3.3.5 Exercices (algorithmes pour l'optimisation sans contraintes)

Exercice 112 (Mise en oeuvre de GPF, GPO)**.** *Corrigé en page 244.*
On considère la fonction $f : \mathbb{R}^2 \to \mathbb{R}$ définie par $f(x_1, x_2) = 2x_1^2 + x_2^2 - x_1x_2 - 3x_1 - x_2 + 4$.

1. Montrer qu'il existe un unique $\bar{x} \in \mathbb{R}^2$ tel que $\bar{x} = \min_{x\in\mathbb{R}^2} f(x)$ et le calculer.
2. Calculer le premier itéré donné par l'algorithme du gradient à pas fixe (GPF) et du gradient à pas optimal (GPO), en partant de $(x_1^{(0)}, x_2^{(0)}) = (0, 0)$, pour un pas de $\alpha = .5$ dans le cas de GPF.

Exercice 113 (Convergence de l'algorithme du gradient à pas optimal)**.** *Suggestions en page 244. Corrigé détaillé en page 245*
Soit $f \in C^2(\mathbb{R}^n, \mathbb{R})$ t.q. $f(x) \to \infty$ quand $|x| \to \infty$. Soit $x_0 \in \mathbb{R}^n$. On va démontrer dans cet exercice la convergence de l'algorithme du gradient à pas optimal.

1. Montrer qu'il existe $R > 0$ t.q. $f(x) > f(x_0)$ pour tout $x \notin B_R$, avec $B_R = \{x \in \mathbb{R}^n, |x| \le R\}$.
2. Montrer qu'il existe $M > 0$ t.q. $|H(x)y \cdot y| \le M|y|^2$ pour tout $y \in \mathbb{R}^n$ et tout $x \in B_{R+1}$ ($H(x)$ est la matrice hessienne de f au point x, R est donné à la question 1).
3. (Construction de "la" suite $(x_k)_{k\in\mathbb{N}}$ de l'algorithme du gradient à pas optimal.) On suppose x_k connu ($k \in \mathbb{N}$). On pose $w_k = -\nabla f(x_k)$. Si $w_k = 0$, on pose $x_{k+1} = x_k$. Si $w_k \ne 0$, montrer qu'il existe $\overline{\rho} > 0$ t.q. $f(x_k + \rho w_k) \le f(x_k + \rho w_k)$ pour tout $\rho \ge 0$. On <u>choisit</u> alors un $\rho_k > 0$ t.q. $f(x_k + \rho_k w_k) \le f(x_k + \rho w_k)$ pour tout $\rho \ge 0$ et on pose $x_{k+1} = x_k + \rho_k w_k$.
   On considère, dans les questions suivantes, la suite $(x_k)_{k\in\mathbb{N}}$ ainsi construite.
4. Montrer que (avec R et M donnés aux questions précédentes)
   (a) la suite $(f(x_k))_{k\in\mathbb{N}}$ est une suite convergente,
   (b) $x_k \in B_R$ pour tout $k \in \mathbb{N}$,
   (c) $f(x_k + \rho w_k) \le f(x_k) - \rho|w_k|^2 + (\rho^2/2)M|w_k|^2$ pour tout $\rho \in [0, 1/|w_k|]$.
   (d) $f(x_{k+1}) \le f(x_k) - |w_k|^2/(2M)$, si $|w_k| \le M$.
   (e) $-f(x_{k+1}) + f(x_k) \ge |w_k|^2/(2\overline{M})$, avec $\overline{M} = \sup(M, \tilde{M})$,
   $\tilde{M} = \sup\{|\nabla f(x)|, x \in B_R\}$.
5. Montrer que $\nabla f(x_k) \to 0$ (quand $k \to \infty$) et qu'il existe une sous suite $(n_k)_{k\in\mathbb{N}}$ t.q. $x_{n_k} \to x$ quand $k \to \infty$ et $\nabla f(x) = 0$.
6. On suppose qu'il existe un unique $\overline{x} \in \mathbb{R}^n$ t.q. $\nabla f(\overline{x}) = 0$. Montrer que $f(\overline{x}) \le f(x)$ pour tout $x \in \mathbb{R}^n$ et que $x_k \to \overline{x}$ quand $k \to \infty$.

