A Mathematical Characterization of the Performance of the "Multi-Slice" Projector

C. Herzet, M. Diallo, P. Héas
INRIA Rennes, Campus de Beaulieu, 35000 Rennes, France

Abstract

We consider an enhanced version of the well-kwown "Petrov-Galerkin" projection in Hilbert spaces. The proposed procedure, dubbed "multi-slice" projector, exploits the fact that the sought solution belongs to the intersection of several high-dimensional slices. This setup is for example of interest in model-order reduction where this type of prior may be computed off-line. In this note, we provide a mathematical characterization of the performance achievable by the multi-slice projector and compare the latter with the results holding in the Petrov-Galerkin setup. In particular, we illustrate the superiority of the multi-slice approach in certain situations.

Nous considérons une version améliorée de la projection de "Petrov-Galerkin" dans un espace de Hilbert. La procédure proposée, appelée "projecteur multi-tranches", exploite le fait que la solution recherchée appartient à l'intersection de plusieurs tranches de hautes dimensions. Dans cette note, nous fournissons une caractérisation mathématique des performances atteignables par le projecteur "multi-tranches" et comparons les résultats obtenus à ceux existants dans le contexte des projections de Petrov-Galerkin. Nous illustrons ainsi la supériorité de l'approche multi-tranches dans certaines situations.

1 Introduction

Let $\mathcal{H}$ be a Hilbert space with inner product $\langle \cdot, \cdot \rangle$ and induced norm $\|\cdot\|$. We focus on the following variational formulation:

$$\text{Find } \boldsymbol{h}^\star \in \mathcal{H} \text{ such that } \ a(\boldsymbol{h}^\star, \boldsymbol{h}) = b(\boldsymbol{h}) \quad \forall \boldsymbol{h} \in \mathcal{H}, \tag{1}$$

where $a : \mathcal{H} \times \mathcal{H} \to \mathbb{R}$ is a bilinear operator and $b : \mathcal{H} \to \mathbb{R}$ a linear operator. Problem (1) is quite common (it appears for example in the weak formulation of elliptic partial differential equations) and has therefore been well-studied in the literature. In particular, it has a unique solution under mild conditions, see Lax-Milgram's and Necas Theorems in [1, Theorems 2.1 and 2.2].

Unfortunately, solving (1) is generaly an intractable problem. A popular alternative to compute an approximation of (1) is known as "Petrov-Galerkin" projection. Formally, this approach consists of approximating (1) by the following problem:

$$\text{Find } \hat{\boldsymbol{h}}_{\text{PG}} \in V_n \text{ such that } \; a(\hat{\boldsymbol{h}}_{\text{PG}}, \boldsymbol{h}) = b(\boldsymbol{h}) \quad \forall \boldsymbol{h} \in Z_m \tag{2}$$

where $V_n \subset \mathcal{H}$ is a linear subspace of dimension n and $Z_m \subset \mathcal{H}$ is a linear subspace of dimension $m \geq n$. Since the dimension of V_n and Z_m are finite, (2) admits a simple algebraic solution under mild conditions. In the literature of model reduction (see *e.g.*, [1]), Petrov-Galerkin approximation is at the core of the family of "projection-based" reduced models.

In this note we elaborate on an alternative projection procedure exploiting several approximation subspaces. Indeed, in the context of model-order reduction, standard strategies to evaluate a good approximation subspace V_n, *e.g.*, reduced basis [1] or proper orthogonal decomposition [2], typically generate a sequence of subspaces $\{V_k\}_{k=0}^n$ and positive scalars $\{\hat{\epsilon}_k\}_{k=0}^n$ such that

$$V_0 \subset V_1 \subset \ldots \subset V_n \tag{3}$$

and

$$\text{dist}(\boldsymbol{h}^\star, V_k) \leq \hat{\epsilon}_k, \quad k = 0 \ldots n. \tag{4}$$

Clearly, (4) provides some useful information about the location of $\boldsymbol{h}^\star$ in $\mathcal{H}$ since it restrains the latter to belong to the intersection of a set of low dimensional slices, *i.e.*,

$$\boldsymbol{h}^\star \in \cap_{k=0}^n \mathcal{S}_k, \tag{5}$$

where

$$\mathcal{S}_k = \{\boldsymbol{h} : \text{dist}(\boldsymbol{h}, V_k) \leq \hat{\epsilon}_k\}, \quad k = 0 \ldots n. \tag{6}$$

In standard Petrov-Galerkin projection (2), only V_n is used and the additional information provided by (5) is discarded. In this work, we consider a simple methodology to exploit the latter additional information into the projection process. More specifically, we focus on the following optimization problem[1]

$$\text{Find } \hat{\boldsymbol{h}}_{\text{MS}} \in \operatorname*{arg\,min}_{\boldsymbol{h}\in V_n} \sum_{j=1}^{m} (b(\boldsymbol{z}_j) - a(\boldsymbol{h}, \boldsymbol{z}_j))^2 \qquad (7)$$
$$\text{subject to } \operatorname{dist}(\boldsymbol{h}, V_k) \leq \hat{\epsilon}_k, \quad k = 0 \dots n,$$

which can be seen as an extension of the standard Petrov-Galerkin approach. In particular, the constraints in (7) exploit the prior information (4) into the projection process: each constraint imposes that the solution belongs to some k-dimensional slice $\mathcal{S}_k$. Hence, in the sequel, we will dub this procedure as "multi-slice" projection.

