Proceedings of the
47th IEEE Conference on Decision and Control
Cancun, Mexico, Dec. 9-11, 2008

Application of Real Rational Modules in System Identification

Tzvetan Ivanov, P.-A. Absil, Brian D. O. Anderson and Michel Gevers

***Abstract*— This paper introduces a real rational module framework in the context of Prediction Error Identification using Box-Jenkins model structures. This module framework, which can easily be extended to other model structures, allows us to solve and/or extend a number of problems related to the computation of error norms that arise in system identification. Our main contribution to system identification is an extension of the asymptotic variance formulas for Box-Jenkins models derived by Ninness and Hjalmarsson to asymptotic auto-covariance with respect to frequency. This is achieved by viewing the sensitivity space of the prediction error as a so-called rational module. The auto-covariance of the transfer function estimates at different frequencies can then be quantified in terms of the poles and zeros of the underlying system and the input spectrum.**

I. INTRODUCTION

Prediction Error Identification (PEI) consists of identifying rational transfer function models by deriving estimators for the coefficients of the numerator and denominator polynomials, stored in a parameter vector $\theta \in \mathbb{R}^m$. Assuming a quadratic cost function for the prediction errors is both essential and interesting, because then the asymptotic covariance of the estimator $\hat{\theta}$ of the parameter vector can be quantified and exhibits a strong relation to the sensitivity of the prediction error w.r.t. the parameters.

Using time-domain techniques Ljung showed in [1] that $\mathrm{var}(\hat{\theta}_N) \approx N^{-1}K^{-1}$. More precisely as the observation length N increases, the error $\hat{\theta}_N - \theta_0$ not only converges to 0 (almost surely, a.s.) but also sufficiently fast for $\sqrt{N}(\hat{\theta}_N - \theta_0)$ to converge to $\gamma \in \mathcal{N}(0, K^{-1})$ (in law), where γ is a zero mean Gaussian limit with covariance K^{-1} s.t.

$$\frac{1}{N\sigma^2}\sum_{t=1}^{N} \frac{\partial \hat{y}(t)}{\partial \theta_i} \cdot \frac{\partial \hat{y}(t)}{\partial \theta_j} \xrightarrow{\text{a.s.}} K_{ij} \quad \text{as} \quad N \to \infty. \tag{1}$$

Here $\theta_0 \in \mathbb{R}^m$ denotes the fixed (or true) parameter vector, $\hat{\theta}_N$ the estimated parameter vector based on N observations $((y_1, u_1), (y_2, u_2), \ldots, (y_N, u_N)) \in (\mathbb{R}^2)^N$ of the process $y = G(\theta_0)u + H(\theta_0)e$, corrupted by white noise e with variance σ^2, and $\frac{\partial \hat{y}(t)}{\partial \theta} \in \mathbb{R}^m$ denotes the negative prediction error gradient at time t.

In the Box-Jenkins model structure the transfer functions $G(\theta)$ and $H(\theta)$ are separately parameterized. Thus, the limit vector can be split as $\gamma = [\gamma_G, \gamma_H]^{\mathrm{T}}$, and the prediction gradient splits similarly into independent parts corresponding to G and H, respectively. Therefore the estimation of G and H, as well as the properties of these estimates, can be treated separately and in order to introduce the main ideas of our contribution we focus on G only in this introduction. In the frequency domain the part of the prediction gradient that corresponds to G is given by $\rho_0(\mathrm{D}_{\theta_0}G)$, where ρ_0 denotes the spectral factor of the signal to noise ratio, $\rho_0\rho_0^* = \Phi_u/\Phi_{v_0}$, while $\mathrm{D}_{\theta_0}(\cdot)$ is the Fréchet derivative at θ_0.

Using Parseval's theorem, Mårtensson and Hjalmarsson [2], [3] expressed the asymptotic covariance matrix K^{-1} as a Gramian, or generalized inner product, in the space of rational transfer function matrices, leading to a nice geometric interpretation based on orthogonality and inner products in L^2. In this paper, we contribute an extension of this geometric interpretation by discussing its relation to uncorrelated random variables, i.e., orthogonality in the stochastic sense, and providing new insights into the interplay of both notions of orthogonality. Understanding this interplay allows us to choose an orthonormal basis (ONB) for the sensitivity space such that the coordinates of the resulting random transfer function $\rho_0(\mathrm{D}_{\theta_0}G)(\gamma_G)$ w.r.t. to that basis have unit variance and are uncorrelated (cf. proof of Theorem 9). This is of fundamental interest since the latter term approximates $\sqrt{N}(G(\hat{\theta}_N) - G(\theta_0))$ weighted with ρ_0. Combining these observations we are able to extend the results on asymptotic variance of transfer function estimates, derived by Ninness and Hjalmarsson in [4], to asymptotic auto-covariance w.r.t. frequency.

The key idea is a very simple concept which says that a random transfer function $\Gamma = \sum \gamma_i m_i$ formed by linear combinations of (deterministic) rational basis functions (m_i) which are orthonormal, with zero mean and unit covariance random coefficients (γ_i), has an auto-covariance w.r.t. frequency, i.e., $E[\Gamma(\omega_1)\Gamma^*(\omega_2)]$, that equals the integral kernel, say $k(\omega_1, \omega_2)$, reproducing the span of the m_i's (cf. Lemma 3). In order to express this integral kernel in terms of the poles and zeros of the underlying dynamical system we use techniques based on rational modules (cf. (23) in Theorem 6). Rational modules are related to polynomial modules; they are abstractly defined as quotient modules and have been studied extensively by Kalman and later by Fuhrmann in system theory; see e.g. [5][Chapter 10].

The paper is organized as follows: In Section II we define

This paper presents research results of the Belgian Network DYSCO (Dynamical Systems, Control, and Optimization), funded by the *Interuniversity Attraction Poles* Programme, initiated by the Belgian State, Science Policy Office. Futhermore this work has been supported in part by ARC Discovery Project Grant DP0664427 and National ICT Australia Ltd. National ICT Australia Ltd. is funded through the Australian Government's *Backing Australia's Ability* initiative, and in part through the Australian Research Council. The scientific responsibility rests with its authors.

Tzvetan Ivanov, P.-A. Absil and Michel Gevers are with the Center for Systems Engineering and Applied Mechanics (CESAME) Universite Catholique de Louvain, B-1348 Louvain-la-Neuve, Belgium. (http://www.inma.ucl.ac.be/~{ivanov,absil,gevers})

Brian D.O. Anderson is with Research School of Information Sciences and Engineering, The Australian National University, Canberra ACT 0200, Australia. The author is also with National ICT Australia, 7 London Circuit, Canberra ACT 0200, Australia. Email:Brian.Anderson@anu.edu.au

978-1-4244-3124-3/08/$25.00 ©2008 IEEE

Schur-stability, real rational subspaces of L^2, polynomial and rational modules as well as integral kernels reproducing these subspaces. In Section III we discuss rational modules as subspaces of the real rational Hardy space. The main results of this paper are provided in Section IV where we derive the asymptotic auto-covariance w.r.t. frequency for the Box-Jenkins model structure. Section V summarizes the key results with special emphasis on the potential of rational modules for system identification.

