Méthodes numériques : optimisation.
Partiel du 8 mars 2017 — Éléments de correction

1 Gradient à pas fixe pour une fonction quadratique

1. On a $f(x) = \sum_{i=1}^{n} \frac{1}{2}\lambda_i t_i^2 + b_i t_i$. Pour i fixé, la fonction $t \mapsto \frac{1}{2}\lambda_i t^2 + b_i t$ admet un minimum si et seulement si $\lambda_i > 0$ ou $\lambda_i = b_i = 0$ (dans les autres cas elle tend vers $-\infty$ en $\pm\infty$, et donc f n'est pas bornée inférieurement).

   Donc si il existe i tel que $\lambda_i < 0$ ou $\lambda_i = 0$ et $b_i \neq 0$, la fonction $t \mapsto f(te_i)$ n'est pas bornée inférieurement, et f n'a donc pas de minimum. Réciproquement si pour tout i on a $\lambda_i > 0$ ou $\lambda_i = b_i = 0$, alors en prenant $x^* = \sum_{i=1}^{n} t_i^* e_i$ où $t_i^* = -\frac{b_i}{\lambda_i}$ si $\lambda_i > 0$ et t_i^* quelconque si $\lambda_i = b_i = 0$, on a $f(x^*) \leqslant f(x)$ pour tout x, et x_* est donc un minimum.

   On peut reformuler cette condition comme ceci : A est symétrique positive (tous les λ_i sont positifs), et $b \in \mathrm{Im}(A)$ (qui est l'espace vectoriel engendré par les e_i pour lesquels $\lambda_i > 0$).

2. Le calcul de $\nabla f(x) = Ax + b$ a déjà été donné dans le cours (c'est dû au fait que A est symétrique).

   On voit par ce qui précède que si l'on a un unique minimum, on doit déjà avoir $\lambda_i > 0$ ou $\lambda_i = b_i = 0$ pour tout i. S'il existe un i pour lequel $\lambda_i = b_i = 0$, alors comme on peut prendre t_i^* arbitraire dans la construction de $x*$, le minimum n'est pas unique. Donc toutes les valeurs propres sont strictement positives. Réciproquement, si toutes les valeurs propres sont strictement positives (donc A symétrique définie positive, donc inversible), on a par ce qui précède l'existence d'un minimum x_*. C'est donc un point critique de f qui vérifie $\nabla f(x_*) = Ax_* + b = 0$, donc $x_* = -A^{-1}b$, ce qui donne l'unicité.

   La condition nécessaire et suffisante demandée est donc A symétrique définie positive.

3. On a $x_{k+1} = x_k - \alpha\nabla f(x_k) = x_k - \alpha r_k$. Donc $r_{k+1} = A(x_k - \alpha r_k) + b = r_k - \alpha A r_k$.

4. On a $r_{k+1,i} = \langle r_{k+1}, e_i\rangle = \langle r_k - \alpha A r_k, e_i\rangle = r_{k,i} - \alpha\langle r_k, Ae_i\rangle$ (par symétrie de A), et donc $r_{k+1,i} = r_{k,i} - \alpha\lambda_i\langle r_k, e_i\rangle = (1 - \alpha\lambda_i)r_{k,i}$.

5. Si $\lambda_i = 0$, on a $b_i = 0$ d'après la première question, et on a $r_{i,0} = \langle r_0, e_i\rangle = \langle Ax_0 + b, e_i\rangle = \lambda_i\langle x_0, e_i\rangle + b_i = 0$. D'après la question précédente, on a donc par récurrence immédiate que $r_{k,i} = 0$ pour tout $k \in \mathbb{N}$. Et donc en particulier $0 = |r_{k+1,i}| \leqslant \rho(\alpha)|r_{k,i}| = 0$

   De même, dès que $\lambda_i > 0$, on a $|r_{k+1,i}| = |1 - \alpha\lambda_i||r_{k,i}| \leqslant \rho(\alpha)|r_{k,i}|$.

   On a donc $\|r_{k+1}\|^2 = \sum_{i=1}^{n} |r_{k+1,i}|^2 \leqslant \rho(\alpha)^2 \sum_{i=1}^{n} |r_{k+1,i}|^2 = \rho(\alpha)^2\|r_k\|$.

