Smoothness of Subspace-valued Maps

Noah Harris, Matthew Robbins, Marie-Hélène Tomé

May 2022

1 Introduction

In this paper we will apply Gaussian Process Subspace (GPS) prediction for general subspace valued maps to reduced order modeling. To begin we will introduce various classical methods of reduced order modeling and prove that when the initial Linear Time Invariant (LTI) system is a C^k function of a parameter space $\Theta \subset \mathbb{R}^d$, the reduced system generated by a particular method of model order reduction has a particular smoothness structure. We will find that for different methods of model order reduction we can create stronger results regarding the smoothness of the reduced system. Once we have proved smoothness, we will begin to make predictions using GPS. We will numerically create sample points, each point consisting of parameter values and a basis for a subspace, for both model order reduction methods and smaller problems. For each case, most sample points will be used to create GPS models for the respective problems and the rest will be used to test the error of our models. We will then quantify our error using the Grassman Distance, a common metric for the Grassman Manifold, and find sample sizes such that our model is reasonably accurate.

2 Smoothness of Rational Krylov Interpolation

Lemma 1. *Let $X \in \mathbb{R}^{n\times n}$, $\mathbf{y} \in \mathbb{R}^n$, and $\mathbf{z} \in \mathbb{R}^{1\times n}$. Assume $(X, \mathbf{y}, \mathbf{z})(\theta) \in C^x$ for $x \in \mathbb{Z}_{\geq 0} \cup \{\infty, \omega\}$. If*

$$\mathcal{V} = Ran(V(\theta)) = \mathcal{K}_m(X(\theta), \mathbf{y}(\theta)) = colsp([\mathbf{y}(\theta)\ X(\theta)\mathbf{y}(\theta), ..., X(\theta)^{m-1}\mathbf{y}(\theta)]),$$

$$\mathcal{W} = Ran(W(\theta)) = \mathcal{K}_m(X(\theta)^T, \mathbf{z}(\theta)^T) = colsp([\mathbf{z}(\theta)^T\ X(\theta)^T\mathbf{z}(\theta)^T, ..., X(\theta)^{T^{m-1}}\mathbf{z}(\theta)^T]),$$

then the bases to $\mathcal{V}, \mathcal{W}$ given by the columns of $V(\theta)$ and $W(\theta)$ respectively are in C^x.

Proof. The vector $\mathbf{y}$ is a C^x function of the parameters by assumption. Since matrix-vector multiplication is a linear combination of the columns of the matrix according to the entries of the vector, given a matrix $X(\theta) \in C^x$ acting on a vector $\mathbf{y}(\theta) \in C^x$, the resulting vector $X(\theta)\mathbf{y}(\theta)$ is C^x. Therefore, the vector $X(\theta)\mathbf{y}(\theta)$ is a C^x function of the parameters: the mapping

$$T : \mathbb{C}^n \to \mathbb{C}^n, \quad T(\mathbf{x}) \to X(\theta)\mathbf{x} \tag{1}$$

is smooth. Define:

$$h_1 : T(\mathbf{y}(\theta)), \quad h_1(\theta) \to X(\theta)\mathbf{y}(\theta). \tag{2}$$

The mapping h_1 is smooth. Then, by smoothness of compositions, the mapping

$$h_2 : T \circ h_1, \quad h_2(\theta) \to X(\theta)^2\mathbf{y}(\theta) \tag{3}$$

is also smooth. By induction, it is then true that the mapping

$$h_n : T \circ h_{n-1}, \quad h_n(\theta) \to X(\theta)^n\mathbf{y}(\theta) \tag{4}$$

is smooth. Therefore, for any n, the vectors $X(\theta)^n\mathbf{y}(\theta)$ are smooth functions of θ. Define:

$$f(\theta) = span(h_0(\theta), h_1(\theta), ..., h_{m-1}(\theta)). \tag{5}$$

Since the span operation is smooth, the span of smooth vectors is a smooth mapping. Therefore, f is a smooth mapping. The basis to the Krylov subspace $\mathcal{K}_m = (X(\theta), \mathbf{y}(\theta))$ is a C^x function of the parameter vector θ.

The transpose operation is smooth. Given that $X(\theta), \mathbf{z}(\theta) \in C^x$, $X(\theta)^T, (z)^T(\theta) \in C^x$. Using the argument above for $X(\theta)^T$ and the vector $\mathbf{z}(\theta)^T$, we conclude the columns of the matrix $W(\theta)$ are C^x functions of the parameter vector. □

Lemma 2. *Let $(X, Y)(\theta) \in C^x$ for $x \in \mathbb{Z}_{\geq 0} \cup \{\infty, \omega\}$, and $\sigma X - Y$ be invertible for $\sigma \in \mathbb{C}$, the mapping*

$$f(X(\theta), Y(\theta)) \to (\sigma X(\theta) - Y(\theta))^{-m}$$

is C^x for any m.

Proof. Since scalar multiplication and matrix addition are smooth functions, by smoothness of compositions, given system matrices X, Y that are C^x functions of a parameter vector θ, the mapping

$$g(\sigma, X(\theta), Y(\theta)) \to X(\theta) - Y(\theta)$$

is C^x. The inverse mapping, when defined, is the composition of determinant functions (Cramer's rule) which are smooth. Since the matrix pencil $\sigma X - Y$ is invertible by assumption, the inverse mapping is smooth.

By the same logic as above, taking powers of a smooth matrix is a smooth function: the mapping

$$h_m(X) \to X^m$$

is smooth. By smoothness of compositions,

$$f(m, \sigma, X(\theta), Y(\theta)) := h_m \circ inv \circ g(\sigma, X(\theta), Y(\theta)) \to [(\sigma X(\theta) - Y(\theta))^{-1}]^m$$

is C^x. □

Theorem 3. *Let $E(\theta), A(\theta), B(\theta), C(\theta) \in C^x$ for $x \in \mathbb{Z}_{\geq 0} \cup \{\infty, \omega\}$. Let $\mathcal{S}_l = \{\mu_1, \mu_2, ..., \mu_K\} \subset \mathbb{C}$ be a set of K left interpolating points closed under conjugation, $\mathcal{S}_l \cap \lambda(A(\theta)) = \emptyset$ and let the left nontrivial tangent directions be given by $\{\mathbf{c}_i\}_{i=1}^K$. Let $\mathcal{S}_r = \{\sigma_1, \sigma_2, ..., \sigma_K\} \subset \mathbb{C}$ be a*

set of K right interpolating points closed under conjugation, $\mathcal{S}_r \cap \lambda(A(\theta)) = \emptyset$ and let the right nontrivial tangent directions be given by $\{\mathbf{b}_i\}_{i=1}^k$. Then the bases given by

$$\left\{(\sigma_i E(\theta) - A(\theta))^{-1} B(\theta)\mathbf{b}_i\right\}_{i=1}^K$$

and

$$\left\{(\mathbf{c}_i^T C(\theta)(\mu_i E(\theta) - A(\theta))^{-1})^T\right\}_{i=1}^K$$

constructed by tangential rational interpolation are C^x functions of θ.

Proof. By Lemma 2, given system matrices $E(\theta), A(\theta), B(\theta), C(\theta) \in C^x$ and considering $m = 1$, the mapping

$$f(E(\theta), A(\theta)) \to (\sigma E(\theta) - A(\theta))^{-1}$$

is C^x. Also, by the same token

$$f(E(\theta), A(\theta)) \to (\mu E(\theta) - A(\theta))^{-1}$$

is C^x. By Lemma 1, the basis vectors of the Krylov subspace are C^x functions of the parameter vector θ. Consider the vectors $\mathbf{v}_i = B(\theta)\mathbf{b}_i$ and $\mathbf{w}_i^T = \mathbf{c}_i^T C(\theta)$ for $i = 1, ..., K$. These vectors are formed from C^x matrices and constant vectors, therefore, they are C^x functions of θ. Applying Lemma 1 to the matrix $(\sigma_i E(\theta) - A(\theta))^{-1}$ and the vector $\mathbf{v}_i$ for $i = 1, ..., K$ arbitrary and considering the Krylov subspace with $m = 1$, we find that each of the basis vectors

$$\left\{(\sigma_i E(\theta) - A(\theta))^{-1}\mathbf{v}_i\right\}_{i=1}^K$$

are C^x. By the same logic, each of the basis vectors given by

$$\left\{\mathbf{w}_i^T(\mu_i E(\theta) - A(\theta))^{-1}\right\}_{i=1}^K$$

are C^x functions. Since span is a smooth function, by smoothness of compositions, for fixed right interpolating frequencies $\{\sigma_i\}_{i=1}^k$, right nontrivial tangent directions $\{\mathbf{b}_i\}_{i=1}^k$, left interpolating frequencies $\{\mu_i\}_{i=1}^k$ and left nontrivial tangent directions $\{\mathbf{c}_i\}_{i=1}^k$, the mappings

$$g : \theta \to span(\left\{(\sigma_i E(\theta) - A(\theta))^{-1} B(\theta)\mathbf{b}_i\right\}_{i=1}^K)$$

and

$$h : \theta \to span(\left\{(\mathbf{c}_i^T C(\theta)(\mu_i E(\theta) - A(\theta))^{-1})^T\right\}_{i=1}^K)$$

are C^x. □

Corollary 4. *Let $A(\theta), E(\theta) \in \mathbb{R}^{n\times n}$ and $\mathbf{b}, \mathbf{c}^T \in \mathbb{R}^n$ be C^x for $x \in \mathbb{Z}_{\geq 0} \cup \{\infty, \omega\}$. Let $\mathcal{S} = \{\sigma_1, \sigma_2, ..., \sigma_K\} \subset \mathbb{C}$ be a set of K distinct interpolating points with multiplicities $\mathcal{M} = \{m_{\sigma_1}, m_{\sigma_2}, ..., m_{\sigma_K}\}$ closed under conjugation and $S \cap \lambda(A(\theta)) = \emptyset$. Define*

$$\mathcal{V} = \bigcup_{i=1}^K \mathcal{V}_{\sigma_i}(\theta) = \bigcup_{i=1}^K \mathcal{K}_{m_{\sigma_i}}(\sigma_i E(\theta) - A(\theta))^{-1}, \mathbf{b})$$

,

$$\mathcal{W} = \bigcup_{i=1}^{K} \mathcal{W}_{\sigma_i}(\theta) = \bigcup_{i=1}^{K} \mathcal{K}_{m_{\sigma_i}}(\bar{\sigma}_i E(\theta)^T - A(\theta)^T)^{-1}, \mathbf{c}^T)$$

and define

$$V_{\sigma_i}(\theta) = [\mathbf{b}\ (\sigma_i E(\theta) - A(\theta))^{-1}\mathbf{b}, ..., (\sigma_i E(\theta) - A(\theta))^{-(m_{\sigma_i}-1)}\mathbf{b}]$$

$$W_{\sigma_i}(\theta) = [\mathbf{c}^T\ (\bar{\sigma}_i E(\theta)^T - A(\theta)^T)^{-1}\mathbf{c}^T, ..., (\bar{\sigma}_i E(\theta)^T - A(\theta)^T)^{-(m_{\sigma_i}-1)}\mathbf{c}^T].$$

