A Short Review in Model Order Reduction Based on Proper Generalized Decomposition
Francisco Chinesta, Pierre Ladevèze, Elías Cueto

To cite this version:

HAL Id: hal-01004940
https://hal.archives-ouvertes.fr/hal-01004940
Submitted on 27 Mar 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
A Short Review on Model Order Reduction Based on Proper Generalized Decomposition

Francisco Chinesta · Pierre Ladeveze · Elías Cueto

Abstract  This paper revisits a new model reduction methodology based on the use of separated representations, the so called Proper Generalized Decomposition—PGD. Space and time separated representations generalize Proper Orthogonal Decompositions—POD—avoiding any a priori knowledge on the solution in contrast to the vast majority of POD based model reduction technologies as well as reduced bases approaches. Moreover, PGD allows to treat efficiently models defined in degenerated domains as well as the multi-dimensional models arising from multidimensional physics (quantum chemistry, kinetic theory descriptions, . . .) or from the standard ones when some sources of variability are introduced in the model as extra-coordinates.

This work has been partially supported by the IUF—Institut Universitaire de France—, the French ANR COSINUS SIMDREAM project and the Spanish Ministry of Science and Innovation, through grants number CICYT-DPI2008-918 and DPI2011-27778-C02-01.

1 Introduction

Today many problems in science and engineering remain intractable, in spite of the impressive progresses attained in mechanical modeling, numerical analysis, discretization techniques and computer science during the last decade, because their numerical complexity, or the restrictions imposed by different requirements (real-time, for instance) make them unaffordable for nowadays technologies.

We can enumerate different challenging scenarios for efficient numerical simulations:

– The first one concerns models that are defined in high dimensional spaces, usually encountered in quantum chemistry and kinetic theory descriptions of complex fluids. Models defined in high dimensional spaces suffer the so-called curse of dimensionality. If one proceeds to the solution of a model defined in a space of dimension $N$ by using a standard mesh based discretization technique, where $M$ nodes are used for discretizing each space coordinate, the resulting number of nodes reaches the astronomical value of $M^N$. With $M = 1000$ (a very coarse description in practice) and $N = 30$ (a very simple model) the numerical complexity results $10^{90}$. It is important to recall that $10^{80}$ is the presumed number of elementary particles in the universe!

Quantum mechanics and molecular modeling of complex fluids are not the only branches of science that suffer from the curse of dimensionality. Consider for example a chemical process involving so few molecules of the reacting species that the use of the continuous concept of concentration is not valid. This situation is often found in genetic processes such as expression of genes. The state of such a discrete system is given by a probability distribution for the number of individual molecules of each
one of the coexisting species. The balance equation governing the evolution of the system, the so-called chemical master equation, is again defined in a high-dimensional space that precludes a direct solution by means of standard mesh-based techniques.

There are of course alternative methods to address these high-dimensional problems indirectly, stochastic simulations being one of the foremost approaches. Stochastic techniques have their own challenges, however. While variance reduction is always an issue, it is quite difficult within the stochastic framework to implement parametric or sensitivity studies that go beyond the brute force approach of computing a large number of expensive, individual simulations.

A second category of problems involves multiscale problems not necessarily defined in high-dimensional spaces, but whose spectrum of characteristic times or lengths is so wide that standard incremental discretization techniques cannot be applied. In such time-multiscale problems, for instance, the time step is extremely small as a consequence of numerical stability requirements. Thus, simulations over the much larger time interval of interest, which typically requires the solution of a large linear algebraic system at each time step, simply become impossible. Multiscale models involving a wide range of characteristic times abound in many fields. Reaction-diffusion models of the degradation of plastic materials, for example, describe chemical reactions occurring within micro-seconds coupled to diffusion of chemical substances taking place over years. In processes involving microwaves, ultrasounds, etc. or materials exhibiting different relaxation times, the difficulty related to time integration is crucial. The same scenario is found in solid mechanics where the constitutive equations are strongly non linear and coupled, involving many scales and different characteristic times.

