



Analyse en composantes
principales
en météorologie &
en mécanique des fluides

O.Pannekoucke
Météo-France/ CNRS, CNRM/GAME,
URA 1357

ISAE – Séminaire CPGE
9-11 Mai 2012, Toulouse



Problématique : pourquoi chercher à réduire l'information?

Si un champ complexe se projette sur quelques composantes, alors on peut:

- accélérer la prévision numérique en réduisant le coût de calcul
 - == Réduction de modèles
 - compression de l'information
 - application dans le contrôle en temps réel
 - modélisation simplifiée d'un système complexe (Turbulence, interaction fluide-structure, coeur humain,...)
- mieux comprendre les interactions entre différentes régions du domaine
 - télé-connexion en climat
 - solution récurrente (régime de temps)

Problème:

- Comment savoir si un champ se prête à une simplification de la dynamique?
 - Il faut que la dynamique soit stationnaire (i.e. qu'on ait toujours le même type de phénomènes)
- Comment déterminer ces motifs récurrents dans le phénomène observé ?
 - Une approche possible : l'**analyse en composante principale (ACP)**

Plan de la présentation

1. Rappel: définition et propriété d'une matrice de covariance
 - Première définition d'une composante principale
2. Direction (s) principale (s) comme direction(s) de plus forte variance
 - Deuxième définition d'une composante principale
3. Exemple d'une réduction de modèle par ACP (== PCA / EOF / POD)
 - Equation de Burger (EDP)
4. Introduction au TP : composantes principales & modèle de Lorenz 1963.

1. Rappel: définition et propriété d'une matrice de covariance

1. Rappel: définition et propriété d'une matrice de covariance

Covariance et distribution Gaussienne

La densité de probabilité d'une telle distribution Gaussienne s'écrit pour un vecteur aléatoire \mathbf{x} de moyenne $\mathbb{E}(\mathbf{x}) = \mathbf{m}$ et de matrice de covariance $\mathbb{E}((\mathbf{x} - \mathbf{m})(\mathbf{x} - \mathbf{m})^T) = \mathbf{C}$

$$p(\mathbf{x}) = \frac{1}{\sqrt{\det(2\pi\mathbf{C})}} \exp \left\{ -\frac{1}{2} |\mathbf{x} - \mathbf{m}|_{\mathbf{C}^{-1}}^2 \right\},$$

où l'on a fait intervenir la distance de la Malahanobis $|\mathbf{x}|_{\mathbf{C}^{-1}}^2 = \mathbf{x}^T \mathbf{C}^{-1} \mathbf{x}$, et où *det* désigne le déterminant d'une matrice.

On rappelle qu'une matrice de covariance est une matrice symétrique, définie positive : $\mathbf{C}^T = \mathbf{C}$ et que pour tout $\mathbf{x} \neq 0$, $\mathbf{x}^T \mathbf{C} \mathbf{x} > 0$.

Il existe une base orthonormale (\mathbf{e}_i) dans laquelle cette métrique \mathbf{C} se représente par une matrice diagonale $\mathbf{D} = \text{diag}(\lambda_i)$

$$\mathbf{C} = \sum_i \lambda_i \mathbf{e}_i \mathbf{e}_i^T$$

$\mathbf{P}_i = \mathbf{e}_i \mathbf{e}_i^T$ représente le projecteur orthogonal suivant la direction \mathbf{e}_i
 $\lambda_i > 0$ est la valeur propre de \mathbf{C} suivant \mathbf{e}_i

