-\exp(-x))(1+\exp(-x))} = \frac{2}{\exp(x) - \exp(-x)}$$

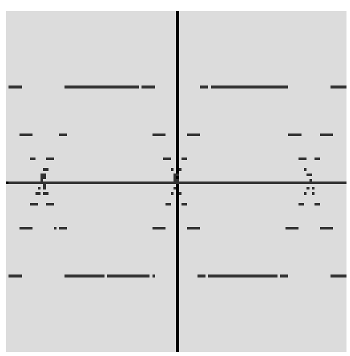
<5 pts>

Nota : Sur mon ordinateur, cette notation flottante, avec $x = 40$, on obtient pour l'expression non corrigée 0 et pour l'expression corrigée $8.49670826228e-18$ (expression la plus précise que l'on puisse avoir).

[-b-] Pour la deuxième expression, puisque $\sin(2x) = 2\sin(x)\cos(x)$, on peut réécrire cette expression par 0 et on aura donc des erreurs numériques dues à la différence de ces deux fonctions qui ont été approximées dans l'ordinateur de deux façons différentes.

<5 pts>

Nota : Normalement on devrait avoir 0 alors que l'on obtient pour cette expression $\sin(2x) - 2\sin(x)\cos(x)$, un véritable monstre numérique dont j'ai tracé la représentation pour pour $x \in [-2 \;\; 2]$ et $y \in [-2 * 1E\!-16 \;\; 2 * 1E\!-16]$.



[-c-] Pour cette expression, on aura un problème d'annulation de cse (chiffres significatifs exacts) due à la soustraction de deux nombres approximés quasi égaux. Cette annulation de cse se passe lorsque $x \to 0$. Il vaudrait mieux réécrire cette expression pour qu'elle soit mathématiquement équivalente et ne fasse plus intervenir ce problème de la façon suivante :

$$\frac{1}{(1-\sin x)} - \frac{1}{(1+\sin x)} = \frac{(1+\sin x) - (1-\sin x)}{(1-\sin x)(1+\sin x)} = \frac{2\sin x}{\cos^2 x}$$

<5 pts>

Nota -1- : $\frac{2\sin x}{\cos^2 x} = \frac{2\tan x}{\cos x}$

Nota -2- : Pour $x = 1E{-}15$, on obtient pour l'expression non corrigée $2.22044604925E{-}15$ et pour l'expression corrigée $2.00000000725E{-}15$ (expression la plus précise que l'on puisse avoir).

III. Recherche des racines d'une équation (59 pts)

On se propose de trouver numériquement dans R une valeur approchée de la racine de la fonction :

$$f(x) = x\,e^x - (1/2) = 0 \tag{1}$$

Où $e^x = \exp(x)$ désigne la fonction exponentielle.

1. Méthode du point fixe
   (a) Après avoir montré qu'il existe une racine **unique** de f dans l'intervalle $[0\ 1]$, donner une forme $x = g_1(x)$ mathématiquement équivalente pour $f(x) = 0$ pour laquelle la suite itérative $r_{n+1} = g_1(r_n)$ converge avec $r_0 = 0.5$, valeur initiale de cette suite et donner les 4 premières estimées $r_1, \ldots, r_4$, obtenue en partant de r_0.
   <12 pts>
   (b) En utilisant la suite itérative $r_{n+1} = g_1(r_n)$ de la question précédente avec toujours $r_0 = 0.5$. Déterminer une borne supérieure du nombre d'itérations nécessaires pour obtenir, par la suite itérative $r_{n+1} = g_1(r_n)$ de la question précédente (avec toujours $r_0 = 0.5$), une valeur approchée de r avec 7 cse après la virgule (commencer par utiliser le théorème de la valeur moyenne comme en démonstration).
   <10 pts>
   (c) Donner une forme $x = g_2(x)$ mathématiquement équivalente pour $f(x) = 0$ pour laquelle la suite itérative $r_{n+1} = g_1(r_n)$ avec $r_0 = 0.5$ ne converge pas et démontrer pourquoi.
   <4 pts>
   (d) Montrer de deux façons différentes (et sans utiliser le nombre d'itérations nécessaires pour obtenir la convergence) que le point fixe $x = g_1(x)$ convergera plus rapidement que la méthode de bissection.
   <8 pts>
2. Méthode de la Bissection
   (a) Donner le nombre d'itérations qui serait nécessaire à la méthode de la bissection pour arriver avec cet intervalle à une estimation de la racine avec 7 cse après la virgule (sans estimer explicitement cette racine par la méthode de la bissection) et en partant toujours de l'intervalle $[0\ 1]$.
   <5 pts>
3. Méthode de Newton
   (a) Donner la relation itérative $r_{n+1} = g_{\text{newt.}}(r_n)$ intervenant dans la méthode itérative de Newton pour la résolution itérative de la racine r de l'équation $f(x) = 0$.
   <5 pts>
   (b) Montrer qu'avec $r_0 = 0.5$, la relation itérative de Newton converge.
   <5 pts>
   (c) Calculer les 4 premières estimées $r_1, \ldots, r_4$, en partant de $r_0 = 3$.
   <5 pts>
   (d) Obtenez une valeur approchée de la racine de $f(x)$ comprise dans J en utilisant l'expansion en série de Taylor de $f(x)$ jusqu'au terme en x^2 (en supposant cette racine proche de 0). Utiliser ensuite cette valeur pour estimer r_0 et calculer les 3 premières estimées $r_1, \ldots, r_3$ de la suite itérative $r_{n+1} = g_{\text{newt.}}(r_n)$ en partant de ce r_0.
   <5 pts>