Let $V(\theta) = [V_{\sigma_1}(\theta), V_{\sigma_2}(\theta), ..., V_{\sigma_k}(\theta)]$ *and* $W(\theta) = [W_{\sigma_1}(\theta), W_{\sigma_2}(\theta), ..., W_{\sigma_k}(\theta)]$ *be such that* $\mathcal{V} = Ran(V(\theta))$ *and* $\mathcal{W} = Ran(W(\theta))$*. Then the bases formed by the columns of* $V(\theta)$ *and the columns of* $W(\theta)$ *are* C^x*.*

Proof. By Lemma 2, given smooth system matrices $E(\theta), A(\theta)$, the mapping

$$h_1 : (E, A)(\theta) \to (\sigma E(\theta) - A(\theta))^{-1}$$

is smooth. Likewise, by Lemma 2, the mapping

$$h_1^{'} : (E, A)(\theta) \to (\bar{\sigma} E(\theta)^T - A(\theta)^T)^{-1}$$

is smooth. Applying Lemma 1 to the matrix $(\sigma_i E(\theta) - A(\theta))^{-1}$ and the vector $\mathbf{b}$, we have that since

$$Ran(V_{\sigma_i}(\theta)) = \mathcal{K}_{m_{\sigma_i}}((\sigma_i E(\theta) - A(\theta))^{-1}, \mathbf{b})$$

the columns of $V(\theta)$ are smooth functions.

Applying Lemma 1 to the matrix $(\bar{\sigma} E(\theta)^T - A(\theta)^T)^{-1}$ and the vector $\mathbf{c}^T$, we have that since

$$Ran(W_{\sigma_i}(\theta)) = \mathcal{K}_{m_{\sigma_i}}((\bar{\sigma}_i E(\theta)^T - A(\theta)^T)^{-1}, \mathbf{c}^T)$$

the columns of $W(\theta)$ are smooth functions.

Therefore, since the frequency σ_i was arbitrary, the columns of $\{V_{\sigma_i}(\theta)\}_{i=1}^{K}$ are smooth functions, and hence the basis formed by their union is a smooth function. The same reasoning holds for $\{W_{\sigma_i}(\theta)\}_{i=1}^{K}$. Therefore, the bases given by the columns of $V(\theta), W(\theta)$ are smooth functions. □

The rational Krylov method takes the union of Krylov subspaces, so, by Theorem 3, the bases given by the columns of the matrices $V(\theta)$ and $W(\theta)$ constructed for $\mathcal{V}, \mathcal{W}$ by the rational Krylov method are C^x functions.

Theorem 5. *Let* $\mathcal{C}(s, \theta)$*,* $\mathcal{K}(s, \theta)$*, and* $\mathcal{B}(s, \theta)$ *be matrix-valued functions which are* C^x *functions of the parameter* θ *for* $x \in \mathbb{Z}_{\geq 0} \cup \{\infty, \omega\}$*. Then the structured transfer function*

$$\mathcal{G}_L = \mathcal{C}(s)\mathcal{K}(s)^{-1}\mathcal{B}(s) \tag{6}$$

is a C^x *function of the parameter* θ*.*

Proof. Given that the individual matrices in the matrix function are smooth functions of the parameter θ, any linear combination of the matrices according to smooth functions of the parameter θ are smooth. Since the matrix function $\mathcal{K}(s)$ is guaranteed to be invertible, the inverse exists and is a smooth function of the parameter θ as it is the composition of smooth functions since $\mathcal{K}(s)$ is a C^x function of the parameter θ (cf. Cramer's Rule). since matrix multiplication preserves smoothness, the product of C^x matrices is a C^x function of the parameter θ. Therefore, the structured transfer function given by equation (1) is a C^x function of the parameter θ. □

Second order rational krylov is a special case of this where the matrix-valued functions are given by

$$\mathcal{C}(s) = C_p + sC_v, \quad \mathcal{K}(s) = s^2M + sE + K, \quad \mathcal{B}(s) = B_u.$$

Therefore, since the transfer function is a C^x function of the parameter θ the bases generated via second order rational krylov interpolation are ^{x} functions of the parameter θ.

3 Smoothness for Balanced Truncation

Theorem 6. *Let $\mathbf{A}(\boldsymbol{\theta}) \in C^\omega(\Theta, \mathcal{S}(n))$ where Θ is an open subset of $\mathbb{R}^d$. If the top-k repeated eigenvalues $(\lambda_i(\boldsymbol{\theta}))_{i=1}^k$ of $\mathbf{A}(\boldsymbol{\theta})$ are separated from the smaller eigenvalues, let $\mathfrak{V}_k$ be the total eigenspace associated with λ_i, $i = 1, \cdots, k$, then $\mathfrak{V}_k(\boldsymbol{\theta}) \in C^\omega(\Theta, G_{k,n})$. If $\mathbf{A}(\boldsymbol{\theta})$ is only continuous at $\boldsymbol{\theta}_0$, then $\mathfrak{V}_k(\boldsymbol{\theta})$ is continuous at $\boldsymbol{\theta}_0$.*

Proof. Since the top-k and bottom-$(n-k)$ repeated eigenvalues of $\mathbf{A}$ are separated throughout Θ, the mapping $\mathfrak{V}_k(\boldsymbol{\theta}) : \Theta \mapsto G_{k,n}$ is well-defined. To prove the first part of the theorem, it suffices to show that for any $\boldsymbol{\theta}_0 \in \Theta$, there exists $\mathbf{U}(\boldsymbol{\theta}) \in C^\omega(B(\boldsymbol{\theta}_0), M^*_{n,k})$ such that $\text{span}(\mathbf{U}) = \mathfrak{V}_k$, where $B(\boldsymbol{\theta}_0) \subset \Theta$ is a neighborhood of $\boldsymbol{\theta}_0$. Since $\mathfrak{V}_k|_{B(\boldsymbol{\theta}_0)} = \text{span} \circ \mathbf{U}$, this implies that $\mathfrak{V}_k|_{B(\boldsymbol{\theta}_0)} \in C^\omega(B(\boldsymbol{\theta}_0), G_{k,n})$. It follows that $\mathfrak{V}_k \in C^\omega(\Theta, G_{k,n})$.

Let $\lambda_1 > \cdots > \lambda_s$ be the largest eigenvalues of $\mathbf{A}(\boldsymbol{\theta}_0)$, with total multiplicity $\sum_{i=1}^s m_i = k$. Applying Theorem 12 to $\mathbf{A}(\boldsymbol{\theta})$ at $\boldsymbol{\theta}_0$, there exist $\mathbf{U}_i(\boldsymbol{\theta}) \in C^\omega(B_i(\boldsymbol{\theta}_0), M^*_{n,m_i})$, $i = 1, \cdots, s$, where $B_i(\boldsymbol{\theta}_0) \subset \Theta$ are neighborhoods of $\boldsymbol{\theta}_0$, such that $\mathbf{U}_i(\boldsymbol{\theta})$ is a basis of the total eigenspace associated with the λ_i-group. Let $\mathbf{U} = (\mathbf{U}_1, \cdots, \mathbf{U}_s)$ and $B(\boldsymbol{\theta}_0) \subset \cap_{i=1}^s B_i(\boldsymbol{\theta}_0)$ such that $\mathbf{U}(\boldsymbol{\theta}_*) \in M^*_{n,k}$ for all $\boldsymbol{\theta}_* \in B(\boldsymbol{\theta}_0)$, then $\mathbf{U}(\boldsymbol{\theta}) \in C^\omega(B(\boldsymbol{\theta}_0), M^*_{n,k})$. This proves the first part of the theorem.

The continuous version can be proved similarly. Let $\mathfrak{U}_i(\boldsymbol{\theta})$ be the total eigenspace associated with the λ_i-group, $i = 1, \cdots, s$. Applying Theorem 9, $\mathfrak{U}_i(\boldsymbol{\theta})$ are continuous at $\boldsymbol{\theta}_0$. Since $\mathfrak{V}_k = \oplus_{i=1}^s \mathfrak{U}_i$, $\mathfrak{V}_k(\boldsymbol{\theta})$ is also continuous at $\boldsymbol{\theta}_0$. □

Theorem 7. *Given a linear time-invariant, like that described in Chapter 6 of Benner et al. (2017), with matrices $A(\theta)$,$B(\theta)$, $C(\theta)$, each that are C^0 functions of the parameter θ. The space,$\mathfrak{V}_k, \mathfrak{W}_k$, onto which classical balanced truncation projects the system onto is a C^0 function of θ. Additionally, if $A(\theta)$,$B(\theta)$, $C(\theta)$, each that are C^ω and the Hankel singular values satify $\sigma_k(\theta) > \sigma_{k+1}(\theta)$, $\forall \theta \in \Theta$, then the space ,$\mathfrak{V}_k, \mathfrak{W}_k$, onto which classical balanced truncation projects the system onto is a C^ω function of θ.*

Proof. The transformations of balanced truncation are characterized by diagonilizing both the controlability gramian matrix, P, and the observability gramian Matrix, Q. To begin we need to prove that these matrices are either C^0 or C^ω functions of θ, when the least smooth of A, B, and C as functions of θ is either C^0 or C^ω, respectively.
We have the following expressions for the Gramian matrices P and Q,

$$P = \int_0^\infty e^{tA} B B^T e^{tA^T} dt \tag{7}$$

$$Q = \int_0^\infty e^{tA^T} C^T C e^{tA} dt \tag{8}$$

To begin, notice that A^T preserves the smoothness of A, because we are simply rearranging the C^0 or C^ω functions that make up A. Now, we look at the matrix exponential. The matrix exponential is best defined in terms of its Taylor series.

