

Chapter 2

Prediction-Error Identification of LPV Systems: Present and Beyond

Roland Tóth, Peter S.C. Heuberger, and Paul M.J. Van den Hof

Abstract The proposed chapter aims at presenting a unified framework of prediction-error based identification of LPV systems using freshly developed theoretical results. Recently, these methods have got a considerable attention as they have certain advantages in terms of computational complexity, optimality in the stochastic sense and available theoretical tools to analyze estimation errors like bias, variance, etc., and the understanding of consistency and convergence. Beside the introduction of the theoretical tools and the prediction-error framework itself, the scope of the chapter includes a detailed investigation of the LPV extension of the classical model structures like ARX, ARMAX, Box–Jenkins, OE, FIR, and series expansion models, like orthonormal basis functions based structures, together with their available estimation approaches including linear regression, nonlinear optimization, and iterative IV methods. Questions of model structure selection and experimental design are also investigated. In this way, the chapter provides a detailed overview about the state-of-the-art of LPV prediction-error identification giving the reader an easy guide to find the right tools of his LPV identification problems.

2.1 Introduction

To design efficient *linear parameter-varying* (LPV) controllers, it is has a paramount importance to have an accurate but at the same time low-complexity LPV model of the underlying behavior of the system at hand. In engineering, it is common to use *first-principle laws* of physics, chemistry, biology, etc., to construct dynamic

R. Tóth (✉) • P.S.C. Heuberger • P.M.J. Van den Hof
Delft Center for Systems and Control, Delft University of Technology, Mekelweg 2,
2628 CD, Delft, The Netherlands
e-mail: r.toth@tudelft.nl; p.s.c.heuberger@tudelft.nl; p.m.j.vandenhof@tudelft.nl

models. However, such a procedure requires a detailed process knowledge from specialists. To assemble the existing knowledge into a coherent and compact mathematical description is not only a challenging task but it usually results in a too complex model as it is hard to distinguish relevant effects from negligible terms. The selection of the *scheduling variable* itself is also often restricted by the way of model construction and likely different choices follow from linearization-based or direct conversion based methods, see [15, 19, 23]. Therefore, modeling is often found to be very laborious and expensive. If the specialist's knowledge is lacking, like in case of poorly understood systems, the derivation of a model from first principles is even impossible. Moreover, certain quantities, like coefficients, rates, etc., required for the model, are likely unknown and have to be estimated by performing dedicated experiments.

Descriptions of systems can alternatively be derived by *system identification* (ID), where the estimation of a dynamical model is accomplished directly from measured input–output data. The *expert's knowledge* still has a major role, as it gives the basic source of information in decisions on parametrization, model-structure selection, experiment design, and the actual way of deriving the estimate. This knowledge also helps in judging the quality and applicability of the obtained models. Even if system identification requires human intervention and expert's knowledge to arrive at appropriate models, it also gives a general framework in which most of the steps can be automated, providing a less laborious and cost intensive modeling process.

In the current literature, many LPV identification approaches have been developed using model structures that are formulated in terms of *state-space* (SS) and *linear-fractional representations* (LFR), e.g., [7, 18, 22, 33, 34], *input–output* (IO) representations, e.g., [1, 3, 11, 36], or *series-expansion* forms, e.g., [23, 27, 28]. Most of the existing approaches use a discrete-time setting and commonly assume *static* dependence on the scheduling variable $p : \mathbb{Z} \rightarrow \mathbb{P}$, with $\mathbb{P} \subseteq \mathbb{R}^m$. Here, static dependence means dependence of the model coefficients only on the instantaneous value of p . For a recent overview of the available methods, see [4, 23].

Recently, LPV–IO model structures based methods have got a considerable attention as they appear to have certain advantages w.r.t. other identification approaches of the field. One of the major benefits is that identification of this representation-based model structures can be addressed via the extension of the LTI *prediction-error* (PE) framework [23, 25]. In opposition with other approaches, this enables the stochastic analysis of the estimates, treatment of general noise models [13, 25], experimental design [6, 12, 36], model structure selection, and direct identification of the involved dependencies [11, 30, 31] often in a computationally attractive manner and also in continuous time [14]. Moreover, in this setting it is also relatively easy to identify models with *dynamic dependence* (dependence of the coefficients on time shifted instances of p), which is often required for high accuracy identification of nonlinear systems (see [23]). However, the main stream LPV control-synthesis approaches are based on models defined in an SS or an LFR form, hence the delivered IO model needs to be converted to such representation forms. Due to the fact that multiplication with any time operator is not commutative over

the p -dependent coefficients, the existing realization theory is more complicated than in the LTI case and often introduces rational dynamic dependence on p in the resulting state-minimal SS forms [26, 32]. It is possible to avoid this phenomenon by aiming at minimal realization in terms of the involved dependency as well, which often requires either auxiliary state variables or special parametrization of the polynomial forms [24]. Improving SS realization of IO models in terms of finding the state basis that provide the simplest scheduling dependence and minimal state dimension is in the focus of current research activities in this area.

In this chapter, we give an overview about the state-of-the-art of LPV prediction-error identification. In particular, we focus on what is feasible by the available approaches, what the practical advantages are, and what developments are still needed. Due to the broad scope of the topic, we will address here only the discrete-time case.

The chapter is organized as follows: First in Sect. 2.2, the concept of LPV series expansion representations is introduced which makes it possible to formulate predictors and model manipulations later on. Next in Sect. 2.3, the basic setting of the LPV prediction-error framework is defined with the concept of the data-generating system, noise models, and one-step-ahead predictors. A general structure of parameterized models and the perspective of estimation in the ℓ_2 -optimal prediction-error sense, identifiability of model structures and informativity of datasets are also studied. Then in Sect. 2.4, the LPV extension of the classical model structures is introduced and their properties are analyzed. This is followed in Sect. 2.5 by a detailed investigation of the available estimation approaches w.r.t. these models in terms of linear regression, nonlinear optimization and iterative instrumental variable methods and their stochastic properties.

2.2 LPV Series-Expansion Representations

In the LTI case, many key concepts and model manipulations in the PE framework are based on a transfer function representation of the dynamic behavior (see [16]). One of the major problems which has so far prevented the analysis of the PE methods in the LPV case has been the lack of a transfer function representation of LPV systems which expresses signal relations in the frequency domain. To illustrate the problem, consider the classical LPV filter form of discrete-time IO representations, often defined in the *single-input single-output* (SISO)¹ case as:

$$\sum_{i=0}^{n_a} a_i(p(k))q^{-i}y(k) = \sum_{j=0}^{n_b} b_j(p(k))q^{-j}u(k), \quad (2.1)$$

¹LPV-IO representations can also be defined for *multiple-input multiple-output* (MIMO) systems in a similar form as (2.1), see [23].

where $u : \mathbb{Z} \rightarrow \mathbb{R}$ is the input, $y : \mathbb{Z} \rightarrow \mathbb{R}$ is the output, and $p : \mathbb{Z} \rightarrow \mathbb{P}$ is the scheduling variable of the LPV system \mathcal{S} represented by (2.1); q is the (forward) time-shift operator, i.e., $q^{-1}u(k) = u(k-1)$, $n_a, n_b \geq 0$ and $a_i : \mathbb{P} \rightarrow \mathbb{R}$ and $b_j : \mathbb{P} \rightarrow \mathbb{R}$ are functions of $p(k)$ (instantaneous value of p) which is called *static dependence*. The functions a_i and b_j can have arbitrary complexity ranging from simple linear to rational or real *meromorphic*² dependence. To guarantee well-posedness of (2.1), it is often assumed that all a_i and b_j are bounded on \mathbb{P} .

In identification, we aim to estimate a dynamical model of the system based on measured data, which corresponds to the estimation of each a_i and b_j in (2.1). To formulate estimation of these functions, it is attractive to introduce

$$A(q^{-1}, p(k)) = \sum_{i=0}^{n_a} a_i(p(k))q^{-i} \quad \text{and} \quad B(q^{-1}, p(k)) = \sum_{j=0}^{n_b} b_j(p(k))q^{-j}$$

as polynomials in q^{-1} with varying coefficients $a_i(p(k))$ and $b_j(p(k))$ and, inspired by the LTI system theory, to write

$$y(k) = F(q^{-1}, p(k))u(k) \quad \text{with} \quad F(q^{-1}, p(k)) = \frac{B(q^{-1}, p(k))}{A(q^{-1}, p(k))}. \quad (2.2)$$

However, $F(q^{-1}, p(k))$ in (2.2) relates to a transfer function if and only if $p(k)$ is a constant signal, i.e., $p(k) = p$ for all k , where $p \in \mathbb{P}$. This is justified by the fact that if q is substituted with the complex z variable, then $F(z^{-1}, p(k))$ is a mixed frequency–time relationship. If $Y(z)$ and $U(z)$ denote the Z -transform of the signals u and y on an appropriate region of convergence, then $Y(z) = F(z^{-1}, p(k))U(z)$ has a meaning if and only if p , associated with (u, y) , is a constant (not-varying with time). Furthermore, $F(q^{-1}, p(k))$ is ill-defined also as an operator because multiplication with q^{-1} is not commutative over time-dependent coefficients such as $b_j(p(k))$, i.e., $q^{-1}b_j(p(k)) = b_j(p(k-1))q^{-1}$. Therefore, multiplication from the left or right has different meaning. In (2.2), it is ambiguous whether left or right multiplication is intended to derive this rational operator form. Currently no theoretical framework is available (to the author’s knowledge) to handle rational time-operator forms with time-dependent coefficients (such a framework does exist in case of constant coefficients, i.e., in the LTI case, see [37]).

To overcome this “representation” problem, it has been shown in [23] that the dynamic mapping between u and y can be characterized as a convolution involving p and u . This so-called *impulse response representation* (IRR) is given as

$$y(k) = \sum_{i=0}^{\infty} (g_i \diamond p)(k)u(k-i) = \left(\sum_{i=0}^{\infty} (g_i \diamond p)q^{-i}u \right)(k) = ((G(q) \diamond p)u)(k), \quad (2.3)$$

² $h : \mathbb{R}^n \rightarrow \mathbb{R}$ is a real meromorphic function if $h = f/g$ with f, g analytic and $g \neq 0$.

where the so-called *impulse response coefficients* g_i are functions of $p(k)$ and of multiple, but finite many, time-shifted instances of p , like $g_i(p(k + \tau_1), \dots, p(k - \tau_2))$ with $\tau_1, \tau_2 \geq 0$. This is called *dynamic dependence*. To express such a broad range of dependencies conveniently, we apply the operator $\diamond: (\mathcal{R}, \mathbb{P}^{\mathbb{Z}}) \rightarrow \mathbb{R}_{\infty}^{\mathbb{Z}}$, where \mathcal{R} is the set of all real meromorphic functions with finite dimensional domain, such that $(g_i \diamond p)(k) = g_i(p(k + \tau_1), \dots, p(k - \tau_2))$. Note that in the sequel, we will use \diamond to express dynamic or general dependence like $(g_i \diamond p)(k)$ whenever it is needed and we will use $g_i(p(k))$ to express if a coefficient has only static dependence. For an illustration, consider the following example.

Example 2.1. Given an asymptotically stable discrete-time LPV–IO representation:

$$y = -0.1 p q^{-1} y - 0.2 q^{-2} y + \sin(p) q^{-1} u, \quad (2.4)$$

with $\mathbb{P} = [0, 1]$. By recursive substitution for $q^{-1}y, q^{-2}y, \dots$, the following IRR, equivalent with (2.4), results

$$\begin{aligned} y = & \underbrace{\sin(p)}_{g_1 \diamond p} q^{-1} u + \underbrace{(-0.1 p \sin(q^{-1} p))}_{g_2 \diamond p} q^{-2} u \\ & + \underbrace{(0.01 p (q^{-1} p) - 0.2) \sin(q^{-2} p)}_{g_3 \diamond p} q^{-3} u + \dots, \end{aligned}$$

where the sequence of functions $g_i \diamond p$ converges to zero as $i \rightarrow \infty$.

Equation (2.3) can be considered as a series expansion of \mathcal{S} in terms of q^{-i} which is convergent if \mathcal{S} is asymptotically stable. Furthermore, (2.3) can be seen as the generalization of LPV–SS and LPV–IO representations with appropriate equivalence transformations available (see [23]).

