CIMPA Research School : Data Science for Engineering and Technology
Tunis 2019

Bregman divergences
a basic tool for pseudo-metrics building for data structured by physics

5- Proper Orthogonal Decomposition (POD) with Bregman divergences

Stéphane ANDRIEUX

ONERA - France

Member of the National Academy of Technologies of France

What is POD , what for ?

Basically : designing a sparse representation of a function $f(x,t)$

$$f(x,t) \approx \tilde{f}(x,t) = \sum_{i=1,N} b_i(t)\Phi_i(t)$$

Modes

An infinity de choices !

But, the POD approach rely on :

The choice of the order of variable

The choice of a product of spaces HxV

H Hilbert space with scalar product $\langle .,. \rangle$, V vector space with a mean .

The demand that the modes Φ are orthogonal

The meaning of "best" representation:

mean (in V) of the norm (in H) of the residual $\left\| f - \tilde{f} \right\|_H^2$

Applications

Compression of data (vectors), representation of fields, identification of structures in vector fields

Construction of reduced models by (spatial) projection of the initial PDE onto $span\ \{\Phi_i\}$

Equivalent approaches to the POD

Consider the case N=1 (one mode POD)

$$\min_{\substack{\Psi \in H \\ \langle \Psi, \Psi \rangle = 1}} \min_{a \in V} \left(\left\| a\Psi - f \right\|^2 \right) \quad \Leftrightarrow \quad \Phi = \arg\max_{\Psi \in H} \frac{\left(\langle \Psi, f \rangle^2 \right)}{\langle \Psi, \Psi \rangle}$$

a) Suppose $\Psi=\Phi$ is fixed in the initial formulation

$$\min_{a \in V} \left(\left\| a\Psi - f \right\|^2 \right) \Leftrightarrow \left(\langle b\Psi - f, \Psi \rangle \delta a \right) = 0 \ \forall \delta a \in V \longrightarrow \boxed{b = \langle f, \Psi \rangle}$$

b) Using this result the initial formulation is : $\min_{\substack{\Psi \in H \\ \langle \Psi, \Psi \rangle = 1}} \left(\left\| \langle \Psi, f \rangle \Psi - f \right\|^2 \right)$

$$\begin{aligned} \left(\left\| \langle \Psi, f \rangle \Psi - f \right\|^2 \right) &= \left(\langle \langle \Psi, f \rangle \Psi - f, \langle \Psi, f \rangle \Psi - f \rangle \right) \\ &= \left(\langle f, f \rangle + \langle \Psi, f \rangle^2 \langle \Psi, \Psi \rangle - 2 \langle \Psi, f \rangle \langle \Psi, f \rangle \right) \\ &= -\left(\langle \Psi, f \rangle^2 \right) + cste \end{aligned}$$

One mode POD

Consider the case N=1 (one mode POD) $\tilde{f} = \langle f, \Phi \rangle \Phi$, $\Phi = \arg\max G$, $\langle \Phi, \Phi \rangle = 1$

$$G(\Psi) = \left[\langle \Psi, f \rangle^2 \right]$$

Stationarity conditions of the Lagrangian $L(\Psi, \mu) = \left[\langle \Psi, f \rangle^2 \right] - \mu\left(\langle \Psi, \Psi \rangle - 1 \right)$

$$\begin{cases} \left[\langle f, \Phi \rangle \langle f, \delta\Psi \rangle \right] - \lambda \langle \Phi, \delta\Psi \rangle = 0 \quad \forall \delta\Psi \in H \\ \langle \Phi, \Phi \rangle = 1 \end{cases}$$

Leads to the definition of the operator

$$\boxed{\begin{array}{l} H \to H \\ \Psi \mapsto A\Psi = \langle f, \Psi \rangle f \end{array}}$$

$$\left. \begin{array}{l} A\Phi = \lambda\Phi \quad \langle \Phi, \Phi \rangle = 1 \\ G(\Phi) = \langle A\Phi, \Phi \rangle = \lambda \end{array} \right\}$$