Réponse

$\boxed{\textbf{1.(a)}}$

Il faut donc démontrer que sur l'intervalle $J = [0\ \ 1]$, $f(a)$ et $f(b)$ sont de signe opposé et démontrer ensuite que la fonction f est continue et **surtout monotone** pour démontrer qu'il existe une racine **unique** de f contenue dans cet intervalle. On trouve $f(0) = -0.5$ et $f(1) \approx 2.218$, qui sont bien de signe opposé. Maintenant, il reste à démontrer qu'il existe une racine unique dans cet intervalle J. $f(x)$ est une fonction continue donc au moins une racine est assurée d'y être. De plus :

$$f'(x) = e^x + x\,e^x = e^x(x+1)\ > 0 \quad \forall x \in [0\ \ 1] = J$$

qui montre qu'il existe une racine **unique** de f dans l'intervalle $J = [0\ 1]$. De plus, nous avons :

$$f(x) = 0 \quad \Leftrightarrow \quad x\,e^x - (1/2) = 0 \quad \Leftrightarrow \quad x = (1/2)e^{-x} = g_1(x)$$

avec $|g_1'(x)| = |-1/2e^{-x}| < 1$ dans l'intervalle $J = [0\ \ 1]$. Donc avec cette suite itérative, et puisque $r_0 = 0.5$ est contenu dans cet intervalle J on est assuré de converger vers la racine unique de f. En partant de $r_0 = 0.5$, on a, $r_n = g_1(r_{n-1})$ et :

$$\begin{aligned} r_1 &= 0.303265329 \\ r_2 &= 0.369201575 \\ r_3 &= 0.345643025 \\ r_4 &= 0.353882548 \end{aligned}$$

<12 pts>

Nota : la valeur a convergence avec le maximum de précision de ma calculatrice est $r_{20} = 0.351733711$

$\boxed{\textbf{1.(b)}}$

Procédons comme en démonstration numéro 2. En utilisant le théorème de la valeur moyenne, on obtient, puisque $g(r) = r$ et $r_n = g(r_{n-1})$ avec r_n la valeur approchée de la racine à la n-ième itération :

$$\begin{aligned} r_n - r &= \frac{g(r_{n-1}) - g(r)}{(r_{n-1} - r)} \times (r_{n-1} - r) \\ &= g_1'(\zeta) \times (r_{n-1} - r) \qquad \text{avec} \quad \zeta \in J \end{aligned}$$

En utilisant le fait que $|g_1'(\zeta)| < 1/2$, on a :

$$|r_n - r| \leq \frac{1}{2}\,|r_{n-1} - r| \leq \left(\frac{1}{2}\right)^2 |r_{n-2} - r| \leq \quad \ldots \quad \leq \left(\frac{1}{2}\right)^n \underbrace{|r_0 - r|}_{<1/2} \leq \left(\frac{1}{2}\right)^{n+1}$$

On obtiendra donc $|r_n - r| < 0.5 \times 10^{-7}$ (7 cse), des que :

$$\left(\frac{1}{2}\right)^{n+1} < \ 0.5 \times 10^{-7}$$

C'est-à-dire pour $n = 24$. Comme il s'agit d'une borne supérieure, on pourrait l'exprimer par $n \leq 24$. On doit donc faire (au plus) 24 itérations pour obtenir (surement), par la suite itérative $r_{n+1} = g_1(r_n)$ (avec toujours $r_0 = 0.5$), une valeur approchée de r avec une précision de 7 cse après la virgule.
<10 pts>

