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 : *24/02/2024*

I	Erreur/Amplification d'erreur en arithmétique flottante (24 pts)
II	Erreur en arithmétique flottante (30 pts)
III	Recherche des racines d'une équation (56 pts)
Total	110 points.

TOUS DOCUMENTS PERSONNELS, CALCULATRICES ET CALCULATEURS AUTORISÉS

I. Amplification d'erreur (24 pts)

Soit la formule suivante :

$$z = \frac{a^b}{b^a}$$

Dans cette formule, $a\!=\!2$ et $b\!=\!3$ sont des valeurs imprécises ou approximées avec (pour ces deux variables) 10% d'erreur relative.

1. En utilisant la méthode de propagation d'erreur (basée sur l'approximation de Taylor de la fonction $z(a,b)$ 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.
   <12 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.
   <12 pts>

Réponse

$\boxed{\textbf{1.}}$

On a donc $\Delta a\!=\!0.2$ et $\Delta b\!=\!0.3$.

La valeur de z approximée ($z^\star$) est la valeur de z en a et b approximée, *i.e.*, $z^\star = 2^3/3^2 = 8/9 \approx 0.\overline{8}$ et pour calculer l'incertitude de $z(a,b)$, fonction de ses deux variables incertaines, en se souvenant du fait que $a^b = \exp(b\ln(a))$:

$$z(a,b) = \frac{a^b}{b^a} = \frac{\exp\big(b\,\ln(a)\big)}{\exp\big(a\,\ln(b)\big)} = \exp\big(b\,\ln(a) - a\,\ln(b)\big)$$

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

$$\begin{aligned}\Delta z(a,b) &= \Big|\big(\frac{b}{a} - \ln(b)\big)\cdot\exp\big(b\,\ln(a) - a\,\ln(b)\big)\Big|\cdot\Delta a \;+\; \Big|\big(\ln(a) - \frac{a}{b}\big)\cdot\exp\big(b\,\ln(a) - a\,\ln(b)\big)\Big|\cdot\Delta b \\ &= \;\Big|\big(\frac{b}{a} - \ln(b)\big)\cdot z\Big|\cdot\Delta a \;+\; \Big|\big(\ln(a) - \frac{a}{b}\big)\cdot z\Big|\cdot\Delta b\end{aligned}$$

Numériquement, on obtient :

$$\begin{aligned}\Delta z(a,b) &= \underbrace{\Big|\big(\frac{3}{2} - \ln(3)\big)\cdot\Big(\frac{8}{9}\Big)\Big|}_{0.356789}\cdot 0.2 \;+\; \underbrace{\Big|\big(\ln(2) - \frac{2}{3}\big)\cdot\Big(\frac{8}{9}\Big)\Big|}_{0.023538}\cdot 0.3 \\ &\approx\; 0.071357815 + 0.0070614703 \approx 0.078419285 < 0.5\times 10^0\end{aligned}$$

Le rang du dernier cse est le rang 0 (celui des unités), donc une façon bien rigoureuse d'écrire cette variable incertaine z en arrondissant correctement au nombre de cse adéquat est d'écrire finalement :

$$z = \tfrac{8}{9} \pm 0.08 = \underline{0}.9 \pm 0.08 = \text{ou mieux} = \underline{1}.0 \pm 0.1$$

et le résultat est exprimé avec $(0.078419285/(8/9)) \approx$ 8.83% d'erreur relative.
<12 pts>

Nota : On doit toujours simplifier le plus possible l'expression avant de calculer sa différentielle. On pouvait écrire la différentielle de la façon suivante (équivalente) :

$$\Delta z(a,b) = \left|\left(a^{b-1}\, b^{-a}\, [b - a\ln(b))\right|\cdot\Delta a \;+\; \left|\left(a^{b}\, b^{-a-1}\, [b\ln(a) - a]\right)\right|\cdot\Delta b$$

ou

$$\left|\frac{(ba^{b-1}) - a^b(\ln(b)\, b^a}{(b^a)^2}\right|\cdot\Delta a \;+\; \left|\frac{(\ln(a)\, a^b\, (b^a) - (a^b(ab^{a-1})}{(b^a)^2}\cdot\Delta b\right.$$

ou

$$\left|\frac{-a^{b-1}\,(\ln(b)\, a - b)}{b^a}\right|\cdot\Delta a \;+\; \left|\frac{b^{-a-1}(\ln(a)\, b - a)}{a^{-b}}\right|\cdot\Delta b$$

On obtient, ce qui est normal, des différentielles très symmétriques par rapport aux variables a et b