The practical interest of the multi-slice approach has already been emphasized in several contributions. In [3, 4] we presented some applications of the multi-slice decoder to the problem of model-order reduction of parametric partial differential equations. In [5] and [6], the authors showed that multi-slice decoder can be of interest to enhance the performance of the "empirical interpolation method" or the simulation of Navier-Stokes equations. "Multi-slice" prior information of the form (5) has also been considered in [7] for data assimilation. However, in the latter contribution, the decoder considered by the authors differs from (7) since the solution is no longer constrained to belong to the low-dimensional subspace V_n.

In this note we provide a mathematical characterization of the performance achievable by the multi-slice decoder (7). More specifically, we derive an "instance optimality property" relating the projection error $\|\hat{\boldsymbol{h}}_{\text{MS}} - \boldsymbol{h}^\star\|$ to the distance between $\boldsymbol{h}^\star$ and the different approximation subspaces V_k. Our result is presented in Theorem 2 in the next section.

2 Performance guarantees

One of the reasons which has ensured the success of Petrov-Galerkin projection is the existence of strong theoretical guarantees, *e.g.*, Cea's Lemma [1, Lemma 2.2] or the Babuska's Theorem [1, Theorem 2.3]. In this section we derive a similar result for the multi-slice decoder (7). The standard result associated to Petrov-Galerkin projection is recalled in Theorem 1 whereas our characterization of the multi-slice decoder (7) is presented in Theorem 2. We conclude this section by providing two examples in which the multi-slice projector leads to bet-

[1] In this note we assume that constraints are available $\forall k \in \{1 \dots n\}$. All the derivations presented in this paper may nevertheless be easily extended to the case where constraints in (7) are only available for *some* $k \in \{1 \dots n\}$.

ter guarantees of reconstruction than the standard Petrov-Galerkin approach.

We first introduce some quantities of interest. First, we let $\{\boldsymbol{v}_j\}_{j=1}^n$ and $\{\boldsymbol{z}_j\}_{j=1}^m$ be orthonormal bases (ONBs) of the subspaces V_n and Z_m, respectively. We define $\{\boldsymbol{a}_j\}_{j=1}^m$ as the Riesz's representers of $\{a(\cdot, \boldsymbol{z}_j)\}_{j=1}^m$. We denote by $\{\sigma_j\}_{j=1}^n$ the set of singular values (sorted in their decreasing order of magnitude) of the Gram matrix

$$\mathbf{G} = [\langle \boldsymbol{a}_i, \boldsymbol{v}_j \rangle]_{i,j} \in \mathbb{R}^{m\times n}. \tag{8}$$

With these notations, the well-known Babuska's theorem (in a Hilbert space) can be formulated as follows:

Theorem 1 (Babuska's Theorem)**.** *If* $\sigma_n > 0$ *then the solution of* (2) *is unique and satisfies*

$$\left\|\boldsymbol{h}^\star - \hat{\boldsymbol{h}}_{\mathrm{PG}}\right\| \leq \frac{\sigma_1}{\sigma_n}\mathrm{dist}(\boldsymbol{h}^\star, V_n). \tag{9}$$

See for example [8] for a proof of this result. Hereafter we provide a similar characterization of the performance of the multi-slice projector (7). In order to state our result we need to introduce the following quantities. We first define the short-hand notations[2]

$$\epsilon_k = \mathrm{dist}(\boldsymbol{h}^\star, V_k), \tag{10}$$

and

$$\gamma = \sup_{\boldsymbol{h}\in V_n^\perp, \|\boldsymbol{h}\|=1} \left(\sum_{j=1}^m \langle \boldsymbol{a}_j, \boldsymbol{h}\rangle^2\right)^{\frac{1}{2}}. \tag{11}$$

Moreover, we define

$$\delta_j = \sum_{k=1}^n |x_{kj}|(\hat{\epsilon}_{k-1} + \epsilon_{k-1}), \tag{12}$$

where x_{kj} are the elements of the matrix $\mathbf{X}$ appearing in the singular value decomposition of $\mathbf{G}$, that is $\mathbf{G} = \mathbf{U}\Lambda\mathbf{X}^{\mathrm{T}}$, where $\mathbf{U} \in \mathbb{R}^{m\times m}$, $\mathbf{X} \in \mathbb{R}^{n\times n}$ are orthogonal matrices and $\Lambda \in \mathbb{R}^{m\times n}$ is the diagonal matrix of singular values $\{\sigma_j\}_{j=1}^n$.

