Comput. Methods Appl. Mech. Engrg. 199 (2010) 1603–1626

Contents lists available at ScienceDirect

Comput. Methods Appl. Mech. Engrg.

journal homepage: www.elsevier.com/locate/cma

A priori model reduction through Proper Generalized Decomposition for solving time-dependent partial differential equations

Anthony Nouy *

GeM – Research Institute in Civil Engineering and Mechanics, UMR CNRS 6183, Ecole Centrale Nantes, University of Nantes, 1 rue de la Noë, 44321 Nantes Cedex 3, France

ARTICLE INFO

Article history:
Received 8 October 2009
Received in revised form 11 January 2010
Accepted 20 January 2010
Available online 28 January 2010

Keywords:
Model reduction
Evolution problems
Proper Orthogonal Decomposition (POD)
Proper Generalized Decomposition (PGD)
Generalized Spectral Decomposition
Separation of variables

ABSTRACT

Over the past years, model reduction techniques have become a necessary path for the reduction of computational requirements in the numerical simulation of complex models. A family of *a priori* model reduction techniques, called Proper Generalized Decomposition (PGD) methods, are receiving a growing interest. These methods rely on the *a priori* construction of separated variables representations of the solution of models defined in tensor product spaces. They can be interpreted as generalizations of Proper Orthogonal Decomposition (POD) for the *a priori* construction of such separated representations. In this paper, we introduce and study different definitions of PGD for the solution of time-dependent partial differential equations. We review classical definitions of PGD based on Galerkin or Minimal Residual formulations and we propose and discuss several improvements for these classical definitions. We give an interpretation of optimal decompositions as the solution of pseudo-eigenproblems. We also introduce a new definition of PGD, called Minimax PGD, which can be interpreted as a Petrov–Galerkin model reduction technique, where test and trial reduced basis functions are related by an adjoint problem. This new definition improves convergence properties of separated representations with respect to a chosen metric. It coincides with a classical POD for degenerate time-dependent partial differential equations. For the numerical construction of each PGD, we propose algorithms inspired from the solution of eigenproblems. Several numerical examples illustrate and compare the different definitions of PGD on transient advection–diffusion–reaction equations.

1. Introduction

The numerical simulation of physical models takes today an important place in numerous branches of science and engineering. Due to the increasing complexity of models, more and more refined discretizations and robust numerical solution techniques are needed in order to obtain reliable predictions of their responses. Furthermore, in the context of optimization, model identification, or parametric and stochastic analyses, the aim is not to predict the response of a unique model but of a family of models. In order to achieve these analyses, traditional solution techniques require the optimal use of constantly evolving computational resources. However, for many applications, innovative methodologies, alternative to the brute force approach, are obviously necessary to access numerical predictions.

The concept of model reduction seems to be a path for solving these computational issues. Model reduction methods exploit the fact that the response of complex models (or of a family of models) can often be approximated with a reasonable precision by the response of a surrogate model, which is the projection of the initial model on a low dimensional reduced basis. The dimension of reduced bases may be of several orders of magnitude lower than the dimension of the classically used numerical models. Model reduction methods distinguish themselves by the way of defining and constructing the reduced bases of functions. Among these methods, model reduction methods based on separation of variables are receiving a growing interest in various fields of scientific computing. In the context of the solution of evolution problems, a separated representation of the solution $u(x,t)$ defined on a space–time domain consists of a sum of products of scalar functions of the time variable by functions of the space variable:

$$u(x,t) \approx u_m(x,t) = \sum_{i=1}^{m} w_i(x)\lambda_i(t). \tag{1}$$

When the solution u is known (or at least an approximation of it), an optimal order m (or rank m) separated representation (1) — also known as tensor product approximation or finite sums decomposition — can be classically defined as the one which minimizes the distance to the solution with respect to a particular norm. This separated representation is optimal in the sense that it minimizes this distance for a given order m of decomposition. Under some

* Tel.: +33 2 40 37 16 76; fax: +33 2 40 37 25 73.
E-mail address: anthony.nouy@ec-nantes.fr

0045-7825/$ - see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.cma.2010.01.009

assumptions on the chosen norm, this is the basic definition of the classical Proper Orthogonal Decomposition (POD), also known as Karhunen–Loève decomposition [18,26], Singular Value Decomposition or Principal Component Analysis in other contexts. This decomposition is classically used as an *a posteriori* model reduction technique for long-time simulations or parametric analyses of evolution problems [5,4,19,20,16,36,24,10,15,3].

In this paper, we focus on the more challenging problem of the *a priori* construction of such separated representations. The aim is to introduce different strategies for the construction of an approximate separated representation (1) of the solution, without *a priori* knowing the solution nor an approximation of it. This requires to adopt another definition of the separated representation and then to propose dedicated algorithms for its construction. The different methods introduced in this article can be interpreted as generalizations of the POD for the *a priori* construction of separated representations. The resulting decompositions have been recently called *Proper Generalized Decompositions* (PGD).[1] This type of methods has been first introduced by Ladevèze in the context of the LATIN method [21] (LArge Time INcrement method) for reducing computational costs (memory requirements and computational times) associated with the solution of multiple linear evolution problems resulting from a nonlinear iterative strategy which is global in time. In this context, separated representation (1) was called "radial approximation". In the literature, two variants of PGD have been proposed for the progressive construction of (1), respectively based on a Galerkin formulation [21,12,22,9] or a Minimal Residual formulation [31,23] of the evolution problem. PGD methods have also been introduced in other contexts: separation of physical variables and parameters (or random variables) in the context of parametrized (or stochastic) PDEs [27,28,32,29],[2] with a possible additional separation of parameters [11,30],[3] separation of coordinate variables in multi-dimensional PDEs [6,1,2,8].[4]

In this paper, we introduce and study different definitions of PGD for the solution of time-dependent partial differential equations. We review classical definitions of PGD based on Galerkin or Minimal Residual formulations and we propose and discuss several improvements for these classical definitions. We give an interpretation of decompositions as the solution of pseudo-eigenproblems and propose algorithms inspired from the solution of eigenproblems for the construction of these decompositions. We also introduce an innovative definition, called Minimax PGD, which allows us to improve convergence properties of decomposition (1) with respect to a chosen metric. This new PGD can be interpreted as an *a priori* Petrov–Galerkin model reduction technique, where test and trial reduced basis functions are related by an adjoint problem involving the chosen metric. For degenerate time-dependent partial differential equations, this new definition coincides with a classical POD with respect to the chosen metric.

The outline of the paper is as follows: In Section 2, we introduce an abstract weak formulation of a class of time-dependent partial differential equations. In Section 3, we recall the principles of model reduction methods with a particular focus on *a posteriori* model reduction methods based on the POD. In Section 4, we focus on PGD methods. We present classical progressive definitions of these decompositions, based on Galerkin or minimal residual formulations. We give an interpretation of these decompositions as the solutions of pseudo-eigenproblems and we propose possible improvements. In Section 5, we introduce and analyze a non-classical definition of PGD, called Minimax PGD. In Section 6, several numerical examples illustrate the behavior of the Proper Generalized Decomposition methods introduced in this article.

2. Time-dependent partial differential equation and discretization

2.1. Model problem: advection–diffusion–reaction equation

As a problem model, we consider a transient advection–diffusion–reaction equation defined on a spatial domain $\Omega \subset \mathbb{R}^d$ and a time interval $I = (0,T)$. The solution $u(x,t)$, with $(x,t) \in \Omega \times I$, solves:

$$\dot{u} - \nabla \cdot (\mu \nabla u) + c \cdot \nabla u + \sigma u = f \quad \text{on } \Omega \times I, \tag{2a}$$

$$u = 0 \quad \text{on } \partial\Omega \times I, \tag{2b}$$

$$u = u_0 \quad \text{on } \Omega \times \{0\}, \tag{2c}$$

where $\dot{u} = \frac{\partial u}{\partial t}$, $u_0(x)$ is the initial condition, $f(x,t)$ is a volumic source, and $\mu(x,t)$, $c(x,t)$ and $\sigma(x,t)$ are diffusion, advection and reaction parameters which are eventually space and time-dependent.

2.2. Space weak formulation

We identify u with a function defined on I with values in Hilbert space $\mathcal{V} = H_0^1(\Omega)$, with $u(t) : x \in \Omega \mapsto u(t)(x) \simeq u(x,t)$. A weak formulation of (2c) writes: find $u : I \to \mathcal{V}$ such that

$$m(\dot{u}(t), v) + a(u(t), v; t) = \ell(v; t) \quad \forall v \in \mathcal{V}, \tag{3a}$$

$$u(0) = u_0, \tag{3b}$$

where $m(\cdot,\cdot)$ and $a(\cdot,\cdot;t)$ are bilinear forms on $\mathcal{V}$ and where $\ell(\cdot;t)$ is a linear form on $\mathcal{V}$, defined by:

$$m(u,v) = \int_\Omega uv\,dx = \langle u, v\rangle_{L^2(\Omega)}, \quad \ell(v;t) = \int_\Omega f(t)v\,dx, \tag{4}$$

$$a(u,v;t) = \int_\Omega \mu(t)\nabla u \cdot \nabla v\,dx + \int_\Omega c(t)\cdot\nabla u v\,dx + \int_\Omega \sigma(t) u v\,dx. \tag{5}$$

2.3. Space–time weak formulation

A space–time weak formulation of (2c) is now introduced [25]. We introduce the following function space

$$L^2(I;\mathcal{V}) = \{v : I \to \mathcal{V}; \int_I \|v(t)\|_{\mathcal{V}}^2\,dt < +\infty\}, \tag{6}$$

where $\|\cdot\|_{\mathcal{V}}$ is a norm on $\mathcal{V}$. We denote $\mathcal{T} = L^2(I;\mathbb{R}) := L^2(I)$ and identify the space $L^2(I;\mathcal{V})$ with the tensor product space $\mathcal{V}\otimes\mathcal{T}$. We denote by $\mathcal{V}' = H^{-1}(\Omega)$ the dual space of $\mathcal{V}$. A weak solution of problem (3b) can then be defined by the following problem: find $u \in \mathcal{V}\otimes\mathcal{T}$ such that $\dot{u} \in L^2(I;\mathcal{V}')$ and[5]

$$B(u,v) = L(v) \quad \forall v \in \mathcal{V}\otimes\mathcal{T}, \tag{7}$$

where B and L are bilinear and linear forms defined by

$$B(u,v) = \int_I m(\dot{u}(t), v(t))dt + \int_I a(u(t), v(t); t)dt + m(u(0^+), v(0^+)), \tag{8}$$

$$L(v) = \int_I \ell(v(t);t)dt + m(u_0, v(0^+)) \tag{9}$$

with $v(0^+) = \lim_{s\downarrow 0} v(s)$. The solution of problem (7) verifies the initial condition in a weak sense.

[1] Roughly speaking, PGD methods introduce different definitions of the separated representation (1) which require only the operator and right-hand side of the PDE, and not the solution itself as in the definition of the POD. With dedicated algorithms, it then allows to build the separated representation without knowing the solution a priori.

[2] $u(x,t,\xi) \approx \sum_{i=1}^m w_i(x,t)\lambda_i(\xi)$, with ξ the (random) parameters. In this context, PGD has been named Generalized Spectral Decomposition as a generalization of spectral decomposition of random processes.

[3] $u(x,t,\xi) \simeq u(x,t,\xi_1,\ldots,\xi_d) \approx \sum_{i=1}^m w_i(x,t)\lambda_i^1(\xi_1)\cdots\lambda_i^d(\xi_d)$.

[4] $u(x_1,\ldots,x_d) \approx \sum_{i=1}^m w_i^1(x_1)\cdots w_i^d(\xi_d)$.

[5] Let us note that $\dot{u}(t) \in \mathcal{V}'$ is assimilated with its Riesz representation in $\mathcal{V}$ in the notation $m(\dot{u}(t), v(t))$.

2.4. Time discontinuous Galerkin approximation

2.4.1. Piecewise polynomial approximation

We consider a time discretization using time discontinuous Galerkin framework [14,38,37]. Let $\mathcal{I} = \{I_k = (t_{k-1}, t_k)\}_{k=1}^r$ denote a partition of $I = (0,T)$, with $t_0 = 0 < t_1 < \cdots < t_r = T$. We denote by $\mathbb{P}^p(I_k;X) = \{v : I_k \to X;\ v(t) = \sum_{i=0}^p v_i t^i,\ v_i \in X\}$ the set of polynomial functions of degree p defined on I_k with values in X, and by $\mathbb{P}^p(\mathcal{I};X)$ the set of piecewise polynomial functions of degree p in time associated with partition $\mathcal{I}$:

$$\mathbb{P}^p(\mathcal{I};X) = \{v : I \to X;\ \forall I_k \in \mathcal{I},\ v_{|I_k} \in \mathbb{P}^p(I_k;X)\}.$$

2.4.2. Definition of the approximation

The time discontinuous Galerkin approximation is sought in the semi-discrete approximation space $\mathbb{P}^p(\mathcal{I};\mathcal{V}) \subset L^2(I;\mathcal{V})$. This approximation space is identified with the tensor product space $\mathcal{V} \otimes \mathcal{T}_P$, with $\mathcal{T}_P := \mathbb{P}^p(\mathcal{I};\mathbb{R}) \subset \mathcal{T}$, where P denote the dimension of $\mathcal{T}_P$. The approximation of problem (7) is then defined by

$$u \in \mathcal{V} \otimes \mathcal{T}_P, \quad B(u,v) = L(v) \quad \forall v \in \mathcal{V} \otimes \mathcal{T}_P, \tag{10}$$

where in the definition (8) of bilinear form B, the time derivative is interpreted as follows:

$$\int_I m(\dot{u}(t), v(t))dt = \sum_{k=1}^r \int_{I_k} m(\dot{u}(t), v(t))dt + \sum_{k=1}^{r-1} m\big(u(t_k^+) - u(t_k^-), v(t_k^+)\big), \tag{11}$$

where $v(t^\pm) = \lim_{s|0} v(t \pm s)$. In practice, problem (10) is separated into r problems, which are local on each time interval I_k and can be solved sequentially. For $k \in \{1,\ldots,r\}$, the restriction $u_{|I_k} \in \mathbb{P}^p(I_k;\mathcal{V})$ is defined by

$$\begin{aligned}&\int_{I_k} m(\dot{u}(t), v(t))dt + \int_{I_k} a(u(t), v(t);t)\,dt \\ &\quad + m\big(u(t_{k-1}^+), v(t_{k-1}^+)\big) \\ &\quad = \int_{I_k} \ell(v(t);t)dt + m(u(t_{k-1}^-), v(t_{k-1}^+)) \quad \forall v \in \mathbb{P}^p(I_k;\mathcal{V}),\end{aligned} \tag{12}$$

where by convention $u(t_0^-) = u_0$. The solution $u_{|I_k} \in \mathbb{P}^p(I_k;\mathcal{V})$ can be interpreted as a time continuous Galerkin approximation (using polynomial approximation in time) of an evolution problem defined on time interval I_k, with a weakly imposed initial condition $u(t_{k-1}^-)$ which is obtained from the solution on the previous time interval.

Remark 1. A time continuous Galerkin approximation would be also defined by problem (10), by choosing for $\mathcal{T}_P$ the space of continuous piecewise polynomial functions in time, *i.e.*, $\mathcal{T}_P = \mathbb{P}^p(\mathcal{I};\mathbb{R}) \cap C^0(I;\mathbb{R})$. This approximation is not used in practice since it requires the solution of a global problem in time. However, it could be used within the context of *a priori* model reduction techniques introduced in the following sections.

Remark 2. Let us note that classical time integration schemes lead to the solution of a problem which can be also recasted under the form (10), by identifying $\mathcal{T}_P \simeq \mathbb{R}^P$, where P is the number of time steps $\{t_k\}_{k=1}^P$. The approximate solution $u = \{u(t_k)\}_{k=1}^P \in (\mathcal{V})^P$, which gathers the solution at the different time steps, can be identified with a function in $\mathcal{V} \otimes \mathcal{T}_P$.

2.5. Spatial approximation

At spatial level, we consider a classical Galerkin approximation (finite element or spectral finite element) by introducing an approximation space $\mathcal{V}_N \subset \mathcal{V}$, with $\mathcal{V}_N = \left\{v = \sum_{i=1}^N \varphi_i v_i; \varphi_i \in \mathcal{V}, v_i \in \mathbb{R}\right\}$. The fully discretized Galerkin approximation is then defined by

$$u \in \mathcal{V}_N \otimes \mathcal{T}_P, \quad B(u,v) = L(v) \quad \forall v \in \mathcal{V}_N \otimes \mathcal{T}_P. \tag{13}$$

We identify $v \in \mathcal{V}_N$ with a vector $\mathbf{v} = (v_1,\ldots,v_N)^T \in \mathbb{R}^N$, such that problem (13) is equivalently written: find $\mathbf{u} \in \mathbb{R}^N \otimes \mathcal{T}_P \simeq \mathbb{P}^p(\mathcal{I};\mathbb{R}^N)$ such that

$$\begin{aligned}&\int_I \mathbf{v}(t)^T \mathbf{M}\dot{\mathbf{u}}(t)dt + \int_I \mathbf{v}(t)^T \mathbf{A}(t)\mathbf{u}(t)dt + \mathbf{v}(0^+)^T \mathbf{M}\mathbf{u}(0^+) \\ &\quad = \int_I \mathbf{v}(t)^T \mathbf{f}(t)dt + \mathbf{v}(0^+)^T \mathbf{M}\mathbf{u}_0, \quad \forall \mathbf{v} \in \mathbb{R}^N \otimes \mathcal{T}_P,\end{aligned} \tag{14}$$

where $\mathbf{M}$ and $\mathbf{A}(t)$ are matrices whose components are defined by $(\mathbf{M})_{ij} = m(\varphi_j, \varphi_i)$ and $(\mathbf{A}(t))_{ij} = a(\varphi_j, \varphi_i; t)$, where $\mathbf{f}(t)$ is a vector whose components are defined by $(\mathbf{f}(t))_i = \ell(\varphi_i;t)$, and where $\mathbf{u}_0 \simeq u_0$. In Eq. (14), the time derivative must be interpreted as in (11) for time discontinuous functions.

3. Proper Orthogonal Decomposition and *a posteriori* model reduction techniques

3.1. Reduced order models

Classical model reduction techniques consist in seeking an approximation of the solution u under the form

$$u_m(x,t) = \sum_{i=1}^m w_i(x)\lambda_i(t), \tag{15}$$

where the $w_i \in \mathcal{V}$ and the $\lambda_i \in \mathcal{T}$ form low dimensional reduced bases of spatial functions and time functions, respectively.

3.1.1. Galerkin projection on a reduced basis of spatial functions

Classical model reduction techniques start with the construction of a low dimensional subspace $\mathcal{V}_m = span\{w_i\}_{i=1}^m \subset \mathcal{V}$ and define the Galerkin approximation $u_m \in \mathcal{V}_m \otimes \mathcal{T}$ by

$$B(u_m, v_m) = L(v_m) \qquad \forall v_m \in \mathcal{V}_m \otimes \mathcal{T}. \tag{16}$$

Problem (16) can be interpreted as a time weak formulation of the following system of m ordinary differential equations: for $i \in \{1,\ldots,m\}$,

$$\sum_{j=1}^m m(w_j, w_i)\dot{\lambda}_j(t) + \sum_{j=1}^m a(w_j, w_i; t)\lambda_j(t) = \ell(w_i;t), \tag{17a}$$

$$\sum_{j=1}^m m(w_j, w_i)\lambda_j(0) = m(u_0, w_i), \tag{17b}$$

where the initial condition corresponds to a projection of the initial condition $u_m(0) = u_0$ on subspace $\mathcal{V}_m$, *i.e.* $m(u_m(0) - u_0, w) = 0,\ \forall w \in \mathcal{V}_m$.

3.1.2. Galerkin projection on a reduced basis of time functions

Another point of view consists in first constructing a low dimensional subspace $\mathcal{T}_m = span\{\lambda_i\}_{i=1}^m \subset \mathcal{T}$. The Galerkin approximation $u_m \in \mathcal{V} \otimes \mathcal{T}_m$ is then defined by

$$B(u_m, v_m) = L(v_m) \quad \forall v_m \in \mathcal{V} \otimes \mathcal{T}_m. \tag{18}$$

Problem (18) can be interpreted as a system of m coupled time-independent partial differential equations:

$$\sum_{j=1}^m b_{ij}(w_j, w^*) = \ell_i(w^*) \quad \forall w^* \in \mathcal{V},\ \ i = 1 \ldots m, \tag{19}$$

where the b_{ij} and ℓ_i are bilinear and linear forms on $\mathcal{V}$, defined by

$$b_{ij}(w, w^*) = B(\lambda_j w, \lambda_i w^*), \quad \ell_i(w^*) = L(w^* \lambda_i). \tag{20}$$

In practise, this approach is not used for large scale applications since the solution of the system of coupled partial differential equations leads to prohibitive computational costs. This limitation will guide the selection of particular Proper Generalized Decomposition methods in Sections 4 and 5.