II. Preliminaries

A. Discrete Time Schur-Stability

We denote the ring consisting of all polynomials in the indeterminate x with coefficients in a field K by $K[x]$. For example $\sum_{i=0}^{n} k_i z^i$ ($k_i \in \mathbb{C}$) defines an element in $\mathbb{C}[z]$ while $\sum_{i=0}^{n} k_i z^{-i}$ ($k_i \in \mathbb{R}$) defines an element in $\mathbb{R}[z^{-1}]$. Regarding discrete-time stability (or Schur-stability), a polynomial $p = \sum p_i x^i \in \mathbb{C}[x]$ is said to be $[x]$-*stable* if

$$p = p_n(x - a_1)(x - a_2)\cdots(x - a_n) \quad (|a_i| < 1), \tag{2}$$

i.e., if all zeros of p are inside the open unit disk $\mathbb{D} \subseteq \mathbb{C}$. Similarly one says that $p \in \mathbb{C}[x]$ is $[x]$-*anti-stable* if there are no zeros of p inside the open unit disk, i.e., $a_i \in \mathbb{D}^{\mathrm{c}}$ for all $i = 1, \dots, n$. Stability and anti-stability for polynomials with real coefficients are defined by viewing $\mathbb{R}[x]$ as a subset of $\mathbb{C}[x]$.[1] In the corresponding field of quotients

$$K(x) = \{p/q \,|\, p, q \in K[x], q \neq 0\},^{2} \tag{3}$$

a rational function $f \in \mathbb{C}(z)$, or $f \in \mathbb{R}(z) \subseteq \mathbb{C}(z)$, is said to be stable (anti-stable) if $q \in \mathbb{C}[z]$ is $[z]$-stable ($[z]$-anti-stable) in the coprime-factorization $f = p/q$ with $p \in \mathbb{C}[z]$.

B. Rational subspaces of L^2

We introduce four rational subspaces of the (complex) Hilbert space $L^2 = L^2(\mathbb{T}, \mathbb{C})$ of square integrable functions on the unit circle. The first is $\mathrm{R}L^2 \subseteq L^2$ consisting of real rational functions with no pole on the unit circle, endowed with the standard 2-norm

$$\|f\|_2^2 = \int |f(\omega)|^2 \,\frac{\mathrm{d}\,\omega}{2\pi}, \tag{4}$$

where we integrate over the unit circle $\mathbb{T} \simeq [0, 2\pi)$ w.r.t. the standard Lebesgue measure. Recall that by choosing the ONB $\{z^k\}_{k\in\mathbb{Z}}$ in L^2 one may identify L^2 with $l^2 = l^2(\mathbb{Z}, \mathbb{C})$ consisting of all sequences $(a_i) \in \mathbb{C}^{\mathbb{Z}}$ with $\sum |a_i|^2 < \infty$ using the Fourier transform

$$\mathscr{F} : l^2 \to L^2, (a_i) \mapsto \sum a_i z^{-i}. \tag{5}$$

The subspace of $\mathrm{R}L^2$ corresponding to strictly causal sequences (having support in $\mathbb{Z}_{\geq 1}$) will be denoted by $\mathrm{R}H^2$ and its orthogonal complement, corresponding to anti-causal sequences (having support in $\mathbb{Z}_{\leq 0}$), by $\mathrm{R}H^2_-$. The subspace of $\mathrm{R}L^2$ consisting of, not necessarily strictly proper, real rational functions with $[z]$-stable denominator endowed with the sup-norm

$$\|f\|_\infty = \sup_{\omega\in\mathbb{T}} |f(\omega)|, \tag{6}$$

will be denoted by $\mathrm{R}H^\infty$. Of all spaces introduced so far there exists complex rational versions which are obtained by replacing $\mathbb{R}[z]$ in the description of Table I by $\mathbb{C}[z]$ and $\mathbb{R}[z^{-1}]$ by $\mathbb{C}[z^{-1}]$. We choose to distinguish them from the real rational versions by a preceding ${}^{\mathrm{c}}(\cdot)$ in the notation, e.g., the complexification of $\mathrm{R}H^2$ is given by ${}^{\mathrm{c}}\mathrm{R}H^2$.

	p/q	q	r/s	s
$\mathrm{R}L^2$	-	Z	-	ⓩ
$\mathrm{R}H^2$	SP	S	O	ⓐ
$\mathrm{R}H^2_-$	-	A	P	ⓢ
$\mathrm{R}H^\infty$	-	S	-	ⓐ

TABLE I. Both $p, q \in \mathbb{R}[z]$ and $r, s \in \mathbb{R}[z^{-1}]$ are assumed to be coprime with $q, s \neq 0$. The following properties are used: SP:: strictly proper , P:: proper, Z:: no $[z]$-zero on $\mathbb{T}$, S:: $[z]$-stable, A:: property Z and $[z]$-anti-stable, ⓩ:: no $[z^{-1}]$-zero on $\mathbb{T}$, O:: $z^{-1} = 0$ is a zero of r, ⓢ:: $[z^{-1}]$-stable, ⓐ:: property ⓩ and $[z^{-1}]$-anti-stable.

C. Polynomial and Rational Modules $(\mathrm{X}_q, \mathrm{X}^q)$

Given a non-zero polynomial $q \in K[x]$ we define its *polynomial module*, denoted by X_q, via

$$\mathrm{X}_q = \{p \,|\, p \in K[x], \deg(p) < \deg(q)\} \subseteq K[x], \tag{7}$$

and its *rational module*, denote by X^q, via

$$\mathrm{X}^q = \{p/q \,|\, p \in \mathrm{X}_q\} \subseteq K(x). \tag{8}$$

Thus X^q consists of all strictly proper rational functions with q as a common denominator. Clearly both spaces are isomorphic, i.e., $\mathrm{X}_q \simeq \mathrm{X}^q$, finitely generated, and have dimension $\deg(q)$ as linear spaces over K. This notation however has a slight ambiguity in our context. Given a $q \in \mathbb{R}[x]\backslash\{0\}$ (where $x = z$ or $x = z^{-1}$) one may associate a real ($K = \mathbb{R}$) or a complex ($K = \mathbb{C}$) polynomial module to it. As before we will distinguish both cases with the ${}^{\mathrm{c}}$-prefix notation, i.e., ${}^{\mathrm{c}}\mathrm{X}_q \subseteq \mathbb{C}[x]$ and $\mathrm{X}_q \subseteq \mathbb{R}[x]$. We do the same for rational modules, i.e., ${}^{\mathrm{c}}\mathrm{X}^q \subseteq \mathbb{C}(z)$ denotes a complex rational module whereas $\mathrm{X}^q \subseteq \mathbb{R}(z)$ denotes a real rational module.

