

CEMRACS 2006

Centre d'Eté Mathématique de Recherche Avancée en Calcul Scientifique

Modélisation de l'aléatoire et propagation d'incertitudes

CIRM Centre International de Rencontres Mathématiques

24/07 - 01/09/2006

Projet : Méthode Spectrale Non Intrusive

Juliette Cahen, Alain Millard, Jean-Marc Martinez

Centre d'Etudes de Saclay DEN/DANS/DM2S

91191 Gif sur Yvette Cedex

Ce projet porte sur le développement des méthodes spectrales fondées sur le chaos polynomial et adaptées à la modélisation des incertitudes en simulation numérique. Elles seront appliquées sur un modèle simulant les écoulements en milieux poreux insaturé.

Les développements en série par polynômes de chaos, seront réalisés en mode non intrusif c'est à dire sans modification du modèle numérique, en considérant le code de calcul comme *boîte noire*. Les coefficients spectraux des variables calculées par le modèle numérique sont déterminés par projection Galerkin ou régression, les premières fondées sur les techniques de quadrature, les secondes sur les moindres carrés associées aux plans d'expériences numériques.

Le travail consistera à développer des composants logiciels permettant d'une part de représenter différentes bases de polynômes (Hermite, Legendre, Laguerre) et d'autre part de calculer les coefficients spectraux par quadrature ou régression. Les chercheurs auront à leur disposition différentes bibliothèques (quadrature tensorisée, régression régularisée par décomposition en valeurs singulières, plans d'expériences par latin hypercube sampling). Les méthodes et les modules seront validés sur des cas tests simples puis sur le code programmé du modèle d'écoulements en milieux poreux.

Table des matières

| | | |
|----------|--|-----------|
| 1 | Présentation du modèle | 3 |
| 2 | Méthode spectrale non intrusive | 3 |
| 2.1 | Quadrature | 4 |
| 2.2 | Régression | 5 |
| 2.3 | Plans d'expériences numériques | 6 |
| 3 | Spécifications fonctionnelles | 6 |
| 3.1 | Variables stochastiques | 7 |
| 3.2 | Développement des polynômes de chaos | 7 |
| 3.3 | Analyse de l'erreur | 8 |
| A | Modèle | 9 |
| B | Polynômes orthogonaux | 11 |
| B.1 | Polynômes d'Hermite | 11 |
| B.2 | Polynôme de Legendre | 12 |
| B.3 | Polynômes de Laguerre | 13 |
| C | Cas tests analytiques | 14 |
| C.1 | Modèle linéaire | 14 |
| C.2 | Modèles non linéaires | 14 |
| C.2.1 | Modèle de Homma-Saltelli | 14 |
| C.2.2 | Modèle de Saltelli-Sobol | 15 |

1 Présentation du modèle

Il s'agit d'étudier un problème d'infiltration dans un milieu poreux non saturé, la phase poreuse étant remplie partiellement d'air et d'eau. La porosité est définie comme étant le rapport entre le volume des vides et le volume total du milieu poreux. On définit la teneur en eau volumique θ du milieu poreux par le rapport entre le volume d'eau et le volume total. L'équation qui régit l'évolution de la teneur en eau dans le milieu découle de la combinaison d'une équation de conservation de masse d'eau, d'une équation de Darcy généralisée et d'une équation liant la teneur en eau à la pression de l'eau.

Dans ce qui suit, on ne s'intéresse qu'à un écoulement vertical, sous l'effet de la pesanteur, dans un milieu poreux alimenté en surface par une condition de débit constant. La coordonnée spatiale (verticale) est notée x . Dans ces conditions, l'équation du problème s'écrit :

$$\frac{\partial \theta}{\partial t} = \frac{\partial}{\partial x} \left[D(\theta) \frac{\partial \theta}{\partial x} \right] - \frac{dK}{d\theta} \frac{\partial \theta}{\partial x}$$

Dans cette équation, $D(\theta)$ représente la diffusivité du milieu et $K(\theta)$ sa conductivité hydraulique. Ces fonctions de la diffusité θ sont déterminées à partir d'essais réalisés sur le milieu poreux et dépendent de l'état du milieu. Elles font intervenir des paramètres incertains. Les incertitudes du problème seront donc introduites en modélisant ces paramètres par des variables aléatoires.

Une description plus détaillée du modèle est donnée en Annexe et notamment la solution de l'équation et la définition des paramètres incertains du problème.

2 Méthode spectrale non intrusive

La Figure 1 illustre la méthode **NISP**, acronyme de *Non Intrusive Spectral method*. On désigne par $y(x, t; u)$ la solution par un solveur numérique d'un modèle déterministe. La solution $y(x, t; u)$ contient l'ensemble des inconnues en espace et temps fonction d'un ensemble de paramètres dont un sous ensemble \mathbf{u} est spécifié avec **incertitudes**. Les incertitudes sur les paramètres incertains u sont modélisées par des variables aléatoires. On supposera que chaque composante u_i du vecteur des paramètres incertains est une variable aléatoire de densité de probabilité $p_i(u_i)$ déterminée. Le problème est donc d'approcher au mieux les caractéristiques probabilistes/statistiques des réponses aléatoires $y(x, t; u)$.