3.2. Proper Orthogonal Decomposition (POD)

The Proper Orthogonal Decomposition (POD) technique (also known as Singular Value Decomposition or Karhunen–Loève decomposition [18,26] in other contexts) consists in defining a separated representation u_m of the solution $u \in \mathcal{V} \otimes \mathcal{T}$ such that it is optimal with respect to a particular metric. An optimal separated representation of order m, denoted

$$u_m = \sum_{i=1}^{m} w_i \lambda_i, \quad w_i \in \mathcal{V}, \ \ \lambda_i \in \mathcal{T}, \tag{21}$$

is classically defined as the one which minimizes the distance to the exact solution u with respect to a given norm $\|\cdot\|$ on $\mathcal{V} \otimes \mathcal{T}$, *i.e.*

$$\|u - u_m\|^2 = \min_{\substack{\{w_i\}_{i=1}^m \in (\mathcal{V})^m \\ \{\lambda_i\}_{i=1}^m \in (\mathcal{T})^m}} \|u - \sum_{i=1}^{m} w_i \lambda_i\|^2. \tag{22}$$

Functions w_i and λ_i in the decomposition u_m can then be considered as optimal reduced basis functions with respect to the chosen norm. A classical choice consists in introducing a natural norm on $\mathcal{V} \otimes L^2(I)$, defined by

$$\|u\|^2 = \int_I \|u(t)\|_{\mathcal{V}}^2 dt, \tag{23}$$

where $\|\cdot\|_{\mathcal{V}}$ is a given norm on Hilbert space $\mathcal{V}$. Denoting by $\langle \cdot, \cdot \rangle_{\mathcal{V}}$ the inner product associated with $\|\cdot\|_{\mathcal{V}}$, the inner product on $\mathcal{V} \otimes L^2(I)$ associated with $\|\cdot\|$ is defined by

$$\langle\!\langle u, v \rangle\!\rangle = \int_I \langle u(t), v(t) \rangle_{\mathcal{V}} dt. \tag{24}$$

Inner product $\langle\!\langle \cdot, \cdot \rangle\!\rangle$ have the following separation property: $\forall \lambda, \lambda^* \in \mathcal{T}$ and $\forall w, w^* \in \mathcal{V}$,

$$\langle\!\langle w\lambda, w^*\lambda^* \rangle\!\rangle = \langle w, w^* \rangle_{\mathcal{V}} \langle \lambda, \lambda^* \rangle_{\mathcal{T}}, \tag{25}$$

where $\langle \cdot, \cdot \rangle_{\mathcal{T}} = \langle \cdot, \cdot \rangle_{L^2(I)}$ is the natural inner product in $L^2(I)$. It corresponds to a classical construction of an inner product on a tensor product space $\mathcal{V} \otimes \mathcal{T}$ by "tensorization" of inner products on $\mathcal{V}$ and $\mathcal{T}$. When the chosen inner product has the separation property (25), we classically show that problem (22) leads to the following eigenproblem:

$$G(w) = \sigma w, \tag{26}$$

where operator $G : \mathcal{V} \to \mathcal{V}$ is defined as follows: for $w, w^* \in \mathcal{V}$,

$$\langle G(w), w^* \rangle_{\mathcal{V}} = \langle \langle w^*, u \rangle_{\mathcal{V}}, \langle u, w \rangle_{\mathcal{V}} \rangle_{\mathcal{T}}. \tag{27}$$

G is called the spatial correlation operator of u associated with inner products $\langle \cdot, \cdot \rangle_{\mathcal{V}}$ and $\langle \cdot, \cdot \rangle_{\mathcal{T}}$ and sometimes called "POD operator". For the particular choice $\langle \lambda, \lambda^* \rangle_{\mathcal{T}} = \langle \lambda, \lambda^* \rangle_{L^2(I)} = \int_I \lambda(t)\lambda^*(t)dt$, G has the following form:

$$G(w) = \int_I u(t) \langle u(t), w \rangle_{\mathcal{V}} dt.$$

G is a symmetric positive linear operator. Under classical regularity assumptions on u, it is a compact operator, so that classical spectral theory applies [34]. Let $\{w_i\}_{i \geqslant 1}$ denote an orthogonal set of eigenfunctions of G, which forms an hilbertian basis of $\mathcal{V}$. The eigenpairs $(w_i, \sigma_i) \in \mathcal{V} \times \mathbb{R}^+$ of G being sorted by decreasing eigenvalues ($\sigma_1 \geqslant \sigma_2 \geqslant \cdots \geqslant 0$), an optimal separated representation u_m of order m of u can then be written as (21), with

$$\lambda_i(t) = \|w_i\|_{\mathcal{V}}^{-2} \langle w_i, u(\cdot, t) \rangle_{\mathcal{V}} \tag{28}$$

and where $\|w_i \lambda_i\| = \|w_i\|_{\mathcal{V}} \|\lambda_i\|_{\mathcal{T}} = \sqrt{\sigma_i}$. The truncation error verifies:

$$\|u - u_m\|^2 = \|u\|^2 - \sum_{i=1}^{m} \sigma_i \underset{m \to \infty}{\to} 0. \tag{29}$$

The above defined functions $w_i \in \mathcal{V}$ and $\lambda_i \in \mathcal{T}$ are orthogonal with respect to inner products $\langle \cdot, \cdot \rangle_{\mathcal{V}}$ and $\langle \cdot, \cdot \rangle_{\mathcal{T}}$, respectively. Decomposition (21) is then called a biorthogonal decomposition of u.

Remark 3. A classical choice for $\langle \cdot, \cdot \rangle_{\mathcal{V}}$ consists in using the natural inner product in $L^2(\Omega)$ or $H^1(\Omega)$ [17]. Let us note that other choices than $\langle \cdot, \cdot \rangle_{L^2(I)}$ could also be made for inner product $\langle \cdot, \cdot \rangle_{\mathcal{T}}$. For example, taking $\langle \lambda, \lambda^* \rangle_{\mathcal{T}} = \int_I \alpha(t)\lambda(t)\lambda^*(t)dt$, with $\alpha(t) > 0$, may allow to better fit the solution in some regions of the time domain. Taking for $\langle \cdot, \cdot \rangle_{\mathcal{T}}$ the natural inner product on $H^1(I)$ may lead to a better approximation of $\dot{u}$ (it is possible to work in $H^1(I) \subset \mathcal{T} = L^2(I)$ since the solution $u \in (\mathcal{V} \otimes L^2(I)) \cap (L^2(\Omega) \otimes H^1(I))$) (see [16]).

Remark 4. The optimal separated representation can be equivalently defined by formulating an eigenproblem on λ,

$$G^{\diamond}(\lambda) = \sigma \lambda, \tag{30}$$

where $G^{\diamond}$, sometimes called the "auxiliary POD operator", is defined as follows: for $\lambda, \lambda^* \in \mathcal{T}$,

$$\langle G^{\diamond}(\lambda), \lambda^* \rangle_{\mathcal{T}} = \langle \langle \lambda^*, u \rangle_{\mathcal{T}}, \langle u, \lambda \rangle_{\mathcal{T}} \rangle_{\mathcal{V}}. \tag{31}$$

Eigenproblems (26) and (30) are equivalent in the following sense (see [16] for more details): if $(w_i, \sigma_i) \in \mathcal{V} \times \mathbb{R}^+$ is an eigenpair of (26), then $(\lambda_i, \sigma_i) \in \mathcal{T} \times \mathbb{R}^+$, with λ_i defined by (28), is an eigenpair of (30). Conversely, if $(\lambda_i, \sigma_i) \in \mathcal{T} \times \mathbb{R}^+$ is an eigenpair of (30), then $(w_i, \sigma_i) \in \mathcal{V} \times \mathbb{R}^+$, with $w_i = \|\lambda_i\|_{\mathcal{T}}^{-2} \langle \lambda_i, u \rangle_{\mathcal{T}}$, is an eigenpair of (26). Let us note that for the particular choice $\langle \lambda, \lambda^* \rangle_{\mathcal{T}} = \int_I \lambda(t)\lambda^*(t)dt$, $G^{\diamond}$ has the following form:

$$G^{\diamond}(\lambda)(t) = \int_I \langle u(t), u(s) \rangle_{\mathcal{V}} \lambda(s) ds,$$

where function $k(t,s) = \langle u(t), u(s) \rangle_{\mathcal{V}}$ is the classical time correlation kernel.[6]

3.3. A posteriori model reduction using Proper Orthogonal Decomposition

Of course, the solution u (or at least an approximation of u) have to be known in order to perform the classical POD. This decomposition can then be considered as an *a posteriori* model reduction technique. However, the POD technique can be used as a model reduction technique for parametric analyses or long-time simulations [19,20,16,36,24,10,15,3].

In a parametric analysis, where operator and right-hand side of the partial differential equation depend on a set of parameters ξ, one may be interested in computing the solution for $\xi \in \Xi$, where Ξ denotes a (discrete) set of parameters values. A simple model reduction technique consists in solving the evolution problem for a subset of parameters values. For each evolution problem, a reduced basis of space (or time) functions is extracted *a posteriori* from the classical POD.[7] The reduced basis obtained for each parameter value can be added to a global reduced basis [36], used for subsequent unresolved parameters values, or can be interpolated in different ways [24,3]. With this procedure, a particular attention must be paid to the sorting and selection of reduced basis functions. Error estimation and reduced basis enrichment strategies are

[6] Note that since $k(t,s) \in L^2(I \times I)$, $G^{\diamond} : L^2(I) \to L^2(I)$ is a Hilbert–Schmidt operator.

[7] Let us also mention the "Reduced Basis Method" [35,7] as an alternative to POD for the definition of optimal reduced basis in the context of parametric or direct stochastic analysis.

necessary in order to control the precision of the reduced order approximation for the unresolved parameters values [19,20].

The POD technique can be also used in long-time simulations. An approximation of the solution is computed with traditional numerical methods on a small time interval $I' = (0, T')$, with $T' \ll T$. The application of POD allows to extract a reduced basis of spatial functions (or time functions) from the restriction of the solution to $I' \times \Omega$. An approximate solution can then be computed on the whole time interval I by solving the reduced order problem (16) (or (18)). Error estimation and reduced basis enrichment strategies are also necessary in order to control the precision of the reduced order approximation on the whole time interval.

4. A priori model reduction through Proper Generalized Decomposition

We now introduce different methods, recently called *Proper Generalized Decomposition* (PGD) methods, for the *a priori* construction of an approximate separated representation of the solution. In this section, we present different possible definitions of PGDs, based on Galerkin orthogonality criteria or minimal residual criteria. We give an interpretation of these decompositions as generalizations of Proper Orthogonal Decompositions. We also propose algorithms for the construction of the different PGDs.

In the following, we denote by $W_m = \{w_i\}_{i=1}^m \in (\mathcal{V})^m$ the set of space functions and by $\Lambda_m = \{\lambda_i\}_{i=1}^m \in (\mathcal{T})^m$ the set of time functions of the decomposition $u_m = \sum_{i=1}^m w_i\lambda_i$, denoted $u_m = W_m \cdot \Lambda_m$.

4.1. PGD based on Galerkin orthogonality criteria

We first introduce a classical definition of the PGD based on Galerkin orthogonality criteria [21,12,22]. Although there is no existence nor convergence results about the proposed definition (in the general case), it appears to be an efficient technique for building separated representations in many applications. Different variants and improvements of this original definition of the PGD are proposed and discussed. These improvements are inspired from recent works in the context of stochastic partial differential equations [27,28,32].

4.1.1. Progressive definition of space and time functions

We assume that a decomposition u_{m-1} of order $(m-1)$ is known. For the definition of the order m decomposition, a new couple $(w, \lambda) \in \mathcal{V} \times \mathcal{T}$ is defined as the optimal couple which verifies the double Galerkin orthogonality criterium:

$$B(u_{m-1} + w\lambda, w\lambda^* + w^*\lambda) = L(w\lambda^* + w^*\lambda), \quad \forall \lambda^* \in \mathcal{T}, \ \forall w^* \in \mathcal{V}. \tag{32}$$

Definition 5. We introduce the following mappings:

- $S_m : \mathcal{T} \to \mathcal{V}$ is the application which maps a time function $\lambda \in \mathcal{T}$ into a space function $w = S_m(\lambda) \in \mathcal{V}$, defined by:
$$B(u_{m-1} + w\lambda, w^*\lambda) = L(w^*\lambda), \quad \forall w^* \in \mathcal{V}. \tag{33}$$
- $T_m : \mathcal{V} \to \mathcal{T}$ is the application which maps a space function $w \in \mathcal{V}$ into a time function $\lambda = T_m(w) \in \mathcal{T}$, defined by:
$$B(u_{m-1} + w\lambda, w\lambda^*) = L(w\lambda^*), \quad \forall \lambda^* \in \mathcal{T}. \tag{34}$$

A couple $(w, \lambda) \in \mathcal{V} \times \mathcal{T}$ then verifies Eq. (32) if and only if $w = S_m(\lambda)$ and $\lambda = T_m(w)$. The progressive Galerkin-based PGD is then defined as follows.

Definition 6 (*Progressive Galerkin PGD*). $(w_m, \lambda_m) \in \mathcal{V} \times \mathcal{T}$ is defined as the "optimal" couple among those $(w, \lambda) \in \mathcal{V} \times \mathcal{T}$ which verify one of the following equivalent properties:

- (w, λ) verifies
$$w = S_m(\lambda) \quad \text{and} \quad \lambda = T_m(w). \tag{35}$$
- $\lambda = T_m(w)$ and w is a fixed point of mapping $G_m = S_m \circ T_m$, *i.e.*
$$w = G_m(w). \tag{36}$$
- $w = S_m(\lambda)$ and λ is a fixed point of mapping $G_m^\diamondsuit := T_m \circ S_m$, *i.e.*
$$\lambda = G_m^\diamondsuit(\lambda). \tag{37}$$

We will see that optimality in Definition 6 is unclearly defined in the general case.

4.1.2. Interpretation as a pseudo-eigenproblem

We can easily show that G_m and $G_m^\diamondsuit$ in (36) and (37) are homogeneous operators of degree 1, *i.e.*

$$\forall \alpha \in \mathbb{R} \setminus \{0\}, \quad G_m(\alpha w) = \alpha G_m(w), \quad G_m^\diamondsuit(\alpha\lambda) = \alpha G_m^\diamondsuit(\lambda). \tag{38}$$

Problems (36) and (37) are then interpreted as pseudo-eigenproblems, the optimal functions w and λ being the dominant eigenfunctions of G_m and $G_m^\diamondsuit$, respectively. A function w (resp. λ) is called an eigenfunction of G_m (resp. $G_m^\diamondsuit$) if it verifies (36) (resp. (37)) (see [28] for further discussions on these pseudo-eigenproblems). The associated eigenvalue can be defined by

$$\sigma_m(w) = B(wT_m(w), wT_m(w)), \tag{39}$$
$$(\text{resp.} \quad \sigma_m^\diamondsuit(\lambda) = B(S_m(\lambda)\lambda, S_m(\lambda)\lambda)). \tag{40}$$

In the general case, the above interpretation needs for further mathematical investigations (it is a non-classical mathematical problem). However, the following degenerate case highlights the above interpretation.

4.1.2.1. Degenerate case. We consider that bilinear form B defined on tensor product space $\mathcal{V} \otimes \mathcal{T}$ can be written as the product of a bilinear form B_S on $\mathcal{V}$ and of a bilinear form B_T on $\mathcal{T}$, *i.e.* $\forall \lambda, \lambda^* \in \mathcal{T}$ and $\forall w, w^* \in \mathcal{V}$,

$$B(w\lambda, w^*\lambda^*) = B_S(w, w^*)B_T(\lambda, \lambda^*). \tag{41}$$

Remark 7. Of course, in the present context, separation property (41) of B is not satisfied for general time-dependent PDEs but it can be seen as a limit case when some parameters of the PDE become negligible.

If separation property (41) is satisfied, one can easily prove that eigenproblems (36) and (37) are equivalent to

$$\sigma_m(w)w = \check{G}_m(w) \quad \text{and} \quad \sigma_m^\diamondsuit(\lambda)\lambda = \check{G}_m^\diamondsuit(\lambda), \tag{42}$$

where

$$\check{G}_m(w) = B_T(u - u_{m-1}, B_S(u - u_{m-1}, w)) = \sigma_m(w)G_m(w), \tag{43}$$
$$\check{G}_m^\diamondsuit(\lambda) = B_S(u - u_{m-1}, B_T(u - u_{m-1}, \lambda)) = \sigma_m^\diamondsuit(\lambda)G_m^\diamondsuit(\lambda). \tag{44}$$

For equivalent eigenfunctions w and λ, *i.e.* related by $\lambda = T_m(w)$ or $w = S_m(\lambda)$, the associated eigenvalues $\sigma_m(w) = \sigma_m^\diamondsuit(\lambda)$ and

$$\sigma_m(w) = B_S(w, w)B_T(T_m(w), T_m(w)) = B_S(w, \check{G}_m(w))B_S(w, w)^{-1}, \tag{45}$$
$$\sigma_m^\diamondsuit(\lambda) = B_S(S_m(\lambda), S_m(\lambda))B_T(\lambda, \lambda) = B_T(\lambda, \check{G}_m^\diamondsuit(\lambda))B_T(\lambda, \lambda)^{-1}. \tag{46}$$

In the case where B_S and B_T define inner products on $\mathcal{V}$ and $\mathcal{T}$, respectively, operators $\check{G}_m$ and $\check{G}_m^\diamondsuit$ are classical correlation operators of $(u - u_{m-1})$ constructed from these inner products. Then, if we choose $w_m \in \mathcal{V}$ and $\lambda_m \in \mathcal{T}$ as the dominant eigenfunctions of $\check{G}_m$ and $\check{G}_m^\diamondsuit$, respectively, $u_{m-1} + w_m\lambda_m$ minimizes the distance to u

measured with the metric induced by B_S and B_T. The obtained decomposition u_m is the classical Proper Orthogonal Decomposition which converges towards u with respect to this metric. It is optimal in the sense that for a given order m, it is the optimal separated representation with respect to this metric.[8] However, if bilinear forms B_S and B_T do not define inner products on $\mathcal{V}$ and $\mathcal{T}$ (in particular if they are non-symmetric), we have no guaranty that these eigenproblems admit real positive eigenvalues and that dominant eigenfunctions will lead to a convergent sequence u_m. Therefore, the obtained decomposition u_m cannot be interpreted as a Proper Orthogonal Decomposition of the solution u.

4.1.2.2. General case. When separation property (41) is not satisfied but when bilinear form B still defines an inner product on $\mathcal{V}\otimes\mathcal{T}$, it is still possible to define pseudo-eigenfunctions and eigenvalues (see [28]) and to prove the convergence of the obtained decomposition u_m with respect to this metric. Denoting by $\|\cdot\|_B = B(\cdot,\cdot)$ the norm induced by B, we have

$$\|u-u_m\|_B^2 = \|u\|_B^2 - \sum_{i=1}^{m}\sigma_i(w_i) = \|u\|_B^2 - \sum_{i=1}^{m}\sigma_i^{\diamondsuit}(\lambda_i), \tag{47}$$

where $\sigma_i(w_i)$ (resp. $\sigma_i^{\diamondsuit}(\lambda_i)$) is the dominant eigenvalue, with σ_i defined by (39) (resp. (40)). In a more general case, when B does not define an inner product (e.g. B non-symmetric as it is the case for the present context of time-dependent PDEs), there is no existence results for solutions (in real Hilbert spaces) to pseudo-eigenproblem (36) (or (37)). However, the interpretation in terms of a pseudo-eigenproblem is a fruitful interpretation in the sense that it allows to propose dedicated algorithms for the construction of the decomposition, these algorithms being inspired from classical algorithms for eigenproblems.

4.1.3. Power algorithm for the construction of the PGD

For a given decomposition u_{m-1}, a new optimal function $w_m\in\mathcal{V}$ (resp. $\lambda_m\in\mathcal{T}$) is defined as the dominant pseudo-eigenfunction of operator G_m (resp. $G_m^{\diamondsuit}$). It leads to the definition of the decomposition $u_m = u_{m-1} + w_m\lambda_m$, with $\lambda_m = T_m(w_m)$ (resp. $w_m = S_m(\lambda_m)$). A natural algorithm to capture the dominant eigenfunction of operator $G_m = S_m\circ T_m$ consists in performing power-type iterations. Starting from an initial function $w^{(0)} = S_m(\lambda^{(0)})$ (in practice, $\lambda^{(0)}$ is generated randomly), we compute the sequence $w^{(k+1)} = G_m(w^{(k)}) = S_m\circ T_m(w^{(k)})$. This leads to the Algorithm 1.

Algorithm 1 (*(PGD-P) (Power iterations algorithm)*).

```
1: for m = 1 to m_max do
2:    Initialize λ
3:    for k = 1 to k_max do
4:       Compute w = S_m(λ)
5:       Normalize  w
6:       Compute λ = T_m(w)
7:       Check convergence of (wλ)
8:    end for
9:    Set w_m = w and λ_m = λ
10:   Set u_m = u_{m-1} + w_m λ_m and check convergence
11: end for
```

Remark 8. In Algorithm 1, there is no need for normalization of functions w or λ. However, we here arbitrarily choose to normalize space functions w.

[8] This construction leads to a uniquely defined decomposition (if eigenvalues are of multiplicity 1), where functions w_i (resp. λ_i) are orthogonal with respect to inner product B_S (resp. B_T).

For particular cases where PGD coincides with POD, Algorithm 1 is a classical power iterations algorithm with deflation for the capture of the dominant eigenspace of the classical eigenproblem associated with the correlation operator of the solution u.

Remark 9. In practise, we select a relatively coarse stagnation criterium for the power iterations ($\approx 10^{-2}$ on $w\lambda$), which is sufficient to obtain a good approximation of the new couple (w_m,λ_m). In general, power iterations reach this criterium very fast (in $k\approx 3$ iterations). A slow convergence may reveal multiple or close eigenvalues. However, also in this case, a good couple (w,λ) (although not converged) is often reached in a few iterations. In practise, we then select $k_{max}\approx 3$.

4.1.4. Optimal Galerkin PGD and its construction

For the general case where PGD does not coincide with POD, Algorithm 1 does not lead to an optimal decomposition $u_m = W_m\cdot\Lambda_m$ in the sense of the Galerkin projection. An optimal Galerkin PGD can be defined by imposing the residual to be simultaneously orthogonal to reduced spaces $\mathcal{V}_m = span(W_m)\subset\mathcal{V}$ and $\mathcal{T}_m = span(\Lambda_m)\subset\mathcal{T}$. These two Galerkin orthogonality criteria are Eqs. (16) and (18) and can be reformulated as follows.

Definition 10. Let $T:(\mathcal{V})^m\to(\mathcal{T})^m$ be the application which maps space functions $W_m\in(\mathcal{V})^m$ into time functions $\Lambda_m = T(W_m)\in(\mathcal{T})^m$ defined by

$$B(W_m\cdot\Lambda_m, W_m\cdot\Lambda_m^*) = L(W_m\cdot\Lambda_m^*),\quad \forall\Lambda_m^*\in(\mathcal{T})^m. \tag{48}$$

Let $S:(\mathcal{T})^m\to(\mathcal{V})^m$ be the application which maps time functions $\Lambda_m\in(\mathcal{T})^m$ into space functions $W_m = S(\Lambda_m)\in(\mathcal{V})^m$ defined by

$$B(W_m\cdot\Lambda_m, W_m^*\cdot\Lambda_m) = L(W_m^*\cdot\Lambda_m),\quad \forall W_m^*\in(\mathcal{V})^m. \tag{49}$$

A decomposition $u_m = W_m\cdot T(W_m)$verifies Eq. (16). The application of mapping T requires the solution of the system of ordinary differential Eq. (17b). A decomposition $u_m = S(\Lambda_m)\cdot\Lambda_m$ verifies Eq. (18). The application of mapping S requires the solution of a set of time-independent partial differential Eq. (19). The optimal Galerkin-based PGD is then defined as follows.

Definition 11 (*Optimal Galerkin PGD*). The couple $(W_m,\Lambda_m)\in(\mathcal{V})^m\times(\mathcal{T})^m$ is defined as the "optimal" couple among those which verify one of the following equivalent properties:

- (W_m,Λ_m) verifies

$$W_m = S(\Lambda_m)\quad\text{and}\quad\Lambda_m = T(W_m). \tag{50}$$

- $\Lambda_m = T(W_m)$ and W_m is a fixed point of mapping $G = S\circ T$, *i.e.*

$$W_m = G(W_m). \tag{51}$$

- $W_m = S(\Lambda_m)$ and Λ_m is a fixed point of mapping $G^{\diamondsuit} = T\circ S$, *i.e.*

$$\Lambda_m = G^{\diamondsuit}(\Lambda_m). \tag{52}$$

Eq. (51) (resp. (52)) can still be interpreted as a pseudo-eigenproblem on operator $G = S\circ T$ (or $G^{\diamondsuit} = T\circ S$), the optimal set of functions W_m (resp. Λ_m) being associated with the dominant eigenspace of G (resp. $G^{\diamondsuit}$) (see [28]). A natural algorithm then consists in performing subspace iterations in order to capture this dominant eigenspace (Algorithm 2). This algorithm leads to optimal sets of functions in the sense of Galerkin projection and leads to a decomposition u_m which can be significantly better than the decomposition obtained with a progressive definition of the PGD, constructed by a power iterations algorithm (Algorithm 1).