Using these notations, our result reads:

[2] ϵ_k thus represents the true distance from $\boldsymbol{h}^\star$ to V_k. We note that this quantity is usually unknown to the practitioner. This is in contrast which $\hat{\epsilon}_k$ which represents the prior information available to the practitioner but is only an upper bound on ϵ_k.

Theorem 2. *Let $\boldsymbol{h}^\star$ be a solution of* (1) *verifying* (5). *Then any solution $\hat{\boldsymbol{h}}_{\mathrm{MS}}$ of* (7) *verifies*

$$\left\|\boldsymbol{h}^\star - \hat{\boldsymbol{h}}_{\mathrm{MS}}\right\| \leq \begin{cases} \left(\sum_{j=\ell+1}^{n} \delta_j^2 + \rho\, \delta_\ell^2 + \epsilon_n^2\right)^{\frac{1}{2}} & \text{if } \sum_{j=1}^{n} \sigma_j^2\delta_j^2 \geq 4\gamma^2\epsilon_n^2, \\ \left(\sum_{j=1}^{n} \delta_j^2 + \epsilon_n^2\right)^{\frac{1}{2}} & \text{otherwise,} \end{cases} \tag{13}$$

where ℓ is the largest integer such that

$$\sum_{j=\ell}^{n} \sigma_j^2\delta_j^2 \geq 4\gamma^2\epsilon_n^2, \tag{14}$$

and $\rho \in [0,1]$ is defined as

$$\rho\sigma_\ell^2\delta_\ell^2 + \sum_{j=\ell+1}^{n} \sigma_j^2\delta_j^2 = 4\gamma^2\epsilon_n^2. \tag{15}$$

Moreover, if $\sigma_n > 0$, (7) *admits a unique solution.*

A proof of Theorem 2 is detailed in Section 3.

We conclude this section by particularizing the results stated in Theorems 1 and 2 to different setups. In particular, we emphasize two situations[3] where the multi-slice projection has much better reconstruction guarantees than its Petrov-Galerkin counterpart. In order to ease the comparison between the bounds stated in Theorems 1 and 2, we consider the case where $\{\boldsymbol{a}_j\}_{j=1}^m$ is an ONB. We note that in such a case, we have $\sigma_1 \leq 1$ and $\gamma \leq 1$.

Example 1. *We first assume that $\mathbf{X} = \mathbf{I}_n$ in the singular-value decomposition of $\mathbf{G}$. We set $\hat{\epsilon}_j = \epsilon_j$ and assume that*

$$\epsilon_j = \begin{cases} 1 & j = 0 \dots n-3, \\ \epsilon^{\frac{1}{2}} & j = n-2, n-1, \\ \epsilon & j = n, \end{cases} \tag{16}$$

for some $\epsilon \ll 1$. Moreover, we let

$$\sigma_j = \begin{cases} 1 & j = 1 \dots n-3, \\ \epsilon^{\frac{1}{2}} & j = n-2, n-1, \\ \epsilon & j = n. \end{cases} \tag{17}$$

[3]The two setups considered below correspond to those exposed in [7, Section 3.2].

In this setup, the upper bound (9) of Theorem 1 becomes:

$$\left\|\hat{\boldsymbol{h}}_{\mathrm{PG}} - \boldsymbol{h}^\star\right\| \leq \sigma_n^{-1}\ \mathrm{dist}(\boldsymbol{h}^\star, V_n) = \epsilon^{-1}\epsilon = 1. \tag{18}$$

On the other hand, because $\mathbf{X} = \mathbf{I}$, we have

$$\delta_j = \hat{\epsilon}_{j-1} + \epsilon_{j-1} = 2\epsilon_{j-1}. \tag{19}$$

The index ℓ appearing in Theorem 2 is smaller or equal to $n-1$ since

$$\begin{aligned} \sigma_n^2\delta_n^2 &= \sigma_n^2(2\epsilon_{n-1})^2 = 4\epsilon^3 \ll 4\epsilon^2, \\ \sigma_{n-1}^2\delta_{n-1}^2 &= \sigma_{n-1}^2(2\epsilon_{n-2})^2 = 4\epsilon^2, \end{aligned}$$

and thus

$$\sigma_{n-1}^2\delta_{n-1}^2 + \sigma_n^2\delta_n^2 \geq 4\epsilon^2 \geq 4\gamma^2\epsilon^2 \tag{20}$$

since $\gamma \leq 1$. The upper bound in Theorem 2 becomes

$$\begin{aligned} \left\|\boldsymbol{h}^\star - \hat{\boldsymbol{h}}_{\mathrm{MS}}\right\| &\leq \left(\delta_{n-1}^2 + \delta_n^2 + \epsilon_n^2\right)^{\frac{1}{2}}, \\ &= \left(4\epsilon + 4\epsilon + \epsilon^2\right)^{\frac{1}{2}}, \\ &\leq 3\epsilon^{\frac{1}{2}}. \end{aligned} \tag{21}$$

Hence the bound in the multi-slice setup (21) can be arbitrarily small as compared to (18) when $\epsilon \to 0$.

