Optimisation
Master 1 Statistique & Data Science, Ingénierie Mathématique, 2020-2021

Examen partiel du mercredi 31 mars 2021
Corrigé

Exercice 1. *(sur environ 4 points)*
Soient $\alpha \in \mathbb{R}$ et $f_\alpha : \mathbb{R}^2 \to \mathbb{R}$ la fonctionnelle définie par

$$f_\alpha(x) = \alpha^2 x_1^2 + \frac{1}{2}x_2^2 + \alpha x_1 x_2 + \alpha^2 x_1 + x_2 - 5.$$

1. Montrer que f est une fonctionnelle quadratique et préciser la matrice symétrique A, le vecteur b et le réel c pour lesquels

$$f_\alpha(x) = \frac{1}{2}\langle Ax, x\rangle - \langle b, x\rangle + c.$$

2. Pour quelles valeurs de α la fonction f_α est-elle
   2.a) convexe ?
   2.b) strictement convexe ?
3. Pour quelles valeurs de α le problème d'optimisation

$$\inf_{x\in\mathbb{R}^2} f_\alpha(x)$$

admet des solutions ? En cas d'existence préciser s'il y a unicité de la solution. *On ne demande pas de calculer les solutions.*

Solution de l'exercice 1.

1. On a

$$f_\alpha(x) = \frac{1}{2}\langle Ax, x\rangle - \langle b, x\rangle + c$$

avec

$$A = \begin{pmatrix} 2\alpha^2 & \alpha \\ \alpha & 1 \end{pmatrix} \quad , \quad b = \begin{pmatrix} -\alpha^2 \\ -1 \end{pmatrix} \quad \text{et} \quad c = -5.$$

2. D'après le cours une fonctionnelle quadratique est :
— convexe si et seulement si A est positive.
— strictement convexe si et seulement si A est définie positive.
Ici $A = \begin{pmatrix} 4\alpha^2 & 2\alpha \\ 2\alpha & 8 \end{pmatrix}$. Notons λ_1 et λ_2 les deux valeurs propres réelles de A. On cherche des conditions sur α pour que λ_1 et λ_2 soient (strictement) positives. On a

$$\mathrm{Tr}(A) = \lambda_1 + \lambda_2 = 2\alpha^2 + 1 > 0$$

et

$$\det(A) = \lambda_1\lambda_2 = 2\alpha^2 - \alpha^2 = \alpha^2 \geqslant 0.$$

   2.a) Donc f est convexe si et seulement si λ_1 et λ_2 sont toutes deux positives, soit si et seulement si $\det(A) = \lambda_1\lambda_2 = 28\alpha^2 \geqslant 0$ ce qui est vrai pour tout $\alpha \in \mathbb{R}$.

2.b) De même f est strictement convexe si et seulement si et seulement si $\det(A) = \lambda_1\lambda 2 = 28\alpha^2 > 0$ soit si et seulement si $\alpha \neq 0$.

3. D'après le cours " $\inf_{x\in\mathbb{R}^n} f(x)$ est fini si et seulement si A est positive et telle que le système linéaire $Ax = b$ admet (au moins) une solution, et alors l'ensemble de solutions de $Ax = b$ est l'ensemble des minimums globaux de f."

   Pour $\alpha \neq 0$, on vient de montrer que A est définie positive. Alors A est inversible et le système linéaire $Ax = b$ admet une unique solution $x = A^{-1}b$ qui est l'unique solution du problème d'optimisation. Pour $\alpha = 0$, A est positive, non définie positive. Il reste à déterminer si le système linéaire $Ax = b$ admet des solutions, c'est-à-dire si $b \in \mathrm{Im}(A)$. Pour $\alpha = 0$ on a
$$A = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$$
   donc $\mathrm{Im}(A) = \mathrm{Vect}(\begin{pmatrix} 0 \\ 1 \end{pmatrix})$ et on a $b = \begin{pmatrix} 0 \\ -1 \end{pmatrix} \in \mathrm{Im}(A)$, donc le problème d'optimisation admet une infinité de solutions.

