

Rapport TP Optimisation bayésienne

Enzo Isnard

22 janvier 2022

1 Génération de réalisations de processus gaussien (1d)

Dans cette partie on s'intéresse à générer des réalisations de processus stochastiques qui sont au coeur de l'optimisation bayésienne : les processus gaussien. Un processus gaussien est un processus stochastique $(X_t)_{t \in T}$ qui vérifie que :

$$\forall S \subset T \text{ fini}, \forall (\lambda_s) \in \mathbb{R}^{|S|}, \exists (\mu, \sigma) \in \mathbb{R} \times \mathbb{R}_+ : \sum_{s \in S} \lambda_s X_s \sim \mathcal{N}(\mu, \sigma)$$

On notera que $X_t \sim \mathcal{GP}(\mu(\cdot), \Sigma(\cdot, \cdot))$ si X_t est un processus gaussien avec $\mathbb{E}[X_t] = \mu(t)$ et que $\text{cov}(X_t, X_s) = \Sigma(t, s)$.

On s'intéresse au cas où Σ est donné par un de ces noyaux (Noyau gaussien, exponentiel et brownien) :

$$\begin{aligned}\Sigma_G(x, x'; \theta, \sigma) &= \sigma^2 \exp(-\|x - x'\|^2 / \theta^2) \\ \Sigma_E(x, x'; \theta, \sigma) &= \sigma^2 \exp(-\|x - x'\| / \theta) \\ \Sigma_B(x, x') &= \min(x, x')\end{aligned}$$

Si on fixe x' on obtient le graphe suivant pour chacun des noyaux :

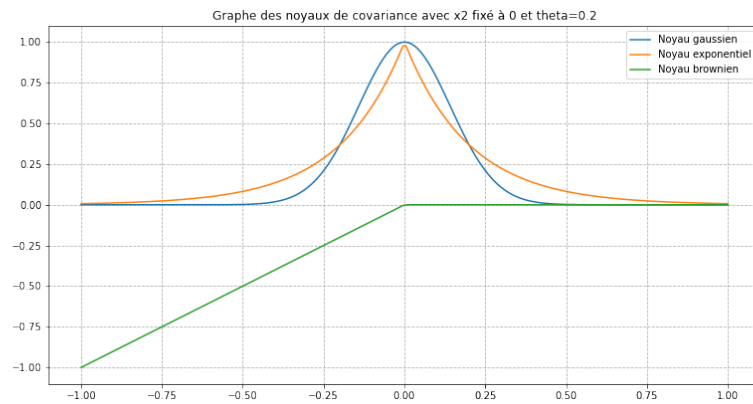
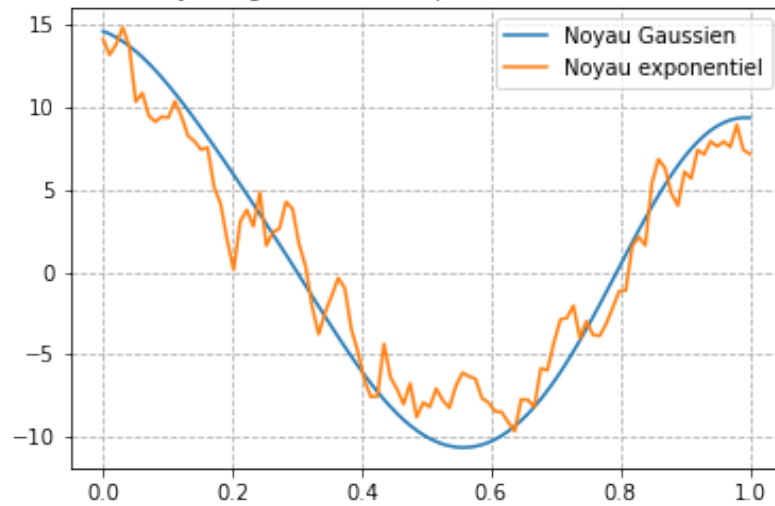