Example 2. *We now consider $\mathbf{X} = n^{-\frac{1}{2}}\,\mathbf{1}_{n\times n}$ where $\mathbf{1}_{n\times n}$ is an $n \times n$ matrix of 1's. We set $\hat{\epsilon}_j = \epsilon_j$ and assume that*

$$\epsilon_j = \begin{cases} \frac{1}{2} & j = 0, \\ \frac{1}{2(n-1)} & j = 1\ldots n-1, \\ \epsilon & j = n, \end{cases} \tag{22}$$

for some $\epsilon \ll n^{-1}$ (Note that we must have: $\epsilon \leq \frac{1}{2(n-1)}$ by definition). Moreover, we let

$$\sigma_j = \begin{cases} \sigma & j = 1\ldots n-1, \\ \epsilon^2 & j = n, \end{cases} \tag{23}$$

for some $1 \geq \sigma > \epsilon$ whose value will be specified below.

With these choices, the upper bound (9) of Theorem 1 becomes:

$$\left\|\hat{\boldsymbol{h}}_{\mathrm{PG}} - \boldsymbol{h}^\star\right\| \leq \sigma_n^{-1}\ \mathrm{dist}(\boldsymbol{h}^\star, V_n) = \epsilon^{-2}\epsilon = \epsilon^{-1}. \tag{24}$$

On the other hand, we have

$$\begin{aligned}\delta_j &= \sum_{k=1}^{n} |x_{kj}| (\hat{\epsilon}_{k-1} + \epsilon_{k-1}), \\ &= 2n^{-\frac{1}{2}} \sum_{k=1}^{n} \epsilon_{k-1}, \\ &= 2n^{-\frac{1}{2}}. \end{aligned} \tag{25}$$

By choosing σ such that (we remind the reader that $\sigma_{n-1} = \sigma$ by definition (23)*)*

$$\sigma_{n-1}^2 \delta_{n-1}^2 + \sigma_n^2 \delta_n^2 = 4\epsilon^2, \tag{26}$$

we obtain that index ℓ appearing in Theorem 2 is smaller or equal to $n-1$ since $\gamma \leq 1$. The upper bound in Theorem 2 then reads

$$\begin{aligned}\left\| \boldsymbol{h}^\star - \hat{\boldsymbol{h}}_{\text{MS}} \right\| &\leq \left(\delta_{n-1}^2 + \delta_n^2 + \epsilon_n^2\right)^{\frac{1}{2}}, \\ &= \left(4n^{-1} + 4n^{-1} + \epsilon^2\right)^{\frac{1}{2}}, \\ &\leq 3n^{-\frac{1}{2}}, \end{aligned} \tag{27}$$

where the last inequality follows from our initial assumption $\epsilon \ll n^{-1}$.

3 Proof of Theorem 2

In this section, we provide a proof of the result stated in Theorem 2. We first note that problem (7) is equivalent to finding the minimum of a quadratic function over a closed bounded subset of V_n. A minimizer thus always exists. Moreover, the unicity of the minimizer stated at the end of Theorem 2 follows from the strict convexity of the cost function when $\sigma_n > 0$.

In the rest of this section, we thus mainly focus on the derivation of the upper bound (13). Our proof is based on the following steps. First, since $\hat{\boldsymbol{h}}_{\text{MS}} \in V_n$, we have that

$$\begin{aligned}\left\| \boldsymbol{h}^\star - \hat{\boldsymbol{h}}_{\text{MS}} \right\|^2 &= \left\| P_{V_n}(\boldsymbol{h}^\star) - \hat{\boldsymbol{h}}_{\text{MS}} \right\|^2 + \left\| P_{V_n}^{\perp}(\boldsymbol{h}^\star) \right\|^2, \\ &= \left\| P_{V_n}(\boldsymbol{h}^\star) - \hat{\boldsymbol{h}}_{\text{MS}} \right\|^2 + \epsilon_n^2, \end{aligned} \tag{28}$$

where $P_{V_n}(\cdot)$ (resp. $P_{V_n}^{\perp}(\cdot)$) denotes the orthogonal projector onto V_n (resp. $V_n^{\perp}$). We then derive an upper bound on $\|P_{V_n}(\boldsymbol{h}^\star) - \hat{\boldsymbol{h}}_{\text{MS}}\|^2$ as follows:

- We identify a set $\mathcal{D}$ such that $P_{V_n}(\boldsymbol{h}^\star) - \hat{\boldsymbol{h}}_{\mathrm{MS}} \in \mathcal{D}$ in Section 3.1. We then have $\|P_{V_n}(\boldsymbol{h}^\star) - \hat{\boldsymbol{h}}_{\mathrm{MS}}\|^2 \leq \sup_{\boldsymbol{d}\in\mathcal{D}} \|\boldsymbol{d}\|^2$.
- We derive the analytical expression of $\sup_{\boldsymbol{d}\in\mathcal{D}} \|\boldsymbol{d}\|^2$ as a function of the parameters $\{\epsilon_k\}_{k=1}^n$, $\{\hat{\epsilon}_k\}_{k=1}^n$ and $\{\sigma_k\}_{k=1}^n$.

