FISE2 – 2021/2022

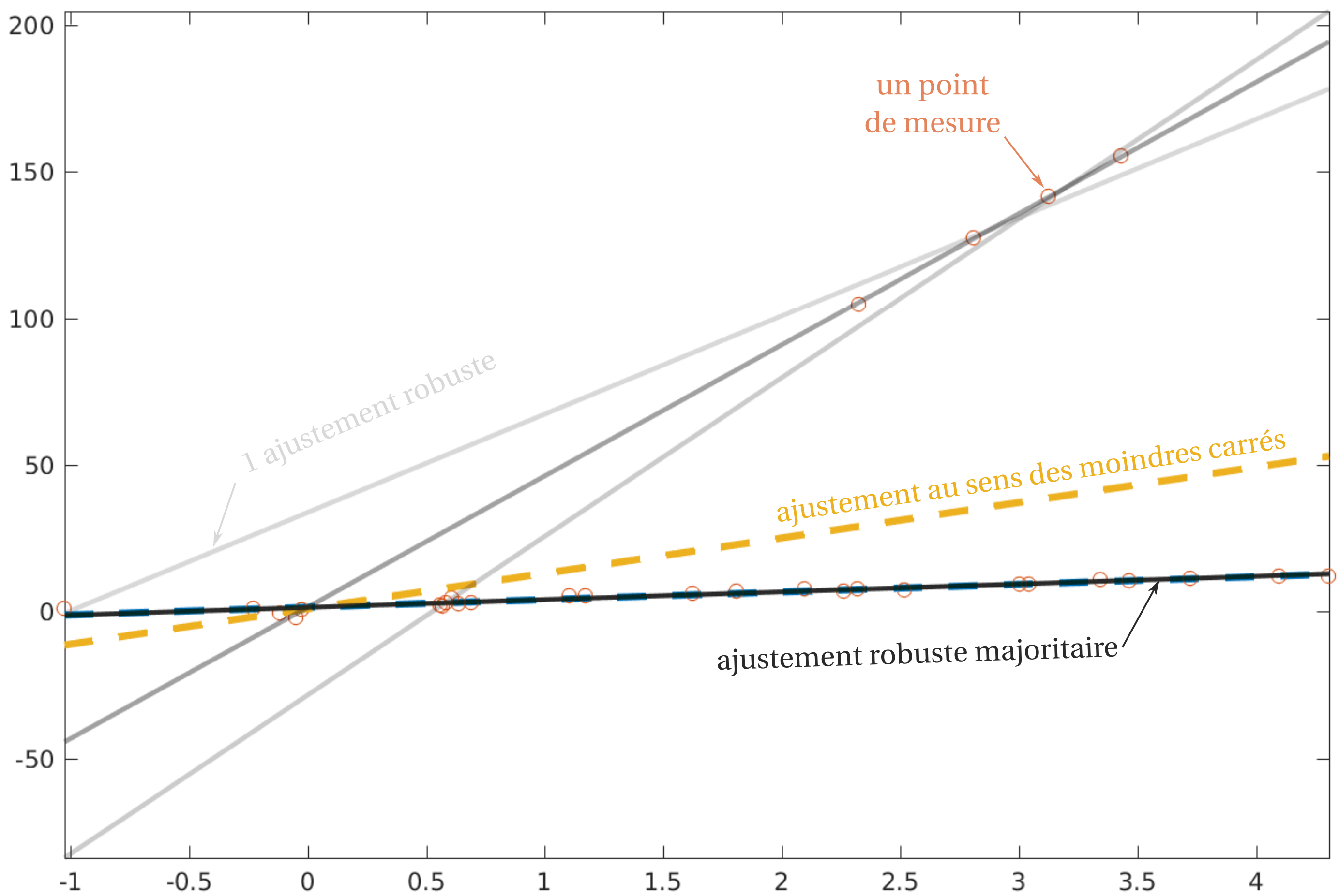
Projet d'optimisation continue
Estimation robuste: les M-estimateurs

Le but de ce projet est d'appliquer les méthodes d'optimisation continue à un cas concret: le problème de l'*estimation robuste*, c'est à dire l'estimation en présence de points aberrants (outliers). Nous nous intéressons à une classe d'estimateurs appelés *M-estimateurs* qui sont définis sous la forme d'un problème d'optimisation continue.