$$e^{tA} = I + (tA) + \frac{(tA)^2}{2!} + \frac{(tA)^3}{3!} + ...$$

Now we use Higham (2008), Thm 4.7, which proves that this series has an infinite radius of convergence and therefore converges for all $A \in M_{n\times n}$ It is also true that $C^\omega(M_n(\mathbb{C}), \mathrm{GL}_n(\mathbb{C}))$. The Taylor series can only exist for a function that is at least C^ω, so we have that this matrix exponential is as smooth as $A(\theta)$. We have that each of the four terms in the product of each integrand is as smooth as A, B, and C. Their product is similarly smooth since matrix multiplication preserves smoothness. Lastly, integrating a function, which is what the integral is doing to the individual component functions that comprise the integrand, outputs a function that is at least as smooth, therefore our Gramian matrices, P and Q, are at least as smooth as the least smooth among A, B, and C.
Now that we have P and Q, we need to prove that the balanced transformation we need can be reach through smooth mappings. To begin, we note that the next step of classical balanced truncation is to find Cholesky decompositions of P and Q. For the purpose of proving a smooth mapping exists, assume that the Cholesky factors are generated in the following way. We can write an eigenvalue decomposition of P as,

$$P = \sum_{i=1}^{k} \lambda_k P_{v_k}$$

where all the P_{v_i} matrices are projection matrices. Now, write

$$P^{\frac{1}{2}} = \sqrt{\lambda_k} P_{v_k}$$

Since we have $P_{v_i}^2 = P_{v_i}$ for all P_{v_i}, we do the Cholesky deomosition as $P = P^{\frac{1}{2}}(P^{\frac{1}{2}})^T$. This is a specific way that we can do the Cholesky decomposition for positive semi-definite matrices, which is unique, unlike general Cholesky decomposition. To prove that the mapping from P to $P^{\frac{1}{2}}$ is smooth, notice that there is a clear smooth, bijection from $P^{\frac{1}{2}}$ to P. The inverse function theorem applies and we see that this is a smooth mapping. The same process yields a smooth Cholesky decomposition of Q. So now as we move forward we denote, $P^{\frac{1}{2}}$ as R_P

and $Q^{\frac{1}{2}}$ as R_Q.
We now can get a basis for the space that we are projecting onto. To begin we take a full Singular Value Decomposition of $R_P R_Q^T$, which we write as

$$R_P R_Q^T = U\Sigma Z^T$$

Then the desired transformations for our system are written as

$$V_k = R_P^T U_k \Sigma_k^{\frac{-1}{2}} \tag{9}$$

$$Q_k = R_Q^T Z_k \Sigma_k^{\frac{-1}{2}} \tag{10}$$

Where U_k and Z_k, are based on the first k rows that are selected of the U and Z. Now we denote the space that the columns of V_k is a basis for as $\mathcal{V}_k$ and similarly define $\mathcal{W}_k$. Now to show that this space is a smooth mapping we notice that we can reduce column space of V_k.

$$\mathcal{V}_k = range(R_P^T U_k \Sigma_k^{\frac{-1}{2}}) = range(R_P^T U_k)$$

Since $\Sigma_k^{\frac{-1}{2}}$ is just a diagonal matrix and therefore does not change the span. So now we know that the subspace V_k is as smooth as the less smooth one of R_P and U_k. We have already seen that R_p is as smooth as A, B, and C. So we finish off the proof by proving the smoothness of U_k, and for the $\mathcal{W}_k$ case, we show that Z_k is also smooth. To do this we do an eigenvalue decomposition of the symmetric matrices $R_Q P R_Q$ and $R_P Q R_P$, which we write as,

$$R_Q P R_Q = Z\Sigma^2 Z^T$$
$$R_P Q R_P = U\Sigma^2 U^T$$

For both the case where R_Q and R_P are C^0 and the case where R_Q and R_P are C^ω we apply Thereom 6 that is proved in [add reference to smooth-PC.tex], in order to prove that the space we are projecting on to is C^0 or C^ω respectively. We apply our assumption that the Hankel singular values satisfy $\sigma_k(\theta) > \sigma_{k+1}(\theta)$, $\forall\theta \in \Theta$, which implies that the top-k eigenvalues of the left hand side of both equations are separated from the smaller eigenvalues and therefore we can apply Thereom 6. So we now have that $\mathfrak{V}_k$, $\mathfrak{W}_k$, the associated eigenspaces are C^0 when R_Q and R_P are C^0 and C^ω when R_Q and R_P are C^ω. □

3.1 Smoothness for Frequency-Limited Balanced Truncation

We follow the setup of Frequency-Limited Balanced Truncation layed out in [add reference Benner 2016]. In the frequency-domain, we have the following transformed Gramians.

$$P = \frac{1}{2\pi}\int_{-\infty}^{\infty}(i\omega I - A)^{-1}BB^T((i\omega I - A)^{-1})^H d\omega$$

$$Q = \frac{1}{2\pi}\int_{-\infty}^{\infty}((i\omega I - A)^{-1})^H C^T C((i\omega I - A)^{-1}) d\omega$$

For frequency-limited balanced truncation, we only select a subset of the real number line to integrate over. We impose that this subset is symmetric around the origin. In other words, let the domain of integration be Ω, defined as below,

$$\Omega = \bigcup_{i=1}^{k} [-\omega_{2i}, -\omega_{2i-1}] \cup [\omega_{2i-1}, \omega_{2i}]$$

where $0 \leq w_1 < w_2 < \ldots < w_{2k}$. Then the frequency limited Gramians are,

$$P = \frac{1}{2\pi} \int_{\Omega} (i\omega I - A)^{-1} B B^T ((i\omega I - A)^{-1})^H d\omega$$

$$Q = \frac{1}{2\pi} \int_{\Omega} ((i\omega I - A)^{-1})^H C^T C ((i\omega I - A)^{-1}) d\omega$$

Once we have our frequency-limited Gramians, we continue in the same way as classical balanced truncation, so we now endeavor to show this entire process is as smooth as the matrices that defined the initial LTI.

Theorem 8. *Given a system with matrices $A(\theta)$,$B(\theta)$, $C(\theta)$, each that are C^0 functions of the parameter θ. The space,$\mathfrak{T}_k, \mathfrak{U}_k$, onto which frequency limited balanced truncation projects the system onto is a C^0 function of θ. Additionally, if $A(\theta)$,$B(\theta)$, $C(\theta)$, each that are C^ω and the frequency-limited Hankel singular values satify $\sigma_k(\theta) > \sigma_{k+1}(\theta)$, $\forall \theta \in \Theta$, then the space ,$\mathfrak{T}_k, \mathfrak{U}_k$, onto which frequency-limited balanced truncation projects the system onto is a C^ω function of θ.*

Proof. When we proved Thereom 7, we first proved that the Gramians were as smooth as the least smooth of $A(\theta)$,$B(\theta)$, and $C(\theta)$. Then we proved that projection constructed from the Gramians is similarly as smooth. The process of creating this projection for the frequency-limited case is identical once we have our Gramians, therefore once we prove that the frequency-limited Gramians are smooth, we can just refer to the proof of Theorem 6 and replace P, Q with P_Ω,Q_Ω.
To show the frequency-limited Gramians are as smooth as the least smooth of $A(\theta)$,$B(\theta)$, and $C(\theta)$, we begin with a reference to Lemma 2, with $X = I$ and $m = 1$, to show $(i\omega I - A)^{-1}$ is as at least as smooth as A. The product in the integrand will be at least as smooth as the least smooth of its factors, which are all as least as smooth as our original system. We then integrate with respect to a variable that is not in the parameter space, which preserves smoothness structure. Therefore, each of our frequency-limited Gramians are at least as smooth as our initial LTI and we can refer to the proof of Theorem 7 to show that the reduced-system follows the same smoothness structure as the time domain case of balanced truncation. □

4 Numerical Results

4.1 Measuring the Accuracy of our Predictions

To measure the error in the predictions made by GPS after training on the sample dataset, we used the Grassman distance to measure the distance between the predicted subspace and

the subspace computed via the model reduction method we wished GPS to approximate. For each of the predictions given matrix representations of the actual and predicted subspaces V_{true} and $V_{predicted}$, respectively, we computed

$$dist(\mathcal{V}_{true}, \mathcal{V}_{predicted}) = || \arccos(\sigma(V_{predicted}^T V_{true}))||$$

where $\sigma(\cdot)$ denotes the vector of singular values of its matrix argument. The maximum value for this distance is $\frac{\pi\sqrt{k}}{2}$, where k is the dimension of the subspace to be predicted. To normalize the error metric on a more interpretable scale of from 0 to 1 we divided the distance metric by this maximal value.

4.2 Predicting Functions on the 1,2-Grassmanian Manifold

To motivate the importance of the smoothness structure of the underlying subspace-valued map in providing theoretical guidance on the accuracy of GPS predictions, we used GPS to approximate the following subspace-valued function:

$$f_\alpha : \mathbb{R} \to G_{1,2}, \;\; f_\alpha(\theta) = span\{\cos(\alpha(\theta)), \sin(\alpha(\theta))\}.$$

We varied the smoothness structure of the function α to observe the effect on the accuracy of GPS in predicting f_α. We investigated the accuracy of GPS predictions of f_α for the following variants of α:

1. $\alpha \in C^\omega$, low sensitivity to perturbations

$$\alpha(\theta) = \frac{\pi}{2}(cos\theta + 1)$$

2. $\alpha \in C^\omega$, high sensitivity to perturbations

$$\alpha(\theta) = \arccos(v_1 + sign(v_2))$$

where $\mathbf{v} = (v_1, v_2)$ is the eigenvector associated with the largest eigenvalue of the following matrix K with functional dependence on the parameter θ

$$K(\theta) = \begin{bmatrix} 0.15 + \frac{2.9}{\pi} * \theta & -0.05 \\ -0.05 & 1.05 \end{bmatrix}.$$

3. $\alpha \in C^0$

$$\alpha(\theta) = |\theta - \frac{\pi}{2}| + \frac{\pi}{4}$$

4. $\alpha \notin C^0$

$$\alpha(\theta) = \begin{cases} -0.8\frac{\pi}{2} & \theta \in [0, \frac{\pi}{2}) \\ 0.8\frac{\pi}{2} & \theta \in [\frac{\pi}{2}, \pi]. \end{cases}$$