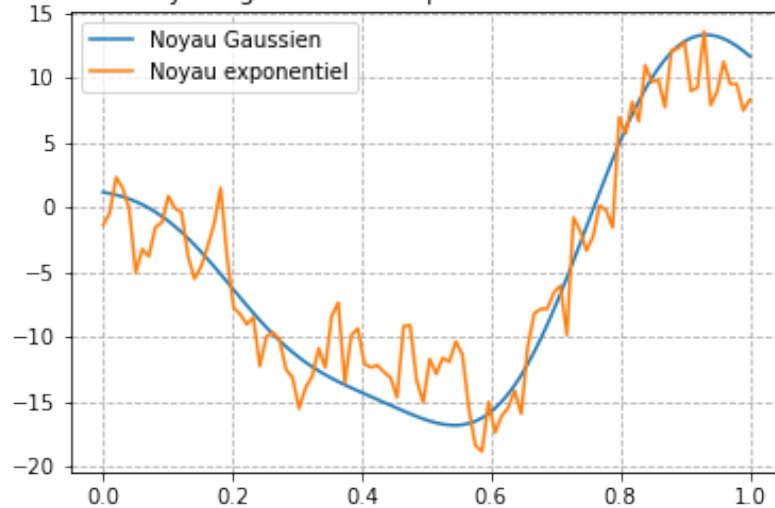
FIGURE 1 – Graphe des noyaux de covariance avec x' fixé à 0, σ à 0 et θ à 0.2

On trace ensuite différentes réalisations de processus gaussiens avec différentes valeurs de σ et θ avec le noyau gaussien et exponentiel.

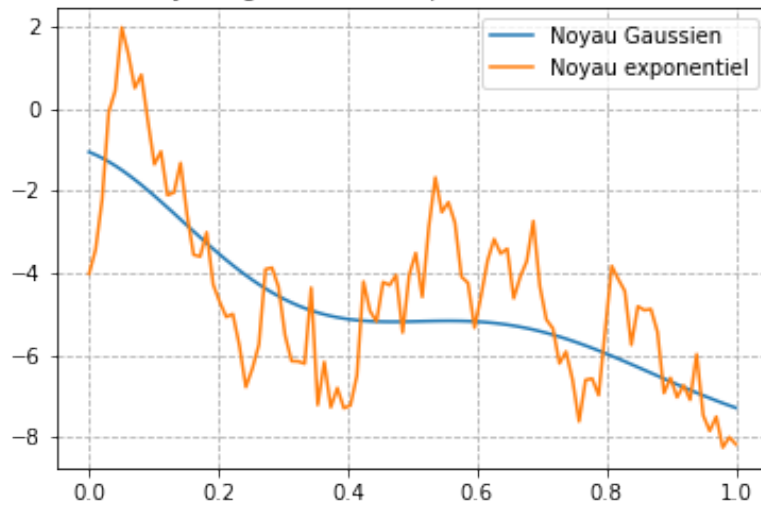
Graphe de réalisations de processus gaussiens
avec les noyaux gaussien et exponentiel avec $\sigma=5$ et $\theta=0.25$