D. Integral Kernels Reproducing Subspaces

Abstractly speaking, given a set U and a complex, complete, inner-product space $(\mathcal{F}, \langle\cdot,\cdot\rangle)$ of functions $f \in \mathcal{F}$ having U as their domain and taking values in $\mathbb{C}$ then for each point $v \in U$ one may consider the *evaluation at* v, denoted by $\mathtt{ev}_v$, as the map

$$\mathtt{ev}_v : \mathcal{F} \to \mathbb{C}, f \mapsto f(v). \tag{9}$$

By the Riesz-representation Theorem, see e.g. [6], $\mathtt{ev}_v$ is continuous if and only if it can be expressed as $\mathtt{ev}_v(f) = \langle f, k_v\rangle$ for some (uniquely determined) $k_v \in \mathcal{F}$. If $\mathtt{ev}_v$ is continuous for all $v \in U$ this gives rise to a function

$$k : U \times U \to \mathbb{C}, (u, v) \mapsto k_v(u). \tag{10}$$

[1] For example the polynomial $(z^{-1} - \mathrm{i}/2)(z^{-1} + \mathrm{i}/2)$ is $[z^{-1}]$-stable in $\mathbb{R}[z^{-1}]$, $(z - \mathrm{i}/2)$ is $[z]$-stable in $\mathbb{C}[z]$, $(z + 3) \in \mathbb{R}[z]$ is $[z]$-anti-stable, and $(z - 1/2)(z - 2) \in \mathbb{R}[z]$ is neither $[z]$-stable nor $[z]$-anti-stable.

[2] Note that the set of rational functions in $\mathbb{C}(z)$, $\mathbb{R}(z)$ coincides with the set of rational functions in $\mathbb{C}(z^{-1})$, $\mathbb{R}(z^{-1})$, respectively, and thus we choose not to distinguish between them.

If we think of taking the inner-product as integration w.r.t. some measure μ on U, i.e., $\langle f, g\rangle = \int f\bar{g}\,\mathrm{d}\mu$, then k admits an interpretation as an integral kernel acting on elements in $\mathcal{F}$ via $T_k : \mathcal{F} \to \mathcal{F}$ defined by

$$(T_k f)(v) = \int f(u)k^*(u, v)\,\mathrm{d}\mu(u). \tag{11}$$

In our abstract setting this becomes $(T_k f)(v) = \langle f, k(\cdot, v)\rangle$. Due to our definition of k we have two key properties:

1) $k(\cdot, v) \in \mathcal{F}$ for all $v \in U$.
2) $T_k f = f$ for all $f \in \mathcal{F}$.

If 1) and 2) are satisfied one says that k has the *reproducing property* w.r.t. $\mathcal{F}$. It is easy to check that k is uniquely determined by these properties and one may thus call it *the* reproducing kernel of $\mathcal{F}$.[3] Without proof we recall the following standard result [7].

Theorem 1: Let k denote the kernel reproducing $\mathcal{F}$. Then

$$k(u, v) = \sum_{i\in I} b_i(u)b_i^*(v), \tag{12}$$

for any ONB $\{b_i\}_{i\in I}$ of $\mathcal{F}$. ■

Remark 2: The reader may wonder why we limit our discussion of reproducing kernels to complex inner product spaces even though we have introduced various $\mathbb{R}$-subspaces $\mathcal{F}$ of L^2 like $\mathcal{F} = \mathrm{R}L^2$, or $\mathcal{F} = \mathrm{R}H^2$, or $\mathcal{F} = \mathrm{X}^q$ for some $q \in \mathbb{R}[z]$ which is $[z]$-stable. The reason is that in any case the evaluation at $\omega \in \mathbb{T}$ is not a real valued function, i.e., we have $\mathtt{ev}_\omega : \mathcal{F} \to \mathbb{C}$, and in particular $\mathtt{ev}_\omega$ is not a linear functional in the sense of the $\mathbb{R}$-linear structure on $\mathcal{F}$.[4] ■

We conclude this section on integral kernels reproducing inner product spaces, e.g., subspaces of L^2, by discussing their relation to random transfer functions.

Lemma 3: Let $M \subseteq \mathrm{R}H^2$ denote a finitely generated subspace and k the reproducing kernel of its complexification ${}^{\mathrm{c}}M \subseteq {}^{\mathrm{c}}\mathrm{R}H^2$. Let $\{m_i\}_{i=1}^n$ denote an orthonormal basis for M and consider the following random variable Γ on M:

$$\Gamma = \gamma_1 m_1 + \gamma_2 m_2 + \cdots + \gamma_n m_n, \tag{13}$$

where $\gamma \in \mathbb{R}^n$ is zero mean and has a unit covariance that is $E[\gamma_i\gamma_j] = \delta_{ij}$. Then there holds

$$E[\Gamma(\omega_1)\Gamma^*(\omega_2)] = k(\omega_1, \omega_2), \tag{14}$$

for all $\omega_1, \omega_2 \in \mathbb{T}$.

Proof: Let $k_{\omega_i} = k(\cdot, \omega_i)$ $(i = 1, 2)$. Due the reproducing property of k we may deduce that

$$\begin{aligned} E[\Gamma(\omega_1)\Gamma^*(\omega_2)] &= E\left[\left\langle \sum \gamma_i m_i, k_{\omega_1}\right\rangle \left\langle \sum \gamma_j m_j, k_{\omega_2}\right\rangle^*\right] \\ &= E\left[\sum\sum \gamma_i \langle m_i, k_{\omega_1}\rangle \gamma_j^* \langle m_j, k_{\omega_2}\rangle^*\right] \\ &= \sum\sum m_i(\omega_1)m_j^*(\omega_2)E[\gamma_i\gamma_j] \end{aligned}$$

which, together with $E[\gamma_i\gamma_j] = \delta_{ij}$, and the result in Theorem 1, concludes the proof. □

III. Rational modules embedded in $\mathrm{R}H^2$

Given a non-zero $[z]$-stable polynomial $q \in \mathbb{C}[z]$ we may regard ${}^{\mathrm{c}}\mathrm{X}^q$ as a subspace of ${}^{\mathrm{c}}\mathrm{R}H^2$. Motivated by our aim to characterize the geometry of the rational module in an algebraic way, i.e., in terms of the zeros of q, we define the $\star$-*operation*. This operation allows us to interpret adjoint operators, formed by complex conjugation arising from the inner-product in ${}^{\mathrm{c}}\mathrm{R}L^2$, in terms of the poles and zeros of the corresponding rational function. For non-zero polynomial $q = q_n\prod(z - a_i) \in \mathbb{C}[z]$ it is given by

$$q^\star = \bar{q}_n(1 - z\bar{a}_1)\cdot(1 - z\bar{a}_2)\cdots(1 - z\bar{a}_n). \tag{15}$$

Note if q is monic $q^\star$ is general not monic. It satisfies $\deg(q^\star) \le \deg(q)$ but not necessarily with equality, e.g., if $q = z^n$ then $q^\star = 1$.[5] The connection to the inner-product (geometry) is due to the fact that for all $z \in \mathbb{T}$ we have

$$f^*(z) = \overline{f(1/\bar{z})} = \frac{\sum \bar{p}_i z^{-i}}{\sum \bar{q}_j z^{-j}} = z^{\deg(q)-\deg(p)}\cdot\frac{p^\star}{q^\star}, \tag{16}$$

for any $f = p/q$ with $p, q \in \mathbb{C}[z]$.[6] This is a standard construction sometimes referred to as *para-adjoint* of f.