Exercice 2. *(sur environ 10 points)*

1. Rappeler la définition d'une fonction convexe sur $\mathbb{R}^n$.
2. Soit $f : \mathbb{R}^n \to \mathbb{R}$. Montrer que f est convexe sur $\mathbb{R}^n$ si et seulement si

$$\forall p \geqslant 2, \forall x_1, \ldots, x_p \in \mathbb{R}^n, \forall \theta \in [0,1]^p \text{ tel que } \sum_{k=1}^{p} \theta_k = 1, \quad f\left(\sum_{k=1}^{p} \theta_k x_k\right) \leqslant \sum_{k=1}^{p} \theta_k f(x_k).$$

3. Soit $p \geqslant 2$. Montrer que

$$\forall x \in \mathbb{R}^p, \forall \theta \in [0,1]^p \text{ tel que } \sum_{k=1}^{p} \theta_k = 1, \quad \langle \theta, x\rangle^2 \leqslant \sum_{k=1}^{p} \theta_k x_k^2.$$

On fixe maintenant un entier $p \geqslant 2$. On note

$$\mathcal{P} = \left\{ q \in]0,1[^p \text{ tel que } \sum_{k=1}^{p} q_k = 1 \right\}.$$

et on fixe un point $q \in \mathcal{P}$. On pose $g : \mathbb{R}^p \to \mathbb{R}$ la fonction définie par

$$g(x) = \ln\left(\sum_{k=1}^{p} e^{x_k}\right) - \langle q, x\rangle$$

et $S : \mathbb{R}^p \to \mathbb{R}^p$ la fonction définie par

$$(S(x))_k = \frac{e^{x_k}}{\displaystyle\sum_{j=1}^{p} e^{x_j}}, \quad k = 1, \ldots, p.$$

4. Montrer que pour tout $x \in \mathbb{R}^p$, $S(x) \in \mathcal{P}$.
5. Justifier que g est différentiable et donner l'expression de son gradient en fonction de q et S.

6. Justifier que g est deux fois différentiable et montrer que pour tout $x \in \mathbb{R}^p$,

$$\nabla^2 g(x) = \operatorname{diag}(S(x)) - S(x)S(x)^T.$$

7. Montrer que g est convexe à l'aide de l'inégalité de la question 3.
8. En déduire que l'ensemble des minima globaux de g est la droite affine

$$\{x \in \mathbb{R}^p,\ \exists \lambda \in \mathbb{R},\ \forall k \in \{1, \dots, p\},\ x_k = \lambda + \ln q_k\}.$$

et préciser la valeur minimale atteinte.
9. La fonction g est-elle strictement convexe ? Justifier.

Solution de l'exercice 2.

1. $f : \mathbb{R}^n \to \mathbb{R}$ est une fonction *convexe* si

$$\forall x, y \in \mathbb{R}^n,\ \forall \theta \in [0,1], \quad f(\theta x + (1-\theta)y) \leqslant \theta f(x) + (1-\theta) f(y).$$

2. L'implication réciproque est triviale en prenant $p = 2$ et en posant $x = x_1$, $y = x_2$, $\theta = \theta_1$ et $\theta_2 = 1 - \theta_1 = 1 - \theta$. On démontre l'implication directe par récurrence sur p.
Pour l'initialisation, on vient de voir que pour $p = 2$ on retrouve la définition de la convexité.
Pour l'hérédité : On suppose que pour un certain $p \geqslant 2$ on a

$$\forall x_1, \dots, x_p \in \mathbb{R}^n, \forall \theta \in [0,1]^p \text{ tel que } \sum_{k=1}^{p} \theta_k = 1, \quad f\left(\sum_{k=1}^{p} \theta_k x_k\right) \leqslant \sum_{k=1}^{p} \theta_k f(x_k).$$

