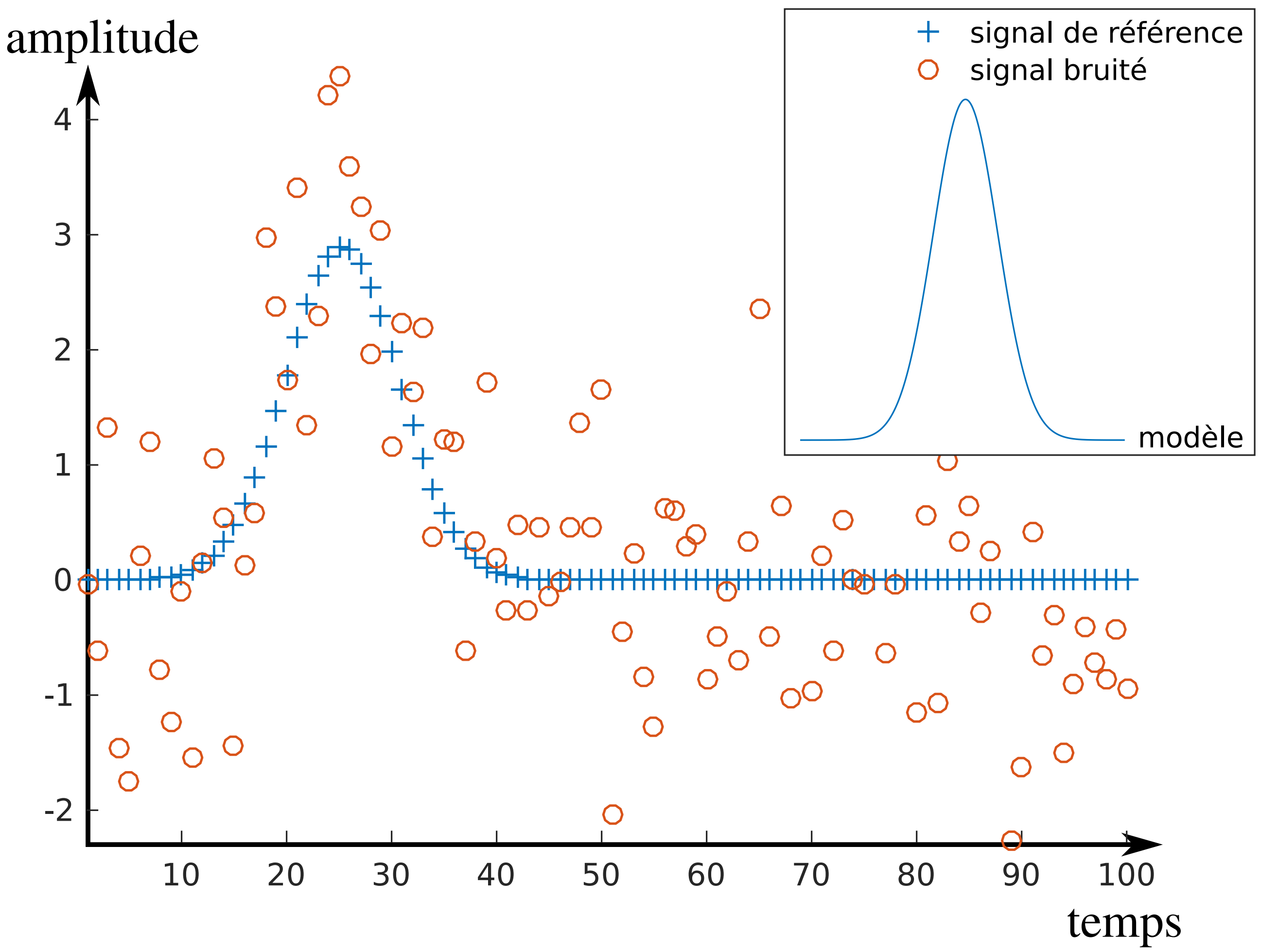
FISE2 – 2016/2017

Projet d'optimisation continue: Estimateur du maximum de vraisemblance

Le but de ce projet est d'appliquer les méthodes d'optimisation continue à un cas concret. Afin de faciliter la compréhension du principe des méthodes d'optimisation, nous commençons par étudier une fonction de coût dépendant d'un seul paramètre avant de traiter le problème de l'estimation conjointe de plusieurs paramètres.

