

Modéliser les risques et aléas en grande dimension

31 janvier 2022

Sylvain Girard (girard@phimeca.com)

Quel genre d'ingénieur e seras-tu ?

Et pour quoi ?

Quel genre d'ingénieur e seras-tu ?

Et pour quoi ?

Complexité

Aléa & Incertitude



Abstraction & Expérience

Pression sélective sur les organisations... et leurs membres !

Phimeca, entreprise à mission d'ingénieurie

Construire ensemble, par une ingénieurie innovante, une industrie respectueuse de l'homme et de son environnement.



- Depuis 2001
- 30 ingénieur·e·s et docteur·e·s
- Conception & Dimensionnement ; Exploitation & Maintenance

Problématique



1. Colmatage des générateurs de vapeur





Le colmatage obstrue le passage de la vapeur



Le colmatage perturbe la réponse dynamique



i évolution avant nettoyage chimique

Le colmatage perturbe la réponse dynamique



- ↗ : nettoyage chimique
- \checkmark : évolution après nettoyage chimique

Modèle de générateur de vapeur



- Entrée : 16 taux de colmatages
- Sortie : signal temporel de pression

Programmé en Modelica avec ThermoSysPro (https://thermosyspro.com/)

Le modèle reproduit l'effet du colmatage



Tentative d'inversion par filtrage particulaire



Le problème inverse de prédiction des taux colmatage à partir des signaux de pression est mal posé.

Réduction de dimension

Réduire la dimension de G plongé dans \mathbb{R}^m , c'est construire un ensemble approché $A \subset \mathbb{R}^m$ doté d'un système de coordonnées de basse dimension $C \subset \mathbb{R}^l$ (l petit).



Réduction non supervisée : A est une (plus ou moins) « bonne approximation » de G (\rightarrow ACP)

Réduction supervisée : C est un « bon point de départ » pour prédire autre chose (\rightarrow SIR)

Analyse en composantes principales (ACP)

Étant donné z_1, \ldots, z_N de G, l'ACP produit une famille d'espaces vectoriels A_1, \ldots, A_N par ajout successif d'un vecteur de base orthogonal.

La projection sur A_k préserve les distances euclidiennes,

 ⇔ minimise la des normes des résidus,

 ⇔ maximise les distances entre projections,

 ⇔ maximise les distances au barycentre.

 Solution algébrique : décomposition aux valeurs singulières (SVD).

ACP de la sortie dynamique

Les résidus d'approximation en dimension 2 sont presque nuls. (=Les 2 premières composantes expliquent toute la variance).



Le colmatage modifie essentiellement la pente, et en moindre mesure la courbure.

Régression inverse par tranches (SIR)

Soit un vecteur X et une variable Y aléatoires.

 $A \text{ est un espace de réduction de dimension efficace si la projection} \\ \psi \text{ est telle que } Y \mid X \ \sim \ Y \mid \psi(X).$

► Sous des conditions peu contraignantes sur X, la courbe de régression inverse E(X | Y) est contenue dans cet espace.



- 1. Approximation par morceaux (slices) de la régression inverse
- 2. Estimation de l'espace la contenant au mieux par ACP.

Ker-Chau Li (1991) « Sliced Inverse Regression for Dimension Reduction »

Résultat : nouvelle méthode de diagnostic

L'inversion est devenue triviale $(1d \rightarrow 1d)$.



Diagnostic du colmatage 1/2



Générateur de vapeur en bon état.

Diagnostic du colmatage 2/2



La fissuration d'un tube a causé un arrêt fortuit.

Synthèse intermédiaires

Les méthodes linéaires fonctionnent souvent très bien !

 Réduction non supervisée : annalyse en composantes principales (ACP)

 Réduction supervisée : régression inverse par tranches (sliced inverse regression, SIR)

Imaginons pixel noir se déplaçant sur un maillage blancs à n cases. Quelle est la dimension de l'espace vectoriel contenant le nuage de point (dans Rⁿ) de ses positions ?

2. Pronostic de fatigue d'ancrages

Objectif : surveiller et prédire l'endommagement en fatigue des ancrages d'une structure *offshore*.



Pliez et dépliez sans cesse un trombonne :

Il se rompt brutalement à cause de la fatigue.



Méthode de pronostic traditionnelle



Réduction de dimension traditionnelle

Une chaîne de modèles empiriques et d'heuristiques.



Utilisons les mesures embarquées



C'est le mouvement du flotteur et non les vagues qui nous intéressent !

Le signal de mouvement est de grande dimension



Réduction de dimension en cascade



Jumeau numérique de surveillance et pronostic



Variable 1

3. Dispersion de polluants dans l'atmosphère



Exemple inspiré de l'accident de l'usine Lubrizol (Rouen, 2013).





Carte de décision probabiliste



Probabilités de dépassement estimées avec 100 simulations

Limite de l'ACP



Inspiré de Fukunaga & Olsen (1971) « An Algorithm for Finding Intrinsic Dimensionality of Data »

De l'ACP aux modèles auto-associatifs (MAA)

- 1. Un algorithme pour trouver un espace vectoriel de projection préservant la « topologie »
 - =les relations locales de voisinage, là où l'ACP préserve globalement les distances.

2. Utilisation séquentielle de cet algorithme en intercallant à chaque étape l'estimation d'une fonction de « rattrapage » liant les coordonnées de projection aux données de départ
On remplace par le résidu à chaque étape

Stéphane Girard & lovleff (2008) « Auto-Associative Models, Nonlinear Principal Component Analysis, Manifolds and Projection Pursuit »

Algorithme des (MAA)

- 1. Ajout d'un vecteur de base. Soit ψ_k la projection associée.
- **2.** Estimation d'une fonction de rattrapage r_k : $\psi_k(x) \mapsto x$,
- **3.** Remplacer les données par les résidus $x r_k(\psi_k(x))$, et itérer.

L'ACP est un MAA

- préservant les distances euclidienne,
- avec l'identité comme fonction de rattrapage

Fukunaga & Olsen avec MAA



Coordonées des projections



Vecteur directeur des projections



Réduction d'un phénomène spatio-temporel

 \hat{A} gauche : simulation de dispersion atmosphérique de radionucléides par modèle pX **IRSN**

À droite : approximation de basse dimension par un MAA.



Individu de test, absent de l'échantillon d'apprentissage.

Points à retenir

La réduction de dimension supervisée gagne à être connue
 Ker-Chau Li (1991) « Sliced Inverse Regression for Dimension Reduction »

MAA = réduction de dimension « auto-supervisée » + estimation de « fonctions de rattrapage »

Notre objectif : aboutir à une « ACP non linéaire » *robuste*.

Notion d'invariant en machine learning : des transformations préservent les distances, les angles, une topologie...

Merci de votre attention !



- + Valentin, François (Stagiaire)
- + Francesco (doctorant)

Tu veux faire de la science ancrée dans le réel ?

Tu est attentif·ve aux autres et au monde qui t'entoure ?

Parlons en ! (girard@phimeca.com)

Merci de votre attention.

Annexes

Fonction de coût pour les MAA

Remplacement du critère métrique global par un critère topologique local :

maximiser
$$rac{\sum_i \psi_k(x_i)^2}{\sum_j \sum_i c_{ij} \; \psi_k(x_i - x_j)^2} = rac{V}{V_c}$$

avec $c_{ij} = 1$ si x_i et x_j contigus, 0 sinon.

Solution : 1^{er} vecteur propre de $V_c^+ V$ (V_c est singulière)