Combining these results, we obtain (13)-(15).

3.1 Definition of $\mathcal{D}$

We express $\mathcal{D}$ as the intersection of two sets $\mathcal{D}_1$ and $\mathcal{D}_2$ that we define in Sections 3.1.2 and 3.1.3 respectively. In order to properly define these quantities, we introduce some particular ONBs for V_n and $W_m = \mathrm{span}\left(\{\boldsymbol{a}_j\}_{j=1}^m\right)$ in Section 3.1.1.

3.1.1 Some particular bases for V_n and W_m

Let

$$\mathbf{G} = \mathbf{U}\Lambda\mathbf{X}^{\mathrm{T}} \tag{29}$$

be the singular value decomposition of the Gram matrix defined in (8), where $\mathbf{U} \in \mathbb{R}^{m\times m}$ and $\mathbf{X} \in \mathbb{R}^{n\times n}$ are orthonormal matrices and $\Lambda \in \mathbb{R}^{m\times n}$ is the diagonal matrix of singular values. We denote by $\{\sigma_j\}_{j=1}^n$ the set of singular values of $\mathbf{G}$ sorted in their decreasing order of magnitude.

We define the following bases for V_n and W_m:

$$\boldsymbol{v}_j^* = \sum_{i=1}^n x_{ij}\boldsymbol{v}_i, \tag{30}$$

$$\boldsymbol{a}_j^* = \sum_{i=1}^m u_{ij}\boldsymbol{a}_i, \tag{31}$$

where $\mathbf{U} \in \mathbb{R}^{m\times m}$ and $\mathbf{X} \in \mathbb{R}^{n\times n}$ are the orthonormal matrices appearing in (29). We note that $\left\{\boldsymbol{v}_j^*\right\}_{j=1}^n$ is an ONB whereas $\left\{\boldsymbol{a}_j^*\right\}_{j=1}^m$ is not necessarily orthonormal. By definition, $\left\{\boldsymbol{v}_j^*\right\}_{j=1}^n$ and $\left\{\boldsymbol{a}_j^*\right\}_{j=1}^m$ enjoy the following desirable property:

$$\left\langle \boldsymbol{a}_i^*, \boldsymbol{v}_j^* \right\rangle = \begin{cases} \sigma_j & \text{if } i = j \\ 0 & \text{otherwise.} \end{cases} \tag{32}$$

3.1.2 Definition of $\mathcal{D}_1$

Let us define $\mathcal{D}_1$ as

$$\mathcal{D}_1 = \left\{ \boldsymbol{d} = \sum_{j=1}^{n} \beta_j \boldsymbol{v}_j^* : \sum_{j=1}^{n} \sigma_j^2 \beta_j^2 \leq 4\gamma^2 \epsilon_n^2 \right\}, \tag{33}$$

where γ is defined in (11). We show hereafter that $P_{V_n}(\boldsymbol{h}^\star) - \hat{\boldsymbol{h}}_{\text{MS}} \in \mathcal{D}_1$.

Let us first consider the intermediate set

$$\mathcal{S} = \left\{ \boldsymbol{h} : f(\boldsymbol{h}) \leq \gamma^2 \epsilon_n^2 \right\}, \tag{34}$$

where $f(\boldsymbol{h}) = \sum_{j=1}^{m} \big(b(\boldsymbol{z}_j) - a(\boldsymbol{h}, \boldsymbol{z}_j)\big)^2$ is the cost function appearing in the variational formulation of multi-slice projector (7).

Clearly $P_{V_n}(\boldsymbol{h}^\star) \in \mathcal{S}$ because

$$\begin{aligned}
f(P_{V_n}(\boldsymbol{h}^\star)) &= \sum_{j=1}^{m} \big(b(\boldsymbol{z}_j) - a(P_{V_n}(\boldsymbol{h}^\star), \boldsymbol{z}_j)\big)^2 \\
&= \sum_{j=1}^{m} \big(\langle \boldsymbol{a}_j, \boldsymbol{h}^\star \rangle - \big\langle \boldsymbol{a}_j, P_{V_n}(\boldsymbol{h}^\star) \big\rangle\big)^2 \\
&= \sum_{j=1}^{m} \big(\big\langle \boldsymbol{a}_j, P_{V_n}^{\perp}(\boldsymbol{h}^\star) \big\rangle\big)^2 \\
&\leq \gamma^2 \big\| P_{V_n}^{\perp}(\boldsymbol{h}^\star) \big\|^2 \\
&\leq \gamma^2 \epsilon_n^2.
\end{aligned} \tag{35}$$

Moreover, $\hat{\boldsymbol{h}}_{\text{MS}} \in \mathcal{S}$. This can be seen from the following arguments. First, $P_{V_n}(\boldsymbol{h}^\star)$ is a feasible point for problem (7), that is

$$\text{dist}(P_{V_n}(\boldsymbol{h}^\star), V_k) \leq \hat{\epsilon}_k \text{ for } k = 0 \ldots n. \tag{36}$$