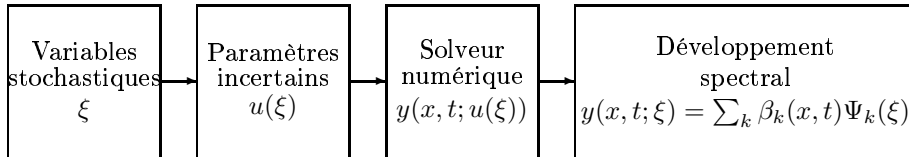


FIG. 1 – *Non Intrusive Spectral Method*

Les variables aléatoires ξ retenues pour la décomposition spectrale, seront appelées *variables stochastiques* qu'on doit distinguer des paramètres incertains u souvent nommés variables ou paramètres aléatoires du calcul. Cette distinction entre les variables stochastiques et les variables aléatoires est importante. Elle permettra d'étudier la convergence des réponses en fonction du choix des variables stochastiques ξ indépendamment des lois de probabilités des paramètres incertains u . La génération des variables incertaines u à partir des variables stochastiques se fera par changement de variables (développé dans la section 4.1 dédiée aux variables stochastiques).

On note $\text{pdf}_i(\xi_i)$ la densité de probabilité des variables stochastiques ξ_i et par ds leur nombre, c'est à dire la dimension stochastique. Compte tenu de leur indépendance :

$$\text{pdf}(\xi_1, \xi_2, \dots, \xi_{\text{ds}}) = \prod_{i=1}^{\text{ds}} \text{pdf}_i(\xi_i)$$

Le choix des variables stochastiques induit la famille des polynômes de chaos $\Psi_k(\xi)$. Les coefficients modaux $\beta_k(x, t)$ seront approchés par quadrature ou régression.

2.1 Quadrature

Le calcul des coefficients modaux par quadrature consiste à approcher l'intégrale :

$$\beta_k(x, t) = \frac{1}{\|\Psi_k\|^2} \int y(x, t; \xi) \Psi_k(\xi) \text{pdf}(\xi) d\xi$$

par méthode d'intégration numérique de **Gauss**.

En simplifiant les notations par suppression des variables espace et temps, on a :

$$\beta_k = \frac{1}{\|\Psi_k\|^2} \int y(\xi) \Psi_k(\xi) \text{pdf}(\xi) d\xi$$

On pourra utiliser la librairie **CubPack** pour le calcul des points ξ^j (points dits de collocation) et des poids ω^j afin d'obtenir l'approximation $\hat{\beta}_k$ des coefficients modaux :

$$\beta_k \simeq \hat{\beta}_k = \frac{1}{\|\Psi_k\|^2} \sum_{j=1}^N y(\xi^j) \Psi_k(\xi^j) \omega^j$$

On pourra également estimer les coefficients modaux par intégration de **Gauss-Legendre** en utilisant la propriété d'uniformisation de toute variable aléatoire continue. On en rappelle le principe. En notant $F(\xi)$ la fonction de répartition d'une variable aléatoire, le changement de variable $\zeta = F(\xi)$ produit une variable aléatoire uniforme ζ dans $[0, 1]$. Et du fait des propriétés de la fonction de répartition (monotonie), la fonction inverse $\xi = F^{-1}(\zeta)$ permet d'établir le lien univoque entre une variable uniforme dans $[0, 1]$ et une variable aléatoire continue de fonction de répartition F . Il suffit donc d'inverser la fonction de répartition.

En notant $F^{-1}(\zeta) = \{F_1^{-1}(\xi_1), F_2^{-1}(\xi_2), \dots\}$ l'ensemble des uniformisations il vient :

$$\beta_k = \frac{1}{\|\Psi_k\|^2} \int_{[0,1]} y(F^{-1}(\zeta)) \Psi_k(F^{-1}(\zeta)) d\zeta$$

On obtient alors l'approximation par quadrature en utilisant l'intégration de Gauss-Legendre :

$$\hat{\beta}_k \simeq \frac{1}{\|\Psi_k\|^2} \sum_{i=1}^N y(F^{-1}(\zeta^i)) \Psi_k(F^{-1}(\zeta^i)) \omega^i$$

Pour cela, on pourra utiliser comme dans le cas précédent la librairie CubPack mais aussi celle développée par Petras basée sur la *cubature* de Smolyak et qui est limitée aux fonctions de poids uniformes (intégration Gauss-Legendre). On propose une comparaison entre ces 2 approches de calcul des coefficients modaux.