$\boxed{\mathbf{2}}$

On a donc $a\in[a_{\min}\,,\,a_{\max}]$ et $b\in[b_{\min}\,,\,b_{\max}]$. Il existe deux raisonnement possibles :

• Si on considère a comme étant la variable et $b\!=\!3$ (comme étant une constante) :
on a $\;\triangleright\;\; z'(a,b\!=\!3) = [b\, a^{b-1} - \ln(b)\, a^b]/b^a = [3a^2 - \ln(3)\, a^3]/3^a > 0$ (pour $a\!=\!2$).
et si on considère b comme étant la variable et $a\!=\!2$ (comme étant une constante) :
on a $\;\triangleright\;\; z'(a\!=\!2,b) = [a^b\,\ln(a) - b^{-1}\, a^{b+1}]/b^a = [\ln(2)\, 2^b - (2^{b+1}/b)]/b^2 > 0$ (pour $b\!=\!3$).

Donc la fonction $z(a,b)$ est croissante pour des valeurs croissantes de a et b donc :

$$z(a,b) \;\in\; [z_{\min} = a_{\min}^{b_{\min}}/b_{\min}^{a_{\min}} = 0.811099577 \quad z_{\max} = a_{\max}^{b_{\max}}/b_{\max}^{a_{\max}} = 0.975582899]$$

• Les fonctions a^b et b^a sont continues et monotones sur l'intervalle de confiance des deux valeurs imprécises. Du coup, comme la fonction $z(a,b) = a^b/b^a$ est une simple opération linéaire ; (*i.e.* ; une division) entre ces deux fonctions, cela veut dire que la fonction $z(.)$ est continue et monotone et donc que l'on peut appliquer la méthode de la fourchette. Dans ce cas de simple combinaison (division) de fonctions monotones, il nous reste à trouver, parmi l'ensemble : $\{a_{\min}^{b_{\min}}/b_{\min}^{a_{\min}}\,,\, a_{\min}^{b_{\max}}/b_{\max}^{a_{\min}}\,,\, a_{\max}^{b_{\min}}/b_{\min}^{a_{\max}}\,,\, a_{\max}^{b_{\max}}/b_{\max}^{a_{\max}}\}$ laquelle de ces bornes est $z_{\min}$ et laquelle est $z_{\max}$ définissant ainsi l'intervalle de confiance $[z_{\min}\;\; z_{\max}]$ de la variable z. Comme on a numériquement : $a_{\min}^{b_{\min}}/b_{\min}^{a_{\min}} \approx 0.811$, $a_{\min}^{b_{\max}}/b_{\max}^{a_{\min}} \approx 0.945$, $a_{\max}^{b_{\min}}/b_{\min}^{a_{\max}} \approx 0.818$, $a_{\max}^{b_{\max}}/b_{\max}^{a_{\max}} \approx 0.975$. On a donc :

$$z(a,b) \;\in\; [z_{\min} = a_{\min}^{b_{\min}}/b_{\min}^{a_{\min}} = 0.811099577 \quad z_{\max} = a_{\max}^{b_{\max}}/b_{\max}^{a_{\max}} = 0.975582899]$$

La connaissance de cet intervalle de confiance nous permet de trouver facilement $\Delta z \approx (z_{\max} - z_{\min})/2 = (0.975582899 - 0.811099577)/2 = 0.164483322/2 = 0.082241661 < 0.5 \times 10^0$, soit aussi une estimation de $z = \underline{0}.9 \pm 0.09$ avec un cse comme dans la première question et une erreur relative de $\approx 9.25\%$.

<12 pts>

Nota : 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 z, son approximation par une linéarité reste encore très bonne et on observe des résultats quasi équivalents ! On rappelle aussi que la méthode de propagation d'erreur basée sur l'approximation de Taylor peut s'appliquer dans tous les cas. Dans cet exemple particulier où on peut obtenir l'intervalle de confiance (ou de variation) de la variable finale imprécise z (*i.e.*, la borne min ou max de z à partir des bornes min et max des différentes variables de la fonction $z(.)$), on peut appliquer la méthode de la fourchette qui est à la fois plus simple mais aussi exacte (non approximée).