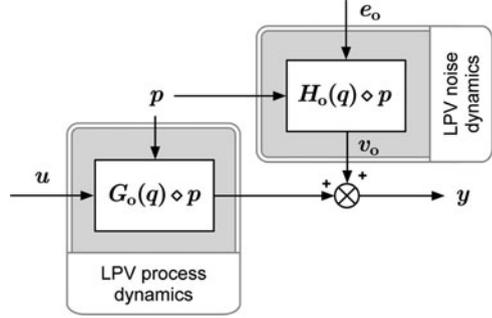
2.3 An LPV Prediction-Error Framework

By using the LPV impulse response representation, established in the previous section, it becomes possible to extend the classical PE framework to the LPV case allowing sophisticated analysis of the estimation of LPV–IO models. To do so, we will first define the concept of an LPV data-generating system. This will be followed by deriving a one-step-ahead predictor for the observed output sequence that we will use to formulate the estimation of parametrized models under a mean-squared prediction-error criterion.

2.3.1 Data-Generating System

According to the classical PE setting, the *data-generating system* is considered as a discrete-time deterministic filter G_o whose output is influenced by a stochastic

Fig. 2.1 LPV concept of the data-generating system



noise process v_o in an additive manner (see Fig. 2.1). It is assumed that v_o is a *quasi-stationary* noise process with a bounded *power spectral density* $\Phi_{v_o}(\omega)$. In case G_o is linear, it is possible to lump many different sources of disturbances in v_o , such as noise-corrupted actuation, uncontrollable inputs, process noise, etc., because, under minor restrictions, all these effects can be propagated through G_o . In the LTI case, this suggests that for many systems it is valid to assume that $\Phi_{v_o}(\omega)$ is a rational function, i.e., that v_o can be represented as a filtered zero-mean white noise process.

Using this concept, a data-generating LPV system \mathcal{S}_o can be analogously formulated as

$$y(k) = (G_o(q) \diamond p)(k) u(k) + v_o(k), \quad (2.5)$$

where the process part is represented by an LPV impulse response form

$$G_o(q) \diamond p = \sum_{i=0}^{\infty} (g_i^o \diamond p) q^{-i}, \quad (2.6)$$

with g_i being bounded w.r.t. every $p \in \mathcal{P}$, where $\mathcal{P} \subseteq \mathbb{P}^{\mathbb{Z}}$ denotes all possible trajectories of p that are allowed during the operation of the system. In order to guarantee that (2.6) is convergent, it is a necessary assumption that G_o under \mathcal{P} represents an asymptotically stable LPV system. Additionally, it is assumed that the disturbance v_o can be described as

$$v_o(k) = (H_o(q) \diamond p)(k) e_o(k), \quad (2.7)$$

where H_o is a convergent LPV-IRR, i.e., it corresponds to an asymptotically stable LPV filter, it is monic, i.e., $H_o(\infty) = 1$, and $e_o(k)$ is a zero-mean white noise process. Similar to the LTI case, asymptotic stability of H_o in the deterministic sense is a necessary assumption of the PE setting (see [16]), otherwise $\Phi_{v_o}(\omega)$ would not be bounded yielding that ID of $G_o(q)$ is an ill-posed problem. Furthermore, it is important to point out that in terms of (2.7), $\mathbb{E}\{v_o(k)\} = 0$ for each k , where \mathbb{E} denotes the expectation operator, but the autocorrelation of v_o , i.e., $R_{v_o, v_o}(k, \tau) = \mathbb{E}\{v_o(k)v_o(k - \tau)\}$, is time-dependent. However in the asymptotic

sense, $\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{k=1}^N R_{v_o, v_o}(k, \tau)$ exists for a given p and for all $\tau \in \mathbb{Z}$ due to the convergence properties of $H_o(q)$ and the independence of p from e_o . Hence v_o qualifies as a quasi-stationary signal (see [25] for the detailed proof).

2.3.2 One-Step-Ahead Prediction of v_o

In order to formulate the estimation of parametrized LPV models of (2.5) in a prediction-error setting, it is necessary to derive a predictor of y . The simplest case is to characterize a *one-step-ahead predictor* of y , for which it is essential to clarify how we can predict $v_o(k)$ at a given time-step k if we have observed $v_o(\tau)$ for $\tau \leq k-1$. In terms of (2.7):

$$v_o(k) = (h_0^o \diamond p)(k) e_o(k) + \sum_{i=1}^{\infty} (h_i^o \diamond p)(k) e_o(k-i), \quad (2.8)$$

meaning that $\{v_o(\tau)\}_{\tau \leq k-1}$ and a given trajectory of p defines $\{e_o(\tau)\}_{\tau \leq k-1}$. Note that if each h_i^o depends only on the current and the backward-time-shifted values of p , e.g., $p(k), p(k-1), p(k-2)$, etc., which is called *causal dynamic dependence*, then only the knowledge of $\{p(\tau)\}_{\tau \leq k}$ and $\{v_o(\tau)\}_{\tau \leq k-1}$ is sufficient to characterize $\{e_o(\tau)\}_{\tau \leq k-1}$. In case the noise process has an LPV-IO representation in terms of (2.7) with only static p -dependence, then the equivalent IRR form (2.8) is guaranteed to have only causal dependence [23]. In most practical applications, causal dynamic dependence on p is quite realistic.

To follow the classical concept of defining the prediction of $v_o(k)$, assume that observations of $e_o^{(k-1)} = \{e_o(\tau)\}_{\tau \leq k-1}$ and $p^{(k)} = \{p(\tau)\}_{\tau \leq k}$ are given. Under this information set, our objective is to compute the one-step-ahead prediction of $v_o(k)$ w.r.t. the ℓ_2 -loss:

$$\hat{v}_o(k|k-1) = \arg \min_{\delta \in \mathbb{R}} \mathbb{E} \left\{ \|v_o(k) - \delta\|_{\ell_2}^2 \mid e_o^{(k-1)}, p^{(k)} \right\}. \quad (2.9)$$

In [23, 25] it is shown that if $p^{(k)}$ is fully known, then (2.9) is equal to

$$\hat{v}_o(k|k-1) = \sum_{i=1}^{\infty} (h_i^o \diamond p)(k) e_o(k-i). \quad (2.10)$$

It easily follows that (2.9) also minimizes the mean-squared error of the prediction. Of course it is not practical to assume that $e_o^{(k-1)}$ is known. To formulate prediction of $v_o(k)$ in terms of observation of $v_o^{(k-1)} = \{v_o(\tau)\}_{\tau \leq k-1}$, it is required that H_o has a stable inverse, i.e., there exists a monic convergent LPV-IRR denoted as H_o^\dagger such that $(H_o^\dagger(q)H_o(q)) \diamond p = 1$. Note that if such a H_o^\dagger exists, then it is a bi-lateral inverse

of H , i.e., $H_0^\dagger(q)H_0(q) = H_0(q)H_0^\dagger(q) = 1$, which can be shown based on telescopic sums, see [25]. This implies that

$$e_o(k) = (H_0^\dagger(q) \diamond p)(k) v_o(k). \quad (2.11)$$

Then using (2.11), we can write (2.10) as

$$\hat{v}_o(k|k-1) = ((H_0(q) - 1) \diamond p)(k) e_o(k) = (1 - H_0^\dagger(q) \diamond p)(k) v_o(k), \quad (2.12)$$

which is the LPV form of the classical one-step-ahead predictor result [16].

2.3.3 One-Step-Ahead Prediction of v_o with Noisy p

In the previous derivation, it was essential that full, i.e., noise-free observation of the sequence $p^{(k)}$ is available, which we will call the “ p -true case”. In the LPV literature, such an assumption is generally taken as a technical necessity regardless of the used identification setting (see [6, 7, 12, 18, 22, 33, 34, 36], exceptions: [3, 5]) and the resulting methods based on it are almost exclusively applied in practical situations where measurements of p are polluted by noise with various stochastic properties. The reason for this theoretical gap lies in the difficulty to establish a conditional expectation of $v_o(k)$ in the situation when instead of $p(k)$ only observations of

$$p_*(k) = p(k) + w_o(k), \quad (2.13)$$

with $w_o(k)$ is an iid noise process, are available as each h_i^o can be a nonlinear function with dynamic dependence on p . For systems with simple dependencies, formulation of $\hat{v}_o(k|k-1)$ is possible but no general formula can be given based on the current results (see [3] for an analysis of consistency for LPV *autoregressive with exogenous input* (ARX) models under stochastic p). It has been only recently shown that a general approach to formulate the one-step-ahead predictor in case of noisy observations of $p(k)$ can be derived from moment-generating functions of the underlying distribution of $w_o(k)$ [25]. Furthermore, a parametrized noise model to capturing nonwhite noise on p can also be applied. However, currently the stochastic properties of the estimated models are not well understood in that case. For the sake of simplicity and coherence of our overview, we will restrict our attention to the p -true case and investigate estimation under such an assumption.

2.3.4 One-Step-Ahead Prediction of y

As a next step, we need to formulate the one-step-ahead predictor of $y(k)$ to address identification of a parametrized model in terms of minimizing the prediction error,

i.e., the difference between $y(k)$ and the predicted model output, which is the primary goal in the PE setting. Consider the p -true case with $y^{(k-1)} = \{y(\tau)\}_{\tau \leq k-1}$, $u^{(k)} = \{u(\tau)\}_{\tau \leq k}$, and $p^{(k)} = \{p(\tau)\}_{\tau \leq k}$ (in case of causal dependence). Since (2.5) implies that

$$v_o(k) = y(k) - \sum_{i=0}^{\infty} (g_i^o \diamond p)(k) u(k-i), \quad (2.14)$$

$v_o(\tau)$ is characterized for $\tau \leq k-1$ w.r.t. the information set $(u^{(k)}, p^{(k)}, y^{(k-1)})$. Based on (2.12), it is not complicated to show that under the given information set, the one-step-ahead prediction of $y(k)$ w.r.t. the ℓ_2 loss is

$$\hat{y}(k|k-1) = ((H_o^\dagger(q)G_o(q)) \diamond p)(k) u(k) + (1 - H_o^\dagger(q) \diamond p)(k) y(k). \quad (2.15)$$

This corresponds to the LPV form of the classical result of the LTI case (see [16]). Note that in a similar manner, k -step-ahead predictors can also be formulated in this setting. For a detailed proof, see [23, 25].

2.3.5 Parametrized Models and Estimation

Now, introduce an LPV parametrized model in the form of

$$(G(q, \theta_g), H(q, \theta_h)), \quad (2.16)$$

where $G(q, \theta_g)$ and $H(q, \theta_h)$ are the IRRs of the process part, denoted as \mathcal{G}_{θ_g} , and the noise part, denoted as \mathcal{H}_{θ_h} , of the model structure, respectively, and $\theta_g \in \Theta_g \subseteq \mathbb{R}^{n_g}$ with $\theta_h \in \Theta_h \subseteq \mathbb{R}^{n_h}$ are the parameters to be estimated. Note that these parameters are not necessarily associated with the parametrization of impulse response coefficients directly, but can correspond to the parametrization of the coefficients of the process and noise models given in an SS or IO form. Then these parametrized structures are represented by the IRRs: G and H . Also introduce $\theta = \text{col}(\theta_g, \theta_h) \in \Theta \subseteq \mathbb{R}^{n_\theta}$, the vector of independent parameters in θ_g and θ_h .

Denote $\mathcal{G} = \{\mathcal{G}_{\theta_g} \mid \theta_g \in \Theta_g\}$ and $\mathcal{H} = \{\mathcal{H}_{\theta_h} \mid \theta_h \in \Theta_h\}$ the collection of all process and noise models with the considered parametrization and similarly denote the overall parametrized model (2.16) as \mathcal{M}_θ . Then, based on the model structure \mathcal{M}_θ , the model set, denoted as \mathcal{M} , takes the form

$$\mathcal{M} = \{ (\mathcal{G}_{\theta_g}, \mathcal{H}_{\theta_h}) \mid \theta = \text{col}(\theta_g, \theta_h) \in \Theta \}. \quad (2.17)$$

This set corresponds to the set of candidate models in which we seek the model that explains data gathered from \mathcal{S}_o the best, under a given criterion. We denote by $\mathcal{S}_o \in \mathcal{M}$, when the data-generating system is in the model set, i.e., $\exists \theta_o = \text{col}(\theta_{o,g}, \theta_{o,h}) \in \Theta$ such that $G_o(q) \diamond p = G(q, \theta_{o,g}) \diamond p$ and $H_o(q) \diamond p = H(q, \theta_{o,h}) \diamond p$.