Φ is the eigenvector of A with greatest eigenvalue

$$\left[\| \tilde{f} - f \|^2 \right] = \left[\langle f, f \rangle^2 \right] - \lambda$$

Examples of A operators

H space of square integrable scalar fields on a domain Ω : $\langle u,v\rangle = \int_\Omega u(x)v(x)d\Omega$

T time domain with averaging operator $\overline{g} = \frac{1}{T}\int_T g(t)dt$

$$A\Phi = \int_\Omega R(x,y)\Phi(y)dy = \lambda\Phi(x) \text{ with } R(x,y) = \frac{1}{T}\int_T u(x,t)u(y,t)dt$$

R is the (space) correlation function.

H space of square integrable scalar functions on a time domain D $\langle u,v\rangle = \int_D u(t)v(t)dt$

T space of integrable functions on a space domain Ω, spatial averaging operator $\overline{g} = \int_\Omega g(x)d\Omega$

$$A\Phi = \int_D R(t,t')\Phi(t)dt' = \lambda\Phi(t) \text{ with } R(t,t') = \int_\Omega u(x,t)u(x,t')dx$$

R is the (time) correlation function.

H space of square integrable scalar fields on a domain Ω : $\langle u,v\rangle = \int_\Omega u(x)v(x)d\Omega$

E probabilistic space, expectation operator associated to the probability measure dp $\overline{g} = \int g dp$

$$A\Phi = \int_\Omega R(x,y)\Phi(y)dy = \lambda\Phi(x) \text{ with } R(x,y) = \int u(x,p)u(y,p)dp$$

R is the (space) correlation function

Main results for N modes POD

For N modes POD, same derivation except the supplementary condition of orthogonality of modes

$$\tilde{f} = \sum_{i=1}^{N} \langle f, \Phi_i \rangle \Phi_i$$

$$A\Phi_i = \lambda_i \Phi_i \quad , \quad \langle \Phi_i, \Phi_j \rangle = \delta_{ij}$$

λ_i first greatest eigenvalues

$$\left\langle \| f - \tilde{f}_N \|^2 \right\rangle = \sum_{i=N+1}^{\infty} \lambda_i$$

Existence of the eigensystem is guaranteed by the spectral theory of Hilbert-Schmidt operators (as A is HS)

The kernels of A are the correlation operators, they benefit from the decomposition

$$R(x, x') = \sum_{i=1}^{\infty} \lambda_i \Phi_i(x) \Phi_i(x')$$

The r-decomposition is exact if the operator's spectrum is zero beyond the rank r

$$\left\langle \| f - \tilde{f}_r \|^2 \right\rangle = 0$$

SVD : an alternative way for the POD in finite dimension

Suppose we have n snapshots of m dimension vectors [f]

$$F = \begin{bmatrix} f_1^1 & f_2^1 & . & f_m^1 \\ f_1^2 & f_2^2 & . & . \\ . & . & . & . \\ f_1^n & f_2^n & . & f_m^n \end{bmatrix}$$

nxm Matrix of snapshots

One mode POD of [f]

$$\max_{\substack{\varphi \in IR^m \\ \sum_{\gamma=1}^{m} \varphi^\gamma \varphi^\gamma = 1}} G(\varphi) = \langle \varphi, f \rangle^2 = \sum_{j=1}^{n} \left(\sum_{\alpha=1}^{m} f_j^\alpha \varphi^\alpha \right)^2$$