II. Erreur/Amplification d'Erreur en Arithmétique Flottante (30 pts)

1. Pour sensibiliser les utilisateurs de machines informatiques que le calcul numérique en notation flottante se fait majoritairement à partir de valeurs approximées (et donc entachées d'erreurs), on aimerait faire le programme en langage C suivant qui, affiche à l'écran :

   $>$ Stockage en float de ? :

   Attend une valeur numérique (avec l'instruction en langage C : scanf), par exemple : 0.85 (entrée au clavier) puis, l'appuie sur la touche $\boxed{\text{RETURN}}$), permet d'afficher à l'écran :

   $>$ 0.850000 est arrondi en float par 0.850000023841858 avec une erreur de 0.000003 pour 100

   Avec 0.850000023841858 la véritable approximation de 0.85 codé (ou représenté) en float avec 12 chiffres après la virgule et 0.000003 pour cent, la vraie erreur relative de cette approximation. Bref, si on entre la valeur 0.1, on devrait avoir comme résultat de ce mini programme :

   $>$ Stockage en float de ? : 0.1 $\boxed{\text{RETURN}}$
   $>$ 0.100000 est arrondi en float par 0.100000001490116 avec une erreur de 0.000001 pour 100

   Faire ce programme en langage C.
   <10 pts>

2. Expliquer pourquoi le calcul numérique de ces quatre expressions en notation flottante :

   (a) $\frac{(1+x)-1}{x} \qquad x\neq 0$
   (b) $\frac{\cos x-\cos 2x}{\sin x+\sin 2x} \qquad x\neq 0$
   (c) $\sqrt{1+x}-\sqrt{1+x^2}$
   (d) $\sin x-(x-\frac{x^3}{6})$

   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 quatres expressions.
   <20 pts>

Réponse

$\boxed{\textbf{1.}}$

Si on considère que :

• [-1-] cette même valeur stockée en double (*i.e.* ; float double précision) est stockée sans erreur (avec une représentation avec 12 chiffres après la virgule comme exigée dans l'énoncé [et une erreur relative exprimée avec une précision en double)], et

• [-2-] en se souvenant que, en C (ou C++), *“% f”* , *“% . 15f”* et *“% lf”* permet d'afficher respectivement un flottant en version arrondi et en version “avec 15 chiffres après la virgule”, et un double en version arrondi, on a donc tout simplement le programme suivant :

```
. double dval ;
. printf("Stockage en float de ? : ") ;
. scanf("% lf",&dval) ;
. float fval=dval ;
. double ErrPourc=fabs((dval-fval)/dval)*100.0 ;
. printf("%f est en fait arrondi  en float par %.15f avec une
.            erreur de % lf pour 100",fval,fval,ErrPourc) ;
```

<10 pts>

Une autre solution (un peu moins simple) MAIS qui n'utilise pas les DOUBLE et faisant référence à l'astuce utilisée dans l'exercice I de l'examen Intra-H23 (avec la notion de PREC*ision*).

```
. int PREC=1000000 ;
. float nbfloat ;
. printf("Stockage en float de ? : ") ;   scanf("%f",&nbfloat) ;
. float bestnbfloat=(float)(((int)(nbfloat*PREC))/PREC) ;
. printf("%f est en fait arrondi  en float par %.15f avec une
.            erreur de %f pour %d",fval,fval,(nbfloat-bestnbfloat)*100.PREC) ;
```

$\boxed{\textbf{2.}}$

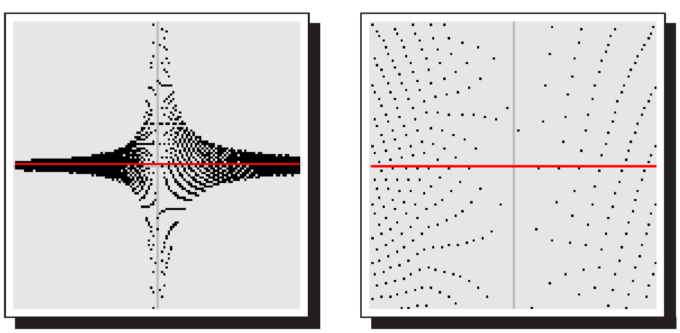
[-a-]

Pour la première expression, lorsque $x \to 0$, on a successivement et tout d'abord :

1. un problème numérique d'ADDITION DE DEUX NOMBRE FLOTTANTS D'ORDRE DE GRANDEUR DIFFÉRENTES avec possiblement un UNDERFLOW lorsque x tend trop vers 0 (x arrondi à 0 par possible manque de précision flottante) puis,
2. la SOUSTRACTION DE DEUX NOMBRES APPROXIMÉS QUASI ÉGAUX (entre $(1+x)$ et 1) (ou PROBLÈME D'ANNULATION DE CHIFFRES SIGNIFICATIFS EXACTS)
3. puis finalement, (et toujours) lorsque x tend trop vers 0, un possible problème d'UNDERFLOW conduisant, dans ce cas, à une EXPRESSION INDÉTERMINÉE et finalement à un NaN (NOT A NUMBER).

Dans tous les cas (et donc pas seulement lorsque $x \to 0$), on a tout intérêt, bien sûr, à réécrire cette expression comme étant tout simplement égale à 1.