Soit maintenant $p+1$ vecteurs $x_1, \dots, x_p, x_{p+1} \in \mathbb{R}^n$ et $\theta \in [0,1]^{p+1}$ tel que $\sum_{k=1}^{p+1} \theta_k = 1$. On va alors découper la combinaison linéaire des $p+1$ vecteurs en deux termes :

$$\sum_{k=1}^{p+1} \theta_k x_k = \sum_{k=1}^{p} \theta_k x_k + \theta_{p+1} x_{p+1} = (1 - \theta_{p+1}) \sum_{k=1}^{p} \frac{\theta_k}{1 - \theta_{p+1}} x_k + \theta_{p+1} x_{p+1}.$$

On remarque au préalable que si $\theta_{p+1} = 1$ on a tous les autres $\theta_k = 0$ et l'inégalité à démontrer est triviale. Si $\theta_{p+1} \neq 1$, par convexité de f,

$$\begin{aligned} f\left(\sum_{k=1}^{p+1} \theta_k x_k\right) &= f\left((1 - \theta_{p+1}) \sum_{k=1}^{p} \frac{\theta_k}{1 - \theta_{p+1}} x_k + \theta_{p+1} x_{p+1}\right) \\ &\leqslant (1 - \theta_{p+1}) f\left(\sum_{k=1}^{p} \frac{\theta_k}{1 - \theta_{p+1}} x_k\right) + \theta_{p+1} f(x_{p+1}) \end{aligned}$$

On remarque que pour tout k $\frac{\theta_k}{1-\theta_{p+1}} \geqslant 0$ et

$$\sum_{k=1}^{p} \frac{\theta_k}{1 - \theta_{p+1}} = \frac{1 - \theta_{p+1}}{1 - \theta_{p+1}} = 1.$$

Donc on peut appliquer l'hypothèse de récurrence avec ces poids. Ainsi,

$$\begin{aligned} f\left(\sum_{k=1}^{p+1} \theta_k x_k\right) &\leqslant (1-\theta_{p+1}) \sum_{k=1}^{p} \frac{\theta_k}{1-\theta_{p+1}} f\left(x_k\right) + \theta_{p+1} f(x_{p+1}) \\ &\leqslant \sum_{k=1}^{p+1} \theta_k f\left(x_k\right). \end{aligned}$$

Ceci est valable pour toute famille de $p+1$ vecteurs et poids θ. On a donc démontré l'hérédité.

Ainsi, par récurrence, la propriété est vraie pour tout $p \geqslant 2$.

3. La fonction $\varphi : t \mapsto t^2$ est convexe sur $\mathbb{R}$. On utilise la caractérisation précédente pour $x \in \mathbb{R}^p$:

$$(\langle \theta, x\rangle)^2 = \left(\sum_{k=1}^{p} \theta_k x_k\right)^2 \leqslant \sum_{k=1}^{p} \theta_k x_k^2.$$

4. Montrons que pour tout $x \in \mathbb{R}^p$, $S(x) \in \mathcal{P}$. Soit $x\mathbb{R}^p$, alors pour tout $k \in \{1, \ldots, p\}$,

$$(S(x))_k = \frac{e^{x_k}}{\displaystyle\sum_{j=1}^{p} e^{x_j}} \geqslant 0$$

Par ailleurs, on a bien

$$\sum_{k=1}^{p} (S(x))_k = \frac{\sum_{k=1}^{p} e^{x_k}}{\displaystyle\sum_{j=1}^{p} e^{x_j}} = 1.$$

5. Pour tout $x \in \mathbb{R}^p$, $\displaystyle\sum_{k=1}^{p} e^{x_k} > 0$ donc $\ln\left(\displaystyle\sum_{k=1}^{p} e^{x_k}\right)$ est bien défini. Cette fonction $x \mapsto \ln\left(\displaystyle\sum_{k=1}^{p} e^{x_k}\right)$ est différentiable en tant que composée de fonctions différentiables. La fonction $x \mapsto \langle q, x\rangle$ est différentiable car linéaire. Finalement g est bien différentiable en tant que somme de fonctions différentiables.