Exercice 114 (Jacobi et optimisation)**.** *Corrigé détaillé en page 248*
Rappel Soit $f \in C^1(\mathbb{R}^n, \mathbb{R})$; on appelle **méthode de descente à pas fixe** $\alpha \in \mathbb{R}_+^*$ pour la minimisation de f, une suite définie par

$$\boldsymbol{x}^{(0)} \in \mathbb{R}^n \text{ donné},$$
$$\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} + \alpha \boldsymbol{w}^{(k)}$$

où $\boldsymbol{w}^{(k)}$ est une **direction de descente stricte** en $\boldsymbol{x}^{(k)}$, c.à.d. $\boldsymbol{w}^{(k)} \in \mathbb{R}^n$ vérifie la condition $\boldsymbol{w}^{(k)} \cdot \nabla f(x^{(k)}) < 0$.
Dans toute la suite, on considère la fonction f de $\mathbb{R}^n$ dans $\mathbb{R}$ définie par

$$f(\boldsymbol{x}) = \frac{1}{2} A\boldsymbol{x} \cdot \boldsymbol{x} - \boldsymbol{b} \cdot \boldsymbol{x}, \tag{3.36}$$

où A une matrice carrée d'ordre n, symétrique définie positive, et $\boldsymbol{b} \in \mathbb{R}^n$. On pose $\bar{x} = A^{-1}b$.

1. Montrer que la méthode de Jacobi pour la résolution du système $A\boldsymbol{x} = \boldsymbol{b}$ peut s'écrire comme une méthode de descente à pas fixe pour la minimisation de la fonction f définie par (3.36). Donner l'expression du pas α et de la direction de descente $\boldsymbol{w}^{(k)}$ à chaque itération k et vérifier que c'est bien une direction de descente stricte si $\boldsymbol{x}^{(k)} \ne A^{-1}\boldsymbol{b}$.

2. On cherche maintenant à améliorer la méthode de Jacobi en prenant non plus un pas fixe dans l'algorithme de descente ci-dessus, mais un pas optimal qui est défini à l'itération k par
$$f(\boldsymbol{x}^{(k)} + \alpha_k \boldsymbol{w}^{(k)}) = \min_{\alpha>0} f(\boldsymbol{x}^{(k)} + \alpha \boldsymbol{w}^{(k)}), \tag{3.37}$$
où $\boldsymbol{w}^{(k)}$ est défini à la question précédente. On définit alors une méthode de descente à pas optimal par :
$$\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} + \alpha_k \boldsymbol{w}^{(k)}.$$
On appelle cette nouvelle méthode "méthode de Jacobi à pas optimal".
   (a) Justifier l'existence et l'unicité du pas optimal défini par (3.37), et donner son expression à chaque itération.
   (b) Montrer que $|f(\boldsymbol{x}^{(k)}) - f(\boldsymbol{x}^{(k+1)})| = \dfrac{|\boldsymbol{r}^{(k)} \cdot \boldsymbol{w}^{(k)}|^2}{2A\boldsymbol{w}^{(k)} \cdot \boldsymbol{w}^{(k)}}$ si $w^{(k)} \neq 0$.
   (c) Montrer que $\boldsymbol{r}^{(k)} \to \boldsymbol{0}$ lorsque $k \to +\infty$, et en déduire que la suite donnée par la méthode de Jacobi à pas optimal converge vers la solution $\bar{\boldsymbol{x}}$ du système linéaire $A\boldsymbol{x} = \boldsymbol{b}$.
   (d) On suppose que la diagonale extraite D de la matrice A (qui est symétrique définie positive) est de la forme $D = \alpha\mathrm{Id}$ avec $\alpha \in \mathbb{R}$.
      i. Ecrire l'algorithme de descente à pas optimal dans ce cas.
      ii. Comparer les algorithmes de descente obtenus par Jacobi et Jacobi à pas optimal avec les algorithmes de gradient que vous connaissez.