Ces directions \mathbf{e}_i sont les composantes principales de \mathbf{C}

1. Rappel: définition et propriété d'une matrice de covariance

Densité de probabilité et densité empirique

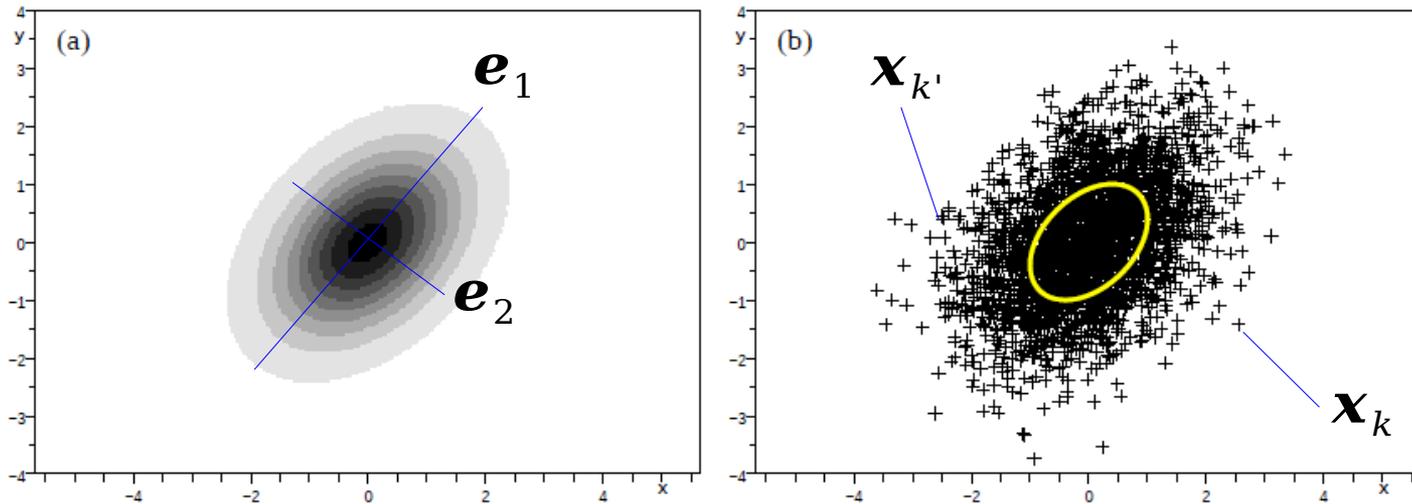


FIG. Représentation de l'allure d'une distribution gaussienne (a) et sa discrétisation à l'aide d'un échantillon de taille finie ($N_s = 3000$) (b). En (a), l'échelle de niveau de gris correspond à l'intensité de la probabilité : noir pour très probable et blanc pour très peu probable. En (b) l'ellipse jaune correspond à l'ellipse caractéristique de la dispersion associée à C (analogue au cercle de rayon σ dans le cas d'une matrice de covariance $C = \sigma^2 I_2$).

1. Rappel: définition et propriété d'une matrice de covariance

Formulation ensembliste des matrices de covariance

Soit $\mathbf{X} = [\mathbf{x}_1 \ \mathbf{x}_2 \ \cdots \ \mathbf{x}_N]$ la matrice dont les colonnes sont des réalisations de la variable aléatoire \mathbf{x}

La matrice de covariance empirique est donnée par

$$\mathbf{C}_N = \frac{1}{N-1} \sum_{k=1}^N (\mathbf{x}_k - \bar{\mathbf{x}})(\mathbf{x}_k - \bar{\mathbf{x}})^T$$

$$\bar{\mathbf{x}} = \frac{1}{N} \sum_{k=1}^N \mathbf{x}_k$$

Dans la suite, on suppose la moyenne empirique comme nulle. La matrice de covariance empirique s'écrit alors

$$\mathbf{C}_N = \frac{1}{N-1} \mathbf{X} \mathbf{X}^T$$

(dans la pratique, on utilise souvent des variables centrées)

1. Rappel: définition et propriété d'une matrice de covariance

Utilisation en assimilation de données

En assimilation de données, $\mathbf{x} \in \mathbb{R}^n$, $n \approx 10^7 - 10^8$ et on rencontre des matrices de très grande dimension

$$\mathbf{C} \in M_n(\mathbb{R}) \Rightarrow n^2 \text{ coefficients i.e. } O(10^{14} - 10^{16})$$

Une manière de réduire leur complexité est d'adopter une réduction de la matrice aux premières composantes principales

$$\mathbf{C} = \sum_{i=1}^n \lambda_i \mathbf{e}_i \mathbf{e}_i^T \approx \mathbf{C}_m = \sum_{i=1}^m \lambda_i \mathbf{e}_i \mathbf{e}_i^T, \text{ où } m \ll n$$

Une autre possibilité est de supposer connues les composantes principales (\mathbf{e}_i) de la matrice, il ne reste alors qu'à estimer la diagonale ce qui réduit la représentation en machine de la matrice à l'équivalent du stockage mémoire d'un seul vecteur

2. Direction (s) principale (s) comme direction (s) de plus forte variance

2. Direction (s) principale (s) comme direction (s) de plus forte variance

Direction de plus grande variance : optimisation sous contraintes

Quelle est la direction suivant laquelle l'information se projette le mieux ?