Algorithm 2 (*(PGD-S) (Subspace iterations algorithm)*).

1: **for** $m = 1$ to m_{max} **do**
2: Initialize Λ_m
3: **for** $k = 1$ to k_{max} **do**
4: Compute $W_m = S(\Lambda_m)$
5: Compute $\Lambda_m = T(W_m)$
6: Check convergence
7: **end for**
8: Set $u_m = W_m \cdot \Lambda_m$ and check convergence
9: **end for**

4.1.5. Approximations of the optimal Galerkin PGD

At each iteration, the application of mapping $G = S \circ T$ requires the application of mappings T and S. In the context of time-dependent PDEs, the application of mapping T is relatively cheap since it corresponds to a set of ODEs (formulation of the initial problem on a reduced basis of space functions). However, the application of mapping S is generally very costly (and not usual) for large scale applications (2D or 3D problems) since it involves the solution of a set of coupled time-independent PDEs. In the present context, this algorithm should be avoided. The question is then: can we build a better decomposition than the progressive PGD without applying the mapping S?

A simple modification of power iterations algorithm (Algorithm 1) is possible. It consists in introducing the application of mapping T in order to update the whole set of time functions Λ_m after each construction of a new couple of functions (w_m, λ_m). This leads to the Algorithm 3, which corresponds to the power iterations algorithm with an additional updating of time functions. Although the obtained decomposition is not the optimal one, this modification significantly improves the quality of the progressive PGD.

Algorithm 3 (*(PGD-P*) Power iterations algorithm with update*).

1: **for** $m = 1$ to m_{max} **do**
2: Perform steps 2 to 9 of Algorithm 1 and obtain the dominant eigenfunction w_m of G_m.
3: Compute $\Lambda_m = T(W_m)$
4: Set $u_m = W_m \cdot \Lambda_m$ and check convergence
5: **end for**

Remark 12. Another algorithm, inspired from Arnoldi algorithm for classical eigenproblems, has been proposed in [28] in the context of stochastic partial differential equations. This algorithm allows the capture of an approximation of the dominant eigenspace of G without applying the mapping S. For many applications, power iterations algorithm with update and Arnoldi algorithm lead to very similar convergence properties. Arnoldi algorithm is usually more efficient since it requires the solution of only m uncoupled time-independent PDEs in order to build an order m decomposition. In this paper, we do not focus on efficiency aspects of PGD constructions but only on convergence properties of the different definitions. This Arnoldi algorithm is not detailed but it should be considered as a way to further improve computational efficiency of PGD methods (see efficiency analyses in [28]).

4.2. PGD based on a minimal residual formulation

In this section, we introduce another possible definition of the PGD based on a minimal residual criterium [31,23]. This construction is more "robust" than the Galerkin PGD in the sense that monotone convergence of the decomposition in the residual norm can be proved. However, it has several drawbacks which are detailed in Section 4.2.4.

4.2.1. Progressive definition of time and space functions

We still consider a progressive definition of the decomposition by assuming that a decomposition u_{m-1} of order $m-1$ is known. Let us denote by $\langle\langle \cdot, \cdot \rangle\rangle$ an inner product on $\mathcal{V} \otimes \mathcal{T}$ and by $\| \cdot \|$ the associated norm. Let us define the residual $\mathcal{R}(u) \in \mathcal{V} \otimes \mathcal{T}$ of Eq. (7) by:

$$\langle\langle v, \mathcal{R}(u) \rangle\rangle = L(v) - B(u, v) = \langle\langle v, \ell - \mathcal{B}(u) \rangle\rangle, \quad \forall v \in \mathcal{V} \otimes \mathcal{T}, \tag{53}$$

where $\ell \in \mathcal{V} \otimes \mathcal{T}$ and operator $\mathcal{B} : \mathcal{V} \otimes \mathcal{T} \to \mathcal{V} \otimes \mathcal{T}$ are defined by using Riesz representations in Hilbert space $\mathcal{V} \otimes \mathcal{T}$.

We introduce the following natural definition of the Minimal Residual PGD.

Definition 13 (*Progressive Minimal Residual PGD*). An optimal couple $(w_m, \lambda_m) \in \mathcal{V} \times \mathcal{T}$ is defined as the one which minimizes the residual norm:

$$(w_m, \lambda_m) \in \arg \min_{(w,\lambda) \in \mathcal{V} \times \mathcal{T}} \|\mathcal{R}(u_{m-1} + w\lambda)\|^2 \tag{54}$$

or equivalently:

$$(w_m, \lambda_m) \in \arg \min_{(w,\lambda) \in \mathcal{V} \times \mathcal{T}} \frac{1}{2} \langle\langle \mathcal{B}(w\lambda), \mathcal{B}(w\lambda) \rangle\rangle - \langle\langle \mathcal{R}(u_{m-1}), \mathcal{B}(w\lambda) \rangle\rangle. \tag{55}$$

Let us note that this definition of PGD is equivalent to the Galerkin PGD applied to a least-square formulation of the problem (symmetrized problem).

4.2.2. Interpretation as a pseudo-eigenproblem

The stationarity conditions (or Euler–Lagrange equations) associated with quadratic optimization problem (54) write: $\forall (w^*, \lambda^*) \in \mathcal{V} \times \mathcal{T}$,

$$\langle\langle \mathcal{B}(w\lambda), \mathcal{B}(w\lambda^* + w^*\lambda) \rangle\rangle = \langle\langle \mathcal{R}(u_{m-1}), \mathcal{B}(w\lambda^* + w^*\lambda) \rangle\rangle. \tag{56}$$

In order to directly apply the analysis and algorithms of Section 4.1, we reformulate the problem with similar notations.

Definition 14. We use the following definitions for mappings S_m and T_m.

- $S_m : \mathcal{T} \to \mathcal{V}$ is the application which maps a time function $\lambda \in \mathcal{T}$ into a space function $w = S_m(\lambda) \in \mathcal{V}$ defined by:

$$\langle\langle \mathcal{B}(w\lambda), \mathcal{B}(w^*\lambda) \rangle\rangle = \langle\langle \mathcal{R}(u_{m-1}), \mathcal{B}(w^*\lambda) \rangle\rangle, \quad \forall w^* \in \mathcal{V}. \tag{57}$$

- $T_m : \mathcal{V} \to \mathcal{T}$ is the application which maps a space function $w \in \mathcal{V}$ into a time function $\lambda = T_m(w) \in \mathcal{T}$ defined by

$$\langle\langle \mathcal{B}(w\lambda), \mathcal{B}(w\lambda^*) \rangle\rangle = \langle\langle \mathcal{R}(u_{m-1}), \mathcal{B}(w\lambda^*) \rangle\rangle, \quad \forall \lambda^* \in \mathcal{T}. \tag{58}$$

Stationarity conditions (57) and (58) can then be written $w = S_m(\lambda)$ and $\lambda = T_m(w)$, respectively. The simultaneous verification of both equations can still be interpreted as pseudo-eigenproblem (36) (or (37)) on operator $G_m = S_m \circ T_m$ (or $G_m^{\diamond} = T_m \circ S_m$). The (pseudo) eigenvalue associated with an eigenfunction w of G_m (*i.e.* such that $w = G_m(w)$) is defined by $\sigma_m(w) = \langle\langle \mathcal{B}(wT_m(w)), \mathcal{B}(wT_m(w)) \rangle\rangle$. The optimal function w_m is the dominant eigenfunction of G_m, which maximizes $\sigma_m(w)$. The obtained couple $(w_m, T_m(w_m))$ is optimal in the sense that it minimizes the residual norm. We easily prove the following property:

$$\|\mathcal{R}(u_m)\|^2 = \|\mathcal{R}(u_{m-1})\|^2 - \sigma_m(w_m) = \|\ell\|^2 - \sum_{i=1}^{m} \sigma_i(w_i) \tag{59}$$

which shows that the residual norm is monotonically decreasing. Decomposition u_m can be constructed by using power iterations algorithm (Algorithm 1), where power iterations allow the capture of the dominant eigenfunctions w_m of successive operators G_m.

Remark 15. Power iterations can also be interpreted as an alternated minimization algorithm for solving (54) (minimizing successively on time functions and space functions).

4.2.3. Other definitions and algorithms

Other definitions and algorithms introduced in Sections 4.1.4 and 4.1.5 for the Galerkin PGD can be naturally applied to the minimal residual PGD, in order to improve the convergence properties of the decomposition. It simply consists in applying the definitions and algorithms of Galerkin-based PGD on the following least-square formulation of the problem:

$$\begin{aligned} &u \in \mathcal{V}\otimes\mathcal{T}, \quad \widehat{B}(u,v) = \widehat{L}(v) \quad \forall v \in \mathcal{V}\otimes\mathcal{T},\\ &\widehat{B}(u,v) = \langle\!\langle \mathcal{B}(v), \mathcal{B}(u)\rangle\!\rangle = \langle\!\langle v, \mathcal{B}^*\mathcal{B}(u)\rangle\!\rangle,\\ &\widehat{L}(v) = \langle\!\langle \mathcal{B}(v), \ell\rangle\!\rangle = \langle\!\langle v, \mathcal{B}^*(\ell)\rangle\!\rangle \end{aligned} \tag{60}$$

where $\mathcal{B}^*$ is the adjoint operator of $\mathcal{B}$. Algorithms introduced in Sections 4.1.4 and 4.1.5 are then simply transposed to this formulation by using bilinear form $\widehat{B}$ and linear form $\widehat{L}$ (instead of B and L) in the definition of mappings T_m, S_m, T and S.

4.2.4. Comments on the minimal residual PGD

An advantage of the minimal residual formulation is that the convergence with m of the PGD u_m is monotonic, if convergence is evaluated with the residual norm (cf. Eq. (59)). In that sense, it is a robust construction of a separated representation and it can be used in cases where Galerkin-based PGD fails. However, it has several drawbacks:

- Although it has a monotonic convergence in residual norm, the resulting decomposition may present very poor convergence properties with respect to usual norms. This will be illustrated in numerical examples. In fact, the convergence rate with respect to usual metrics strongly depends on the choice of the residual norm $\|\cdot\|$. Taking for the residual norm the natural norm in $L^2(\Omega)\otimes L^2(I)$ usually leads to bad convergence rates in usual solution norms. A suitable residual norm, constructed from the operator of the problem, may improve this convergence rate.[9] However, the construction of a suitable norm is not straightforward and the use of optimal norms may induce additional computational issues.
- This formulation can be easily implemented in a discretized framework, where the residual $\mathcal{R}(u) \in \mathcal{V}_N\otimes\mathcal{T}_P$ is defined by replacing function space $\mathcal{V}\otimes\mathcal{T}$ by $\mathcal{V}_N\otimes\mathcal{T}_P$ in (53). The residual norm then measures the error with respect to the classical Galerkin solution $u \in \mathcal{V}_N\otimes\mathcal{T}_P$ (*i.e.* the solution of Eq. (13)). Algorithms are easily implemented in an algebraic setting (see Appendix B). However, in a continuous framework, it leads to non-classical formulations, which require the introduction of more refined function spaces in order to guaranty existence and uniqueness of solutions (*e.g.* by introducing $H^2(\Omega)\otimes H^1(I)$ in place of $H^1(\Omega)\otimes L^2(I)$ for the advection–diffusion–reaction equation). The *a posteriori* construction of approximation spaces for the approximation of time and space functions is then non-trivial (at least at the space level). In practice, a minimal residual formulation should then be applied to the discretized problem (13) (after the introduction of space and time approximation spaces $\mathcal{V}_N$ and $\mathcal{T}_P$). Let us note that time problems are now global in time[10] (equivalent to the weak formulation of a second order differential equation with initial and final conditions) and therefore, this formulation does not allow to take part of the causality of the initial problem for the computation of time functions. This can be a major drawback if a very high dimensional approximation space $\mathcal{T}_P$ is used.
- This formulation requires much more computational efforts than the Galerkin PGD. Indeed, PGD algorithms take part of the low order (or low rank) separated representation of operator $\mathcal{B}$ and right-hand side ℓ (see Appendix A on computational aspects). This minimal residual formulation being equivalent to a Galerkin-based PGD on operator $\mathcal{B}^*\mathcal{B}$ and right-hand side $\mathcal{B}^*\ell$, where $\mathcal{B}^*$ denotes the adjoint operator of $\mathcal{B}$, the separation orders are dramatically increased.

For all the above reasons, this PGD based on a minimal residual formulation should be avoided — in the opinion of the author — in practical applications for which Galerkin-based PGDs work.

5. Minimax Proper Generalized Decomposition

In this section, we propose a new definition of PGD which allows to improve the convergence properties of Galerkin-based PGD with respect to a desired metric. This new definition can be interpreted as a PGD based on Petrov–Galerkin criteria, where the orthogonality of the residual is imposed with respect to another set of space and time functions, which are solution of an adjoint problem. We also propose an algorithm for its construction.

Remark 16. Let us note that this new definition could be also applied to the Minimal Residual PGD by considering that the initial formulation is the symmetrized Least-Square formulation of the time-dependent PDE.

5.1. Definition of the Minimax Proper Generalized Decomposition

We here assume that a decomposition u_{m-1} of order $(m-1)$ is known (previously computed). The idea is to define a new couple $(w,\lambda) \in \mathcal{V}\times\mathcal{T}$ with the two following orthogonality criteria:

$$B(u_{m-1} + w\lambda, \tilde{w}^*\tilde{\lambda}) = L(\tilde{w}^*\tilde{\lambda}), \quad \forall \tilde{w}^* \in \mathcal{V}, \tag{61a}$$

$$B(u_{m-1} + w\lambda, \tilde{w}\tilde{\lambda}^*) = L(\tilde{w}\tilde{\lambda}^*), \quad \forall \tilde{\lambda}^* \in \mathcal{T}, \tag{61b}$$

where $(\tilde{w},\tilde{\lambda}) \in \mathcal{V}\times\mathcal{T}$ is another couple of space and time functions. Eq. (61) impose the cancelation of the projection of the residual on subspaces $\mathcal{V}\otimes\{\tilde{\lambda}\}$ and $\{\tilde{w}\}\otimes\mathcal{T}$, instead of $\mathcal{V}\otimes\{\lambda\}$ and $\{w\}\otimes\mathcal{T}$ for the classical Galerkin-based PGD (Eq. (32)). Of course, additional equations must be added in order to define functions $(\tilde{w},\tilde{\lambda}) \in \mathcal{V}\times\mathcal{T}$. We use the following additional orthogonality criteria:

$$B(w^*\lambda, \tilde{w}\tilde{\lambda}) = \langle\!\langle w^*\lambda, w\lambda\rangle\!\rangle \quad \forall w^* \in \mathcal{V}, \tag{62a}$$

$$B(w\lambda^*, \tilde{w}\tilde{\lambda}) = \langle\!\langle w\lambda^*, w\lambda\rangle\!\rangle \quad \forall \lambda^* \in \mathcal{T}, \tag{62b}$$

where $\langle\!\langle\cdot,\cdot\rangle\!\rangle$ is an inner product on $\mathcal{V}\otimes\mathcal{T}$. In practice, we select a classical inner product having the following separation property: $\forall\lambda,\lambda^* \in \mathcal{T}$ and $\forall w, w^* \in \mathcal{V}$,

$$\langle\!\langle w\lambda, w^*\lambda^*\rangle\!\rangle = \langle w, w^*\rangle_{\mathcal{V}}\langle\lambda,\lambda^*\rangle_{\mathcal{T}}, \tag{63}$$

[9] If we want to obtain a good convergence of the solution with respect to a certain norm $\|\cdot\|$ (e.g. L^2 norm), the residual norm $\|\mathcal{R}(u_m)\|$ should give a measure of the error as close as possible to $\|u-u_m\|$. A natural L^2-norm of the residual, *i.e.* $\|\mathcal{R}(u_m)\|^2 = \langle\!\langle\mathcal{R}(u_m),\mathcal{R}(u_m)\rangle\!\rangle$, is clearly not adapted. One could choose $\|\mathcal{R}(u)\|^2 = \langle\!\langle\mathcal{R}(u),\mathcal{M}\mathcal{R}(u)\rangle\!\rangle$, where $\mathcal{M}$ is a suitable symmetric bounded coercive operator. For example, if we denote $\mathcal{B} = \frac{\partial}{\partial t} + \mathcal{A}$, with $\mathcal{A}$ a bounded coercive operator, one could choose for $\mathcal{M}$ the inverse of the symmetric part of $\mathcal{A}$ (or an approximation of it, for computational reasons).

[10] Let us briefly illustrate the fact that time problems are global in time. A time problem $\min_\lambda\|\mathcal{R}(u_{m-1}+w\lambda)\|^2$ can be recast as $\min_\lambda\|l-\dot{\lambda}-a\lambda\|^2$. For simplicity, suppose that $\lambda(0)=0$ is imposed in a strong sense. If we select a L^2-norm for the residual norm, we have $\|l\|^2 = \int_I l(t)^2\,dt$ and the Euler–Lagrange equation associated with the above quadratic minimization problem is $\int_I(\dot{\lambda}^*+a\lambda^*)(\dot{\lambda}+a\lambda-l)dt = 0,\ \forall\lambda^*$. After an integration by part, one can easily prove that the associated strong form is a second order differential equation with initial and final conditions.

where $\langle\cdot,\cdot\rangle_{\mathcal{V}}$ and $\langle\cdot,\cdot\rangle_{\mathcal{T}}$ are inner products on $\mathcal{V}$ and $\mathcal{T}$, respectively. The idea is to construct simultaneously two sets of functions $(w,\lambda)\in\mathcal{V}\otimes\mathcal{T}$ and $(\tilde{w},\tilde{\lambda})\in\mathcal{V}\otimes\mathcal{T}$ such that they verify Eqs. (61) and (62) simultaneously.

Let us now introduce a rigorous definition of optimal sets of functions (w,λ) and $(\tilde{w},\tilde{\lambda})$.

Definition 17. We introduce the functional $\mathcal{L}_m:(\mathcal{V}\otimes\mathcal{T})\times(\mathcal{V}\otimes\mathcal{T})\to\mathbb{R}$ defined for $v,\tilde{v}\in\mathcal{V}\otimes\mathcal{T}$ by

$$\mathcal{L}_m(v,\tilde{v})=\frac{1}{2}\langle\!\langle v,v\rangle\!\rangle-B(u_{m-1}+v,\tilde{v})+L(\tilde{v}). \tag{64}$$

We also introduce $\widehat{\mathcal{L}}_m:(\mathcal{V}\times\mathcal{V})\times(\mathcal{T}\times\mathcal{T})\to\mathbb{R}$, the functional associated with the restriction of $\mathcal{L}_m$ on the set of rank-one separated functions, defined for $\{w,\tilde{w}\}\in\mathcal{V}\times\mathcal{V}$ and $\{\lambda,\tilde{\lambda}\}\in\mathcal{T}\times\mathcal{T}$ by

$$\widehat{\mathcal{L}}_m(\{w,\tilde{w}\},\{\lambda,\tilde{\lambda}\})=\mathcal{L}_m(w\lambda,\tilde{w}\tilde{\lambda}).$$

We then propose the following definition of the Minimax PGD.[11]

Definition 18 (*Progressive Minimax PGD*). The set of functions $(w_m,\tilde{w}_m,\lambda_m,\tilde{\lambda}_m)\in(\mathcal{V}\times\mathcal{V}\times\mathcal{T}\times\mathcal{T})$ is defined by

$$(w_m\lambda_m,\tilde{w}_m\tilde{\lambda}_m)\in\arg\max_{\tilde{w}\in\mathcal{V},\tilde{\lambda}\in\mathcal{T}}\min_{w\in\mathcal{V},\lambda\in\mathcal{T}}\mathcal{L}_m(w\lambda,\tilde{w}\tilde{\lambda}). \tag{65}$$

We notice that (61a), (61b), (62a) and (62b) are the stationarity conditions of functional $\mathcal{L}_m$ with respect to $\tilde{w}$, $\tilde{\lambda}$, w and λ, respectively. The set of functions $(w,\tilde{w},\lambda,\tilde{\lambda})$ then satisfies Eqs. (61) and (62) if and only if it makes stationary the functional $\mathcal{L}_m$. For fixed $(w,\tilde{w})$ (resp. $(\lambda,\tilde{\lambda})$), $\tilde{\lambda}$ (resp. $\tilde{w}$) appears as a Lagrange multiplier which imposes the orthogonality of the residual $\mathcal{R}(u_{m-1}+\lambda w)$ with respect to subspace $\{\tilde{w}\}\otimes\mathcal{T}$ (resp. $\mathcal{V}\otimes\{\tilde{\lambda}\}$).

In the following, we introduce operator $\mathcal{B}:\mathcal{V}\otimes\mathcal{T}\to\mathcal{V}\otimes\mathcal{T}$ and element $\ell\in\mathcal{V}\otimes\mathcal{T}$ associated with B and L by Riesz representation (using inner product $\langle\!\langle\cdot,\cdot\rangle\!\rangle$), as defined in Section 4.2. Functional $\mathcal{L}_m$ can equivalently be written

$$\mathcal{L}_m(v,\tilde{v})=\frac{1}{2}\langle\!\langle v,v\rangle\!\rangle-\langle\!\langle\mathcal{B}(v),\tilde{v}\rangle\!\rangle+\langle\!\langle\mathcal{R}(u_{m-1}),\tilde{v}\rangle\!\rangle, \tag{66}$$

where $\mathcal{R}(u_{m-1})=\ell-\mathcal{B}(u_{m-1})$ is the residual associated with u_{m-1}. If we denote by $e_m=u-u_{m-1}\in\mathcal{V}\otimes\mathcal{T}$, with $u=\mathcal{B}^{-1}\ell$, and $\tilde{e}_m=\mathcal{B}^{*-1}(e_m)$, where $\mathcal{B}^*$ is the adjoint operator of $\mathcal{B}$, we easily see that

$$(e_m,\tilde{e}_m)=\arg\max_{\tilde{v}\in\mathcal{V}\otimes\mathcal{T}}\min_{v\in\mathcal{V}\otimes\mathcal{T}}\mathcal{L}_m(v,\tilde{v}).$$

We can then interpret $w_m\lambda_m$ (resp. $\tilde{w}_m\tilde{\lambda}_m$) as the best rank-one separated representation of $e_m=u-u_{m-1}$ (resp. $\mathcal{B}^{*-1}(e_m)$), in the sense of the minimax problem.[12]

5.2. Reformulation as a fixed point problem

Definition 19. We introduce the following mappings:

- $S_m:\mathcal{T}\times\mathcal{T}\to\mathcal{V}$ is the application which maps time functions $\{\lambda,\tilde{\lambda}\}\in\mathcal{T}\times\mathcal{T}$ into a space function $w=S_m(\lambda,\tilde{\lambda})\in\mathcal{V}$ defined by Eq. (61a).
- $T_m:\mathcal{V}\times\mathcal{V}\to\mathcal{T}$ is the application which maps space functions $\{w,\tilde{w}\}\in\mathcal{V}\times\mathcal{V}$ into a time function $\lambda=T_m(w,\tilde{w})\in\mathcal{T}$ defined by Eq. (61b).
- $\widetilde{S}:\mathcal{V}\times\mathcal{T}\times\mathcal{T}\to\mathcal{V}$ is the application which maps functions $\{w,\lambda,\tilde{\lambda}\}\in\mathcal{V}\times\mathcal{T}\times\mathcal{T}$ into a space function $\tilde{w}=\widetilde{S}(w;\lambda,\tilde{\lambda})\in\mathcal{V}$ defined by Eq. (62a).
- $\widetilde{T}:\mathcal{T}\times\mathcal{V}\times\mathcal{V}\to\mathcal{T}$ is the application which maps functions $\{\lambda,w,\tilde{w}\}\in\mathcal{T}\times\mathcal{V}\times\mathcal{V}$ into a time function $\tilde{\lambda}=\widetilde{T}(\lambda;w,\tilde{w})\in\mathcal{T}$ defined by Eq. (62b).