   Pour l'égalité, on prend i_0 tel que λ_i correspond à la valeur propre (ℓ ou L) telle que $|1-\alpha\lambda_i| = \rho(\alpha)$. On pose alors $x_0 = \sum_{i=1}^{n} t_i e_i$, où $t_i = -\frac{b_i}{\lambda_i}$ si $\lambda_i > 0$ et $i \neq i_0$, $t_{i_0} = \frac{1-b_{i_0}}{\lambda_{i_0}}$, et $t_i = 0$ si $\lambda_i = 0$. De sorte que l'on a $r_{0,i} = \langle Ax_0 + b, e_i\rangle = 0$ sauf si $i = i_0$ pour lequel $r_{0,i_0} = 1$. Autrement dit $r_0 = e_{i_0}$. Alors toutes les inégalités précédentes sont des égalités (tous les termes sont nuls pour $i \neq i_0$, et on a $|1 - \alpha\lambda_{i_0}| = \rho(\alpha)$).

6. Si $\alpha \in]0, \frac{2}{L}[$, on a $\rho(\alpha) < 1$ (cf. cours), donc les r_k convergent linéairement vers 0 à un taux inférieur ou égal à $\rho(\alpha)$. On a $x_k = x_0 - \alpha\sum_{0\leqslant j<k} r_j$, et comme la série converge

absolument, on a que x_k converge vers $x_* = x_0 - \alpha \sum_{j\geqslant 0} r_j \in \mathbb{R}^n$. D'autre part on obtient donc directement que $\|x_k - x_*\| = \alpha \| \sum_{j\geqslant k} r_k\| \leqslant \alpha \|r_k\| \frac{1}{1-\rho(\alpha)} \leqslant \frac{\alpha\|r_0\|}{1-\rho(\alpha)}\rho(\alpha)^k$, ce qui donne la convergence linéaire de x_k à un taux inférieur ou égal à $\rho(\alpha)$. Enfin, puisque $r_k = \nabla f(x_k)$ converge vers 0, on en déduit par continuité que $\nabla f(x_*) = 0$. On conclut soit par convexité de f, soit directement en utilisant l'expression des points de minimum de f obtenus à la question 1.

On a vu en cours que le α qui minimise $\rho(\alpha)$ est $\frac{2}{L+\ell}$: faire un graphe des deux fonctions $\alpha \mapsto |1-\alpha\ell|$ et $\alpha \mapsto |1-\alpha L|$, $\alpha = \frac{2}{L+\ell}$ est le point pour lequel $1-\alpha\ell = L\alpha - 1$. Sur $]0, \frac{2}{L+\ell}]$, on a donc $1-\alpha\ell \geqslant L\alpha - 1$, et d'autre part on a toujours $1-\alpha\ell \geqslant 1-\alpha L$. Donc $1-\alpha\ell \geqslant |1-\alpha L| \geqslant 0$, donc $\rho(\alpha) = |1-\alpha\ell| = 1-\alpha\ell$ et décroit avec α. Sur $[\frac{2}{L+\ell}, +\infty[$, on a cette fois-ci $L\alpha - 1 \geqslant 1-\ell\alpha$, et on a toujours $L\alpha - 1 \geqslant \ell\alpha - 1$ donc $L\alpha - 1 \geqslant |1-\ell\alpha| \geqslant 0$, donc $\rho(\alpha) = |L\alpha - 1| = L\alpha - 1$ et croit avec α, on a donc bien que $\alpha = \frac{2}{L+\ell}$ est le pas qui minimise $\rho(\alpha)$.