With respect to (2.16), we can define the *one-step-ahead prediction error* as

$$e_\theta(k) = y(k) - \hat{y}(k | \theta, k-1), \quad (2.18)$$

where

$$\hat{y}(k | \theta, k-1) = (W_u(q, \theta) \diamond p)(k) u(k) + (W_y(q, \theta) \diamond p)(k) y(k). \quad (2.19)$$

is the one step-ahead predictor of the model output with *subpredictors*

$$W_u(q, \theta) = H^\dagger(q, \theta_g)G(q, \theta_h) \quad \text{and} \quad W_y(q, \theta) = 1 - H^\dagger(q, \theta_h) \quad (2.20)$$

according to (2.15). Denote a data sequence of \mathcal{S}_o by $\mathcal{D}_N = \{y(k), u(k), p(k)\}_{k=1}^N$. Then the basic philosophy of PE-based identification is to find θ w.r.t. a given parametrized model \mathcal{M}_θ with parameter space $\Theta \subseteq \mathbb{R}^{n_\theta}$ and a dataset \mathcal{D}_N such that the one-step-ahead predictor (2.19) associated with θ provides a prediction error e_θ which resembles a zero-mean white noise “as much as possible”.

Based on the predictor form (2.19), many different classical *identification criteria* can be applied to achieve this objective in terms of “minimization” of e_θ subject to θ . A particularly interesting choice is the *mean-squared prediction error* or more often called as the *least-squares* (LS) criterion:

$$V(\mathcal{D}_N, \theta) = \frac{1}{N} \sum_{k=1}^N e_\theta^2(k) = \frac{1}{N} \|e_\theta\|_{\ell_2}^2, \quad (2.21)$$

such that the parameter estimate is

$$\hat{\theta} = \underset{\theta \in \Theta}{\operatorname{argmin}} V(\mathcal{D}_N, \theta). \quad (2.22)$$

Other criteria can also be used to characterize estimation of θ in (2.16) as a minimization of (2.18) w.r.t. a chosen measure (see [16]) or to introduce other objectives, e.g., minimization of the support of θ , or weights, like forgetting factors. However for sake of simplicity, we restrict the main stream of discussion to the classical LS case.

2.3.6 Identifiability and Informativity

To guarantee uniqueness of (2.22), one condition is that \mathcal{M}_θ is *globally identifiable*.

Definition 2.1 (Identifiability, based on [8]). The model structure \mathcal{M}_θ , defined by (2.16) with a parameter domain $\Theta \subseteq \mathbb{R}^{n_\theta}$, is called locally identifiable at a $\theta_1 \in \Theta$, if $\exists \delta > 0$ such that for all $\theta_2 \in \Theta$ in $\|\theta_1 - \theta_2\| \leq \delta$:

$$W_y(q, \theta_1) = W_y(q, \theta_2) \quad \text{and} \quad W_u(q, \theta_1) = W_u(q, \theta_2) \quad \Rightarrow \quad \theta_1 = \theta_2. \quad (2.23)$$

The model structure is said globally identifiable at θ_1 if the same holds for arbitrary large δ . It is called globally identifiable if it is globally identifiable at all $\theta_1 \in \Theta$.

Another condition for uniqueness (2.22) is the *informativity* of the data set \mathcal{D}_N .

Definition 2.2 (Informative data, based on [8]). For a model structure \mathcal{M}_θ , defined by (2.16) with a parameter domain $\Theta \subseteq \mathbb{R}^{n_\theta}$, a quasi-stationary data set $\mathcal{D}_N = \{u(k), y(k), p(k)\}_{k=1}^N$ is called informative, if for any $\theta_1, \theta_2 \in \Theta$

$$\begin{aligned} & \bar{\mathbb{E}}\{(W_y(q, \theta_1) \diamond p)y - (W_y(q, \theta_2) \diamond p)y + (W_u(q, \theta_1) \diamond p)u \\ & \quad - (W_u(q, \theta_2) \diamond p)u\}^2 = 0, \end{aligned} \quad (2.24)$$

with $\bar{\mathbb{E}}$ being the generalized expectation operator,³ implies that

$$W_y(q, \theta_1) = W_y(q, \theta_2) \quad \text{and} \quad W_u(q, \theta_1) = W_u(q, \theta_2). \quad (2.25)$$

In terms of these definitions, if the model set is globally identifiable (no two different parameters θ_1 and θ_2 give rise to the same predictor) and the data set \mathcal{D}_N is informative, then $V(\theta, \mathcal{D}_N)$ has a global optimum in the statistical sense.

2.3.7 Consistency and Convergence

When applying the quadratic ID criterion (2.21), the asymptotic properties of the resulting parameter estimate can be analyzed in the situation when $N \rightarrow \infty$, similarly as in the LTI case. Consider the following definitions of consistency and convergence.

Definition 2.3 (Convergence). For an informative data set \mathcal{D}_N and model structure \mathcal{M}_θ , the parameter estimate $\hat{\theta}$ is called convergent if $N \rightarrow \infty$ implies that $\hat{\theta} \rightarrow \theta^*$ with probability one, i.e., $P(\hat{\theta} = \theta^*) = 1$, where $\theta^* = \arg \min_\theta \bar{\mathbb{E}}\{e_\theta^2\}$.

Note that convergence implies that the asymptotic parameter estimate is independent of the particular noise realization in the data sequence and \mathcal{M}_θ is locally identifiable at θ^* .

Definition 2.4 (Consistency). For model structure \mathcal{M}_θ with model set \mathcal{M} and a data set \mathcal{D}_N which is informative w.r.t. \mathcal{M}_θ , a convergent parameter estimate $\hat{\theta} \rightarrow \theta^*$ is called consistent if either of the following conditions holds:

- If $\mathcal{S}_o \in \mathcal{M}$, then $G_o(q) = G(q, \theta_g^*)$ and $H_o(q) = H(q, \theta_h^*)$.
- If $\mathcal{S}_o \notin \mathcal{M}$ but $\mathcal{S}_o \in \mathcal{G}$, then $G_o(q) = G(q, \theta_g^*)$.

We will investigate these properties w.r.t. the particular identification approaches we consider in Sect. 2.5.

³The notation $\bar{\mathbb{E}}\{x\} = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{k=1}^N \mathbb{E}\{x(k)\}$ is adopted from the prediction-error framework of [16].

2.4 Classical Model Structures

Next we can investigate how the classical model structures can be formulated in the introduced LPV–PE setting. To follow the classical formulations, we will construct both the process and the noise components using an LPV–IO representation form. For the sake of simplicity, we treat these model structures such that their coefficients have only static dependence on p . Extension of these definitions using coefficients with dynamic dependence follows naturally.

2.4.1 Process Model

Consider the parametrized model \mathcal{M}_θ , where the process part \mathcal{G}_{θ_g} , whose IRR is given by $G(q, \theta_g)$, is defined by

$$A(q^{-1}, p(k), \theta_g) \check{y}(k) = B(q^{-1}, p(k), \theta_g) q^{-\tau_d} u(k). \quad (2.26)$$

Here \check{y} is the noise-free output of the process part, $\tau_d \geq 0$ is the input delay and the p -dependent polynomials

$$\begin{aligned} A(q^{-1}, p(k), \theta_g) &= 1 + \sum_{i=1}^{n_a} a_i(p(k), \theta_g) q^{-i}, \\ B(q^{-1}, p(k), \theta_g) &= \sum_{j=0}^{n_b} b_j(p(k), \theta_g) q^{-j}, \end{aligned}$$

with $n_a, n_b, \tau_d \geq 0$, are parametrized as:

$$a_i(p(k), \theta_g) = \sum_{l=0}^{n_\alpha} a_{i,l} \alpha_{i,l}(p(k)), \quad b_j(p(k), \theta_g) = \sum_{l=0}^{n_\beta} b_{j,l} \beta_{j,l}(p(k)), \quad (2.27)$$

with $\alpha_{i,0}(\cdot) = \beta_{j,0}(\cdot) = 1$. In this particular parametrization, which is called *linear parametrization*, $\{\alpha_{i,l}\}_{i=1, l=1}^{n_a, n_\alpha}$ and $\{\beta_{j,l}\}_{j=0, l=1}^{n_b, n_\beta}$ are priori given functions (chosen by the user) which are bounded on \mathbb{P} and

$$\theta_g = [a_{1,0} \cdots a_{n_a, n_\alpha} \ b_{0,0} \cdots b_{n_b, n_\beta}]^\top \in \Theta_g \subseteq \mathbb{R}^{n_g},$$

with $n_g = n_a(n_\alpha + 1) + (n_b + 1)(n_\beta + 1)$, represents the unknown parameters to be estimated for the process part. Note that parametrizations other than (2.27) are possible; however, the advantage of (2.27) is that a large number of functional depen-

dencies can be represented, based on the choice of $\{\alpha_{i,l}\}_{i=1,l=1}^{n_a,n_\alpha}$ and $\{\beta_{j,l}\}_{j=0,l=1}^{n_b,n_\beta}$, and it is linear in θ_g . Additionally for the structure (2.26) with static coefficient dependence, the impulse response coefficients $(g_i \diamond p)(k)$ have only causal dynamic dependence on p [23]. A particular advantage of linear parametrization is that it not only reduces the complexity of the associated estimation problem but at the same time it also makes the adequate selection of the structural dependencies well posed [23]. In terms of (2.27), this selection problem translates to a search for a set of functions $\{\alpha_{i,l}\}_{i=1,l=1}^{n_a,n_\alpha}$ and $\{\beta_{j,l}\}_{j=0,l=1}^{n_b,n_\beta}$ such that the true p -dependent coefficients a_i^o and b_j^o , associated with the IO representation of the underlying system, satisfy $a_i^o \in \text{Span}(\{\alpha_{i,l}\}_{l=0}^{n_\alpha})$ and $b_j^o \in \text{Span}(\{\beta_{j,l}\}_{l=0}^{n_\beta})$. In case of a black-box scenario, the choice of these functions can be arbitrary. One can consider all $\alpha_{i,l}$ and $\beta_{j,l}$ to be rational functions or polynomials with a fixed degree and a fixed order of dynamic dependence. However, the number of possible choices is enormous. Including a too large set of functions can easily lead to over-parametrization, while restriction of $\alpha_{i,l}$ and $\beta_{j,l}$ to only a few basic functions can lead to serious structural bias. In order to assist the selection of an efficient set of functional dependencies in the parametrization of linear regression models, recently practically applicable approaches have been proposed in [11, 30, 31]. In [11] a *dispersion functions* based method while in [30] a *support vector machine* approach, both originating from the *machine learning* field, have been developed. These approaches aim to learning the underlying static or dynamic nonlinear dependence of the coefficients with great efficiency. In [31], a coefficient shrinkage method, the so-called *nonnegative garrote* (NNG) approach originating from statistics, has been introduced for this purpose. The NNG uses regularization in terms of weights to penalize individual elements of the parameter vector θ . In this way, the approach starts with a relatively large set of possible functional dependencies from which those functions that do not contribute significantly to the validity of the estimated model are eliminated by decreasing their weights.

2.4.2 Noise Model

The noise model \mathcal{H}_{θ_h} , whose IRR is given by $H(q, \theta_h)$, is defined as

$$C(q^{-1}, p(k), \theta_h) v(k) = D(q^{-1}, p(k), \theta_h) e(k), \quad (2.28)$$

where $e(k)$ is a zero-mean white noise process and the p -dependent polynomials C and D are defined similarly as A and B with order $n_c, n_d \geq 0$. These polynomials are also considered to be monic in the sense that $c_0(\cdot) = d_0(\cdot) = 1$ and to have linear parameterization in terms of the functions $\{\gamma_{i,l}\}_{i=1,l=1}^{n_c,n_\gamma}$, $\{\delta_{j,l}\}_{j=1,l=1}^{n_d,n_\delta}$ bounded on \mathbb{P} and parameters

$$\theta_h = [c_{1,0} \cdots c_{n_c, n_\gamma} \ d_{1,0} \cdots d_{n_d, n_\delta}] \in \Theta_h \subseteq \mathbb{R}^{n_h},$$

with $n_h = n_c(n_\gamma + 1) + n_d(n_\delta + 1)$, respectively. Furthermore, it is assumed that the IRR $H(q, \theta_h)$ of (2.28) has a stable inverse denoted as $H^\dagger(q, \theta_h)$ for all values of $\theta_h \in \Theta_h$. Note that (2.28) is able to express a large variety of different noise processes. It also includes LTI noise models which admit a polynomial description since for $n_\gamma = 0$ and $n_\delta = 0$, C and D correspond to polynomials with constant coefficients and hence $H(q, \theta_h)$ can also be expressed as a transfer function. Similar to the process part, static coefficient dependence of (2.28) implies that $H(q, \theta_h)$ has causal dynamic dependence.