The set of functions $\{w,\lambda\}\in\mathcal{V}\times\mathcal{T}$ and $\{\tilde{w},\tilde{\lambda}\}\in\mathcal{V}\times\mathcal{T}$ is then searched as the optimal set of functions verifying simultaneously

$$w=S_m(\lambda,\tilde{\lambda}),\quad \lambda=T_m(w,\tilde{w}),\quad \tilde{w}=\widetilde{S}(w;\lambda,\tilde{\lambda}),\quad \tilde{\lambda}=\widetilde{T}(\lambda;w,\tilde{w}). \tag{67}$$

Definition 20. We denote by $\widehat{S}_m:\mathcal{T}\times\mathcal{T}\to\mathcal{V}\times\mathcal{V}$ and $\widehat{T}_m:\mathcal{V}\times\mathcal{V}\to\mathcal{T}\times\mathcal{T}$ the mappings defined by

$$\widehat{S}_m(\lambda,\tilde{\lambda})=\{S_m(\lambda,\tilde{\lambda}),\widetilde{S}(S_m(\lambda,\tilde{\lambda});\lambda,\tilde{\lambda})\}, \tag{68}$$

$$\widehat{T}_m(w,\tilde{w})=\{T_m(w,\tilde{w}),\widetilde{T}(T_m(w,\tilde{w});w,\tilde{w})\}. \tag{69}$$

The following definition of mappings $\widehat{S}_m$ and $\widehat{T}_m$ also holds.

Proposition 21. *Mappings $\widehat{S}_m$ and $\widehat{T}_m$ are uniquely characterized by*

$$\widehat{\mathcal{L}}_m(\{w,\tilde{w}\},\widehat{T}_m(w,\tilde{w}))=\max_{\tilde{\lambda}\in\mathcal{T}}\min_{\lambda\in\mathcal{T}}\widehat{\mathcal{L}}_m(\{w,\tilde{w}\},\{\lambda,\tilde{\lambda}\}):=\frac{1}{2}J(w,\tilde{w}), \tag{70}$$

$$\widehat{\mathcal{L}}_m(\widehat{S}_m(\lambda,\tilde{\lambda}),\{\lambda,\tilde{\lambda}\})=\max_{\tilde{w}\in\mathcal{V}}\min_{w\in\mathcal{V}}\widehat{\mathcal{L}}_m(\{w,\tilde{w}\},\{\lambda,\tilde{\lambda}\}):=\frac{1}{2}J^{\diamond}(\lambda,\tilde{\lambda}). \tag{71}$$

Definition 22. We introduce the composed mapping

$$\widehat{G}_m(w,\tilde{w})=\widehat{S}_m\circ\widehat{T}_m(w,\tilde{w}) \tag{72}$$

and we let $\widehat{G}^1_m$ and $\widehat{G}^2_m$ be the component mappings of $\widehat{G}_m$, *i.e.* such that $\widehat{G}_m=\left\{\widehat{G}^1_m,\widehat{G}^2_m\right\}$, defined by

$$\widehat{G}^1_m(w,\tilde{w})=S_m(\widehat{T}_m(w,\tilde{w})), \tag{73}$$

$$\widehat{G}^2_m(w,\tilde{w})=\widetilde{S}(S_m(\widehat{T}_m(w,\tilde{w}));\widehat{T}_m(w,\tilde{w})). \tag{74}$$

The following proposition immediately follows from previous definitions.

Proposition 23. *For $\{w,\tilde{w}\}\in\mathcal{V}\times\mathcal{V}$ and $\{\lambda,\tilde{\lambda}\}\in\mathcal{T}\times\mathcal{T}$, the following assertions are equivalent:*

- $\{w,\tilde{w}\}$ and $\{\lambda,\tilde{\lambda}\}$ verify Eqs. (61) and (62)
- $\{w,\tilde{w}\}$ and $\{\lambda,\tilde{\lambda}\}$ verify

$$\{w,\tilde{w}\}=\widehat{S}_m(\lambda,\tilde{\lambda})\quad\text{and}\quad\{\lambda,\tilde{\lambda}\}=\widehat{T}_m(w,\tilde{w}). \tag{75}$$

- $\{\lambda,\tilde{\lambda}\}=\widehat{T}_m(w,\tilde{w})$ and $\{w,\tilde{w}\}$ is a fixed point of mapping $\widehat{G}_m$:

$$\{w,\tilde{w}\}=\widehat{G}_m(w,\tilde{w})\iff w=\widehat{G}^1_m(w,\tilde{w})\quad\text{and}\quad\tilde{w}=\widehat{G}^2_m(w,\tilde{w}). \tag{76}$$

The problem is then to find an optimal fixed point $\{w,\tilde{w}\}\in\mathcal{V}\times\mathcal{V}$ of operator $\widehat{G}_m$ and to define the associated time functions by $\{\lambda,\tilde{\lambda}\}=\widehat{T}_m(w,\tilde{w})$. Such a fixed point verifies

$$\begin{aligned}J(w,\tilde{w})&=2\widehat{\mathcal{L}}_m(\{w,\tilde{w}\},\widehat{T}_m(w,\tilde{w}))\\&=\langle\!\langle wT_m(w,\tilde{w}),wT_m(w,\tilde{w})\rangle\!\rangle,\end{aligned} \tag{77}$$

where functional $J:\mathcal{V}\times\mathcal{V}\to\mathbb{R}^+$ has been defined in (70). Then, thanks to the Definition 18, we search for an optimal fixed point $\{w_m,\tilde{w}_m\}$, where optimality is characterized by

[11] For the mathematical analysis of minimax problems, the reader can refer to [13].

[12] In particular, in the sense of the minimax problem, $w_1\lambda_1$ is the best rank-one separated representation of u, $w_2\lambda_2$ is the best rank-one separated representation of $u-w_1\lambda_1$, etc.

$$J(w_m,\tilde{w}_m)=\max_{\tilde{w}\in\mathcal{V}}\min_{w\in\mathcal{V}}.J(w,\tilde{w}). \tag{78}$$

5.3. Interpretation as a pseudo-eigenproblem

First, let us note that if $\{w,\tilde{w}\}\in\mathcal{V}\times\mathcal{V}$ is a fixed point of $\widehat{G}_m$, then $\{\alpha w,\beta\tilde{w}\}\in\mathcal{V}\times\mathcal{V}$ is also a fixed point of $\widehat{G}_m$, for all $\alpha,\beta\in\mathbb{R}\setminus\{0\}$. This result comes from the following homogeneity properties of mappings $\widehat{G}^1_m$ and $\widehat{G}^2_m:\forall\alpha,\beta\in\mathbb{R}\setminus\{0\}$,

$$\widehat{G}^1_m(\alpha w,\beta\tilde{w})=\alpha\widehat{G}^1_m(w,\tilde{w}),\quad \widehat{G}^2_m(\alpha w,\beta\tilde{w})=\beta\widehat{G}^2_m(w,\tilde{w}). \tag{79}$$

Then, if $\{w,\tilde{w}\}=\widehat{G}_m(w,\tilde{w})$, we have

$$\begin{aligned}\widehat{G}_m(\alpha w,\beta\tilde{w})&=\{\widehat{G}_1(\alpha w,\beta\tilde{w}),\widehat{G}_2(\alpha w,\beta\tilde{w})\}\\&=\left\{\alpha\widehat{G}_1(w,\tilde{w}),\beta\widehat{G}_2(w,\tilde{w})\right\}=\{\alpha w,\beta\tilde{w}\}.\end{aligned}$$

We conjecture that fixed point problem (76) can still be interpreted as a pseudo-eigenproblem. If $\{w,\tilde{w}\}$ is a fixed point of $\widehat{G}_m$, w is interpreted as an eigenfunction, associated with eigenvalue $\min_{w\in\mathcal{V}}J(w,\tilde{w})$. The optimal fixed point $\{w_m,\tilde{w}_m\}$ is associated with the dominant eigenvalue, defined by $J(w_m,\tilde{w}_m)=\max_{\tilde{w}\in\mathcal{V}}\min_{w\in\mathcal{V}}J(w,\tilde{w})$.

This interpretation is motivated by the following analysis of a particular case. In the general case, further mathematical investigations are necessary in order to better understand this non-classical problem.

5.4. The POD as a degenerate case of the Minimax PGD

Let us consider the case where bilinear form B on tensor product space $\mathcal{V}\otimes\mathcal{T}$ can be written as the product of a bilinear form B_S on $\mathcal{V}$ and of a bilinear form B_T on $\mathcal{T}$, *i.e.* $\forall\lambda,\tilde{\lambda}\in\mathcal{T}$ and $\forall w,\tilde{w}\in\mathcal{V}$,

$$B(w\lambda,\tilde{w}\tilde{\lambda})=B_S(w,\tilde{w})B_T(\lambda,\tilde{\lambda}). \tag{80}$$

Let $\mathcal{B}:\mathcal{V}\otimes\mathcal{T}\to\mathcal{V}\otimes\mathcal{T}$, $\mathcal{B}_S:\mathcal{V}\to\mathcal{V}$ and $\mathcal{B}_T:\mathcal{T}\to\mathcal{T}$ be the operators associated with bilinear forms B, B_S and B_T, respectively. Eq. (80) means that operator $\mathcal{B}$ admits an order 1 separated representation $\mathcal{B}=\mathcal{B}_S\otimes\mathcal{B}_T$.

Proposition 24. *If bilinear form B satisfies (80), $\{w,\tilde{w}\}$ is a fixed point of $\widehat{G}_m$ if and only if w is an eigenfunction of eigenproblem:*

$$\sigma w=G_m(w), \tag{81}$$

where operator G_m is defined by $G_m(w)=\langle u-u_{m-1},\langle u-u_{m-1},w\rangle_{\mathcal{V}}\rangle_{\mathcal{T}}$, where $u=\mathcal{B}^{-1}\ell$ is the solution of the problem, and

$$\tilde{w}=\alpha\mathcal{B}_S^{*-1}(w), \tag{82}$$

where $\alpha\in\mathbb{R}\setminus\{0\}$ is an arbitrary scalar and $\mathcal{B}_S^*$ denotes the adjoint operator of $\mathcal{B}_S$. The optimal set of functions, defined by Definition 18, is associated with the dominant eigenvalue of operator G_m.

Proof. Let $e_m=u-u_{m-1}$. The following expressions hold:

$$\widehat{G}^1_m(w,\tilde{w})=B_S(w,\tilde{w})B_T(B_S(e_m,\tilde{w}),B_S(e_m,\tilde{w}))^{-1}B_T(e_m,B_S(e_m,\tilde{w})), \tag{83}$$

$$\widehat{G}^2_m(w,\tilde{w})=B_S(w,\tilde{w})\langle w,w\rangle_{\mathcal{V}}^{-1}\mathcal{B}_S^{*-1}(w). \tag{84}$$

Eq. (76) can be rewritten as follows:

$$w=\widehat{G}^1_m(w,\widehat{G}^2_m(w,\tilde{w})), \tag{85}$$

$$\tilde{w}=\widehat{G}^2_m(w,\tilde{w}), \tag{86}$$

Eq. (85) gives $\sigma_m(w)w=G_m(w)$, where $\sigma_m(w)=\frac{\langle w,G_m(w)\rangle_{\mathcal{V}}}{\langle w,w\rangle_{\mathcal{V}}}$ is the classical Rayleigh quotient associated with the symmetric eigenproblem (81). w then verifies Eq. (85) if and only if it is an eigenfunction of correlation operator G_m. From homogeneity property (79) of operator $\widehat{G}^2_m$ (which is trivially observed in Eqs. (84)), Eq. (86) implies Eq. (82) with an arbitrary scalar $\alpha\in\mathbb{R}\setminus\{0\}$.

Finally, let us prove that the optimal set is associated with the dominant eigenvalue of this problem. We have $T_m(w,\tilde{w})=B_S(w,\tilde{w})^{-1}B_S(e_m,\tilde{w})$ and therefore, functional J, defined in (70), writes:

$$\begin{aligned}J(w,\tilde{w})&=\langle\!\langle wT_m(w,\tilde{w}),wT_m(w,\tilde{w})\rangle\!\rangle\\&=\langle w,w\rangle_{\mathcal{V}}\langle w,\mathcal{B}_S^*(\tilde{w})\rangle_{\mathcal{V}}^{-2}\langle\langle e_m,\mathcal{B}_S^*(\tilde{w})\rangle_{\mathcal{V}},\langle e_m,\mathcal{B}_S^*(\tilde{w})\rangle_{\mathcal{V}}\rangle_{\mathcal{T}}\\&=\langle w,w\rangle_{\mathcal{V}}\langle w,\mathcal{B}_S^*(\tilde{w})\rangle_{\mathcal{V}}^{-2}\langle\mathcal{B}_S^*(\tilde{w}),G_m(\mathcal{B}_S^*(\tilde{w}))\rangle_{\mathcal{V}}.\end{aligned} \tag{87}$$

We then have

$$\min_{w\in\mathcal{V}}J(w,\tilde{w})=J(\mathcal{B}_S^*(\tilde{w}),\tilde{w})=\frac{\langle\mathcal{B}_S^*(\tilde{w}),G_m(\mathcal{B}_S^*(\tilde{w}))\rangle_{\mathcal{V}}}{\langle\mathcal{B}_S^*(\tilde{w}),\mathcal{B}_S^*(\tilde{w})\rangle_{\mathcal{V}}}=\sigma_m(\mathcal{B}_S^*(\tilde{w})), \tag{88}$$

$$\Rightarrow\quad\max_{\tilde{w}\in\mathcal{V}}\min_{w\in\mathcal{V}}J(w,\tilde{w})=\max_{\tilde{w}\in\mathcal{V}}\sigma_m(\mathcal{B}_S^*(\tilde{w}))=\max_{w\in\mathcal{V}}\sigma_m(w) \tag{89}$$

so that the optimal set of functions is associated with the dominant eigenfunction of eigenproblem (81). □

Eq. (81) is a classical deflated version of the eigenproblem (26) which defines the Proper Orthogonal Decomposition associated with the (separated) metric $\langle\!\langle\cdot,\cdot\rangle\!\rangle$. Let w_m denote the rightmost eigenfunction of operator G_m associated with eigenvalue $\sigma_m=\sigma_m(w_m)$. Let $\tilde{w}_m=\mathcal{B}_S^{*-1}(w_m)$ and let $\lambda_m=T_m(w_m,\tilde{w}_m)=\langle w_m,w_m\rangle_{\mathcal{V}}^{-1}\langle u-u_{m-1},w_m\rangle_{\mathcal{V}}$. The couple (w_m,λ_m) is optimal in the sense that it minimizes the error $\|u-u_{m-1}-w\lambda\|^2$, where $\|\cdot\|$ is the norm associated with $\langle\!\langle\cdot,\cdot\rangle\!\rangle$, which is the chosen metric.

With this progressive construction of the decomposition $u_m=\sum_{i=1}^m w_i\lambda_i$, we can classically prove that functions $\{w_i\}_{i=1}^m$ (resp. $\{\lambda_i\}_{i=1}^m$) are orthogonal with respect to inner product $\langle\cdot,\cdot\rangle_{\mathcal{V}}$ (resp. $\langle\cdot,\cdot\rangle_{\mathcal{T}}$). Operator G_m writes

$$G_m(w)=\langle u,\langle u,w\rangle_{\mathcal{V}}\rangle_{\mathcal{T}}-\sum_{i=1}^{m-1}\sigma_i\frac{1}{\langle w_i,w_i\rangle_{\mathcal{V}}}w_i\langle w_i,w\rangle_{\mathcal{V}}. \tag{90}$$

G_m has the same eigenfunctions of operator $G(w)=\langle u,\langle u,w\rangle_{\mathcal{V}}\rangle_{\mathcal{T}}$ and the rightmost eigenvalue σ_m of G_m is the mth rightmost eigenvalue of G (eigenfunctions $\{w_i\}_{i=1}^{m-1}$ of G_m are associated with zero eigenvalues).

The above results prove that the Minimax PGD leads to the same decomposition as the classical POD for rank-one operator $\mathcal{B}=\mathcal{B}_S\otimes\mathcal{B}_T$, even if $\mathcal{B}_S$ and $\mathcal{B}_T$ do not define inner products on $\mathcal{V}$ and $\mathcal{T}$. Let us recall that it was not the case of the classical Galerkin-based PGD introduced in Section 4.1.2. An important point is that the obtained decomposition is optimal with respect to the *a priori* chosen metric $\langle\!\langle\cdot,\cdot\rangle\!\rangle$.

5.5. Power iterations algorithm

We now propose the Algorithm 4 for the progressive construction of the Minimax PGD. Different interpretations can be given to this algorithm. First, it is a fixed point algorithm on operator $\widehat{G}_m=\widehat{S}_m\circ\widehat{T}_m$, which can be interpreted as a power iterations algorithm for the capture of the dominant eigenfunction of the associated pseudo-eigenproblem. In the case where B has the separation property (80), this algorithm exactly coincides with a power iterations algorithm for capturing the dominant eigenfunction of classical eigenproblem (81). Then, in this case, the decomposition u_m obtained with Algorithm 4 corresponds to a classical POD. Starting from the Definition 18, Algorithm 4 can also be seen as an algorithm which alternates the solution of $\arg\max_{\tilde{\lambda}}\min_{\lambda}\widehat{\mathcal{L}}_m(\{w,\tilde{w}\},\{\lambda,\tilde{\lambda}\})=\widehat{T}_m(w,\tilde{w})$ for fixed $\{w,\tilde{w}\}$ and the solution of $\arg\max_{\tilde{w}}\min_{w}\widehat{\mathcal{L}}_m(\{w,\tilde{w}\},\{\lambda,\tilde{\lambda}\})=\widehat{S}_m(\lambda,\tilde{\lambda})$ for fixed $\{\lambda,\tilde{\lambda}\}$.

Algorithm 4 (*(D-PGD-P***) Power iterations algorithm (with update)*).

1: **for** $m = 1$ to m_{max} **do**
2: Initialize $\lambda, \tilde{\lambda}$
3: **for** $k = 1$ to k_{max} **do**
4: Compute $\{w, \tilde{w}\} = \widehat{S}_m(\lambda, \tilde{\lambda})$
5: Normalize w and $\tilde{w}$
6: Compute $\{\lambda, \tilde{\lambda}\} = \widehat{T}_m(w, \tilde{w})$
7: Check convergence of $(w\lambda)$ and $(\tilde{w}\tilde{\lambda})$
8: **end for**
9: Set $w_m = w$ and $\tilde{w}_m = \tilde{w}$
10: (without update) Set $\lambda_m = \lambda$
(with update) Compute $\Lambda_m = T(W_m, \widetilde{W}_m)$
11: Set $u_m = W_m \cdot \Lambda_m$ and check convergence
12: **end for**

The application of mapping $\widehat{S}_m$ (step 4), defined by (68), is decomposed into two steps. First, for given time functions $(\lambda, \tilde{\lambda})$, we compute a space function $w = S_m(\lambda, \tilde{\lambda})$ by solving a classical time-independent PDE. Secondly, we compute a space function $\tilde{w} = \widetilde{S}(w; \lambda, \tilde{\lambda})$ by solving an adjoint time-independent PDE. The application of mapping $\widehat{T}_m$ (step 6), defined by (69), is also decomposed into two steps. First, for given space functions $(w, \tilde{w})$, we compute a time function $\lambda = T_m(w, \tilde{w})$ by solving an ordinary differential equation in time (forward equation in time). Secondly, we compute a time function $\tilde{\lambda} = \widetilde{S}(\lambda; w, \tilde{w})$ by solving an adjoint ordinary differential equation in time (backward equation in time). For computational aspects related to the application of these mappings, see Appendices A and B.

Remark 25. In practice, for the initialization (step 2 of Algorithm 4), we randomly generate λ and we let $\tilde{\lambda} = \lambda$.

Following the arguments of Sections 4.1.4 and 4.1.5, which also hold for Minimax PGD, we propose to include an update of time functions Λ_m (step 10 of Algorithm 4). This update can significantly improve the accuracy of the PGD u_m and often provides a good approximation of the optimal PGD which could be obtained by searching simultaneously the whole set of space and time functions (see Section 6). This updating step consists in applying the mapping $T : (\mathcal{V})^m \times (\mathcal{V})^m \to (\mathcal{T})^m$, which maps space functions $W_m = \{w_i\}_{i=1}^m \in (\mathcal{V})^m$ and $\widetilde{W}_m = \{\tilde{w}_i\}_{i=1}^m \in (\mathcal{V})^m$ into time functions $\Lambda_m = T(W_m, \widetilde{W}_m) \in (\mathcal{T})^m$ defined by

$$B(W_m \cdot \Lambda_m, \widetilde{W}_m \cdot \tilde{\Lambda}_m^*) = L(\widetilde{W}_m \cdot \tilde{\Lambda}_m^*), \quad \forall \tilde{\Lambda}_m^* \in (\mathcal{T})^m. \tag{91}$$

The decomposition $u_m = W_m \cdot T(W_m, \widetilde{W}_m)$ is then a Petrov–Galerkin projection of the solution u, defined by

$$u_m \in \mathcal{V}_m \otimes \mathcal{T}, \quad B(u_m, v_m) = L(v_m), \quad \forall v_m \in \widetilde{\mathcal{V}}_m \otimes \mathcal{T}, \tag{92}$$

where $\mathcal{V}_m = span(W_m)$ and $\widetilde{\mathcal{V}}_m = span(\widetilde{W}_m)$.

6. Numerical examples

Several numerical examples will illustrate and compare the behavior of the different definitions of Proper Generalized Decompositions: Galerkin or minimal residual PGDs, progressive or optimal PGDs. It will also illustrate the behavior of the new Minimax Galerkin PGD.

6.1. Preliminaries

6.1.1. Definitions of PGDs and associated algorithms

For the different definitions of the PGDs u_m, we use the notation

$$(\alpha)\text{PGD} - \beta \quad \text{or} \quad \text{D} - (\alpha)\text{PGD} - \beta. \tag{93}$$

The letter α indicates the initial formulation on which the PGD method is applied: $\alpha = \text{G}$ for Galerkin formulation or $\alpha = \text{R}$ for Minimal Residual formulation. The letter β indicates the algorithm which is used for the construction of the PGD, each algorithm being associated with a different definition of the PGD: $\beta = \text{P}$ for progressive PGD constructed with power iterations algorithm, $\beta = \text{P}^*$ for progressive PGD with update constructed with power iterations algorithm with update, $\beta = \text{S}$ for optimal PGD constructed with subspace iterations algorithm. Finally, the additional letter D before the definition indicates that the Minimax formulation of the PGD is used. For the definition of the Minimax PGD, we use the classical inner product in $L^2(\Omega) \otimes L^2(I)$, defined for $u, v \in L^2(\Omega) \otimes L^2(I)$ by

$$\langle\langle u, v \rangle\rangle = \langle\langle u, v \rangle\rangle_{L^2(\Omega)\otimes L^2(I)} = \int_I \int_\Omega u(x,t)\, v(x,t) dx\, dt.$$

We denote by $\| \cdot \|_{L^2(\Omega)\otimes L^2(I)}$ the associated norm. The Proper Orthogonal Decomposition (POD) of the reference solution is also computed with respect to this classical L^2 metric.