1 Description du problème

On a réalisé un ensemble de n mesures $\boldsymbol{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$ représentées ci-dessous par les ronds rouges (signal `data.mat`).



On souhaite estimer aussi précisément que possible l'instant auquel est survenue l'impulsion modélisée par la gaussienne représentée en trait continu bleu.

On dispose d'un modèle du signal recherché:

$$f_{d,a,s}(t) = a^2 \exp\left[-\frac{(t-d)^2}{2(1+s^2)^2}\right], \tag{1}$$

avec d la data à laquelle l'impulsion est localisée, a^2 son amplitude et s un paramètre de forme contrôlant la largeur du modèle (l'écart-type de la gaussienne vaut $1+s^2$). On dispose également d'un modèle des fluctuations des mesures:

$$\mathrm{p}(\boldsymbol{y}|\boldsymbol{x}) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} \exp\left[\frac{-(y_i - x_i)^2}{2\sigma^2}\right], \tag{2}$$

avec $\boldsymbol{x}$ le signal de référence (signal représenté par des croix bleues sur la figure) et σ l'écart-type du bruit. On considère donc que les erreurs sont indépendantes d'une mesure à l'autre et distribuées selon une loi normale centrée.

2 Formulation du problème d'estimation

Pour estimer la date à laquelle est localisé le signal d'intérêt, nous allons chercher les paramètres d, a et s les plus vraisemblables:

$$(\hat{d}, \hat{a}, \hat{s}) \in \underset{d,a,s}{\arg\max}\ \mathrm{p}(\boldsymbol{y}|f_{d,a,s}(\boldsymbol{t})), \tag{3}$$

ce qui revient à résoudre le problème d'optimisation ci-dessous:

$$(\hat{d}, \hat{a}, \hat{s}) \in \underset{d,a,s}{\arg\min} \sum_i \big(y_i - f_{d,a,s}(t_i)\big)^2. \tag{4}$$

On appelle cette procédure d'estimation une estimation au sens du maximum de vraisemblance. Dans le cas considéré d'erreurs distribuées selon une loi normale, cette procédure est appelée *moindres carrés non linéaires* (car la dépendance en d, a et s est non linéaire).

Q1: Pourquoi les problèmes (3) et (4) sont-ils équivalents?

3 Résolution du problème d'optimisation: cas où seule la date d est inconnue

Nous allons supposer dans cette partie que $s = 2$ et $a = 2$.
Le problème d'optimisation est donc ramené à la minimisation d'une fonction $\mathscr{C}(d)$ d'une variable:

$$\underset{d}{\arg\min}\, \mathscr{C}(d) \qquad \text{avec } \mathscr{C}(d) = \sum_i \left(y_i - a^2 \exp\left[-\frac{(t_i-d)^2}{2(1+s^2)^2}\right]\right)^2. \tag{5}$$

Q2: Représenter la fonction de coût $\mathscr{C}(d)$ pour d variant entre 0 et 100.

Q3: Calculer sur le papier la dérivée de la fonction de coût par rapport à la variable d: $\partial \mathscr{C} / \partial d$ (on fera apparaître explicitement les variables a et s, pas leurs valeurs numériques). Quelle condition nécessaire doit-on vérifier pour que $d^\star$ soit un minimiseur de $\mathscr{C}$ ($d^\star \in \arg\min_d \mathscr{C}(d)$)?

Q4: Représenter la dérivée de la fonction de coût $\partial \mathscr{C} / \partial d$ obtenue à la question précédente pour d variant entre 0 et 100. Vérifier le calcul en comparant avec la dérivée obtenue par différences finies (voir la fonction Matlab `diff`).

La méthode de Newton consiste à approximer localement la fonction de coût par un développement de Taylor à l'ordre 2.