7. * On a en fait

$$x_k - x_* = \alpha \sum_{i=1}^n \sum_{j\geqslant k} r_{j,i} e_i = \alpha \sum_{i=1}^n \sum_{j\geqslant k} (1-\alpha\lambda_i)^j r_{0,i} e_i = \alpha \sum_{i\in I} \frac{r_{0,i}}{\alpha\lambda_i}(1-\alpha\lambda_i)^k e_i,$$

où I est l'ensemble des indices tels que $\lambda_i > 0$ (pour tous les autres i, le $r_{0,i}$ est nul). Donc

$$\|x_k - x_*\|^2 = \sum_{i\in I} \frac{r_{0,i}^2}{\lambda_i^2}|1-\alpha\lambda_i|^{2k} = \rho(a)^{2k} \sum_{i\in I_1} \frac{r_{0,i}^2}{\lambda_i^2} + \sum_{i\in I_2} \frac{r_{0,i}^2}{\lambda_i^2}|1-\alpha\lambda_i|^{2k} \geqslant C_1^2 \rho(\alpha)^{2k},$$

où I_1 est l'ensemble (non vide) des indices de I tels que $|1-\alpha\lambda_i| = \rho(\alpha)$, et $C_1 = \sqrt{\sum_{i\in I_1} \frac{r_{0,i}^2}{\lambda_i^2}}$. On obtient donc que dès que $C_1 > 0$, le taux de convergence ne peut pas être strictement inférieur à $\rho(\alpha)$, sinon on aurait $C_1\rho(\alpha)^k \leqslant \|x_k - x_*\| \leqslant C\delta^k$ avec $\delta < \rho(\alpha)$. Ce qui donnerait $C_1 \leqslant C(\frac{\delta}{\rho(\alpha)})^k \to 0$, et donc $C_1 = 0$.

On a que $C_1 = 0$ si et seulement si $r_{0,i} = 0$ pour les $i \in I_1$, autrement dit que $r_0 \in V$ où $V = \mathrm{Vect}(\{e_i, 1 \leqslant i \leqslant n, i \notin I_1\})$. Cela équivaut à ce que $x_0 \in \sum_{i\in I} \frac{-b_i}{\lambda_i} e_i + V$, qui est un espace affine de dimension strictement inférieure à n donc bien de mesure nulle pour la mesure de Lebesgue en dimension n.

2 Méthode de Newton approchée par différence finies

1. On a $f'(x) = 1 - \frac{2}{x^2}$, strictement négative sur $]0, \sqrt{2}[$ et strictement positive sur $]\sqrt{2}, +\infty[$. Donc f est unimodale, avec un minimum global strict en $x_* = \sqrt{2}$. Puis $f''(x) = \frac{4}{x^3}$.
2. L'approximation par différences finies centrées de paramètre ε pour une fonction g est $g(x) \approx \frac{g(x+\varepsilon)-g(x-\varepsilon)}{2\varepsilon}$. On approxime donc $f''(x)$ par $\frac{f'(x+\frac{\varepsilon}{2})-f'(x-\frac{\varepsilon}{2})}{\varepsilon}$, et en approximant chacune des dérivées cela devient $\frac{\frac{f(x+\varepsilon)-f(x)}{\varepsilon} - \frac{f(x)-f(x-\varepsilon)}{\varepsilon}}{\varepsilon} = \frac{f(x+\varepsilon)+f(x-\varepsilon)-2f(x)}{\varepsilon^2}$.
3. Voilà le code qui a donné le graphique du sujet :

```
def f(x):
    return x+2/x
def fseconde(x):
    return 4/x**3

def derivesecond(f,x,eps=1e-8):
```

```
    return ((f(x+eps)+f(x-eps)-2*f(x)))/eps**2

x=1.5
leps=logspace(-1,-9,100)
lerr=[abs(derivesecond(f,x,eps)-fseconde(x)) for eps in leps]
loglog(leps,lerr)
xlabel("$\\varepsilon$")
ylabel("$|f''(x)-f''_{\mathrm{approx},\\varepsilon}(x)|$")
```

On observe que l'erreur est grande si ε est trop petit ou trop grand : si ε est grand, la formule d'approximation par différences finies est mauvaise, et si ε est trop petit, les erreurs numériques dûes au fait que les calculs machines de $f(x+\varepsilon)$, $f(x-\varepsilon)$, et $f(x)$ ne sont pas exacts, et cette erreur numérique est amplifiée par la formule dans laquelle on divise par ε^2. Les erreurs numériques sont donc amplifiées lorsque ε diminue.