2.2 Régression

Dans notre cas, l'approche par régression consiste à calculer les coefficients modaux par moindres carrés des écarts entre valeur calculée par le code et valeur approchée par le développement tronqué sur n simulations avec n supérieur au nombre de coefficients modaux à calculer. En notant β le vecteur des coefficients modaux $(\beta_0, \beta_1, \dots, \beta_p)'$, son estimation $\hat{\beta}$ est donnée par :

$$\hat{\beta}(x, t) = \text{ArgMin}_{\beta} \sum_{j=1}^n [(y(x, t; \xi^j) - \sum_{k=0}^p \beta_k(x, t) \Psi(\xi^j))]^2$$

La recherche optimale se fait évidemment en considérant plus d'équations que d'inconnues, c'est à dire $p + 1 < n$. En notant $Z_{n,p+1}$ la matrice d'éléments $Z_{i,j} = \Psi_j(\xi^i)$ (détaillée dans la section plan d'expériences) et $y' = (y^1, y^2, \dots, y^n)$ le vecteur des réponses $y(\xi^i)$, la résolution du problème fournit les équations *normales* :

$$(Z'Z)\hat{\beta} = Z'y$$

Et lorsque la matrice Z est de rang plein, on obtient la solution classique des moindres carrés :

$$\hat{\beta} = (Z'Z)^{-1} Z'y$$

La solution correspond donc à la minimisation de $\|Z\beta - y\|^2$. Lorsque la matrice Z est mal conditionnée, on peut régulariser la solution par la recherche du minimum de $\|Z\beta - y\|^2 + \lambda\|\hat{\beta}\|^2$. D'autres méthodes peuvent être envisagées comme par exemple celle basée sur une décomposition en valeurs singulières, *Singular Value Decomposition* et ne retenir que les valeurs singulières les plus importantes.

2.3 Plans d'expériences numériques

Les jeux de données des simulations numériques à réaliser correspondent à un plan d'expériences. On note ξ_j^i la valeur de la composante j de la variable stochastique ξ lors de l'expérience i . Le plan d'expériences qui sera spécifié (en quadrature ou régression) sera représenté par la matrice d'expériences X de dimension $n \times \mathbf{ds}$:

$$X = \begin{pmatrix} \xi_1^1 & \xi_2^1 & \cdots & \xi_{\mathbf{ds}}^1 \\ \xi_1^2 & \xi_2^2 & \cdots & \xi_{\mathbf{ds}}^2 \\ \vdots & \ddots & & \vdots \\ \xi_1^n & \xi_2^n & \cdots & \xi_{\mathbf{ds}}^n \end{pmatrix}$$

Pour générer la matrice d'expériences X , on pourra utiliser soit les méthodes de quadrature soit les méthodes de plans d'expériences numériques. Le calcul des points par quadrature ou régression prennent en compte les densités de probabilités $\text{pdf}_i(\xi_i)$. En régression on pourra envisager les méthodes de type Monte Carlo comme par exemple *Simple Random Sampling* ou *Latin Hypercube Sampling*. On pourra envisager d'autres méthodes de planification.

La matrice Z précédemment définie correspond aux valeurs des polynômes multidimensionnels sur les points de la matrice d'expériences X

$$Z = \begin{pmatrix} \Psi_0 & \Psi_1(\xi^1) & \cdots & \Psi_p(\xi^1) \\ \Psi_0 & \Psi_1(\xi^2) & \cdots & \Psi_p(\xi^2) \\ \vdots & \ddots & & \vdots \\ \Psi_0 & \Psi_1(\xi^n) & \cdots & \Psi_p(\xi^n) \end{pmatrix}$$

En notant Ω la matrice diagonale des poids de la quadrature, pour une base de polynômes normalisés, le formalisme de quadrature et de régression de l'estimation des coefficients modaux β est :

$$\begin{aligned} \hat{\beta}_{\text{quadrature}} &= Z' \Omega y \\ \hat{\beta}_{\text{régression}} &= (Z' Z)^{-1} Z' y \end{aligned}$$

Ce formalisme est donc générique et facile à programmer. Il permettra une comparaison aisée des 2 méthodes de calcul des coefficients modaux.

3 Spécifications fonctionnelles

Cette partie définit le *cahier des charges* du composant logiciel **NISP** sur les variables stochastiques, les polynômes de chaos monodimensionnels et les polynômes multidimensionnels associés. La dernière section porte sur l'analyse de l'erreur de modélisation sur les incertitudes des réponses.

3.1 Variables stochastiques

La convergence quadratique de l'approximation $\hat{y}(x, t; z(\xi)) \rightarrow y(x, t; z(\xi))$ dépend évidemment du modèle *physique*, de la modélisation probabiliste des paramètres incertains u_i et du choix des variables stochastiques ξ .

