Mini-projet de simulation MAP 568 : métamodélisation par moindres carrés et par processus gaussiens

Vous pouvez résoudre les tâches qui suivent en python ou en matlab. Vous avez juste besoin d'un générateur de nombres pseudo-aléatoires de loi $\mathcal{N}(0,1)$.
En matlab, vous pouvez utiliser `randn`.
En python, vous pouvez utiliser dans la librairie `numpy` la fonction `numpy.random.randn`.

1 Moindres carrés classiques

La fonction réelle est :

$$Y_{\rm reel}(x) = x.$$

La fonction observée est :

$$Y_{\rm obs}(x) = Y_{\rm reel}(x) + \epsilon_{\rm mes}(x),$$

où l'erreur de mesure $\epsilon_{\rm mes}(x)$ est gaussienne de moyenne nulle et de variance $\sigma^2_{\rm mes}$. Les erreurs de mesure sont indépendantes entre elles.

On dispose de n observations $\boldsymbol{y}_{\rm obs} = (Y_{\rm obs}(x^{(j)}))_{1\leq j\leq n}$ en des points $(x^{(j)})_{1\leq j\leq n}$ equirépartis sur l'intervalle $[0,1]$.

Le métamodèle à $p=2$ paramètres $\boldsymbol{\beta} = (\beta_1, \beta_2)^T$ est :

$$Y_{\rm meta}(x) = \beta_1 + \beta_2 x.$$

L'objectif est de calibrer le métamodèle à l'aide de $\boldsymbol{y}_{\rm obs}$ et de quantifier l'incertitude de ses prédictions.

1. On suppose qu'on connaît la variance $\sigma^2_{\rm mes}$ des erreurs de mesure.
   (a) Générer un jeu de données pour $n = 10$ et $\sigma_{\rm mes} = 0.1$.
   (b) Quel est le tenseur de susceptibilité ?
   Le tenseur de susceptibilité $\mathbf{H}$ *aux points d'observation est :*

$$H_{ji} = (x^{(j)})^{i-1}, \quad j = 1,\dots,n, \quad i = 1,2,$$

(c) Déterminer un ellipsoïde de confiance à 95% pour $\boldsymbol{\beta}$.
On a :

$$\mathbb{P}\Big(\boldsymbol{\beta}_{\mathrm{vrai}} \in \big\{\boldsymbol{\beta} \in \mathbb{R}^2, (\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}})^T \mathbf{H}^T \mathbf{H} (\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}) \leq 6\sigma_{\mathrm{mes}}^2\big\}\Big) = 95\%.$$

pour

$$\widehat{\boldsymbol{\beta}} = (\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T \boldsymbol{y}_{\mathrm{obs}}$$

où 6 *est le* $(1-\alpha)$*-quantile de la loi du* χ^2 *à* 2 *degrés de liberté pour* $n = 10$ *et* $\alpha = 0.05$.

(d) Déterminer un tube de confiance à 95% pour $Y_{\mathrm{reel}}(x)$ et pour $Y_{\mathrm{obs}}(x)$.
On a :

$$\mathbb{P}\Big(Y_{\mathrm{reel}}(x) \in \big[\widehat{Y}(x) - 2\sigma_{\mathrm{mes}} q_{\mathrm{pred}}(x), \widehat{Y}(x) + 2\sigma_{\mathrm{mes}} q_{\mathrm{pred}}(x)\big]\Big) = 95\%,$$

pour

$$\widehat{Y}(x) = (\boldsymbol{h}(x))^T \widehat{\boldsymbol{\beta}}$$
$$q_{\mathrm{pred}}^2 = (\boldsymbol{h}(x))^T (\mathbf{H}^T\mathbf{H})^{-1} \boldsymbol{h}(x)$$
$$h_i(x) = x^{i-1}, \quad i = 1, 2$$

et 2 *est le* $(1-\alpha/2)$*-quantile de la loi de* $\mathcal{N}(0,1)$ *pour* $\alpha = 0.05$.
Idem pour $Y_{\mathrm{obs}}(x)$ *en remplaçant* q_{pred}^2 *par* $1 + q_{\mathrm{pred}}^2$.