<5 pts>



De gauche à droite : expression brute ou original en FLOAT au voisinage de 0 pour x (en noire) montrant un superbe MONSTRE NUMÉRIQUE et expr. math. equiv. en (rouge) pour l'expression : lorsque $[x \in \pm 1.0 \cdot 10^{-4}]$ et $[y \in 1.0 \pm 1.0 \cdot 10^{-4}]$, et lorsque $[x \in \pm 1.0 \cdot 10^{-5}]$ et $[y \in 1.0 \pm 1.0 \cdot 10^{-5}]$, Donc, en résumé, analytiquement, on a la courbe en rouge mais numériquement (ou sur ordinateur et en FLOAT), on a la courbe en noire.

[-b-]

Pour la deuxième expression, lorsque $x \to 0$ (mod. $2\,k\,\pi$ [avec k entier]).

on a au numérateur la SOUSTRACTION DE DEUX NOMBRES APPROXIMÉS QUASI ÉGAUX et un possible UNDERFLOW au dénominateur conduisant à une EXPRESSION INDÉTERMINÉE du type division par zéro.

On propose deux solutions ;

• Une optimale dans laquelle on se propose de trouver une expression mathématiquement équivalente minimisant ces problèmes numériques :

$$\frac{\cos x - \cos 2x}{\sin x + \sin 2x} = \frac{2\sin(x/2)\,\sin(3x/2)}{2\cos(x/2)\,\sin(3x/2)} = \tan(x/2)$$

• Une solution légèrement sous-optimale **(moitié des points [3/5])** dans laquelle on remplace l'expression originale au voisinage de 0 par une expression approximativement mathématiquement équivalente, qui atténuerait fortement ces problèmes numériques, en utilisant un développement limité au voisinage de zéro (par ex. d'ordre 2) :

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

<5 pts>

[-c-]

Pour la troisième expression, on aura un problème d'annulation de cse (chiffres significatifs exacts) due à la soustraction de deux nombres approximés quasi égaux lorsque $x \to 0^+$ (et un calcul impossible pour $x < 0$). Il vaudrait mieux réécrire cette expression, pour qu'elle soit mathématiquement équivalente et ne fasse plus intervenir ce problème, lorsque $x \to 0^+$, en utilisant l'expression conjuguée, de la façon suivante :

$$\begin{aligned}\sqrt{1+x} - \sqrt{1+x^2} &= \frac{(\sqrt{1+x} - \sqrt{1+x^2})(\sqrt{1+x} + \sqrt{1+x^2})}{\sqrt{1+x} + \sqrt{1+x^2}} \\ &= \frac{x - x^2}{\sqrt{1+x} + \sqrt{1+x^2}} = \frac{x(1-x)}{\sqrt{1+x} + \sqrt{1+x^2}}\end{aligned}$$

<5 pts>

Nota : On a toujours un problème d'annulation de cse lorsque $x \to 0$ au numérateur $(x - x^2)$ mais cette forme est beaucoup moins grave qu'initialement (cf. Graphe ci contre) surtout lorsqu'elle est écrite comme étant équivalente à $x(1-x)$.

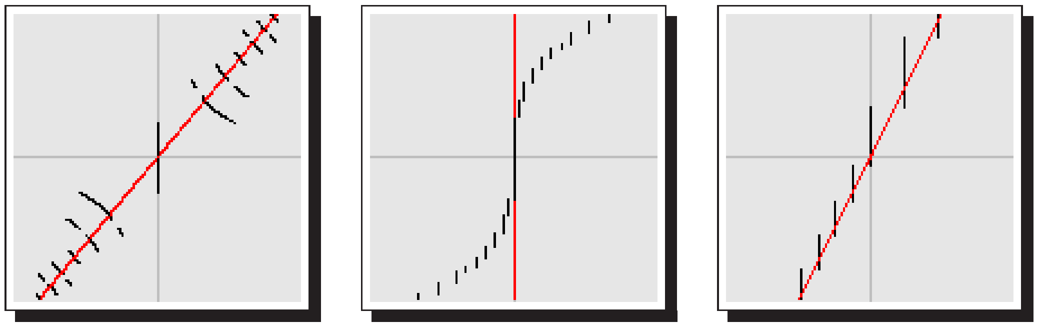
[-d-]

Pour la deuxième expression, lorsque $x \to 0$,

on a la SOUSTRACTION DE DEUX NOMBRES APPROXIMÉS QUASI ÉGAUX On ne peut pas trouver une expression strictement mathématiquement équivalente mais on peut proposer au voisinage de 0, une expression approximativement équivalente qui augmenterait la précision de ce calcul en utilisant un développement limité au voisinage de zéro :

$$\begin{aligned}\sin x - (x - \frac{x^3}{6}) &\approx (x - \frac{x^3}{3!} + \frac{x^5}{5!}) - (x - \frac{x^3}{6}) \\ &= \frac{x^5}{120}\end{aligned}$$