Other challenging problems are defined in degenerated geometrical domains. By this we mean that at least one of the characteristic dimensions of the domain is smaller by several orders of magnitude than the others. This is the case of bar, plate or shell-like domains typical of structures or materials processing applications. In simple situations, such problems are readily transformed into reduced, one or two-dimensional approximate theories (e.g. the classical elastic plate theory). When geometrical or material non-linearities are present, however, it is usually impossible to derive lower-dimensional models of sufficient validity. Standard mesh-based discretization methods then quickly become impractical, in view of the compulsory discretization of the small length scales that yield extremely fine meshes.

Many problems in process control, parametric modeling, inverse identification, and process or shape optimization, usually require, when approached with standard techniques, the direct computation of a very large number of solutions of the concerned model for particular values of the problem parameters. Consider for example the optimization of a process where optimal parameter values must be determined for process operating conditions (e.g. speed, position and temperature of heaters) and material properties (e.g. thermal and rheological properties of the materials). Clearly, it would be useful to be able to simulate this process at once for all values of these parameters within a prescribed interval, and then perform data mining within this rather general solution to identify optimal values.

Traditionally, Simulation-based Engineering Sciences—SBES—relied on the use of static data inputs to perform the simulations. These data could be parameters of the model(s) or boundary conditions. The word static is intended to mean here that these data could not be modified during the simulation. A new paradigm in the field of Applied Sciences and Engineering has emerged in the last decade. Dynamic Data-Driven Application Systems (DDDAS) constitute nowadays one of the most challenging applications of simulation-based Engineering Sciences. By DDDAS we mean a set of techniques that allow the linkage of simulation tools with measurement devices for real-time control of simulations. DDDAS entails the ability to dynamically incorporate additional data into an executing application, and in reverse, the ability of an application to dynamically steer the measurement process.

In this context, real time simulators are needed in many applications. One of the most challenging situations is that of haptic surgery simulators, where forces acting on the surgical tool must be translated to the peripheral device at a rate of 500 Hz. Control, malfunctioning identification and reconfiguration of malfunctioning systems also need to run in real time. All these problems can be seen as typical examples of DDDAS.

Augmented reality is another area in which efficient (fast and accurate) simulation is urgently needed. The idea is supplying in real time appropriate information to the reality perceived by the user. Augmented reality could be an excellent tool in many branches of science and engineering.

Light computing platforms are appealing alternatives to heavy computing platforms that in general are expensive and whose use requires technical knowledge. One can imagine that the off-line parametric solution of many models will make possible the on-line manipulation of those general solutions by using very light computing platforms, as for example smartphones or tablets.

While the previous list is by no means exhaustive, it includes a set of problems with apparent no relationship
among them that can be, however, treated in an unified manner as will be shown in this paper. Their common ingredient is our lack of capabilities (or knowledge) to solve them numerically in a direct, traditional way. In order to obtain a solution, some kind of model order reduction is thus compulsory.

Many years ago, in the 80’s, Pierre Ladeveze proposed a separated representation of the space and time coordinates for performing efficient solutions of complex non-linear thermo-mechanical models. The radial approximation (1) was one of the main blocks of the powerful non-incremental and non-linear LArge Time IΝcrement (LATIN) solver. The corpus of literature devoted to this technique is vast, as proved in the book by Ladeveze on the topic [67], but remained in the form of space-time separations for many years.

A more general separated representation (in this case in the conformation space of complex fluids descriptions) was more recently employed by A. Ammar and F. Chinesta [3] for approximating the solution of multi-dimensional partial differential equations. A. Nouy considered also such separated representations for solving stochastic equations were the deterministic coordinates and the stochastic ones were separated, very much like in the radial, space-time approximation. Proper Generalized Decomposition—PGD—is the common name recently coined for techniques using such separated representations. The nature of the problem, the involved coordinates and the constructor of such approximations can be very different, as will be seen.

The general form of the separated representation involved in the PGD reads:

\[ u(x,t) \approx \sum_{i=1}^{Q} X_i(x) \cdot T_i(t) \]  

(1)

for performing efficient solutions of complex non-linear thermo-mechanical models. The radial approximation (1) was one of the main blocks of the powerful non-incremental and non-linear LArge Time IΝcrement (LATIN) solver. The corpus of literature devoted to this technique is vast, as proved in the book by Ladeveze on the topic [67], but remained in the form of space-time separations for many years.