1 Description du problème

On a réalisé n points de mesure (x_1, y_1), (x_2, y_2), $\ldots (x_n, y_n)$. On souhaiterait ajuster une droite $y = a\,x + b$ aux points mesurés. Cependant, la présence de points aberrants rend assez mauvais l'ajustement au sens des moindres carrés. On préfère alors utiliser d'autres estimateurs, robustes aux mesures aberrantes: les M-estimateurs.

La figure ci-dessous montre des points de mesure, la droite que l'on souhaiterait idéalement retrouver (pointillés bleus) ainsi que les droites obtenues par 2 estimateurs différents que nous allons étudier dans le projet: l'estimateur des moindres carrés (pointillés jaunes) et plusieurs solutions d'un estimateur robuste.



2 Estimation au sens des moindres carrés

L'ajustement au sens des moindres carrés consiste à chercher les paramètres a et b de la droite qui minimisent la somme des carrés des écarts entre les mesures y_i et les valeurs prédites par le modèle linéaire $ax_i + b$:

$$\hat{\theta} = \begin{pmatrix} \hat{a} \\ \hat{b} \end{pmatrix} = \arg\min_{a,b} \underbrace{\sum_{i=1}^{n} (ax_i + b - y_i)^2}_{\mathscr{C}_{\text{moindres carrés}}(a,b)} . \tag{1}$$

Q1: Représentez la fonction de coût $\mathscr{C}_{\text{moindres carrés}}(a, b)$ dans l'espace des paramètres (a, b). On choisira une plage de valeurs pertinente pour les paramètres a et b d'après l'analyse de la figure en page 1.

Q2: Déterminez par échantillonnage régulier des paramètres a et b une solution approchée au problème.

La fonction de coût $\mathscr{C}_{\text{moindres carrés}}(a, b)$ peut se réécrire sous la forme ci-dessous:

$$\mathscr{C}_{\text{moindres carrés}}(a, b) = \left\| \begin{pmatrix} x_1 & 1 \\ \vdots & \vdots \\ x_n & 1 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} - \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \right\|_2^2 \tag{2}$$

Q3: Développez cette expression pour identifier la formulation générale d'un problème de minimisation quadratique et déduisez-en l'expression analytique du minimiseur. Appliquez cette formule pour déterminer les valeurs de $\hat{a}$ et $\hat{b}$ avec 6 chiffres significatifs.

Q4: Représentez le couple trouvé à la question précédente dans l'espace des paramètres pour vérifier visuellement qu'il se trouve bien au minimum de la fonction de coût.

Q5: Tracez la courbe d'équation $y = \hat{a}x + \hat{b}$ obtenue. Que pensez-vous de cet estimateur?

3 Estimation robuste

Pour améliorer l'estimation, on peut considérer une généralisation de la fonction de coût:

$$\hat{\theta} = (\hat{a}, \hat{b}) = \arg\min_{a,b} \underbrace{\sum_{i=1}^{n} \rho(r_i)}_{\mathscr{C}_{\text{robuste}}(a,b)} , \quad \text{avec } r_i = ax_i + b - y_i \tag{3}$$

et où ρ est une fonction de $\mathbb{R}$ dans $\mathbb{R}$. Les résidus $r_i = a\,x_i + b - y_i$ représentent l'écart entre le modèle et les données en chacun des points de mesure.
Nous allons étudier la pénalisation de Cauchy définie par:

$$\forall r \in \mathbb{R},\ \rho_\sigma(r) = \frac{1}{2}\log(1 + (r/\sigma)^2)\,, \tag{4}$$

où σ est un paramètre de la pénalisation.

La résolution du problème de minimisation (3) nécessite un algorithme itératif.