$$\Updownarrow$$

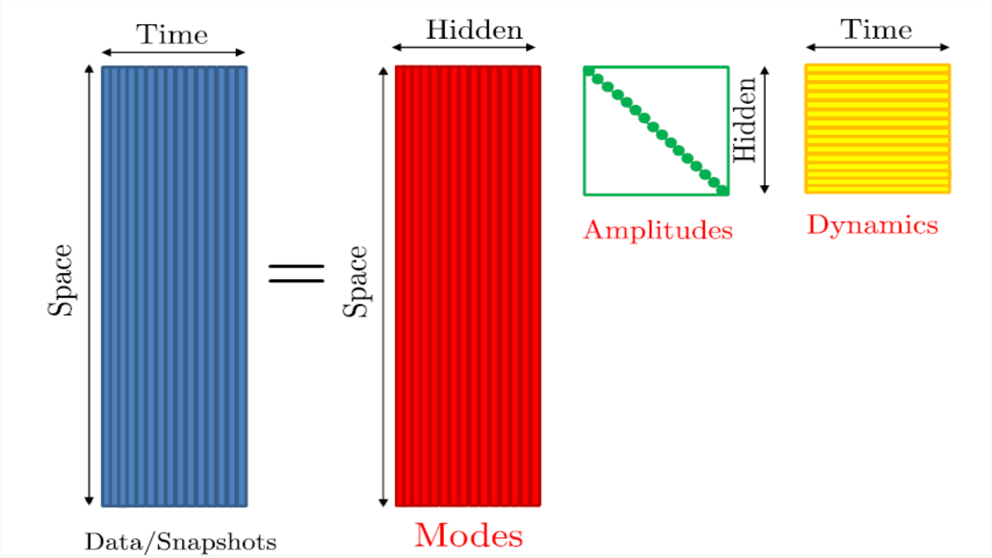
$$[M]\varphi = \lambda\varphi \qquad M^{\alpha\beta} = \sum_{j=1}^{m} f_j^\alpha f_j^\beta \quad or \quad [M] = [F][F]^t$$

SVD of matrix [F]= $U\Sigma V^t$,

$$M = [F][F]^t = \left(U\Sigma V^t\right)\left(V\Sigma U^t\right) = U\Sigma V^t V\Sigma U^t = U\Sigma^2 U^t \Rightarrow \boxed{\lambda_i = \sigma_i^2 \ , \ [\Phi_i] = [U_i]}$$

Eckart-Young theorem $\min_{X\, nxm,\, rank\, X \leq k} \|A - X\|_F \ , \ A nxm$

SVD in practice



Interpretation of SVD on snapshots for spaces modes

Compression of storage for a image $n_x x n_y$ pixels when the first k modes are retained

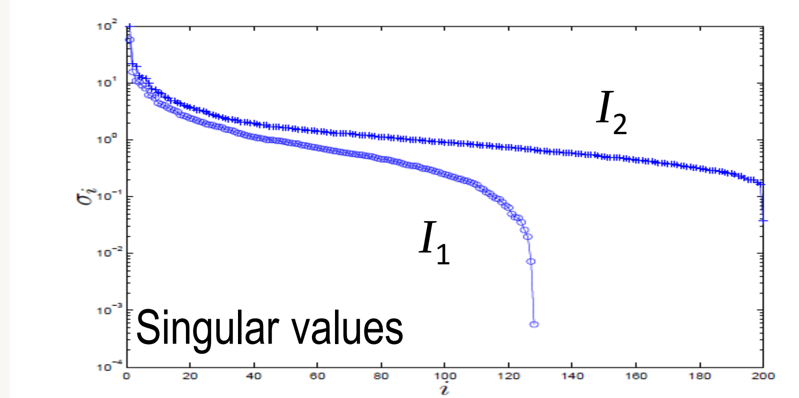
$$C_k = k\frac{1+n_x+n_y}{n_x n_y} \approx k\frac{n_x+n_y}{n_x n_y} \quad C_k \approx \frac{2k}{n_x}$$

Quality factor
(for $N\ max$ modes computed)

$$E(N) = 1 - \left(\sum_{i=N+1}^{N\max} \lambda_i\right) / \left(\sum_{i=1}^{N\max} \lambda_i\right)$$

