Méthodes numériques : optimisation.
Partiel du 10 mars 2016 — Éléments de correction

1 Méthode de la section dorée

1. La fonction $f : x \mapsto \sqrt{1+x^2}$ est strictement décroissante sur $]-\infty, 0]$ et strictement croissante sur $[0, +\infty[$, elle est donc unimodale sur $[-1, 2]$. D'après le cours, les suites données par la méthode de la section dorée convergent linéairement vers le minimum (ici 0) avec un taux de convergence $\frac{1}{\varphi} = \frac{\sqrt{5}-1}{2}$ (l'inverse du nombre d'or).
   Sur $[1, 2]$, la fonction n'est pas unimodale : elle est strictement croissante. On aura donc toujours $f(c_k) < f(d_k)$, l'algorithme choisit donc toujours le triplet de gauche, donc la suite $(a_k)_{k\in\mathbb{N}}$ est stationnaire égale à 1, les autres suites convergent donc aussi vers 1, qui n'est pas le minimum de la fonction sur $\mathbb{R}$, mais qui est bien le minimum sur $[1, 2]$. La convergence est linéaire avec le même taux $\frac{1}{\varphi}$.
2. Voilà un exemple de ce que l'on pourrait écrire :

```
alpha=(sqrt(5)-1)/2
def SectionDoree(f,a,b,eps):
    c=a+(1-alpha)*(b-a)
    d=a+alpha*(b-a)
    fc,fd=f(c),f(d)
    while b-a>eps:
        if fc>fd:
            a,c=c,d
            d=a+alpha*(b-a)
            fc,fd=fd,f(d)
        else:
            d,b=c,d
            c=a+(1-alpha)*(b-a)
            fc,fd=f(c),fc
    return (a+b)/2
```

3. Voici le code de ce qui a été utilisé pour produire le graphique donné dans l'énoncé.

```
def fcompteur(x):
    global compteur
    compteur+=1
    return sqrt(1+x**2)

tolerances=[10**(-i) for i in range(1,15)]
nevaluations=[]
erreurs=[]
```

```
for eps in tolerances:
    compteur=0
    c=minDoree(fcompteur,-1,2,eps)
    erreurs.append(abs(c))
    nevaluations.append(compteur)

semilogy(nevaluations,tolerances,'--k')
semilogy(nevaluations,erreurs,'k')
semilogy(nevaluations,alpha**nevaluations,':k')

legend(['Tolérances','Erreurs','$\\alpha^n$'])
xlabel('Nombre d'évaluations~$n$')
show()
```

4. On note $\hat{c}_k$ et $\hat{d}_k$ les valeurs des variables `c` et `d` après k passages dans la boucle. Lorsque $|\hat{c}_k|$ et $|\hat{d}_k|$ sont plus petits que 10^{-8}, alors $\hat{c}_k^2$ et $\hat{d}_k^2$ sont plus petits que 10^{-16} qui est la précision machine. Donc pour la machine, $1+\hat{c}_k^2$ et $1+\hat{d}_k^2$ sont tous les deux codés par le même nombre à virgule flottante : 1. Donc le calcul de $\sqrt{1+\hat{c}_k^2}$ et $\sqrt{1+\hat{d}_k^2}$ renvoie la même valeur 1. On se retrouve donc dans tous les cas dans la deuxième partie de la boucle, le triplet choisi est toujours celui de gauche, et la variable `a` (de l'ordre de 10^{-8}) n'est plus modifiée, et les valeurs des variables `b`, `c`, et `d`, se rapprochent de cette valeur jusqu'à atteindre la tolérance demandée.

5. * La théorie est la même, la fonction étant unimodale sur $[-1, 2]$. Par contre cette fois-ci, il suffit que $|\hat{c}_k|$ et $|\hat{d}_k|$ soient de l'ordre de 10^{-4} pour que la fonction retourne la même chose pour les variables `c` et `d`. Les erreurs se stabiliseraient donc autour de 10^{-4}, la méthode est donc moins précise dans ce cas-là. Cela est dû au fait que le minimum est dégénéré.