Q6: Implémentez la méthode des plus fortes pentes en réalisant l'étape de recherche linéaire (ligne `arg min` de l'algorithme décrit en page 13 du chapitre 3) avec l'algorithme de Fletcher-Lemaréchal. Vous pourrez tester dans un premier temps cet algorithme sur la fonction de coût $\mathscr{C}_{\text{moindres carrés}}$ que vous avez étudiée dans la partie précédente. Il faudra pour cela utiliser la fonction de coût et son gradient:

$$\mathscr{C}_{\text{moindres carrés}}(a,b) = \left\| \begin{pmatrix} x_1 & 1 \\ \vdots & \vdots \\ x_n & 1 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} - \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \right\|_2^2$$

$$\nabla \mathscr{C}_{\text{moindres carrés}} = 2 \begin{pmatrix} x_1 & \cdots & x_n \\ 1 & \cdots & 1 \end{pmatrix} \left(\begin{pmatrix} x_1 & 1 \\ \vdots & \vdots \\ x_n & 1 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} - \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \right).$$

Q7: Représentez la fonction de pénalisation ρ_σ pour $\sigma = 1$. Calculez les deux premières dérivées de cette fonction par rapport au résidu r. La fonction ρ_σ est-elle convexe?

Q8: Représentez la fonction de coût robuste définie dans les équations (3) et (4) dans l'espace des paramètres (a, b).

Q9: Calculez (à la main) le gradient de la fonction de coût. Représentez le comme un champ de vecteurs avec la fonction Matlab `quiver` ainsi que les lignes de niveaux de la fonction de coût. Vérifiez que le gradient est orthogonal aux lignes de niveau (important: il faut utiliser un repère orthonormé avec la commande `axis equal`).

Q10: Représentez la succession des itérés de l'algorithme des plus fortes pentes appliqué à la fonction de coût robuste $\mathscr{C}_{\text{robuste}}(a,b)$. Utilisez la représentation de la fonction de coût par lignes de niveaux (fonction Matlab `contour`) ainsi que les courbes suivantes illustrant la convergence de la méthode:

- distance entre itérés en fonction de l'itération,

- évolution de la fonction de coût et de la norme du gradient en fonction de l'itération,
- distance en fonction de l'itération à la solution obtenue après un grand nombre d'itérations, par ex: 100.

Commentez les courbes obtenues.

Q11: Répétez cette étude pour d'autres points de départ (valeur initiale des paramètres a et b). Interprétez.

Q12: Appliquez la méthode de quasi-Newton à cette fonction de coût. Représentez, comme à la question 9, les courbes suivantes illustrant la convergence de la méthode:

- distance entre itérés en fonction de l'itération,
- évolution de la fonction de coût et de la norme du gradient en fonction de l'itération,
- distance en fonction de l'itération à la solution obtenue après un grand nombre d'itérations, par ex: 100.

Commentez et comparez aux résultats obtenus avec la méthode des plus fortes pentes.

Q13: Représentez les droites ajustées au nuage de points dans le cas de l'ajustement au sens des moindres carrés et dans le cas de la pénalisation robuste de Cauchy pour différentes valeurs initiales des paramètres a et b. Commentez, puis étudiez l'influence du paramètres σ de la pénalisation de Cauchy.

Q14: Montrez que l'estimateur robuste que nous avons étudié correspond à l'estimateur du maximum de vraisemblance vu dans le cours d'estimation pour le signal d'Olivier Alata pour une modélisation probabiliste des erreurs indépendantes et identiquement distribuées: $Y \sim \text{Cauchy}(ax + b, \sigma)$.

Q15: (bonus) Comparez les droites obtenues en minimisant $\mathscr{C}_{\text{robuste}}$ à celles que donne l'algorithme RANSAC décrit ci-dessous:

1. choisir deux points de mesure aléatoirement et définir la droite $y = a_t\, x + b_t$ passant par ces deux points.
2. sélectionner les points de mesure (x_i, y_i) qui sont bien expliqués par la droite construite à l'étape 1: points tels que $|y_i - (a_t\, x_i + b_t)| < \epsilon$

3. si le nombre de points sélectionnés est supérieur à αn (où α est un paramètre compris entre 0 et 1 et n est le nombre de points de mesure), alors
   - ré-estimer les paramètres a et b au sens des moindres carrés en n'utilisant que les points sélectionnés à l'étape 2
   - calculer l'erreur moyenne de modélisation sur ce sous-ensemble des points de mesure
   - conserver les paramètres estimés a et b si cette erreur est la plus faible obtenue jusque là
4. recommencer à l'étape 1 jusqu'à avoir réalisé le nombre prévu d'itérations.

On pourra utiliser les valeurs suivantes pour les paramètres de l'algorithme RANSAC: nombre d'itérations=1000, $\epsilon = 1$ et $\alpha = 1/2$.