A more general separated representation (in this case in the conformation space of complex fluids descriptions) was more recently employed by A. Ammar and F. Chinesta [3] for approximating the solution of multi-dimensional partial differential equations. A. Nouy considered also such separated representations for solving stochastic equations were the deterministic coordinates and the stochastic ones were separated, very much like in the radial, space-time approximation. Proper Generalized Decomposition—PGD—is the common name recently coined for techniques using such separated representations. The nature of the problem, the involved coordinates and the constructor of such approximations can be very different, as will be seen.

The general form of the separated representation involved in the PGD reads:

\[ u(x_1, \ldots, x_N) \approx \sum_{i=1}^{Q} F_i^1(x_1) \cdots F_i^N(x_N) \]  

(2)

where \(x_i\) denote a scalar or vector coordinate defined in a domain \(\Omega_i\) of moderate dimension \(\Omega_i \subset \mathbb{R}^d\), with \(d \leq 3\), in general.

2 Review on PGD Foundations and Applications

Proper Generalized Decomposition techniques allow to circumvent most of the challenges mentioned in the introduction in a unified way. They also make it possible to solve models never until now solved, or speeding up the solution of others up to a limit not yet reached by other techniques. However, because the technique is still in its infancy, many applications (possibly the most challenging and impressive ones) are still in progress, while many and mathematical foundations remain to be fully understood. See [37] for a recent review. In what follows we are describing the PGD state of the art.

2.1 Advanced Non-linear Solvers Making Use of Space-Time Separated Representations

Even if non-linear models can be solved by using standard linearization strategies in the context of separated representations, similarly to [8] or [113], more robust strategies exist for addressing efficiently history-dependent non-linearities, making use of the natural separation of the local and non-linear thermo-mechanical behavior from the global and linear equilibrium in which a space-time separated representation constitutes an essential ingredient allowing spectacular computing time savings.

Thus, space-time separated representations can be viewed as a part of the LArge Time IΝcrement method (LATIN) which has been introduced by P. Ladeveze in 1985 [57, 58] and was the core of two books [64, 67]. These books relate the story of the PGD until their publication. Originally designed to deal with nonlinearities, such as plasticity or viscoplasticity, in time dependent problems [19–21, 27, 40–42, 59–61, 68, 72], during the past three decades, many theoretical works have been done to enhance the strategy and to deal with fundamental problems in computational mechanics. The strategy has been applied in the context of large displacements in [22, 64]. A domain decomposition technique has been proposed in [16, 17, 30, 62, 63] and multiscale features in space have been introduced in [66, 69, 73]. These features have been extended to time in [74–77, 79, 80, 100] and the scalability of the strategy illustrated in [81, 95, 110]. A multiparametric strategy for taking into account the variability parameters or to be used in identification techniques has been introduced in [2, 23–25]. The strategy has been applied to multiphysics problems in [46, 93, 94]. Some techniques specific to the manipulation of fields approximated by the PGD have been proposed in [65]. The problem of the verification of model reduction methods within the PGD was addressed in [82].

The strategy was also successfully applied to many studies, strongly related to industrial problems: the prediction of damage in composites [1, 14, 70, 78, 118], the computation of assemblies [31], the simulation of dynamic shocks [26, 71, 86], the simulation of porous media [47, 92] and the virtual testing of joints for the prediction of damping in space launchers [28].

2.2 Multidimensional Models

As discussed in the introduction, some models are inherently defined in a configurational or conformation space
that is high-dimensional in nature. Its discretization leads to the well-known curse of dimensionality if traditional, mesh-based techniques are applied. Separated representations, on the contrary, provide with an efficient means of dealing with these requirements. It is thus natural that the origin of PGD techniques, has one of its two legs in this framework, the other being the space-time separation of the LATIN method.