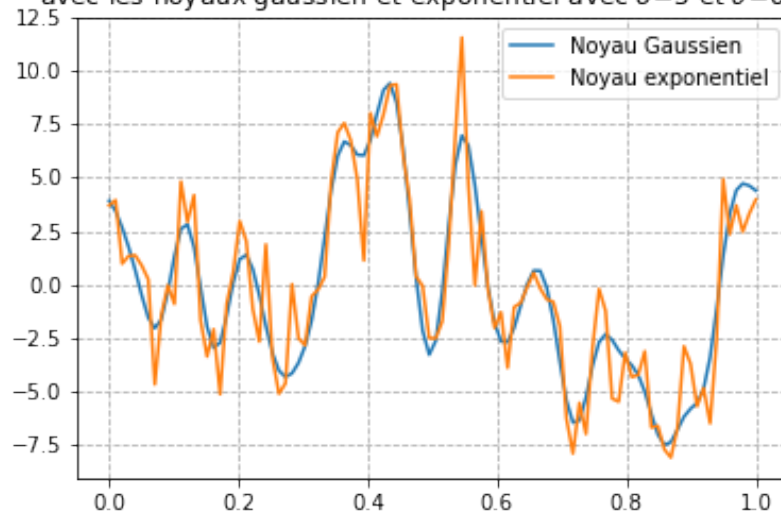
Graphe de réalisations de processus gaussiens
avec les noyaux gaussien et exponentiel avec $\sigma=10$ et $\theta=0.25$



Graphe de réalisations de processus gaussiens
avec les noyaux gaussien et exponentiel avec $\sigma=5$ et $\theta=0.5$



Graphe de réalisations de processus gaussiens
avec les noyaux gaussien et exponentiel avec $\sigma=5$ et $\theta=0.05$



On voit que σ influe sur l'amplitude des variations et θ le nombre d'oscillations.

2 Métamodélisation

Dans cette partie on s'intéresse à créer un méta-modèle d'une fonction f , ie approcher f par une réalisation d'un processus stochastique. On étudie plus précisément une méthode appelée le krigeage qui approche une fonction f par un processus gaussien stationnaire.

On se donne au départ un ensemble de points \mathcal{X} et on évalue la fonction f en chacun des points. On pose $y = f(\mathcal{X})$. Pour estimer une nouvelle valeur x_0 on suppose que toutes les réalisations sont issues d'un processus gaussien et que $f(x_0)$ est de la forme :

$$f(x_0) = \sum_{x \in \mathcal{X}} \lambda_x f(x)$$

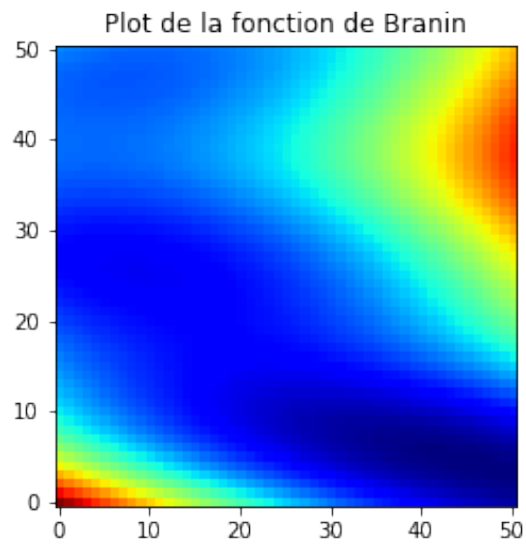
où les λ_x sont des réels. On cherche ensuite à ce que notre estimateur soit le meilleur estimateur non-biaisé au sens qu'il minimise la variance de l'erreur entre la vraie valeur de celle de la valeur estimée.

Si on suppose que le processus gaussien modélisant à une moyenne nulle on trouve que les λ_x vérifient le système linéaire suivant :
 $K_n \lambda = k_n$ où $K_n = (\Sigma(x, x'))_{(x, x') \in \mathcal{X}^2}$ et $k_n = (\Sigma(x, x_0))_{x \in \mathcal{X}}$.