Pour $x \in \mathbb{R}^p$ et un indice j quelconque,

$$\frac{\partial g}{\partial x_j}(x) = \frac{e^{x_j}}{\displaystyle\sum_{k=1}^{p} e^{x_k}} - q_j = (S(x))_j - q_j.$$

Finalement on a $\nabla g(x) = S(x) - q$ pour tout $x \in \mathbb{R}^p$.

6. L'expression $\nabla g(x) = S(x) - q$ montre que g est deux fois différentiable. Calculons la

matrice hessienne de g. On a pour un indice i quelconque,

$$\frac{\partial^2 g}{\partial x_j \partial x_i}(x) = \frac{\partial}{\partial x_i}\left[\frac{e^{x_j}}{\sum_{k=1}^{p} e^{x_k}} - q_j\right] = \begin{cases} -\dfrac{e^{x_j}e^{x_i}}{\left(\sum_{k=1}^{p} e^{x_k}\right)^2} & \text{si } i \neq j, \\ \dfrac{e^{x_i}}{\sum_{k=1}^{p} e^{x_k}} - \dfrac{(e^{x_i})^2}{\left(\sum_{k=1}^{p} e^{x_k}\right)^2} & \text{si } i = j. \end{cases}$$

Finalement,

$$\frac{\partial^2 g}{\partial x_j \partial x_i}(x) = \begin{cases} -(S(x))_i(S(x))_j & \text{si } i \neq j, \\ (S(x))_i - (S(x))_i^2 & \text{si } i = j. \end{cases}$$

On obtient bien que

$$\nabla^2 g(x) = \operatorname{diag}(S(x)) - S(x)S(x)^T.$$

7. Montrons que g est convexe à l'aide de l'inégalité de la question 3. Comme g est deux fois différentiable, g est convexe si et seulement si pour tout $x \in \mathbb{R}^p$, $\nabla^2 g(x)$ est positive. Soit $h \in \mathbb{R}^n$, on a

$$\begin{aligned}\langle \nabla^2 g(x)h, h\rangle &= \langle (\operatorname{diag}(S(x)) - S(x)S(x)^T)h, h\rangle \\ &= \langle \operatorname{diag}(S(x))h, h\rangle - \langle S(x)S(x)^T h, h\rangle\end{aligned}$$

Or on a d'une part

$$\langle \operatorname{diag}(S(x))h, h\rangle = \sum_{k=1}^{p} (S(x))_k h_k^2.$$

D'autre part

$$\langle S(x)S(x)^T h, h\rangle = \langle S(x)^T h, S(x)^T h\rangle = (S(x)^T h)^2 = \langle S(x), h\rangle^2.$$

En appliquant l'inégalité de la question 3 avec $x = h$ et $S(x) = \theta$ (qui vérifie bien les hypothèses car $S(x) \in \mathcal{P}$), on a

$$\langle S(x), h\rangle^2 \leqslant \sum_{k=1}^{p} (S(x))_k h_k^2.$$

Ainsi on a bien $\langle \nabla^2 g(x)h, h\rangle \geqslant 0$.

8. Comme g est convexe, g atteint un minimum global en $x \in \mathbb{R}^p$ si et seulement si $\nabla g(x) = 0$. Il reste à montrer que

$$\{x \in \mathbb{R}^p,\ \nabla g(x) = 0\} = \{x \in \mathbb{R}^p,\ \exists \lambda \in \mathbb{R},\ \forall k \in \{1, \ldots, p\},\ x_k = \lambda + \ln q_k\}.$$

On sait que $\nabla g(x) = S(x) - q$. On va raisonner par double inclusion. Soit x tel que $\nabla g(x) = 0$. Alors, $S(x) = q$, donc pour tout $k \in \{1, \ldots, p\}$,

$$\frac{e^{x_k}}{\sum_{j=1}^{p} e^{x_j}} = q_k \Leftrightarrow x_k = \ln\left(\sum_{j=1}^{p} e^{x_j}\right) + \ln(q_k),$$

donc $\lambda = \ln\left(\sum_{j=1}^{p} e^{x_j}\right)$ convient. Réciproquement, soit $x \in \mathbb{R}^p$ de la forme $x_k = \lambda + \ln q_k$ avec $\lambda \in \mathbb{R}$. Alors, pour tout $k \in \{1, \ldots, p\}$,

$$(S(x))_k = \frac{e^{x_k}}{\sum_{j=1}^{p} e^{x_j}} = \frac{e^{\lambda} q_k}{\sum_{j=1}^{p} e^{\lambda} q_j} = q_k$$

et donc $\nabla g(x) = 0$.