4. On écrit la formule de Taylor à l'ordre 4 :

$$f(x+h) = f(x) + hf'(x) + \frac{h^2}{2}f''(x) + \frac{h^3}{6}f^{(3)}(x) + \frac{h^4}{24}f^{(4)}(x+\theta_h h),$$

où $\theta_h \in]0,1[$. En prenant $h=\varepsilon$ puis $h=-\varepsilon$ et en sommant les deux égalités obtenues, on obtient

$$f(x+\varepsilon) + f(x-\varepsilon) = 2f(x) + \varepsilon^2 f''(x) + \frac{\varepsilon^4}{24}[f(x+\theta_\varepsilon\varepsilon) + f(x-\theta_{-\varepsilon}\varepsilon)],$$

et donc $|\frac{f(x+\varepsilon)+f(x-\varepsilon)-2f(x)}{\varepsilon^2} - f''(x)| = \frac{\varepsilon^2}{24}|f(x+\theta_\varepsilon\varepsilon) + f(x-\theta_{-\varepsilon}\varepsilon)| \leqslant \frac{\varepsilon^2}{12}C_4$.
On a donc ensuite

$$\begin{aligned}\left|\frac{\hat f(x+\varepsilon)+\hat f(x-\varepsilon)-2\hat f(x)}{\varepsilon^2} - f''(x)\right| &\leqslant \left|\frac{\hat f(x+\varepsilon)+\hat f(x-\varepsilon)-2\hat f(x)}{\varepsilon^2} - \frac{f(x+\varepsilon)+f(x-\varepsilon)-2f(x)}{\varepsilon^2}\right| + \frac{C_4}{12}\varepsilon^2\\ &\leqslant \frac{|\hat f(x+\varepsilon)-f(x+\varepsilon)|+|\hat f(x-\varepsilon)-f(x-\varepsilon)|+2|f(x)-\hat f(x)|}{\varepsilon^2} + \frac{C_4}{12}\varepsilon^2\\ &\leqslant \frac{4\eta}{\varepsilon^2} + \frac{C_4\varepsilon^2}{12}.\end{aligned}$$

Pour minimiser cette erreur totale, les deux termes doivent être du même ordre de grandeur (si le deuxième terme est négligeable devant le premier, en augmentant ε on diminue l'erreur totale, et vice-versa), donc on obtient que cette erreur totale est minimale lorsque $\varepsilon^2 \approx \eta\varepsilon^2$, soit encore ε de l'ordre de $\eta^{\frac{1}{4}}$. Plus précisément, la fonction $h(t) = \frac{a}{t} + bt$ avec a, b strictement positifs admet son unique minimum en $t_* = \sqrt{\frac{a}{b}}$. Dans notre cas on obtient le minimum de $\frac{4\eta}{\varepsilon^2} + \frac{C_4\varepsilon^2}{12}$ pour $\varepsilon^2 = \sqrt{\frac{12\cdot 4\eta}{C_4}} = 4\sqrt{\frac{3\eta}{C_4}}$, soit encore $\varepsilon = 2(\frac{3\eta}{C_4})^{\frac{1}{4}}$, ce qui est bien de la forme $\varepsilon = C\eta^{\frac{1}{4}}$ avec C de l'ordre de grandeur de 1.
Dans le cas pratique du graphique, on sait que la précision machine (pour des nombres à virgule flottante) est de l'ordre de 10^{-16}. On obtient donc bien une erreur minimale lorsque ε est de l'ordre de 10^{-4}, et l'erreur totale est de l'ordre de $\frac{\eta}{\varepsilon^2} \approx \varepsilon^2$ soit ici 10^{-8}. On voit également que l'erreur est proportionnelle à ε^2 lorsque $\varepsilon \gg 10^{-4}$: lorsque ε est multiplié par 10, l'erreur est multipliée par 100 (la partie droite du graphique est une droite de pente 2, pour les échelles logarithmiques en abscisses et en ordonnées). On voit également que l'erreur se comporte proportionnellement à $\frac{1}{\varepsilon^2}$ lorsque $\varepsilon \ll 10^{-4}$: cette fois-ci c'est une droite de pente -2 dans les échelles logarithmiques. La courbe est moins proche d'une droite régulière : les erreurs numériques peuvent être plus ou moins grande, elles ne sont pas toutes exactement égales à η.