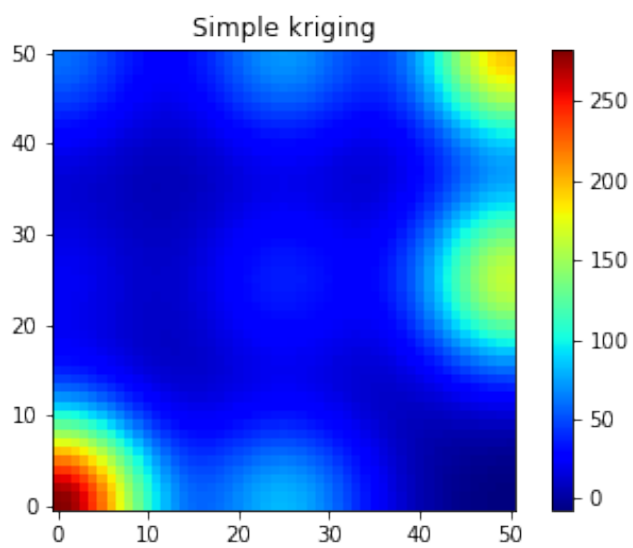
On prend comme fonction la fonction de Branin qui est une fonction benchmark en optimisation :

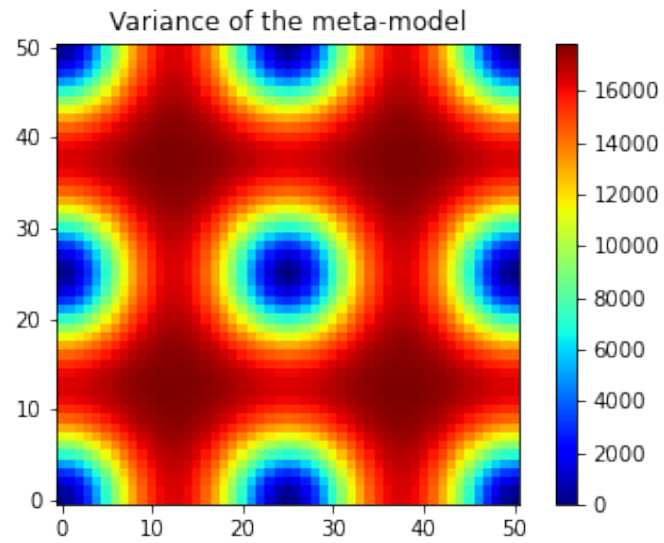
$$f_x(x_1, y_1) = (x_2 - 5.1/(4\pi^2)x_1^2 + 5/\pi x_1 - 6)^2 + 10((1 - 1/(8\pi))\cos(x_1) + 1) + 5x_1$$

Le graphe de cette fonction est le suivant :