Pour finir déterminons la valeur minimale de g. Pour $x_k = \ln(q_k)$ on a

$$g(x) = \ln\left(\sum_{k=1}^{p} e^{x_k}\right) - \langle q, x\rangle = \ln\left(\sum_{k=1}^{p} q_k\right) - \sum_{k=1}^{p} q_k \ln(q_k) = -\sum_{k=1}^{p} q_k \ln(q_k).$$

La valeur minimale de g est donc $-\sum_{k=1}^{p} q_k \ln(q_k)$.

9. On vient de montrer que l'ensemble des minimums globaux de g est une droite affine. Or une fonction strictement convexe admet au plus un minimum global donc g n'est pas strictement convexe.

Exercice 3. *(sur environ 11 points)*

Soit $n \geqslant 2$ un entier. Soit $P \in \mathcal{M}_n(\mathbb{R})$ une matrice symétrique définie positive. On note $\|x\| = \sqrt{\langle x, x\rangle}$ la norme euclidienne usuelle de $\mathbb{R}^n$, $\|x\|_P = \sqrt{\langle Px, x\rangle}$ la norme associée à P et $\|x\|_{P^{-1}} = \sqrt{\langle P^{-1}x, x\rangle}$ la norme associée à P^{-1} la matrice inverse de P. On note par la suite $\gamma = \lambda_{\min}(P) > 0$ la valeur propre minimale de P et $\delta = \lambda_{\max}(P)$ la valeur propre maximale de P.

1. Démontrer que pour tout $x \in \mathbb{R}^2$ on a les inégalités

$$\gamma\|x\|^2 \leqslant \|x\|_P^2 \leqslant \delta\|x\|^2 \quad \text{et} \qquad \frac{1}{\delta}\|x\|^2 \leqslant \|x\|_{P^{-1}}^2 \leqslant \frac{1}{\gamma}\|x\|^2.$$

Dans la suite de l'exercice, $f : \mathbb{R}^n \to \mathbb{R}$ désigne une fonction deux fois différentiable sur $\mathbb{R}^n$ et fortement convexe telle que

$$\forall x, h \in \mathbb{R}^n, \quad m\|h\|^2 \leqslant \langle \nabla^2 f(x)h, h\rangle \leqslant M\|h\|^2$$

avec $0 < m \leqslant M$. On note $x^\star$ le point où f atteint son minimum global et $p^\star = f(x^\star)$ la valeur minimale de f. On rappelle que pour tout $x \in \mathbb{R}^n$, on a

$$p^\star \geqslant f(x) - \frac{1}{2m}\|\nabla f(x)\|^2.$$

Le but de l'exercice est d'étudier la convergence d'un algorithme de descente qui prend pour direction de descente à l'étape k le vecteur

$$d^{(k)} = -P\nabla f(x^{(k)}).$$

Plus précisément, on considère l'Algorithme 1 ci-dessous.

Algorithme 1 : Algorithme de descente à étudier

Données : Un point initial $x^{(0)} \in \mathbb{R}^n$, un seuil de tolérance $\varepsilon > 0$
Résultat : Un point $x \in \mathbb{R}^n$ proche de $x^\star$
Initialiser x :
  $x \leftarrow x^{(0)}$;
  $k \leftarrow 0$;
tant que $\|\nabla f(x)\| > \varepsilon$ **faire**

1. Calculer $d^{(k)} = -P\nabla f(x^{(k)})$.
2. Déterminer un pas de descente $t^{(k)} > 0$ pour la direction de descente $d^{(k)}$ (par la méthode exacte ou la méthode de rebroussement de paramètres α et β).
3. Mettre à jour x :
   $x \leftarrow x^{(k+1)} = x^{(k)} + t^{(k)} d^{(k)}$;
   $k \leftarrow k + 1$;

fin

Implémentation de l'Algorithme 1 avec la méthode de rebroussement On suppose que les fonctions

`function fx = f(x)` et `function gfx = gradf(x)`

sont définies dans scilab.