Indeed, rewriting $\boldsymbol{h}^\star$ as

$$\boldsymbol{h}^\star = \sum_{j=1}^{n} \langle \boldsymbol{v}_j, \boldsymbol{h}^\star \rangle \boldsymbol{v}_j + \boldsymbol{z}, \tag{37}$$

where $\boldsymbol{z} \in V_n^{\perp}$, we have

$$\begin{aligned}
\hat{\epsilon}_k &\geq \operatorname{dist}(\boldsymbol{h}^\star, V_k) \\
&= \left\| P_{V_k}^{\perp}(\boldsymbol{h}^\star) \right\| \\
&= \left\| \sum_{j=k+1}^{n} \langle \boldsymbol{v}_j, \boldsymbol{h}^\star \rangle \boldsymbol{v}_j + \boldsymbol{z} \right\| \\
&= \sqrt{\left\| \sum_{j=k+1}^{n} \langle \boldsymbol{v}_j, \boldsymbol{h}^\star \rangle \boldsymbol{v}_j \right\|^2 + \|\boldsymbol{z}\|^2} \\
&\geq \left\| \sum_{j=k+1}^{n} \langle \boldsymbol{v}_j, \boldsymbol{h}^\star \rangle \boldsymbol{v}_j \right\| \\
&= \left\| P_{V_k}^{\perp}\left(P_{V_n}(\boldsymbol{h}^\star)\right) \right\| \\
&= \operatorname{dist}(P_{V_n}(\boldsymbol{h}^\star), V_k).
\end{aligned} \tag{38}$$

The first inequality follows from our initial assumption $\boldsymbol{h}^\star \in \cap_{k=0}^{n} \mathcal{S}_k$. The third equality is true because $\boldsymbol{z} \in V_n^{\perp}$. Now, since $\hat{\boldsymbol{h}}_{\mathrm{MS}}$ is a minimizer of $f(\boldsymbol{h})$ over the set of feasible points, we have $f(\hat{\boldsymbol{h}}_{\mathrm{MS}}) \leq f(P_{V_n}(\boldsymbol{h}^\star)) \leq \gamma^2 \epsilon_n^2$ and therefore $\hat{\boldsymbol{h}}_{\mathrm{MS}} \in \mathcal{S}$.

We finally show that $\hat{\boldsymbol{h}}_{\mathrm{MS}} \in \mathcal{S}$ and $P_{V_n}(\boldsymbol{h}^\star) \in \mathcal{S}$ implies $P_{V_n}(\boldsymbol{h}^\star) - \hat{\boldsymbol{h}}_{\mathrm{MS}} \in \mathcal{D}_1$. Let us first note that, if $\boldsymbol{h} \in V_n$, the cost function $f(\boldsymbol{h})$ can be rewritten as:

$$\begin{aligned}
f(\boldsymbol{h}) &= \sum_{j=1}^{m} (b(\boldsymbol{z}_j) - a(\boldsymbol{h}, \boldsymbol{z}_j))^2 \\
&= \sum_{j=1}^{m} (\langle \boldsymbol{a}_j, \boldsymbol{h}^\star \rangle - \langle \boldsymbol{a}_j, \boldsymbol{h} \rangle)^2, \\
&= \sum_{j=1}^{m} \left(\left\langle \boldsymbol{a}_j^*, \boldsymbol{h}^\star \right\rangle - \left\langle \boldsymbol{a}_j^*, \boldsymbol{h} \right\rangle\right)^2, \\
&= \sum_{j=1}^{n} \left(\left\langle \boldsymbol{a}_j^*, \boldsymbol{h}^\star \right\rangle - \sigma_j \left\langle \boldsymbol{v}_j^*, \boldsymbol{h} \right\rangle\right)^2 + \sum_{j=n+1}^{m} \left\langle \boldsymbol{a}_j^*, \boldsymbol{h}^\star \right\rangle^2,
\end{aligned} \tag{39}$$

where the third equality follows from the fact that $\{\boldsymbol{a}_j\}_{j=1}^{m}$ and $\left\{\boldsymbol{a}_j^*\right\}_{j=1}^{m}$ differ up to an orthonormal transformation; the last equality is a consequence of (32) and the fact that $\boldsymbol{h} \in V_n$ by hypothesis.