(e) Faire un test des résidus.
On doit avoir :

$$\frac{1}{\sigma_{\mathrm{mes}}^2} \|\widehat{\boldsymbol{\varepsilon}}\|^2 \sim \chi_{n-2}^2$$

avec

$$\widehat{\boldsymbol{\varepsilon}} = \big(\mathbf{I} - \mathbf{H}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\big)\boldsymbol{y}_{\mathrm{obs}}.$$

(f) Recommencer avec d'autres valeurs de n et de σ_{mes}.

2. On suppose qu'on ne connaît pas la variance σ_{mes}^2 des erreurs de mesure.

(a) Proposer un estimateur de la variance de l'erreur de mesure.
On peut proposer :

$$\widehat{\sigma}^2 = \frac{1}{n-2} \|\widehat{\boldsymbol{\varepsilon}}\|^2$$

(b) Déterminer un ellipsoïde de confiance pour $\boldsymbol{\beta}$.
On a :

$$\mathbb{P}\Big(\boldsymbol{\beta}_{\mathrm{vrai}} \in \big\{\boldsymbol{\beta} \in \mathbb{R}^2, (\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}})^T \mathbf{H}^T \mathbf{H} (\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}) \leq 8.9\widehat{\sigma}^2\big\}\Big) = 95\%.$$

où 8.9 *est* $p = 2$ *fois le* $(1-\alpha)$*-quantile de la loi de Fisher* $\mathcal{F}_{2,n-2}$ *pour* $n = 10$ *et* $\alpha = 0.05$.

(c) Déterminer un tube de confiance pour $Y_{\text{reel}}(x)$ et pour $Y_{\text{obs}}(x)$.
On a :

$$\mathbb{P}\Big(Y_{\text{reel}}(x) \in \big[\widehat{Y}(x) - 2.3\widehat{\sigma}q_{\text{pred}}(x)\,,\, \widehat{Y}(x) + 2.3\widehat{\sigma}q_{\text{pred}}(x)\big]\Big) = 1 - \alpha,$$

où 2.3 *est le* $(1-\alpha/2)$*-quantile de la loi de Student* $\mathcal{T}_{n-2}$ *pour* $n = 10$ *et* $\alpha = 0.05$.

(d) Recommencer avec d'autres valeurs de n et de σ_{mes}.

3. Recommencer un certain nombre de fois en retirant le jeu de données. Vérifier empiriquement que le nombre de fois où le vrai $\boldsymbol{\beta}$, i.e. $(0,1)^T$, tombe dans l'ellipsoïde de confiance est bien -à peu près- 95%. Fixer un x_{test} qui ne soit pas dans la grille d'apprentissage (par exemple, $x_{\text{test}} = 1/2$ pour n pair) et vérifier empiriquement que le nombre de fois où le vrai $Y_{\text{reel}}(x_{\text{test}})$, i.e. x_{test}, tombe dans l'intervalle de confiance est bien -à peu près- 95%.

2 Problème avec les moindres carrés

La fonction réelle est :

$$Y_{\text{reel}}(x) = x^2.$$

La fonction observée est :

$$Y_{\text{obs}}(x) = x^2 + \epsilon_{\text{mes}}(x),$$

où l'erreur de mesure $\epsilon_{\text{mes}}(x)$ est gaussienne de moyenne nulle et de variance σ^2_{mes}.

On dispose de n observations $\boldsymbol{y}_{\text{obs}} = (Y_{\text{obs}}(x^{(j)}))_{1\leq j\leq n}$ en des points $(x^{(j)})_{1\leq j\leq n}$ equirépartis sur l'intervalle $[0,1]$.

Le métamodèle à $p = 2$ paramètres $\boldsymbol{\beta} = (\beta_1, \beta_2)^T$ est :

$$Y_{\text{meta}}(x) = \beta_1 + \beta_2 x.$$

4. Générer un jeu de données pour $n = 10$ et $\sigma_{\text{mes}} = 0.1$.
5. On suppose qu'on connaît la variance σ^2_{mes} des erreurs de mesure.
   (a) Déterminer un ellipsoïde de confiance pour $\boldsymbol{\beta}$.
   (b) Déterminer un tube de confiance pour $Y_{\text{reel}}(x)$ et pour $Y_{\text{obs}}(x)$.
   (c) Faire un test des résidus.
   (d) Recommencer avec d'autres valeurs de n et de σ_{mes}.