2. ***(sur 4 points)*** Ecrire une fonction scilab

```
function [Xk,FXk] = algo_descente(x0, eps, alph, bet, f, gradf, P)
```

qui prend en entrées un point initial $x^{(0)} \in \mathbb{R}^n$, un seuil de tolérance $\varepsilon > 0$, et deux paramètres α et β, les fonctions pour évaluer f et ∇f et la matrice P. La fonction renvoie la matrice `Xk` (de taille n par nombre d'itérations) contenant la liste des vecteurs $x^{(k)}$ donnée par l'Algorithme 1 **avec le calcul du pas par méthode de rebroussement** ainsi que la liste des valeurs `Fxk` de la fonctionnelle f aux points $x^{(k)}$. On veillera à ajouter une condition pour limiter le nombre d'itérations à 10 000.

Etude de la convergence de l'Algorithme 1 avec la méthode exacte Le but de la fin de l'exercice est de démontrer la convergence de l'Algorithme 1 **avec la méthode exacte pour le calcul du pas de descente**.

On suppose qu'à l'étape k de l'Algorithme 1 le point $x^{(k)}$ est différent de $x^\star$.

3. Montrer que
$$\|d^{(k)}\|^2 \leqslant \delta \|\nabla f(x^{(k)})\|_P^2.$$

4. Montrer que pour tout $t > 0$,
$$f(x^{(k+1)}) \leqslant f(x^{(k)} + td^{(k)}) \leqslant f(x^{(k)}) + \left(-t + \frac{M\delta t^2}{2}\right) \|\nabla f(x^{(k)})\|_P^2.$$

5. En déduire que
$$f(x^{(k+1)}) \leqslant f(x^{(k)}) - \frac{\gamma}{2M\delta}\|\nabla f(x^{(k)})\|^2.$$
6. Conclure sur la convergence de l'Algorithme 1.

Solution de l'exercice 3.

1. P est symétrique donc diagonalisable dans une base orthonormée. Soit $(e_1, \dots, e_n)$ une telle base orthornormée et $(\lambda_1, \dots, \lambda_n)$ les valeurs propres associées. Alors, en décomposant x dans cette base
$$x = \sum_{k=1}^{n} \langle x, e_k \rangle e_k$$
$$Px = \sum_{k=1}^{n} \langle x, e_k \rangle P e_k = \sum_{k=1}^{n} \lambda_k \langle x, e_k \rangle e_k$$
d'où
$$\|x\|_P^2 = \langle Px, x \rangle = \sum_{k=1}^{n} \lambda_k \langle x, e_k \rangle^2$$
Comme pour tout k on a $\lambda_k \leqslant \delta$,
$$\|x\|_P^2 = \sum_{k=1}^{n} \lambda_k \langle x, e_k \rangle^2 \leqslant \delta \sum_{k=1}^{n} \langle x, e_k \rangle^2 = \delta \|x\|^2.$$
De même, pour tout k on a $\lambda_k \geqslant \gamma$, d'où $\|x\|_P \geqslant \gamma \|x\|^2$. L'encadrement pour P^{-1} est le même puisque les valeurs propres de P^{-1} sont les $\frac{1}{\lambda_k}$.
2.

```
function [Xk, FXk] = algo_descente(x0, eps, alph, bet, f, gradf, P)
x=x0;
k=0;
gfx=gradf(x);
Xk = x;
FXk = f(x);
while(gfx'*gfx > eps^2 & k < 10^5)
    d = - P*gfx;
    t=1;
    while(f(x+td) > f(x) + alph*t*gfx'*d)
        t = bet*t;
    end
    x = x + t*d;
    k=k+1;
    gfx=gradf(x);
    Xk = [Xk, x];
    FXk = [FXk, f(x)];
end
endfunction
```

