Problème n°1
Méthode des moindres carrés

⋆ Exercice 1 – Fonctions quadratiques généralisées

Mots-clefs : Différentiabilité • Fonctions convexes • Fonctions coercives

Soit $n \in \mathbb{N}^*$. Soit $A \in \mathcal{M}_{n,n}(\mathbb{R})$ une matrice réelle symétrique, $b \in \mathbb{R}^n$ un vecteur et $c \in \mathbb{R}$ un scalaire. Alors on appelle*fonction quadratique généralisée* la fonction définie par

$$f : \left\{ \begin{array}{ccc} \mathbb{R}^n & \to & \mathbb{R} \\ x & \mapsto & \dfrac{1}{2} \langle A\,x, x \rangle - \langle b, x \rangle + c \end{array} \right.$$

(a) Lorsque $n = 1$, quelles sont les fonctions quadratiques généralisées ?

(b) Montrer que toute fonction quadratique généralisée est de classe $\mathcal{C}^2$. Calculer son gradient et sa matrice hessienne.

(c) En déduire que f est convexe si et seulement si A est semi-définie positive.

(d) Montrer que f est strictement convexe si et seulement si A est définie positive.

(e) Montrer que f est coercive si et seulement si f est strictement convexe.

> Attention, les résultats démontrés aux questions **(d)** et **(e)** sont spécifiques aux fonctions quadratiques généralisées. En général, il existe des fonctions strictement convexes avec une matrice hessienne non définie positive sur son ensemble de définition, ainsi que des fonctions strictement convexes non coercives.

⋆ Exercice 2 – Optimisation des fonctions quadratiques généralisées

Mots-clefs : Fonctions régulières • Fonctions convexes • Règle de Fermat • Méthode du gradient

*On reprend les notations de l'**Exercice 1**.*

(a) On suppose que A est semi-définie positive. Montrer que chercher les minimiseurs de f est équivalent à résoudre le système linéaire suivant :

$$A\,x = b$$

(b) Montrer que les fonctions quadratiques généralisées sont régulières.

(c) Écrire les itérations de la méthode du gradient (explicite) pour la minimisation de f. Sous quelles conditions sur le pas de temps la méthode converge-t-elle ?

(d) On suppose que A est définie positive. Montrer que f est fortement convexe.

(e) En déduire une condition sur le pas de temps de la méthode du gradient pour obtenir un taux de convergence linéaire.

> À première vue, pour trouver les minimiseurs de f, il est plus simple de passer par la résolution du système linéaire $A\,x = b$ que d'appliquer la méthode du gradient qui, même dans le cas fortement convexe, ne donne en un nombre d'itérations fini qu'une approximation du minimiseur $A^{-1}b$. Cependant, si A est mal conditionnée, l'inversion du système linéaire peut engendrer des erreurs importantes. Les itérations de la méthode du gradient ne reposant sur aucune inversion de matrice, elle ne souffre pas de cet écueil. Ainsi, les itérations de cette méthode fournissent une méthode intéressante pour évaluer l'inverse de A en tout point x lorsque A est mal conditionnée.

⋆ Exercice 3 – Projection sur un convexe

Mots-clefs : Ensembles convexes • Projection sur un convexe

Soit $m, n \in \mathbb{N}^*$. Soit $M \in \mathcal{M}_{m,n}(\mathbb{R})$.

(a) Montrer que $\mathrm{Im}(M)$ est un ensemble convexe, non vide et fermé.

(b) En déduire que la projection sur $\mathrm{Im}(M)$ existe et est unique.

(c) Expliciter l'écriture du problème de projection sur $\mathrm{Im}(M)$.