Dans ce projet, on souhaite pouvoir évaluer au moins 3 types de variables stochastiques :

1. normale centrée réduite \rightarrow polynômes d'Hermite
2. exponentielle de paramètre $\lambda = 1 \rightarrow$ polynômes de Laguerre
3. uniforme dans $[0, 1] \rightarrow$ polynômes de Legendre

L'Annexe rappelle la méthode de calcul par récurrence de ces polynômes orthogonaux et leur fonction de poids associée.

Il faudra donc *coder* les transformations entre la loi d'une paramètres incertains u_i et celle de la variable stochastique associée ξ_i . Lorsque les variables stochastiques ont la même loi que les paramètres incertains, la transformation est évidemment très simple. Par exemple, lorsque l'incertitude du paramètre u_i est modélisée par une loi normale $p_i(u_i) = \mathcal{N}(m, \sigma)$, la variable *naturelle* stochastique ξ_i associée est une variable aléatoire centrée réduite de loi $\text{pdf}_i(\xi_i) = \mathcal{N}(0, 1)$. La transformation permet d'exprimer u_i en fonction de ξ_i :

$$u_i = m + \sigma \xi_i$$

Dans le cas où les lois entre les paramètres incertains z_i et les variables stochastiques ξ_i sont différentes, c'est un peu plus compliqué. Par exemple, dans le cas où z_i est une variable aléatoire uniforme sur $[a, b]$ et où ξ_i retenue est la variable aléatoire normale centrée réduite, la transformation inverse s'exprime par :

$$u_i = a + \frac{(b-a)}{2} \left(1 + \text{erf}\left(\frac{\xi_i}{\sqrt{2}}\right)\right)$$

Et enfin, il existera des combinaisons entre lois où l'expression analytique de la fonction inverse n'existe pas comme par exemple celui d'une variable stochastique uniforme associée à une variable normale. La solution consiste alors à faire une interpolation en des points de la fonction $\text{erf}()$ puis à inverser cette fonction tabulée.

3.2 Développement des polynômes de chaos

On rappelle la décomposition des réponses $y(x, t)$ calculées par les code est :

$$y(x, t; \xi) = \sum_k \beta_k(x, t) \Psi_k(\xi)$$

Le nombre des coefficients modaux à calculer dépend de la dimension stochastique ds (dimension du vecteur ξ) et de l'ordre de la troncature no . On rappelle que le nombre de coefficients modaux est donné par :

$$p + 1 = \frac{(\text{ds} + \text{no})!}{\text{ds}! \text{no}!}$$

En notant $\hat{\beta}(x, t)$ l'approximation des coefficients modaux obtenus par quadrature ou régression, le développement spectral tronqué de la solution à l'ordre `no` fournit un *méta-modèle stochastique*, méta-modèle car modèle du modèle numérique et stochastique car représentatif d'une variable aléatoire.

$$\hat{y}(x, t; \xi) = \sum_{k=0}^p \hat{\beta}_k(x, t) \Psi_k(\xi)$$

Lorsque plusieurs variables stochastiques sont retenues, les polynômes de chaos Ψ_k sont multidimensionnels et seront calculés par tensorisation de polynômes de chaos monodimensionnels.

Le polynôme de chaos monodimensionnel de degré d associé à la variable stochastique ξ_i sera noté $\varphi_i^d(\xi_i)$. Et en notant α l'indice multidimensionnel, les polynômes de chaos multidimensionnels s'expriment par :

$$\begin{aligned} \alpha &= \{\alpha_1, \alpha_2, \dots, \alpha_{\mathbf{ds}}\} \\ \alpha_i &\in \{0, 1, \dots, \mathbf{no}\}, i = 1, 2, \dots, \mathbf{ds} \\ \sum_{i=1}^{\mathbf{ds}} \alpha_i &\leq \mathbf{no} \\ \Psi_{k(\alpha)}(\xi_1, \xi_2, \dots, \xi_{\mathbf{ds}}) &= \prod_{i=1}^{\mathbf{ds}} \varphi_i^{\alpha_i}(\xi_i) \end{aligned}$$

L'indice k est lié à l'indice multiple α . On rappelle que les variables stochastiques peuvent être de lois différentes, il suffit qu'elles soient indépendantes.

Le degré du polynôme multidimensionnel est :

$$\text{degré}(\Psi_{k(\alpha)}) = \sum_{i=1}^{\mathbf{ds}} \alpha_i$$

Dans la suite, pour simplifier les notations on ne rappellera pas cette dépendance, en notant k l'indice des polynômes multidimensionnels. La relation entre k et α est évidemment indépendante du choix des variables stochastiques. Elle ne dépend que de la dimension stochastique \mathbf{ds} et de l'ordre `no` et peut donc être codée de façon générique. Nous vous proposons un programme (fonction `C`) qui permet d'obtenir le tableau des équivalences entre les $p + 1$ valeurs de k et les composantes α_i de l'indice multiple α . En entrée on précisera la dimension stochastique \mathbf{ds} et l'ordre `no`. En sortie on récupère dans un fichier (au format `ascii`) un tableau $p + 1 \times \mathbf{ds}$ où les lignes du tableau regroupent les $p + 1$ différentes valeurs de l'indice multiple α .