→ On cherche une direction \mathbf{e}_1 telle que la variance de la projection des états \mathbf{x}_i suivant cette direction est maximale.

En particulier, \mathbf{e}_1 est unitaire, i.e. $\|\mathbf{e}_1\|^2 = 1$

Ces composantes forment un vecteur ligne $\mathbf{X}_{\mathbf{e}_1} = \mathbf{e}_1^T \mathbf{X}$

la variance n'est autre que

$$V(\mathbf{X}_{\mathbf{e}_1}) = \frac{1}{N-1} \mathbf{X}_{\mathbf{e}_1} \mathbf{X}_{\mathbf{e}_1}^T = \frac{1}{N-1} \mathbf{e}_1^T \mathbf{X} \mathbf{X}^T \mathbf{e}_1 = \mathbf{e}_1^T \mathbf{C}_N \mathbf{e}_1 = J(\mathbf{e}_1)$$

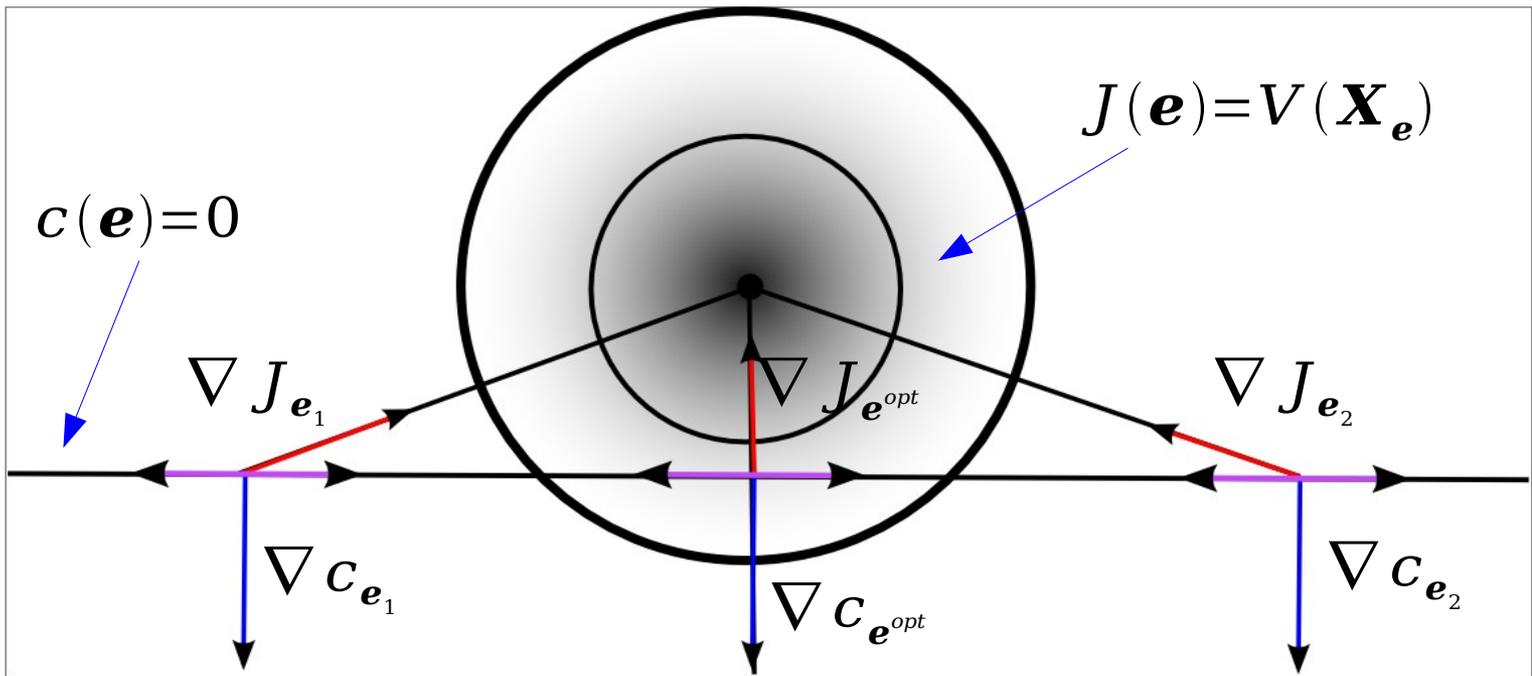
Le problème est donc de maximiser $J(\mathbf{e}_1)$ sous la contrainte $c(\mathbf{e}_1) = 0$