5. Méthode de Newton : $x_{k+1} = x_k - \frac{f'(x_k)}{f''(x_k)}$. Si on approxime f' par la formule des différences finies centrées en x_k : $\frac{f(x+\varepsilon)-f(x-\varepsilon)}{2\varepsilon}$, et $f''(x_k)$ par la formule de la question 2, on obtient exactement la formule de récurrence modifiée par rapport à la méthode de Newton : $x_{k+1} = x_k - \frac{\varepsilon[f(x_k+\varepsilon)-f(x_k-\varepsilon)]}{2[f(x_k+\varepsilon)+f(x_k-\varepsilon)-2f(x_k)]}$.
6. Voilà le code qui a donné le graphique de l'énoncé :

```
def newton(f,x,eps=1e-5):
    fp=f(x+eps)
    fm=f(x-eps)
    return x-eps/2*(fp-fm)/(fp+fm-2*f(x))

for eps in [1e-3,1e-4,1e-5,1e-6,1e-7]:
    x=2
    l=[x]
    for j in range(12):
        x=newton(f,x,eps)
        l.append(x)

    semilogy(abs(array(l)-sqrt(2)))
legend(["$\\varepsilon=10^{-"+str(i)+"}$" for i in range(3,8)])
xlabel("itération $k$")
ylabel("erreur $|x_k-x_*|$")
```

   Lorsqu'on exécute ce code en prenant $\varepsilon = 10^{-8}$, on devrait avoir $f(x+\varepsilon)+f(x+\varepsilon)-2f(x)$ de l'ordre de 10^{-16} (puisqu'en divisant par ε^2 on obtient une approximation de $f''(x)$). Mais comme la précision machine est aussi de l'ordre de 10^{-16}, on atteint la précision machine, et il se peut donc que $\hat{f}(x+\varepsilon)+\hat{f}(x+\varepsilon)-2\hat{f}(x)$ soit vraiment nul. Lorsqu'on effectue la division, on obtient donc une erreur.
7. * Le début de la question a été fait en TD (on fait une formule de Taylor à l'ordre 3 pour f, en $x+\varepsilon$ et $x-\varepsilon$, que l'on soustrait pour obtenir $|\frac{f(x+\varepsilon)-f(x-\varepsilon)}{2\varepsilon} - f'(x)| \leqslant \frac{C_3}{6}\varepsilon^2$, où C_3 est une borne sur $f^{(3)}$, puis on fait la même inégalité triangulaire pour comparer par rapport à la même formule avec $\hat{f}$, ce qui fait apparaître un terme $\frac{\eta}{\varepsilon}$ en plus). On obtient donc que

$$\frac{\varepsilon[\hat{f}(x+\varepsilon)-\hat{f}(x-\varepsilon)]}{2[\hat{f}(x+\varepsilon)+\hat{f}(x-\varepsilon)-2\hat{f}(x)]} = \frac{f'(x)+O(\frac{\eta}{\varepsilon}+\varepsilon^2)}{f''(x)+O(\frac{\eta}{\varepsilon^2}+\varepsilon^2)}.$$

   Lorsqu'on suppose que $f''(x)$ est bornée inférieurement et supérieurement par une constante ℓ de l'ordre de 1 (ce qui est le cas au voisinage d'un minimum non dégénéré), on obtient

$$\frac{\varepsilon[\hat{f}(x+\varepsilon)-\hat{f}(x-\varepsilon)]}{2[\hat{f}(x+\varepsilon)+\hat{f}(x-\varepsilon)-2\hat{f}(x)]} = \frac{f'(x)}{f''(x)} + O(\frac{\eta}{\varepsilon}+\varepsilon^2+f'(x)(\frac{\eta}{\varepsilon^2}+\varepsilon^2)).$$