⋆ Exercice 4 – Méthode des moindres carrés

Mots-clefs : Fonctions convexes • Règle de Fermat

On reprend les notations des ***Exercices 1*** *et* ***3****.* Soit un vecteur $v \in \mathbb{R}^m$. On s'intéresse au problème d'optimisation suivant :

$$\min_{x \in \mathbb{R}^n} \left\{ J(x) = \frac{1}{2} \, \|M\,x - v\|_2^2 \right\} \qquad (\mathcal{P}_{\mathrm{LS}})$$

(a) Montrer que la fonction objectif J du problème $(\mathcal{P}_{\mathrm{LS}})$ est une fonction quadratique généralisée. En déduire des conditions sur M pour assurer la convexité ou la stricte convexité de J.

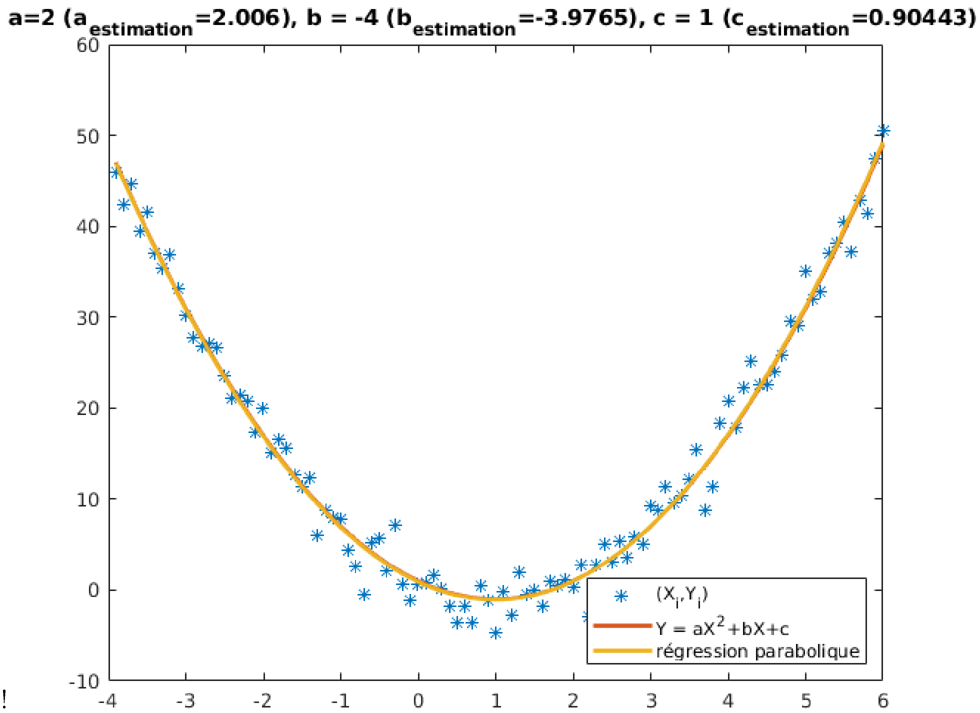
(b) On pose
$$M^\dagger = (M^\top M)^{-1}\, M$$
Montrer que les solutions du problème $(\mathcal{P}_{\mathrm{LS}})$ sont les solutions du système linéaire
$$M^\top M\, x = M^\top b$$
Sous quelles conditions sur M le problème $(\mathcal{P}_{\mathrm{LS}})$ admet-il une unique solution ?

(c) Montrer que $(\mathcal{P}_{\mathrm{LS}})$ peut s'écrire comme le problème de la projection sur $\mathrm{Im}\, M$. En déduire que $(\mathcal{P}_{\mathrm{LS}})$ admet au moins une solution.

(d) Vérifier que si $M\,x^* = v$, alors x^* est solution de $(\mathcal{P}_{\mathrm{LS}})$.

> La matrice $M^\dagger$ est appelée *pseudo-inverse* de M.