For optimal PGDs, constructed with subspace iterations algorithm, convergence curves will be limited to low orders m because of limited computational resources (limitation due to the application of mapping S). It reflects the fact that this optimal PGD is unaccessible for practical applications.

6.1.2. Parameters of algorithms

Default values for the parameters of algorithms are chosen such that the algorithm captures the associated PGD with a good accuracy: we use a convergence criterium of 10^{-2} for stopping power and subspace iterations and we set the maximum number of iterations k_{max} to 4.

6.1.3. Approximation and error indicator

For approximation space $\mathcal{V}_N \subset \mathcal{V}$, we use classical finite elements or spectral finite elements. For the approximation space $\mathcal{T}_P \subset \mathcal{T}$, we use piecewise polynomials of degree p on a uniform partition of $I = (0, T)$. The reference solution u is the classical Galerkin approximation, which is the solution of (13). We introduce the following error indicator between the reference solution and a separated representation u_m of order m:

$$\varepsilon_m = \frac{\|u - u_m\|_{L^2(\Omega)\otimes L^2(I)}}{\|u\|_{L^2(\Omega)\otimes L^2(I)}}. \tag{94}$$

6.1.4. Computational costs

In these numerical examples, we focus on convergence properties of PGD definitions and not on their relative computational efficiencies. Let us just mention that for constructing a decomposition of a given order m, Galerkin PGD and Minimax Galerkin PGD give similar computational costs (see Appendix A), much lower than with minimal residual PGD. Let us also mention that PGD algorithms share nice computational properties. Computational times are unaffected by the non-uniformity of the time discretization or the time-dependence of the operator, contrary to traditional incremental schemes.

6.2. Example 1: pure diffusion

We consider a pure diffusion problem defined on $\Omega \times I$, with $\Omega = (0,1) \times (0,1)$ and $I = (0,1)$:

$$\dot{u} - \mu \Delta u = f \quad \text{on } \Omega \times I, \tag{95}$$

$$u = 0 \quad \text{on } \partial\Omega \times I, \tag{96}$$

$$u = 0 \quad \text{on } \Omega \times \{0\} \tag{97}$$

with $f(x, y, t) = 1 + 2xt$ and $\mu = 1$. At the space level, we introduce a regular mesh of Ω with triangular linear finite elements and $N = 1354$ nodes. At the time level, we use a degree $p = 0$ piecewise

polynomial approximation on 100 uniform intervals. The reference solution u is shown in Fig. 1.

6.2.1. Analysis of PGDs

Convergence curves of the different PGDs are illustrated in Fig. 2 (Galerkin PGDs in Fig. 2(a) and Minimal Residual PGDs in Fig. 2(b)). We observe that optimal PGDs ((G)PGD-S or (R)PGD-S) are very close to the POD. We observe that progressive Galerkin PGDs have also good convergence properties and that the update significantly improves the convergence. Let us note that in this example, D-(G)PGD has quite the same properties as (G)PGD. Finally, we observe that (R)PGD gives a slower convergence (and also leads to much higher computational costs) than Galerkin PGDs.

Fig. 3 illustrates the first six modes of the POD and of the (G)PGD-P*. We observe that the first 4 modes are very similar. It illustrates the fact that PGD method allows the *a priori* construction of reduced basis which are very similar to the *a posteriori* POD reduced basis. In fact, we generally observe such a similarity between the dominant modes obtained with POD and (G)PGD-P*. However, we clearly observe that the subsequent modes are quite different.

6.2.2. Properties of algorithms

In this example, we briefly analyze the properties of PGD algorithms and motivate the selection of their default parameters. Fig. 4 shows the convergence curves of progressive PGDs (constructed by power iterations algorithms) for different values of parameter k_{max}. In these numerical tests, we choose a very low convergence criterium for power iterations such that we exactly perform k_{max} power iterations. Each PGD algorithm is run five times. Since power iterations are initialized randomly and since we perform a low number of iterations, we should obtain different curves for each PGD, as it can be observed for $k_{max} = 1$. However, we observe that these curves are almost superimposed for $k_{max} = 3$. It reveals that for all m, only three power iterations are sufficient for capturing accurately the dominant couple (w_m, λ_m). Let us note that these results indicate that the PGDs obtained with these algorithms are unaffected by the randomness of the initialization. However, for the progressive (R) PGD without update, we observe that the curves are not exactly superimposed for $k_{max} = 6$. It reveals a slower convergence of power iterations in the case of the Minimal Residual PGD.

6.2.3. Influence of the diffusion parameter

(Fig. 5) shows the influence of the diffusion parameter μ on the convergence properties of POD, (G)PGD and D-(G)PGD. We observe similar behaviors for D-(G)PGD-P* and (G)PGD-P*, with a slight superiority of D-(G)PGD-P* for each value of μ. We also observe that for $\mu \to \pm\infty$, decompositions u_m converges very quickly and tends to be exact with only $m = 2$. In fact, as mentioned in Remark

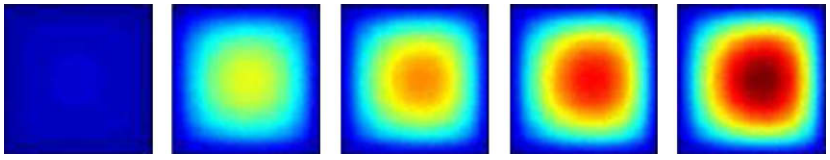


Fig. 1. Example 1. Reference solution $u(t)$ at instants $t = iT/4$, for $i = 0$ (left) to $i = 4$ (right).

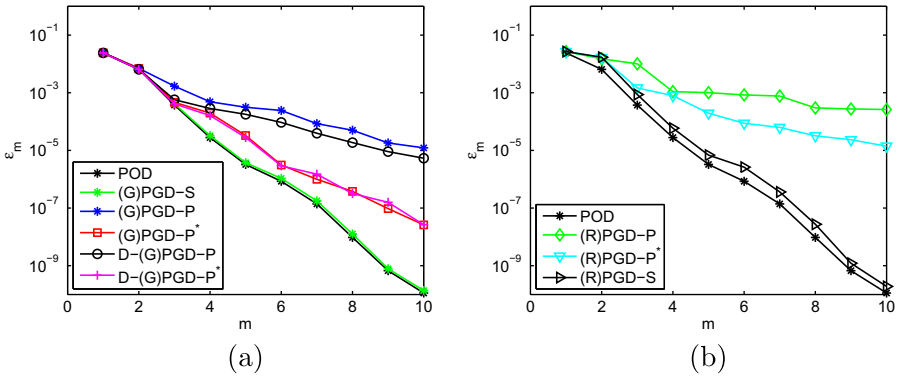


Fig. 2. Example 1. Convergence of PGDs.

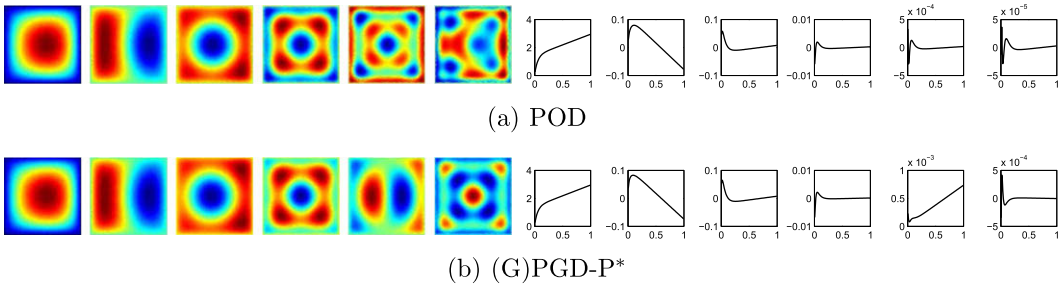
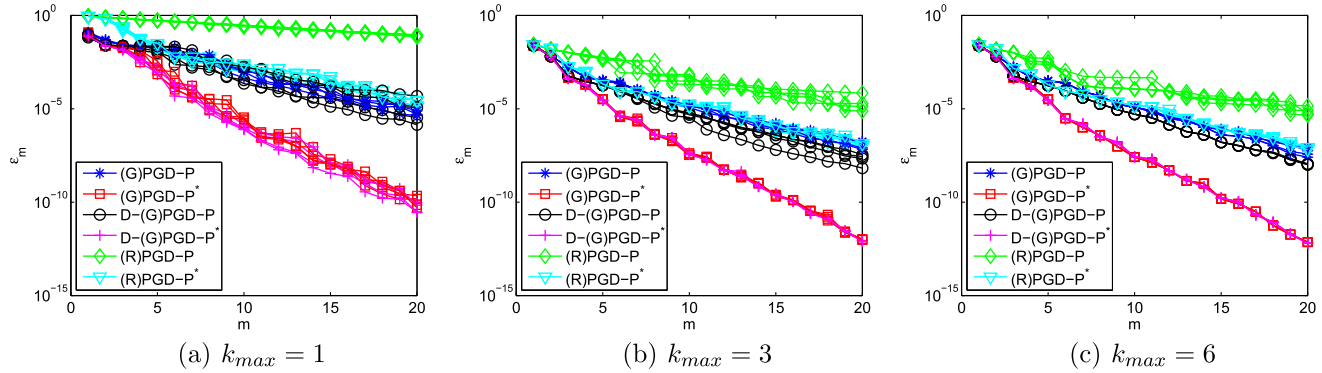
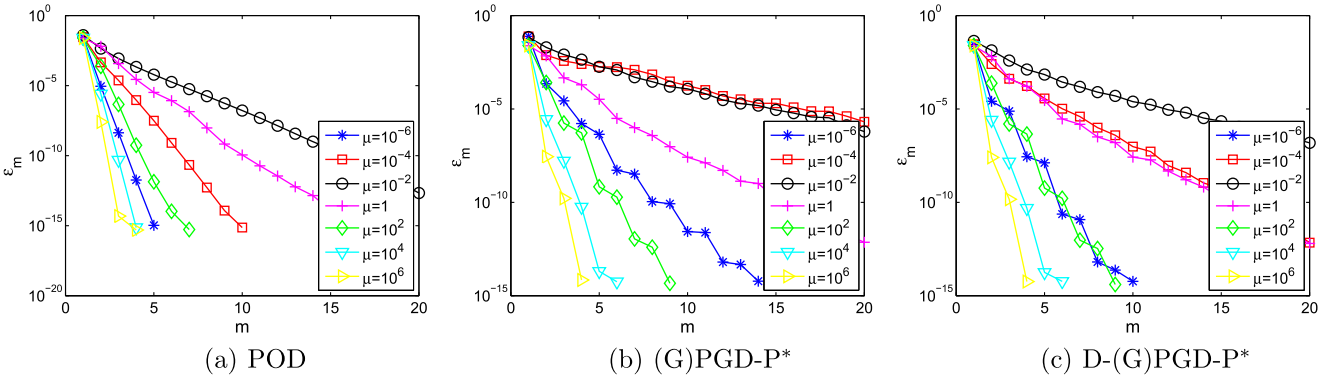


Fig. 3. Example 1. Modes w_i and λ_i of u_m, for $i = 1$ (left) to $i = 6$ (right).



(a) $k_{max} = 1$ (b) $k_{max} = 3$ (c) $k_{max} = 6$

Fig. 4. Example 1. Influence of parameter k_{max} of power iterations algorithms. Each PGD algorithm is run five times. Power iterations are initialized randomly.



(a) POD (b) (G)PGD-P* (c) D-(G)PGD-P*

Fig. 5. Example 1. influence of diffusion parameter μ on the convergence of PGDs and POD.

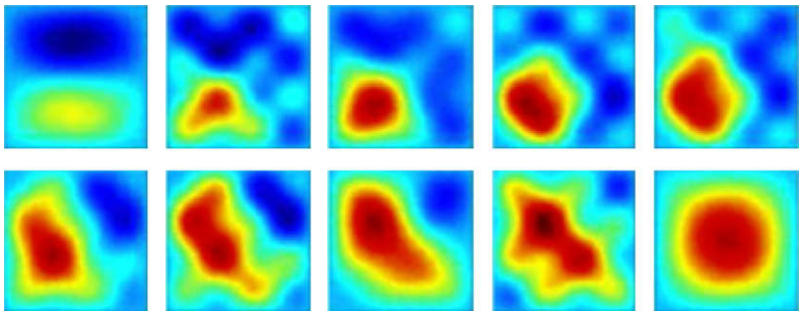


Fig. 6. Example 2. Reference solution $u(t)$ at instants $t = iT/9$, for $i = 0$ (top left) to $i = 9$ (bottom right).

7, that corresponds to two limit cases where the operator is a rank-one operator (*i.e.* having separation property (41)) and therefore, the exact solution admits a separated representation of order 2 which is exactly captured by PGD algorithms (Minimax or not).

6.3. Example 2: pure diffusion, manufactured solution

We consider the same operator and discretization as in Example 1, except for the definition of the right-hand side f. We set $f = \mathcal{B}(u)$, where u is the following manufactured solution (see Fig. 6), which admits an exact separated representation of order 5:

$$u = \sin(\pi x)\sin(\pi y)t + \frac{1}{2}\sin(\pi x)\sin(2\pi y)(1-t) + \frac{1}{2}\sin(2\pi x)\sin(\pi y)\sin(\pi t) + \frac{1}{3}\sin(2\pi x)\sin(2\pi y)\sin(2\pi t) + \frac{1}{5}\sin(4\pi x)\sin(4\pi y)\sin(4\pi t). \quad (98)$$

In Fig. 7, we illustrate the convergence curves of PGDs. In Fig. 7(a) and (b), we observe that optimal PGDs (Galerkin or Minimal Residual) are very close to the POD and that the subspace iterations algorithm allows to capture the exact solution for $m = 5$. We also observe that progressive PGDs do not lead to the exact decomposition in $m = 5$ modes. However, progressive PGDs with update almost capture the exact solution for $m = 5$. In this example, we still observe that the Minimax Galerkin PGD is very similar to the Galerkin PGD.

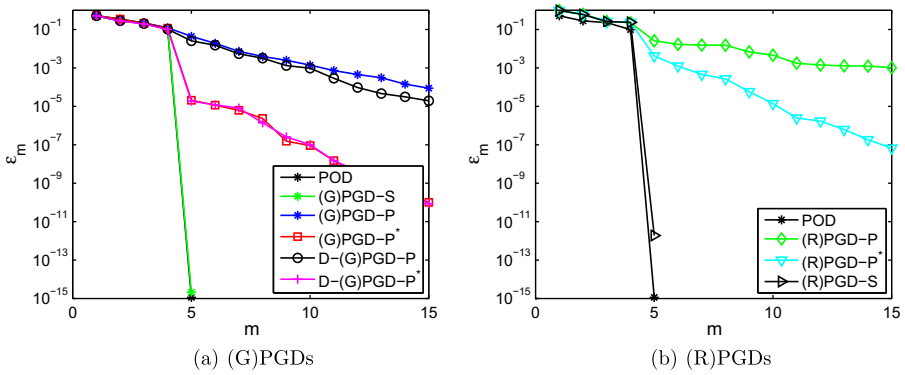
6.4. Example 3: advection–diffusion

We consider the following advection–diffusion problem defined on $\Omega \times I$, with $\Omega = (0,1)\times(0,1)$ and $I = (0,1)$:

$$\dot{u} - \mu\Delta u + c\cdot\nabla u = 0 \quad \text{on } \Omega\times I, \quad (99)$$

$$u = 0 \quad \text{on } \partial\Omega \times I, \quad (100)$$

$$u = u_0 \quad \text{on } \Omega\times\{0\} \quad (101)$$



(a) (G)PGDs (b) (R)PGDs

Fig. 7. Example 2. Manufactured solution with separation order 5. Convergence of PGDs.

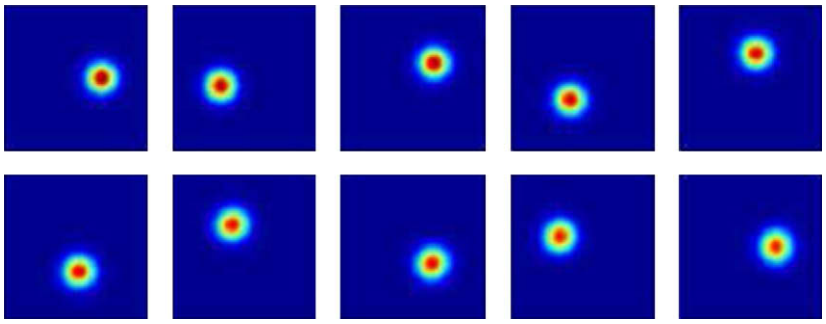
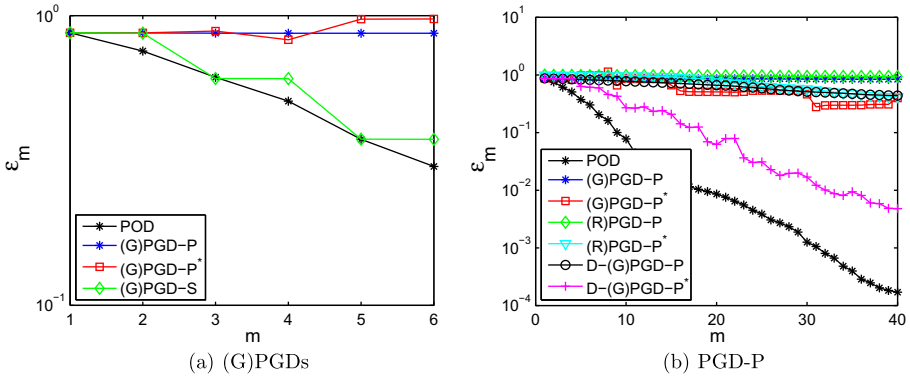


Fig. 8. Example 3. Reference solution $u(t)$ at instants $t = iT/9$, for $i = 0$ (top left) to $i = 9$ (bottom right).



(a) (G)PGDs (b) PGD-P

Fig. 9. Example 3. Convergence of PGDs.

with $\mu = 10^{-3}$, $u_0(x,y) = \exp\left(-\frac{\left(x-\frac{2}{3}\right)^2+\left(y-\frac{1}{2}\right)^2}{0.07^2}\right)$ and $c(x,y,t) = 10\pi(-y+\frac{1}{2}, x-\frac{1}{2})$. At the space level, we use a uniform cartesian mesh with 8×8 quadrangular spectral elements of degree 8. It corresponds to a dimension $N = 3249$ for $\mathcal{V}_N$. At the time level, we use a degree $p = 1$ piecewise polynomial approximation on 300 uniform intervals. The reference solution u is shown in Fig. 8.

6.4.1. Analysis of PGDs

In Fig. 9, we observe the convergence of the different PGDs. Fig. 9(a) illustrates that the optimal Galerkin PGD is still very close to the POD. In Fig. 9(b), we observe that classical progressive PGDs are very far from this optimal decomposition. Very slow convergences are observed for progressive (G)PGD and (R)PGD without update and the convergence is only slightly improved when the update is performed. This example is an illustration where classical PGDs fail at constructing *a priori* a good separated representation u_m. However, this limitation is cured by the newly proposed progressive Minimax Galerkin PGD, which gives a satisfactory convergence when update is performed (D-(G)PGD-P*).

Fig. 10 illustrates the first 12 modes of the POD and progressive PGDs. We observe (as expected) that the POD modes present shorter and shorter length scales. We also observe that (G)PGD-P* and (R)PGD-P* capture a reduced basis of modes which is very different from the one obtained with POD (and far less pertinent). However, we observe that D-(G)PGD-P* allows the capture of reduced bases of modes presenting almost the same spatial and temporal features as the POD reduced basis.

6.4.2. Properties of algorithms

As done for the example in Section 6.2, we here briefly study the behavior of PGD algorithms for different values of parameter k_{max}

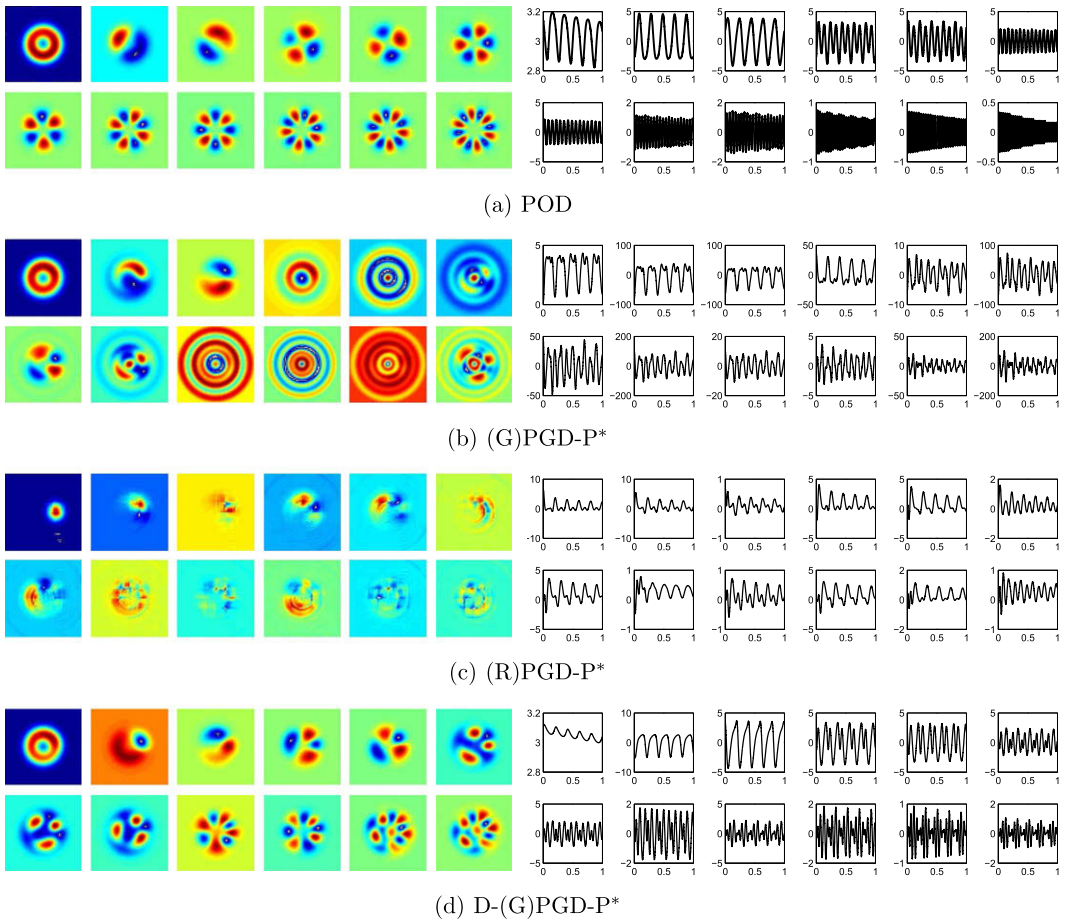


Fig. 10. Example 3. Modes w_i and λ_i of u_m, for $i = 1$ (top left) to $i = 12$ (bottom right).