3 Régression par processus gaussien

La fonction réelle est :
$$Y_{\mathrm{reel}}(x) = x^2.$$

La fonction observée est :

$$Y_{\mathrm{obs}}(x) = Y_{\mathrm{reel}}(x) + \epsilon_{\mathrm{mes}}(x),$$

où l'erreur de mesure $\epsilon_{\mathrm{mes}}(x)$ est gaussienne de moyenne nulle et de variance σ^2_{mes} connue. Les erreurs de mesure sont indépendantes entre elles.

On dispose de n observations $\boldsymbol{y}_{\mathrm{obs}} = (Y_{\mathrm{obs}}(x^{(j)}))_{1\leq j\leq n}$ en des points $(x^{(j)})_{1\leq j\leq n}$ equirépartis sur l'intervalle $[0, 1]$.

Le métamodèle à $p = 2$ paramètres, $\boldsymbol{\beta} = (\beta_1, \beta_2)^T$, est :

$$Y_{\mathrm{meta}}(x) = \beta_1 + \beta_2 x + Z_{\mathrm{mod}}(x).$$

On prend une erreur de modèle sous la forme d'un processus gaussien de fonction d'autocorrélation gaussienne :

$$C_{\mathrm{mod}}(x - \tilde{x}) = \sigma^2_{\mathrm{mod}} \exp\Big(- \frac{(x - \tilde{x})^2}{l_c^2} \Big).$$

L'objectif est de calibrer le métamodèle à l'aide de $\boldsymbol{y}_{\mathrm{obs}}$ et de quantifier l'incertitude de ses prédictions.

1. On suppose $\sigma_{\mathrm{mes}} = 0.1$, $\sigma_{\mathrm{mod}} = 0.2$, $l_c = 0.5$.
   (a) Générer un jeu de données pour $n = 10$ et $\sigma_{\mathrm{mes}} = 0.1$.
   (b) Déterminer un ellipsoïde de confiance pour $\boldsymbol{\beta}$.
   La distribution a posteriori de $\boldsymbol{\beta}$ sachant $\boldsymbol{y}_{\mathrm{obs}}$ est gaussienne de moyenne $\boldsymbol{\beta}_{\mathrm{post}}$ et de matrice de covariance $\mathbf{Q}_{\mathrm{post}}$ données par :

$$\begin{aligned} \boldsymbol{\beta}_{\mathrm{post}} &= (\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{R}^{-1}\boldsymbol{y}_{\mathrm{obs}}, \\ \mathbf{Q}_{\mathrm{post}} &= (\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})^{-1}, \end{aligned}$$

   avec

$$R_{jl} = \sigma^2_{\mathrm{mes}}\mathbf{1}_0(j - l) + C_{\mathrm{mod}}(x^{(j)} - x^{(l)}), \quad j, l = 1, \ldots, n.$$

(c) Déterminer un tube de confiance pour $Y_{\text{reel}}(x)$ et $Y_{\text{obs}}(x)$.
La distribution a posteriori de $Y_{\text{reel}}(x)$ sachant $\boldsymbol{y}_{\text{obs}}$ est gaussienne. La moyenne a posteriori est

$$\hat{Y}_{\text{post}}(x) \quad = \quad (\boldsymbol{h}(x))^T \boldsymbol{\beta}_{\text{post}} + (\boldsymbol{r}(x))^T \mathbf{R}^{-1} \big(\boldsymbol{y}_{\text{obs}} - \mathbf{H}\boldsymbol{\beta}_{\text{post}}\big),$$

où $\boldsymbol{h}^{(0)}$ est le vecteur de susceptibilité au point $\boldsymbol{x}^{(0)}$:

$$h_i(x) = x^{i-1}, \quad i = 1, 2$$

$\boldsymbol{r}(x)$ est le vecteur de covariance

$$r_j(x) = C_{\text{mod}}(x - x^{(j)}), \quad j = 1, \ldots, n,$$

La variance a posteriori est

$$\text{Var}_{\text{post}}\big(Y(x)\big) = \sigma_{\text{mod}}^2 - \begin{pmatrix} \boldsymbol{h}(x) \\ \boldsymbol{r}(x) \end{pmatrix}^T \begin{pmatrix} \mathbf{0} & \mathbf{H}^T \\ \mathbf{H} & \mathbf{R} \end{pmatrix}^{-1} \begin{pmatrix} \boldsymbol{h}(x) \\ \boldsymbol{r}(x) \end{pmatrix}.$$

(d) Tracer des réalisations de $Y_{\text{reel}}(x)$ selon la loi a posteriori.
Utiliser la méthode de Choleski.