2 Séparation des variables pour la descente de gradient à pas fixe

1. Il suffit de voir que f est C^1 puisque g et h le sont, et que l'on a $\partial_i f(x,y) = \partial_i g(x)$ pour $1 \leqslant i \leqslant n$ et $\partial_{n+i} f(x,y) = \partial_i h(y)$ pour $1 \leqslant i \leqslant p$.
   On a donc $\nabla f(x,y) = (\nabla g(x), \nabla h(y))$, et les itérées de la méthode de descente de gradient à pas fixe sont donc données par
   $$(x_{k+1}, y_{k+1}) = (x_k, y_k) - \alpha \nabla f(x_k, y_k) = (x_k - \alpha \nabla g(x_k), y_k - \alpha \nabla h(y_k)).$$
   Les suites $(x_k)_{k\in\mathbb{N}}$ et $(y_k)_{k\in\mathbb{N}}$ correspondent donc à la descente de gradient à pas fixe pour la fonction g partant de x_0 et pour la fonction h partant de y_0.
   Ces fonctions vérifient bien les hypothèses du Théorème 1 du cours : en effet les ensembles S_0^g et S_0^h sont bien fermés dans $\mathbb{R}^n$ et $\mathbb{R}^p$ car g et h sont continues (vu qu'elles sont C^1). On sait donc que si $\alpha < \frac{2}{L_g}$ et $\alpha < \frac{2}{L_h}$:
   — Pour tout k, $x_k \in S_0^g$ et $y_k \in S_0^h$.
   — Les suites $(g(x_k))_{k\in\mathbb{N}}$ et $(h(y_k))_{k\in\mathbb{N}}$ sont décroissantes et convergent vers deux limites finies g_∞ et h_∞.
   — Les suites $(\nabla g(x_k))_{k\in\mathbb{N}}$ et $(\nabla h(y_k))_{k\in\mathbb{N}}$ convergent vers 0 dans $\mathbb{R}^n$ et dans $\mathbb{R}^p$.
   On peut donc prendre $L = \max(L_g, L_h)$, on a donc bien que si $\alpha < \frac{2}{L}$, alors $\alpha < \frac{2}{L_g}$ et $\alpha < \frac{2}{L_h}$, et les conclusions précédentes nous donnent ce qui est demandé : il suffit d'obtenir que $(f(x_k, y_k))_{k\in\mathbb{N}}$ est décroissante avec une limite finie, mais c'est le cas puisque c'est la somme de deux suites décroissantes qui ont une limite finie.

2. (a) Si A_g a une valeur propre strictement négative λ avec un vecteur propre v associé, alors $g(tv) = \lambda t^2 \langle v, v\rangle_n + t\langle b_g, v\rangle_n$ qui tend vers $-\infty$ lorsque $t \to +\infty$ donc g n'est pas bornée inférieurement. Sinon toutes les valeurs propres sont positives ou nulles, donc g est convexe (car sa hessienne A_g est symétrique positive), un point de minimum est donc équivalent à un point critique, c'est à dire un x tel que $A_g x + b_g = 0$. Si une des valeurs propres est nulle, alors s'il y a une solution elle ne peut pas être unique (si $A_g v = 0$ avec $v \neq 0$ alors pour toute solution x, $x + v$ est aussi solution). Donc si la fonction admet un unique minimum alors toutes les valeurs propres sont strictement positives, c'est-à-dire que $\ell_g > 0$. Réciproquement si $\ell_g > 0$, alors A_g est symétrique définie positive, donc inversible, donc l'équation $A_g x + b = 0$ a une unique solution, il y a donc un unique point critique qui est un minimum puisque la fonction est alors convexe. C'est la même chose pour h, qui admet un unique minimum si et seulement si $\ell_h > 0$.