Nota : Dans ce raisonnement, on a considéré le pire des cas pour estimer le nombre d'itérations nécessaires, c'est a dire estimer une borne supérieure pour ce nombre d'itérations. En effet, on a dit que dans

le pire des cas, $g_1'(\zeta) = 1/2$ lorsque $\zeta \in J = [0 \quad 1]$ et commencer le raisonnement en utilisant ce pire des cas. Pour s'en convaincre, si on fait itérer la suite, on obtient 7 cse pour r après 15 itérations en obtenant $r_{15} = 0.351733689$ (la valeur a convergence avec le maximum de précision de ma calculatrice est $r_{19} = 0.351733711$).
Si on voulait calculer le nombre d'itérations exact nécessaire il faudrait remplacer dans le raisonnement $g_1'(\zeta)$ par $g_1'(r)$ ($g_1'(.)$ à convergence), *i.e.*, $g_1'(r = 0.35173368) = 0.351733722$ ce qui nous aurait permis de trouver qu'en 16 itérations $(r_0, \ldots, r_{15})$, on obtiendrait la précision voulue demandée. Mais cela demanderait de connaitre d'abord r et si on connaît r exactement, on connaît aussi le nombre d'itérations nécessaires pour l'obtenir.

1.(c)

$$f(x) = 0 \quad \Leftrightarrow \quad x\,e^x - (1/2) = 0 \quad \Leftrightarrow \quad x = (1/2) - \ln(x) = g_2(x)$$

avec $|g_2'(x)| = |-1/x| > 1$ dans l'intervalle $[0 \quad 1]$ donc cette suite itérative on n'est pas assuré de converger vers la racine unique de f qui se trouve dans l'intervalle $[0 \quad 1]$. En partant de $r_0 = 0.5$, on a, $r_n = g_2(r_{n-1})$ et :

$$\begin{aligned} r_1 &= 1.193147181 \\ r_2 &= 0.323405494 \\ r_3 &= 1.628848343 \\ r_4 &= 0.012126773 \end{aligned}$$

qui diverge (trois itérations plus tard, ma calculatrice donne MATH ERROR).
<4 pts>

Nota :
▷ Une autre possibilité aurait été de donner : $f(x)=0 \Leftrightarrow x\,e^x - (1/2) + x = x$ qui aurait donné : $g_2(x) = x\,e^x + x - (1/2)$, pour laquelle $|g_2'(x)| = |e^x(x+1) + 1| > 1$ dans l'intervalle $[0 \quad 1]$ donnant ainsi une suite itérative qui aurait été aussi divergente.
▷ Une autre possibilité aurait été de donner : $f(x)=0 \Leftrightarrow x = \ln(0.5/x)$, pour laquelle $|g_2'(x)| = |-1/x| > 1$ dans l'intervalle $[0 \quad 1]$ donnant ainsi une suite itérative qui aurait été aussi divergente.

1.(d)

Les deux méthodes (point fixe et bissection) sont d'ordre linéaire. Pour savoir si $x = g_1(x)$ est plus rapide que la méthode de bissection, la première façon consiste à calculer sa constante asymptotique : $|g_1'(r)| = |-1/2e^{-r}| = |-1/2e^{-0.351733711}| \approx 0.3517$. Comme $|g_1'(r)| < 0.5$, on a une procédure itérative plus rapide que la méthode de la bissection. On aurait pu le montrer avec la valeur estimée à la quatrième itération $|g_1'(r_4 = 0.35)| \approx 0.35 < 0.5$

La deuxième méthode va utiliser la relation $|e_{n+1}| = C\,|e_n|$ pour une relation itérative d'ordre linéaire avec $e_n = r_n - r$. Utilisons les résultats de la question 1.(a), *i.e.*, $r_1 = 0.303265329$, $r_2 = 0.369201575$, $r_3 = 0.345643025$, $r_4 = 0.353882548$ et le fait que $r = 0.351733711$. On a $|e_1| = 0.048468382$, $|e_2| = 0.017467864$, $|e_3| = 0.006090686$, $|e_4| = 0.002148837$.
Maintenant, faisons les rapports $|e_{n+1}|/|e_n|$

$$\begin{aligned} |e_2|/|e_1| &= 0.017467864/0.048468382 \approx 0.360397093 \\ |e_3|/|e_2| &= 0.006090686/0.017467864 \approx 0.348679495 \\ |e_4|/|e_3| &= 0.002148837/0.006090686 \approx 0.352807056 \end{aligned}$$

Puisqu'on a la relation $|e_{n+1}| = C\,|e_n|$ pour une relation itérative d'ordre linéaire, on a donc $C \in$ l'intervalle de confiance $[0.348679495 \quad 0.360397093]$ pour lequel $C < 0.5$ et qui montre que cette méthode du point fixe est plus rapide (en fait $1.42 \approx 0.5/0.35$ fois plus rapide) que la méthode de la bissection.