We note that $P_{V_n}(\boldsymbol{h}^\star) - \hat{\boldsymbol{h}}_{\mathrm{MS}}$ can be written as $\sum_{j=1}^{n} \beta_j \boldsymbol{v}_j^*$ by setting

$\beta_j = \langle \boldsymbol{v}_j^*, P_{V_n}(\boldsymbol{h}^\star)\rangle - \left\langle \boldsymbol{v}_j^*, \hat{\boldsymbol{h}}_{\mathrm{MS}} \right\rangle$. Therefore, we have

$$\begin{aligned}
\sum_{j=1}^n \sigma_j^2 \beta_j^2 =& \sum_{j=1}^n \Big(\sigma_j \langle \boldsymbol{v}_j^*, P_{V_n}(\boldsymbol{h}^\star)\rangle - \sigma_j \left\langle \boldsymbol{v}_j^*, \hat{\boldsymbol{h}}_{\mathrm{MS}} \right\rangle \Big)^2, \\
=& \sum_{j=1}^n \Big(\sigma_j \langle \boldsymbol{v}_j^*, P_{V_n}(\boldsymbol{h}^\star)\rangle - \langle \boldsymbol{a}_j^*, \boldsymbol{h}^\star \rangle - \sigma_j \left\langle \boldsymbol{v}_j^*, \hat{\boldsymbol{h}}_{\mathrm{MS}} \right\rangle + \langle \boldsymbol{a}_j^*, \boldsymbol{h}^\star \rangle \Big)^2, \\
\leq& 2 \sum_{j=1}^n \big(\sigma_j \langle \boldsymbol{v}_j^*, P_{V_n}(\boldsymbol{h}^\star)\rangle - \langle \boldsymbol{a}_j^*, \boldsymbol{h}^\star \rangle \big)^2 + 2 \sum_{j=1}^n \Big(\sigma_j \left\langle \boldsymbol{v}_j^*, \hat{\boldsymbol{h}}_{\mathrm{MS}} \right\rangle - \langle \boldsymbol{a}_j^*, \boldsymbol{h}^\star \rangle \Big)^2, \\
\leq& 2 f(P_{V_n}(\boldsymbol{h}^\star)) + 2 f(\hat{\boldsymbol{h}}_{\mathrm{MS}}), \\
\leq& 4\gamma^2 \epsilon_n^2,
\end{aligned}$$

where the first inequality follows from the standard inequality $(a+b)^2 \leq 2(a^2+b^2)$, the second from (39), and the last one from the fact that $\hat{\boldsymbol{h}}_{\mathrm{MS}} \in \mathcal{S}$ and $P_{V_n}(\boldsymbol{h}^\star) \in \mathcal{S}$.

3.1.3 Definition of $\mathcal{D}_2$

Let

$$\delta_j = \eta_j + \hat{\eta}_j, \tag{40}$$

where

$$\begin{aligned}
\eta_j &= \sum_{i=1}^n |x_{ij}| \epsilon_{i-1}, \\
\hat{\eta}_j &= \sum_{i=1}^n |x_{ij}| \hat{\epsilon}_{i-1},
\end{aligned} \tag{41}$$

and the x_{ij}'s are the elements of the matrix $\mathbf{X}$ appearing in the SVD decomposition (29). We define $\mathcal{D}_2$ as

$$\mathcal{D}_2 = \left\{ \boldsymbol{d} = \sum_{j=1}^n \beta_j \boldsymbol{v}_j^* : |\beta_j| \leq \eta_j \right\}. \tag{42}$$

We show hereafter that $P_{V_n}(\boldsymbol{h}^\star) - \hat{\boldsymbol{h}}_{\mathrm{MS}} \in \mathcal{D}_2$.

We first note that if $\boldsymbol{h}$ is feasible for problem (7), we must have

$$\left| \langle \boldsymbol{v}_j^*, \boldsymbol{h} \rangle \right| \leq \hat{\eta}_j. \tag{43}$$

Indeed, if $\boldsymbol{h}$ is feasible, the constraint $\mathrm{dist}(\boldsymbol{h}, V_k) \leq \hat{\epsilon}_k$ simply writes as

$$\sum_{j=k+1}^n \langle \boldsymbol{v}_j, \boldsymbol{h} \rangle^2 \leq \hat{\epsilon}_k^2.$$

In particular, this implies that

$$|\langle \boldsymbol{v}_{k+1}, \boldsymbol{h} \rangle| \leq \hat{\epsilon}_k.$$

Using the fact that

$$\boldsymbol{v}_j^* = \sum_{k=1}^{n} x_{kj} \boldsymbol{v}_k,$$

we obtain (43). In a similar way, we can find that

$$\left| \left\langle \boldsymbol{v}_j^*, P_{V_n}(\boldsymbol{h}^\star) \right\rangle \right| \leq \eta_j, \tag{44}$$

by using the fact that $\mathrm{dist}(P_{V_n}(\boldsymbol{h}^\star), V_k) \leq \epsilon_k$ from (38).