2.4.3 Overall Model Structure

Finally, we can define the overall model structure in terms of (2.26) and (2.28). Let $\theta = \text{col}(\theta_g, \theta_h) \in \Theta_g \times \Theta_h \subseteq \mathbb{R}^{n_\theta}$ with $n_\theta = n_g + n_h$ in case the parametrizations of \mathcal{G}_{θ_g} and \mathcal{H}_{θ_h} are independent. Otherwise θ is constructed from θ_g and θ_h such that it contains only independent parameters. Then the signal relations of the overall parametrized model \mathcal{M}_θ are given as

$$A(q^{-1}, p(k), \theta_g) \check{y}(k) = B(q^{-1}, p(k), \theta_g) q^{-\tau_d} u(k), \quad (2.29a)$$

$$C(q^{-1}, p(k), \theta_h) v(k) = D(q^{-1}, p(k), \theta_h) e(k), \quad (2.29b)$$

$$y(k) = \check{y}(k) + v(k). \quad (2.29c)$$

Note that in this general model structure, we can distinguish specific cases which correspond to the classical model structures used in the LTI setting.

2.4.3.1 LPV-ARX and ARMAX structures

By considering $C(q^{-1}, p(k), \theta_h) \triangleq A(q^{-1}, p(k), \theta_g)$ and $D(q^{-1}, p(k), \theta_h) \triangleq 1$ in (2.29a–c), the LPV version of the so-called *autoregressive with exogenous input* (ARX) model structure results:

$$A(q^{-1}, p(k), \theta_g) y(k) = B(q^{-1}, p(k), \theta_g) q^{-\tau_d} u(k) + e(k), \quad (2.30)$$

with $\theta = \theta_g$. It is important to acknowledge here that w.r.t. (2.19):

$$\hat{y}(k | \theta, k-1) = \underbrace{B(q^{-1}, p(k), \theta_g) q^{-\tau_d} u(k)}_{(W_u(q, \theta) \diamond p)(k)} + \underbrace{1 - A(q^{-1}, p(k), \theta_g)}_{(W_y(q, \theta) \diamond p)(k)} y(k). \quad (2.31)$$

This means that if A and B are linearly parametrized then the predictor (2.19) (and hence (2.18)) is linear in θ , giving that the solution of (2.22) can be analytically computed (see Sect. 2.5.1.1). A particularly interesting feature of this structure in

the LPV case is that it explicitly assumes that v is correlated with p through the same nonlinearities as the noise-free output signal. Hence, it generally applies a more restrictive assumption on the noise than in the LTI case. This type of model structure is most commonly applied in LPV-ID approaches that are based on least-squares estimates (e.g., [1, 9, 35]).

An extension of this model structure can be achieved by introducing a *moving average* part on e , i.e., when $D(q^{-1}, p(k), \theta_h) \neq 1$. This significantly reduces the restrictiveness of the modeling assumption on the noise, but the price to be paid is that the predictor (2.19) is not linear in θ any more as

$$(W_u(q, \theta) \diamond p)(k) = \sum_{i=0}^{\infty} (1 - D(q^{-1}, p(k), \theta_h))^i B(q^{-1}, p(k), \theta_g) q^{-\tau_d}, \quad (2.32a)$$

$$(W_y(q, \theta) \diamond p)(k) = 1 - \sum_{i=0}^{\infty} (1 - D(q^{-1}, p(k), \theta_h))^i A(q^{-1}, p(k), \theta_g). \quad (2.32b)$$

Note that here the infinite sum term results due to the inversion of the scheduling dependent noise model. This term is convergent as H^\dagger is defined to be stable. Furthermore, due to noncommutativity of q in this setting, the polynomial terms do not commute, e.g., (2.32a) is not equal to $B(q^{-1}, p(k), \theta_g) q^{-\tau_d} \sum_{i=0}^{\infty} (1 - D(q^{-1}, p(k), \theta_h))^i$.

2.4.3.2 LPV-FIR and Series Expansion Structures

Considering the IRR form (2.6) of the process part of \mathcal{S}_o , a particularly interesting idea is to truncate this series expansion to get an approximation of the original system (2.5) in the following form:

$$y(k) \approx \sum_{i=0}^{n_f} (g_i^o \diamond p)(k) u(k-i) + v_o(k). \quad (2.33)$$

Note that if n_f is large enough, the approximation error is negligible in (2.33). This gives that by introducing a polynomial model structure⁴:

$$y(k) = \sum_{i=0}^{n_f} g_i(p(k), \theta_g) u(k-i) + v(k) = F(q^{-1}, p(k), \theta_g) u(k) + v(k), \quad (2.34)$$

where each coefficient g_i of the polynomial F is linearly parameterized in terms of the functions $\{\alpha_{i,l}\}_{i=0, l=1}^{n_f, n_\alpha}$ and parameters θ_g , a rather simple but effective approximation of the original system can be achieved. This model structure is

⁴It is more natural to use dynamic dependence in the parametrization of the coefficients in (2.34), but for the sake of simplicity we use only static dependence here.

the LPV form of the well-known LTI *finite impulse response* (FIR) models and has the attractive property of being *linear-in-the-coefficients* if $v(k)$ is a white noise, as in this case, its associated one-step-ahead predictor is $\hat{y}(k | \theta, k-1) = F(q^{-1}, p(k), \theta_g)u(k)$. Note that in the latter case, which corresponds to an *output-error* (OE) noise model, this model structure can also be derived from (2.29a–c) by setting A, B, C to 1 and $B(q^{-1}, p(k), \theta_g) \triangleq F(q^{-1}, p(k), \theta_g)$.

FIR models have many advantages in identification due to their simple structure. However, their well-known disadvantage, both in the LTI and the LPV cases, is that the expansion may have a slow convergence rate, meaning that it requires a relatively large number of parameters for an adequate approximation of the system. In order to benefit from the same properties, but achieve faster convergence rate of the expansion, it is attractive to use basis functions which, opposite to q^{-i} , have infinite impulse responses. A particular choice of such a basis follows through the use of *orthonormal basis functions* (OBFs), which are specific basis functions in \mathcal{H}_2 (*Hardy space* of square integrable complex functions) and have already proven their usefulness in LTI identification (see [10]). Based on this idea, it is possible to show that any asymptotically stable LPV system has a series expansion in terms of an OBF basis set $\{\phi_i(z)\}_{i=1}^{\infty} \subset \mathcal{H}_2$ (see [23]), via the expansion of each q^{-i} in (2.3) in terms of $\{\phi_i(q)\}_{i=1}^{\infty}$. Thus, (2.5) can be written as

$$y(k) = (w_0^o \diamond p)(k)u(k) + \sum_{i=1}^{\infty} (w_i^o \diamond p)(k)\phi_i(q)u + v_o(k), \quad (2.35)$$

where w_i^o are functions with dynamic dependence on p . An obvious choice of model structure is to use a truncated expansion, i.e., truncating (2.35) to a finite sum in terms of $\{\phi_i\}_{i=1}^{n_w}$, and introducing a parametrization of the expansion coefficients:

$$y(k) = w_0(p(k), \theta_g)u(k) + \sum_{i=1}^{n_w} w_i(p(k), \theta_g)\phi_i(q)u + v(k), \quad (2.36)$$

where each w_i is parametrized as in the FIR case using functions with static dependence on p . Similar to the FIR, this structure is linear in the coefficients $\{w_i\}_{i=1}^{n_w}$ if $v(k)$ is assumed to be white, but with $n_w \ll n_f$ for the same approximation capability. Furthermore, it is proven that structures like (2.36), i.e., a OBF filter bank followed by a static nonlinearity are general approximators of nonlinear systems with fading memory, i.e., nonlinear dynamic systems with convolution representation [2]. Additionally, OBF and FIR models have a direct and simple SS and LFR realization, a somewhat unique property among polynomial IO structures [23]. An important question that arises is whether the basis functions ϕ_i can be chosen such that a fast convergence rate can be achieved for all possible trajectories of p , i.e., how $\{\phi_i(q)\}_{i=1}^{n_w}$ with minimal n_w should be chosen such that the approximation error is adequate for the problem at hand. For this purpose the Kolmogorov n -width theory gives an effective way to characterize optimal

convergence rate as an optimization problem in terms of the poles of the basis functions $\{\phi_i\}_{i=1}^{N_w}$ [28]. Using this concept, data-driven basis selection methods can be introduced to optimize the basis set w.r.t. to the dynamical behavior of the system at hand, see [23, 28] for further details.

2.4.3.3 LPV–BJ and OE Structures

Independent parametrization of \mathcal{G}_{θ_g} and \mathcal{H}_{θ_h} with the complete structure of (2.29a–c) corresponds to an LPV *Box–Jenkins* (BJ) model structure, which allows to describe a wide range of noise dynamics in a general sense. Of course this generality has a heavy price in terms of a complicated one-step-ahead predictor characterized by

$$\begin{aligned} (W_u(q, \theta) \diamond p)(k) &= \sum_{j=0}^{\infty} (1 - D(q^{-1}, p(k), \theta_h))^j C(q^{-1}, p(k), \theta_h) \\ &\quad \times \sum_{i=0}^{\infty} (1 - A(q^{-1}, p(k), \theta_g))^i B(q^{-1}, p(k), \theta_g) q^{-\tau_d}, \end{aligned} \quad (2.37a)$$

$$(W_y(q, \theta) \diamond p)(k) = 1 - \sum_{j=0}^{\infty} (1 - D(q^{-1}, p(k), \theta_h))^j C(q^{-1}, p(k), \theta_h). \quad (2.37b)$$

A simplification of this structure in terms of $C(q^{-1}, p(k), \theta_h) \triangleq D(q^{-1}, p(k), \theta_h) \triangleq 1$ leads to the so-called *output-error* (OE) type of model structure with

$$(W_u(q, \theta) \diamond p)(k) = \sum_{i=0}^{\infty} (1 - A(q^{-1}, p(k), \theta_g))^i B(q^{-1}, p(k), \theta_g) q^{-\tau_d}, \quad (2.38a)$$

$$(W_y(q, \theta) \diamond p)(k) = 1. \quad (2.38b)$$

Regarding LPV–BJ models an extra distinction can be introduced w.r.t. p -independent noise models. For instance, assuming $H(q, \theta_h)$ to be a rational LTI transfer function leads to a simplified LPV–BJ model for which a refined instrumental variable estimation approach has been developed recently (see [13]).

2.4.4 Informativity and Identifiability

Regarding the introduced model structures, we can investigate conditions under which informativity of a given data set and identifiability of a particular model structure itself can be assured. As these are the basic ingredients for a successful identification of the system, it is important to review the available results even if they are rather sparse.

For the case of LPV–ARX models with polynomial dependence of the coefficients on the parameters, conditions for identifiability and informativity have been studied in [6, 9, 36]. Based on these results, the following theorem holds true.

Theorem 2.1 (Identifiability, LPV–ARX case). *The LPV–ARX model structure (2.30) with linear parametrization (2.27) is globally identifiable, if and only if each set of functions $\{\alpha_{i,l}(\cdot)\}_{l=1}^{n_\alpha}$ and $\{\beta_{j,l}(\cdot)\}_{l=1}^{n_\beta}$ for all $i \in \{1, \dots, n_a\}$ and $j \in \{0, \dots, n_b\}$, used in the parametrization (2.27), are linearly independent on \mathbb{P} .*

For a detailed proof see [6]. Identifiability conditions for other model structures require linear independence (necessary condition) of the functions used in the parametrization of each p -dependent coefficient and also co-primeness of certain pairs of polynomials just like in the LTI case, see [8]. However, the sufficient conditions to guarantee identifiability in these cases have not been established yet.