A. The Orthogonal Complement of X^q and ${}^{\mathrm{c}}\mathrm{X}^q$

In the following we discuss the real-rational case but assure the reader that Theorem 4 and Corollary 5 and the proofs remain true under *complexification*, i.e., replacing $\mathbb{R}[z]$ by $\mathbb{C}[z]$, $\mathrm{R}H^2$ by ${}^{\mathrm{c}}\mathrm{R}H^2$ and X^q by ${}^{\mathrm{c}}\mathrm{X}^q$. Theorem 4 is related to a famous theorem of Beurling [8] characterizing invariant subspaces of H^2. In a follow up paper we will provide an elementary (algebraic) proof for the rational case which does require functional analysis.

Theorem 4: Let $q \in \mathbb{R}[z]$ be non-zero and $[z]$-stable. There holds

$$\mathrm{R}H^2 = \frac{q^\star}{q}\,\mathrm{R}H^2 \oplus \mathrm{X}^q, \tag{17}$$

and both summands are orthogonal.[7] Moreover $m^*m = 1$ with $m = q^\star/q$. ■

Corollary 5: Let $q_i \in \mathbb{R}[z]$ be non-zero and $[z]$-stable $(i = 1, 2)$. Then we have the following decomposition

$$\mathrm{X}^{q_1q_2} = \mathrm{X}^{q_1} \oplus m_1\,\mathrm{X}^{q_2}, \tag{18}$$

with $m_1 = q_1^\star/q_1$. Moreover both summands are orthogonal w.r.t. the inner product on $\mathrm{X}^{q_1q_2}$ induced by $\mathrm{R}H^2$.

Proof: It is clear that $\mathrm{X}^{q_1} + m_1\mathrm{X}^{q_2} \subseteq \mathrm{X}^{q_1q_2}$ since $\deg(q_1^\star) \le \deg(q_1)$. Due to Theorem 4 it follows that X^{q_1} is orthogonal to $m_1\mathrm{X}^{q_2}$ and in particular $\mathrm{X}^{q_1} \cap m_1\mathrm{X}^{q_2} = \{0\}$. It remains to check that $\mathrm{X}^{q_1} + \mathrm{X}^{q_2} \supseteq \mathrm{X}^{q_1q_2}$. Note that $m_1\mathrm{X}^{q_2} \simeq X^{q_2}$ since m_1 is unitary and in particular

$$\dim(\mathrm{X}^{q_1} + m_1\mathrm{X}^{q_2}) = \deg(q_1) + \deg(q_2), \tag{19}$$

which proves the claim due to $\dim(\mathrm{X}^{q_1q_2}) = \deg(q_1q_2)$. This concludes the proof. □

[3] We find it convenient to denote the complex conjugate of $z \in \mathbb{C}$ by $\bar{z}$ and the complex conjugate of a complex valued function $f \in \mathcal{F}$ by f^*, i.e., $f^*(z) := \overline{f(z)}$.

[4] In other words no integral kernel can have the reproducing property w.r.t. a real inner-product on $\mathcal{F}$ and one has to consider its complexification ${}^{\mathrm{c}}\mathcal{F}$, i.e., ${}^{\mathrm{c}}\mathrm{R}L^2$, ${}^{\mathrm{c}}\mathrm{R}H^2$ or ${}^{\mathrm{c}}\mathrm{X}^q$, respectively.

[5] Please do not confuse $p^\star = \bar{p}_n\prod(1 - \bar{a}_i z)$ with $p^* = \bar{p}_n\prod(\bar{z} - \bar{a}_i)$.

[6] Note that (16) holds because the two maps $z \mapsto \bar{z}$ and $z \mapsto z^{-1}$ coincide on the unit circle $\mathbb{T}$.

[7] We use the symbol $\oplus$ to denote a general, not necessarily orthogonal, direct sum of two subspaces.

B. Canonical ONB and the Reproducing Kernel of ${}^{\mathrm{c}}\mathrm{X}^q$

One should emphasize the importance of Corollary 5. Since m_1 is unitary, i.e., $m_1^* m_1 = 1$, it preserves angles and in particular $\mathrm{X}^{q_1 q_2} \simeq \mathrm{X}^{q_1} \dot{\oplus} \mathrm{X}^{q_2}$ holds canonically.[8] Thus in order to construct an ONB for X^q it suffices to be able to do this for rational modules of irreducible polynomials. In the case of $K[x] = \mathbb{R}[z]$ these are the polynomials of degree less than or equal two. For $K[x] = \mathbb{C}[z]$ these are the monomials. In particular if $q \in \mathbb{C}[z]$ is given by $q = \prod(z - a_i)$ for some finite family $(a_i | i \in I) \subseteq \mathbb{D}$, there holds

$$ {}^{\mathrm{c}}\mathrm{X}^q \simeq \dot{\bigoplus_{i \in I}} {}^{\mathrm{c}}\mathrm{X}^{(z-a_i)}. \tag{20}$$

This greatly simplifies the construction of ONBs and integral kernels reproducing ${}^{\mathrm{c}}\mathrm{X}^q$.

Theorem 6: Let $q \in \mathbb{C}[z]$ be $[z]$-stable given by $q = \prod_{i=1}^n (z - a_i)$. The following statements are true:

1) There holds
$$\left\| \frac{1}{z - a_i} \right\|_2^2 = \frac{1}{1 - |a_i|^2}. \tag{21}$$
2) An orthonormal basis $\{b_i\}$ of ${}^{\mathrm{c}}\mathrm{X}^q$ is given by
$$b_i = \frac{q_i^\star}{q_i} \frac{\sqrt{1 - |a_i|^2}}{z - a_i} \quad i = 1, \ldots, n, \tag{22}$$
with $q_1 = 1$ and $q_i = \prod_{j=1}^{i-1} (z - a_j)$ for all $i > 1$.
3) There holds
$$k(z, z) = \sum_{i=1}^n \frac{1 - |a_i|^2}{|z - a_i|^2}, \tag{23}$$
where $k : \mathbb{T}^2 \to \mathbb{C}$ is the reproducing kernel of ${}^{\mathrm{c}}\mathrm{X}^q$.