   (b) On peut soit réécrire $f(z) = \frac{1}{2}\langle z, Az\rangle + \langle b, z\rangle$ dans $\mathbb{R}^{n+p}$ (où on a noté $z = (x_1, \dots x_n, y_1, \dots y_p)$ avec $A = \begin{pmatrix} A_g & 0 \\ 0 & A_h \end{pmatrix}$ et $b = \binom{b_g}{b_h}$, dans ce cas on obtient que A est symétrique définie positive, sa plus grande valeur propre étant $L = \max(L_g, L_h)$ et sa plus petite $\ell = \min \ell_g, \ell_h$ et on applique le cours : la méthode converge dans tous les cas de figure si $\alpha < \frac{2}{L}$ et le taux de convergence est inférieur à $r(\alpha) = \max(|1 - \alpha\ell|, |1 - \alpha L|)$. De plus $r(\alpha)$ atteint son minimum pour $\alpha = \frac{2}{L+\ell}$.
   Sinon on peut directement appliquer le résultat à g et h (puisque la descente de gradient pour f revient à faire la descente de gradient avec g pour les x_k et avec h pour les y_k), pour lesquelles la méthode converge si $\alpha < \frac{2}{L}_g$ et $\alpha < \frac{2}{L}_h$. Le taux de convergence de x_k est inférieur à $r_g(\alpha) = \max(|1 - \alpha\ell_g|, |1 - \alpha L_g|)$, de même celui pour les y_k est inférieur à $r_h(\alpha) = \max(|1 - \alpha\ell_h|, |1 - \alpha L_h|)$. Donc le taux de convergence pour (x_k, y_k) est

inférieur à $\max(r_g(\alpha), r_h(\alpha))$ et en regardant les différents cas possibles, on a que $r(\alpha) = \max(|1-\alpha\ell|, |1-\alpha L|)$ où $\ell = \min \ell_g, \ell_h$ et $L = \max(L_g, L_h)$.

On n'a pas forcément montré dans le cours que si $\alpha \geqslant \frac{2}{L}$, alors il existe effectivement des points initiaux pour lesquels la méthode ne converge pas. Montrons-le pour g : si v est un vecteur propre associé à L_g, on sait qu'on a

$$x_{k+1} - x_k = x_k - \alpha(A_g x_k + b_g) - x_{k-1} + \alpha(A_g x_k + b_g) = (I_n - \alpha A_g)(x_k - x_{k-1}).$$

Si on arrive à avoir $x_1 - x_0 = v$, alors par récurrence on aura $x_{k+1} - x_k = (1-\alpha L_g)^k v$ donc si $\alpha >= \frac{2}{L_g}$, alors $1 - \alpha L_g \leqslant -1$ et donc la suite $(1-\alpha L_g)^k$ ne converge pas, donc on ne peut pas avoir $x_{k+1} - x_k \to 0$, donc la suite $(x_k)_{k\in\mathbb{N}}$ ne converge pas non plus. C'est effectivement possible de prendre x_0 tel que $x_1 - x_0 = v$ puisque cela équivaut à $v = -\alpha(A_g x_0 + b)$ et que A_g est inversible (et $\alpha \neq 0$).