To establish a condition on informativity in the ARX case, define

$$\Delta W_y(q) = W_y(q, \theta_1) - W_y(q, \theta_2), \quad \Delta W_u(q) = W_u(q, \theta_1) - W_u(q, \theta_2).$$

Then it follows that (2.24) equals to

$$\bar{\mathbb{E}} \{ ([\Delta W_u(q) - \Delta W_y(q)G_o(q)] \diamond p)u + ([\Delta W_y(q)H_o(q)] \diamond p)e \}^2 = 0. \quad (2.39)$$

Straightforward application of \mathbb{E} in $\bar{\mathbb{E}}$ gives that (2.39) is equivalent with

$$\bar{\mathbb{E}} \{ ([\Delta W_u(q) - \Delta W_y(q)G_o(q)] \diamond p)u \}^2 = 0, \quad (2.40a)$$

$$\bar{\mathbb{E}} \{ ([\Delta W_y(q)H_o(q)] \diamond p)e \}^2 = 0. \quad (2.40b)$$

Now we can seek for conditions on u and p for which the above conditions imply that $\Delta W_u(q) = \Delta W_y(q) = 0$. As p is independent of e , (2.40b) holds if and only if $\bar{\mathbb{E}}\{\Delta W_y(q) \diamond p\} = 0$ whenever $e \neq 0$. However, $\bar{\mathbb{E}}\{\Delta W_y(q) \diamond p\} = 0$ does not necessarily imply $\Delta W_y(q) = A(q^{-1}, p(k), \theta_1) - A(q^{-1}, p(k), \theta_2) = 0$. In case of global identifiability of \mathcal{M}_θ , the necessary and sufficient condition to guarantee this is that the data matrix $\Phi_y = [\varphi_y(1) \cdots \varphi_y(N)]^\top$ satisfies that $\bar{\mathbb{E}}\{\Phi_y^\top \Phi_y\} \succ 0$ where

$$\varphi_y(k) = [\alpha_{1,1}(p(k)) \cdots \alpha_{n_a, n_\alpha}(p(k))]^\top.$$

Next we need to find necessary and sufficient conditions on u and p such that

$$\bar{\mathbb{E}}\{(\Delta W_u(q) \diamond p)u\}^2 = 0 \quad \Rightarrow \quad \Delta W_u(q) = 0. \quad (2.41)$$

In case of an LPV–ARX model:

$$\Delta W_u(q) = B(q^{-1}, p(k), \theta_1)q^{-\tau_d} - B(q^{-1}, p(k), \theta_2)q^{-\tau_d}. \quad (2.42)$$

Assume that the LPV-ARX model is globally identifiable. Now in order to guarantee that (2.41) holds with (2.42), a necessary and sufficient condition is that $\Phi_u = [\varphi_u(1) \cdots \varphi_u(N)]^\top$ satisfies $\mathbb{E}\{\Phi_u^\top \Phi_u\} \succ 0$ where

$$\varphi_u(k) = \left[u(k - \tau_d) \quad \beta_{0,1}(p(k))u(k - \tau_d) \quad \cdots \quad \beta_{n_b, n_\beta}(p(k))u(k - \tau_d - n_b) \right]^\top.$$

Theorem 2.2 (Informative dataset, LPV-ARX case). *Given a globally identifiable LPV-ARX model structure (2.30), denoted by \mathcal{M}_θ , with linear parametrization (2.27), then a quasi-stationary data set \mathcal{D}_N generated by $\mathcal{S}_0 \in \mathcal{M}$ is informative w.r.t. \mathcal{M}_θ , if for $\Phi_u = [\varphi_u(1) \cdots \varphi_u(N)]^\top$ and $\Phi_y = [\varphi(1) \cdots \varphi(N)]^\top$ it holds that $\mathbb{E}\{\Phi_u^\top \Phi_u\} \succ 0$ and $\mathbb{E}\{\Phi_y^\top \Phi_y\} \succ 0$.*

Note that in case of a given data set \mathcal{D}_N with finite N , the above conditions translate to the existence of a set of time instances $\mathcal{K}_y, \mathcal{K}_u \subseteq \{1, \dots, N\}$ with $\text{Card}(\mathcal{K}_y) = n_a n_\alpha = n_y$ and $\text{Card}(\mathcal{K}_u) = (n_b + 1)(n_\beta + 1) = n_u$ such that $\Phi_y = [\varphi_y(k_1) \cdots \varphi_y(k_{n_y})]^\top$ with $\{k_1, \dots, k_{n_y}\} = \mathcal{K}_y$ and $\Phi_u = [\varphi_u(k_1) \cdots \varphi_u(k_{n_u})]^\top$ with $\{k_1, \dots, k_{n_u}\} = \mathcal{K}_u$ satisfy that $\Phi_u^\top \Phi_u \succ 0$ and $\Phi_y^\top \Phi_y \succ 0$. Note that these conditions are the minimal conditions which are required to be satisfied by the data set \mathcal{D}_N to guarantee informativity. It is also possible to formulate these conditions w.r.t. typical scheduling trajectories to derive easily understandable design rules. This has been investigated in [6] where the above conditions are interpreted in case of piecewise and periodic scheduling trajectories. Informativity conditions for other type of model structures are in the focus of current research.

2.5 Identification with the LS Criterion

Using the LS criterion (2.21), in this section we will investigate the estimation of the LPV model structures introduced in Sect. 2.4 with linear parametrization. According to the available approaches in the literature, identification of (2.29a-c) can be investigated from two perspectives: *local* setting (identification for constant p and interpolation) and the *global* setting (identification with varying p). Here we will concentrate on global approaches as only this setting allows to address the minimization of the prediction error e_θ which is our aim with the introduced PE framework. We will see that in the considered global PE setting, the predictor (2.19), w.r.t. each of the introduced model structures, can be rewritten as a linear or a pseudolinear regression. This yields that estimation of these structures is available via a (iterative) least-squares estimate. Alternatively, other nonlinear optimization schemes can also be applied in the absence of the linear-in-the-coefficients property of (2.19). Furthermore, we will explore identification with *instrumental variables* (IV) in this context as well, showing why such approaches can be rather advantageous in the LPV case. In addition, consistency and variance properties of the estimated parameters will be also investigated.

2.5.1 Prediction-Error Based Identification in the ARX Case

In terms of the considered global setting, we aim at the direct minimization of (2.21) in terms of the parametrized model structure (2.29a–c) using a dataset \mathcal{D}_N where p is varying. This data set is assumed to be informative w.r.t. (2.29a–c) to have a well-posed problem for identification. To fulfill our estimation objective, several approaches can be introduced for the various model structures given in Sect. 2.4. For the sake of clarity, we will study these estimation approaches step-by-step, starting from the most simplest case of ARX models where the estimation can be addressed via simple linear regression.

2.5.1.1 Linear Regression

Consider the LPV–ARX model structure (2.30). A particular property of this structure with the linear parametrization (2.27) of A and B is that the predictor (2.19) is linear in the parameters $\theta = \theta_g$, see (2.31), and hence can be written as

$$\hat{y}(k | \theta, k-1) = \varphi^\top(k) \theta, \quad (2.43)$$

where

$$\begin{aligned} \varphi(k) = & \left[-y(k-1) - \alpha_{1,1}(p(k))y(k-1) \quad \cdots \quad -\alpha_{n_a, n_a}(p(k))y(k-n_a) \right. \\ & \left. u(k-\tau_d) \quad \beta_{0,1}(p(k))u(k-\tau_d) \quad \cdots \quad \beta_{n_b, n_b}(p(k))u(k-\tau_d-n_b) \right]^\top. \end{aligned} \quad (2.44)$$

As (2.43) is a linear regression equation, thus by defining $\Phi = [\varphi(1) \cdots \varphi(N)]^\top$ and $Y = [y(0) \cdots y(N)]^\top$, the minimum of (2.21) is unique and equal to

$$\hat{\theta}_{LS} = \Phi^+ Y, \quad (2.45)$$

if $\text{rank}(\Phi) = n_\theta$, where $\Phi^+ = \left(\frac{1}{N} \Phi^\top \Phi\right)^{-1} \frac{1}{N} \Phi^\top$ is the regularized *Moore–Penrose pseudoinverse*. This approach is summarized in terms of Algorithm 1. Equation (2.45) has been used in many works, e.g., [1], to estimate LPV–IO models, however, in the introduced PE framework it is justified that (2.45) is the minimizer of (2.21) in case of an LPV–ARX model structure. It is also important to mention that (2.43) can be also considered as an LTI *multiple-output multiple-input* (MIMO) ARX model with “virtual” input and output signals $\{\beta_{j,l}(p)u\}$ and $\{\alpha_{i,l}(p)y\}$.

To get an insight of the stochastic behavior of the LS estimator (2.45), assume that $\mathcal{S}_0 \in \mathcal{G}$ and consider the “optimal” residual error e_{θ_0} , which based on (2.43) is

$$e_{\theta_0}(k) = y(k) - \hat{y}(k | \theta_0, k-1) = y(k) - \varphi^\top(k) \theta_0. \quad (2.46)$$

Algorithm 1: LPV-ARX identification, LS global method

-
- Require:** a data record $\mathcal{D}_N = \{u(k), y(k), p(k)\}_{k=1}^N$ of \mathcal{S}_0 , the LS identification criterion V , and the LPV-ARX model structure (2.30) with linear parametrization (2.27) and parameters $\theta = [a_{1,0} \cdots a_{n_a, n_a} \ b_{0,0} \cdots b_{n_b, n_b}] \in \mathbb{R}^{n_\theta}$. Assume that \mathcal{D}_N is informative w.r.t. (2.30) and (2.30) is globally identifiable.
- 1: calculate the signals $x_{i,l} = -\alpha_{i,l}(p)q^{-i}y$ and $x_{j+n_a+1,l} = \beta_{j,l}(p)q^{-j-\tau_d}u$ and let $\varphi = [x_{0,0} \cdots x_{n_a+n_b+1, n_b}]^\top$ giving that $y(k) = \varphi^\top(k)\theta + e_\theta(k)$.
 - 2: estimate θ in terms of $\hat{\theta} = \arg \min_{\theta} V(\theta, \mathcal{D}_N)$. In case of (2.21), $\hat{\theta} = \Phi^+ Y$ with $Y = [y^\top(1) \cdots y^\top(N)]^\top$ and $\Phi = [\varphi(1) \cdots \varphi(N)]^\top$.
 - 3: **return** estimated model (2.30).
-

As a consequence, (2.45) satisfies that

$$\hat{\theta}_{\text{LS}} = \theta_0 + \left(\sum_{k=1}^N \varphi(k) \varphi^\top(k) \right)^{-1} \cdot \sum_{k=1}^N \varphi(k) e_{\theta_0}(k). \quad (2.47)$$

Equation (2.47) yields that $\hat{\theta}_{\text{LS}}$ is a consistent estimate of θ_0 (unbiased for finite data) if the following conditions are respected:

$$\bar{\mathbb{E}} \left\{ \varphi(k) \varphi^\top(k) \right\} \succ 0 \quad \text{and} \quad \bar{\mathbb{E}} \left\{ \varphi(k) e_{\theta_0}(k) \right\} = 0. \quad (2.48)$$

While the first condition is satisfied in case \mathcal{D}_N is informative w.r.t. (2.30), the second condition only holds if e_{θ_0} is a white noise. Unfortunately, this is only true when the data-generating system \mathcal{S}_0 itself has an ARX noise structure. In that case, based on the classical proofs, it is possible to show that if the model set \mathcal{M} is uniformly asymptotically stable w.r.t. a compact Θ and it is globally identifiable, then under the assumption that the data set \mathcal{D}_N is informative and quasi-stationary, the estimates $\hat{\theta} \rightarrow \hat{\theta}_* = \arg \min_{\theta \in \Theta} \bar{\mathbb{E}} \{ \|e_\theta\|_{\ell_2}^2 \}$ with probability 1 where θ_* is unique [25]. Furthermore, if $\theta_0 = \theta_{0,g}$ associated with G_0 satisfies that $\theta_0 \in \Theta$, then $\theta_* = \theta_0$. This proves consistency and convergence of the estimation. In case of noisy observation of p , convergence of the LS estimates can be shown, but in general, consistency does not hold (see [3] for a motivation). It has been recently shown that this problem can be circumvented by using a regressor with estimated moment functions associated with p_* , see [25] for further details.

In practice, the ARX modeling assumption often appears to be rather restrictive. Even though it might be a fair assumption to consider that the process is well parametrized by (2.29a), the noise model $A(q^{-1}, p(k), \theta_g)v(k) = e(k)$ is often not rich enough to capture v_0 . Indeed, in most cases, there is no reasonable explanation to justify why the noise v_0 and the process part of \mathcal{S}_0 should contain the same dynamics and nonlinearities. In terms of estimation, this means that using the LS method in practice will most often lead to biased estimates. Consequently, some methods have been developed in order to cope with the error induced by this invalid assumption on the noise. A method proposed in [3] and relying on an instrumental variable approach is described in the next section.