avec $c(\mathbf{e}_1) = \|\mathbf{e}_1\|^2 - 1$

2. Direction (s) principale (s) comme direction (s) de plus forte variance

Résolution de l'optimisation sous contrainte & Lagrangien - 1

$$J(\mathbf{e} + \delta \mathbf{e}_c) = J(\mathbf{e}) + \langle \nabla J_{\mathbf{e}} | \delta \mathbf{e}_c \rangle + O(\delta \mathbf{e}_c^2)$$



$$\nabla J_{\mathbf{e}^{opt}} \in (\text{Ker } dc_{\mathbf{e}^{opt}})^{\perp} = \text{Image}(dc_{\mathbf{e}^{opt}}^T) \Rightarrow \{\exists \mathbf{w}, dc_{\mathbf{e}^{opt}}^T \mathbf{w} = \nabla J_{\mathbf{e}^{opt}}\}$$

$$\nabla J_{\mathbf{e}^{opt}} + \lambda \nabla c_{\mathbf{e}^{opt}} = 0$$

2. Direction (s) principale (s) comme direction (s) de plus forte variance

Résolution de l'optimisation sous contrainte & Lagrangien - 2

\mathbf{e}^{opt} maximise $J(\mathbf{e})$ sous la contrainte $c(\mathbf{e})=0$

implique:

$$(1) \quad \nabla J_{\mathbf{e}^{opt}} + \lambda^{opt} \nabla c_{\mathbf{e}^{opt}} = 0$$

$$(2) \quad c(\mathbf{e}^{opt}) = 0$$

Ce qui se résume à la stationnarité en $(\mathbf{e}^{opt}, \lambda^{opt})$ du « Lagrangien »

$$L(\mathbf{e}, \lambda) = J(\mathbf{e}) + \lambda c(\mathbf{e})$$

λ est appelé le « multiplicateur de Lagrange »

En effet, $\delta L = (\partial_{\mathbf{e}} L)^{opt} \delta \mathbf{e} + (\partial_{\lambda} L)^{opt} \delta \lambda = 0$