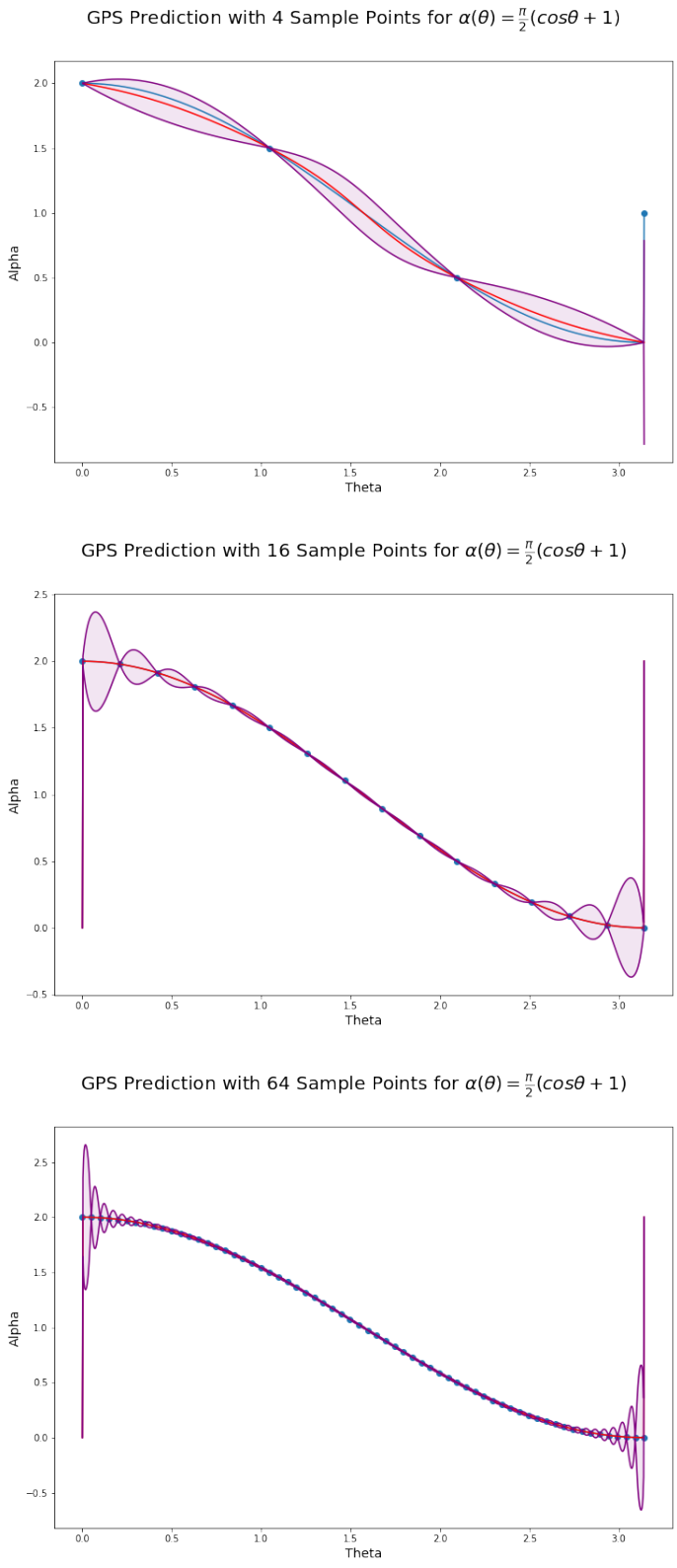


Figure 1: GPS Predictions in Case (1) Analytic Function, Low Sensitivity

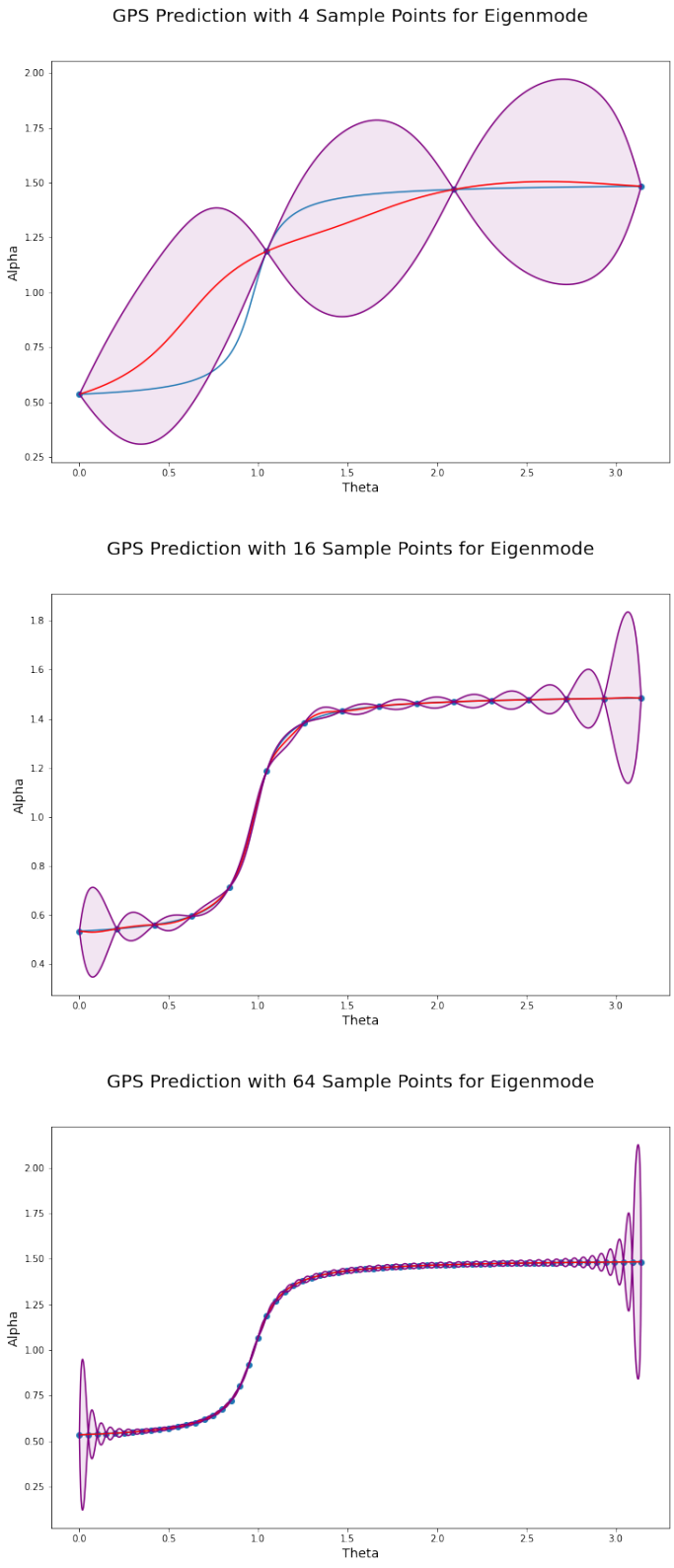


Figure 2: GPS Predictions in Case (2) Analytic Function, High Sensitivity

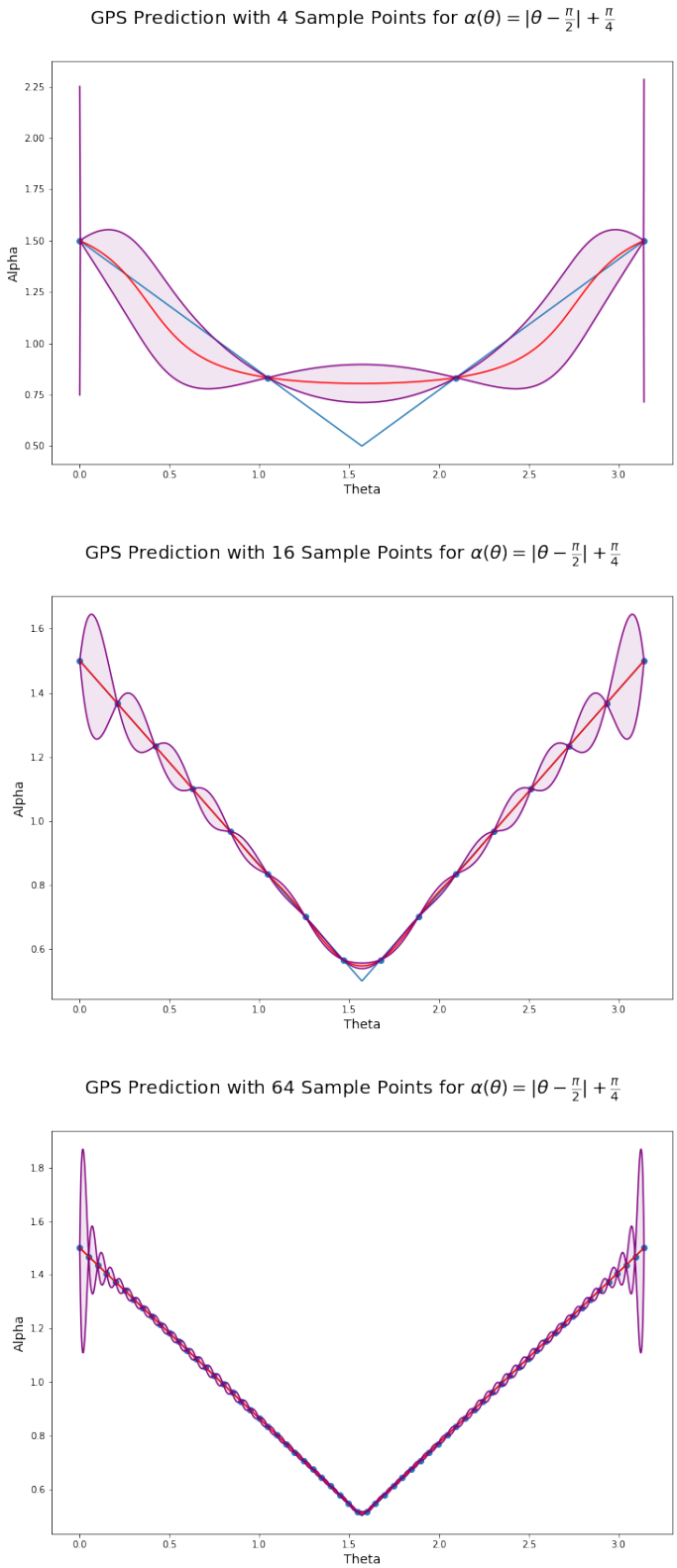
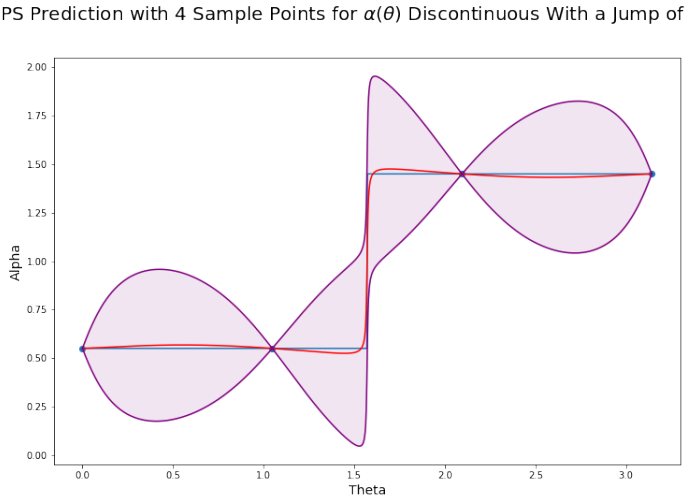
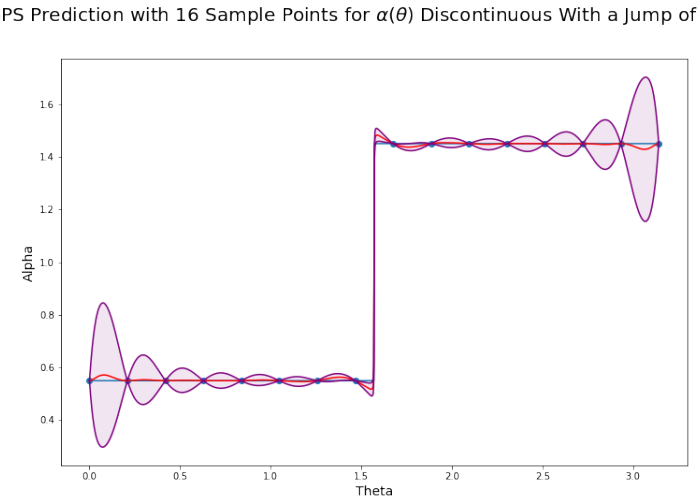