   Donc tant que $f'(x)$ n'est pas trop proche de 0, le ε qui minimise cette erreur entre la méthode de Newton et la formule approchée est lorsque $\frac{\eta}{\varepsilon^2}$ est du même ordre de grandeur que ε^2, soit $\varepsilon \approx \eta^{\frac{1}{4}}$.
   Cependant lorsque $f'(x)$ est trop proche de 0 (donc quand on est proche du minimum), le raisonnement n'est plus valide, puisque cette fois-ci, par exemple si $f'(x)$ est de l'ordre de ε, l'erreur par rapport à la méthode de Newton est plutôt de l'ordre de $\frac{\eta}{\varepsilon}+\varepsilon^2$, qui correspond juste à l'erreur qu'on fait en approchant $f'(x)$ par différence finie centrée. Quitte à perdre

de la proximité avec la méthode de Newton sur les premières itérations, on a donc intérêt à minimiser cette erreur finale si on veut obtenir plus de précision pour localiser le minimum, ce qui correspond plutôt à prendre un ε de l'ordre de $\eta^{\frac{1}{3}}$, ce qui correspond bien à quelque chose entre 10^{-5} et 10^{-6} quand les calculs se font en virgule flottante avec une précision de l'ordre de 10^{-16}.

3 Méthode de descente de gradient à pas fixe pour une fonction convexe

1. La formule de Taylor à l'ordre 2 donne

$$f(x+h) = f(x) + \langle \nabla f(x), h\rangle + \int_0^1 t\langle h, H_f(x+th)h\rangle \mathrm{d}t.$$

Et comme on a $0 \leqslant \int_0^1 t\langle h, H_f(x+th)h\rangle \mathrm{d}t \leqslant \int_0^1 tL\|h\|^2 \mathrm{d}t = \frac{L}{2}\|h\|^2$, on obtient l'encadrement indiqué.
En remplaçant x par x_* on obtient que $0 \leqslant f(x_*+h) - f(x_*)$ pour tout $h \in \mathbb{R}^n$, autrement dit que $f(x) \geqslant f(x_*)$ pour tout $x \in \mathbb{R}^n$. C'est donc un point de minimum global, qui n'est pas forcément unique : toute fonction constante vérifie les hypothèses données.

2. On applique la partie de droite de l'encadrement avec $h = -\alpha\nabla f(x_k)$ et $x = x_k$. On a donc $x_{k+1} = x + h$, et donc

$$f(x_{k+1}) - f(x_k) + \alpha\langle \nabla f(x_k), \nabla f(x_k)\rangle \leqslant \frac{L\alpha^2}{2}\|\nabla f(x_k)\|^2,$$

ce qui après simplification, est exactement l'estimation demandée.

3. En posant $F(x) = x - \alpha\nabla f(x)$, on a une fonction F qui est C^1 dont la dérivée en x est $I_n - \alpha H_f(x)$. On a $x_{k+1} = F(x_k)$, et $x_* = F(x_*)$.
De sorte que $x_{k+1} - x_* = F(x_k) - F(x_*) = \int_0^1 F'(x_* + t(x_k - x_*)) \cdot (x_k - x_*) \mathrm{d}t$ (par la formule de Taylor à l'ordre 1), ce qui correspont exactement à la formule demandée.
Ensuite, on a d'après le lemme du cours que $\|[I_n - \alpha H_f(x)]h\| \leqslant \max_i |\lambda_i| \|h\|$, où les λ_i sont les valeurs propres de $I_n - \alpha H_f(x)$, qui sont donc dans $[1-\alpha L, 1]$ d'après l'hypothèse principale. Donc comme $\alpha < \frac{2}{L}$, on a $1 - \alpha L > -1$, et donc les λ_i sont dans $[-1, 1]$. On en conclut donc que $\|[I_n - \alpha H_f(x)]h\| \leqslant \|h\|$. On a donc (avec l'indication donnée), que

$$\|x_{k+1} - x_*\| \leqslant \int_0^1 \|[I_n - \alpha H_f(x_* + t(x_k - x_*))](x_k - x_*)\| \mathrm{d}t \leqslant \int_0^1 \|x_k - x_*\| \mathrm{d}t = \|x_k - x_*\|.$$