est équivalent à

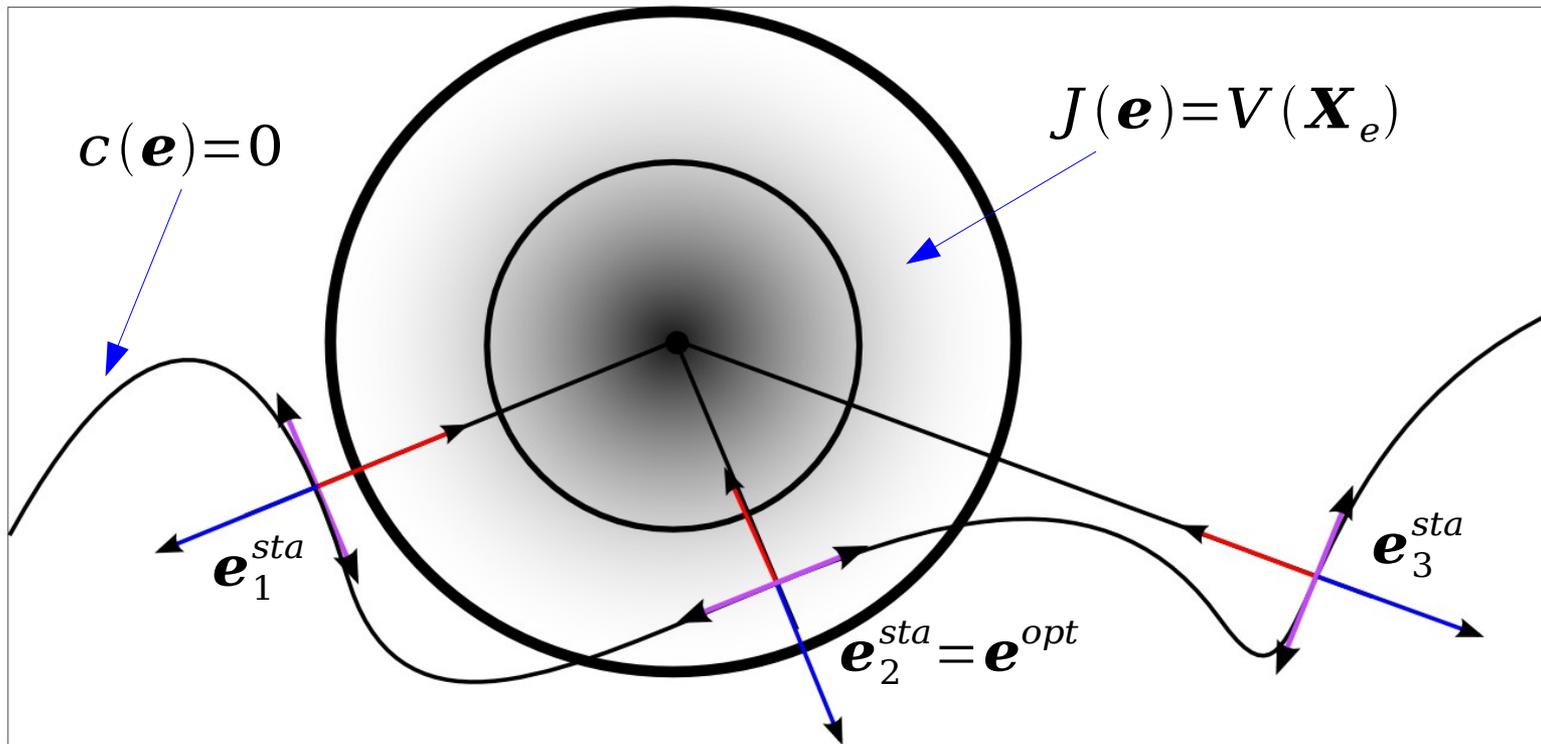
$$(1) \quad (\partial_{\mathbf{e}} L)^{opt} = \nabla J_{\mathbf{e}^{opt}} + \lambda^{opt} \nabla c_{\mathbf{e}^{opt}} = 0$$

$$(2) \quad (\partial_{\lambda} L)^{opt} = c(\mathbf{e}^{opt}) = 0$$

2. Direction (s) principale (s) comme direction (s) de plus forte variance

Résolution de l'optimisation sous contrainte & Lagrangien - 2

Une condition nécessaire et non suffisante :
il peut y avoir plusieurs points stationnaires (\mathbf{e}_i^{sta})
reste à trouver lequel vérifie l'optimalité parmi ces points



2. Direction (s) principale (s) comme direction (s) de plus forte variance

Résolution du problème avec contrainte pour l'ACP

$$L(\mathbf{e}, \lambda) = \mathbf{e}^T \mathbf{C}_N \mathbf{e} - \lambda (\|\mathbf{e}\|^2 - 1)$$

- (1) $\mathbf{C}_N \mathbf{e} = \lambda \mathbf{e}$ \mathbf{e} est un vecteur propre de \mathbf{C}_N associé à λ
(2) $\|\mathbf{e}\|^2 = \mathbf{e}^T \mathbf{e} = 1$

$$(1) \quad V(X_{\mathbf{e}}) = \mathbf{e}^T \mathbf{C}_N \mathbf{e} = \lambda$$

λ s'interprète comme la variance suivant la direction \mathbf{e}

\mathbf{e}_1^{opt} est donc le vecteur propre de \mathbf{C}_N associé à la plus grande valeur propre