Separated representations were applied for solving the multidimensional Fokker-Planck equation describing complex fluids in the kinetic theory framework. In [3], the authors addressed the solution of the linear Fokker-Planck equation describing multi-bead-spring molecular models in the steady state and in homogeneous flows. In that case the distribution function describing the molecular configuration only depends on the molecule configurational coordinates. The solution procedure was extended to non-linear kinetic theory descriptions of more complex molecular descriptions in [90]. The transient solution was addressed in [5] in which the time was added as an extra-coordinate. Transient solutions making use of reduced bases in the context of an adaptive proper orthogonal decomposition [116] were considered in the case of low dimensional configuration spaces: the FENE model was addressed in [4] and liquid crystalline polymers in [6]. In [87], the authors considered multi-bead-spring models but considered spectral approximation for representing all the functions involved in the finite sums decomposition. A deeper analysis of non-linear and transient models was considered in [8]. Complex fluid models were coupled with complex flows in [112] and [91] opening very encouraging perspectives and claiming the necessity of defining efficient stabilizations. A first tentative of convective stabilization was proposed in [43]. Finally, in [33] the PGD was applied for solving the stochastic equation within the Brownian Configuration Field framework. The interested reader can refer to [39] and the references therein for an exhaustive overview of PGD in computational rheology.

Multidimensional models encountered in the finer descriptions of matter (ranging from quantum chemistry to statistical mechanics descriptions) were revisited in [35]. The multidimensional chemical master equation was efficiently solved in [38]. The Langer’s equation governing phase transitions was considered in [85]. Finally, models coming from financial mathematics were addressed in [48].

2.3 Separating the Physical Space

Other models are not inherently defined in a high-dimensional space, but can nevertheless be treated efficiently in a separated manner. Models defined in cubic (or hypercubic, but moderate-dimensional) domains suggest the following separated representation

\[ u(x, y, z) \approx \sum_{i=1}^{Q} X_i(x) \cdot Y_i(y) \cdot Z_i(z) \]  

Thus, the 3D solution results in a sequence of 1D solutions for computing functions \( X_i(x), Y_i(y) \) and \( Z_i(z) \). This kind of domains are very frequent when addressing homogenization problems in which the elementary volume element is a cube. The interested reader can refer to [34]. A full decomposition was also efficiently applied for solving the Navier-Stokes equations in a cube in [45]. Fully separated representations in complex, non-hypercubic domains can be equally performed by using an appropriate X-FEM like strategy [54]. In that work the enforcement of non-homogeneous Dirichlet boundary conditions was also analysed, giving rise to a generalized form of the PGD approach.

In the case of plates, shells or extruded geometries, one could consider the separated approximation [18] also advantageously:

\[ u(x, y, z) \approx \sum_{i=1}^{Q} X_i(x, y) \cdot Z_i(z) \]

In the case of models defined in plate geometries \((x, y)\) represent the in-plane coordinates and \(z\) the thickness direction. In the case of extruded profiles \((x, y)\) represents the surface extruded in the \(z\) direction. This kind of representation makes possible fully 3D solutions with numerical complexities (and therefore computation times) characteristic of 2D solutions, without any simplifying kinematic assumption.

2.4 Parametric Models: A Route to Efficient Optimization, Inverse Identification and Real Time Simulation

Previous multidimensional models may seem too far from the usual computational mechanics practice. However, usual computational mechanics models could be enriched by a PGD treatment. Thus, adding some new coordinates to models initially non high-dimensional, could lead to new, never explored insights in the physics. Imagine for instance solving the heat (Fourier) equation and assume unknown the material thermal conductivity. This could happen because it has a stochastic nature, for instance, or simply because prior to solve the thermal model it should be measured. You have three possibilities: (i) wait for the conductivity to be measured (a conservative solution, but impractical in many engineering situations); (ii) solve the equation for many values of the conductivity (a sort of brute force approach) to get an overall idea of the behavior of the solution; or (iii) solve the heat equation only once for any value of the conductivity.
Obviously the third alternative is the most exciting one, but up to our knowledge, no result has been presented in the literature so far that could eventually allow to do so. To compute this general solution, however, it suffices to introduce the conductivity \( k \) as an extra-coordinate, playing the same role as the standard space and time coordinates, even if there are not derivatives concerning this extra-coordinate in the heat equation. A possible separated approximation could be:

\[
 u(x, t, k) \approx \sum_{i=1}^{Q} X_i(x) \cdot T_i(t) \cdot K_i(k)
\]

This procedure works very well in practice, and can be extended to introduce many other extra-coordinates: source term, boundary conditions, initial condition, etc. The price to pay is the increase of the dimensionality of the resulting model that now contains the standard physical coordinates (space and time) plus all the other extra-coordinates that we decided to introduce. However, the separated representation of the PGD allows to treat these models advantageously.