3.3 Analyse de l'erreur

L'erreur du méta-modèle stochastique se décompose en erreur d'estimation et de troncature¹ :

¹Elle correspond à l'erreur d'approximation en apprentissage statistique

$$y(x, t; \xi) - \hat{y}(x, t; \xi) = \underbrace{\sum_{k=0}^p (\beta_k(x, t) - \hat{\beta}_k(x, t)) \Psi_k(\xi)}_{\text{erreur d'estimation}} + \underbrace{\sum_{k=p+1}^{\infty} \beta_k(x, t) \Psi_k(\xi)}_{\text{erreur de troncature}}$$

Compte tenu de la propriété d'orthogonalité et en notant $\|f\|$ la norme associée au produit scalaire $\langle f, g \rangle = \int f(\xi)g(\xi)\text{pdf}(\xi)d\xi$, on obtient la norme de l'erreur :

$$\|(y(x, t) - \hat{y}(x, t))\|^2 = \sum_{k=0}^p (\beta_k(x, t) - \hat{\beta}_k(x, t))^2 \|\Psi_k\|^2 + \sum_{k=p+1}^{\infty} \beta_k(x, t)^2 \|\Psi_k\|^2$$

Le minimum est donc atteint lorsque l'erreur d'estimation est nulle c'est à dire $\forall k = \{0, 1, \dots, p\}, \hat{\beta}_k(x, t) = \beta_k(x, t)$. Cette analyse basée sur la norme L_2 est donc équivalente à celle de la variance généralement utilisée comme critère d'incertitude dans les analyse d'incertitudes autour d'une valeur nominale².

A Modèle

On rappelle l'équation du problème, où $\theta(x, t)$ est la teneur en eau volumique, et $D(\theta), K(\theta)$ respectivement la diffusivité du milieu et sa conductivité hydraulique :

$$\frac{\partial \theta}{\partial t} = \frac{\partial}{\partial x} [D(\theta) \frac{\partial \theta}{\partial x}] - \frac{dK}{d\theta} \frac{\partial \theta}{\partial x}$$

Dans cette équation, $D(\theta)$ représente la diffusivité du milieu et $K(\theta)$ sa conductivité hydraulique. Ces fonctions sont déterminées à partir d'essais réalisés sur le milieu poreux et dépendent de l'état du milieu. On caractérise l'état du milieu par la teneur en eau à charge nulle θ_s et à l'état initial θ_n . On définit la teneur en eau relative :

$$\Theta = \frac{\theta - \theta_n}{\Delta\theta} \quad \text{où} \quad \Delta\theta = \theta_s - \theta_n$$

Les fonctions $K(\theta)$ et $D(\theta)$ sont prises de la forme :

$$\begin{aligned} K(\theta) &= K_n + \Delta K \Theta^2 \frac{C-1}{C-\Theta} \\ D(\Theta) &= h(C) \left[\frac{S}{\Delta\theta(C-\Theta)} \right]^2 \end{aligned}$$

Dans ces expressions :

$$\begin{aligned} \Delta K &= K_s - K_n \\ h(C) &= \frac{C(C-1)(\pi(C-1) + B)}{4(C-1) + 2B} \quad \text{où} \quad B = 1.46147 \end{aligned}$$

²à distinguer des études de fiabilité visant à calculer la probabilité de dépassement d'un seuil où les réalisations extrêmes se situent en général sur les *queues* de distribution, loin de la valeur moyenne.

Les fonctions $K(\Theta)$, $D(\Theta)$ dépendent donc des paramètres C, K_s, K_n, S . C'est sur ces paramètres que porteront les incertitudes de la modélisation.

Le milieu est soumis à une condition de débit constant R en $x = 0$, ce qui se traduit par :

$$K(\Theta) - D(\theta) \frac{\partial \theta}{\partial x} = R$$

Par ailleurs, au temps $t = 0$, l'ensemble du milieu est dans un état non saturé, caractérisé par la condition :

$$\theta(x, t = 0) = \theta_n, x \geq 0$$

Une solution analytique de ce problème est donnée dans la référence suivante : P. Broadbridge, I. White, 'Constant rate rainfall infiltration : A versatile nonlinear model 1. Analytic solution', Water Resources Research, vol 24, N°1, pp 145-154, 1988. On se propose d'utiliser cette solution analytique pour les calculs déterministes. On donne ci-après les principaux éléments de cette solution.