2. Direction (s) principale (s) comme direction (s) de plus forte variance

Résolution du problème avec contrainte pour l'ACP

Ayant trouvé la première direction principale \mathbf{e}_1 , on cherche une deuxième direction \mathbf{e}_2 , normalisée, avec comme **autre contrainte** qu'elle soit orthogonale à la première direction.

$$L(\mathbf{e}_2, \lambda, \mu) = \mathbf{e}_2^T \mathbf{C}_N \mathbf{e}_2 - \lambda (\|\mathbf{e}_2\|^2 - 1) - \mu \mathbf{e}_1^T \mathbf{e}_2$$

L'optimum est un point vérifiant la stationnarité du Lagrangien:

$$(1) \quad (\partial_{\mathbf{e}_2} L) = \mathbf{C}_N \mathbf{e}_2 - \lambda \mathbf{e}_2 - \mu \mathbf{e}_1 = 0$$

$$(2) \quad (\partial_{\lambda} L) = \|\mathbf{e}_2\|^2 - 1 = 0$$

$$(3) \quad (\partial_{\mu} L) = \mathbf{e}_1^T \mathbf{e}_2 = 0$$

$\mathbf{e}_1^T \mathbf{C}_N \mathbf{e}_2 = \mu = 0$ s'interprète comme la **covariance** entre \mathbf{e}_1 et \mathbf{e}_2

$\mathbf{e}_2^T \mathbf{C}_N \mathbf{e}_2 = \lambda$ s'interprète comme la **variance** suivant la direction \mathbf{e}_2

\mathbf{e}_2^{opt} est donc le vecteur propre de \mathbf{C}_N associé à la deuxième plus grande valeur propre.

L'analyse en composantes principales construit une base orthonormale dans laquelle l'information se projette le mieux. Ces composantes sont les vecteurs propres de la matrice de covariance \mathbf{C}_N

3. Exemple d'une réduction de modèle par ACP

aussi appelé:

Analyse en Composantes Principales (ACP)

Principal Component Analysis (PCA)

Empirical Orthogonal Function (EOF)

Proper Orthogonal Decomposition (POD)

3. Exemple d'une réduction de modèle par ACP

Réduction de modèle dans le cas d'une EDP : l'éq. de Burger

$$(E) \quad \partial_t u + u \partial_x u = \nu \partial_x^2 u$$

Sur un domaine circulaire de périmètre L

Supposons que les composantes principales soient les exponentielles circulaires

$$(1) \quad e_n(x) = e^{ik_n x}, \quad k = \frac{2\pi}{L}$$

La dynamique s'exprime par la description de l'évolution en temps des composantes dans cette base

$$(2) \quad u(x, t) = \sum_{n=-\infty}^{+\infty} u_n(t) e_n(x)$$

$$(3) \quad u_n(t) = \langle e_n | u \rangle = \frac{1}{L} \int_0^L u(x, t) \bar{e}_n(x) dx$$

$$(4) \quad \partial_x u(x, t) = \sum_{n=-\infty}^{+\infty} u_n(t) \partial_x e_n(x) = \sum_{n=-\infty}^{+\infty} u_n(t) (ink) e_n(x)$$

$$(5) \quad u(x, t) \approx u_T(x, t) = \sum_{n=-T}^T u_n(t) e_n(x)$$

3. Exemple d'une réduction de modèle par ACP

Réduction de modèle dans le cas d'une EDP : l'éq. de Burger

$$(E) \quad \partial_t u + u \partial_x u = \nu \partial_x^2 u$$

La projection de Galerkin consiste à ré-écrire la dynamique (E) dans cette représentation:

$$(1) \quad \frac{d}{dt} u_n + \langle e_n | u_T \partial_x u_T - \nu \partial_x^2 u_T \rangle = 0, \forall n \in [-T, T]$$

$$(2) \quad \frac{d}{dt} u_n + \langle e_n | u_T \partial_x u_T \rangle = -\nu n^2 k^2 u_n,$$