Example 1: Image compression

Two images I_1 and I_2, described by a matrix M of "gray level $h(i,j)$ at pixel (i,j)" respectively 200x300 and 128




Original image I_1


Rank 6 representation


Original image I_2


Rank 6 representation


Rank 12 representation


Rank 20 representation

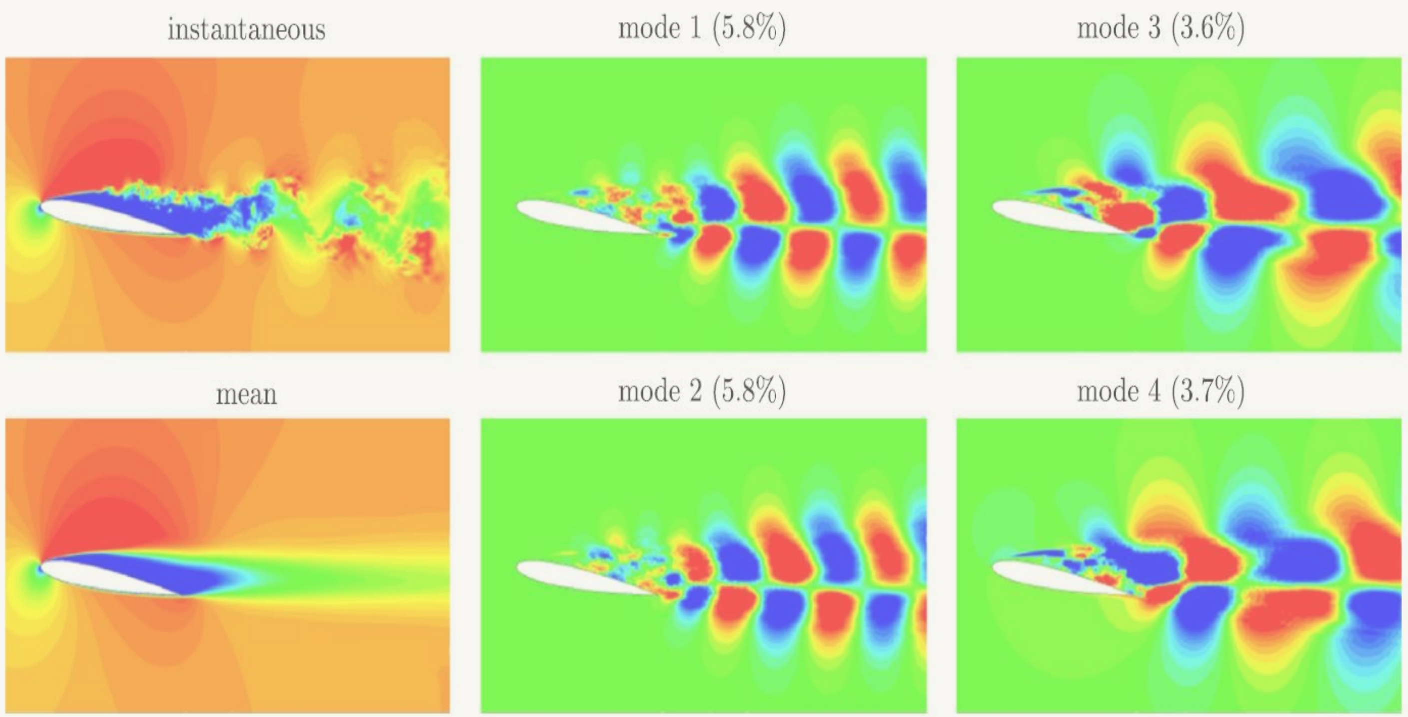

Rank 12 representation


Rank 20 representation

Example 2: Patterns recognition

Because of an energetic interpretation of the scalar product in H

Identification of energy dominant modes in a flow around a airfoil



I. - Bregman POD for quadratic generating functions

If J is quadratic $\quad D_J(e_1,e_2) = J(e_1) - J(e_2) - \left\langle \nabla J(e_2), e_1 - e_2 \right\rangle$

$$D_J(e_1,e_2) = J(e_1 - e_2) \qquad D_J(e_1,e_2) = \left\langle Qe_1, e_2 \right\rangle = \left\langle qe_1, qe_2 \right\rangle \quad q = Q^{1/2}$$