3. On a $x_{k+1} = x_k(1 - \frac{\alpha}{\sqrt{1+x_k^2}})$ et $y_{k+1} = y_k(1 - \frac{2\alpha y_k^2}{\sqrt{1+y_k^4}})$.
   (a) Si on suppose $0 < \alpha \leqslant 2$, alors dès que $x_k \neq 0$, $-1 < (1 - \frac{\alpha}{\sqrt{1+x_k^2}}) < 1$, donc on obtient $|x_{k+1}| < |x_k|$. La suite des $|x_k|$ est donc décroissante (et même strictement décroissante si elle n'est pas stationnaire à 0), et converge vers une limite ℓ qui vérifie par continuité $\ell = \ell|1 - \frac{\alpha}{\sqrt{1+\ell^2}}|$ qui n'a comme solution que $\ell = 0$ (car $-1 < (1 - \frac{\alpha}{\sqrt{1+\ell^2}}) < 1$ pour $\ell \neq 0$ et $0 < \alpha \leqslant 2$).
   Si $\alpha > 2$ alors il y a une autre solution à cette équation : $\ell = \sqrt{\frac{\alpha}{2} - 1}$. Si on part de $x_0 = \ell$, alors on obtient $x_k = (-1)^k \ell$ qui ne converge pas.
   La suite des x_k converge donc quelque soit x_0 si et seulement si $0 < \alpha \leqslant 2$ (et dans ce cas-là elle converge vers 0).
   Si $0 < \alpha < 2$, la convergence est linéaire et le taux de convergence est égal à $|1-\alpha|$ (sauf si l'un des x_k est nul). En effet, on a $\frac{|x_{k+1}|}{|x_k|} = |1 - \frac{\alpha}{\sqrt{1+x_k^2}}| \to |1-\alpha| < 1$ lorsque $k \to \infty$.
   Si $\alpha = 2$, si les x_k sont tous non-nuls, alors on a $\frac{|x_{k+1}|}{|x_k|} \to 1$, et la convergence n'est pas linéaire (on dit qu'elle est sous-linéaire).
   (b) Soit δ tel que si $|y| < \delta$, on ait $(1 - \frac{2\alpha y^2}{\sqrt{1+y^4}}) \in]-1, 1[$. On obtient facilement par récurrence que si $|y_0| < \delta$ alors la suite $|y_k|$ est décroissante et converge vers une limite $\ell < \delta$ qui vérifie par continuité $\ell = \ell|1 - \frac{2\alpha\ell^2}{\sqrt{1+\ell^4}}|$, et on obtient donc $\ell = 0$.
   Si tous les y_k sont non-nuls, on a que $\frac{|y_{k+1}|}{|y_k|} \to 1$, et la convergence n'est pas linéaire.
   (c) On a vu à la première question qu'il suffit de prendre $0 < \alpha \leqslant 2$ pour être assuré de la convergence de (x_k). Pour s'assurer de la convergence des y_k il suffit de prendre α suffisamment petit pour que $\alpha \frac{y_0^2}{\sqrt{1+y_0^4}} < 1$. En effet, on montre alors par récurrence que la suite des $|y_k|$ est décroissante et vérifie $\alpha \frac{y_0^2}{\sqrt{1+y_0^4}} < 1$ pour tout k, et on montre de même que sa limite ℓ vaut zéro. Comme précédemment, si aucun des y_k n'est nul, la convergence de $(y_k)_k \in \mathbb{N}$ n'est pas linéaire, donc la convergence des points (x_k, y_k) ne l'est pas non plus.
   (d) * On fait un développement limité pour obtenir la première limite. Ensuite on écrit

$$\frac{1}{k}\frac{1}{y_k^2} = \frac{1}{k}\left[\frac{1}{y_0^2} + \sum_{i=1}^{k-1} \frac{1}{y_{i+1}^2} - \frac{1}{y_i^2}\right] = \frac{1}{ky_0^2} - \frac{1}{k}\sum_{i=1}^{k-1} u_i,$$

   qui est une moyenne de Cesàro d'une suite convergent vers $\sqrt{2\alpha}$, et qui converge donc aussi vers $\sqrt{2\alpha}$.

3 Méthodes de Newton et de la sécante en dimension 1

1. (a) On a $f'(x) = \frac{x}{\sqrt{1+x^2}}$ et $f''(x) = \frac{1}{(1+x^2)^{\frac{3}{2}}}$.

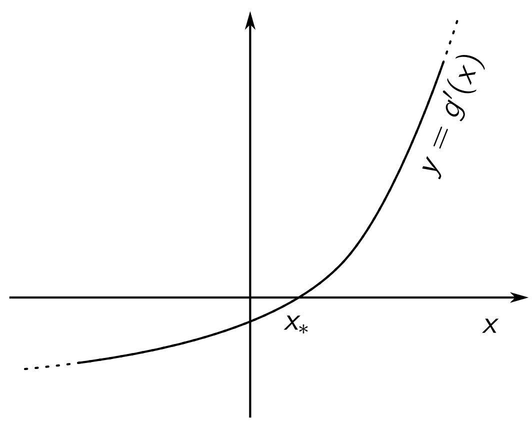
   (b) On a donc $f''(x) > 0$ pour tout x, donc les itérées sont bien définies pour tout k et données par $x_{k+1} = x_k - \frac{f'(x_k)}{f''(x_k)} = x_k - x_k(1+x_k^2) = -x_k^3$.