This kind of parametric modeling was addressed in [11, 18, 114] where material parameters were introduced as extra-coordinates. In [83] and [84] thermal conductivities, macroscopic temperature and its time evolution were introduced as extra-coordinates for computing linear and nonlinear homogenization. In [18] the anisotropy direction of plies involved in a composite laminate were considered as extra-coordinates. By assuming a certain uncertainty in the real orientation of such plies, authors evaluate the envelope of the resulting distorted structures due to the thermomechanical coupling.

The futurist concept of “virtual chart for dimensioning” consists of a PGD-reduced model describing the physics for a family of structures. It allows engineers to proceed to complex designs in some seconds. This concept, fully compatible with the most recent design procedures, is being actively developed by the Ladeveze’s group.

2.5 Real-Time Simulation, DDDAS and More

It is easy to understand that after performing this type of calculations, in which parameters are considered advantageously as new coordinates of the model, a posteriori inverse identification or optimization can be easily handled. This new PGD framework allows to perform this type of calculations very efficiently, because in fact all the result have been previously computed in the form of a separated, high-dimensional solution so that they constitute a simple post-processing of this general solution. Process optimization was considered in [52], for instance. Shape optimization was performed by considering all the geometrical parameters as extra-coordinates, leading to the model solution in any of the geometries generated by the parameters considered as extra-coordinates [88]. This strategy could be an alternative to the POD-based shape optimization considered in [117]. Inverse methods in the context of real-time simulations were addressed in [55] and were coupled with control strategies in [53] as a first step towards DDDAS (dynamic data-driven application systems). Moreover, because the general parametric solution was precomputed off-line, it can be used on-line under real time constraints and using light computing platforms like smartphones [18, 53], that constitutes a first step towards the use of this kind of information in augmented reality platforms.

As mentioned before, surgical simulators must operate at frequencies higher than 500 Hz. The use of model reduction seems to be an appealing alternative for reaching such performances. However, techniques based on the use of the POD, PODI, even combined with an asymptotic numerical methods to avoid the computation of the tangent matrix, exhibit serious difficulties to fulfil such requirements as discussed in [96–99]. The use of parametric solutions in which the applied load and its point of application are considered as extra-coordinates open an unimaginable field of applications.

2.6 Uncertainty Quantification and Stochastic Parametric Analyses

The PGD method was introduced in [101] in the context of uncertainty quantification. In this context, the method was initially called Generalized Spectral Decomposition (GSD) and was considered to be a generalization of Karhunen-Loève decomposition for the a priori construction of separated representations of the solution of stochastic partial differential equations (SPDE).

The problem that defines the reduced bases of functions in the decomposition of the solution has been interpreted as a pseudo eigenproblem. This interpretation has led to the development of dedicated algorithms inspired from solution techniques for classical eigenproblems [103]. In this context of stochastic problems, a major interest of PGD algorithms is that they decouple the solution of deterministic problems and algebraic stochastic equations, thus making the PGD a promising alternative to traditional methods for uncertainty propagation [105]. PGD has also been extended to stochastic nonlinear problems with two alternative strategies: the use of nonlinear iterative solvers with the PGD method as a solver for the succession of linearized stochastic problems [102], or the direct construction of a separated representation of the solution of nonlinear problems [104]. More recently, the PGD has been successfully applied to the solution of high dimensional stochastic parametric problems, with the introduction of suitable hierarchical tensor representations and associated algorithms [106].
PGD has also been exploited for the construction of robust reduced order models in structural solid mechanics [32], thus providing an alternative to traditional modal superposition approaches. In [109], PGD method was coupled with fictitious domain methods in order to handle the solution of PDEs with geometrical uncertainties.