Figure 3: GPS Predictions in Case (3) Continuous Function





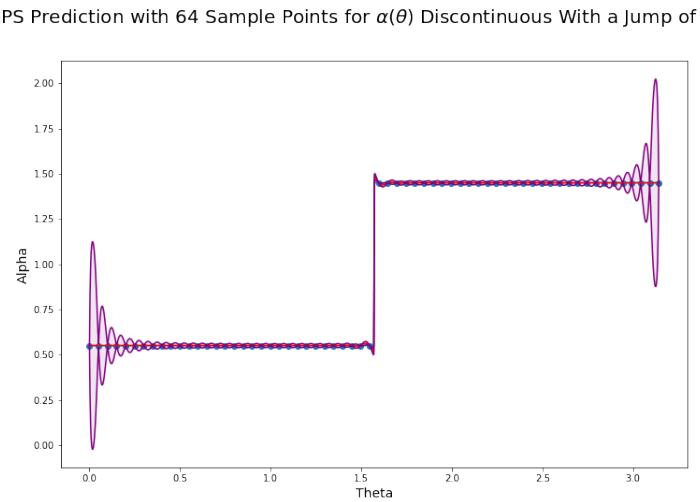


Figure 4: GPS Predictions in Case (4) Discontinuous Functions

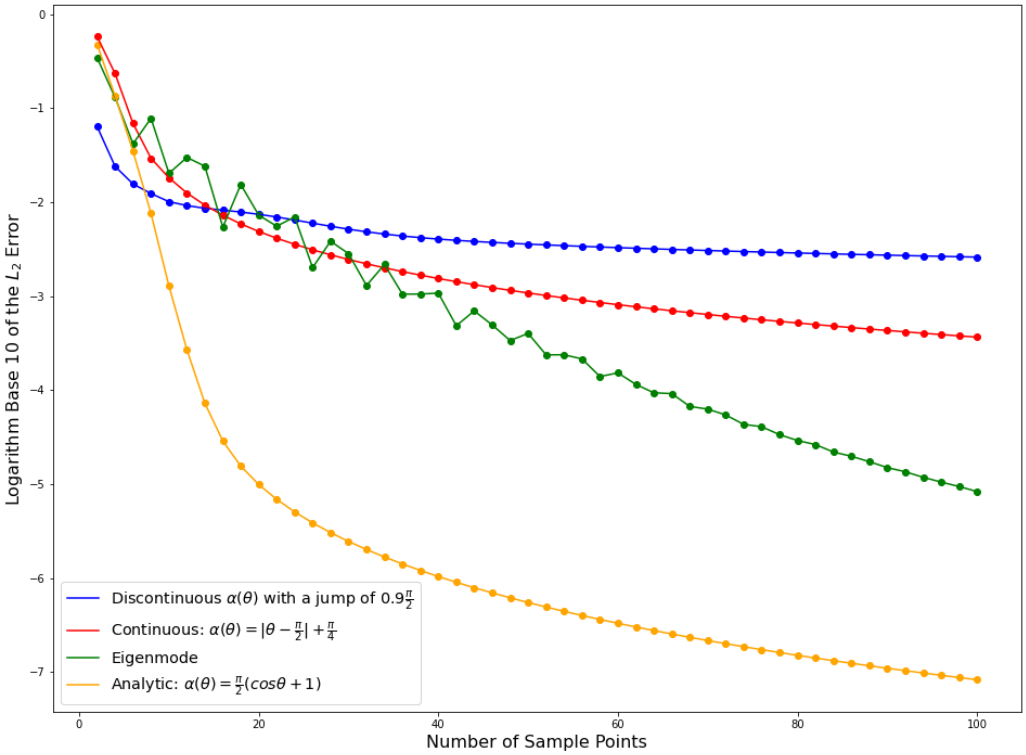


Figure 5: $\mathcal{L}_2$ Error Convergence for Subspace-valued Functions with Varying Smoothness Structure

Qualitatively, GPS predictions can be seen to improve as the number of sample points ℓ used to build the model increases. GPS can be seen to be most accurate in predicting Case (1) (cf. Figure 1). The 95% predictive interval is narrowest for the analytic, low sensitivity case for $\ell = 4$ compared to the remaining cases. In the case where $\alpha \in C^0$, qualitatively, the error in predicting the non-differentiable point at $\frac{\pi}{2}$ decreases as we increase the number of points sampled (cf. Figure 2). In the discontinuous case, the error made in approximating the discontinuity improves as we increase the number of sample points ℓ. However, since GPS predicts f_α as a smooth function, there will always be nonzero error in approximating a discontinuous function (cf. Figure 4).

Comparing the convergence rate in the $\mathcal{L}_2$ error metric (cf. Figure 5), it can be seen that the higher the sensitivity of the subspace-valued function to perturbations, the slower the rate of convergence of the error. The convergence rate is fastest for the analytic case with low sensitivity, i.e., $\alpha(\theta) = \frac{\pi}{2}(cos\theta + 1)$, followed by the analytic case with high sensitivity, the Eigenmode problem, followed by the merely continuous case, and slowest for the discontinuous case. For discontinuous functions, the error fails to converge to zero since there will always be nonzero error in approximating the finite jump in the function. Here the error is

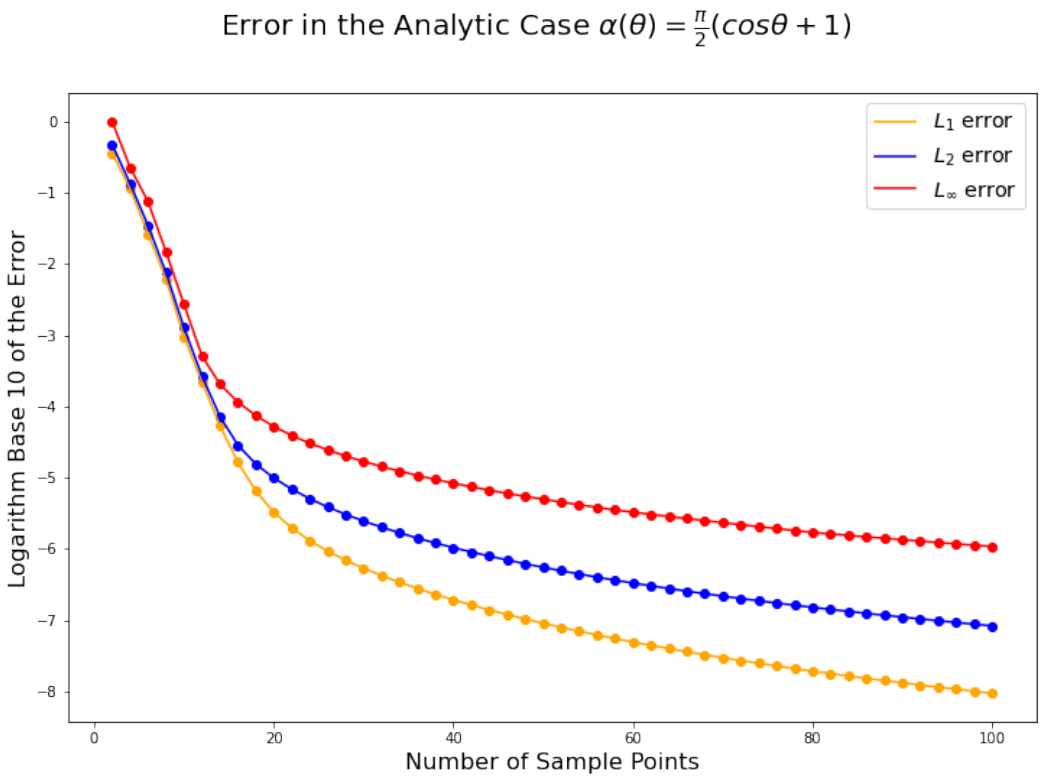


Figure 6: Error Convergence for Analytic, Low Sensitivity f_α

measured as

$$\epsilon = ||d_g(f(\theta), \tilde{f}(\theta))||_p$$

where p = $\mathcal{L}_1$, $\mathcal{L}_2$, $\mathcal{L}_\infty$.

For each number of sample points, ℓ, we selected a fixed length scale β that depended on the reciprocal of the number of sample points ℓ. As the number of sample points ℓ increase, the error in the $\mathcal{L}_1$, $\mathcal{L}_2$ and $\mathcal{L}_\infty$ norms decreases as the number of sample points l increases. In the analytic case with low sensitivity to perturbations, we can expect the fastest convergence in the error to the lowest values. On a logarithm base 10 scale, the error in approximating the subspace-valued map f_α converges linearly (cf. Figure 6).

4.3 Anemometer

Now we move to a larger problem, beginning with a LTI of order greater than 20000, namely 29008. A realization of the LTI is given by system matrices E, A, B, and C. The data we begin with consists of matrices A_1 and A_2. We define the system matrix A in terms of a parameter $p \in [0, 1]$ in the following way,

$$A = p(A_1) + (1 - p)(A_2)$$

We wish to use GPS to approximate the mapping from p to the bases computed to perform model reduction for each of the methods explored. In the sections below, we will explain

how samples were generated for each case, check any necessary conditions to ensure GPS is applicable, and use our sample data to create a GPS model and predict the subspace-mapping. We will also analyze how many samples are necessary to get good predictions as well as provide numerical results on sensitivity of the subspace-mapping when the conditions for GPS to work well are violated or nearly violated.