Proof: Let $g_i = 1/(z - a_i)$ for all $i = 1, \ldots, n$. Then 1) follows from computation of $\langle g_i, g_i \rangle$, e.g., by using the Residue Theorem. 2) It is clear that $g_i / \|g_i\|_2$ is an ONB for ${}^{\mathrm{c}}\mathrm{X}^{(z-a_i)}$. Using the canonical identification (20) the result follows from 1). Finally 3) follows from Theorem 1 and the fact that $q_i^\star / q_i$ is unitary. This concludes the proof. □

Example 1: Let $q = z^2 + 1/9 = (z - \mathrm{i}/3)(z + \mathrm{i}/3)$. Let $a_1 = \mathrm{i}/3$ and $a_2 = -\mathrm{i}/3$. We can compute an ONB, say (b_1, b_2), of ${}^{\mathrm{c}}\mathrm{X}^q$ using statement 2) in Theorem 6. Note that $q_2 = \prod_{j=1}^{2-1} (z - a_j) = (z - a_1) = z - \mathrm{i}/3$ and thus, due to the definition of the $\star$-operation in (15), there holds $q_2^\star = 1 + z\,\mathrm{i}/3$. The ONB is thus given by

$$(b_1, b_2) = \left(\frac{2\sqrt{2}}{3z - \mathrm{i}}, \frac{3 + \mathrm{i}z}{-\mathrm{i} + 3z} \cdot \frac{2\sqrt{2}}{3z + \mathrm{i}} \right) \tag{24}$$

Note that our ONB depends on the way we choose to order the poles.[9] We obtain the diagonal, say $k(z, z)$, of the kernel reproducing ${}^{\mathrm{c}}\mathrm{X}^q$ using statement 3) in Theorem 6

$$k(z, z) = \frac{2\sqrt{2}}{|3z - \mathrm{i}|^2} + \frac{2\sqrt{2}}{|3z + \mathrm{i}|^2}. \tag{25}$$

One should note that (24) and (25) can be re-used in the computation of the ONB and integral kernel reproducing any rational module $\mathrm{X}^{\tilde{q}}$ that contains X^q. Such a larger module might be, for example, given by $\tilde{q} = q \cdot (z - 1/2)$.

IV. Asymptotic Auto-Covariance of Transfer Function Estimates in PEI

A. Model Class and Parametrization

The dynamical systems that we consider can be represented by a set of transfer functions $\mathcal{D} = \mathcal{G} \times \mathcal{H} \subseteq \mathrm{R}H^\infty \times \mathrm{R}H^\infty$ of the *Box-Jenkins* model structure given by a parametrization

$$\Pi : \Theta \to \mathcal{D} \quad \text{with} \quad \Pi = [G, H] = [B/A, C/D], \tag{26}$$

mapping θ to $[G_\theta, H_\theta] = [B_\theta / A_\theta, C_\theta / D_\theta]$ according to

$$\Pi : \begin{bmatrix} \theta_A \\ \theta_B \\ \theta_C \\ \theta_D \end{bmatrix} \mapsto \begin{bmatrix} \dfrac{\theta_{B,1} z^{-1} + \cdots + \theta_{B,b} z^{-b}}{1 + \theta_{A,1} z^{-1} + \cdots + \theta_{A,a} z^{-a}} \\[2ex] \dfrac{1 + \theta_{C,1} z^{-1} + \cdots + \theta_{C,c} z^{-c}}{1 + \theta_{D,1} z^{-1} + \cdots + \theta_{D,d} z^{-d}} \end{bmatrix}^{\mathrm{T}}, \tag{27}$$

where $\Theta \subseteq \mathbb{R}^a \times \mathbb{R}^b \times \mathbb{R}^c \times \mathbb{R}^d = \mathbb{R}^m$.[10] Moreover Θ can be chosen such that 1) Π is bijective (no pole-zero cancellations) 2) C_θ is $[z^{-1}]$-anti-stable for all $\theta \in \Theta$, i.e., H_θ is minimum-phase and 3) Θ is bounded. Note that $\mathcal{D} \subseteq \mathrm{R}H^\infty \times \mathrm{R}H^\infty$, together with Table I, implies that we have $A_\theta, D_\theta \in \mathbb{R}[z^{-1}]$ are $[z^{-1}]$-anti-stable for all $\theta \in \Theta$.

B. Probabilistic Setup

Let $\mathbb{N} = \mathbb{Z}_{\geq 1}$ denote our (one-sided) time axis and $\mathcal{A}$ denote the standard product σ-field on $\mathbb{R}^{\mathbb{N}}$. We assume that an observation $y \in \mathbb{R}^{\mathbb{N}}$ can be explained by

$$y = Gu + v \quad \text{with} \quad u = \mathrm{S}_u w \quad \text{and} \quad v = \sigma H e. \tag{28}$$

where the driving noise e is distributed according to some probability measure P_{e} on $\mathcal{A}$ making it an i.i.d. process of unit variance and bounded 8-th order moments. Moreover one assumes the following: (1) zero-initial conditions; (2) u is a known and fixed sequence that has been obtained as a filtered (by S_u) version of *one particular realization* w of a zero-mean white noise process with unit variance; (3) $\mathrm{S}_u \in \mathrm{R}H^\infty$, and (4) $\sigma \in \mathbb{R}_{>0}$. Thus the input u is assumed to be deterministic, having been obtained as one particular realization of a process with spectrum $\Phi_u = \mathrm{S}_u \mathrm{S}_u^*$; we denote by $\mathcal{U} \subseteq \mathbb{R}^{\mathbb{N}}$ the set of inputs defined this way. Analogously the spectrum of v is given by $\Phi_v = \sigma^2 H H^*$. One observes $(y, u) \in (\mathbb{R}^2)^{\mathbb{N}}$. The probability measure P_{e} endows the observation space $((\mathbb{R}^2)^{\mathbb{N}}, \mathcal{A} \otimes \mathcal{A})$, for each $\mu = (\theta, u) \in \Theta \times \mathcal{U}$, with a probability measure P_μ, expectation $E_\nu[\cdot]$, variance $\mathrm{var}_\nu[\cdot]$, determined by $P_\mu(F)$ for all $F \in \mathcal{A}$[11]

$$P_\mu((y, u) \in F \times \mathbb{R}^{\mathbb{N}}) = P_{\mathrm{e}}(G_\theta u + H_\theta e \in F). \tag{29}$$

[8] By $\mathrm{X}^{q_1} \dot{\oplus} \mathrm{X}^{q_2}$ $\dot{\oplus}$ we mean $\mathrm{X}^{q_1} \times \mathrm{X}^{q_2}$ endowed with the usual linear space and inner product structure. The canonical identification of $\mathrm{X}^{q_1} \dot{\oplus} \mathrm{X}^{q_2}$ to $\mathrm{X}^{q_1} \oplus m_1 \mathrm{X}^{q_2}$ is given by $(f_1, f_2) \mapsto f_1 + m_1 f_2$.

[9] If we had chosen $a_1 = -\mathrm{i}/3$ and $a_2 = \mathrm{i}/3$ we would have obtained a different basis.

[10] That is to say that, e.g., the coefficients of the polynomial $B \in \mathbb{R}[z^{-1}]$ of degree b are stored in the column vector $\theta_B = [\theta_{B,1}, \cdots, \theta_{B,b}]^{\mathrm{T}}$ which is possible since $\theta_{B,0} = 0$ by definition. Also note that, by construction, $|H(z)| \to 1$ as $|z| \to \infty$, i.e., H is monic.