<8 pts>

2.(a)

L'intervalle de longueur 1 se divise de moitié par la méthode de la bissection donc il faut que celui-ci en largeur soit :

$$\frac{1}{2^n} < 0.5 \times 10^{-7} \Leftrightarrow n > \frac{\log 2 + 7 \log 10}{\log 2} = 24.25 \quad \text{soit} \;\; 25 \;\; \text{itérations}$$

<5 pts>

Nota : Contrairement à la question 1.(b), il s'agit non pas d'une borne supérieure mais du nombre exact d'itérations pour arriver à la précision voulue sur r.

3.(a)

La fonction $f(x)$ est dérivable sur $\mathbb{R}$ et la méthode itérative de Newton permet d'écrire (cf. question 1.(a)) :

$$r_{n+1} = r_n - \frac{f(r_n)}{f'(r_n)} = r_n - \frac{r_n\, e^{r_n} - 0.5}{e^{r_n} + r_n\, e^{r_n}} = r_n - \frac{r_n - 0.5\, e^{-r_n}}{1 + r_n} = g_{\text{newt.}}(r_n)$$

<5 pts>

3.(b)

Dans ce cas, le plus simple est de montrer que la fonction $f(x)$, sur l'intervalle $J = [0, 1]$ (cf. question 1.(a)), intervalle dans lequel on prendra $r_0 = 1/2$ et dans lequel une racine unique s'y trouve, ne présente pas d'*extrema*, i.e., de valeurs pour laquelle, $f'(x)$ s'annule. Dans notre cas, la dérivée a été déjà calculé à la question 1.(a) ; $f'(x) = e^x(x+1)$ qui ne s'annule jamais sur $J = [0 \;\; 1]$ donc aucun problème, la méthode de Newton va converger.
<5 pts>

Nota : Il aurait été beaucoup beaucoup plus difficile de démontrer que cette suite converge à partir de la condition de convergence du point fixe, *i.e.*, en démontrant que $|g'_{\text{newt.}}(\xi)| < 1 \quad \forall \xi \in J = [0, 1]$.

3.(c)

Remarquons que $r_0 = 3$ ne fait pas partie de $J = [0 \;\; 1]$, en partant de cette valeur de départ, on a, $r_n = g_{\text{newt.}}(r_{n-1})$ et :

$$\begin{aligned} r_1 &= 2.256223384 \\ r_2 &= 1.579411487 \\ r_3 &= 1.007047086 \\ r_4 &= 0.596294865 \end{aligned}$$

<5 pts>

Nota -1- :
En en ayant la curiosité d'itérer encore...

$$\begin{aligned} r_5 &= 0.395285354 \\ r_3 &= 0.353328747 \\ r_4 &= 0.351735921 \\ r_5 &= 0.351733711 \quad \text{max. de présision de ma calculatrice} \\ r_6 &= 0.351733711 \end{aligned}$$

Nota -2- : Très très peu d'étudiant (en fait un seul) ont eu la curiosité de calculer ces trois itérations pour $r_0 = 0.5$!. C'est vrai, je ne le demandais pas... mais cela en valait la peine par curiosité et pas seulement pour un **BONUS de 5 points supplémentaires :**

En partant de $r_0 = 0.5$, on a, $r_n = g_{\text{newt.}}(r_{n-1})$ et :

$$
\begin{aligned}
r_1 &= 0.368843553 \\
r_2 &= 0.351984973 \\
r_3 &= 0.351733766 \\
r_4 &= 0.351733711 \quad \text{valeur a convergence avec le maximum de précision de ma calculatrice}
\end{aligned}
$$

$\boxed{\textbf{3.(d)}}$

$$
\begin{aligned}
f(x) = x\,e^x - (1/2) &= 0 \\
x\big(1 + x + o(x^2)\big) - (1/2) &= 0 \\
x^2 + x - (1/2) + o(x^3) &= 0 \\
\Leftrightarrow \quad x_{1,2} &= -1/2 \pm \sqrt{3}/2 \approx 0.366025403 \quad \text{ou} \quad -1.077350269
\end{aligned}
$$

En prenant la valeur positive $x \approx 0.366025403$ qui est dans J et en itérant trois fois $r_n = g_1(r_{n-1})$, on obtient :

$$
\begin{aligned}
r_1 &= 0.351909405 \\
r_2 &= 0.351733738 \\
r_3 &= 0.351733711
\end{aligned}
$$

qui convergera en trois itérations ; en 2 itérations avec 7 cse et en 3 itérations avec toute la précision de ma calculatrice.

<5 pts>