Tout d'abord on introduit les variables unidimensionnelles suivantes :

$$\begin{aligned} x^* &= \frac{x}{\lambda_s} & \text{où } \lambda_s &= \frac{h}{C(C-1)} \frac{S^2}{\Delta \theta \Delta K} \\ t^* &= \frac{t}{t_s} & \text{où } t_s &= \frac{h}{C(C-1)} \frac{S^2}{(\Delta K)^2} \end{aligned}$$

On introduit :

$$\begin{aligned} K^* &= \frac{(K - K_n)}{\Delta K} \\ D^* &= \frac{D}{D_r} \\ D_r &= \frac{\lambda_s^2}{t_s} \\ R^* &= \frac{R - K_n}{\Delta K} \end{aligned}$$

On pose enfin :

$$\begin{aligned} m &= 4C(C-1) \\ \rho &= \frac{R^*}{m} \\ \tau &= mt^* \end{aligned}$$

La solution est donnée sous forme paramétrée en ζ : $\Theta(\zeta, \tau), x^*(\zeta, \tau)$. En posant :

$$\begin{aligned} A &= \exp[-\zeta^2 \tau^{-1}] \\ B_1 &= (\zeta - \rho \tau) \tau^{-1/2} \\ B_2 &= (\zeta + \rho \tau) \tau^{-1/2} \\ C_1 &= \zeta - \rho(1 + \rho^{-1})^{1/2} \tau] \tau^{-1/2} \\ C_2 &= \zeta + \rho(1 + \rho^{-1})^{1/2} \tau] \tau^{-1/2} \end{aligned}$$

il vient :

$$u = \frac{1}{2}A\{2 \exp(B_2^2) + f(C_1) - f(B_1) + f(C_2) - f(B_2)\}$$

$$\frac{\partial u}{\partial \zeta} = \rho A[2 \exp(B_2^2) - (1 + \rho^{-1})^{1/2}\{f(C_1) - f(C_2)\} + f(B_1) - f(B_2)]$$

où la fonction $f(x) = \exp(x^2) \operatorname{erfc}(x)$.

Et en, particulier, la solution au niveau de la surface libre ($x = 0$) est donnée par :

$$\Theta_0 = C\left\{1 - \frac{1}{1 + 2\rho\{1 - \exp(-\rho\tau) \operatorname{erfc}(-\rho\tau^{1/2}) + (1 + \rho^{-1})^{1/2} \operatorname{erf}[\rho(\rho + 1)\tau]^{1/2}\}}\right\}$$

Les propriétés du milieu susceptibles de variabilité sont :

- C : variation entre 1.02 et 1.5 (peut aller jusqu'à l' ∞). On utilisera une densité de probabilité exponentielle ($\lambda e^{-\lambda x}$).
- K_n : ordre de grandeur pour de l'argile : $10^{-10}ms^{-1}$,
- K_s : ordre de grandeur pour de l'argile : $10^{-17}ms^{-1}$. Chacune de ces conductivités hydrauliques peut varier d'un ordre de grandeur dans les 2 sens, par exemple pour K_n de 10^{-11} à $10^{-9}ms^{-1}$. Par contre ces 2 paramètres pourront être corrélés. On utilisera, pour ces paramètres, des lois permettant de se limiter à des grandeurs nécessairement positives (loi normale tronquée, loi lognormale).
- S : ordre de grandeur pour de l'argile : $10^{-4}ms^{-1/2}$. On utilisera, pour ce paramètre, une loi permettant de se limiter à une grandeur nécessairement positive (loi normale tronquée, loi lognormale).

B Polynômes orthogonaux

B.1 Polynômes d'Hermite

Les polynômes d'Hermite d'ordre n sont les polynômes définis par la relation de récurrence :

$$H_{n+1}(x) = 2xH_n(x) - 2nH_{n-1}(x)$$

et les conditions d'initialisation :

$$H_0(x) = 1, H_1(x) = 2x$$

La fonction de poids est $\omega(x) = e^{-x^2}$ sur $]-\infty, +\infty[$

$$\int_{-\infty}^{+\infty} H_n(x)H_m(x)e^{-x^2} dx = \sqrt{\pi}2^n n! \delta_{n,m}$$

Pour améliorer la précision des calcul, il est préférable d'avoir des polynômes normalisés. En posant :

$$\tilde{H}_n(x) = \frac{H_n(x)}{\sqrt{2^n n!}}$$

on obtient alors une nouvelle relation de récurrence des polynômes \tilde{H}_n :

$$\tilde{H}_{n+1}(x) = x\tilde{H}_n(x)\sqrt{\frac{2}{n+1}} - \tilde{H}_{n-1}\sqrt{\frac{n}{n+1}}$$

avec $\tilde{H}_0 = 1$ et $\tilde{H}_1(x) = \sqrt{2}x$.