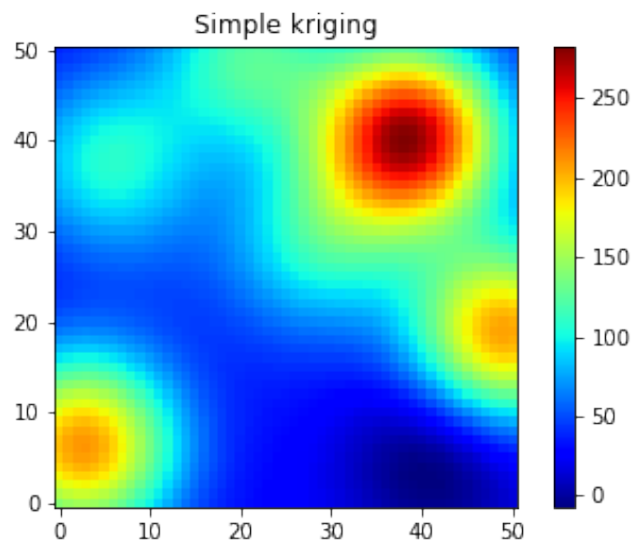


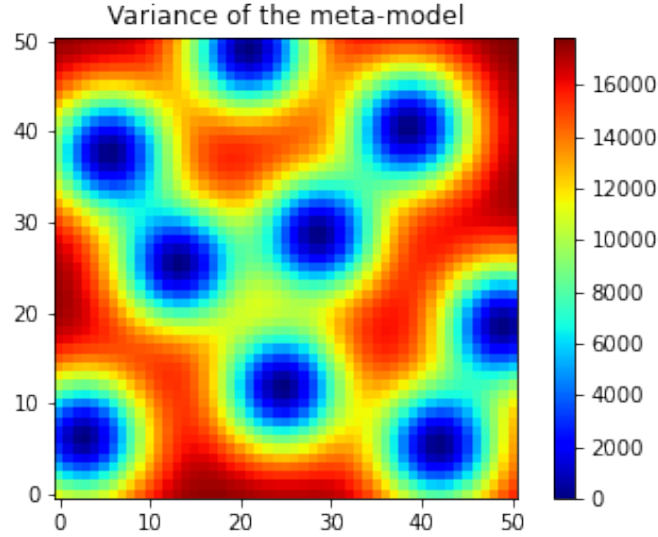
En prenant comme ensemble \mathcal{X} une grille régulière on obtient :





On essaie aussi avec un LHS avec un critère maximin optimisé :





On a que la norme L2 entre le krigeage simple avec la grille uniforme et la vraie fonction est de 2494 alors qu'avec le LHS elle est de 2351, donc on a un résultat un peu meilleur avec le LHS.

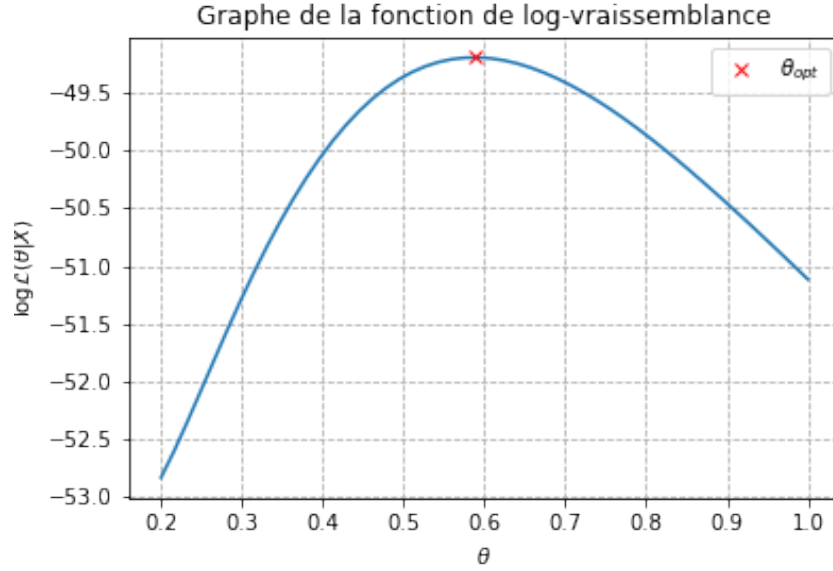
On souhaite maintenant optimiser la valeur du θ du noyau gaussien. Pour cela on calcule la log-vraisemblance des observations avec un paramètre θ et on la maximise pour obtenir le meilleur θ .

La log-vraisemblance a comme expression :

$$\log \mathcal{L}(\theta | \mathcal{X}, y) = -\frac{n}{2}(\log(2\pi) + \log \hat{\sigma}_n + 1) - \frac{1}{2} \log |R_n|$$

où R_n est la matrice de corrélations de \mathcal{X} et $\hat{\sigma}_n$ est un estimateur de la variance.

On a comme graphe pour la log-vraisemblance avec le LHS :



On a $\theta_{opt} = 0.54$.

