Constrained SPICE in Volterra-Laguerre Modeling of Human Smooth Pursuit

Viktor Bro[1] and Alexander Medvedev[1]

Abstract— **The Volterra model is a well-established option in nonlinear black-box system identification. However, the estimated model is often over-parametrized. This paper presents an approach to reducing the number of parameters of a Volterra model with the kernels parametrized in the orthonormal basis of Laguerre functions by estimating it with a sparse estimation algorithm subject to constraints. The resulting parameter estimates are scrutinized for parameter redundancy and functional dependence by principal component analysis. The benefits of this approach are illustrated by identifying the human smooth pursuit system. Previous studies have suggested that the Volterra model structure is suitable for modeling the human smooth pursuit system both in health and disease. The data sets are obtained by eye tracking in a study performed on 7 test subjects diagnosed with Parkinson's disease and 22 healthy control subjects. In terms of output error, the reduced model has similar performance to that of the full model.**

I. INTRODUCTION

Nonlinear system identification has been an actively developing research field for at least two decades, see e.g. [1], [14], [15], [16]. Modeling of nonlinear dynamical systems from data is necessary in design of control and estimation algorithms that operate under significant changes in the plant and the environment and thus clearly reveal the nonlinear nature of real-life processes and systems. Further, nonlinear identification is gaining momentum as a tool of quantifying properties and discerning regulation mechanisms of living organisms. The latter function becomes increasingly important due to the role that systems biology and computational medicine play in revolutionizing health care.

There are two main approaches to nonlinear system identification with regard to model structure. One is block-oriented modeling, where the plant to be identified is stipulated to possess a certain internal structure, [3]. Wiener and Hammerstein models, that are composed of linear dynamical blocks cascaded with a nonlinear static mapping of the output or input, correspondingly, are popular and practically useful examples of block-oriented modeling. In engineered systems, a block scheme of the system design is readily available and selecting an adequate model topology for system identification is typically a straightforward task. Of course, defining the orders of the linear dynamical block and parameterizing the nonlinearities can still be challenging.

*This work was partly supported by Vinnova, Sweden's Innovation Agency through the project "Multimodal motor symptoms quantification platform for individualized Parkinson's disease treatment", MuSyQ

[1]Division of Systems and Control, Department of Information Technology, Uppsala University, Uppsala, Sweden {viktor.bro, alexander.medvedev}@it.uu.se

Nonlinear black-box system identification is another approach to estimating models from data. In this case, also the internal topology of the systems is assumed unknown and has to be established from input-output data. Two mathematical paradigms appear to be useful in this respect: Volterra models [7] and artificial neural networks [20]. Both modeling vehicles introduce massive over-parametrization to capture the nonlinear dynamics that results in identifiability issues. Sparse estimation in Volterra models [5], [4], [18] and pruning in neural networks [19] are commonly enforced to keep model complexity reasonably low.

Applying nonlinear system identification to living organisms is not straightforward. Dynamical variability between subjects of the same species and in the same subject at different time instants is orders of magnitude higher than that one faces in technical systems, where most of the variability can be attributed to measurement noise. It is also often impossible to decouple a biological subsystem from the rest of the organism without losing essential functions, a property creating unavoidable cross-talk between the connected loops.

All biological systems operate in closed loop. The mechanisms underlying biological regulations are often unknown, as well as the topology of the system itself. In this sense, the situation with system uncertainty is reversed compared to that in engineered systems. While in engineered systems there is no uncertainty in control laws since they are readily defined by design, biological control is still not well understood, especially when it is implemented by neural circuits.

The present paper aims at demonstrating how sparse estimation can be exploited for elucidating the structure of a discrete Volterra model with kernels parametrized in the orthonormal basis of Laguerre functions. The human smooth pursuit system (SPS) is selected as a meaningful example of a biological plant to be modeled. The main contribution of the presented study is twofold. On the one hand, it is suggested that the sparsity-promoting mechanism of a parameter estimation algorithm is to be used for constraining the structural degrees of freedom in Volterra models. On the other hand, the efficacy of the proposed approach is exemplified on a practically important application of SPS modeling that is involved in the areas of biometrics, medical diagnostics, and disability aids.

The paper is composed as follows. First, a brief summary of Volterra-Laguerre models and the sparse parameter estimation algorithm SPICE is provided. Next, the human SPS system and video-based eye tracking are described, followed up by relevant facts regarding the experimental setup. Further, model estimation results from eye tracking

data are presented along with their principal component analysis. Finally, some conclusions are drawn.

II. PRELIMINARIES

A. The human smooth pursuit system

There are two ways in which humans shift their gaze involuntarily: saccades and smooth pursuit. While saccades are quick, episodic movements where the gaze shifts focus from one instant to another, the smooth pursuit system (SPS) allows the gaze to continuously track some object of interest, constantly keeping it within focus.