[11] $\mathcal{A} \otimes \mathcal{A}$ denotes the standard product σ-field. We assume identifiability in the sense that the map given by $\Theta \to \{\text{measures on } (\mathbb{R}^2)^{\mathbb{N}}\}, \theta \mapsto P_{(\theta, u)}$ is injective for all $u \in \mathcal{U}$.

C. Prediction Error Estimates

The prediction error estimator based on the first $((y_1, u_1), \ldots, (y_N, u_N)) \in (\mathbb{R}^2)^N$ observations will be denoted by $\hat{\theta}_N : (\mathbb{R}^2)^N \to \Theta$ with

$$\hat{\theta}_N : (y, u) \mapsto \arg\min_{\theta \in \Theta} \frac{1}{N} \sum_{t=1}^{N} |\hat{y}_\theta(t) - y(t)|^2, \tag{30}$$

where the map $\hat{y} : (\mathbb{R}^2)^\mathbb{N} \times \Theta \to \mathbb{R}^\mathbb{N}$ denotes the *one-step-ahead predictor* given by[12]

$$\hat{y} : ((y, u), \theta) \mapsto H_\theta^{-1} G_\theta u + (1 - H_\theta^{-1}) y. \tag{31}$$

It is natural to consider the sensitivity of $\hat{y}((y, u), \theta)$ w.r.t. small perturbations around a fixed $\theta_0 \in \Theta$ as an a priori measure of how good the estimate $\hat{\theta}_N$ will resemble θ_0 if one observes (y, u) with $y = G_0 u + \sigma H_0 e$ explaining the output.[13] The higher the sensitivity the better the estimate. For technical convenience we will characterize the sensitivity in terms of transfer functions. The proof is standard and therefore skipped.

Theorem 7: Let $\mu_0 = (\theta_0, u)$ with $\theta_0 \in \Theta$, $u \in \mathcal{U}$. Moreover let $\Pi_0 = \Pi(\theta_0)$ and $\rho_0 = \mathrm{S}_u/(\sigma H_0)$. Note that the signal to noise ratio Φ_u/Φ_{v_0} is given by $\rho_0 \rho_0^*$. Then $\hat{y}_\theta = \hat{y}((G_0 u + \sigma H_0 e, u), \theta)$, with $\hat{y}$ denoting the function defined in (31), can be explained by

$$\hat{y}_\theta = \sigma(R_\theta w + S_\theta e) \quad \text{for all} \quad \theta \in \Theta, \tag{32}$$

with $R, S : \Theta \to \mathrm{R}H^\infty$ well defined by $R = \sigma^{-1} S_u (G_0 - H^{-1}(G_0 - G))$ and $S = H_0 - H^{-1} H_0$ since H_θ has a stable inverse due to assumption 2) stated after (27). There holds[14]

$$\mathrm{D}_{\theta_0} R = \rho_0 \cdot \mathrm{D}_{\theta_0} G \tag{33a}$$

$$\mathrm{D}_{\theta_0} S = H_0^{-1} \cdot \mathrm{D}_{\theta_0} H. \tag{33b}$$

D. Unit Covariance in the Sensitivity Space

The following Theorem 8 has been proven by Caines and Ljung in [1]; see also [9][Chapter 9].

Theorem 8: Under weak regularity assumptions the estimator $\hat{\theta}_N$ converges in law according to

$$\sqrt{N}(\hat{\theta}_N - \theta_0) \xrightarrow{\mathcal{L}} \gamma = [\gamma_G, \gamma_H] \in \mathcal{N}(0, K^{-1}), \tag{34}$$

as $N \to \infty$, with the limit γ being zero mean, Gaussian and block-diagonal covariance K^{-1}. Further[15]

$$e_i^{\mathrm{T}} (E_{\mu_0}[\gamma_G \gamma_G^{\mathrm{T}}])^{-1} e_j = \langle (\mathrm{D}_{\theta_0} R)(e_i), (\mathrm{D}_{\theta_0} R)(e_j) \rangle, \tag{35a}$$

$$e_k^{\mathrm{T}} (E_{\mu_0}[\gamma_H \gamma_H^{\mathrm{T}}])^{-1} e_l = \langle (\mathrm{D}_{\theta_0} S)(e_k), (\mathrm{D}_{\theta_0} S)(e_l) \rangle, \tag{35b}$$

for all $i, j \in \{1, \ldots, a+b\}$ and $k, l \in a+b+\{1, \ldots, c+d\}$, where $\langle \cdot, \cdot \rangle$ denotes the inner product on $\mathrm{R}H^2$, and $\{e_i\}$ the standard basis of $\mathbb{R}^m$. ■

We seek to characterize $\mathrm{var}_{\mu_0}[\sqrt{N} G(\hat{\theta}_N)]$ for large N in a way that allows us to exploit the preceding Theorem 8. We do this with a first order Taylor expansion of G around θ_0 and the following approximation in law

$$\begin{aligned} \sqrt{N}(G(\hat{\theta}_N) - G_0) \ &= \ (\mathrm{D}_{\theta_0} G)(\sqrt{N}(\hat{\theta}_N - \theta_0)) \ + \\ + \, \mathrm{O}(\|\sqrt{N}(\hat{\theta}_N - \theta_0)\|^2/\sqrt{N}) \ &\xrightarrow{\mathcal{L}} \ (\mathrm{D}_{\theta_0} G)(\gamma_G), \end{aligned} \tag{36}$$

as $N \to \infty$, and similarly

$$\sqrt{N}(H(\hat{\theta}_N) - H_0) \ \xrightarrow{\mathcal{L}} \ (\mathrm{D}_{\theta_0} H)(\gamma_H). \tag{37}$$

We are now ready to state the main result in Theorem 9 which exploits Theorem 8 by utilizing (36), (37) and Lemma 3.[16]

Theorem 9: Let $G_\gamma = (\mathrm{D}_{\theta_0} G)(\gamma_G)$ and $H_\gamma = (\mathrm{D}_{\theta_0} H)(\gamma_H)$. For all $\omega_i \in \mathbb{T}$ $(i = 1, 2)$ there holds

$$E_{\mu_0}[G_\gamma(\omega_1) \cdot G_\gamma^*(\omega_2)] = \frac{k_{0,R}(\omega_1, \omega_2)}{\rho_0(\omega_1) \rho_0^*(\omega_2)}, \tag{38a}$$

$$E_{\mu_0}[H_\gamma(\omega_1) \cdot H_\gamma^*(\omega_2)] = \frac{k_{0,S}(\omega_1, \omega_2)}{H_0^{-1}(\omega_1) H_0^{-1,*}(\omega_2)}, \tag{38b}$$

and thus for all $\omega \in \mathbb{T}$, as $N \to \infty$, there holds

$$N \cdot \mathrm{var}_{\mu_0}[G(\hat{\theta}_N)(\omega)] \to \frac{\Phi_{v_0}(\omega)}{\Phi_u(\omega)} \cdot k_{0,R}(\omega, \omega), \tag{39a}$$

$$N \cdot \mathrm{var}_{\mu_0}[H(\hat{\theta}_N)(\omega)] \to |H_0(\omega)|^2 \cdot k_{0,S}(\omega, \omega), \tag{39b}$$

where $k_{0,R}$ and $k_{0,S}$ denote the reproducing kernels of the complexifications respectively of $\mathrm{Im}(\mathrm{D}_{\theta_0} R) \subseteq \mathrm{R}H^2$ and $\mathrm{Im}(\mathrm{D}_{\theta_0} S) \subseteq \mathrm{R}H^2$.