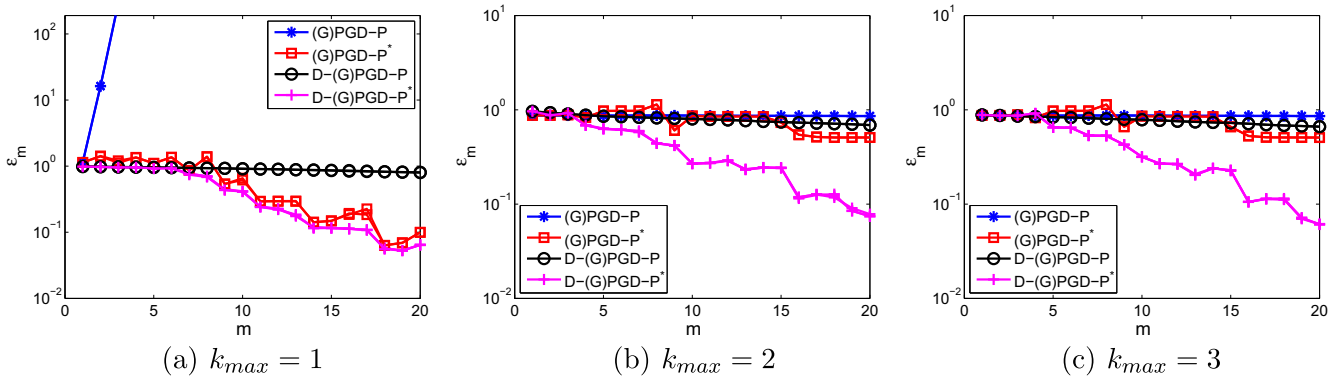


Fig. 11. Example 3. Influence of parameter k_{max} on power iterations algorithms. Each PGD algorithm is run three times. Power iterations are initialized randomly.

(maximum number of power iterations). Fig. 11 indicates that for each progressive PGD algorithm, a very low number of power iterations ($k_{max} \approx 3$) is sufficient to capture the dominant couple (w_m, λ_m). However, we note that for (G)PGD-P (without update of time functions), a coarse approximation of the dominant couple (with $k_{max} = 1$) leads to a divergence of the decomposition. (G)PGD-P* (with update) does not diverge but presents a strange property. Indeed, the decomposition u_m obtained with $k_{max} = 1$ (coarse approximation of the dominant couples) is better than the converged (G)PGD-P*. The Minimax PGD seems to circumvent these robustness issues.

6.4.3. Influence of the diffusion parameter

We now consider problems with different ratios between diffusion and advection terms, from pure advection ($\mu = 0$) to high diffusion ($\mu = 10^{-1}$). In Fig. 12(a), we observe that in order to obtain an accurate POD of the solution, more and more modes are required when decreasing the diffusion, which reveals a richer spectral content of the solution for low diffusion problems. In Fig. 12(b), we observe that for high diffusion problems, (G)PGD-P* works but converges slowly. For zero diffusion, (G)PGD-P* has a very slow convergence. Finally, In Fig. 12(c), we observe that the Minimax Galerkin PGD works pretty well regarded to

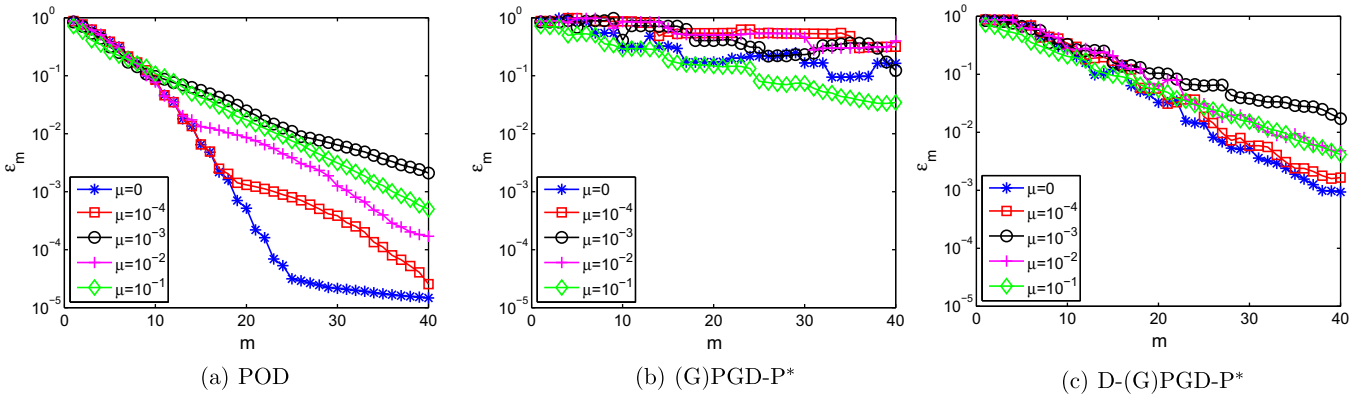


Fig. 12. Example 3. Influence of the diffusion μ on the convergence of POD and progressive PGDs.

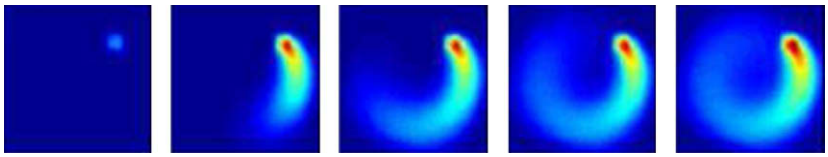


Fig. 13. Example 4. Reference solution $u(t)$ at instants $t = iT/4$, for $i = 0$ (left) to $i = 4$ (right).

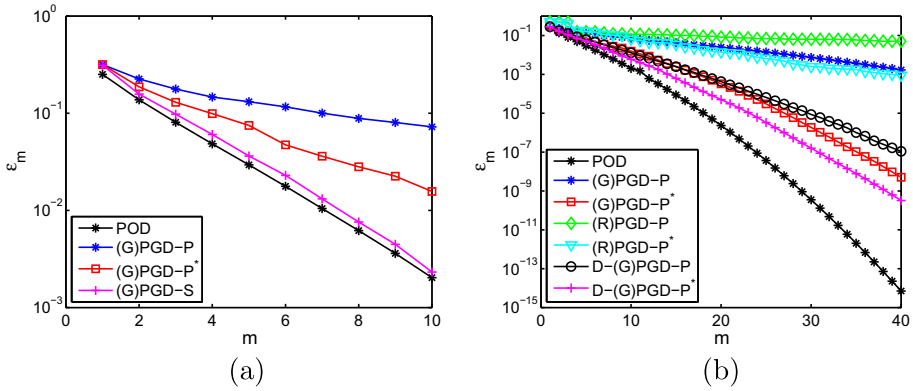


Fig. 14. Example 4. Convergence of PGDs.

the expected optimal decomposition (the POD), whatever the diffusion value.

6.5. Example 4: advection–diffusion–reaction

We consider the following advection–diffusion–reaction problem, taken from [33], defined on $\Omega \times I$, with $\Omega = (0,1) \times (0,1)$ and $I = (0, 0.03)$:

$$\dot{u} - \mu\Delta u + c \cdot \nabla u + \sigma u = f \quad \text{on } \Omega \times I, \tag{102}$$

$$u = 0 \quad \text{on } \partial\Omega \times I, \tag{103}$$

$$u = 0 \quad \text{on } \Omega \times \{0\} \tag{104}$$

with $\mu = 1$, $\sigma = 10$, $c = 250(y - \frac{1}{2}, \frac{1}{2} - x)$ and $f(x,y,t) = 100 I_{\Omega_1}(x,y)$, where I_{Ω_1} is the indicator function of a subdomain $\Omega_1 = (0.7, 0.8) \times (0.7, 0.8)$. At the space level, we use a regular mesh of Ω with triangular linear finite elements and $N = 2774$ nodes. At the time level, we use a degree $p = 0$ piecewise polynomial approximation on 100 uniform intervals. The reference solution u is shown in Fig. 13.

6.5.1. Analysis of PGDs

On Fig. 14, we observe the convergence of the different PGDs. Fig. 14(a) illustrates that the optimal Galerkin PGD is very close to the POD. In Fig. 14(b), we first observe that progressive PGDs based on a Minimal Residual formulation give poor convergence properties (with or without update), such as the progressive Galerkin PGD without update. We also observe that the best convergence properties are obtained with the Minimax Galerkin PGD. We notice that D-(G)PGD-P (without update) is similar to the

Table 1
Example 4. Minimal order m such that $\epsilon_m < 10^{-2}$ for $\alpha) - \beta$ PGDs.

	(G)-S	(G)-P	(G)-P*	(R)-P	(R)-P*	D-(G)-P	D-(G)-P*
m	8	28	12	>100	23	11	9

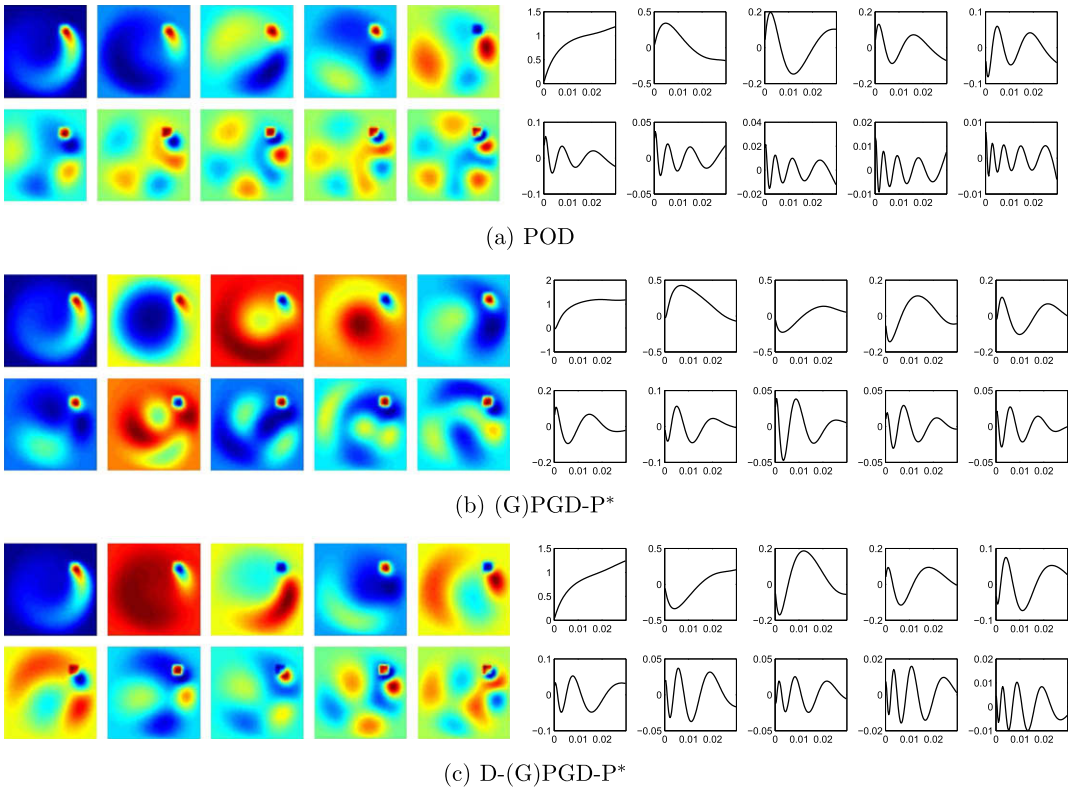
classical (G)PGD-P* (with update), which is a very nice property of the Minimax PGD.

Table 1 indicates for each PGD the order m (or dimension of reduced basis) required to obtain a desired precision $\epsilon_m = 10^{-2}$. It illustrates that the Minimax PGD leads to the construction of quasi-optimal reduced basis.

Fig. 15 illustrates the first 10 modes of the POD and of progressive PGDs. We observe that the modes of the progressive Minimax Galerkin PGD are very close from the ones of POD. It illustrates that this Minimax PGD extracts more rapidly the spectral content of the solution, compared to other PGDs.

6.6. Example 5: advection–diffusion–dispersion

We consider the following advection–diffusion–dispersion problem defined on $\Omega \times I$, with $\Omega = (0,1) \times (0,1)$ and $I = (0,1)$:



(a) POD

(b) (G)PGD-P*

(c) D-(G)PGD-P*

Fig. 15. Example 4. Modes w_i and λ_i of u_m, for $i = 1$ (top left) to $i = 10$ (bottom right).

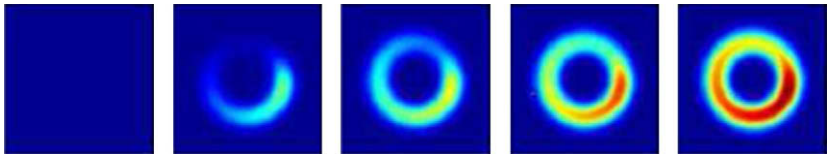


Fig. 16. Example 5. Reference solution $u(t)$ at instants $t = iT/4$, for $i = 0$ (left) to $i = 4$ (right).

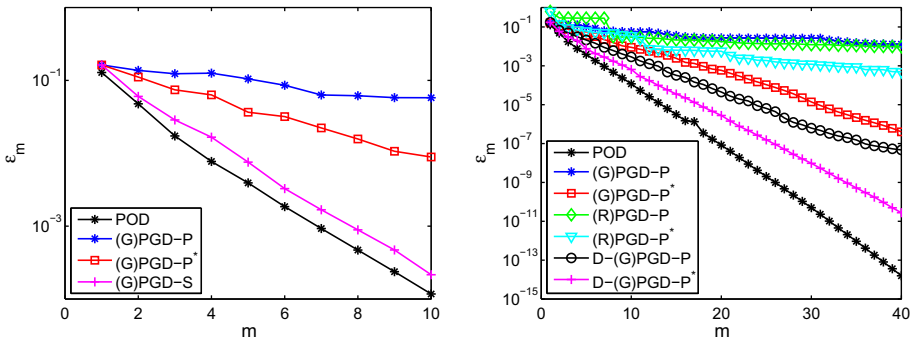


Fig. 17. Example 5. Convergence of PGDs.

$$\dot{u} - \nabla(D\nabla u) + c \cdot \nabla u = 0 \quad \text{on } \Omega \times I, \tag{105}$$

$$u = 0 \quad \text{on } \partial\Omega \times I, \tag{106}$$

$$u = 0 \quad \text{on } \Omega \times \{0\}, \tag{107}$$

where $c(x,y,t) = 3\pi\left(-y + \frac{1}{2}, x - \frac{1}{2}\right)$, $f(x,y) = \exp\left(-\frac{\left(x-\frac{3}{4}\right)^2 + \left(y-\frac{1}{2}\right)^2}{0.07^2}\right)$ and where D is a dispersion–diffusion tensor defined by

$$D = \mu I_d + \|c\|(\alpha_l \tilde{c} \otimes \tilde{c} + \alpha_t(I_d - \tilde{c} \otimes \tilde{c})), \quad \tilde{c} = \frac{c}{\|c\|}, \tag{108}$$

where I_d is the identity tensor, $\mu = 10^{-4}$ is the effective diffusion, $\alpha_l = 10^{-1}$ is the longitudinal dispersion and $\alpha_t = 10^{-4}$ is the transversal dispersion. At the space level, we use a regular mesh of Ω with triangular linear finite elements and $N = 5127$ nodes. At the time level, we use a degree $p = 1$ piecewise polynomial approximation on 100 uniform intervals. The reference solution u is shown in Fig. 16.

Table 2
Example 5. Minimal order m such that $\epsilon_m < 10^{-3}$ for $(\alpha) - \beta$ PGDs.

	(G)-S	(G)-P	(G)-P*	(R)-P	(R)-P*	D-(G)-P	D-(G)-P*
m	8	>150	18	>150	33	13	10

6.6.1. Analysis of PGDs

In Fig. 17, we observe the convergence of the different PGDs. Fig. 17(a) illustrates that the optimal Galerkin PGD is still very close to the POD. In Fig. 14(b), we note that progressive PGDs based on a Minimal Residual formulation give poor convergence properties, such as the progressive Galerkin PGD without update. With the update, D-(G)PGD-P* leads to a decomposition which is close to the optimal one. We also notice that D-(G)PGD-P (without update) is better than the classical (G)PGD-P* (with update). It is a very nice property of the Minimax PGD.

Table 2 indicates for each PGD the order m required to obtain a desired precision $\epsilon_m = 10^{-3}$. It reveals that for classical progressive Galerkin or Minimal Residual PGDs, the updating of time functions is necessary to obtain satisfactory reduced basis. However, the Minimax PGD leads to the construction of quasi-optimal reduced basis with or without update of time functions. For this desired precision, D-(G)PGD-P* allows to construct a reduced order model with a dimension two times lower than (G)PGD-P* and three times lower than (R)PGD-P*.

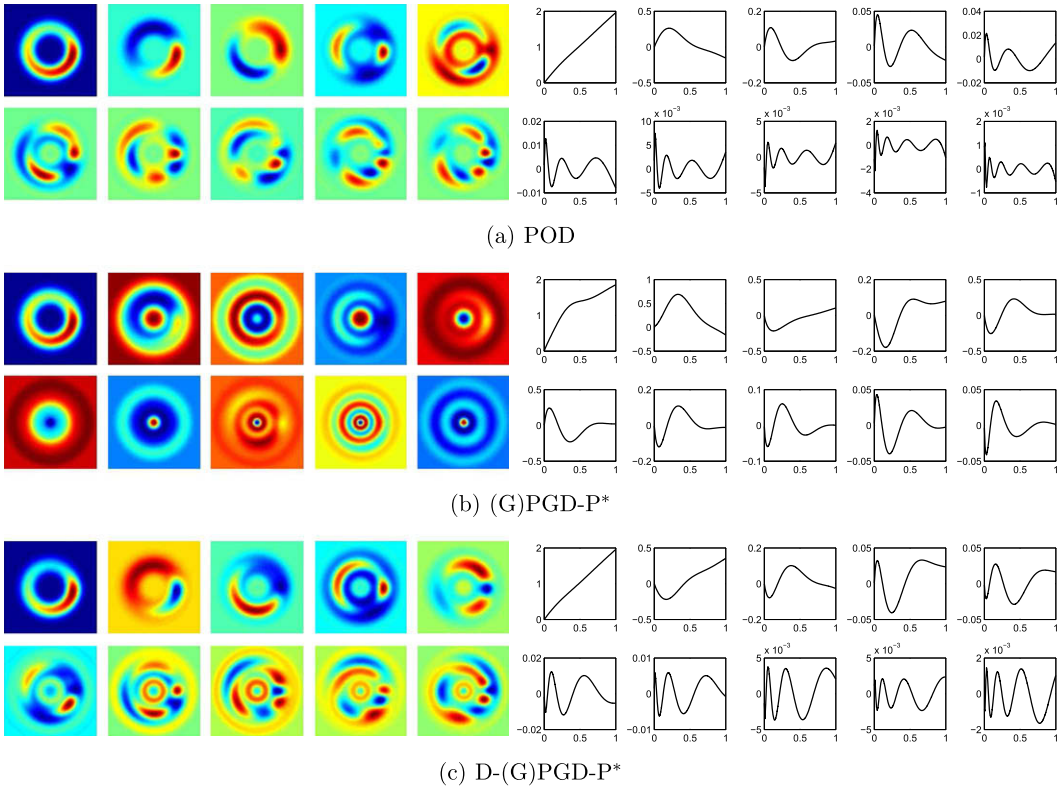


Fig. 18. Example 5. Modes w_i and λ_i of u_m, for $i = 1$ (top left) to $i = 10$ (bottom right).

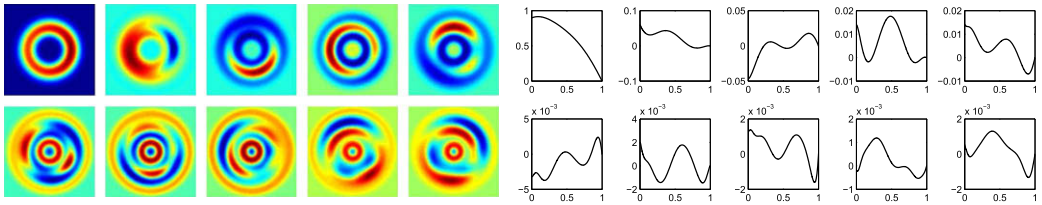


Fig. 19. Example 5. Modes $\bar{w}_i$ and $\bar{\lambda}_i$ of D-(G)PGD-P*, for $i = 1$ (top left) to $i = 10$ (bottom right).

Fig. 18 illustrates the first 10 modes of the POD and of the progressive PGDs. We still observe that the modes of the progressive Minimax Galerkin PGD are very close from the ones of POD. Classical (G)PGD-P* clearly fails at capturing accurately the upper spectrum of the solution. Fig. 19 illustrates the first 10 dual modes $(\tilde{w}_i, \tilde{\lambda}_i)$ of D-(G)PGD-P*.

6.7. Example 6: canister

This example is taken from [33] and represents the transport of pollutant inside an active carbon filter. The concentration of pollutant u satisfies the following advection–diffusion–reaction problem defined on $\Omega \times I$, with $I = (0, 2)$ and $\Omega = \Omega_1 \cup \Omega_2$ (see Fig. 20(a)):

$$\dot{u} - \mu \Delta u + c \cdot \nabla u + \sigma u = 0 \quad \text{on } \Omega \times I \tag{109a}$$

with $\mu = 0.01$, $\sigma = 0.01 I_{\Omega_1} + 10 I_{\Omega_2}$. The advection field $c = \nabla \psi$, where ψ is a potential which is obtained by solving a Laplace equation on Ω. For boundary conditions, we impose $u = 1$ on Γ_1 and $u = 0$ on Γ_2. On the complementary part of the boundary $\partial\Omega \setminus (\Gamma_1 \cup \Gamma_2)$, we impose homogeneous Neumann boundary conditions. The initial condition u_0 is such that $u_0 = 1$ on Γ_1 and 0 elsewhere in Ω.

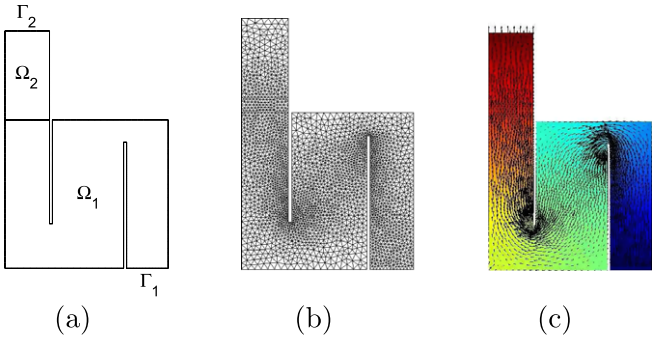


Fig. 20. Example 6. Geometry (a), finite element mesh (b) and potential flow c (c).