2.7 Numerical Analysis

From the point of view of the numerical analysis of separated representation and its associated constructors we can cite some recent works. There exist several approaches to the numerical analysis of PGD. They combine the existence of a best approximation and a greedy algorithm. The idea of using greedy algorithms to construct successive approximations was considered in [3] in the context of the numerical solution of multidimensional PDEs. One can note, however, that the strategy considered in [3] can be interpreted as a greedy algorithm for the error minimization only for elliptic symmetric problems since they can be rewritten in terms of minimization of an energy functional. In [9] the analysis of the convergence of the greedy rank-one update algorithm for solving full-rank linear systems was addressed.

A more complete theoretical study was addressed in [89] for symmetric elliptic problems by exploiting its connection to greedy algorithms from nonlinear approximation theory explored, for example, in [44]. Hence, the variational version of the algorithm, based on the minimization of a sequence of Dirichlet energies, was shown to converge. In [51] the authors extend the convergence analysis of the pure greedy and orthogonal greedy algorithms considered in [89] to the technically more complicated case where the Laplace operator is replaced by a high-dimensional Ornstein-Uhlenbeck operator with unbounded drift. This case bears many similarities with that appearing in Fokker-Planck equations arising in bead-spring chain type kinetic polymer models with finitely extendible non-linear elastic potentials. These are frequently posed on a high-dimensional Cartesian product configuration space $D = D_1 \times \cdots \times D_N$ contained in $\mathbb{R}^{Nd}$, where each set $D_i$, $i = 1, \ldots, N$, is a bounded open ball in $\mathbb{R}^d$, $d = 2, 3$.

General elliptic problems were considered in [49]. The extension of these results to non-linear symmetric coercive problems is the subject of the paper [29] that also contains specific results when the algorithm is applied to finite-dimensional problems. Current research tracks include for example the extension of the convergence analysis to non-symmetric problems and the analysis of the rate of convergence for non-linear equations. In [50], it is provided a mathematical analysis of a family of progressive and updated PGDs for a particular class of convex optimization problems in reflexive tensor Banach spaces. The above results have led to a generalization of the concept of Singular Value Decomposition (SVD). In particular, a constrained version of the SVD has been proposed in [108].

2.8 Miscellaneous

There is a variety of open questions when applying the PGD based discretization techniques. We discussed previously the question related to the convective stabilization [43]. The issue related to the stabilization of mixed formulation has not been addressed until now.

Other important question lies in the error estimation. A first attempt was considered in [10]. Coupling standard mesh based discretization techniques with reduced bases (POD and PGD) is an issue of major interest in order to represent localized behaviors (e.g. discontinuities, singularities, boundary layers, ...). Some coupling strategies were proposed in [7] and [12].

Multi-scale and multi-physics non-linear coupled models involving different characteristic times were efficiently coupled in [36] where a globalization of local problems to speed-up the simulation in a separated representation framework has been proposed. Later, in [13] the time axis was transformed in a two-dimensional domain to account for two very different time scales. The coupling of other multi-physics models in the context of composites manufacturing processes was performed in [113]. A monolithic approach of coupled thermo-mechanical models was employed in [15]. In [111] the solution of the electromagnetic equations was considered.

In [56] the PGD was introduced in the boundary element method framework for solving transient models where few works concerning the use of reduced bases exist [115]. Because the use of a space-time separated representation the time-dependent kernel is no more needed, and only the steady state kernel functions are needed. The space-time separation implies significant computational time savings.

In [107], alternative definitions of PGD and dedicated algorithms are provided in the context of time-dependent PDEs. In particular, a new definition of PGD called Minimax PGD is introduced for handling non symmetric variational problems.

3 Discussion

As can be noticed from the pages before, the Proper Generalized Decomposition technique has been advantageously applied to a plethora of different problems and in different ways. Ranging from the original space-time separated representations to more recent applications in which models are
efficiently cast into a high-dimensional framework to overcome real-time requirements by just making real-time post-processing, the potential field of applications seems not to have been completely explored.

In fact, some crucial aspects at the very root of the method continue to be not fully understood. For instance, the relationship among POD/SVD and PGD techniques, has not been analysed to the whole extent. Optimality conditions of the separated representation, for instance, are intimately related to this question and have not been completely solved so far.

We believe, therefore, that there is a lot of room in the PGD field to perform research, both at the theoretical level and from the point of view of applications, whose range seems to be unlimited.

References