2.5.1.2 Instrumental Variable Approach

The original aim of instrumental variable methods is to cope with the fact that in most cases, e_{θ_0} is a colored process. The idea is to introduce an instrument $\zeta : \mathbb{Z} \rightarrow \mathbb{R}^{n_\theta}$ which is used to produce a consistent estimate independently on the noise model taken. The IV estimate is given as

$$\hat{\theta}_{\text{IV}} = \left(\sum_{k=1}^N \zeta(k) \varphi^\top(k) \right)^{-1} \cdot \sum_{k=1}^N \zeta(k) y(k), \quad (2.49)$$

which implies that

$$\hat{\theta}_{\text{IV}} = \theta_0 + \left(\sum_{k=1}^N \zeta(k) \varphi^\top(k) \right)^{-1} \cdot \sum_{k=1}^N \zeta(k) e_{\theta_0}(k). \quad (2.50)$$

Therefore, and similarly to the LS solution, $\hat{\theta}_{\text{IV}}$ is a consistent estimate of θ_0 if

$$\mathbb{E} \left\{ \zeta(k) \varphi^\top(k) \right\} \succ 0 \quad \text{and} \quad \mathbb{E} \left\{ \zeta(k) e_{\theta_0}(k) \right\} = 0. \quad (2.51)$$

There is a considerable amount of freedom in the choice of an instrument respecting these conditions. In the LTI context, the choice of the instrument has been widely studied and most of the advanced IV methods offer similar performance as extended LS methods or other PE minimization methods (see [17, 20]). A particularly interesting fact is that, under the ARX model assumption, the variance of the IV estimate is minimal if the instrument is chosen as the noise-free version of the regressor [21]. In other words, when directly applying the IV theory to the LPV-ARX model (2.30) (the LPV-ARX model can be seen as an LTI model), the optimal IV estimate is given by

$$\hat{\theta}_{\text{IV}}^{\text{opt}} = \left(\sum_{k=1}^N \zeta_{\text{opt}}(k) \varphi^\top(k) \right)^{-1} \cdot \sum_{k=1}^N \zeta_{\text{opt}}(k) y(k), \quad (2.52)$$

where the optimal instrument is defined as:

$$\zeta_{\text{opt}}(k) = \left[-\check{y}_o(k-1) \quad -\alpha_{1,1}(p(k))\check{y}_o(k-1) \quad \cdots \quad -\alpha_{n_a, n_a}(p(k))\check{y}_o(k-n_a) \right. \\ \left. u(k-\tau_d) \quad \beta_{0,1}(p(k))u(k-\tau_d) \quad \cdots \quad \beta_{n_b, n_b}(p(k))u(k-\tau_d-n_b) \right]. \quad (2.53)$$

Here \check{y}_o denotes the noise-free output of the data-generating system \mathcal{S}_o which is a priori unknown in practice. Consequently, often an estimate of \check{y}_o is applied as an instrument, like the simulated output of a previously obtained model estimate which in turn can be refined iteratively. Note that if $\mathcal{S}_o \in \mathcal{M}$, then both the IV solution

(2.52) and the LS solution (2.47) are consistent and statistically optimal (minimum variance and unbiased). Furthermore, consistency of (2.52) also holds in the p -noisy case, see [3].

To construct the optimal instrument ζ_{opt} , a particular approach is the IV4 method [16], proposed in the LPV case in [3]. In this approach, the instrument is built using the simulated data generated from an estimated auxiliary ARX model. This method is given in detail by Algorithm 2. In [3], it was shown that in case

Algorithm 2: LPV–ARX identification, IV4 global method

Require: a data record $\mathcal{D}_N = \{u(k), y(k), p(k)\}_{k=1}^N$ of \mathcal{S}_o , the LS identification criterion V , and the LPV–ARX model structure (2.30) with linear parametrization (2.27) and parameters $\theta = [a_{1,0} \cdots a_{n_a, n_a} \ b_{0,0} \cdots b_{n_b, n_b}] \in \mathbb{R}^{n_\theta}$. Assume that \mathcal{D}_N is informative w.r.t. (2.30) and (2.30) is globally identifiable on Θ .

- 1: estimate an LPV–ARX model by Algorithm 1.
 - 2: generate an estimate $\hat{y}_o(k)$ of $\check{y}_o(k)$ based on the resulting ARX model of the previous step.
 - 3: build an instrument based on $\hat{y}_o(k)$ and then estimate θ via (2.49).
 - 4: **return** estimated model (2.30).
-

\mathcal{S}_o corresponds to an LPV–OE model ($v_o = e_o$), Algorithm 2 leads to an unbiased estimate. Moreover, like in the LTI case, any structural modeling error of the noise results in a bias for the LS estimate while, using this IV method, only the variance of $\hat{\theta}_{\text{IV}}$ is increased. Nevertheless, the bigger the difference between the true noise process and the noise model assumed is, the higher the resulting variance in the IV estimate is. Depending on the size N of the dataset, the variance increase of the IV estimate can lead to worse results than by the LS estimator (for which the variance is known to remain low). Consequently, it is highly important to assume a noise model as realistic as possible in the first place. In the LTI case, many IV methods are dedicated to more general noise models such as OE or BJ [39]. In Sect. 2.5.2.3, we consider some available methods for LPV–OE and LPV–BJ model structures which were introduced in [13].

2.5.1.3 Estimation of FIR and OBF Models

Even if LPV–FIR and OBF models have different representation capabilities than ARX models, if the noise v_o is assumed to be zero mean and white, the one-step-ahead predictor can be written as a liner regression similar to (2.43) where

$$\varphi(k) = [u(k) \ \alpha_{0,1}(p(k))u(k) \ \cdots \ \alpha_{n_f, n_\alpha}(p(k))u(k - n_f)]^\top. \quad (2.54)$$

Due to this fact, the LS approach can be used to estimate such models and as the regressor does not contain any output terms, the LS estimate is consistent even if

the noise v_0 is not white or p is polluted by noise, till these noise effects are not correlated with u . However, in case of OBF models, the selection of the basis set $\{\phi_i(q)\}_{i=1}^{n_w}$ has a paramount importance as it governs the size of parametrization and the approximation error directly. To provide a computationally attractive data-driven selection of efficient sets of basis functions, a *Fuzzy clustering* based approach has been proposed in [28]. For more on LPV–OBF models and issues of basis selection and identification, see [23, 28].

2.5.2 Prediction-Error Based Identification with General Noise Models

Next we consider the global identification of model structures which utilize extended noise models to increase the validity of the noise modeling. Unfortunately, due to the more complicated noise models, the linear-in-the-coefficient property is lost in these cases, causing (2.22) to be a nonlinear optimization problem whose solution is more complicated than in the previous case. First a particular idea of rewriting the one-step-ahead predictor as a pseudolinear regression is extended from the LTI framework to the LPV case, allowing to derive a computationally attractive solution. Then we also consider general nonlinear optimization to solve the estimation problem. Finally, it is shown how the IV approach can offer a simple solution for the estimation of θ in case of a more general noise model than in Sect. 2.5.1.2.

2.5.2.1 Pseudolinear Regression

Consider first the LPV–ARMAX case, where A, B, D are (linearly) parametrized as in (2.27). By multiplying (2.19) with $D(q^{-1}, p(k), \theta_h)$ on the left, it follows that $D(q^{-1}, p(k), \theta_h) \sum_{j=0}^{\infty} (1 - D(q^{-1}, p(k), \theta_h))^j = 1$ and hence

$$\begin{aligned} D(q^{-1}, p(k), \theta_h) \hat{y}(k | \theta, k-1) &= B(q^{-1}, p(k), \theta_g) u(k - \tau_d) \\ &\quad + (D(q^{-1}, p(k), \theta_h) - A(q^{-1}, p(k), \theta_g)) y(k) \end{aligned} \quad (2.55)$$

in terms of the subpredictors (2.32a–b). By adding $(1 - D(q^{-1}, p(k), \theta_h)) \hat{y}(k | \theta)$ to both sides of (2.55) and using (2.18), (2.19) is equivalent with

$$\begin{aligned} \hat{y}(k | \theta, k-1) &= B(q^{-1}, p(k), \theta_g) u(k - \tau_d) + (1 - A(q^{-1}, p(k), \theta_g)) y(k) \\ &\quad + (D(q^{-1}, p(k), \theta_h) - 1) e_{\theta}(k). \end{aligned} \quad (2.56)$$

Then by considering a regressor $\varphi^\top(k | \theta)$ defined as before (see (2.44)), but extended with $\delta_{1,1}(p(k))e_\theta(k-1), \dots, \delta_{n_d, n_\delta}(p(k))e_\theta(k-n_d)$, the predictor (2.56) can be rewritten as

$$\hat{y}(k | \theta, k-1) = \varphi^\top(k | \theta)\theta. \quad (2.57)$$

This equation corresponds to a *pseudolinear regression*, hence minimization of (2.21) follows by an iterative LS approach where an estimate of $e_\theta(k)$ is generated by a model obtained in a previous iteration, see Algorithm 3.

Algorithm 3: LPV-ARMAX identification, iterative LS global method

Require: a data record $\mathcal{D}_N = \{u(k), y(k), p(k)\}_{k=1}^N$ of \mathcal{S}_0 , the LS identification criterion V , and the LPV-ARMAX model structure (2.29a-c) with $C(q^{-1}, p(k), \theta_h) \triangleq A(q^{-1}, p(k), \theta_g)$ and linear parametrization (2.27) with parameters $\theta = [a_{1,0} \cdots b_{n_b, n_\beta} \ d_{1,0} \cdots d_{n_d, n_\delta}] \in \mathbb{R}^{n_\theta}$. Assume that \mathcal{D}_N is informative w.r.t. (2.29a-c) and (2.29a-c) is globally identifiable on Θ .

1: estimate an ARX model by Algorithm 1 resulting in $\hat{\theta}_g^{(0)}$. Set $D(q^{-1}, p(k), \theta_h^{(0)}) = 1$,

$$\hat{\theta}^{(0)} = \left[\left(\hat{\theta}_g^{(0)} \right)^\top \left(\hat{\theta}_h^{(0)} \right)^\top \right]^\top \text{ and } \tau = 0.$$

2: **repeat**

3: generate an estimate $\hat{e}_\theta^{(\tau)}(k)$ based on the resulting model of the previous step, i.e., $\hat{\theta}^{(\tau)}$.

4: calculate the signals $x_{i,l} = -\alpha_{i,l}(p)q^{-i}y$, $x_{j+n_a+1,l} = \beta_{j,l}(p)q^{-j-\tau_d}u$ and

$$x_{i+n_a+n_b+1,l} = -\delta_{i,l}(p)q^{-i}\hat{e}_\theta^{(\tau)} \text{ and let } \varphi_\tau = [x_{0,0} \cdots x_{n_a+n_b+n_d+1, n_\delta}]^\top.$$

5: estimate θ in terms of $\hat{\theta}^{(\tau+1)} = \Phi_\tau^\dagger Y$ with $Y = [y^\top(1) \cdots y^\top(N)]^\top$ and

$$\Phi_\tau = [\varphi_\tau(1) \cdots \varphi_\tau(N)]^\top. \text{ Increase } \tau \text{ by } 1.$$

6: **until** $\hat{\theta}_\tau$ has converged or the maximum number of iterations is reached.

7: **return** estimated model (2.29a-c).

Now consider the LPV-OE case. In this case, $\theta_h = \emptyset$ and (2.19) read as

$$\hat{y}(k | \theta, k-1) = \sum_{i=0}^{\infty} (1 - A(q^{-1}, p(k), \theta_g))^i B(q^{-1}, p(k), \theta_g) q^{-\tau_d} u(k). \quad (2.58)$$

Define $\check{y} = (G(q, \theta_g) \diamond p)u$ as the noise-free output of the LPV-OE model. Then, (2.58) can be rewritten as

$$\hat{y}(k | \theta, k-1) = (1 - A(q^{-1}, p(k), \theta_g))\check{y}(k) + B(q^{-1}, p(k), \theta_g) q^{-\tau_d} u(k). \quad (2.59)$$

This gives the idea again to introduce the regressor

$$\varphi^\top(k | \theta) = [-\check{y}(k-1) \quad -\alpha_{1,1}(p(k))\check{y}(k-1) \quad \cdots \quad -\alpha_{n_a, n_\alpha}(p(k))\check{y}(k-n_a) \\ u(k-\tau_d) \quad \beta_{0,1}(p(k))u(k-\tau_d) \quad \cdots \quad \beta_{n_b, n_\beta}(p(k))u(k-\tau_d-n_b)], \quad (2.60)$$

to write (2.59) in the form of (2.57). Again an iterative LS algorithm, similar to Algorithm (3), can be introduced to obtain an estimate.