4.3.1 Generating a Sample Dataset Via Rational Krylov Interpolation

Using Rational Krylov interpolation, we generated left and right modeling subspaces with dimension k = 20. Both modeling subspaces were computed by choosing ten complex shift frequencies at which to interpolate the transfer function $H(s) = C^T(sE - A)^{-1}B$ of the system. For each value of the parameter p, the vectors giving a basis to the left modeling subspace were computed as $(sE-A)^{-1}B$ for each of the ten chosen complex shift frequencies. Similarly, to compute the vectors giving a basis to the right modeling subspace, we computed $(\bar{s}E-A)^{-T}C^T$ for each of the ten chosen frequencies. To generate a real set of basis vectors, the real part of the vector was taken as one basis vector and the imaginary part of the computed vector was taken as another basis vector. Hence, given 10 frequencies at which to interpolate the transfer function, we arrive at a basis of dimension 20. To generate an orthonormal set of basis vectors we normalized each of vectors and computed the singular value decomposition of the matrix of basis vectors for the left and right modeling subspaces, respectively. We took the top 20 left singular vectors as orthonormal representations of the respective subspaces they span.

4.3.2 Predicting Subspace-Valued Mappings Computed Via Rational Krylov

We predicted the subspace-valued map given by Rational Krylov interpolation with two real shifts $\{0, 10^4\}$ so that $k = 2$, two imaginary shifts $\pm 10^4 i$ so that $k = 2$, and for 20 imaginary shift frequencies $\{\pm j * 10^4 i\}_{j=1}^{10}$ so that $k = 20$.

For 50 evenly spaced values of the parameter p between 0 and 1, we generated a basis to the left and right modeling subspaces, respectively using rational krylov interpolation as discussed above. We selected 10 of the values and corresponding bases as target points for which we wanted to predict the subspace mapping using GPS. To generate the set of target values for which to predict the mapping, we selected every modulo 5 point of the 50 evenly spaced points on the interval from (0, 1). We used the remaining 40 points and corresponding bases for training. We predicted the mapping from the parameter p to the left modeling subspace and the mapping from the parameter to the right modeling subspace separately. Although information is lost in predicting the mappings separately, there is currently no methodology for using GPS to predict the mapping from a parameter p to a tuple of subspaces, i.e., the mapping

$$f : p \to (V, W).$$

The error converges linearly for each of the cases explored.

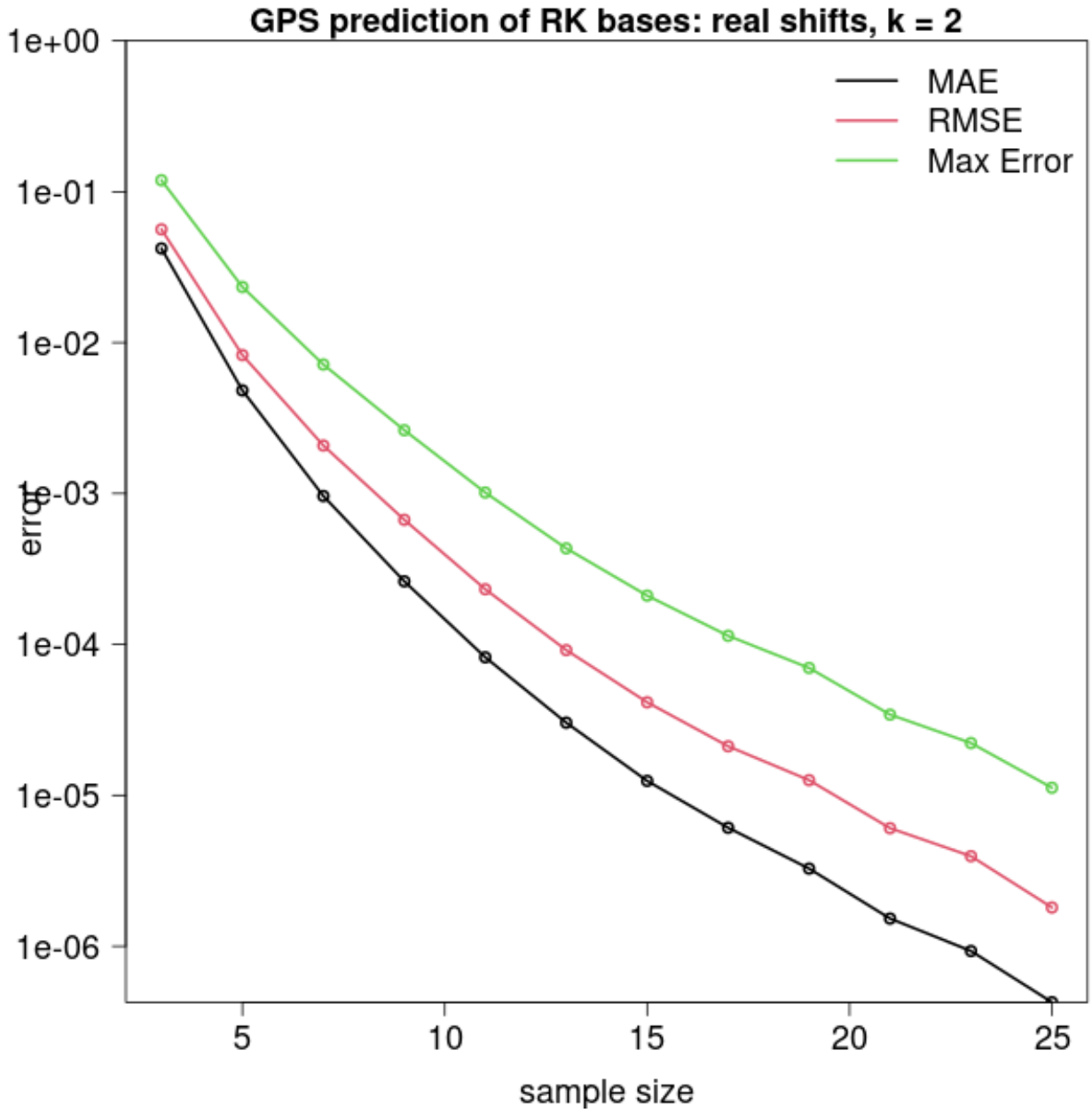


Figure 7: Error Convergence in Predicting the Subspace-Valued Map given by Rational Krylov Interpolation for $k = 2$ Real Shifts

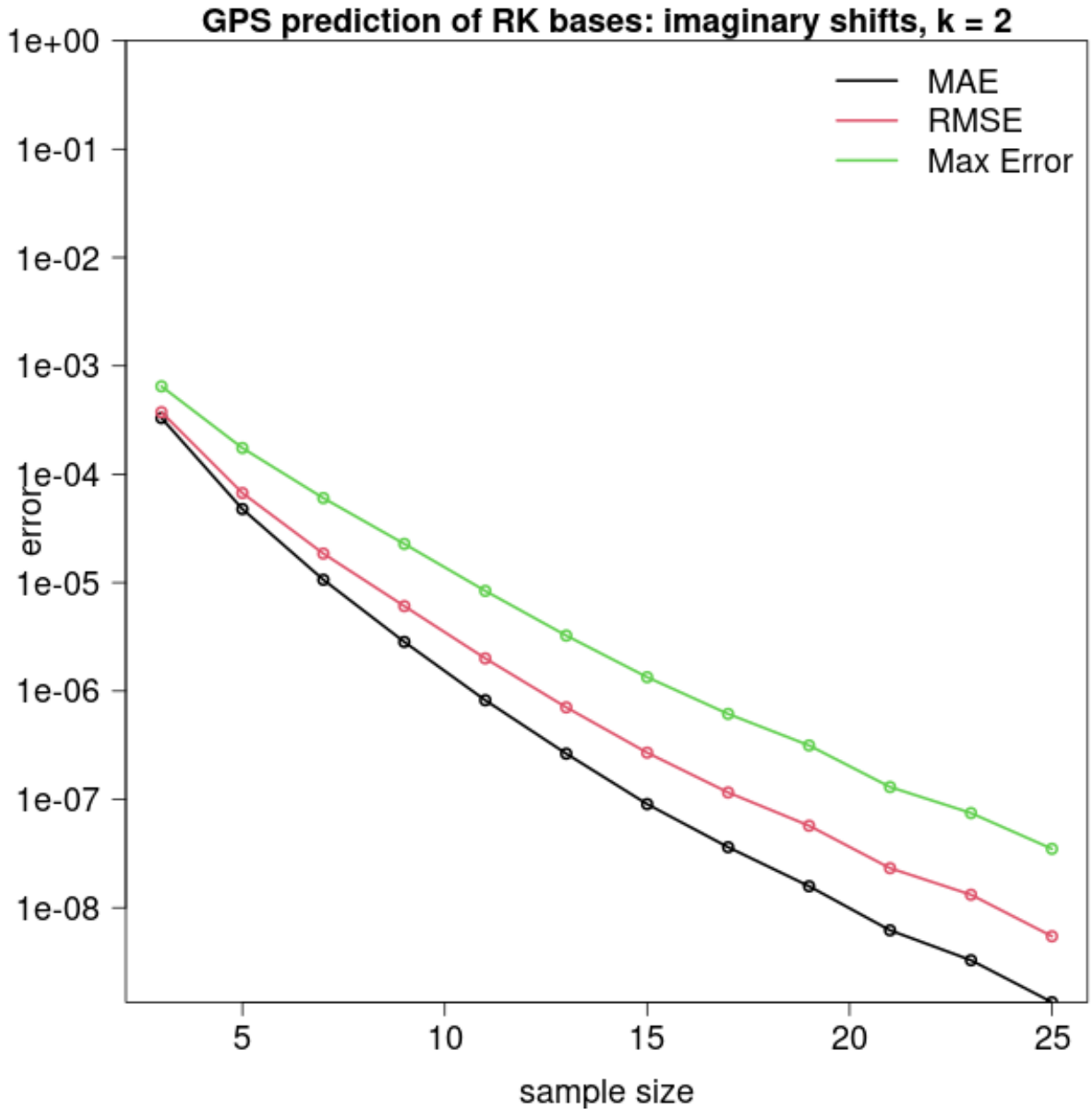


Figure 8: Error Convergence in Predicting the Subspace-Valued Map given by Rational Krylov Interpolation for $k = 2$ Imaginary Shifts