   (c) On a donc par récurrence $|x_k| = |x_0|^{3^k}$, qui converge vers 0 si $|x_0| < 1$, vers 1 si $|x_0| = 1$ (mais dans ce cas on obtient $x_k = (-1)^k x_0$, donc $(x_k)_{k\in\mathbb{N}}$ ne converge pas), qui diverge si $|x_0| > 1$. La suite converge donc si et seulement si $x_0 \in]-1,1[$. Si $x_0 = 0$ la suite est stationnaire, sinon elle converge à l'ordre 3. Ce n'est pas en contradiction avec les résultats donnés dans le cours, puisque la convergence est censée être d'ordre supérieur ou égal à deux (en fait c'est ici parce que $f^{(3)}(0) = 0$ que cela converge encore plus rapidement). On dit que la méthode de Newton n'est pas robuste puisqu'on voit qu'elle ne converge pas pour n'importe quelle donnée initiale x_0, bien que l'on ait des bonnes propriétés sur la fonction f (de classe C^∞, unimodale sur $\mathbb{R}$, avec un minimum non dégénéré).

2. On rappelle que comme la fonction g' est dérivable, elle est strictement convexe si et seulement si sa dérivée g'' est strictement croissante. On s'intéresse donc à une fonction g strictement décroissante sur $]-\infty, x_*]$ et strictement croissante sur $[x_*, +\infty[$, telle que g'' soit strictement croissante sur $\mathbb{R}$.

   (a) La fonction f de la première question est bien unimodale, mais comme f'' est croissante sur $]-\infty, 0]$ et décroissante sur $[0, +\infty[$, alors f ne satisfait pas ces hypothèses.

   Pour obtenir une telle fonction g, il suffit d'avoir une dérivée g' qui soit C^1, strictement convexe et prenne des valeurs strictement négatives pour $x < x_*$ et positives pour $x > x_*$. Voici un dessin d'une telle dérivée g' :



   Si on prend pour g n'importe quelle primitive d'une fonction dessinée ci-dessus, on obtient bien une fonction unimodale (puisque la dérivée est strictement négative puis strictement positive), et sa dérivée est strictement convexe.

   Par exemple en voyant ce dessin, on peut imaginer prendre $g'(x) = e^x - 1$, et donc par exemple $g(x) = e^x - x$. On a alors bien $g'(x)$ strictement convexe, strictement négative sur $]-\infty, 0[$ et strictement positive sur $]0, +\infty[$, donc g est unimodale avec un unique minimum en 0.

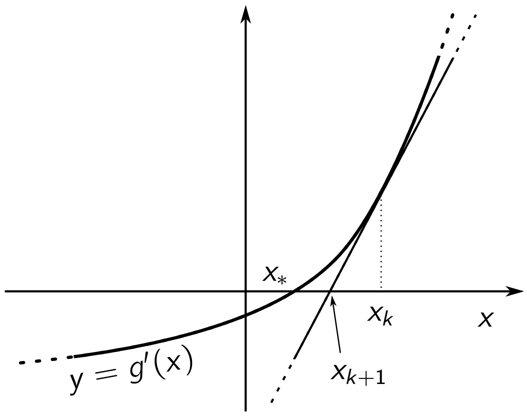
   (b) On voit bien sur le graphique précédent que puisque la courbe de g' est au-dessus de ses tangentes, elle ne doit pas avoir de tangente de pente négative, sinon on ne pourrait pas avoir $g'(x) < 0$ pour $x \to \infty$.