3. On a
$$\|d^{(k)}\|^2 = \| - P\nabla f(x^{(k)})\|^2 \leqslant \delta\|P\nabla f(x^{(k)})\|^2_{P^{-1}} = \delta\langle P^{-1}P\nabla f(x^{(k)}), P\nabla f(x^{(k)})\rangle$$
$$= \delta\langle \nabla f(x^{(k)}), P\nabla f(x^{(k)})\rangle$$
$$= \delta\|\nabla f(x^{(k)})\|^2_P.$$

4. Par définition de la méthode exacte pour le calcul du pas de descente,
$$f(x^{(k+1)}) = \min_{t>0} f(x^{(k)} + td^{(k)}),$$
donc pour tout $t > 0$, on a bien
$$f(x^{(k+1)}) \leqslant f(x^{(k)} + td^{(k)}).$$
D'après la formule de Taylor-Maclaurin, il existe $z \in [x^{(k)}, x^{(k)} + td^{(k)}]$ tel que
$$f(x^{(k)} + td^{(k)}) = f(x^{(k)}) + t\langle \nabla f(x^{(k)}), d^{(k)}\rangle + \frac{t^2}{2}\langle \nabla^2 f(z)d^{(k)}, d^{(k)}\rangle.$$
Or par hypothèse et d'après la question précédente,
$$\langle \nabla^2 f(z)d^{(k)}, d^{(k)}\rangle \leqslant M\|d^{(k)}\|^2 \leqslant M\delta\|\nabla f(x^{(k)})\|^2_P.$$
De plus,
$$\langle \nabla f(x^{(k)}), d^{(k)}\rangle = -\langle \nabla f(x^{(k)}), P\nabla f(x^{(k)})\rangle = -\|\nabla f(x^{(k)})\|^2_P.$$
Ainsi on a bien
$$f(x^{(k)} + td^{(k)}) \leqslant f(x^{(k)}) + \left(-t + \frac{M\delta t^2}{2}\right)\|\nabla f(x^{(k)})\|^2_P.$$

5. On choisit le $t > 0$ qui minimise $\left(-t + \frac{M\delta t^2}{2}\right)$, c'est-à-dire
$$t = \frac{1}{M\delta}.$$
Pour cette valeur de t on a $-t + \frac{M\delta t^2}{2} = -\frac{1}{2M\delta}$. On a donc
$$f(x^{(k+1)}) \leqslant f(x^{(k)}) - \frac{1}{2M\delta}\|\nabla f(x^{(k)})\|^2_P.$$
On conclut en utilisant l'inégalité $\|\nabla f(x^{(k)})\|^2_P \geqslant \gamma\|\nabla f(x^{(k)})\|^2$ pour obtenir la majoration
$$f(x^{(k+1)}) \leqslant f(x^{(k)}) - \frac{\gamma}{2M\delta}\|\nabla f(x^{(k)})\|^2.$$

6. Etant donné la rappel en début d'énoncé,
$$\|\nabla f(x^{(k)})\|^2 \geqslant 2m(f(x^{(k)}) - p^\star).$$
Ainsi, en soustrayant par $p^\star$,
$$f(x^{(k+1)}) - p^\star \leqslant f(x^{(k)}) - p^\star - \frac{\gamma}{2M\delta}\|\nabla f(x^{(k)})\|^2 \leqslant \left(1 - \frac{m\gamma}{M\delta}\right)\left(f(x^{(k)}) - p^\star\right).$$
Par récurrence immédiate on a
$$f(x^{(k)}) - p^\star \leqslant \left(1 - \frac{m\gamma}{M\delta}\right)^k \left(f(x^{(0)}) - p^\star\right)$$
et $\left(1 - \frac{m\gamma}{M\delta}\right) \in [0, 1[$ donc l'Algorithme 1 converge linéairement.