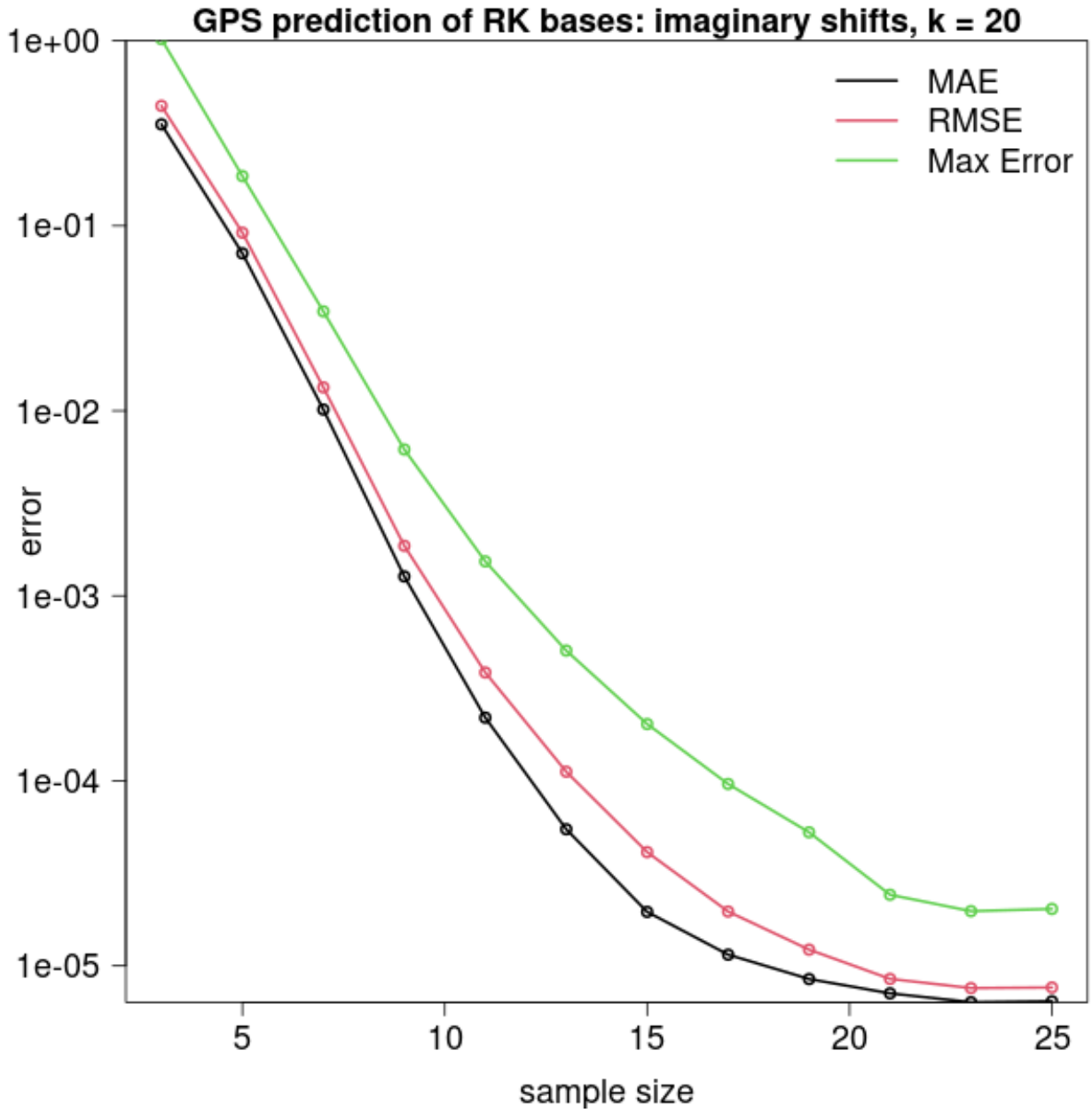
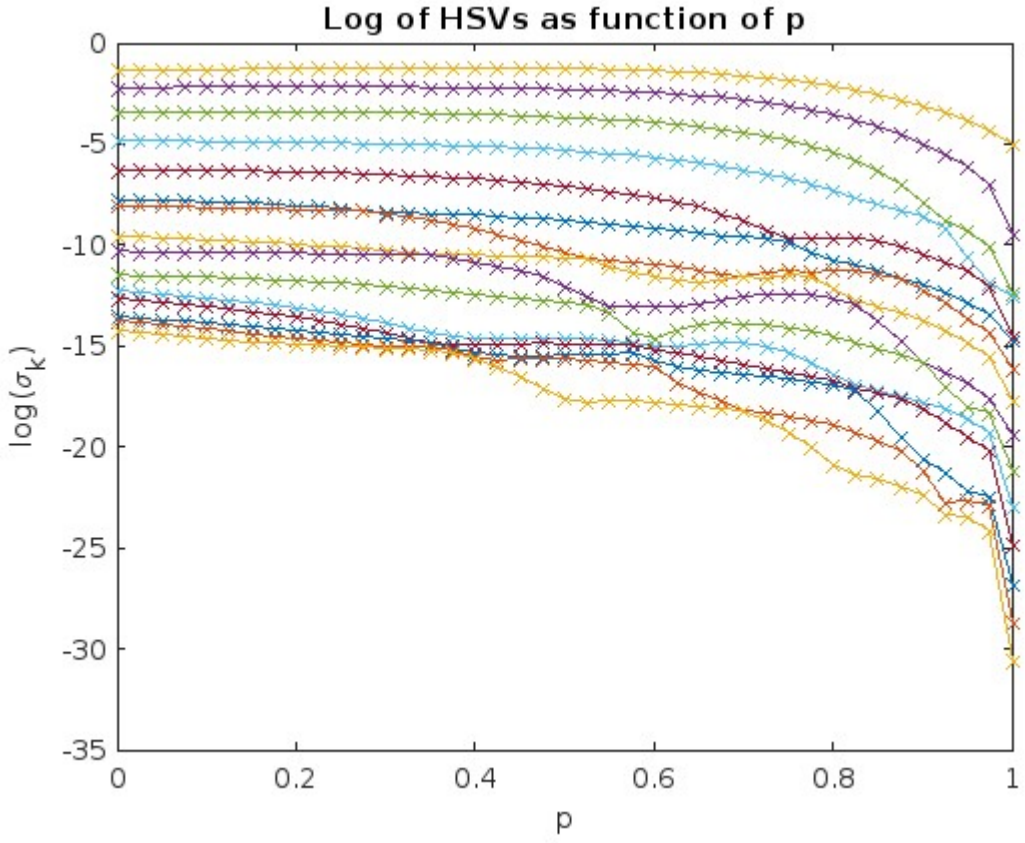


Figure 9: Error Convergence in Predicting the Subspace-Valued Map given by Rational Krylov Interpolation for $k = 20$ Imaginary Shifts

4.3.3 Predicting Subspace-Valued Mappings Computed via Balanced Truncation

Similar to the Rational Krylov, we want to create sample points consisting of a value of p and a basis for our reduced system. However, we now have to check the Hankel Singular Values condition from Theorem 6, $\sigma_k(\theta) > \sigma_{k+1}(\theta)$, to ensure that the projection is smooth. We will verify this numerically by computing the Hankel singular values for a sample of values of p. We chose to compute 21 samples for evenly spaced values of p on $[0, 1]$. We will reduce the system to order k, such that the top k Hankel singular values do not cross over our parameter space. The following plot is of the top 15 Hankel singular values over the parameter space. In the plot, we take the logarithm of the Hankel singular value to make the differences visible.



It is clear that some of the Hankel singular values that we are plotting crossing over our parameter space as there are apparent points where the singular values jump. However, the top four Hankel singular values do not appear to cross over the parameter space and therefore the transformation that reduces the system to order 4 should be C^ω according to Theorem 6. For each of the values of p that we have plotted Hankel singular values we record the matrices V and W that reduce the system to order 4, these are matrices are basis for the subspaces that we want to apply GPS prediction to. As mentioned for the Rational Krylov, we apply GPS to the left transformations, V, and right transformations, W, separately.

The error converges linearly in the case where $k = 5$ but as the persistent spectral gap assumption is relaxed (cf. Figure 11 too see the singular value crossings), the error fails to

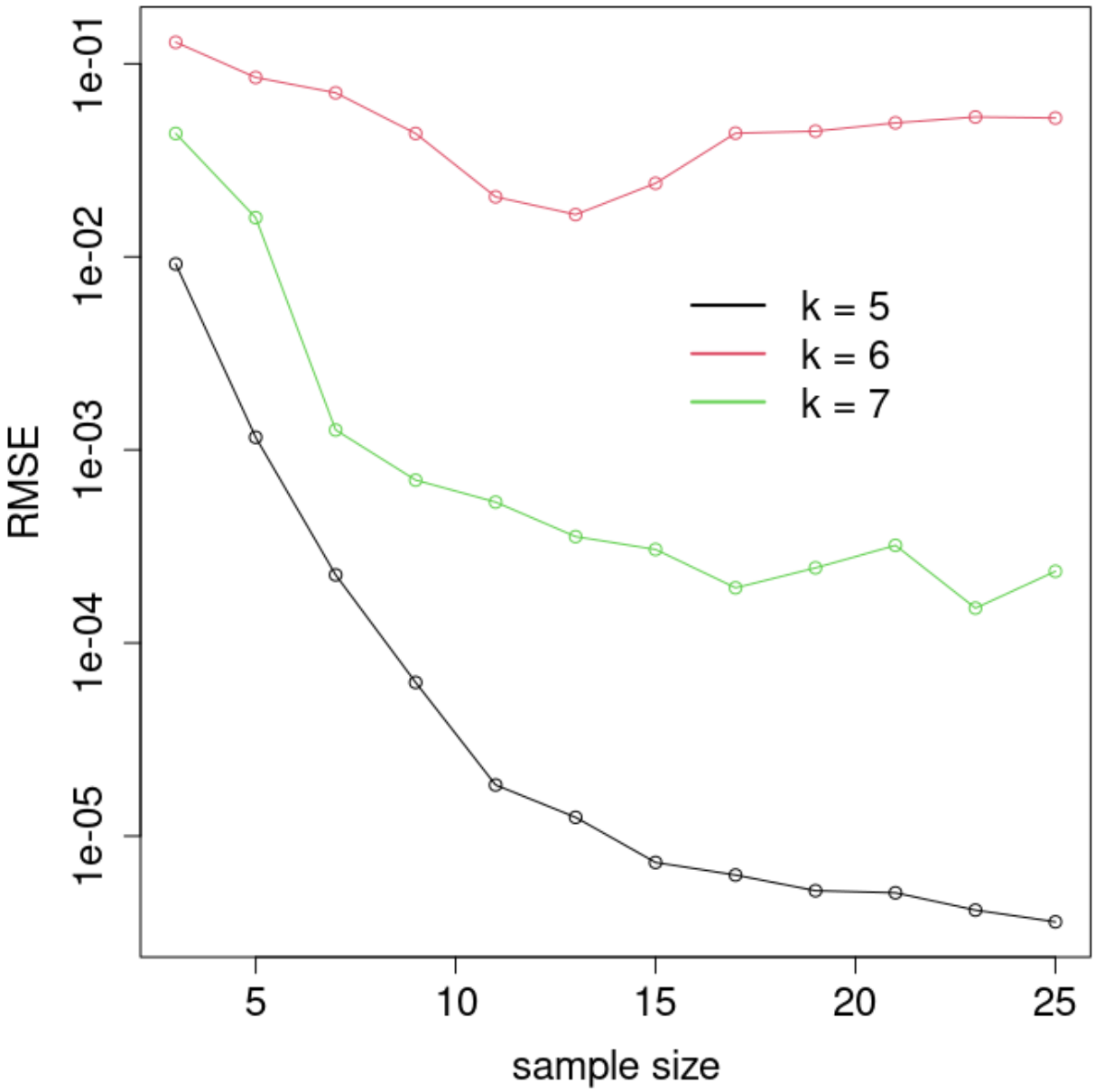


Figure 10: Error Convergence for GPS Prediction of the Subspace-valued Mapping given by Balanced Truncation