Proof: Let $\beta^{-1} \in \mathbb{R}^{(a+b)\times(a+b)}$ denote the positive definite matrix defined by

$$[\beta^{-1}]_{ij} = E_{\mu_0}[\gamma_{G,i} \cdot \gamma_{G,j}]. \tag{40}$$

Then $\tilde{\gamma}_G = \beta^{1/2} \gamma_G$ has unit covariance. Denoting the standard basis of $\mathbb{R}^{a+b}$ by $\{e_i\}$ we define $r_i \in \mathrm{R}H^2$ via $r_i = (\mathrm{D}_{\theta_0} R)(e_i)$ for all $1 \le i \le a+b$ and let $r = [r_1, \ldots, r_{a+b}]^{\mathrm{T}}$. Then $\{r_i\}$ is a basis and $\{\tilde{r}_i\}$ with $\tilde{r} = \beta^{-1/2} r$ is an ONB of $\mathrm{Im}(\mathrm{D}_{\theta_0} R)$. Thus $G_\gamma = \rho_0^{-1} (\mathrm{D}_{\theta_0} R)(\gamma_G) = \rho_0^{-1} \cdot \gamma_G^{\mathrm{T}} r$, due to (33). The results quantifying the asymptotic auto-covarinace of $G(\hat{\theta}_N)$ now follow from Lemma 3 and

$$\gamma_G^{\mathrm{T}} r = \gamma_G (\beta^{+1/2} \beta^{-1/2}) r = \sum \tilde{\gamma}_{G,i} \tilde{r}_i, \tag{41}$$

since we already proved that $\tilde{\gamma}_{G,i}$ are uncorrelated, unit variance and $(\tilde{r}_i)$ is an ONB. The result for the auto-covariance of $H(\hat{\theta}_N)$ is checked analogously. □

[12] Note that (30) is well defined because the $\hat{y}$ in (31) is a causal function of the observed sequence (y, u).

[13] We use the notation G_0 for $G_{\theta_0} = G(\theta_0)$ and H_0 for $H_{\theta_0} = H(\theta_0)$ to prevent sub sub indices.

[14] Here $\mathrm{D}_{\theta_0}(\cdot)$ denotes the Fréchet-derivative at θ_0 w.r.t. the Banach space H^∞.

[15] The result is usually stated in the time domain $\mathrm{cov}_{\mu_0}(\gamma_G) = \beta^{-1}$ with

$$\beta_{ij} = \lim_{N\to\infty} \frac{1}{N\sigma^2} \sum_{t=1}^{N} \frac{\partial \hat{y}(t)}{\partial \theta_i} \cdot \frac{\partial \hat{y}(t)}{\partial \theta_j} =: \bar{E}[r_i(t) r_j(t)] \quad P_{\mu_0}\text{-a.s.},$$

with $r_i(t) = (\mathrm{D}_{\theta_0} R)(e_i) w(t)$ and $r_j(t) = (\mathrm{D}_{\theta_0} R)(e_j) w(t)$ by Parseval's Theorem. The results for γ_H are stated analogously.

[16] The results about asymptotic variance expressed in (39) have been derived by Ninness and Hjalmarsson in [4] and further studied by Mårtensson in [3]. Our contribution is the result on asymptotic-covariance w.r.t. frequency in (38) which explains the off-diagonal elements, i.e., $k(\omega_1, \omega_2)$ for $\omega_1 \neq \omega_2$, of the corresponding integral kernels.

E. Rational Modules for the Sensitivity Space

In formula (39a) of the preceding section the kernel $k_{0,R}(\omega,\omega)$ was just implicitly defined. It would be preferable to derive explicit formulas like

$$\mathrm{var}_{\mu_0}[G(\hat{\theta}_N)] \approx \frac{\Phi_{v_0}(\omega)}{N\Phi_u(\omega)}\left(k + 2\sum_{i=1}^{a}\frac{1-|\xi_i|^2}{|\mathrm{e}^{\mathrm{i}\omega}-\xi_i|^2}\right) \tag{42}$$

with $k = b-a \geq 0$, $A_0 = \prod_{i=1}^{a}(1-z^{-1}\xi_i)$, quantifying $k_{0,R}$ in terms of the poles of true system G_0, see e.g. [4]. A look at Theorem 6 reveals this is possible when the sensitivity space forms a rational module, i.e., $\mathrm{Im}(\mathrm{D}_{\theta_0}R) = \mathrm{X}^q$ for some $[z]$-stable $q \in \mathbb{R}[z]$. A necessary and sufficient condition for this in the case $a \leq b$ is given in Theorem 11. On the other hand it turns out that $\mathrm{Im}(\mathrm{D}_{\theta_0}S)$ always forms a rational module (cf. Theorem 12) and thus analytic expressions for (39b) similar to (42) in terms of the poles and zeros of H_0 always exist.

Remark 10: For an arbitrary polynomial $p \in K[x]$ of degree n we use the notation ∇p for $x^{-n}\cdot p \in K(x)$.[17] Let $p = \prod_{i=1}^{n}(z-a_i) \in \mathbb{R}[z]$ with $0 < |a_i| < 1$ and k denote a non-negative integer and $q = z^k p$. Then $\{z^i q^{-1}\}_{i=0}^{n+k-1}$ is a basis of $\mathrm{X}^q \subseteq \mathrm{R}H^2$ with $z^i q^{-1} = z^{-(n+k-i)}L^{-1}$, where $L = \nabla p \in \mathbb{R}[z^{-1}]$ equals $z^{-n}p$ by definition. In particular (i) For every pair (k, L) with $L \in \mathbb{R}[z^{-1}]$ being $[z^{-1}]$-antistable and k non-negative integer the vectors $z^{-1}/L, \ldots, z^{-m}/L$ with $m = n + k$ form a basis for the rational module $\mathrm{X}^{z^k\nabla L} \subseteq \mathrm{R}H^2$. (ii) Conversely a necessary condition for $\{z^{-i}/L\}_{i=1}^{m}$ with $L^{-1} \in \mathrm{R}H^\infty$ to span a rational module $\mathrm{X}^q \subseteq \mathrm{R}H^2$ is that $L \in \mathbb{R}[z^{-1}]$ such that it is $[z^{-1}]$-anti-stable and $m \geq n$ where $n = \deg(L)$. In this case $q = z^m L = z^k\nabla L$ with $k = m-n$. ■