3 EGO

Dans cette partie on souhaite f_b en utilisant le krigage. Pour cela on commence par faire un krigage avec un ensemble \mathcal{X} donné. On peut calculer ensuite l'espérance qu'un point améliore le minimum des points qu'on a évalués. Plus précisément on calcule la fonction suivante :

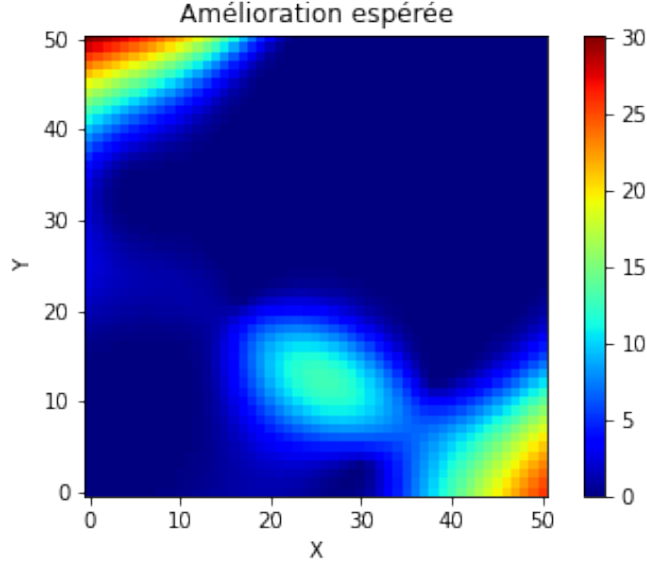
$$EI(x) = \mathbb{E}[\max(y_{min} - X, 0) | \mathcal{X}, x]$$

On peut montrer que EI a en fait comme expression simplifiée dans :

$$EI(x) = (y_{min} - m_n(x))\phi\left(\frac{y_{min} - m_n(x)}{s_n(x)}\right) + s_n(x)\Phi\left(\frac{y_{min} - m_n(x)}{s_n(x)}\right)$$

où est l'espérance du processus gaussien au point x , $s_n(x)$ l'écart-type au point x , ϕ la densité de la loi normale centrée réduite et Φ la fonction de répartition de la loi normale centrée réduite.

On obtient comme graphe de EI en utilisant le LHS et θ_{opt} .



Pour optimiser EI on utilise $BFGS$ car il est assez rapide et que la fonction ne semble pas trop irrégulière. On pourrait aussi utiliser un algorithmes d'optimisation globale comme $CMA - ES$.

En optimisant avec 20 évaluations en utilisant $BFGS$, Nelder-Mead et EGO(en prenant le meme LHS qu'avant comme plan d'expérience initial) on obtient que Nelder-Mead et $BFGS$ obtiennent une valeur de 14 alors que EGO a une valeur de -14. Cela dépend bien sûr de la run.

4 Propagation du feu de forêt

Dans cette partie on optimise une pare-feu pour minimiser la quantité de végétations brûlées lors d'un incendi dans une simulation numérique.

Le domaine de simulation est $\Omega = [0, 1]^2$.

Le pare-feu est une zone rectangulaire de végétation défrichée. Il est déterminé par 4 coordonnées $(x_{min}, x_{max}, y_{min}, y_{max}) = x$.

On impose que le pare-feu soit au dessus la ligne $y = 0.2$.

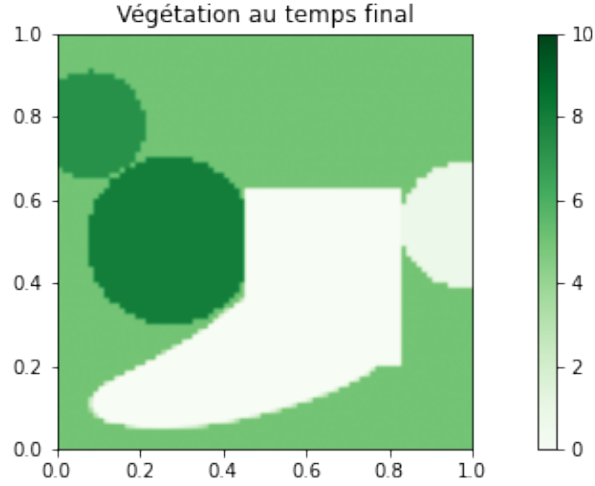
Au début on utilise une fonction pénalisée pour prendre en compte les contraintes :