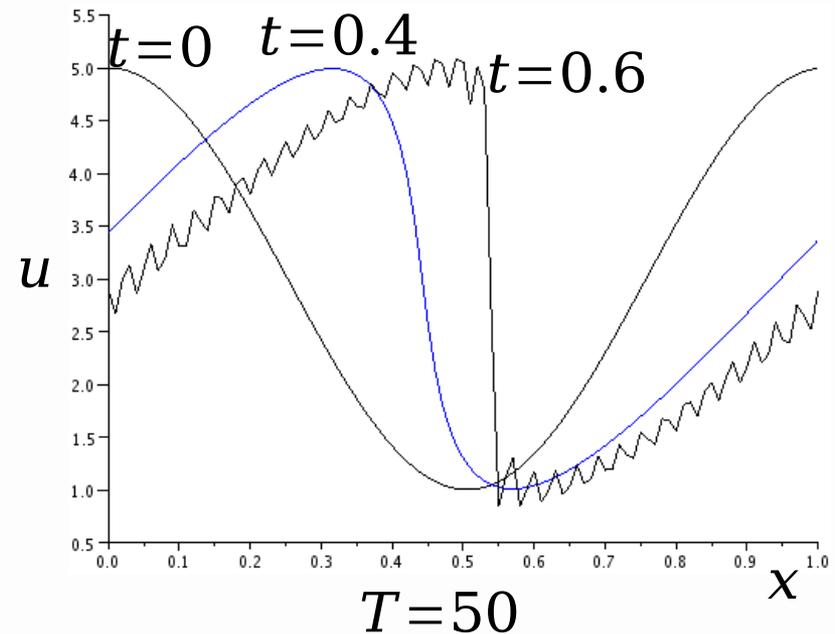
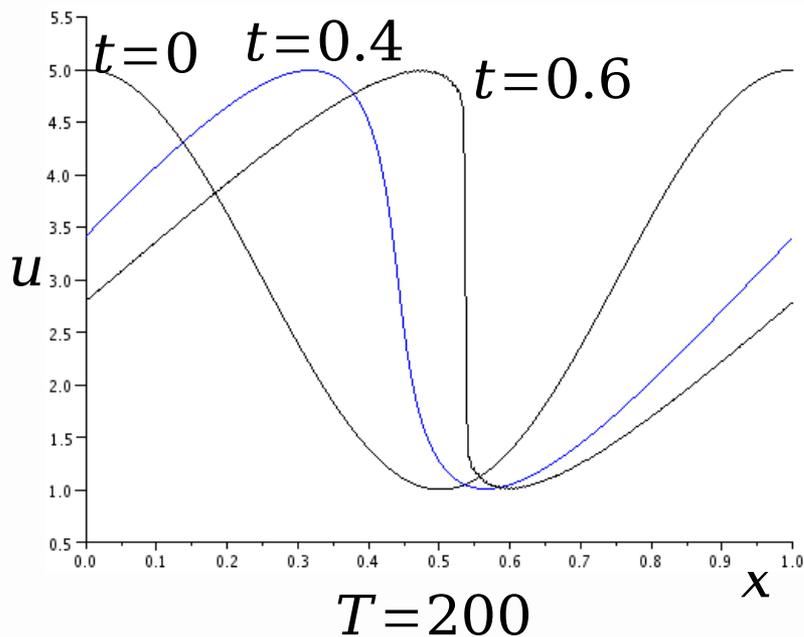
$$(3) \quad \frac{d}{dt} u_n + \sum_{p+q=n} u_p (i q k u_q) = -\nu n^2 k^2 u_n,$$

$$(4) \quad \frac{d}{dt} u_n = -\sum_{p+q=n} u_p (i q k u_q) - \nu n^2 k^2 u_n.$$