<5 pts>



De gauche à droite : expression brute en FLOAT au voisinage de 0 (en noire) & expr. math. equiv. en (rouge) pour l'expression : (b) $[x \in \pm 0.5 \cdot 10^{-4}]$, et $[y \in \pm 0.3 \cdot 10^{-4}]$ (c) $[x \in \pm 10^{-8}$ et $y \in \pm 10^{-23}]$, (d) $[(x, y) \in \pm 10^{-7}]$. Donc, en résumé, analytiquement, on a la courbe en rouge mais numériquement (ou sur ordinateur et en FLOAT), l'expression (a), (b) et (c) originale est en fait équivalente à (et nous donne) la courbe en noire.

III. Méthode du Point Fixe/Newton (56 pts)

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

$$f(x) = e^{\cos x} - x = [\exp(\cos x)] - x = 0 \tag{1}$$

1. Méthode du Point Fixe
   (a) Montrer qu'il existe une racine unique r pour cette Eq. (1) dans l'intervalle $J = [\pi/3\,,\,\pi/2]$ et donner une forme $x = g_1(x)$, mathématiquement équivalente de $f(x) = 0$ (Eq. (1)), pour laquelle la suite itérative $r_{n+1} = g_1(r_n)$ converge lorsque, le premier élément de cette suite est le milieu de l'intervalle J, c.-à-d. ; $r_0 \approx 1.3$.
   <8 pts>
   (b) Utiliser le résultat de la question précédente pour calculer les 6 premières estimées $r_1, \ldots, r_6$ en partant de $r_0 = 1.3$.
   <6 pts>
   (c) Pour obtenir à l'avance le nombre d'itérations nécessaires de la méthode du point fixe pour arriver à la racine souhaitée avec la précision voulue, on peut essayer de trouver (en utilisant le théorème des accroissements finis ou de la valeur moyenne) une majoration du type $|r_n - r| \leq K$, ou r_n désigne la valeur approchée, à la nième itération, de cette racine. Trouver cette majoration et en déduire le nombre d'itérations nécessaires pour obtenir, par cette méthode, une valeur approchée à 10^{-4} près de cette racine.
   <5 pts>
   (d) Une autre possibilité consiste à arrêter la procédure itérative du point fixe lorsque r_{n+1} et r_n sont suffisamment proches, *i.e.*, $|r_{n+1} - r_n|$ suffisamment petits. Dans ce cas, trouver une relation du type :
   $$|r_n - r| \leq K\,|r_{n+1} - r_n|\,,$$
   avec K un majorant que l'on précisera et en déduire le nombre d'itérations nécessaire pour obtenir une valeur approchée de r à 10^{-4} près.
   <6 pts>
   (e) Donner une forme $x = g_2(x)$, mathématiquement équivalente de $f(x) = 0$ (Eq. (1)), pour laquelle la suite itérative $r_{n+1} = g_2(r_n)$ ne converge pas (lorsque le premier élément de la suite est $r_0 = 1.3$ fixé au milieu de l'intervalle considéré) et expliquer mathématiquement pourquoi.
   <6 pts>

(f) Utiliser le résultat de la question précédente pour calculer les 6 premières estimées $r_1, \ldots, r_6$ de la suite itérative $r_{n+1} = g_2(r_n)$ en partant de $r_0 = \pi/3$.
<**5 pts**>

2. Méthode de la Bissection

(a) Si on avait utilisé la méthode de la bissection, avec combien d'itérations serait-on arrivé à une valeur approchée de la racine à 10^{-4} près ? (Nota : on vous demande d'obtenir une estimation de ce nombre en utilisant les propriété de la méthode de la bissection mais sans calculer les différentes estimation $r_0, r_1, \ldots,$ donné par cette méthode)
<**5 pts**>

3. Méthode de Newton

(a) Donner la relation $r_{n+1} = g_{\text{newt.}}(r_n)$ 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 théoriquement qu'avec $r_0=1.3$, la relation itérative de Newton converge.
<**5 pts**>

(c) Calculer les 3 premières estimées $r_1, \ldots, r_3$ en partant de $r_0=1.3$.
<**5 pts**>

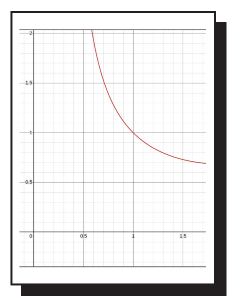
Réponse

1.(a)