2. On suppose $\sigma_{\text{mes}} = 0.1$ et $l_c = 0.5$.

(a) Déterminer σ_{mod} par une méthode de maximum de vraisemblance.
On cherche à minimiser en σ_{mod} la fonction

$$\log\det(\mathbf{R}) + (\boldsymbol{y}_{\text{obs}} - \mathbf{H}\boldsymbol{\beta}_{\text{post}})^T \mathbf{R}^{-1} (\boldsymbol{y}_{\text{obs}} - \mathbf{H}\boldsymbol{\beta}_{\text{post}})$$

où $\mathbf{R}$ (et donc $\boldsymbol{\beta}_{\text{post}}$) dépend de σ_{mod}.

(b) Reprendre les questions précédentes.

4 Détermination d'une longueur

On désire poser un cable de télécommunication sur le fond marin. On travaille sur une section droite du fond marin reliant deux points sous-marins, et on note $x \in [0, 1]$ la position sur ce segment et $f(x)$ la profondeur. On ne connaît pas la profondeur en chaque point, mais on peut l'évaluer à l'aide d'une sonde en quelques points. On veut évaluer, à l'aide du modèle par processus gaussien, la longueur minimale de cable nécessaire pour relier les deux points sous-marins $x = 0$ et $x = 1$ de telle sorte que le cable repose sur le fond marin. La sortie scalaire d'intérêt est donc :

$$L_f = \int_0^1 \sqrt{1 + f'(x)^2} dx.$$

On prendra

$$f(x) = 2 + \cos(4x) + 2x + x^2 - \exp(x) + 0.3\sin(12x).$$

1. Tracer la fonction f, en utilisant une grille régulière de taille (au moins) 1000. Evaluer la quantité d'intérêt L_f numériquement. (Par exemple, approcher l'intégrale par une quadrature et les dérivés par des différences finies.)
2. *Métamodèle par processus gaussien.* On observe f en les points d'observations $(x^{(j)})_{j=1}^{6} = (0, 0.2, 0.4, 0.6, 0.8, 1)$ sans erreur de mesure. On modélise f comme la réalisation d'un processus gaussien $Y(x)$ de fonction moyenne nulle et de fonction de covariance de Matérn $3/2$ avec $\sigma^2 = 0.2$ et $\ell_c = 0.4$. Tracer les points d'observations, la moyenne a posteriori du processus gaussien et les courbes des bornes inférieures et supérieures des intervalles de confiance à 95% pour les valeurs de $f(x)$. Interpréter brièvement le tracé obtenu. En notant $\hat{f}$ la fonction moyenne a posteriori, évaluer numériquement $L_{\hat{f}}$. Vous devriez obtenir une sous-estimation de L_f. Expliquer pourquoi.
3. *Evaluation de la longueur minimale par simulations conditionnelles.* Toujours avec les 6 points d'observations précédents, tracer 5 réalisations conditionnelles du processus gaussien précédent, en superposant ce tracé à celui de f et de la moyenne a posteriori et des intervalles de confiance. Estimer L_f par méthode de Monte Carlo. C'est-à-dire effectuer (au moins) 1000 simulations conditionnelles, et évaluer $L_{\cdot}$ pour chacune d'entre elle. L'estimation de L_f est alors la moyenne empirique de ces évaluations. Un intervalle de confiance conditionnel à 90% pour L_f est défini par les quantiles empiriques d'ordres 5% et 95% de ces évaluations. Donner les valeurs de l'estimation et de l'intervalle de confiance. Les choses devraient mieux se passer que pour la méthode précédente. Expliquer pourquoi.

On peut tracer facilement des réalisations du processus $Y(x)$ avec la loi a posteriori, car il s'agit d'un processus à statistique gaussienne de moyenne et covariance connues analytiquement.