D_J defines a norm and a scalar product (Mahalanobis distance in IR^n)

$$\|h\|_Q^2 = D_Q(h,0) = Q(h) \quad , \quad \langle h, g \rangle_Q = \frac{1}{2}\left(\|h+g\|_Q^2 - \|h\|_Q^2 - \|q\|_Q^2\right) = \frac{1}{2}\left(Q(h+g) - Q(h) - Q(g)\right)$$

POD with Bregman divergence generated by quadratic function J_Q

$$\tilde{f}_Q^{\,N} = \sum_{i=1}^{N} \langle f, \Phi_i \rangle_Q \, \Phi_i$$

N first orthonormal eigenmodes and eigenvalues of self-adjoint compact operator A

$$\Box \left\| f - \tilde{f}_N^{\,J} \right\|_Q^2 \Box = \sum_{i=N+1}^{\infty} \lambda_i$$

$$A : H \to H$$

$$\Psi \mapsto A\Psi = \Box \langle f, \Psi \rangle_Q \, f \Box$$

II. - POD for μ-similar Bregman divergence

μ-similar Bregman divergences

A Bregman divergence D_J on a domain $K \subset \mathrm{IR}^n$ is μ-similar for some $\mu > 0$ if there exists a nxn positive definite matrix Q such that, for each pair (e_1, e_2) belonging to K^2

$$\mu D_Q(e_1,e_2) \le D_J(e_1,e_2) \le D_Q(e_1,e_2) \quad \forall (e_1,e_2) \in K^2$$

$$\mu Q(e_1-e_2) \le D_J(e_1,e_2) \le Q(e_1-e_2)$$

$$\|h\|_Q^2 = D_Q(h,0)$$

$$\downarrow$$

$$\mu \sum_{i=N+1}^{\infty} \lambda_i \le \left\langle D_J\left(f - \tilde{f}_N^J, 0\right)\right\rangle \le \sum_{i=N+1}^{\infty} \lambda_i$$

POD with μ-similar Bregman divergence

$$\tilde{f}_Q^N = \sum_{i=1}^{N} \langle f, \Phi_i \rangle_Q \Phi_i$$

$$\left\langle \left\| f - \tilde{f}_N^J \right\|_Q^2 \right\rangle = \sum_{i=N+1}^{\infty} \lambda_i$$

N first orthonormal eigenmodes and eigenvalues of self-adjoint compact operator A

$$A : H \to H$$
$$\Psi \mapsto A\Psi = \left\langle \langle f, \Psi \rangle_Q f \right\rangle$$

III. - POD with general Bregman divergence

POD strongly rely on scalar products and norms at various steps of derivation

1- Define a pseudo-norm and pseudo-scalar product form Bregman divergence

Define a pseudo-norm $\|h\|_J^2 = D_J(h,0)$

Use a symmetric polarization formula $\langle h,g\rangle_Q = \frac{1}{2}\left(\|h+g\|^2 - \|h\|^2 - \|q\|^2\right)$

2- Revisit all the steps in order to decide when substituting the scalar products and norms with corresponding Bregman induced pseudo-*

BD do not enjoy the triangle inequality

No hope that we have equivalence between the two formulations of POD

$$\min_{\substack{\Psi \in H \\ \langle \Psi,\Psi\rangle = 1}} \min_{a \in V} \left(\|a\Psi - f\|_J^2 \right) \qquad \Phi = \arg\max_{\Psi \in H} \frac{\left(\langle \Psi, f\rangle_J^2 \right)}{\langle \Psi,\Psi\rangle}$$