At the space level, we use a regular mesh of Ω with triangular linear finite elements and $N = 2826$ nodes (Fig. 20(b)). At the time level, we use a degree $p = 1$ piecewise polynomial approximation on 100 uniform intervals. The potential ψ is computed on the same finite element mesh. The solution ψ, and the corresponding advection flow $c = \nabla \psi$, are shown in Fig. 20(c). The reference solution u is shown in Fig. 21.

6.7.1. Analysis of PGDs

In Fig. 22, we observe the convergence of the different PGDs. As in the other examples, Fig. 22(a) illustrates that the optimal Galerkin PGD is still very close to the POD. Fig. 22(b) illustrates the superiority of the Minimax strategy for the progressive construction of PGD. Indeed, D-(G)PGD-P and D-(G)PGD-P* lead to very similar decompositions, close to the optimal one. Let us note that the fact that D-(G)PGD-P (without update) gives an almost optimal

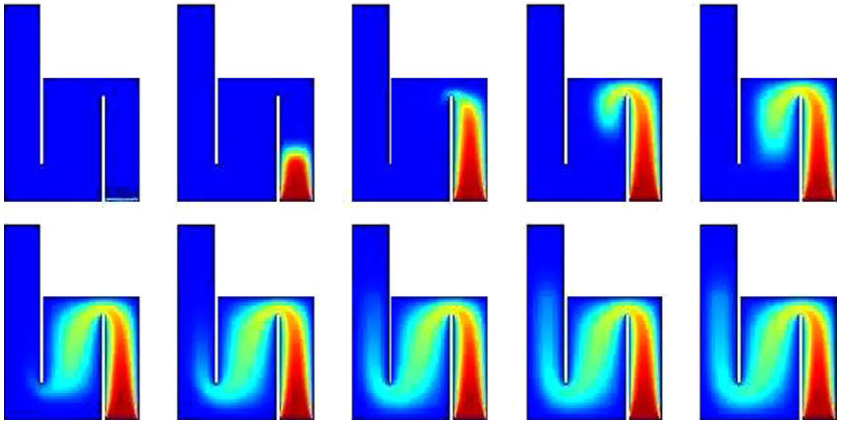


Fig. 21. Example 6. Reference solution $u(t)$ at instants $t = iT/9$, for $i = 0$ (top left) to $i = 9$ (bottom right).

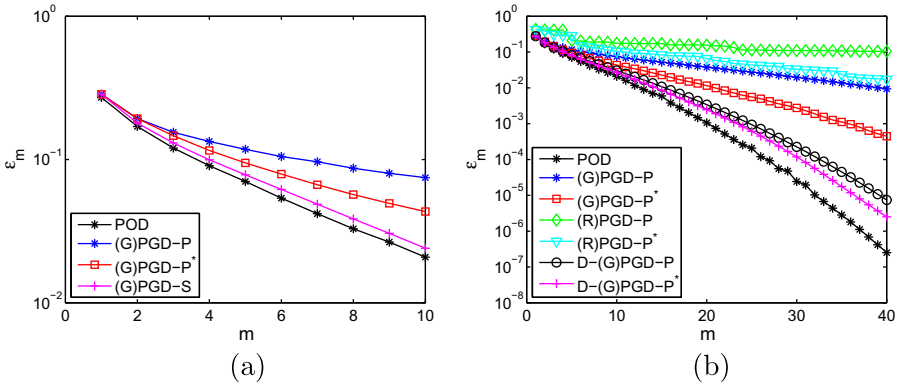


Fig. 22. Example 6. Convergence of PGDs.

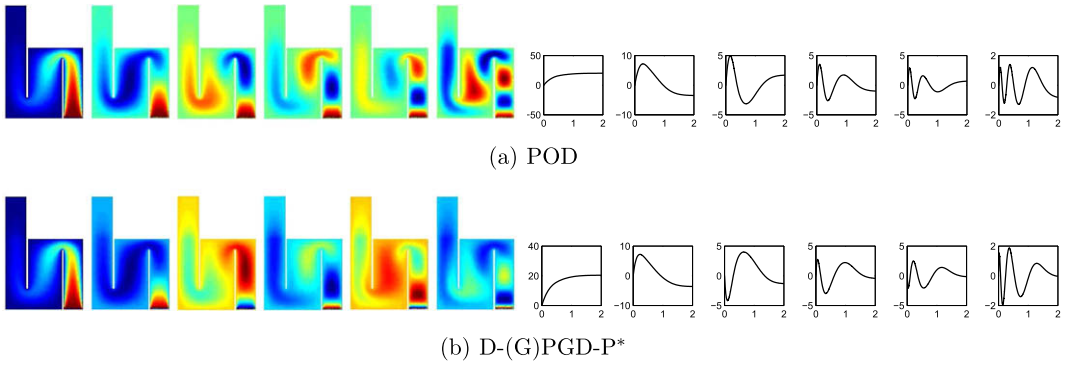


Fig. 23. Example 6. Modes w_i and λ_i of u_m, for $i = 1$ (left) to $i = 6$ (right).

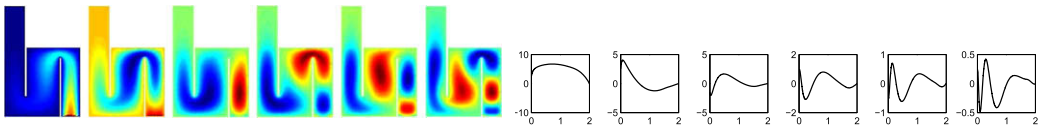


Fig. 24. Example 6. Modes $\tilde{w}_i$ and $\tilde{\lambda}_i$ of D-(G)PGD-P*, for $i = 1$ (left) to $i = 6$ (right).

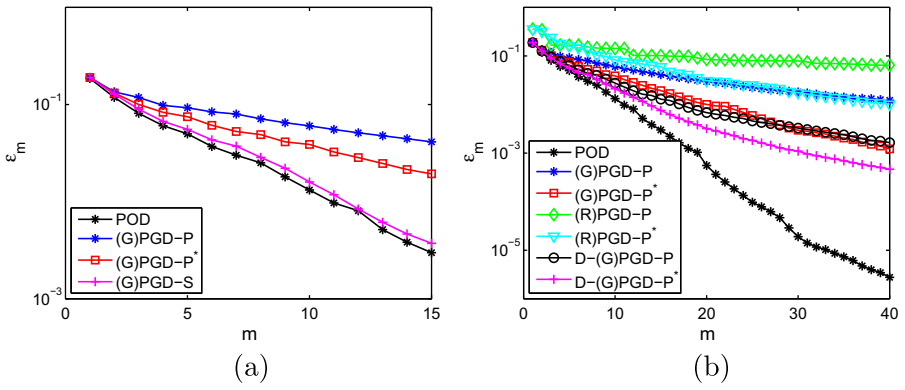


Fig. 25. Example 6. Time-dependent advection. Convergence of PGDs.

decomposition is a very nice property of the Minimax PGD. The progressive PGDs based on a Minimal Residual formulation give poor convergence properties, such as the progressive Galerkin PGD without update. Classical (G)PGD-P* (with update) gives intermediate results.

Let us emphasize that the PGD method allows to build *a priori* a very low dimensional reduced order model which represents very accurately the solution (precision 10^{-3} with a dimension $m \approx 30$ for D-(G)PGD-P* and $m \approx 50$ for (G)PGD-P*).

Fig. 23 illustrates the first six modes of the POD and D-(G)PGD-P*. Fig. 24 illustrates the first six dual modes $(\tilde{w}_i, \tilde{\lambda}_i)$ of D-(G)PGD-P*.

6.7.2. Time-dependent advection flow

We consider the same problem as previously except for $I = (0, 4)$ and the advection flow c which is now time-dependent. We take $c(t) = \alpha(t)c_0$ where c_0 is the previously used time-independent advection velocity (shown in Fig. 20(a)), and $\alpha(t) = 2 - t(1 - t/4)$. Fig. 25 still illustrates the superiority of the Minimax PGD for the progressive definition of the decomposition.

7. Conclusion

We have introduced different possible definitions of Proper Generalized Decompositions (PGD) for the *a priori* construction of separated variables representation of the solution of time-dependent PDEs. For each definition of PGD, a dedicated algorithm has been proposed. The different definitions are based on a Galerkin or a Minimal Residual formulation of the evolution problem. For each formulation, different variants of PGD have been proposed, which correspond to progressive or simultaneous (optimal) constructions of the decomposition. We have also proposed an innovative definition of PGD, called Minimax PGD, which can be interpreted as a Petrov–Galerkin model reduction technique. This new PGD preserves the computational advantages of Galerkin PGD (compared to Minimal Residual PGD) and improves the convergence properties of the decomposition with respect to a chosen metric. The numerical examples have illustrated that an optimal PGD, constructed by a subspace iterations algorithm, have quite the same convergence properties as a classical *a posteriori* POD. However, in the context of the solution of time-dependent PDEs,

an optimal PGD is not of practical interest since its construction requires the solution of systems of coupled PDEs whose computational costs are prohibitive for large scale applications. Then, progressive PGDs, which are of practical interest, have been compared to this optimal PGD or to an optimal *a posteriori* POD. We have illustrated that the progressive Galerkin PGD often gives good convergence properties, better than the progressive minimal residual PGD, which is more robust but also more computationally expansive. We have also illustrated that the newly proposed Minimax Galerkin PGD (or Petrov–Galerkin PGD) significantly improves the convergence properties of classical progressive Galerkin PGDs. In particular, for some problems, we have observed that the progressive Minimax PGD allows the *a priori* construction of a decomposition close to an optimal POD. Significant computational savings can then be expected for large scale applications.

PGD methods constitute promising alternative computational techniques for solving a large class of problems defined in tensor product spaces. In the context of time-dependent PDEs, computational costs are not affected by the time-dependence of the operator or the non-uniformity of the time discretization. However, further mathematical analyses and numerical experiments are still necessary. For some problems where an *a posteriori* POD tells us that a few modes are sufficient to describe the solution, the different PGD definitions proposed in this article may fail at capturing these modes. A better understanding of the underlying pseudo-eigenproblem could allow us to propose improved progressive PGD definitions and more efficient algorithms for the capture of quasi-optimal decompositions.

Future works will be devoted to the extension of the newly proposed PGD for the construction of multi-dimensional separated representations. Another important perspective concerns the introduction of robust error indicators for the adaptive construction of PGD decompositions.

Acknowledgement

This work is supported by the French National Research Agency (Grant ANR-06-JCJC-0064) and by GdR MoMaS with partners ANDRA, BRGM, CEA, CNRS, EDF, IRSN.

Appendix A. Computational aspects of Proper Generalized Decomposition algorithms

A.1. Preliminaries

We consider that bilinear and linear forms a and ℓ in Eq. (3b) admit the following separated representations: $\forall w, w^* \in \mathcal{V}$

$$a(w,w^*;t)=\sum_{i=1}^{m_a}\alpha_i(t)a_i(w,w^*),\quad \ell(w^*;t)=\sum_{i=1}^{m_\ell}\gamma_i(t)\ell_i(w^*), \tag{A.1}$$

where a_i and ℓ_i are time-independent bilinear and linear forms on $\mathcal{V}$, and where α_i and γ_i are real-valued time functions. For convenience, we introduce the following notations:

$$a_0(w,w^*)=m(w,w^*),\quad \alpha_0=1,\quad \ell_0(w^*)=m(u_0,w^*),\quad \gamma_0=1. \tag{A.2}$$

If a space discretization $\mathcal{V}_N\subset\mathcal{V}$ is introduced (see Section 2.5), matrix $\mathbf{A}(t)$ and vector $\mathbf{f}(t)$, defined in Section 2.5, admit the following separated representations:

$$\mathbf{A}(t)=\sum_{i=1}^{m_a}\alpha_i(t)\mathbf{A}_i,\quad \mathbf{f}(t)=\sum_{i=1}^{m_\ell}\gamma_i(t)\mathbf{f}_i \tag{A.3}$$

and for convenience, we define the matrix $\mathbf{A}_0=\mathbf{M}$ and the vector $\mathbf{f}_0=\mathbf{M}\mathbf{u}_0$, which are respectively associated with bilinear form a_0 and linear form ℓ_0. We introduce the following notations: for $v,\tilde{v}\in\mathcal{T}_P$,

$$(v,\tilde{v})_I=\int_I v(t)\tilde{v}(t)dt, \tag{A.4}$$

$$\begin{aligned}(\dot{v},\tilde{v})_I&=\int_I \dot{v}(t)\tilde{v}(t)dt+v(0^+)\tilde{v}(0^+) \qquad (A.5)\\ &=\sum_{k=1}^{r}\int_{I_k}\dot{v}(t)\tilde{v}(t)dt+\sum_{k=1}^{r-1}\left(v(t_k^+)-v(t_k^-)\right)\tilde{v}(t_k^+)\\ &\quad+v(0^+)\tilde{v}(0^+). \qquad (A.6)\end{aligned}$$

A.2. Application of mappings

A.2.1. Computing $w=S_m(\lambda,\tilde{\lambda})$ or $w=S_m(\lambda)$ (spatial problem)

First of all, let us note that mapping $S_m(\lambda)$ associated with the Galerkin PGD (defined by Eq. (33)) corresponds to mapping S_m of Minimax Galerkin PGD (defined by (61a)) for $\tilde{\lambda}=\lambda$, *i.e.* $S_m(\lambda)\equiv S_m(\lambda,\lambda)$. Let $u_{m-1}(x,t)=\sum_{i=1}^{m-1}w_i(x)\lambda_i(t)$. For given $\lambda,\tilde{\lambda}\in\mathcal{T}$, $w=S_m(\lambda,\tilde{\lambda})$ is the solution of the following problem:

$$B_\lambda(w,w^*)=L_\lambda^m(w^*),\quad \forall w^*\in\mathcal{V}, \tag{A.7}$$

where B_λ and L_λ^m are bilinear and linear forms on $\mathcal{V}$ defined by

$$B_\lambda(w,w^*)=B(w\lambda,w^*\tilde{\lambda})=\sum_{i=0}^{m_a}\alpha_i^\lambda a_i(w,w^*), \tag{A.8}$$

$$\begin{aligned}L_\lambda^m(w^*)&=L(w^*\tilde{\lambda})-B(u_{m-1},w^*\tilde{\lambda})\\ &=\sum_{i=0}^{m_\ell}\gamma_i^\lambda\ell_i(w^*)-\sum_{i=0}^{m_a}\sum_{j=1}^{m-1}\alpha_i^{\lambda,j}a_i(w_j,w^*),\end{aligned} \tag{A.9}$$

where

$$\alpha_0^\lambda=(\dot{\lambda},\tilde{\lambda})_I,\quad \alpha_i^\lambda=(\alpha_i\lambda,\tilde{\lambda})_I\quad \text{for } i=1\dots m_a, \tag{A.10}$$

$$\gamma_0^\lambda=(\dot{\gamma}_0,\tilde{\lambda})_I=\tilde{\lambda}(0^+),\quad \gamma_i^\lambda=(\gamma_i,\tilde{\lambda})_I\quad \text{for } i=1\dots m_\ell, \tag{A.11}$$

$$\alpha_0^{\lambda,j}=(\dot{\lambda}_j,\tilde{\lambda})_I,\quad \alpha_i^{\lambda,j}=(\alpha_i\lambda_j,\tilde{\lambda})_I\quad \text{for } i=1\dots m_a,\ j=1\dots m-1. \tag{A.12}$$

When using a spatial discretization, problem (A.7) can be rewritten as the following system of equations: find $\mathbf{w}\in\mathbb{R}^N$ such that

$$\left(\sum_{i=0}^{m_a}\alpha_i^\lambda\mathbf{A}_i\right)\mathbf{w}=\sum_{i=0}^{m_\ell}\gamma_i^\lambda\mathbf{f}_i-\sum_{i=0}^{m_a}\sum_{j=1}^{m-1}\alpha_i^{\lambda,j}\mathbf{A}_i\mathbf{w}_j. \tag{A.13}$$

A.2.2. Computing $\lambda=T_m(w,\tilde{w})$ or $\lambda=T_m(w)$ (time problem)

First of all, let us note that mapping $T_m(w)$ associated with the Galerkin PGD (defined by Eq. (34)) corresponds to mapping $T_m(w,\tilde{w})$ of Minimax Galerkin PGD (defined by (61b)) for $\tilde{w}=w$, *i.e.* $T_m(w)\equiv T_m(w,w)$. Let $u_{m-1}(x,t)=\sum_{i=1}w_i(x)\lambda_i(t)$. For given $w,\tilde{w}\in\mathcal{V}$, $\lambda=T_m(w,\tilde{w})$ is the solution of the following problem:

$$B_w(\lambda,\lambda^*)=L_w^m(\lambda^*),\quad \forall\lambda^*\in\mathcal{T}, \tag{A.14}$$

where B_w and L_w^m are bilinear and linear forms on $\mathcal{T}$ defined by

$$B_w(\lambda,\lambda^*)=B(w\lambda,\tilde{w}\lambda^*)=\alpha_0^w(\dot{\lambda},\lambda^*)_I+\sum_{i=1}^{m_a}(\alpha_i^w\lambda,\lambda^*)_I, \tag{A.15}$$

$$\begin{aligned}L_w^m(\lambda^*)&=L(\tilde{w}\lambda^*)-B(u_{m-1},\tilde{w}\lambda^*)=\sum_{i=1}^{m_\ell}(\gamma_i^w,\lambda^*)_I+\gamma_0^w\lambda^*(0^+)\\ &\quad-\sum_{j=1}^{m-1}\alpha_0^{w,j}(\dot{\lambda}_j,\lambda^*)_I-\sum_{i=1}^{m_a}\sum_{j=1}^{m-1}(\alpha_i^{w,j}\lambda_j,\lambda^*)_I,\end{aligned} \tag{A.16}$$

where

$$\alpha_i^w(t) = \alpha_i(t) a_i(w, \tilde{w}) \quad \text{for } i = 0 \dots m_a, \tag{A.17}$$

$$\gamma_i^w(t) = \gamma_i(t) \ell_i(\tilde{w}) \quad \text{for } i = 0 \dots m_\ell, \tag{A.18}$$

$$\alpha_i^{w,j} = \alpha_i(t) a_i(w_j, \tilde{w}) \quad \text{for } i = 0 \dots m_a, \ j = 1 \dots m-1. \tag{A.19}$$

Problem Eq. (A.14) then appears as a time weak formulation of the following ordinary differential equation: find $\lambda(t)$ such that

$$\alpha_0^w \dot{\lambda}(t) + \left(\sum_{i=1}^{m_a} \alpha_i^w(t) \right) \lambda(t) = \sum_{i=1}^{m_\ell} \gamma_i^w(t) - \sum_{j=1}^{m-1} \alpha_0^{w,j} \dot{\lambda}_j(t) - \sum_{i=1}^{m_a} \sum_{j=1}^{m-1} \alpha_i^{w,j}(t) \lambda_j(t) \tag{A.20}$$

$$\alpha_0^w \lambda(0) = \gamma_0^w - \sum_{j=1}^{m-1} \alpha_0^{w,j} \lambda_j(0) \tag{A.21}$$

The initial condition writes $m(w\lambda(0), \tilde{w}) = m\left(u_0 - \sum_{j=1}^{m-1} w_j \lambda_j(0), \tilde{w}\right)$, and corresponds to a weak imposition of the initial condition $u_{m-1}(0) + w\lambda(0) = u_0$, which is projected on $\tilde{w}$ with respect to inner product $m(\cdot,\cdot)$.

A.2.3. Computing $\tilde{w} = \widetilde{S}(w; \lambda, \tilde{\lambda})$ (adjoint spatial problem)

For given $w \in \mathcal{V}$ and given $\lambda, \tilde{\lambda} \in \mathcal{T}$, $\tilde{w} = \widetilde{S}(w; \lambda, \tilde{\lambda})$ is the solution of the following problem:

$$B_\lambda(w^*, \tilde{w}) = \widetilde{L}_{\lambda,w}(w^*), \quad \forall w^* \in \mathcal{V}, \tag{A.22}$$

where B_λ and $\widetilde{L}_{\lambda,w}$ are bilinear and linear forms on $\mathcal{V}$. B_λ is defined by Eq. (A.8) and $\widetilde{L}_{\lambda,w}$ is defined by

$$\widetilde{L}_{\lambda,w}(w^*) = \langle w^*, \langle \lambda, \lambda \rangle_{\mathcal{T}} w \rangle_{\mathcal{V}}. \tag{A.23}$$

Remark 26. Let us note that for given $\lambda, \tilde{\lambda} \in \mathcal{T}$, bilinear form B_λ is the same as bilinear form of problem $w = S_m(\lambda, \tilde{\lambda})$. Then, after the computation of $w = S_m(\lambda, \tilde{\lambda})$, the bilinear form B_λ can be directly reused (without additional computational efforts) for the solution of $\tilde{w} = \widetilde{S}(w; \lambda, \tilde{\lambda})$.

When using a spatial discretization, problem (A.22) can be rewritten as the following system of equations: find $\tilde{\mathbf{w}} \in \mathbb{R}^N$ such that

$$\left(\sum_{i=0}^{m_a} \alpha_i^\lambda \mathbf{A}_i^T \right) \tilde{\mathbf{w}} = \langle \lambda, \lambda \rangle_{\mathcal{T}} \mathbf{M}_{\mathcal{V}} \mathbf{w}, \tag{A.24}$$

where $\mathbf{M}_{\mathcal{V}} \in \mathbb{R}^{N \times N}$ is the matrix associated with inner product $\langle \cdot, \cdot \rangle_{\mathcal{V}}$ on $\mathcal{V}_N$.

A.2.4. Computing $\tilde{\lambda} = \widetilde{T}(\lambda; w, \tilde{w})$ (adjoint time problem)

For a given $\lambda \in \mathcal{T}$ and given $w, \tilde{w} \in \mathcal{V}$, $\tilde{\lambda} = \widetilde{T}(\lambda; w, \tilde{w})$ is the solution of the following problem:

$$B_w(\lambda^*, \tilde{\lambda}) = \widetilde{L}_{w,\lambda}(\lambda^*), \quad \forall \lambda^* \in \mathcal{T}. \tag{A.25}$$

where B_w and $\widetilde{L}_{w,\lambda}$ are bilinear and linear forms on $\mathcal{T}$. B_w is defined by Eq. (A.15) and $\widetilde{L}_{w,\lambda}$ is defined by

$$\widetilde{L}_{w,\lambda}(\lambda^*) = \langle \lambda^*, \langle w, w \rangle_{\mathcal{V}} \lambda \rangle_{\mathcal{T}}. \tag{A.26}$$

Remark 27. Let us note that for given $w, \tilde{w} \in \mathcal{V}$, bilinear form B_w is the same as bilinear form of problem $\lambda = T_m(w, \tilde{w})$. Then, after the computation of $\lambda = T_m(w, \tilde{w})$, the bilinear form B_w can be directly reused (without additional computational efforts) for the solution of $\tilde{\lambda} = \widetilde{T}(\lambda; w, \tilde{w})$.