L'étude des variations de la fonction f sur $J = [\pi/3\,,\,\pi/2]$ (*on oublie pas de mettre sa calculatrice en radian*) montre que la fonction est continue et décroissante sur J, donc monotone car $f'(x) = -\sin x \cdot \exp(\cos x) - 1 < 0$ sur J (puisque $\sin x$, $\cos x$ et $\exp(\cos x)$ sont toutes les trois des fonctions positives sur J, donc $-\sin x \cdot \exp(\cos x)$ et $-\sin x \cdot \exp(\cos x) - x = f'(x)$ sont donc < 0 sur J).
De plus, puisque $f(\pi/3) \approx 0.60$ est positive et $f(\pi/2) \approx -0.57$ est négative, alors $f(\pi/3)f(\pi/2) < 0$ et, puisque la fonction f est continue et décroissante sur J, il existe donc une racine r unique de f dans l' intervalle J.
<**4 pts**>

De plus $f(x)=0$ est équivalent à l'équation $x = \arccos(\ln x) = g_1(x)$ (l'autre forme $x = g(x)$ possible [cf. Question 1.(e).] ne permet pas d'assurer la contractance, *i.e.*, une convergence itérative) avec :

$$|g'(x)| = \left| \frac{-\frac{1}{x}}{\sqrt{1-(\ln x)^2}} \right|$$



On aimerait maintenant savoir quel est l'*extrema* de cette fonction sur l'intervalle $J = [\pi/3\,,\,\pi/2] = [\approx 1.05\;,\;\approx 1.57]$. Plutôt que de calculer la dérivée de $g'(x)$ et ensuite étudier son tableau de variation qui nous prendrait beaucoup d'effort, on peut calculer les valeurs de cette fonction pour un ensemble de valeurs discrétisé entre la borne min et max de J. On trouve $g'(\pi/3 \approx 1.05) \approx \mathbf{0.956}$, $g'(1.25) \approx 0.779$ et $g'(\pi/2) \approx 0.714$ qui semble nous montrer que la fonction $|g'(x)|$ est aussi décroissante et évoluera donc entre $g'(\pi/3) \approx 0.956$ et $g'(\pi/2) \approx 0.714$. On peut s'en convaincre en utilisant sa calculatrice graphique ou le calculateur de graphe *online* que j'ai mis a disposition sur ma page Web de ce cours [cf. graphe à doite de g' qui montre que g' est décroissante sur J avec sa plus haute valeur donnée donc par $g'(\pi/3 \approx 1.05) \approx 0.956$].

donc :

$$|g'(x)| < 0.956 < 1 \quad \forall\, x \in J = [\pi/3\,,\,\pi/2]$$

La fonction $g(x)$ est donc contractante sur J et la convergence est assurée puisque le premier élément $r_0 = 1.3$ de la suite $r_{n+1} = g_1(r_n)$ est aussi dans J.
<**4 pts**>

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

En partant de $r_0 = 1.3$, on a, $r_n = g_1(r_{n-1}) = \arccos\big(\ln(r_{n-1})\big)$, et ;

$$
\begin{aligned}
r_1 &= 1.305324843 \\
r_2 &= 1.301086327 \\
r_3 &= 1.304459131 \\
r_4 &= 1.301774582 \\
r_5 &= 1.303910917 \\
r_6 &= 1.302210589
\end{aligned}
$$

Avec ces six itérations, on peut seulement dire que l'on converge vers la valeur $r = 1.3\underline{0}$ avec le chiffre des centièmes qui semble significatif.
<6 pts>

Nota : On peut constater que l'initialisation ($r_0 = 1.3$) est bonne mais que la convergence de cette méthode d'estimation itérative est extrêmement lente. En 6 itérations, seule le chiffre des centièmes de cette estimation de racine ne change plus. Cette lenteur dans la convergence sera démontrée dans la suite de cette exercice. Plus précisément, il faut faire 68 itérations pour finalement converger vers $r = 1.302964001$ et avoir le maximum des 9 chiffres significatifs de ma calculatrice *CASIO fx-991MS !*

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

En utilisant le théorème de la valeur moyenne, on obtient, puisque $g(r) = r$ et $r_n = g_1(r_{n-1})$ avec r_n la valeur approchée de la racine à la nième itération ;

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

En utilisant l'inégalité $|g_1'(\xi)| < 0.956$ (cf. question 1.(a)), on obtient les inégalités suivantes ;

$$
|r_n - r| \leq \; 0.956\,|r_{n-1} - r| \;\leq\; 0.956^2\,|r_{n-2} - r| \;\leq \ldots \leq\; 0.956^n\,\underbrace{(r_0 - r)}_{\leq \pi/6}
$$