> L'idée derrière le problème $(\mathcal{P}_{\mathrm{LS}})$ est de trouver un vecteur $x^* \in \mathbb{R}^n$ satisfaisant *au mieux*, et plus précisément, au sens des *moindres carrés*, la relation
> $$M\,x = v$$
> où M est connu mais v est entaché d'erreur, de sorte que la relation linéaire ci-dessus n'est pas vérifiée de manière exacte.
> Il faut noter qu'il existe d'autres manières de définir le caractère optimal de x^* vis-à-vis de cette relation, ce qui se traduit par un choix de fonction objectif différent, et donc par une solution (*a priori*) différente. Le choix des moindres carrés, outre l'avantage qu'il présente de conduire à un problème d'optimisation lisse, est motivé par des raisons relevant du domaine des probabilités. On peut montrer en effet que, si $v = M\,x_0 + n$ avec n un vecteur de $\mathbb{R}^m$ tel que chaque coefficient est généré par la même loi normale (et donc, de manière indépendante pour chaque coefficient), alors la solution x^* obtenue à l'aide de la méthode des moindres carrés permet d'obtenir l'estimation de x_0 la plus probable sachant v et M.
> On signale enfin qu'en anglais, *moindres carrés* se lit *least squares*.



h!

Figure 1 – En bleu, le nuage de points, en rouge la parabole théorique, en jaune, l'estimation obtenue au sens des moindres carrés.

⋆ Exercice 5 – Application : Régression parabolique

Mots-clefs : Fonctions convexes · Règle de Fermat

*On reprend les notations de l'**Exercice 4**.* Soit $p \in \mathbb{N}^*$. On suppose qu'on possède un nuage de points (X_i, Y_i) pour $i = 1, \ldots, p$ (voir Figure 1). On aimerait trouver la parabole qui approche au mieux ce nuage au sens des moindres carrés. Plus précisément, on suppose que

$$\forall\, i \in [\![1\,;\, p]\!], \qquad Y_i \approx a\,X_i^2 + b\,X_i + c$$

avec a, b et c trois réels, ce qui revient à résoudre le problème d'optimisation

$$\min_{(a,b,c)\in\mathbb{R}^3} \frac{1}{2} \sum_{i=1}^{p} (a\,X_i^2 + b\,X_i + c - Y_i)^2$$

(a) Montrer que le problème considéré peut s'écrire comme un problème d'optimisation quadratique de la forme $(\mathcal{P}_{\mathrm{LS}})$.

(b) Justifier que ce problème admet au moins une solution optimale, parmi les solutions du système linéaire

$$\left\{\begin{array}{lllllll}
a\displaystyle\sum_{i=1}^{p} X_i^4 & + & b\displaystyle\sum_{i=1}^{p} X_i^3 & + & c\displaystyle\sum_{i=1}^{p} X_i^2 & = & \displaystyle\sum_{i=1}^{p} X_i^2\,Y_i \\
a\displaystyle\sum_{i=1}^{p} X_i^3 & + & b\displaystyle\sum_{i=1}^{p} X_i^2 & + & c\displaystyle\sum_{i=1}^{p} X_i & = & \displaystyle\sum_{i=1}^{p} X_i Y_i \\
a\displaystyle\sum_{i=1}^{p} X_i^2 & + & b\displaystyle\sum_{i=1}^{p} X_i & + & c\,p & = & \sum_{i=1}^{p} Y_i
\end{array}\right.$$

(c) Montrer que ce système admet une unique solution si et seulement s'il existe trois indices $1 \leq i_1, i_2, i_3 \leq p$ tels que la matrice de Vandermonde

$$\begin{pmatrix} X_{i_1}^2 & X_{i_1} & 1 \\ X_{i_2}^2 & X_{i_2} & 1 \\ X_{i_3}^2 & X_{i_3} & 1 \end{pmatrix}$$

soit inversible.

(d) Justifier que la matrice de VANDERMONDE apparaissant à la question précédente est inversible si les nombres X_{i_1}, X_{i_2} et X_{i_3} sont deux à deux distincts.

> Une interprétation de ce résultat est la suivante : pour obtenir une estimation uniquement déterminée d'une parabole, il faut et il suffit d'avoir au moins trois points distincts dans le nuage de points.