Let us now show that $P_{V_n}(\boldsymbol{h}^\star) - \hat{\boldsymbol{h}}_{\mathrm{MS}} \in \mathcal{D}_2$. We first note that $P_{V_n}(\boldsymbol{h}^\star) - \hat{\boldsymbol{h}}_{\mathrm{MS}}$ can be written as $\sum_{j=1}^{n} \beta_j \boldsymbol{v}_j^*$ by setting $\beta_j = \left\langle \boldsymbol{v}_j^*, P_{V_n}(\boldsymbol{h}^\star) \right\rangle - \left\langle \boldsymbol{v}_j^*, \hat{\boldsymbol{h}}_{\mathrm{MS}} \right\rangle$. This leads to

$$\begin{aligned}
|\beta_j| &= \left| \left\langle \boldsymbol{v}_j^*, P_{V_n}(\boldsymbol{h}^\star) \right\rangle - \left\langle \boldsymbol{v}_j^*, \hat{\boldsymbol{h}}_{\mathrm{MS}} \right\rangle \right|, \\
&\leq \left| \left\langle \boldsymbol{v}_j^*, P_{V_n}(\boldsymbol{h}^\star) \right\rangle \right| + \left| \left\langle \boldsymbol{v}_j^*, \hat{\boldsymbol{h}}_{\mathrm{MS}} \right\rangle \right|, \\
&\leq \hat{\eta}_j + \eta_j = \delta_j,
\end{aligned}$$

where the last inequality follows from (43) and (44).

3.2 Expression of $\sup_{\boldsymbol{d} \in \mathcal{D}} \|\boldsymbol{d}\|^2$

We consider the following problem:

$$\sup_{\boldsymbol{d} \in \mathcal{D}} \|\boldsymbol{d}\|^2 = \sup_{\boldsymbol{\beta}} \|\boldsymbol{\beta}\|^2 \quad \text{subject to} \quad \left\{ \begin{array}{l} \sum_{j=1}^{n} \sigma_j^2 \beta_j^2 \leq 4\gamma^2 \epsilon_n^2 \\ |\beta_j| \leq \delta_j \end{array} \right. . \tag{45}$$

If $\sum_{j=1}^{n} \sigma_j^2 \delta_j^2 \leq 4\gamma^2 \epsilon_n^2$, the first constraint in (45) is always inactive and the solution simply reads

$$\sup_{\boldsymbol{d} \in \mathcal{D}} \|\boldsymbol{d}\|^2 = \sum_{j=1}^{n} \delta_j^2. \tag{46}$$

If $\sum_{j=1}^{n} \sigma_j^2 \delta_j^2 \geq 4\gamma^2 \epsilon_n^2$, the solution of (45) is given by

$$\sup_{\boldsymbol{d} \in \mathcal{D}} \|\boldsymbol{d}\|^2 = \sum_{j=\ell+1}^{n} \delta_j^2 + \rho\, \delta_\ell^2, \tag{47}$$

where ℓ is the largest integer such that

$$\sum_{j=\ell}^{n} \sigma_j^2 \delta_j^2 \geq 4\gamma^2 \epsilon_n^2, \tag{48}$$

and $\rho \in [0, 1]$ is defined as

$$\rho \sigma_\ell^2 \delta_\ell^2 + \sum_{j=\ell+1}^{n} \sigma_j^2 \delta_j^2 = 4\gamma^2 \epsilon_n^2. \tag{49}$$

This can be seen by verifying the optimality condition of problem (45). We note that problem (45) is the same (up to some constants) to the one considered in [7, Section 3.1]. The solution (47) is therefore similar, up to some different constants, to the one obtained in that paper.

Acknowledgement

The authors thank the "Agence nationale de la recherche" for its financial support through the Geronimo project (ANR-13-JS03-0002).

References

[1] A. Quarteroni, A. Manzoni, and F. Negri, *Reduced Basis Methods for Partial Differential Equations*, vol. 92. Springer International Publishing, 2016.

[2] G. Berkooz, P. Holmes, and J. L. Lumley, "The proper orthogonal decomposition in the analysis of turbulent flows," *Annual Review of Fluid Mechanics*, vol. 25, no. 1, pp. 539–575, 1993.

[3] C. Herzet, M. Diallo, and P. Héas, "Beyond petrov-galerkin projection by using multi-space prior," in *European Conference on Numerical Mathematics and Advanced Applications (Enumath'17)*, (Voss, Norway), September 2017.

[4] C. Herzet, M. Diallo, and P. Héas, "Beyond Petrov-Galerkin projection by using multi-space prior," in *Model Reduction of Parametrized Systems IV (MoRePaS'18)*, (Nantes, France), April 2018.

[5] J. Argaud, B. Bouriquet, H. Gong, Y. Maday, and O. Mula, "Stabilization of (G)EIM in presence of measurement noise: application to nuclear reactor physics," in *Spectral and High Order Methods for Partial Differential Equations ICOSAHOM 2016*, pp. 133–145, Springer, 2017.

[6] L. Fick, Y. Maday, A. T. Patera, and T. Taddei, "A Reduced Basis Technique for Long-Time Unsteady Turbulent Flows," *ArXiv e-prints*, Oct. 2017.

[7] P. Binev, A. Cohen, W. Dahmen, R. DeVore, G. Petrova, and P. Wojtaszczyk, "Data assimilation in reduced modeling." arXiv:1506.04770, June 2015.

[8] J. Xu and L. Xikatanov, "Some observation on Babuska and Brezzi theories," *Numer. Math.*, vol. 94, no. 1, pp. 195–202, 2003.