   Démontrons le résultat demandé par l'absurde, en supposant qu'il existe un $x_0 \in \mathbb{R}$ tel que $g''(x_0) \leqslant 0$. Comme g' est strictement convexe, alors g'' est strictement croissante, et donc on prend $x_1 < x_0$ pour et on obtient que $g''(x_1) < 0$. On a donc que pour tout $x \leqslant x_1$,

$$g(x_1) - g'(x) = \int_x^{x_1} g''(x)\mathrm{d}x \leqslant \int_x^{x_1} g''(x_1)\mathrm{d}x \leqslant (x - x_1)g''(x_1).$$

On obtient donc $g'(x) \geqslant g'(x_1) - g''(x_1)(x - x1)$, donc $g'(x) \to +\infty$ lorsque $x \to -\infty$. Or comme g est unimodale, on doit avoir $g'(x) \leqslant 0$ pour tout $x < x_*$, ce qui est en contradiction avec le fait que $g'(x) \to +\infty$ lorsque $x \to -\infty$.

(c) D'après la question précédente, l'itérée à l'étape k est donc toujours bien définie et donnée par $x_{k+1} = x_k - \frac{g'(x_k)}{g''(x_k)}$. On suppose par récurrence que $x_k > x_*$. On sait que x_{k+1} est l'abscisse de l'intersection de l'axe des abscisses et de la tangente à la courbe $y = g'(x)$ au point d'abscisse x_k :

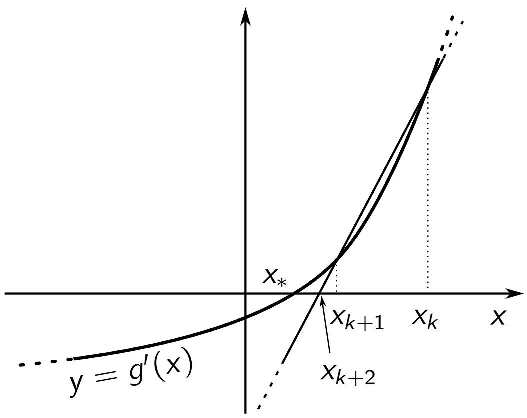


Comme cette courbe est strictement convexe, elle est au dessus de sa tangente (strictement, sauf au point de tangence), et comme le point $(x_{k+1}, 0)$ appartient à la tangente, on a donc $g'(x_{k+1}) > 0$ et donc $x_{k+1} > x_*$ puisque g' est strictement croissante. Comme la pente de la tangente est strictement positive (d'après la question précédente) et que $g'(x_k) > 0$, on obtient bien que $x_{k+1} < x_k$.

(d) On en déduit (par récurrence) que si $x_0 > x_*$, la suite $(x_k)_{k\in\mathbb{N}}$ est décroissante et minorée par x_*, et qu'elle converge donc vers une limite ℓ qui par continuité vérifie $\ell = \ell - \frac{g'(\ell)}{g''(\ell)} = 0$, c'est à dire $g'(\ell) = 0$, donc $\ell = x_*$ (puisqu'on a vu que g' était strictement croissante à la question (b), et que donc la seule solution à $g'(x) = 0$ est x_*).

Si x_0 est négatif, à l'aide du même genre de graphique on peut obtenir que $x_1 > 0$, et donc que l'on se retrouve dans le cas précédent à partir du rang 1.

(e) Ici, on suppose également par récurrence que l'on a $x_* < x_{k+1} < x_k$. Alors x_{k+2} est donné par l'intersection de la sécante à la courbe aux points d'abscisses x_k et x_{k+1} :



On voit bien également que l'on doit avoir $x_* < x_{k+2} < x_{k+1}$. Le fait que $x_{k+2} < x_{k+1}$ provient du fait que $0 < g'(x_{k+1}) < g'(x_k)$ (puisque g' est strictement croissante). Le fait que $x_* < x_{k+2}$ provient de la convexité de g' : si on avait $x_{k+2} < x_*$, on aurait les points sur le segment passant par $(x_{k+2}, 0)$ et $(x_k, g'(x_k))$ au dessus de la courbe (et même strictement sauf pour l'extrémité droite), donc le point $(x_{k+1}, g'(x_{k+1}))$ ne serait pas sur ce segment.