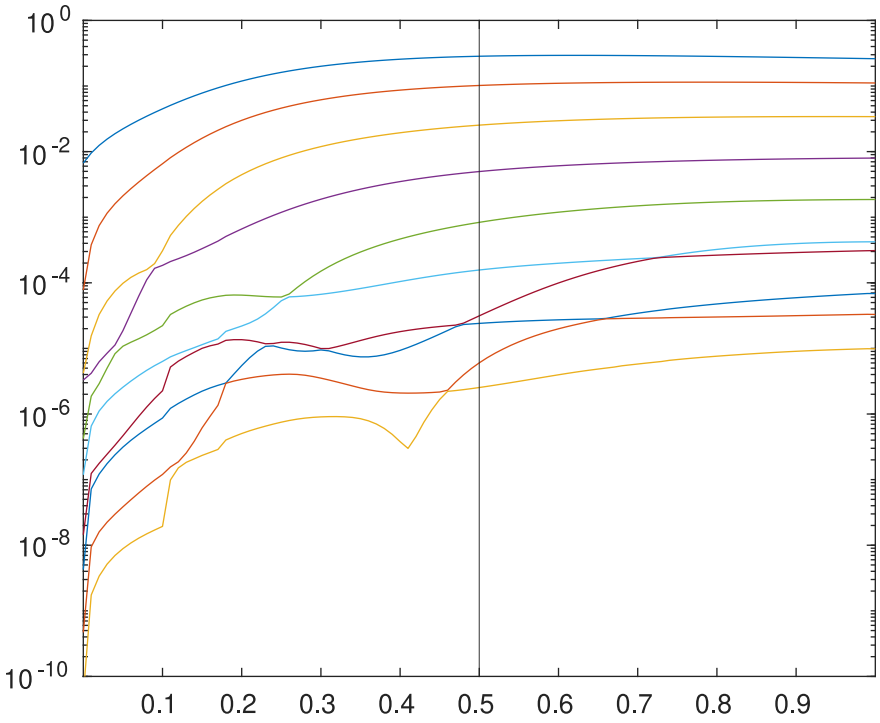


Figure 11: Hankel Singular Value Plot

converge. In the case where $k = 6$, we no longer have a persistent spectral gap between the sixth and seventh largest Hankel singular values. Therefore, convergence is worst in the case k = 6. In the case $k = 7$, the spectral gap condition is satisfied but its convergence is worse than the $k = 5$ case, which can be explained, in part, by the diminishing spectral gap as the parameter approaches 0.5 (cf. Figure 11). This demonstrates the importance of the spectral gap assumption in assuring the accuracy of GPS predictions when using Balanced Truncation.

The convergence is best for the Rational Krylov case, illustrating that the bases constructed through Rational Krylov Interpolation preserve smoothness structure for arbitrary differentiability class, while with Balanced Truncation we can only show this for continuous and analytic functions.

5 Conclusion

In this paper, we have shown various smoothness guarantees for classical methods of reduced order modeling when applied to LTI systems as functions of a parameter space $\Theta \subset \mathbb{R}^d$. For any given differentiability class (C^x for $x \in \mathbb{Z}_{\geq 0} \cup \{\infty, \omega\}$) of system matrices as functions of the parameter $\theta \in \Theta$, bases constructed through Rational Krylov Interpolation preserve this underlying differentiability class as functions of θ. This also holds for Structured Rational Krylov Interpolation, from which the result holds for second-order Rational Krylov as a special case of the above. We have also shown that bases constructed via Balanced Truncation are C^x functions of the parameter θ for $x\{0, \omega\}$. Putting this all together, we have demon-

strated that GPS will work well in approximating any mapping from a parameter θ to the bases needed for model order reduction when the individual model order reduction methods used in building the mapping are smooth functions of θ. These guarantees on smoothness enable faster computations speeds that can allow for increased sample sizes and improved prediction accuracy.

References

P. Benner, M. Ohlberger, A. Cohen, and K. Willcox. *Model Reduction and Approximation*. Society for Industrial and Applied Mathematics, Philadelphia, PA, 2017.

G. Berkooz, P. Holmes, and J. L. Lumley. The proper orthogonal decomposition in the analysis of turbulent flows. *Annual Review of Fluid Mechanics*, 25(1):539–575, Jan. 1993.

K.-W. E. Chu. On multiple eigenvalues of matrices depending on several parameters. *SIAM Journal on Numerical Analysis*, 27(5):1368–1385, Oct. 1990.

N. J. Higham. *Functions of Matrices: Theory and Computation*. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 2008.

T. Kato. *Perturbation Theory for Linear Operators*. Springer, 2nd edition, 1980.

B. Peherstorfer and K. Willcox. Data-driven operator inference for nonintrusive projection-based model reduction. *Computer Methods in Applied Mechanics and Engineering*, 306: 196–215, July 2016.

P. J. Schmid. Dynamic mode decomposition and its variants. *Annual Review of Fluid Mechanics*, 54, Jan. 2022.

G. W. Stewart. Error and perturbation bounds for subspaces associated with certain eigenvalue problems. *SIAM Review*, 15(4):727–764, 1973.

J.-g. Sun. Multiple eigenvalue sensitivity analysis. *Linear Algebra and its Applications*, 137-138:183–211, Aug. 1990.

R. Zhang, S. Mak, and D. Dunson. Gaussian process subspace prediction for model reduction. *SIAM Journal on Scientific Computing*, 44(3):A1428–A1449, 2022.

Appendix A: Perturbation Theory Used for Proofs of Smoothness

Theorem 9 (Kato (1980), Thm 5.1, Sec 5.7)**.** *Let* $\mathbf{A}(\boldsymbol{\theta}) : \mathbb{C}^d \mapsto M_n(\mathbb{C})$ *be continuous at* $\boldsymbol{\theta} = 0$*. The unordered* n*-tuple of repeated eigenvalues* $\mathfrak{s} = (\lambda_i)_{i \in n}$ *of* $\mathbf{A}$ *is continuous at* $\boldsymbol{\theta} = 0$*. For any eigenvalue* λ *of* $\mathbf{A}(0)$ *with algebraic multiplicity* l*, let* $\Gamma \subset \mathbb{C}$ *be a closed curve enclosing* λ *but no other eigenvalues of* $\mathbf{A}(0)$*. For sufficiently small neighborhoods*

of the origin in $\mathbb{C}^d$*, the number of repeated eigenvalues of* $\mathbf{A}(\boldsymbol{\theta})$ *equals* l*; the* λ***-group*** *is the unordered* n*-tuple of these repeated eigenvalues.* ***Total eigenspace*** *associated with the* λ*-group is the sum of the eigenspaces associated with the eigenvalues in the* λ*-group, which is continuous at* $\boldsymbol{\theta} = 0$*.*

Theorem 10 (Kato (1980), Sec 5.8)**.** *The unordered* n*-tuple of repeated eigenvalues* $\mathfrak{s} = (\lambda_i)_{i \in n}$ *as a function of an order-*n *matrix* $\mathbf{A}$ *is partially differentiable at* $\mathbf{A} = \mathbf{A}_0$ *if and only if* $\mathbf{A}_0$ *is diagonable, and is holomorphic in a neighborhood of* $\mathbf{A}_0$ *if* $\mathbf{A}_0$ *has* n *distinct eigenvalues.*

Theorem 11 (Kato (1980), Sec 6.4)**.** *The unordered* n*-tuple of repeated eigenvalues* $\mathfrak{s}$ *as a function of an order-*n *symmetric matrix* $\mathbf{A}$ *is partially continuously differentiable.*

Theorem 12 (Chu (1990) Sec 4.1)**.** *Let* $\mathbf{A}(\boldsymbol{\theta}) \in C^\omega(B, M_n(\mathbb{C}))$ *be an analytic function where* B *is a neighborhood of the origin in* $\mathbb{C}^d$*. For any eigenvalue* λ *of* $\mathbf{A}(0)$*, which has algebraic multiplicity* l*, let* $\mathbf{X}_0, \mathbf{Y}_0 \in M^*_{n,l}(\mathbb{C})$ *be bases of the corresponding right- and left-invariant subspaces. There exist analytic functions* $\mathbf{X}(\boldsymbol{\theta}), \mathbf{Y}(\boldsymbol{\theta}) \in C^\omega(B_0, M^*_{n,l}(\mathbb{C}))$ *in a neighborhood* $B_0 \subset B$ *that are bases of the right and left total eigenspaces associated with the* λ*-group such that* $\mathbf{X}(0) = \mathbf{X}_0$ *and* $\mathbf{Y}(0) = \mathbf{Y}_0$*.*

Theorem 13 (Sun (1990) Thm 3.2)**.** *Let* $\mathbf{A}(\boldsymbol{\theta}) \in C^\omega(B, \mathcal{S}(n))$ *and* $\mathbf{B}(\boldsymbol{\theta}) \in C^\omega(B, \mathcal{S}_+(n))$ *be analytic functions, where* B *is a neighborhood of the origin in* $\mathbb{R}^d$*,* $\mathcal{S}(n)$ *is the set of symmetric order-*n *matrices, and* $\mathcal{S}_+(n)$ *is the set of symmetric positive definite order-*n *matrices. For any* l*-multiple eigenvalue* λ *of the generalized eigenvalue problem (GEVP)* $\mathbf{Ax} = \lambda\mathbf{Bx}$ *at* $\boldsymbol{\theta} = 0$*, let* $\mathbf{X}_0 \in M^*_{n,l}(\mathbb{R})$ *be a basis of the corresponding eigenspace such that* $\mathbf{X}_0^T\mathbf{A}(0)\mathbf{X}_0 = \lambda\mathbf{I}_l$ *and* $\mathbf{X}_0^T\mathbf{B}(0)\mathbf{X}_0 = \mathbf{I}_l$*. There exists an analytic function* $\mathbf{X}(\boldsymbol{\theta}) \in C^\omega(B_0, M^*_{n,l}(\mathbb{R}))$ *in a neighborhood* $B_0 \subset B$ *that is a basis of the total eigenspace associated with the* λ*-group such that* $\mathbf{X}(0) = \mathbf{X}_0$*.*