On a donc :

$$\int_{-\infty}^{+\infty} \tilde{H}_n(x)\tilde{H}_m(x)e^{-x^2} dx = \sqrt{\pi}\delta_{n,m}$$

Lorsqu'on souhaite utiliser comme fonction de poids, la loi normale centrée réduite $\mathcal{N}(0, 1) = (2\pi)^{-1/2} \int e^{-x^2/2} dx$, le changement de variables $x = \xi/\sqrt{2}$ permet d'obtenir les polynômes associés à la densité normale :

$$\varphi_n(\xi) = \tilde{H}_n(\xi/\sqrt{2})$$

On obtient alors un système de polynômes φ_n orthogonaux pour la densité normale :

$$\int_{-\infty}^{+\infty} \varphi_n(x)\varphi_m(\xi)\frac{e^{-\xi^2/2}}{\sqrt{2\pi}} d\xi = \delta_{n,m}$$

On rappelle que lorsque la base est normalisée, la variance est égale à la somme des carrés des coefficients modaux (à l'exclusion du premier β_0 représentant la moyenne). Et d'autre part pour les analyses de sensibilité globale par décomposition de la variance, on peut également calculer facilement les variances partielles et reconstituer facilement les indices de Sobol.

D'autre part, lorsque l'intégration de **Gauss-Hermite** est définie pour la fonction de poids $w(x) = e^{-x^2}$, il faut corriger les points et poids de la quadrature :

$$\int_{-\infty}^{+\infty} f(x)e^{-x^2} dx = \sum_j \omega_j f(x_j)$$

Par le changement de variables $x = \xi/\sqrt{2}$ on obtient :

$$\int_{-\infty}^{+\infty} f(\xi)\frac{1}{\sqrt{2\pi}}e^{-\frac{\xi^2}{2}} d\xi = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{+\infty} f(\sqrt{2}x)e^{-x^2} dx = \sum_j \frac{\omega_j}{\sqrt{\pi}} f(\sqrt{2}x_j)$$

La correction consiste donc à diviser les poids par $\sqrt{\pi}$ et à effectuer un changement d'échelle sur les points de collocation x_j d'un facteur $\sqrt{2}$.

B.2 Polynôme de Legendre

Les polynômes de Legendre d'ordre n sont définis par la relation de récurrence :

$$(n+1)P_{n+1}(x) = (2n+1)xP_n(x) - nP_{n-1}(x)$$

initialisée par

$$P_0(x) = 1, P_1(x) = x$$

La fonction de poids est $\omega(x) = 1$ sur $[-1, +1]$. Le produit scalaire :

$$\langle P_n, P_m \rangle = \int_{-1}^{+1} P_n(x)P_m(x)dx = \frac{2}{2n+1}\delta_{n,m}$$

Un procédé de normalisation équivalent à celui effectué précédemment sur les polynômes d'Hermite, conduit à une nouvelle relation de récurrence sur les polynômes \tilde{P}_n :

$$\tilde{P}_{n+1} = \frac{\sqrt{(2n+1)(2n+3)}}{n+1}x\tilde{P}_n(x) - \frac{n}{n+1}\sqrt{\frac{2n+3}{2n-1}}\tilde{P}_{n-1}$$

avec $\tilde{P}_0 = 1$ et $\tilde{P}_1(x) = \sqrt{3}x$. Les polynômes \tilde{P}_n vérifient donc la relation :

$$\int_{-1}^{+1} \tilde{P}_n(x)\tilde{P}_m(x)dx = 2\delta_{n,m}$$

Par changement de variable $x = 2\xi - 1$, on obtient les polynômes de Legendre associés à la densité uniforme sur $[0, 1]$:

$$\varphi_n(\xi) = \tilde{P}_n(2\xi - 1)$$

Les polynômes φ_n sont donc orthogonalisés sur la densité uniforme $[0, 1]$:

$$\int_0^{+1} \varphi_n(\xi)\varphi_m(\xi)dx = \delta_{n,m}$$

A noter également, le changement à opérer lorsqu'on sera amené à calculer une intégration de **Gauss-Legendre**. Lorsque le support de la fonction de poids $w(x) = 1$ est $[-1, +1]$, les points x_j et poids ω_j de la quadrature calculés permettent d'approcher l'intégrale sur $[-1, +1]$ par :

$$\int_{-1}^{+1} f(x)dx \simeq \sum_{j=1}^n \omega_j f(x_j)$$

Lorsqu'on souhaite calculer l'intégrale sur $[0, +1]$, il faut *corriger* les points et poids de la quadrature calculés précédemment par le même changement de variables $x = 2\xi - 1$:

$$\int_0^{+1} f(\xi)d\xi = \int_{-1}^{+1} f\left(\frac{x+1}{2}\right)\frac{1}{2}dx \simeq \sum_{j=1}^n \frac{\omega_j}{2} f\left(\frac{x_j+1}{2}\right)$$