On obtiendra donc $|r_n - r| < 10^{-4}$ dès que $(\pi/6) \times (0.956)^n < 10^{-4}$, *i.e.*, dès que $n \geq 190.3$, *i.e.*, dès que $\mathbf{n = 191}$. Il s'agit bien sûr d'une borne supérieure. Si on veut quelque chose de plus précis, alors on doit prendre $g_1'(\zeta = r = 1.303) \approx 0.796$ conduisant à $n = 37$ itérations (les deux réponses sont acceptables).
<5 pts>

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

En utilisant le théorème des accroissement finis, on a (cf. démo2 ou examen IntraH16 [III.1.(d)]) :

$$
r_{n+1} - r_n \;=\; (g'(\zeta) - 1) \times (r_n - r) \quad \text{avec } \zeta \text{ compris entre } r_n \text{ et } r
$$

De ce fait, on a : $\quad |r_n - r| = \dfrac{1}{|1 - g'(\zeta)|} \times |r_{n+1} - r_n| \quad$ avec $\;\forall\, x \in J\;$ et $\;|g'(x)| < 0.956$

$$
\text{d'où} \quad \frac{1}{|1 - g'(\zeta)|} \leq 22.73 \quad \text{et on peut écrire :} \quad |r_n - r| \leq 22.73 \cdot |r_{n+1} - r_n|
$$

Si on veut $|r_n - r|$ inférieur à 10^{-4} ;
il suffirra de choisir n tel que : ▶ $\quad |r_{n+1} - r_n| < \frac{10^{-4}}{22.73}$ ($\approx 0.0000044 \approx \mathbf{4.4 \times 10^{-6}}$).
<6 pts>

Nota -1- : Il s'agit bien sûr d'une borne supérieure. Si on veut quelque chose de plus précis, alors on doit prendre $g_1'(\zeta = r = 1.303) \approx 0.796$ conduisant à $|r_{n+1} - r_n| < \frac{10^{-4}}{4.902} \approx 0.00002$ (les deux réponses sont acceptables).

Nota -2- : Dans notre, on arrive à $r_{15} \approx 1.303060517$ et $r_{16} \approx 1.302887191$ qui a une erreur $< 10^{-4}$ (par rapport à la racine $r = 1.302964001$) et la différence $|r_{16} - r_{15}| \approx 0.000173326$. Dans notre exemple, $r_{24} \approx 1.302951639$ et $r_{25} \approx 1.30297384$ et la différence $|r_{24} - r_{25}| \approx 0.0000222$ qui atteint le critère souhaité.

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

$f(x) = 0$ est équivalent aussi à l'équation $x = \exp\big(\cos(x)\big) = g_2(x)$ qui ne converge pas, car :

$$|g_2'(x)| = \Big| -\sin(x)\,\exp(\cos(x))\Big| > 1 \quad \forall\, x \in J = [\pi/3\,,\,\pi/2]$$

On peut donner un exemple, en particulier en $x = \pi/3$, $|\sin(x)\,\exp(\cos(x))| = 1.4278 > 1$.
<6 pts>

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

En partant de $r_0 = \pi/3 \approx 1.047$, on a, $r_n = g_2(r_{n-1}) = \exp(\cos(r_{n-1}))$ et ;

$$\begin{aligned}
r_1 &= 1.648721271\\
r_2 &= 0.9255106785\\
r_3 &= 1.825309036\\
r_4 &= 0.777420535\\
r_5 &= 2.039541964\\
r_6 &= 0.636502088
\end{aligned}$$

<5 pts>

Nota -1- : En partant de $r_0 = 1.3$, cela diverge mais il faut **beaucoup** d'itération et donc pour s'en apercevoir il faut être capable de programmer sa calculatrice. Sur ma calculatrice, j'obtiens ; en partant de $r_0 = 1.3$, on a $r_1 = 1.306692099$, $r_2 = 1.298285646$, $r_3 = 1.308851866, \ldots$, $r_{16} = 1.191107547$, $r_{17} = 1.448649049$, $r_{18} = 1.129577609$, $r_{19} = 1.532716706$, $\ldots$, $r_{40} = 0.464437172$, $r_{41} = 2.445070137$ et semble osciller après entre deux valeurs : 0.464334594 et 2.445182472. En fait ces deux valeurs particulières sont liées par les relations : $0.464334594 = \exp(\cos(2.445182472))$ et $2.445182472 = \exp(\cos(0.464334594))$.