Exercice 115 (Fonction non croissante à l'infini). *Suggestions en page 244.*

Soient $n \geq 1$, $f \in C^2(\mathbb{R}^n, \mathbb{R})$ et $a \in \mathbb{R}$. On suppose que $A = \{x \in \mathbb{R}^n;\ f(x) \leq f(a)\}$ est un ensemble borné de $\mathbb{R}^n$ et qu'il existe $M \in \mathbb{R}$ t.q. $|H(x)y \cdot y| \leq M|y|^2$ pour tout $x,\ y \in \mathbb{R}^n$ (où $H(x)$ désigne la matrice hessienne de f au point x).

1. Montrer qu'il existe $\overline{x} \in A$ t.q. $f(\overline{x}) = \min\{f(x),\ x \in \mathbb{R}^n\}$ (noter qu'il n'y a pas nécessairement unicité de $\overline{x}$).
2. Soit $x \in A$ t.q. $\nabla f(x) \neq 0$. On pose $T(x) = \sup\{\alpha \geq 0; [x, x - \alpha\nabla f(x)] \subset A\}$. Montrer que $0 < T(x) < +\infty$ et que $[x, x - T(x)\nabla f(x)] \subset A\}$ (où $[x, x - T(x)\nabla f(x)]$ désigne l'ensemble $\{tx + (1-t)(x - T(x)\nabla f(x)), t \in [0,1]\}$.
3. Pour calculer une valeur appochée de $\overline{x}$ (t.q. $f(\overline{x}) = \min\{f(x), x \in \mathbb{R}^n\}$), on propose l'algorithme suivant :
Initialisation : $x_0 \in A$,
Itérations : Soit $k \geq 0$.
Si $\nabla f(x_k) = 0$, on pose $x_{k+1} = x_k$. Si $\nabla f(x_k) \neq 0$, on choisit $\alpha_k \in [0, T(x_k)]$ t.q. $f(x_k - \alpha_k \nabla f(x_k)) = \min\{f(x_k - \alpha\nabla f(x_k)),\ 0 \leq \alpha \leq T(x_k)\}$ (La fonction T est définie à la question 2) et on pose $x_{k+1} = x_k - \alpha_k \nabla f(x_k)$.

   (a) Montrer que, pour tout $x_0 \in A$, l'algorithme précédent définit une suite $(x_k)_{k\in\mathbb{N}} \subset A$ (c'est–à–dire que, pour $x_k \in A$, il existe bien au moins un élément de $[0, T(x_k)]$, noté α_k, t.q. $f(x_k - \alpha_k\nabla f(x_k) = \min\{f(x_k - \alpha\nabla f(x_k)), 0 \leq \alpha \leq T(x_k)\}$).
   (b) Montrer que cet algorithme n'est pas nécessairement l'algorithme du gradient à pas optimal. [on pourra chercher un exemple avec $n = 1$.]
   (c) Montrer que $f(x_k) - f(x_{k+1}) \geq \frac{|\nabla f(x_k)|^2}{2M}$, pour tout $k \in \mathbb{N}$.
4. On montre maintenant la convergence de la suite $(x_k)_{k\in\mathbb{N}}$ construite à la question précédente.
   (a) Montrer qu'il existe une sous suite $(x_{k_\ell})_{\ell\in\mathbb{N}}$ et $x \in A$ t.q. $x_{k_\ell} \to x$, quand $\ell \to \infty$, et $\nabla f(x) = 0$.
   (b) On suppose, dans cette question, qu'il existe un et un seul élément $z \in A$ t.q. $\nabla f(z) = 0$. Montrer que $x_k \to z$, quand $k \to \infty$, et que $f(z) = \min\{f(x), x \in A\}$.