Theorem 11: Assume $k = b-a \geq 0$ in (27) and $\rho_0 = \mathrm{S}_u/(\sigma H_0) \in \mathrm{R}H^\infty$ holds.[18] Then $\mathrm{Im}(\mathrm{D}_{\theta_0}R)$ is a rational module X^q if and only if $A_0^2/\rho_0 \in \mathbb{R}[z^{-1}]$ has degree n with $n \leq a+b$, in which case we have $q = z^k\nabla(A_0^2/\rho_0)$. As a special case for a white input sequence, i.e., $\mathrm{S}_u = \sigma_u^2$, and a trivial noise model $H_0 = 1$ we have $q = z^k(\nabla A_0)^2$.[19]

Proof: Consider the curve $\Gamma : (-1, 1) \to \mathcal{G}$

$$\Gamma : \tau \mapsto p(\tau)/d(\tau), \quad \text{s.t.} \quad p(\tau), d(\tau) \in \mathbb{R}[z] \tag{43}$$

where, w.l.o.g., we may assume that $p(\tau), d(\tau)$ are coprime for all τ, see e.g. [10], and $\Gamma(0) = p_0/d_0 = G(\theta_0)$. Note that $p = \nabla B$ and $d = z^k\nabla A$ define the curves $A, B : (-1, 1) \to \mathbb{R}[z^{-1}]$. Denote $\frac{\partial}{\partial\tau}\Gamma$ evaluated at $\tau = 0$ by $\dot{\Gamma}(0)$ then

$$\begin{aligned}\dot{\Gamma}(0) &= \frac{\dot{p}(0)}{d_0} - \frac{p_0\dot{d}(0)}{d_0^2} \\ &= \frac{\dot{p}(0)z^k\nabla A_0 - p_0 z^k(\dot{\nabla}A)(0)}{z^{2k}\nabla A_0^2} \in \mathrm{X}^{z^k\nabla A_0^2}.\end{aligned} \tag{44}$$

This proves $\mathrm{Im}(\mathrm{D}_{\theta_0}G) \subseteq \mathrm{X}^{z^k\nabla A_0^2}$. Since both are linear spaces of dimension $a+b$ this also implies equality. By part (i) in Remark 10 we have that $\{\frac{\rho_0 z^{-i}}{A_0^2}\}_{i=1}^{a+b}$ is a basis of $\mathrm{Im}(\mathrm{D}_{\theta_0}R)$ and, together with $\rho_0/A_0^2 \in \mathrm{R}H^\infty$ and part (ii) of Remark 10, the latter is a rational module in $\mathrm{R}H^2$ if and only if $A_0^2/\rho_0 \in \mathbb{R}[z^{-1}]$ has degree less than or equal $a+b$. □

Theorem 12: The linear space $\mathrm{Im}(\mathrm{D}_{\theta_0}S)$ forms the rational module $\mathrm{X}^q \subseteq \mathrm{R}H^2$ with $q = \nabla C_0\nabla D_0$.

Proof: Let I denote the Cartesian product of $\{1, \ldots, c\}$ and $\{1, \ldots, d\}$. A basis for $\mathrm{D}_{\theta_0}H$ is given by $\left\{\frac{\partial H(\theta_0)}{\partial\theta_{C,k}}, \frac{\partial H(\theta_0)}{\partial\theta_{D,l}}\right\}$ with $(k, l) \in I$. This basis is given by z^{-k}/D_0 and $-C_0 z^{-l}/D_0^2$, respectively. The elements in the basis for $\mathrm{Im}(\mathrm{D}_{\theta_0}S)$ are thus given by z^{-k}/C_0 and z^{-l}/D_0 with $(k, l) \in I$ respectively. By (i) in Remark 10 these span the rational modules $\mathrm{X}^{\nabla C_0}$ and $\mathrm{X}^{\nabla D_0}$, respectively. Since $\nabla C_0, \nabla D_0$ are coprime it follows that

$$\mathrm{X}^{\nabla C_0} + \mathrm{X}^{\nabla D_0} = \mathrm{X}^{\nabla C_0\cdot\nabla D_0} \tag{45}$$

which concludes the proof. □

V. Conclusions and Future Work

We have derived the asymptotic auto-covariance w.r.t. to frequency of the estimated transfer function based on the prediction error framework for the Box-Jenkins model structure. We saw that the asymptotic auto-covariance equals the integral kernel of the sensitivity space. The latter can be fully characterized in terms of the poles and zeros of the underlying dynamical system if the sensitivity space forms a rational module. In follow up papers we will demonstrate the power of the real rational module framework which we believe leads to a coordinate free and natural treatment of the asymptotic Fisher information metric and integral kernels reproducing it on the tangent space of the model manifold.

VI. Acknowledgments

The authors would like to thank Paul Van Dooren for helpful discussions.

References

[1] P. Caines and L. Ljung, "Asymptotic Normality of Prediction Error Estimators for Approximate System Models," in *Stochastics*, vol. 3, January 1979, pp. 29–46.

[2] J. Mårtensson and H. Hjalmarsson, "Variance error quantification for identified poles and zeros," *Automatica*, July 2007, submitted.

[3] J. Mårtensson, "Geometric analysis of stochastic model errors in system identification," Ph.D. dissertation, KTH, 2007.

[4] B. Ninness and H. Hjalmarsson, "Variance error quantifications that are exact for finite-model order," in *Transactions on Automatic Control*, vol. 49, no. 8, August 2004, pp. 1275–1290.

[5] R. E. Kalman, P. L. Falb, and M. A. Arbib, *Topics in Mathematical System Theory*. McGraw-Hill Book Company, 1968.

[6] W. Rudin, *Real and complex analysis, 3rd ed.* New York, NY, USA: McGraw-Hill, Inc., 1987.

[7] B. Okutmustur, "Reproducing Kernel Hilbert Spaces," Ph.D. dissertation, Bilkent University, August 2005.

[8] A. Beurling, "On two problems concerning linear transformations in Hilbert space," in *Acta Math.*, vol. 81, 1949, pp. 239–255.

[9] L. Ljung, *System Identification: Theory for the User (2nd Edition)*. Prentice Hall PTR, 1999.

[10] U. Helmke and P. A. Fuhrmann, "Tangent spaces of rational matrix functions," *Linear Algebra and its Applications*, vol. 271, pp. 1–40, 1998.

[17] For example if $p \in \mathbb{R}[z^{-1}]$ we have $\nabla p = (z^{-1})^{-n}p = z^n p$ whereas if $p \in \mathbb{R}[z]$ we have $\nabla p = z^{-n}p$. Note $\nabla\nabla p = p$ ($p \in K[x]$ s.t. $x \nmid p$).

[18] The case $b < a$, i.e., $k = a-b$ being positive, is more involved since then $\mathrm{Im}(\mathrm{D}_{\theta_0}G)$ becomes a rational module if and only if $z^k | A_0^2$ which is false in general.

[19] The term A_0^2/ρ_0 was denoted by $A_\dagger$ in [4] where one imposed it to be a polynomial in z^{-1} of degree less than or equal $a+b$ just as in Theorem 11 here. Thus our results clarify the main motivation behind such a technical assumption.