B.3 Polynômes de Laguerre

Les polynômes de Laguerre d'ordre n sont définis par la relation de récurrence :

$$(n+1)L_{n+1}(x) = (2n+1-x)L_n(x) - nL_{n-1}(x)$$

initialisée par

$$L_0(x) = 1, L_1(x) = 1 - x$$

La fonction de poids est $\omega(x) = e^{-x}$ sur $[0, +\infty[$

$$\int_0^{+\infty} L_n(x)L_m(x)e^{-x}dx = \delta_{n,m}$$

Dans ce cas, la fonction de poids correspond à la densité de la loi exponentielle de paramètre $\lambda = 1$. De plus cette famille de polynômes orthogonaux est *naturellement* normalisée. Les polynômes φ_n orthogonaux sur la densité exponentielle correspondent donc aux polynômes de Laguerre :

$$\varphi_n(\xi) = L_n(\xi)$$

Il n'y a donc pas modification de l'intégration de **Gauss-Laguerre** :

$$\int_0^{+\infty} f(\xi)e^{-\xi}d\xi = \sum_j \omega_j f(x_j)$$

C Cas tests analytiques

C.1 Modèle linéaire

Dans un premier temps, on vous propose de valider la programmation sur des fonctions simples, comme par exemple les modèles linéaires en les paramètres incertains :

$$y(u) = \alpha_0 + \sum_k \alpha_i u_i$$

On pourra tester le programme sur plusieurs modèles en faisant varier le nombre des paramètres incertains, les coefficients α_i mais aussi les lois de probabilité des paramètres u_i . Lorsqu'il y a même loi entre les paramètres u_i et leur variable stochastique ξ_i associée, on devrait par cubature et régression reconstituer le modèle. On pourra également tester l'approximation lorsque les lois des ξ_i diffèrent de celles des u_i .

C.2 Modèles non linéaires

C.2.1 Modèle de Homma-Saltelli

Ce cas test correspond à une fonction non linéaire et non monotone proposé pour illustrer l'intérêt de la méthode de Sobol en analyse de sensibilité comparée à d'autres méthodes basées sur des approximation linéaires (Pearson) ou monotones (Spearman).

$$f(u_1, u_2, u_3) = \sin(u_1) + a\sin^2(u_2) + bu_3^4\sin(u_1)$$

où u_1, u_2, u_3 sont des variables aléatoires de densité uniforme dans $[-\pi, +\pi]$.

La variance et les variances partielles sont :

$$\begin{aligned} D &= \frac{a^2}{8} + \frac{b\pi^4}{5} + \frac{b^2\pi^8}{18} + \frac{1}{2} \\ D_1 &= \frac{b\pi^4}{5} + \frac{b^2\pi^8}{50} + \frac{1}{2} \\ D_2 &= \frac{a^2}{8} \\ D_3 &= 0 \\ D_{12} &= 0 \\ D_{13} &= \frac{b^2\pi^8}{18} - \frac{b^2\pi^8}{50} \\ D_{23} &= 0 \\ D_{123} &= 0 \end{aligned}$$

On pourra prendre par exemple $a = 7$ et $b = 0.1$. Les valeurs exactes de la variance et des variances partielles seront comparées aux valeurs approchées fournies par le méta-modèle stochastique. Les valeurs approchées de la variance et des variances partielles seront évidemment obtenues à partir des coefficients modaux estimés par quadrature ou régression.

C.2.2 Modèle de Saltelli-Sobol

Ce dernier test comme le précédent correspond également à une fonction non linéaire et non monotone. Dans notre cas, il est intéressant dans la mesure où la fonction n'est pas *régulière* et donc l'approximation par des polynômes peut poser des problèmes.

$$\begin{aligned} f(\mathbf{u}) &= \prod_{i=1}^p g_i(u_i) \text{ où } u_i \text{ uniforme dans } [0, 1] \\ g_i(x_i) &= \frac{|4u_i - 2| + a_i}{1 + a_i} \text{ avec } a_i \leq 0 \end{aligned}$$

La variance et les variances partielles sont :

$$\begin{aligned} D &= \prod_{i=1}^p (D_i + 1) - 1 \text{ avec } D_i = \frac{1}{3(1 + a_i)^2} \\ D_{i_1, i_2, \dots, i_s} &= \prod_{i=1}^s D_{i_i} \end{aligned}$$

La sensibilité de la réponse y en fonction des paramètres incertains u_i dépend des coefficients a_i :

- $a_i = 0 \rightarrow 0 \leq g_i \leq 2$, u_i est une variable influente

- $a_i = 9 \rightarrow 0.9 \leq g_i \leq 1.1$, u_i est une variable peu influente
- $a_i = 99 \rightarrow 0.99 \leq g_i \leq 1.01$, u_i est une variable pas influente

Pour cette fonction aussi, les valeurs exactes de la variance et des variances partielles seront comparées aux valeurs approchées fournies par les méta-modèles obtenus par quadrature ou régression.