⋆⋆ Exercice 6 – Régularisation de TYKHONOV

Mots-clefs : Fonctions convexes • Règle de FERMAT

*On reprend les notations des **Exercices 1**, **3** et **4**.* Soit $\lambda > 0$. On s'intéresse au problème d'optimisation suivant :

$$\min_{x \in \mathbb{R}^n} \frac{1}{2} \|M\,x - v\|_2^2 + \lambda \, \|x\|_2^2 \tag{$\mathcal{P}_1$}$$

(a) Justifier que la fonction objectif de ce problème est fortement convexe. En déduire le nombre de solutions pour le problème $(\mathcal{P}_1)$.

(b) On note x_λ^* la solution du problème $(\mathcal{P}_1)$. Montrer qu'on a

$$(M^\top M + 2\,\lambda\,\mathrm{I}_n)\,x_\lambda^* = M^\top v$$

(c) On note x_{LS}^* une solution du problème $(\mathcal{P}_{\mathrm{LS}})$. Montrer que

$$\|x_\lambda^*\|_2^2 \leq \|x_{\mathrm{LS}}^*\|_2^2$$

(d) On suppose que M est injective. Montrer que

$$\lim_{\lambda \to 0} x_\lambda^* = x_{\mathrm{LS}}^*$$

> En général, si J_λ admet un minimiseur x_λ^* pour tout $\lambda \geq 0$, rien n'assure que la suite des x_λ^* converge vers x_0^* même si les J_λ convergent vers J_0. Une propriété suffisante pour obtenir une telle convergence est la Γ-convergence des J_λ vers J_0.

⋆⋆ Exercice 7 – Optimisation quadratique sous contraintes

Mots-clefs : Dualité de LAGRANGE • Qualification de contraintes • Conditions KKT

*On reprend les notations des **Exercices 1**, **3**, **4** et **6**.* On suppose que M est injective. Soit $\delta > 0$. On s'intéresse maintenant au problème d'optimisation sous contraintes suivant :

$$\min_{x \in \mathbb{R}^n} \frac{1}{2} \|M\,x - v\|_2^2 \text{ sous la contrainte } \|x\|_2 \leq \delta \tag{$\mathcal{P}_2$}$$

(a) Après avoir écrit les contraintes du problème $(\mathcal{P}_2)$ comme des contraintes différentiables, écrire le lagrangien associé.

(b) Justifier que le problème $(\mathcal{P}_2)$ est convexe et que les contraintes sont qualifiées.

(c) Écrire les conditions d'optimalité de KARUSH–KUHN–TUCKER pour le problème $(\mathcal{P}_2)$.

(d) Montrer que si $\|x^*_{\mathrm{LS}}\|_2 < \delta$, alors x^*_{LS} est solution du problème $(\mathcal{P}_2)$.

(e) Montrer que si $\|x^*_{\mathrm{LS}}\|_2 \geq \delta$, alors il existe $\lambda_0 > 0$ tel que $\delta \geq \|x^*_{\lambda_0}\|$. En déduire que, dans ce cas, $x^*_{\lambda_0}$ est solution du problème $(\mathcal{P}_2)$.

> Autrement dit, on vient de démontrer que si δ et λ sont choisis suffisamment petits, alors les deux problèmes $(\mathcal{P}_1)$ et $(\mathcal{P}_2)$ ont même solution.

⋆⋆⋆ Exercice 8 – Total Least Squares **TODO**

Mots-clefs : Règle de FERMAT • Dualité de LAGRANGE • Conditions KKT

On reprend les notations des ***Exercices 1****,* ***3****,* ***4****,* ***6*** *et* ***7****.* Soit $p, q \in \mathbb{N}^*$. Pour tout $S = (s_{i,j})_{\substack{1\leq i\leq p\\1\leq j\leq q}} \in \mathcal{M}_{p,q}(\mathbb{R})$, on note

$$\|S\|_F = \sqrt{\sum_{i=1}^{p}\sum_{j=1}^{q} {s_{i,j}}^2}$$

la norme de FRÖBENIUS de S. On s'intéresse au problème d'optimisation suivant :

$$\min_{\substack{x\in\mathbb{R}^n\\ \tilde{M}\in\mathcal{M}_{n,n}(\mathbb{R})\\ \tilde{v}\in\mathbb{R}^n}} \frac{1}{2}\|(M,v)-(\tilde{M},\tilde{v})\|_F^2 \text{ sous les contraintes } \tilde{M}\,x = \tilde{v} \qquad (\mathcal{P}_{\mathrm{TLS}})$$