4. Comme la suite des $\|x_k - x_*\|$ est décroissante, on obtient que la suite des x_k est bornée, on peut en extraire une suite $(x_{\varphi(k)})$ convergente vers $\bar{x}$. D'après la question 2, comme $0 < \alpha < \frac{2}{L}$, la suite des $f(x_k)$ est décroissante et converge vers une valeur finie (f est bornée inférieurement par $f(x_*)$), et donc par la même question 2, on en déduit que $\|\nabla f(x_k)\| \leqslant \sqrt{\frac{f(x_k) - f(x_{k+1})}{\alpha(1 - \alpha\frac{L}{2})}}$ converge vers 0. En passant à la limite dans la suite extraite, comme f est C^1, on obtient donc que $\|\nabla f(\bar{x})\| = 0$.
On obtient donc un point $\bar{x}$ qui satisfait exactement les mêmes hypothèses que x_*. Et donc d'après la question 3, la suite des $\|x_k - \bar{x}\|$ est décroissante et converge vers une limite r. Et donc $\|x_{\varphi(k)} - \bar{x}\|$ converge vers r. Comme on avait $x_{\varphi(k)}$ qui convergeait vers $\bar{x}$, on en déduit que $r = 0$, et que donc x_k converge vers $\bar{x}$.

5. On utilise la partie gauche de l'encadrement de la question 1 avec $x = x_k$ et $h = x_* - x_k$ pour obtenir que $\varepsilon_k \leqslant \langle \nabla f(x_k), x_k - x_* \rangle \leqslant \|\nabla f(x_k)\| \|x_k - x_*\|$ par Cauchy-Schwarz. Le fait que la suite des $\|x_k - x_*\|$ soit décroissante donne le résultat voulu.

6. On a $\varepsilon_{k+1} - \varepsilon_k = f(x_{k+1}) - f(x_k) \leqslant -\alpha(1-\alpha\frac{L}{2})\|\nabla f(x_k)\|^2 \leqslant -\alpha(1-\alpha\frac{L}{2})\| \frac{\varepsilon_k^2}{\|x_0-x_*\|^2}$ d'après les questions 2 et 5, ce qui donne l'estimation demandée.

   En divisant des deux côtés par $\varepsilon_k \varepsilon_{k+1}$ (on suppose qu'ils sont non nuls sinon l'inégalité à démontrer n'a pas de sens, et on aurait alors la suite des $f(x_k)$ stationnaire, et donc $\nabla f(x_k)$ nul d'après la question 2, et donc enfin la suite des x_k stationnaire, ce qui rend la conclusion de la question 7 évidemment valide), on obtient $\frac{1}{\varepsilon_k} - \frac{1}{\varepsilon_{k+1}} \leqslant -\alpha(1-\alpha\frac{L}{2})\frac{\varepsilon_k}{\varepsilon_{k+1}\|x_0-x_*\|^2} \leqslant -\frac{\alpha(1-\alpha\frac{L}{2})}{\|x_0-x_*\|^2}$, puisque $\varepsilon_{k+1} \leqslant \varepsilon_k$ et que $-\alpha(1-\alpha\frac{L}{2}) < 0$. On obtient donc la deuxième estimation.

7. La récurrence est immédiate pour la première estimation. En l'inversant, on obtient

$$f(x_k) - f(x_*) = \varepsilon_k \leqslant \frac{1}{\frac{1}{\varepsilon_0} + k\frac{\alpha(1-\frac{L}{2}\alpha)}{\|x_0-x_*\|^2}} = \frac{\varepsilon_0}{1 + k\frac{\alpha(1-\frac{L}{2}\alpha)\varepsilon_0}{\|x_0-x_*\|^2}} = \frac{f(x_0) - f(x_*)}{1 + C\,k},$$

   où $C = \frac{\alpha(1-\frac{L}{2}\alpha)[f(x_0)-f(x_*)]}{\|x_0-x_*\|^2}$.