Nota -2- : Avec une calculatrice du type *CASIO fx-991MS*, par exemple [et en utilisant le mode radian], je rappelle que cette suite itérative se programme facilement en affichant 1.3 puis en appuyant sur la touche $\boxed{=}$ puis en appuyant sur la touche $\boxed{EXP}$ puis sur $\boxed{COS}$ puis sur $\boxed{ANS}$. Chaque appel à la touche $\boxed{=}$ nous permet ensuite d'avoir une itération de la suite.

$\boxed{\textbf{2.}}$

La méthode de la bissection permet de construire à partir de l'intervalle $[a\,,\,b]$ contenant r, un nouvel intervalle de longueur moitié contenant r. En appliquant n fois consécutives la méthode, on obtient un intervalle de longueur divisée par 2^n contenant r. Dans notre cas, à l'itération n, on a un majorant de l'erreur absolu donné par $\Delta r = \frac{|\pi/6|}{2^n}$. Or on veut $\Delta r < 10^{-4}$, donc $2^n > 10^4/(\pi/6)$, ce qui, dans notre cas est vrai des que $\underline{n = 13}$.
Ce nombre d'itération est à comparer avec le $n = 191$ ou $n = 37$ obtenu par les deux critères différents utilisés pour la méthode itérative du point fixe précédente. La méthode de la bissection est quelquefois beaucoup plus intéressante que la méthode du point fixe !
<5 pts>

3.(a)

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

$$r_{n+1} = r_n - \frac{f(r_n)}{f'(r_n)} = r_n + \frac{\big(\exp(\cos(r_n)) - r_n\big)}{\big([\sin(r_n) \cdot \exp(r_n)] + 1\big)}$$

<5 pts>

3.(b)

Dans ce cas, le plus simple est de montrer que la fonction $f(x)$, sur l'intervalle $J = [\pi/3\,,\, \pi/2]$ (intervalle dans lequel on prendra $r_0 = 1.3$ et dans lequel une racine unique s'y trouve) ne présente pas d'*extrema*, *i.e.*, de valeurs pour laquelle, $f'(x)$ s'annule (cf. Question 1.(a)).
On rappelle que dans notre cas, $f'(x) = -\sin x \cdot \exp(\cos x) - 1 < 0$ sur J et donc ne s'annule jamais ; donc aucun problème de convergence.
<5 pts>

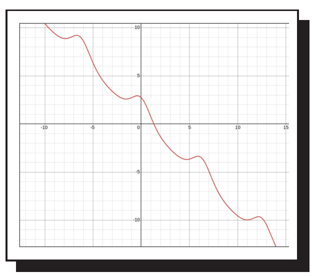
3.(c)

En partant de $r_0 = 1.3$, on a, $r_n = g_{\text{newt.}}(r_{n-1})$ et :

$$\begin{aligned} r_1 &= 1.30296232 \\ r_2 &= 1.302964001 \\ r_3 &= 1.302964001 \end{aligned}$$

<5 pts>

Nota -1- : Sur ma calculatrice, l'estimation se stabilise très rapidement, à partir de $r_2 = 1.302964001$! (à comparer avec la méthode du point fixe précédente qui se stabilisait à partir de $r_{i>68}$). Cela nous indique aussi que la fonction dont on cherche la racine est très très linéaire au voisinage de la racine.
On peut voir l'effet d'une racine simple sur la convergence de Newton qui, dans ce cas, est quadratique.



On peut voir que cette fonction $e^{\cos x} - x$ est très linéaire au voisinage de la racine.

Nota -2- : Sur ma calculatrice, si on partait d'un peu plus loin, par exemple $r_0 = 1.6$, on trouverait succéssivement, $r_0=\underline{1}.6$, $r_1=\underline{1}.44085295$, $r_2=\underline{1.3}45984623$, $r_3=\underline{1.3}10920725$, $r_4=\underline{1.30}3985831$, $r_5=\underline{1.30}308306$, $r_6=\underline{1.3029}77687$, $r_7=\underline{1.3029}65572$, $r_8=\underline{1.302964}181$, $r_9=\underline{1.30296402}2$, $r_{10}=\underline{1.30296400}4$, $r_{11}=\underline{1.302964001}$.
À partir de r_6, (stabilité), on a 4 cse après la virgule et on gagne 2 cse pour les deux prochaines itérations, puis, à partir de r_8, 4 cse pour les deux suivantes itérations donnant pour r_{10} ; 8 cse après la virgule ce qui est caractéristique et typique d'une convergence quadratique (pour la méthide de Newton liée à une racine simple).

Nota Finale : La méthode du point fixe peut s'avérer très lente (beaucoup plus lente que la méthode de la bissection) et la méthode de Newton peut s'avérer très rapide surtout si la fonction dont on cherche la racine est très linéaire au voisinage de cette racine.