(a) Montrer que le problème $(\mathcal{P}_{\mathrm{TLS}})$ peut se réécrire sous la forme du problème d'optimisation convexe non contraint

$$\min_{\substack{x\in\mathbb{R}^n\\ \tilde{M}\in\mathcal{M}_{n,n}(\mathbb{R})}} \frac{1}{2}\|M-\tilde{M}\|_F^2 + \frac{1}{2}\|\tilde{M}\,x - v\|_2^2$$

(b) En déduire que les solutions $(x^*_{\mathrm{TLS}}, \tilde{M}^*_{\mathrm{TLS}})$ du problème $(\mathcal{P}_{\mathrm{TLS}})$ sont les solutions su système d'équations

$$\begin{cases} \tilde{M} + (\tilde{M}\,x - v)\,x^\top = M \\ \tilde{M}^\top\tilde{M}\,x = \tilde{M}^\top v \end{cases}$$

(c) Vérifier que, si on ajoute la contrainte $\tilde{M} = M$, alors on retrouve le problème $(\mathcal{P}_{\mathrm{LS}})$.

(d) Écrire le lagrangien associé au problème $(\mathcal{P}_{\mathrm{TLS}})$, ainsi que les conditions d'optimalité de KARUSH–KUHN–TUCKER .

On va procéder comme dans l'**Exercice 7**, et ajouter la contrainte $\|x\|_2 \leq \delta$ au problème $(\mathcal{P}_{\mathrm{TLS}})$. On s'intéresse donc au problème

$$\min_{\substack{x\in\mathbb{R}^n\\ \tilde{M}\in\mathcal{M}_{n,n}(\mathbb{R})}} \frac{1}{2}\|M-\tilde{M}\|_F^2 + \frac{1}{2}\|\tilde{M}\,x - v\|_2^2 \text{ sous la contrainte } \|x\|_2 \leq \delta \qquad (\mathcal{P}_3)$$

(e) Après avoir écrit les contraintes du problème $(\mathcal{P}_3)$ comme des contraintes différentiables, écrire le lagrangien associé. Justifier que le problème $(\mathcal{P}_3)$ est convexe et que les contraintes sont qualifiées. Écrire les conditions d'optimalité de KARUSH–KUHN–TUCKER pour le problème $(\mathcal{P}_3)$.

(f) On note x^*_{TLS} une solution (si elle existe) du problème $(\mathcal{P}_{\mathrm{TLS}})$. Montrer que si $\|x^*_{\mathrm{TLS}}\|_2 \leq \delta$, alors x^*_{TLS} est solution du problème $(\mathcal{P}_3)$.

(g)

(h)

(i)

La méthode présentée dans cet exercice est utilisée pour donner une estimation d'un vecteur $x \in \mathbb{R}^n$ qui doit théoriquement vérifier le modèle linéaire

$$M\, x = v$$

À nouveau, l'estimation de x se fait aux moindres carrés, puisqu'on cherche à minimiser une distance au carré. Contrairement à la méthode des moindres carrés présentée dans l'**Exercice 4**, on suppose ici que M et v sont tous deux connus de manière incertaine. Si l'on applique cette hypothèse au problème de la régression parabolique de l'**Exercice 5**, cela revient à considérer que non seulement les ordonnées Y_i ne sont pas exactes (les points (X_i, Y_i) ne sont pas situés exactement sur la parabole), mais également que les abscisses X_i ne sont pas connues exactement.

Pour aller plus loin

Référence principale : TODO