IV. - POD with general Bregman divergence

POD strongly rely on scalar products and norms at various steps of derivation

1- Define a pseudo-norm and pseudo-scalar product form Bregman divergence

Divergence	Squared norm $\|\| . \|\|_J^2$	Scalar product $\langle .,. \rangle_{D_J}$
Extended Bregman divergence	$D_J(e,0) = J(e)$	$\langle e, f \rangle_{D_J} = \frac{1}{2}\left(J(e+f,0) - D(e,0) - D(f,0)\right)$
Symmetric Bregman divergence	$\cancel{D_J^s(e,0) = \langle \nabla J(e), e \rangle}$	$\langle e, f \rangle_{D_J^s} = \frac{1}{2}\left(\langle \nabla J(e+f), e+f \rangle - \langle \nabla J(e), e \rangle - \langle \nabla J(f), f \rangle\right)$

2- We <u>choose</u>

To keep the orthogonality of the modes Φ_i in H with its own scalar product

$$\langle \Phi_i, \Phi_j \rangle = \delta_{ij}$$

To adopt as objective of the POD

$$\tilde{f}_J = \langle f, \Phi \rangle \Phi \quad , \quad \Phi = \underset{\substack{\Psi \in H \\ \langle \Psi, \Psi \rangle = 1}}{\arg\min} \left[\| \langle f, \Psi \rangle \Psi - f \|_J^2 \right] = G(\Psi)$$

One mode Bregman POD

V. - POD with general Bregman divergence

Bregman divergence orthogonal decomposition - BDOD

The BDOD *of order* N $\tilde{f}_N^J(x)$ *of a function* $f(x,t)$ *defined on the product* $H\mathrm{x}E$ *where* H *is a Hilbert space of functions on a spatial domain* Ω *and* E *a time domain, with* $\langle .,. \rangle$ *the scalar product of* H *and* $\cdot$ *the time averaging function on* E*, is*

$$\tilde{f}_J^N = \sum_{i=1}^{N} \langle f, \Phi_i \rangle \Phi_i$$

where the functions Φ_i *are sequentially determined by the minimization problem :*

$$\Phi_i = \underset{\substack{\Psi \in H, \langle \Psi, \Psi \rangle = 1 \\ \langle \Psi, \Phi_j \rangle = 0 \; j = 1, i-1}}{\arg\min} \left(J\left[\sum_{j=1}^{i-1} \langle f, \Phi_j \rangle \Phi_j + \langle f, \Psi \rangle \Psi - f \right] \right)$$

and J *is the convex generating function of the Bregman divergence* D_J

V. - POD with general Bregman divergence

Characterization of Bregman divergence orthogonal decomposition

Using an appropriate Lagrangian

One mode Bregman POD

$$\begin{array}{l}\left\langle \nabla J\left[\langle f,\Phi\rangle\Phi - f\right], \langle f,\delta\Psi\rangle\Phi + \langle f,\Phi\rangle\delta\Psi\right\rangle = \lambda\langle\Phi,\delta\Psi\rangle \quad \forall\,\delta\Psi\in H \\ \langle\Phi,\Phi\rangle = 1\end{array}$$

A form of eigenproblem

Eliminating the Lagrange multiplier

$$2\langle f,\Phi\rangle\langle\nabla J_{\Phi},\Phi\rangle = \lambda \qquad \begin{cases} \langle\Phi, f\rangle\left\langle \nabla J\left(\langle\Phi, f\rangle\Phi - f\right), \langle 2f, \langle\Phi, f\rangle\Phi - f\rangle\Phi + f\right\rangle = 0 \\ \langle\Phi,\Phi\rangle = 1\end{cases}$$

More work to be done !

1- Repeat sequentially for N-decomposition : Better algorithm ?

2- POD on Product space for multiphysics applications

Thanks for your attention