On note notre ensemble d'admissibilité

$$\mathcal{A} = \{x \in \mathbb{R}^4 : y_{min} \geq 0.2 \text{ et } y_{min} \leq y_{max} \text{ et } x_{min} \leq x_{max}\}$$

Notre fonction coût pénalisé est ainsi donné par :

$$J_{penalized}(x) = J(x) + 100(1 - \mathbf{1}_{\mathcal{A}}(x))$$

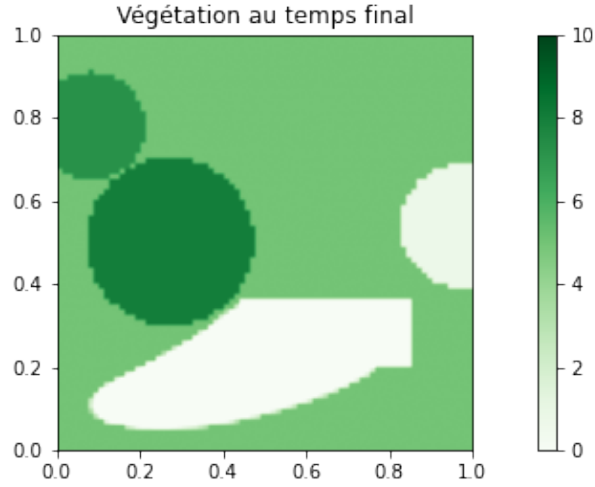


On optimise maintenant avec la contrainte incluse dans la fonction d'acquisition. La nouvelle fonction d'acquisitions est donnée par :

$$EIC(x) = EI(x)\mathbb{P}(x \in \mathcal{A})$$

où l'ensemble des points admissibles est $\{x \in \mathbb{R}^4 : x \in \mathcal{A}\}$.

D'un point de vue pratique on réécrit notre ensemble d'admissibilité comme étant : $\mathcal{A} = \{x \in \mathbb{R}^4 : c(x) \leq h\}$ où $c : \mathbb{R}^4 \rightarrow \mathbb{R}$ et $h \in \mathbb{R}$.



On obtient un bien meilleur résultat avec cette méthode.

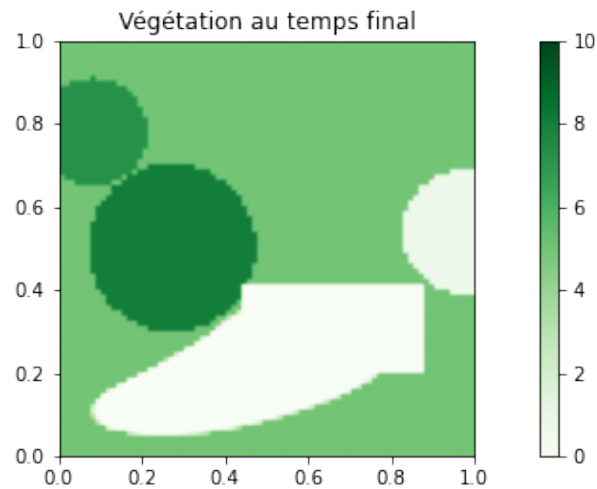
On regarde pas maintenant le cas maintenant où nos entrées sont perturbées par un bruit $\epsilon \sim \mathcal{U}(-\delta, \delta)$ avec $\delta = 0.025$. On veut maintenant optimiser

l'espérance la fonction coût. Pour ce faire on estime l'espérance avec un estimateur de Monte-Carlo :

$$\mathbb{E}[J(X)] = \frac{1}{n} \sum_{i=1}^n J(x_i)$$

où les x_i sont des réalisations i.i.d. de nos entrées.

On obtient en utilisant 10 échantillons pour l'estimateur de MC :



Le résultat dépend beaucoup de la run. Dans ma run j'obtiens un pare-feu plus grand que précédemment.