8. * On écrit $x_{k+1} = x_k - \alpha p x_k^{p-1}$. La fonction satisfait bien les hypothèses, au moins sur un intervalle $[-r, r]$ donné (on pourrait modifier la fonction en dehors de cet intervalle pour qu'elle satisfasse les conditions sur $\mathbb{R}$), et si on prend $x_0 > 0$ suffisamment petit (de telle sorte que $\alpha p x_0^{p-2} < 1$ et que $x_0 \in]0, r]$), on a $x_{k+1} = x_k(1 - \alpha p x_k^{p-2})$, ce qui donne par récurrence que la suite des x_k est décroissante et strictement positive, donc x_k reste dans l'intervalle donné (et donc peu importe comment on aurait modifié la fonction en dehors de $[-r, r]$ si on voulait rentrer exactement dans le cadre des hypothèses sur $\mathbb{R}$). On obtient également que x_k converge vers ℓ, et à la limite $\ell = \ell(1 - \alpha p \ell^{p-2})$, ce qui n'est possible que si $\ell = 0$. Donc $x_k \to 0$ (on pouvait aussi directement l'obtenir en utilisant le résultat des questions précédentes, vu qu'on sait alors que $f(x_k) \to f(x_*)$ avec ici $x_* = 0$, donc $x_k^p \to 0$). Ensuite on écrit $x_{k+1}^{-p+2} = x_k^{-p+2}(1-\alpha p x_k^{p-2})^{-p+2} = x_k^{-p+2}(1-\alpha p(-p+2)x_k^{p-2} + o(x_k^{p-2}))$ puisque $x_k \to 0$. Soit encore $x_{k+1}^{-p+2} = x_k^{-p+2} + \alpha p(p-2) + o(1)$ donc $\frac{1}{(x_{k+1})^{p-2}} - \frac{1}{(x_k)^{p-2}}$ converge vers $\alpha p(p-2)$.

   En faisant une somme téléscopique, on espère donc obtenir que $\frac{1}{(x_k)^p} - \frac{1}{(x_0)^{p-2}}$ se comporte comme $\alpha p(p-2)k$ (c'est la convergence des moyennes de Cesáro) et que donc $x_k \sim [ap(p-2)k]^{\frac{-1}{p-2}}$. Pour le montrer, on prend $\varepsilon > 0$ et on obtient un k_0 tel que $\frac{1}{(x_{k+1})^{p-2}} - \frac{1}{(x_k)^{p-2}} \in [\alpha p(p-2) - \frac{\varepsilon}{2}, \alpha p(p-2) + \frac{\varepsilon}{2}]$ pour $k \geqslant k_0$. On obtient donc en faisant la somme entre k_0 et k (exclus) que pour tout $k \geqslant k_0$

$$(\alpha p(p-2) - \frac{\varepsilon}{2})(k-k_0) \leqslant \frac{1}{(x_k)^{p-2}} - \frac{1}{(x_{k_0})^{p-2}} \leqslant (\alpha p(p-2) + \frac{\varepsilon}{2})(k-k_0),$$

   soit encore $\alpha p(p-2) - \frac{\varepsilon}{2} + \frac{a}{k} \leqslant \frac{1}{k(x_k)^{p-2}} \leqslant \alpha p(p-2) - \frac{\varepsilon}{2} + \frac{b}{k}$, ou a et b dépendent de k_0,ε,x_0, mais pas de k. Donc en prenant $k_1 \geqslant k_0$ assez grand, on a pour tout $k \geqslant k_1$ $\frac{1}{k(x_k)^{p-2}} \in [\alpha p(p-2) - \varepsilon, \alpha p(p-2) + \varepsilon]$. On obtient donc bien que $\frac{1}{k(x_k)^{p-2}} \to \alpha p(p-2)$, soit encore que $k^{\frac{1}{p-2}} x_k \to [\alpha p(p-2)]^{\frac{-1}{p-2}}$, ce qui était ce que l'on voulait.

   En conclusion on obtient une suite pour laquelle $f(x_k) - f(0) = x_k^p \sim \frac{C}{k^{\frac{p}{p-2}}}$. On ne peut donc pas obtenir une puissance meilleure que 1 (ce qui est le résultat de la question 7) au dénominateur en toute généralité, puisque en prenant p assez grand, on a $\frac{p}{p-2}$ aussi proche de 1 que l'on veut.