A pseudolinear regression form can be obtained for the LPV–BJ case by combining the approaches of the ARMAX and OE cases. In the LPV–BJ case

$$\hat{y}(k | \theta, k-1) = (W_u(q, \theta) \diamond p)(k)u(k) + (W_y(q, \theta) \diamond p)(k)y(k), \quad (2.61)$$

where the subpredictors are defined by (2.37a–b). Again introduce $\check{y} = (G(q, \theta_g) \diamond p)u$ as the noise-free model output and multiply (2.61) with $D(q^{-1}, p(k), \theta_h)$ on the left giving:

$$\begin{aligned} D(q^{-1}, p(k), \theta_h) \hat{y}(k | \theta, k-1) &= C(q^{-1}, p(k), \theta_h) \check{y}(k) + D(q^{-1}, p(k), \theta_h) y(k) \\ &\quad - C(q^{-1}, p(k), \theta_h) y(k). \end{aligned} \quad (2.62)$$

As $e_\theta(k) = y(k) - \hat{y}(k | \theta, k-1)$, we can write

$$e_\theta(k) = (1 - D(q^{-1}, p(k), \theta_h))e_\theta(k) + C(q^{-1}, p(k), \theta_h) (y(k) - \check{y}(k)). \quad (2.63)$$

Note that C is monic. We also know that

$$\check{y} = (1 - A(q^{-1}, p(k), \theta_g))\check{y}(k) + B(q^{-1}, p(k), \theta_g) u(k). \quad (2.64)$$

Let $\xi = y - \check{y}$, then (2.63) reads as

$$e_\theta(k) = y(k) - \varphi^\top(k | \theta)\theta, \quad (2.65)$$

where

$$\begin{aligned} \varphi^\top(k) &= [-\check{y}(k-1) \quad -\alpha_{1,1}(p(k))\check{y}(k-1) \quad \cdots \quad -\alpha_{n_a, n_a}(p(k))\check{y}(k-n_a) \\ &\quad u(k-\tau_d) \quad \beta_{0,1}(p(k))u(k-\tau_d) \quad \cdots \quad \beta_{n_b, n_b}(p(k))u(k-\tau_d-n_b) \\ &\quad -\xi(k-1) \quad -\gamma_{1,1}(p(k))\xi(k-1) \quad \cdots \quad -\gamma_{n_c, n_c}(p(k))\xi(k-n_c) \\ &\quad e_\theta(k-1) \quad \delta_{1,1}(p(k))e_\theta(k-1) \quad \cdots \quad \delta_{n_d, n_d}(p(k))e_\theta(k-n_d)]. \end{aligned} \quad (2.66)$$

As $e_\theta(k) = y(k) - \hat{y}(k | \theta, k-1)$, (2.65), can be written as (2.57). Again an iterative LS algorithm, similar to Algorithm (3) can be introduced to obtain an estimate.

As we could see, despite noncommutativity of the time operator q , the rather complicated LPV model structures in the considered PE setting could have been relatively easily transformed to a pseudo linear regression form, hence their estimation is available by different iterative schemes. However, there is a particular difficulty to establish consistency and convergence results w.r.t. these schemes. Namely, the optimal regressor is required by the linear regression form to achieve such properties. However, as these items are approximated, we can analyze the estimation properties only in the case if the corresponding iterative approximations have converged. Such convergence properties, just like in the LTI case, are not fully

understood in general. In terms of application of these approaches, it has a prime importance that convergence is quite sensitive on the modeling assumption and the largeness of parametrization which are typically ill-chosen in most applications. Nevertheless, the introduced schemes, if they converge, provide computationally efficient estimation approaches in the considered context.

2.5.2.2 Nonlinear Optimization

Alternatively, minimization of (2.21) is available by general nonlinear optimization methods, like gradient-based minimization which can be applied directly computing the partial derivatives of the predictor (2.19) w.r.t. θ . Even the advanced LSQNONLIN approach of MATLAB can be directly used to obtain an estimate. As the application of these nonlinear optimization schemes only extends to the solution of the underlying optimization problem, these approaches are not presented in detail. However, there are two particular difficulties that can hinder the application of nonlinear optimization schemes:

1. In case of over-parametrization of the scheduling dependencies, the number of possible saddle points of (2.21) can seriously increase which can slow down or even prevent the convergence to the global optimum.
2. In case of large-scale systems, the computational time can be substantial compared to other approaches.

2.5.2.3 Instrumental Variable Approach

As the alternative of the previous estimation method, we can also introduce an instrumental variable approach that makes possible the direct identification of LPV–BJ models with p -independent noise part. Hence it improves considerably the achievable variance of the IV4 method in case of more complicated noise processes. To derive such an improved IV scheme, we first start with rewriting the process equation (2.29a) as

$$F(q^{-1}, \theta_g) \check{y}(k) + \sum_{i=1}^{n_a} \sum_{l=1}^{n_\alpha} a_{i,l} \tilde{y}_{i,l}(k) = \sum_{j=0}^{n_b} \sum_{l=0}^{n_\beta} b_{j,l} \tilde{u}_{j,l}(k), \quad (2.67)$$

where $F(q^{-1}, \theta_g) = 1 + \sum_{i=1}^{n_a} a_{i,0} q^{-i}$ is an LTI filter, $\tilde{y}_{i,l}(k) = \alpha_{i,l}(p(k)) \check{y}(k-i)$ and $\tilde{u}_{j,l}(k) = \beta_{j,l}(p(k)) u(k-j-\tau_d)$. Note that in this way the process part is rewritten as a MISO LTI model with $n_a n_\alpha + (n_b + 1)(n_\beta + 1)$ input signals $\{\tilde{y}_{i,l}\}_{i=1, l=1}^{n_a, n_\alpha}$ and $\{\tilde{u}_{j,l}\}_{j=0, l=0}^{n_b, n_\beta}$. However, this is not a representation of the original LPV behavior of (2.29a) as it contains lumped output terms. As a second step, assume that the noise part is not dependent on p , hence it is modeled as a stable LTI filter $H(q, \theta_h) = D(q^{-1}, \theta_h) / C(q^{-1}, \theta_h)$, which is a technical assumption we need to take to derive the intended approach. Given the fact that the polynomial operator F commutes in

(2.67) over the constant coefficients, thus (2.67) can be rewritten as

$$y(k) = - \sum_{i=1}^{n_a} \sum_{l=1}^{n_\beta} \frac{a_{i,l}}{F(q^{-1}, \theta_g)} \tilde{y}_{i,l}(k) + \sum_{j=0}^{n_b} \sum_{l=0}^{n_\beta} \frac{b_{j,l}}{F(q^{-1}, \theta_g)} \tilde{u}_{j,l}(k) + H(q, \theta_h) e(k), \quad (2.68)$$

which can be considered again as a “pseudolinear” form. This reformulation makes possible to introduce IV estimation of the considered LPV–BJ models. An approach that has been successfully applied in this context is an extended version of the *refined instrumental variable* (RIV) approach of the LTI identification framework [13].

As a refinement of the IV scheme presented in Sect. 2.5.1.2, IV methods have been developed to cope with more general noise structures such as the BJ case, where the data equation, under the assumption of $\mathcal{S}_o \in \mathcal{M}$, can be written as

$$y(k) = \varphi^\top(k) \theta_{o,g} + Q(q, \theta_o) e_{\theta_o}(k), \quad (2.69)$$

with $Q(q, \theta_o)$ being an LTI transfer function, $Q^{-1}(q, \theta_o)$ being stable, and $e_{\theta_o}(k)$ is a white noise. In the LPV–BJ case with p -independent noise part, this can be achieved by taking

$$\varphi(k) = \left[-y(k-1) \cdots -y(k-n_a) \quad -\tilde{y}_{1,1}(k) \cdots -\tilde{y}_{n_a, n_\alpha}(k) \quad \tilde{u}_{0,0}(k) \cdots \tilde{u}_{n_b, n_\beta}(k) \right]^\top$$

and

$$Q(q, \theta_o) = F(q^{-1}, \theta_{o,g}) \frac{D(q^{-1}, \theta_{o,h})}{C(q^{-1}, \theta_{o,h})}.$$

Based on this form, the extended-IV estimate in the ℓ_2 sense can be given as [21]:

$$\hat{\theta}_{g, \text{XIV}} = \arg \min_{\theta_g \in \mathbb{R}^{n_g}} \left\| \frac{1}{N} \sum_{k=1}^N L(q) \zeta(k) L(q) \varphi^\top(k) \theta_g - \frac{1}{N} \sum_{k=1}^N L(q) \zeta(k) L(q) y(k) \right\|_{\ell_2}^2, \quad (2.70)$$

where $\zeta(k)$ is the instrument and $L(q)$ is a stable prefilter. The conditions for consistency in this case read as:

$$\bar{\mathbb{E}} \left\{ L(q) \zeta(k) L(q) \varphi^\top(k) \right\} \succ 0 \quad \text{and} \quad \bar{\mathbb{E}} \{ L(q) \zeta(k) L(q) e_{\theta_o}(k) \} = 0. \quad (2.71)$$

Again, there is a considerable amount of freedom in the choice of the instruments. In [21, 38] it has been shown that the minimum variance estimator can be achieved in the BJ case if ζ is chosen as the noise-free version of the extended regressor φ and $L(q) = Q^{-1}(q, \theta_o)$. Furthermore, it holds true that in case of noise modeling error, the extended IV method is consistent and the variance of the estimates should

be significantly decreased with respect to the IV4 method: even if the noise process is not in the noise model set defined, it is more likely to be better described by an BJ model than by an ARX model.

In terms of the estimation, it is important to notice that φ contains the noise-free output terms $\{\tilde{y}_{i,l}\}$. Therefore, by momentarily assuming that $\{\tilde{y}_{i,l}(k)\}_{i=1,l=0}^{n_a,n_\alpha}$ are known a priori and that the data-generating system \mathcal{S}_o is in the model set, then the previously discussed conditions for optimal estimates lead to the choice of optimal instrument [13]:

$$\zeta_{\text{opt}}(k) = \begin{bmatrix} -\check{y}_o(k-1) & \cdots & -\check{y}_o(k-n_a) & -\tilde{y}_{1,1}^o(k) & \cdots & -\tilde{y}_{n_a,n_\alpha}^o(k) \\ \tilde{u}_{0,0}(k) & \cdots & \tilde{u}_{n_b,n_\beta}(k) \end{bmatrix}^\top \quad (2.72)$$

while the optimal filter is given as

$$L_{\text{opt}}(q) = Q^{-1}(q, \theta_o) = \frac{C(q^{-1}, \theta_{o,h})}{D(q^{-1}, \theta_{o,h})F(q^{-1}, \theta_{o,g})}. \quad (2.73)$$

In a practical situation, the optimal instrument (2.72) and filter (2.73) are unknown a priori. Therefore, the RIV estimation normally involves an iterative (or relaxation) algorithm in which, at each iteration, an “auxiliary model” is used to generate an estimate of (2.72) and (2.73). This auxiliary model is based on the parameter estimates obtained at the previous iteration. Consequently, if convergence occurs, the optimal instrument and filter are obtained. Based on the previous considerations, the RIV algorithm dedicated to the LPV case is summarized in Algorithm 4.

Using a similar concept, the so-called *simplified* RIV (SRIV) method can also be developed for the estimation of LPV–OE models. As in this case $C(q^{-1}, \theta_h) = D(q^{-1}, \theta_h) = 1$, Step 7 of Algorithm 4 can be skipped. In practical cases, it is a fair assumption to consider that the noise model assumed is incorrect for both LPV–OE and LPV–BJ models. In this case, the LPV–SRIV algorithm might perform as well as the LPV–RIV algorithm: the BJ assumption might be more realistic, but this is compensated by the reduced number of parameters to be estimated under the OE assumption. Additionally, both the RIV and SRIV algorithms can be also extended to be applicable in a closed-loop setting [29].