Pour obtenir des réalisations du processus $Y'(x)$, on a deux moyens :
- Les réalisations de $Y(x)$ étant de classe $\mathcal{C}^1$, on peut tirer une réalisation de $Y(x)$, calculer numériquement sa dérivée (par différences finies), ce qui donne une réalisation de $Y'(x)$.
- la loi a posteriori du processus $Y'(x)$ est gaussienne de moyenne :

$$\hat{Y'}_{\rm post}(x) = \frac{d}{dx}\hat{Y}_{\rm post}(x) = (\boldsymbol{h}'(x))^T \boldsymbol{\beta}_{\rm post} + (\boldsymbol{r}'(x))^T \mathbf{R}^{-1}\big(\boldsymbol{y}_{\rm obs} - \mathbf{H}\boldsymbol{\beta}_{\rm post}\big)$$

et de matrice de covariance (de taille 2×2) :

$$\mathbf{Cov}_{Y',\rm post}(x,\tilde{x}) = \frac{\partial^2}{\partial x \partial \tilde{x}} \mathbf{Cov}_{Y,\rm post}(x,\tilde{x})$$

On peut donc tirer directement des réalisations du processus $Y'(x)$ selon la loi a posteriori ainsi définie.

5 Minimisation d'une fonction

On veut minimiser une fonction coûteuse (on cherche le minimiseur global $x_{\min}$ de $f(x)$, supposé unique). On prendra :

$$f(x) = 1 - \sin(2\pi x + 8\exp(2\pi x - 7))(1 + 0.1x), \quad x \in [0, 1]$$

On modélise f comme la réalisation d'un processus gaussien $Y(x)$ de fonction moyenne nulle et de fonction de covariance gaussienne avec $\sigma^2 = 1$ et $\ell_c = 0.1$.

On commence par lancer un plan d'expérience de type grille régulière grossière sur $[0, 1]$ (disons, trois points en 0, 0.5 et 1).

On cherche ensuite à enrichir de manière séquentielle le plan d'expériences pour déterminer au mieux $x_{\min}$. On applique la stratégie de l'"expected improvement" :

- à l'étape n, on a évalué f aux points $(x^{(j)})_{j=1}^n$ et on connaît donc $\mathcal{F}_n = (x^{(j)}, f(x^{(j)}))_{j=1}^n$. On note $f_{\min,n} = \min_{1\leq j\leq n} f(x^{(j)})$ l'estimateur du minimum de f. La loi de $Y(x)$ sachant $\mathcal{F}_n$ est $\mathcal{N}(m_n(x), \sigma_n^2(x))$ avec les formules habituelles.

- si on appelle f en un nouveau point x, alors l'"improvement" $I(x) = (f_{\min,n} - f(x))^+$ augmente si $f(x)$ est plus petit que $f_{\min,n}$ (ici, $y^+ = \max(y, 0)$). On cherche donc un x qui augmente $I(x)$, mais on ne connaît pas f et donc pas I. On propose donc de maximiser l'"expected improvement" :

$$\mathbb{E}\big[(f_{\min,n} - Y(x))^+ \big| \mathcal{F}_n\big]$$

On pose :

$$x^{(n+1)} = \underset{x\in[0,1]}{\operatorname{argmin}} \ \mathbb{E}\big[(f_{\min,n} - Y(x))^+ \big| \mathcal{F}_n\big]$$

et on appelle f en $x^{(n+1)}$.

1. Vérifier qu'on a :

$$\begin{aligned}\mathbb{E}\big[(f_{\min,n} - Y(x))^+ \big| \mathcal{F}_n\big] =&(f_{\min,n} - m_n(x))\Phi\Big(\frac{f_{\min,n} - m_n(x)}{\sigma_n(x)}\Big)\\ &+ \sigma_n(x)\phi\Big(\frac{f_{\min,n} - m_n(x)}{\sigma_n(x)}\Big)\end{aligned}$$

   avec Φ la fonction de répartition et ϕ la densité de la loi $\mathcal{N}(0, 1)$ (pour Φ, penser à la fonction erf).

2. Implémenter la méthode d'"expected improvement", discuter le comportement des points sélectionnés par la méthode, étudier (numériquement) la convergence de l'algorithme vers $x_{\min}$.