Q5: Calculer sur le papier la dérivée seconde de la fonction de coût par rapport à la variable d: $\partial^2 \mathscr{C} / \partial d^2$. Quelle condition doit-on vérifier pour que $d^\star$ soit un minimiseur de $\mathscr{C}$ ($d^\star \in \arg\min_d \mathscr{C}(d)$)?

Q6: Représenter la dérivée seconde de la fonction de coût $\partial^2 \mathscr{C} / \partial d^2$ pour d variant entre 0 et 100. Vérifier le calcul avec la dérivée seconde calculée par différences finies (avec `diff`).

Q7: Sélectionner un point de départ au hasard. Utilisez les valeurs des dérivées première et seconde de la fonction de coût afin de représenter l'approximation quadratique donnée par le développement de Taylor à l'ordre 2 autour du point de départ.

Q8: Représenter les itérés de l'algorithme de Newton sur la fonction de coût et la dérivée première, pour différents points de départ.

Q9: Quelle valeur $\hat{d}$ obtient-on si on s'initialise à la position du maximum du signal?

4 Résolution du problème d'optimisation: cas où les autres paramètres sont mals connus

Nous allons maintenant chercher à minimiser la fonction de coût $\mathscr{C}(\boldsymbol{\theta})$ par rapport au vecteur d'inconnues $\boldsymbol{\theta} = (d, a, s)^t$.

Q10: Représenter la fonction de coût $\mathscr{C}(\boldsymbol{\theta})$.

Q11: Déterminer une valeur approchée des paramètres optimaux en cherchant le minimum parmi les valeurs de la fonction de coût calculées pour une grille de paramètres (c'est à dire, pour les triplets (d, s, a) tels que $d \in \{d_{\min} + k\Delta_d \,|\, k \in \mathbb{N} \text{ et } d_{\min} + k\Delta_d \leq d_{\max}\}$, $s \in \{s_{\min} + k\Delta_s \,|\, k \in \mathbb{N} \text{ et } s_{\min} + k\Delta_s \leq s_{\max}\}$ et $a \in \{a_{\min} + k\Delta_a \,|\, k \in \mathbb{N} \text{ et } a_{\min} + k\Delta_a \leq a_{\max}\}$).

Q12: Pour diviser par deux l'erreur absolue commise sur le triplet $\widehat{(d, s, a)}$ obtenu à la question précédente, combien de calculs sont nécessaires? Combien de calculs seraient nécessaires si n_θ paramètres étaient inconnus?

Q13: Calculer sur le papier le gradient de la fonction de coût $\nabla_{\boldsymbol{\theta}} \mathscr{C}$.

Q14: Vérifier le calcul du vecteur gradient en comparant avec les dérivées partielles obtenues par différences finies (avec `diff`). Quelle erreur relative maximale est commise?

Q15: Implémenter la méthode des plus fortes pentes décrite dans le chapitre 3 du cours. Représenter les itérés $\{\boldsymbol{\theta}_k\}_{k=1..N_{\text{iter}}}$, ainsi que l'écart entre deux itérés successifs $\|\boldsymbol{\theta}_k - \boldsymbol{\theta}_{k-1}\|$, l'écart à l'optimum $\boldsymbol{\theta}_\infty$ (obtenu au bout d'un grand nombre d'itérations) $\|\boldsymbol{\theta}_k - \boldsymbol{\theta}_\infty\|$, l'écart en terme de fonctions de coût $\|\mathscr{C}(\boldsymbol{\theta}_k) - \mathscr{C}(\boldsymbol{\theta}_\infty)\|$ et la norme infinie du gradient $\|\nabla_{\boldsymbol{\theta}} \mathscr{C}(\boldsymbol{\theta}_k)\|_\infty$. On arrêtera l'algorithme lorsque la norme du gradient $\|\nabla_{\boldsymbol{\theta}} \mathscr{C}(\boldsymbol{\theta}_k)\|_\infty$ sera inférieure à 10^{-2}.

Q16: Etudier l'influence des paramètres de la fonction de recherche linéaire ainsi que le choix du point de départ.

Q17: Implémenter la méthode de quasi-Newton et comparer la convergence à celle des plus fortes pentes.

Q18: Utiliser la fonction Matlab `fminunc` pour déterminer le minimum.