On en conclut par récurrence que la suite $(x_k)_{k\in\mathbb{N}}$ est décroissante et minorée, et converge donc vers une limite ℓ. Pour montrer que $\ell = x_*$, on utilise le théorème des accroissement finis pour g' : il existe $c_k \in]x_{k+1}, x_k[$ tel que $\frac{g'(x_{k+1})-g'(x_k)}{x_{k+1}-x_k} = g''(c_k)$. Par continuité de g'', on obtient donc que $g''(c_k) \to g''(\ell)$ lorsque $k \to \infty$, et donc en passant à la limite dans la définition de x_{k+2} on obtient $\ell = \ell - \frac{g'(\ell)}{g''(\ell)}$, et donc $g'(\ell) = 0$, c'est à dire $\ell = x_*$.

(f) D'après le dessin précédent, on voit également que si on a initialement $x_* < x_0 < x_1$, alors on aura $x_* < x_2 < x_1$ et on se ramène au cas précédent à partir de la première étape.

En faisant les mêmes dessins pour les configurations différentes où on a toujours au moins une itérée parmi x_k et x_{k+1} strictement inférieure à x_*, on obtient les cas suivants :

(i) si $x_k < x_{k+1} < x_*$, alors on a $x_k < x_{k+1} < x_* < x_{k+2}$,

(ii) si $x_{k+1} < x_k < x_*$, alors on a $x_{k+1} < x_k < x_* < x_{k+2}$,

(iii) si $x_k < x_* < x_{k+1}$, alors on a $x_k < x_{k+2} < x_* < x_{k+1}$,

(iv) si $x_{k+1} < x_* < x_k$, alors on a $x_{k+1} < x_{k+2} < x_* < x_k$.

On voit alors que les configurations s'enchainent dans cet ordre : (i)→(iii)→(iv)→(i), et que d'autre part la confuration (ii) mène à la configuration (iii). Donc quelle que soit la configuration de départ, on se retrouve en une, deux ou trois étapes dans la configuration (i). On a donc seulement à traiter le cas où $x_0 < x_1 < x_*$, et on obtient donc $x_0 < x_1 < x_* < x_2$, puis $x_0 < x_1 < x_3 < x_* < x_2$, et enfin après la troisième étape $x_0 < x_1 < x_3 < x_4 < x_* < x_2$.

Il faut ici encore faire un dessin pour voir que la pente de la sécante entre les points d'abscisses x_3 et x_4 est plus grande que celle entre les points d'abscisses x_0 et x_1, et que donc on doit avoir $x_5 < x_2$.

Au final, par récurrence, on obtient donc

$$x_0 < x_1 < x_3 < x_4 < \cdots < x_{3k} < x_{3k+1} < \cdots < x_* < \cdots < x_{3k+2} < \cdots < x_5 < x_2,$$

ce qui donne la convergence comme précédemment de la suite des x_{3k} et des x_{3k+1} vers x_*, et de même pour la suite des x_{3k+2}.

3. Si $x_k \neq x_{k-1}$ on a

$$x_{k+1} = x_k - \frac{x_k - x_{k-1}}{f'(x_k) - f'(x_{k-1})} f'(x_k) = \frac{x_{k-1}f'(x_k) - x_k f'(x_{k-1})}{f'(x_k) - f'(x_{k-1})},$$

qui est bien défini puisque f' est strictement croissante et que donc $f'(x_k) \neq f'(x_{k-1})$. En remplaçant avec l'expression $f'(x) = \frac{x}{\sqrt{1+x^2}}$, on obtient

$$x_{k+1} = x_k x_{k-1} \frac{\frac{1}{\sqrt{1+x_k^2}} - \frac{1}{\sqrt{1+x_{k-1}^2}}}{\frac{x_k}{\sqrt{1+x_k^2}} - \frac{x_{k-1}}{\sqrt{1+x_{k-1}^2}}} = x_k x_{k-1} \frac{\sqrt{1+x_{k-1}^2} - \sqrt{1+x_k^2}}{x_k\sqrt{1+x_{k-1}^2} - x_{k-1}\sqrt{1+x_k^2}}.$$