3. Exemple d'une réduction de modèle par ACP

Réduction de modèle dans le cas d'une EDP : l'éq. de Burger

$$(E) \quad \partial_t u + u \partial_x u = \nu \partial_x^2 u$$

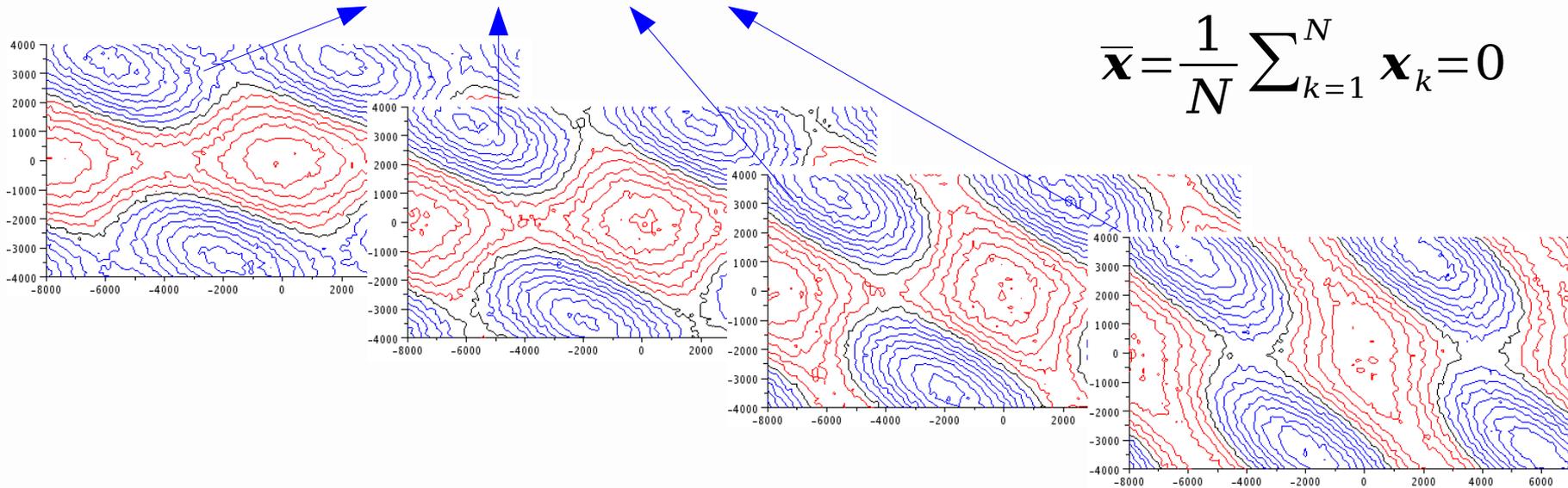


4. Introduction au TP : composantes principales et modèle de Lorenz 1963.

4. Introduction au TP : composantes principales & modèle de Lorenz 1963.

ACP à partir d'une simulation ou d'observations

$$\mathbf{X} = [\mathbf{x}_1 \quad \mathbf{x}_2 \quad \mathbf{x}_3 \quad \mathbf{x}_4 \quad \dots \quad \mathbf{x}_N]$$



- 1- Diagonalisation de $\mathbf{C}_N = \frac{1}{N-1} \mathbf{X} \mathbf{X}^T$ pour construire les composantes principales
- 2- Calcul des amplitudes du champ le long des premières composantes principales
- 3- Représentation de cette dynamique dans l'espace réduit

Conclusion

L'analyse en composante principale est un outil d'**algèbre linéaire** simple à mettre en oeuvre.

Le formalisme qui repose sur l'**optimisation sous contrainte d'égalité** (Lagrangien, multiplicateur de Lagrange).

Cet outil permet de **simplifier la représentation** d'un phénomène complexe (champ de la mécanique des fluides, coeur humain,..)

On peut en déduire une **dynamique réduite** qui permet de **diminuer le coût** d'une **prévision** et **faciliter l'assimilation des données** en recherchant la solution dans un petit sous-espace.

Quelques questions à traiter :

Réduction de modèle pour quelques EDP simples

Étude de la convergence vers la solution théorique avec le nombre de composantes sélectionnées

Assimilation de données dans le modèle réduit, comparaison avec celle dans le modèle complet

Déduction de la dynamique dans l'espace réduit à partir d'une séquence d'observations



Merci