2.6 Conclusion

By using an impulse response representation of LPV systems, it has been shown in this chapter that a unified prediction-error framework for the identification of LPV polynomial models can be established. We have seen that this framework allows to understand the role of general noise models in the LPV setting, making possible to formulate the LPV extensions of classical model structures of the LTI case, like ARX, ARMAX, Box–Jenkins, OE, FIR, and series expansion models. Furthermore estimation of these models is computationally rather attractive and allows

Algorithm 4: LPV–BJ identification, RIV global method

Require: a data record $\mathcal{D}_N = \{u(k), y(k), p(k)\}_{k=1}^N$ of \mathcal{S}_0 , the LS identification criterion V , and the LPV–BJ model structure (2.29a–c) with linear parametrization (2.27), an LTI noise model characterized by $C(q^{-1}, \theta_h)$ and $D(q^{-1}, \theta_h)$ and with parameters $\theta_g = [a_{1,0} \cdots b_{n_b, n_\beta}]^\top \in \mathbb{R}^{n_g}$ and $\theta_h = [c_1 \cdots d_{n_d}]^\top \in \mathbb{R}^{n_h}$. Assume that \mathcal{D}_N is informative w.r.t. (2.29a–c) and (2.29a–c) is globally identifiable on Θ .

1: estimate an ARX model by Algorithm 1 resulting in $\hat{\theta}_g^{(0)}$. Set

$$C(q^{-1}, \hat{\theta}_h^{(0)}) = D(q^{-1}, \hat{\theta}_h^{(0)}) = 1, \hat{\theta}^{(0)} = \left[\left(\hat{\theta}_g^{(0)} \right)^\top \left(\hat{\theta}_h^{(0)} \right)^\top \right]^\top \text{ and } \tau = 0.$$

2: **repeat**

- 3: compute an estimate of $\check{y}(k)$ via $A(p_k, q^{-1}, \hat{\theta}_g^{(\tau)})\check{y}(k) = B(p_k, q^{-1}, \hat{\theta}_g^{(\tau)})u(k)$ and generate the auxiliary signals $\{\check{y}_{i,l}\}_{i=1, l=1}^{n_a, n_\alpha}$ and $\{\check{u}_{j,l}\}_{j=0, l=0}^{n_b, n_\beta}$ according to (2.67).
 4: compute the filter:

$$L(q^{-1}, \hat{\theta}^{(\tau)}) = \frac{C(q^{-1}, \hat{\theta}_h^{(\tau)})}{D(q^{-1}, \hat{\theta}_h^{(\tau)})F(q^{-1}, \hat{\theta}_g^{(\tau)})}$$

and the associated filtered signals $\{\check{u}_{j,l}^f(k)\}_{j=0, l=0}^{n_b, n_\beta}$, $y_f(k)$ and $\{\check{y}_{i,l}^f(k)\}_{i=1, l=0}^{n_a, n_\alpha}$.

5: build the filtered estimated regressor $\hat{\phi}_f^{(\tau)}(k)$ and the filtered instrument $\hat{\zeta}_f^{(\tau)}(k)$ as:

$$\hat{\phi}_f^{(\tau)}(k) = \left[-y_f(k-1) \cdots -y_f(k-n_a) \quad -\check{y}_{1,1}^f(k) \cdots -\check{y}_{n_a, n_\alpha}^f(k) \quad \check{u}_{0,0}^f(k) \cdots \check{u}_{n_b, n_\beta}^f(k) \right]^\top$$

$$\hat{\zeta}_f^{(\tau)}(k) = \left[-\check{y}_f(k-1) \cdots -\check{y}_f(k-n_a) \quad -\check{y}_{1,1}^f(k) \cdots -\check{y}_{n_a, n_\alpha}^f(k) \quad \check{u}_{0,0}^f(k) \cdots \check{u}_{n_b, n_\beta}^f(k) \right]^\top$$

6: calculate $\hat{\theta}_g^{(\tau+1)} = \left(\frac{1}{N} \Gamma_\tau^\top \Phi_\tau \right)^{-1} \left(\frac{1}{N} \Gamma_\tau^\top Y \right)$ with $Y = [y(1) \cdots y(N)]^\top$,

$$\Phi_\tau = \left[\hat{\phi}_f^{(\tau)}(1) \cdots \hat{\phi}_f^{(\tau)}(N) \right]^\top \text{ and } \Gamma_\tau = \left[\hat{\zeta}_f^{(\tau)}(1) \cdots \hat{\zeta}_f^{(\tau)}(N) \right]^\top.$$

7: estimate the noise signal v_o as $\hat{v}(k) = y(k) - \check{y}(k)$. Based on \hat{v} , the estimation of $\hat{\theta}_h^{(\tau+1)}$ follows using the ARMA estimation algorithm of the MATLAB identification toolbox (an IV approach can also be used for this purpose, see [39]). Increase τ by 1.

8: **until** $\theta^{(\tau)}$ has converged or the maximum number of iterations is reached.

9: **return** estimated model (2.29a–c).

to derive stochastic properties of the model estimates which is a unique property among the available approaches of the LPV identification literature. We could see that the present research focus in this framework not only lies in further developing results on these stochastic properties or understanding the rather challenging case of the p -noisy case but also establishing model structure selection tools and experiment design methods which allow better application of data-driven LPV modeling by the engineering community.

References

1. Bamieh B, Giarré L (2002) Identification of linear parameter varying models. *Int Journal of Robust Nonlin Contr* 12:841–853
2. Boyd S, Chua LO (1985) Fading memory and the problem of approximating nonlinear operators with Volterra series. *IEEE Trans Circ Syst* 32(11):1150–1161
3. Butcher M, Karimi A, Longchamp R (2008) On the consistency of certain identification methods for linear parameter varying systems. In: *Proceedings of the 17th IFAC world congress, Seoul, Korea*, pp 4018–4023
4. Casella F, Lovera M (2008) LPV/LFT modelling and identification: overview, synergies and a case study. In: *Proceedings of IEEE international symposium on computer-aided control system design, San Antonio, TX, USA*, pp 852–857
5. Cerone V, Regruto D (2008) Set-membership identification of LPV models with uncertain measurements of the time-varying parameter. In: *Proceedings of the 47th IEEE conference on decision and control, Cancun, Mexico*, pp 4491–4496
6. Dankers AG, Tóth R, Heuberger PSC, Bombois X, Van den Hof PMJ (2011) Identifiability and the informativity of data sets for LPV–ARX identification. *Proceedings of the 50th IEEE conference on decision and control, Orlando, FL, USA*, pp 799–804
7. dos Santos PL, Ramos JA, de Carvalho JLM (2007) Identification of linear parameter varying systems using an iterative deterministic-stochastic subspace approach. In: *Proceedings of the European Control Conf, Kos, Greece*, pp 4867–4873
8. Gevers M, Bazanella AS, Bombois X, Mišković L (2009) Identification and the information matrix: how to get just sufficiently rich? *IEEE Trans Automat Contr* 54(12):2828–2840
9. Giarré L, Bauso D, Falugi P, Bamieh B (2006) LPV model identification for gain scheduling control: an application to rotating stall and surge control problem. *Contr Eng Prac* 14(4):351–361
10. Heuberger PSC, Van den Hof PMJ, Bo Wahlberg (2005) *Modeling and Identification with Rational Orthonormal Basis Functions*. Springer-Verlag, London
11. Hsu K, Vincent TL, Poolla K (2008) Nonparametric methods for the identification of linear parameter varying systems. In: *Proceedings of the international symposium on computer-aided control system design, San Antonio, TX, USA*, pp 846–851
12. Khalate AA, Bombois X, Tóth R, Babuška R (2009) Optimal experimental design for LPV identification using a local approach. In: *Proceedings of the 15th IFAC symposium on system identification, Saint-Malo, France*, pp 162–167
13. Laurain V, Gilson M, Tóth R, Garnier H (2010) Refined instrumental variable methods for identification of LPV Box–Jenkins models. *Automatica* 46(6):959–967
14. Laurain V, Tóth R, Gilson M, Garnier H (2011) Direct identification of continuous-time LPV input/output models. *Special issue, IET Contr Theor Appl* 4(10):2082–2096
15. Leith DJ, Leithhead WE (1998) Gain-scheduled controller design: an analytic framework directly incorporating non-equilibrium plant dynamics. *Int J Contr* 70:249–269
16. Ljung L (1999) *System Identification, theory for the user*. Prentice Hall, London
17. Ljung L (2009) Experiments with identification of continuous time models. In: *Proceedings of the 15th IFAC symposium on system identification, Saint-Malo, France*
18. Lovera M, Mercère G (2007) Identification for gain-scheduling: a balanced subspace approach. In: *Proceedings of the American Control Conf, New York City, USA*, pp 858–863
19. Murray-Smith R, Johansen TA (1997) *Multiple model approaches to modeling and control*. Taylor and Francis, London
20. Rao GP, Unbehauen H (2004) Identification of continuous-time systems: direct or indirect? *Syst Sci* 30(3):25–50
21. Söderström T, Stoica P (1983) *Instrumental variable methods for system identification*. Springer-Verlag, New York
22. Sznaier M, Mazzaro C, Inanc T (2000) An LMI approach to control oriented identification of LPV systems. In: *Proceedings of the American Control Conf, Chicago, IL, USA*, pp 3682–3686

23. Tóth R (2010) Modeling and Identification of Linear Parameter-Varying Systems. Lecture notes in control and information sciences, Vol. 403, Springer, Germany
24. Tóth R, Abbas H, Werner W (2011a) On the state-space realization of LPV input–output model: practical approaches. *IEEE Trans Contr Syst Technol* 20(1):139–153
25. Tóth R, Bitar E, Heuberger PSC, Van den Hof PMJ, Poolla K (2011b) A prediction-error identification framework for linear parameter-varying systems. In prep.
26. Tóth R, Felici F, Heuberger PSC, Van den Hof PMJ (2007) Discrete time LPV I/O and state space representations, differences of behavior and pitfalls of interpolation. In: Proceedings of the European Control Conf, Kos, Greece, pp 5418–5425
27. Tóth R, Heuberger PSC, Van den Hof PMJ (2008) Flexible model structures for LPV identification with static scheduling dependency. In: Proceedings of the 47th IEEE conference on decision and control, Cancun, Mexico, pp 4522–4527
28. Tóth R, Heuberger PSC, Van den Hof PMJ (2009a) Asymptotically optimal orthonormal basis functions for LPV system identification. *Automatica* 45(6):1359–1370
29. Tóth R, Laurain V, Gilson M, Garnier H (2011c) On the closed loop identification of LPV models using instrumental variables. In: Proceedings of the 18th IFAC World Congress, Milano, Italy
30. Tóth R, Laurain V, Zheng W, Poolla K (2011d) A support vector machine approach for LPV linear-regression models. Proceedings of the 50th IEEE Conf. on Decision and Control, Orlando, FL, USA, pp 3192–3197
31. Tóth R, Lyzell C, Enqvist M, Heuberger PSC, Van den Hof PMJ (2009b) Order and structural dependence selection of LPV–ARX models using a nonnegative garrote approach. In: Proceedings of the 48th IEEE conference on decision and control, Shanghai, China, pp 7406–7411
32. Tóth R, Willems JC, Heuberger PSC, Van den Hof PMJ (2011e) The behavioral approach to linear parameter-varying systems. *IEEE Trans Autom Contr* 56(11):2499–2514
33. van Wingerden JW, Verhaegen M (2009) Subspace identification of bilinear and LPV systems for open- and closed-loop data. *Automatica* 45(2):372–381
34. Verdult V, Verhaegen M (2005) Kernel methods for subspace identification of multivariable LPV and bilinear systems. *Automatica* 41(9):1557–1565
35. Wei X (2006) Advanced LPV techniques for diesel engines. PhD thesis, Johannes Kepler University, Linz
36. Wei X, Del Re L (2006) On persistent excitation for parameter estimation of quasi-LPV systems and its application in modeling of diesel engine torque. In: Proceedings of the 14th IFAC symposium on system identification, Newcastle, Australia, pp 517–522
37. Willems JC, Yamamoto Y (2007) Behaviors defined by rational functions. *Linear algebra and its applications* 425:226–241
38. Young PC (1984) Recursive estimation and time-series analysis. Springer-Verlag, Berlin
39. Young PC (2008) The refined instrumental variable method: unified estimation of discrete and continuous-time transfer function models. *J Euro Syst Autom* 42:149–179



<http://www.springer.com/978-1-4614-1832-0>

Control of Linear Parameter Varying Systems with
Applications

Mohammadpour, J.; Scherer, C.W. (Eds.)

2012, XVIII, 550 p., Hardcover

ISBN: 978-1-4614-1832-0