On multiplie en haut et en bas par les expressions conjuguées $\sqrt{1+x_{k-1}^2}-\sqrt{1+x_k^2}$ (conjuguée du numérateur) et $x_k\sqrt{1+x_{k-1}^2}+x_{k-1}\sqrt{1+x_k^2}$ (pour le dénominateur), et on obtient donc

$$x_{k+1}=x_kx_{k-1}\frac{x_k\sqrt{1+x_{k-1}^2}+x_{k-1}\sqrt{1+x_k^2}}{\sqrt{1+x_{k-1}^2}+\sqrt{1+x_k^2}}\frac{1+x_{k-1}^2-(1-x_k^2)}{x_k^2(1+x_{k-1}^2)-x_{k-1}^2(1+x_k^2)},$$

et la fraction la plus à droite se simplifie en $\frac{x_{k-1}^2-x_k^2}{x_k^2-x_{k-1}^2}=-1$, on obtient donc bien le résultat demandé :

$$x_{k+1}=-x_kx_{k-1}\frac{x_k\sqrt{1+x_{k-1}^2}+x_{k-1}\sqrt{1+x_k^2}}{\sqrt{1+x_{k-1}^2}+\sqrt{1+x_k^2}}.$$

On a que $|x_k\sqrt{1+x_{k-1}^2}+x_{k-1}\sqrt{1+x_k^2}|\leqslant\max(|x_{k-1}|,|x_k|)(\sqrt{1+x_{k-1}^2}+\sqrt{1+x_k^2})$, et donc en passant à la valeur absolue on obtient bien que $|x_{k+1}|\leqslant|x_k||x_{k-1}|\max(|x_k|,|x_{k-1}|)$ dans l'expression précédente.

Montrons par récurrence que x_{k+1} est bien défini, que $|x_{k+1}|<|x_k|\leqslant|x_0|$, et que $|x_{k+1}|\neq 0$ (ce qui est vrai au rang $k=0$). Si c'est vrai jusqu'au rang $k-1$ (avec $k\geqslant 1$) alors comme $|x_k|<|x_{k-1}|\leqslant|x_0|$, on obtient que $x_k\neq x_{k-1}$, et donc x_{k+1} est bien défini et vérifie $|x_{k+1}|\leqslant|x_k||x_0|^2<|x_k|$ (puisque $|x_k|\neq 0$). Et si on avait $x_{k+1}=0$, d'après l'expression précédente de x_{k+1}, on aurait $x_k\sqrt{1+x_{k-1}^2}=-x_{k-1}\sqrt{1+x_k^2}$. En élevant au carré, on obtiendrait $x_k^2=x_{k-1}^2$ ce qui est impossible puisque $|x_k|<|x_{k-1}|$.

On a donc bien une suite $(|x_k|)_{k\in\mathbb{N}}$ bien définie, strictement décroissante et strictement positive. Et on a donc que $\max(|x_k|,|x_{k-1}|)=|x_{k-1}|$ et que donc $|x_{k+1}|\leqslant|x_k||x_{k-1}|^2$ pour tout $k\geqslant 1$.

Comme $|x_0|<1$ et $|x_1|<1$, on peut trouver $0<\alpha<1$ suffisamment proche de 1 de sorte que $|x_0|\leqslant\alpha$ et $|x_1|\leqslant\alpha^2$. Alors par récurrence si on suppose $|x_{k-1}|\leqslant\alpha^{2^{k-1}}$ et $|x_k|\leqslant\alpha^{2^k}$, on obtient bien que $|x_{k+1}|\leqslant\alpha^{2^k}(\alpha^{2^{k-1}})^2=\alpha^{2^{k+1}}$.

On obtient donc que la suite converge vers 0 à un ordre supérieur ou égal à 2. Ce n'est pas en contradiction avec les résultats donnés dans le cours qui disent que dans les bonnes hypothèses, en partant suffisamment proche du minimum, l'ordre de convergence est supérieur ou égal à $\varphi\approx 1,618$. C'est ici parce que la dérivée troisième s'annule au point de minimum que l'on a une convergence plus rapide.