Let $\bar{\lambda} \in \mathcal{T}$ be defined by $(\lambda^*, \bar{\lambda})_I = \langle \lambda^*, \lambda \rangle_{\mathcal{T}} \forall \lambda^* \in \mathcal{T}$ (for $\langle \cdot, \cdot \rangle_{\mathcal{T}} = (\cdot, \cdot)_I$, we simply have $\bar{\lambda} = \lambda$). Then, problem (A.25) appears as a time weak formulation of the following ordinary differential equation (backward in time): find $\tilde{\lambda}(t)$ such that

$$-\alpha_0^w \dot{\tilde{\lambda}}(t) + \left(\sum_{i=1}^{m_a} \alpha_i^w(t) \right) \tilde{\lambda}(t) = \langle w, w \rangle_{\mathcal{V}} \bar{\lambda}(t), \tag{A.27}$$

$$\alpha_0^w \tilde{\lambda}(T) = 0. \tag{A.28}$$

Let us note that in Eq. (A.25), the final condition is taken into account in a weak sense. In practice, when using time discontinuous Galerkin approximation, this problem is solved by using a backward in time incremental scheme.

A.2.5. Computing $\Lambda_m = T(W_m, \widetilde{W}_m)$ or $\Lambda_m = T(W_m)$ (time problem)

First of all, let us note that mapping $T(\cdot)$ associated with the Galerkin PGD (defined by Eq. (48)) corresponds to mapping $T(W_m, \widetilde{W}_m)$ of Minimax Galerkin PGD for $\widetilde{W}_m = W_m$, *i.e.* $T(W_m) \equiv T(W_m, W_m)$. For given $W_m, \widetilde{W}_m \in (\mathcal{V})^m$, $\Lambda_m = T(W_m, \widetilde{W}_m)$ is the solution of the following problem:

$$B_W(\Lambda_m, \Lambda_m^*) = L_W(\Lambda_m^*), \quad \forall \Lambda_m^* \in (\mathcal{T})^m, \tag{A.29}$$

where B_W and L_W are bilinear and linear forms on $(\mathcal{T})^m$ defined by

$$B_W(\Lambda_m, \Lambda_m^*) = \sum_{i,j=1}^{m} B\left(\lambda_i w_i, \lambda_j^* \tilde{w}_j\right) = \sum_{i,j=1}^{m} \alpha_0^{w,i,j} \left(\dot{\lambda}_i, \lambda_j^*\right)_I + \sum_{i,j=1}^{m} \sum_{l=1}^{m_a} \left(\alpha_l^{w,i,j} \lambda_i, \lambda_j^*\right)_I, \tag{A.30}$$

$$L_W(\Lambda_m^*) = \sum_{j=1}^{m} L\left(\lambda_j^* \tilde{w}_j\right) = \sum_{j=1}^{m} \sum_{l=1}^{m_\ell} \left(\gamma_l^{w,j}, \lambda_j^*\right)_I + \sum_{j=1}^{m} \gamma_0^{w,j} \lambda_j^*(0^+), \tag{A.31}$$

where

$$\alpha_l^{w,i,j}(t) = \alpha_l(t) a_l(w_i, \tilde{w}_j) \quad \text{for } l = 0 \dots m_a, \tag{A.32}$$

$$\gamma_l^{w,j}(t) = \gamma_l(t) \ell_l(\tilde{w}_j) \quad \text{for } l = 0 \dots m_\ell. \tag{A.33}$$

Problem (A.29) appears as a time weak formulation of the following set of ordinary differential equations: find $\Lambda_m(t) = (\lambda_i(t))_{i=1}^m \in (\mathcal{T})^m$ such that for $j = 1 \dots m$,

$$\sum_{i=1}^{m} \alpha_0^{w,i,j} \dot{\lambda}_i(t) + \sum_{i=1}^{m} \left(\sum_{l=1}^{m_a} \alpha_l^{w,i,j}(t) \right) \lambda_i(t) = \sum_{l=1}^{m_\ell} \gamma_l^{w,j}(t), \tag{A.34}$$

$$\sum_{i=1}^{m} \alpha_0^{w,i,j} \lambda_i(0) = \gamma_0^{w,j}. \tag{A.35}$$

The initial condition writes $m\left(\sum_{i=1}^m w_i \lambda_i(0), \tilde{w}_j\right) = m(u_0, \tilde{w}_j)$, and corresponds to a weak imposition of the initial condition $u_m(0) = u_0$, which is projected on $span(\widetilde{W}_m)$ with respect to inner product $m(\cdot,\cdot)$.

Appendix B. Algebraic formulation of Proper Generalized Decomposition algorithms

In this appendix, we describe the computational aspects of PGD from an fully algebraic point of view, after the introduction of space and time discretizations. The notations of Appendix A are used.

B.1. Preliminaries

A time function $\lambda \in \mathcal{T}_P$ is identified with a vector $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_P) \in \mathbb{R}^P$ such that $\lambda(t) = \sum_{i=1}^P \psi_i(t) \lambda_i$, where $\mathcal{T}_P = span\{\psi_i\}_{i=1}^P$. For all $(\lambda, \tilde{\lambda}) \in \mathcal{T}_P \times \mathcal{T}_P \simeq (\boldsymbol{\lambda}, \tilde{\boldsymbol{\lambda}}) \in \mathbb{R}^P \times \mathbb{R}^P$, time derivatives and integrals are interpreted in algebraic terms in the following way:

$$(\dot{\lambda},\tilde{\lambda})_I = \tilde{\boldsymbol{\lambda}}^T\mathbf{D}_\tau\boldsymbol{\lambda}, \quad (\lambda,\tilde{\lambda})_I = \tilde{\boldsymbol{\lambda}}^T\mathbf{M}_\tau\boldsymbol{\lambda}, \tag{B.1}$$

$$(\alpha_l\lambda,\tilde{\lambda}) = \tilde{\boldsymbol{\lambda}}^T\mathbf{A}_{\tau,l}\boldsymbol{\lambda} \quad \text{for } l = 1\dots m_a, \tag{B.2}$$

$$(\gamma_l,\tilde{\lambda})_I = \tilde{\boldsymbol{\lambda}}^T\mathbf{f}_{\tau,l} \quad \text{for } l = 1\dots m_\ell, \tag{B.3}$$

where matrices $\mathbf{D}_\tau, \mathbf{M}_\tau, \mathbf{A}_{\tau,l} \in \mathbb{R}^{P\times P}$ and vectors $\mathbf{f}_{\tau,l} \in \mathbb{R}^P$ are defined by

$$(\mathbf{D}_\tau)_{ij} = (\dot{\psi}_j,\psi_i)_I, \quad (\mathbf{M}_\tau)_{ij} = (\psi_j,\psi_i)_I, \tag{B.4}$$

$$(\mathbf{A}_{\tau,l})_{ij} = (\alpha_l\psi_j,\psi_i)_I \quad \text{for } l = 1\dots m_a, \tag{B.5}$$

$$(\mathbf{f}_{\tau,l})_i = (\gamma_l,\psi_i)_I \quad \text{for } l = 1\dots m_\ell. \tag{B.6}$$

For the Minimax PGD, we introduce an inner product $\langle\!\langle\cdot,\cdot\rangle\!\rangle$ on $\mathcal{V}\otimes\mathcal{T}$ whose restriction on $\mathcal{V}_N\otimes\mathcal{T}_P$ can be represented by matrices $\mathbf{M}_\mathcal{V} \in \mathbb{R}^{N\times N}$ and $\mathbf{M}_\mathcal{T} \in \mathbb{R}^{P\times P}$: for $(w,\tilde{w}) \in \mathcal{V}_N\times\mathcal{V}_N \simeq (\mathbf{w},\tilde{\mathbf{w}}) \in \mathbb{R}^N\times\mathbb{R}^N$ and $(\lambda,\tilde{\lambda}) \in \mathcal{T}_P\times\mathcal{T}_P \simeq (\boldsymbol{\lambda},\tilde{\boldsymbol{\lambda}}) \in \mathbb{R}^P\times\mathbb{R}^P$,

$$\langle\!\langle w\lambda,\tilde{w}\tilde{\lambda}\rangle\!\rangle = \langle w,\tilde{w}\rangle_\mathcal{V}\langle\lambda,\tilde{\lambda}\rangle_\mathcal{T} = (\tilde{\mathbf{w}}^T\mathbf{M}_\mathcal{V}\mathbf{w})(\tilde{\boldsymbol{\lambda}}^T\mathbf{M}_\mathcal{T}\boldsymbol{\lambda}). \tag{B.7}$$

B.2. Algebraic reformulation in tensor product spaces

The fully discretized problem writes: find $\mathbf{u} \in \mathbb{R}^N\otimes\mathbb{R}^P$ such that

$$\left(\mathbf{M}\otimes\mathbf{D}_\tau + \sum_{l=1}^{m_a}\mathbf{A}_l\otimes\mathbf{A}_{\tau,l}\right)\cdot\mathbf{u} = \sum_{l=0}^{m_\ell}\mathbf{f}_l\otimes\mathbf{f}_{\tau,l}, \tag{B.8}$$

where $\mathbf{f}_0\otimes\mathbf{f}_{\tau,0}$ takes into account the initial condition, with $\mathbf{f}_0 = \mathbf{M}\mathbf{u}_0$ and $\mathbf{f}_{\tau,0} \in \mathbb{R}^P$ such that $\forall\tilde{\lambda} \in \mathcal{T}_P$, $\tilde{\boldsymbol{\lambda}}^T\mathbf{f}_{\tau,0} = (\dot{\gamma}_0,\tilde{\lambda})_I = \tilde{\lambda}(0^+)$ (we recall that $\gamma_0 = 1$). For simplification, we let $\mathbf{A}_0 = \mathbf{M}$ and $\mathbf{A}_{\tau,0} = \mathbf{D}_\tau$. The tensor product approximation of order m writes:

$$\mathbf{u}_m = \sum_{i=1}^{m}\mathbf{w}_i\otimes\boldsymbol{\lambda}_i, \quad \mathbf{w}_i \in \mathbb{R}^N, \ \ \boldsymbol{\lambda}_i \in \mathbb{R}^P.$$

Remark 28. For the PGD based on a minimal residual formulation, we start with the following equation, instead of (B.8):

$$\left(\sum_{l=0}^{m_a}\sum_{l'=0}^{m_a}\left(\mathbf{A}_l^T\mathbf{A}_{l'}\right)\otimes\left(\mathbf{A}_{\tau,l}^T\mathbf{A}_{\tau,l'}\right)\right)\cdot\mathbf{u} = \sum_{l=0}^{m_a}\sum_{l'=0}^{m_\ell}\left(\mathbf{A}_l^T\mathbf{f}_{l'}\right)\otimes\left(\mathbf{A}_{\tau,l}^T\mathbf{f}_{\tau,l'}\right). \tag{B.9}$$

B.3. Application of mappings

B.3.1. Computing $w = S_m(\lambda,\tilde{\lambda})$ or $w = S_m(\lambda) \equiv S_m(\lambda,\lambda)$

For given $\lambda,\tilde{\lambda} \in \mathcal{T}_P$, computing $w = S_m(\lambda,\tilde{\lambda}) \simeq \mathbf{w} \in \mathbb{R}^N$ requires the solution of the following system of equations:

$$\left(\sum_{l=0}^{m_a}(\tilde{\boldsymbol{\lambda}}^T\mathbf{A}_{\tau,l}\boldsymbol{\lambda})\mathbf{A}_l\right)\mathbf{w} = \sum_{l=0}^{m_\ell}(\tilde{\boldsymbol{\lambda}}^T\mathbf{f}_{\tau,l})\mathbf{f}_l - \sum_{i=1}^{m-1}\sum_{l=0}^{m_a}(\tilde{\boldsymbol{\lambda}}^T\mathbf{A}_{\tau,l}\boldsymbol{\lambda}_i)\mathbf{A}_l\mathbf{w}_i. \tag{B.10}$$

B.3.2. Computing $\lambda = T_m(w,\tilde{w})$ or $\lambda = T_m(w) \equiv T_m(w,w)$

For given $w,\tilde{w} \in \mathcal{V}_N$, computing $\lambda = T_m(w,\tilde{w}) \simeq \boldsymbol{\lambda} \in \mathbb{R}^P$ requires the solution of the following system of equations:

$$\left(\sum_{l=0}^{m_a}(\tilde{\mathbf{w}}^T\mathbf{A}_l\mathbf{w})\mathbf{A}_{\tau,l}\right)\boldsymbol{\lambda} = \sum_{l=0}^{m_\ell}(\tilde{\mathbf{w}}^T\mathbf{f}_l)\mathbf{f}_{\tau,l} - \sum_{i=1}^{m-1}\sum_{l=0}^{m_a}(\tilde{\mathbf{w}}^T\mathbf{A}_l\mathbf{w}_i)\mathbf{A}_{\tau,l}\boldsymbol{\lambda}_i. \tag{B.11}$$

B.3.3. Computing $\tilde{w} = \widetilde{S}(w;\lambda,\tilde{\lambda})$

For a given $w \in \mathcal{V}_N$ and given $\lambda,\tilde{\lambda} \in \mathcal{T}_P$, computing $\tilde{w} = \widetilde{S}(w;\lambda,\tilde{\lambda}) \simeq \tilde{\mathbf{w}} \in \mathbb{R}^N$ requires the solution of the following system of equations:

$$\left(\sum_{l=0}^{m_a}(\tilde{\boldsymbol{\lambda}}^T\mathbf{A}_{\tau,l}\boldsymbol{\lambda})\mathbf{A}_l^T\right)\tilde{\mathbf{w}} = (\boldsymbol{\lambda}^T\mathbf{M}_\mathcal{T}\boldsymbol{\lambda})\mathbf{M}_\mathcal{V}\mathbf{w}. \tag{B.12}$$

B.3.4. Computing $\tilde{\lambda} = \widetilde{T}(\tilde{\lambda};w,\tilde{w})$

For a given $\lambda \in \mathcal{T}_P$ and given $w,\tilde{w} \in \mathcal{V}_N$, computing $\tilde{\lambda} = \widetilde{T}(\lambda;w,\tilde{w}) \simeq \tilde{\boldsymbol{\lambda}} \in \mathbb{R}^P$ requires the solution of the following system of equations:

$$\left(\sum_{l=0}^{m_a}(\tilde{\mathbf{w}}^T\mathbf{A}_l\mathbf{w})\mathbf{A}_{\tau,l}^T\right)\tilde{\lambda} = (\mathbf{w}^T\mathbf{M}_\mathcal{V}\mathbf{w})\mathbf{M}_\mathcal{T}\boldsymbol{\lambda}. \tag{B.13}$$

B.3.5. Computing $\Lambda_m = T(W_m,\widetilde{W}_m)$ or $\Lambda_m = T(W_m) \equiv T(W_m,W_m)$

For given $W_m,\widetilde{W}_m \in (\mathcal{V})^m$, identified with matrices $\mathbf{W},\widetilde{\mathbf{W}} \in \mathbb{R}^{N\times m}$, $\Lambda_m = T(W_m,\widetilde{W}_m) \simeq \boldsymbol{\Lambda} \in \mathbb{R}^m\otimes\mathbb{R}^P$ is the solution of the following problem:

$$\left(\sum_{l=0}^{m_a}(\widetilde{\mathbf{W}}^T\mathbf{A}_l\mathbf{W})\otimes\mathbf{A}_{\tau,l}\right)\cdot\boldsymbol{\Lambda} = \sum_{l=0}^{m_\ell}(\widetilde{\mathbf{W}}^T\mathbf{f}_l)\otimes\mathbf{f}_{\tau,l}. \tag{B.14}$$

References

[1] A. Ammar, B. Mokdad, F. Chinesta, R. Keunings, A new family of solvers for some classes of multidimensional partial differential equations encountered in kinetic theory modelling of complex fluids, J. Non-Newton. Fluid Mech. 139 (3) (2006) 153–176.

[2] A. Ammar, B. Mokdad, F. Chinesta, R. Keunings, A new family of solvers for some classes of multidimensional partial differential equations encountered in kinetic theory modelling of complex fluids. Part II: Transient simulation using space-time separated representations, J. Non-Newton. Fluid Mech. 144 (2–3) (2007) 98–121.

[3] D. Amsallem, C. Farhat, Interpolation method for adapting reduced-order models and application to aeroelasticity, AIAA J. 46 (7) (2008) 1803–1813.

[4] J.A. Atwell, B.B. King, Proper orthogonal decomposition for reduced basis feedback controllers for parabolic equations, Math. Comput. Modell. 33 (1–3) (2001) 1–19.

[5] G. Berkooz, P. Holmes, J.L. Lumley, The proper orthogonal decomposition in the analysis of turbulent flows, Annu. Rev. Fluid Mech. 25 (1993) 539–575.

[6] G. Beylkin, M.J. Mohlenkamp, Algorithms for numerical analysis in high dimensions, SIAM J. Sci. Comput. 26 (6) (2005) 2133–2159.

[7] S. Boyaval, C. Le Bris, Y. Maday, N.C. Nguyen, A.T. Patera, A reduced basis approach for variational problems with stochastic parameters: application to heat conduction with variable robin coefficient, Comput. Methods Appl. Mech. Engrg. 198 (41–44) (2009) 3187–3206.

[8] C. Le Bris, T. Lelievre, Y. Maday, Results and questions on a nonlinear approximation approach for solving high-dimensional partial differential equations, Construct. Approx. 30 (3) (2009) 621–651.

[9] F. Chinesta, A. Ammar, F. Lemarchand, P. Beauchene, F. Boust, Alleviating mesh constraints: model reduction, parallel time integration and high resolution homogenization, Comput. Methods Appl. Mech. Engrg. 197 (5) (2008) 400–413.

[10] E.A. Christensen, M. Brons, J.M. Sorensen, Evaluation of proper orthogonal decomposition-based decomposition techniques applied to parameter-dependent nonturbulent flows, SIAM J. Sci. Comput. 21 (4) (2000) 1419–1434.

[11] A. Doostan, G. Iaccarino, A least-squares approximation of partial differential equations with high-dimensional random inputs, J. Comput. Phys. 228 (12) (2009) 4332–4345.

[12] D. Dureisseix, P. Ladevèze, B.A. Schrefler, A computational strategy for multiphysics problems — application to poroelasticity, Int. J. Numer. Methods Engrg. 56 (10) (2003) 1489–1510.

[13] I. Ekeland, R. Teman, Convex Analysis and Variational Problems, Classics in Applied Mathematics, SIAM, 1999.

[14] K. Eriksson, C. Johnson, V. Thomée, Time discretization of parabolic problems by the discontinuous Galerkin method, RAIRO Model. Math. Anal. 19 (1985) 611–643.

[15] M.D. Gunzburger, J.S. Peterson, J.N. Shadid, Reduced-order modeling of time-dependent PDEs with multiple parameters in the boundary data, Comput. Methods Appl. Mech. Engrg. 196 (4–6) (2007) 1030–1047.

[16] T. Henri, J.-P. Yvon, Convergence estimates of pod-Galerkin methods for parabolic problems, in: Proceedings of the 21st IFIP TC7 Conference on System Modeling and Optimization, July 21–25, 2003, Sophia Antipolis, France, Springer-Verlag, 2003, pp. 295–306.

[17] A. Iollo, S. Lanteri, J.-A. Désidéri, Stability properties of POD-Galerkin approximations for the compressible Navier–Stokes equations, Technical Report 3589, INRIA, 1998.

[18] K. Karhunen, Zur spektraltheorie stochastischer prozesse, Ann. Acad. Sci. Fenn. 34 (1946).

[19] K. Kunisch, S. Volkwein, Galerkin proper orthogonal decomposition methods for parabolic problems, Numer. Math. 90 (1) (2001) 117–148.

[20] K. Kunisch, S. Volkwein, Galerkin proper orthogonal decomposition methods for a general equation in fluid dynamics, SIAM J. Numer. Anal. 40 (2) (2002) 492–515.
[21] P. Ladevèze, Nonlinear Computational Structural Mechanics – New Approaches and Non-Incremental Methods of Calculation, Springer-Verlag, 1999.
[22] P. Ladevèze, A. Nouy, On a multiscale computational strategy with time and space homogenization for structural mechanics, Comput. Methods Appl. Mech. Engrg. 192 (2003) 3061–3087.
[23] P. Ladevèze, J.C. Passieux, D. Néron, The LATIN multiscale computational method and the Proper Generalized Decomposition, Comput. Methods Appl. Mech. Engrg. 199 (21–22) (2010) 1287–1296.
[24] T. Lieu, M. Lesoinne, Parameter adaptation of reduced order models for three-dimensional flutter analysis, AIAA Paper 888, 2004.
[25] J.L. Lions, E. Magenes, Non-homogenous boundary value problems and applications, 1972.
[26] M. Loève, Fonctions aléatoires du second ordre, C.R. Acad. Sci. Paris 220 (1945).
[27] A. Nouy, A generalized spectral decomposition technique to solve a class of linear stochastic partial differential equations, Comput. Methods Appl. Mech. Engrg. 196 (45–48) (2007) 4521–4537.
[28] A. Nouy, Generalized spectral decomposition method for solving stochastic finite element equations: invariant subspace problem and dedicated algorithms, Comput. Methods Appl. Mech. Engrg. 197 (2008) 4718–4736.
[29] A. Nouy, Recent developments in spectral stochastic methods for the numerical solution of stochastic partial differential equations, Arch. Comput. Methods Engrg. 16 (3) (2009) 251–285.
[30] A. Nouy, Proper generalized decompositions and separated representations for the numerical solution of high dimensional stochastic problems, Arch. Comput. Methods Engrg., in press.
[31] A. Nouy, P. Ladevèze, Multiscale computational strategy with time and space homogenization: a radial-type approximation technique for solving micro problems, Int. J. Multiscale Comput. Engrg. 170 (2) (2004) 557–574.
[32] A. Nouy, O.P. Le Maître, Generalized spectral decomposition method for stochastic non-linear problems, J. Comput. Phys. 228 (1) (2009) 202–235.
[33] N. Pares, P. Diez, A. Huerta, Bounds of functional outputs for parabolic problems. Part 1: Exact bounds of the discontinuous Galerkin time discretization, Comput. Methods Appl. Mech. Engrg. 197 (19–20) (2008) 1641–1660.
[34] F. Riesz, B. Sz.-Nagy, Functional Analysis, Dover Publications, New York, NY, 1990.
[35] D.V. Rovas, L. Machiels, Y. Maday, Reduced-basis output bound methods for parabolic problems, IMA J. Numer. Anal. 26 (3) (2006) 423–445.
[36] R. Schmidt, M. Glauser, Improvements in low dimensional tools for flow–structure interaction problems: using global pod, AIAA Paper 2004-0889, 2004.
[37] D. Schötzau, C. Schwab, T. Werder, K. Gerdes, hp-Discontinuous Galerkin time stepping for parabolic problems, Comput. Methods Appl. Mech. Engrg. 49–50 (190) (2001) 6685–6708.
[38] V. Thomee, Galerkin Finite Element Methods for Parabolic Problems, Chapter: The Discontinuous Galerkin Time Stepping Method, Springer Series in Computational Mathematics, Springer, 1997, pp. 181–208.