The SPS is a complex neurally controlled feedback system, involving the eyes, the extraocular muscles, and several parts of the brain [22]. Because of this, the SPS can be affected by for example alcohol and drugs [24], as well as by mental and neurological conditions such as schizophrenia [13] and Parkinson's disease [2], [6], [8], [9], [10], [11].

B. The Volterra Model

The Volterra series is a functional expansion of a dynamical, nonlinear, time-invariant system. A discrete system with input $u(k) \in \mathbb{R}$ and output $y(k) \in \mathbb{R}$, $k = 0, \cdots, K-1$ may be approximated by the truncated Volterra series

$$y(k) = y_0 + \sum_{n=1}^{N} H_n u(k) + e(k), \tag{1}$$

where $e(k) \in \mathbb{R}$ is a noise term, $N \in \mathbb{N}$ is the *Volterra order*, and

$$H_n u(k) = \sum_{i_1=0}^{\infty} \cdots \sum_{i_n=0}^{\infty} h_n(i_1, \cdots, i_n) u(k-i_1) \cdots u(k-i_n) \tag{2}$$

are the *Volterra functionals*. The functions h_n are *Volterra kernels*. In most practical cases, the Volterra kernels are cumbersome to calculate explicitly. Therefore, the kernel functions are usually expanded in terms of some orthogonal basis. For the kernels that are in $\ell^2[0, \infty)$ and not excessively oscillative, the discrete Laguerre functions present a popular choice.

C. Laguerre series representation of the Volterra Kernels

The j:th Laguerre function is defined in the $\mathcal{Z}$-domain as

$$\Phi_j(z) = \frac{\sqrt{1-\alpha}z}{z-\sqrt{\alpha}} \left(\frac{1-\sqrt{\alpha}z}{z-\sqrt{\alpha}} \right)^j, \tag{3}$$

where $0 < \alpha < 1$ is the Laguerre parameter. Denote the corresponding functions in the time domain through the inverse $\mathcal{Z}$-transform by $\phi_j(k) = \mathcal{Z}^{-1}\{\Phi_j(z)\}$. The properties of the Laguerre functions are described in detail in e.g. [7]. Furthermore, these functions form an orthonormal basis in $\ell_2[0, \infty)$ so that a function $h(\cdot) \in \ell_2[0, \infty)$ can be unambiguously written as

$$h(k) = \sum_{j=0}^{\infty} \gamma_j \phi_j(k). \tag{4}$$

The Volterra functionals can then be parametrized using the Laguerre functions as

$$h_n(i_1, \cdots, i_n) = \sum_{j_1=0}^{\infty} \cdots \sum_{j_n=0}^{\infty} \gamma_n(j_1, \cdots, j_n) \phi_{j_1}(i_1) \cdots \phi_{j_n}(i_n). \tag{5}$$

Note that, in practice, only a finite number of basis functions can be reliably calculated. Thus, a truncated Laguerre expansion usually becomes an approximation of the true kernel.

Let L denote the *Laguerre order* of a truncated series, i.e. the highest order of the included Laguerre functions. Then

$$h_n(i_1, \cdots, i_n) \approx \sum_{j_1=0}^{L} \cdots \sum_{j_n=0}^{L} \gamma_n(j_1, \cdots, j_n) \phi_{j_1}(i_1) \cdots \phi_{j_n}(i_n) \tag{6}$$

and the Volterra functionals become

$$H_n u(k) = \sum_{j_1=0}^{\infty} \cdots \sum_{j_n=0}^{\infty} \gamma_n(j_1, \ldots, j_n) \psi_{j_1}(k) \cdots \psi_{j_n}(k), \tag{7}$$

where the convolution $\psi_j(k) = (\phi_j * u)(k)$ denotes the *Laguerre filter output*. The Volterra-Laguerre (VL) model is finally written as

$$y(k) = y_0 + \sum_{n=1}^{N} \sum_{j_1=0}^{L} \cdots \sum_{j_n=0}^{L} \gamma_n(j_1, \ldots, j_n) \prod_{l=1}^{n} \psi_{j_l}(k) + e(k). \tag{8}$$

The kernel functions are symmetric with respect to index, since the Laguerre functions commute. Therefore, many VL coefficients are redundant and do not have to be estimated individually. This reduces the number of coefficients in a Volterra model of order N parametrized in L Laguerre functions from $(L^{N+1}-1)/(L-1)$ to

$$N_c = \binom{L+N+1}{N}.$$

D. VL models in linear regression form

Let $\mathbf{c}$ be a vector of all N_c Volterra-Laguerre coefficients, $\mathbf{y} = [y(0), \ldots, y(K-1)]^T$ be the vector of measurements, and $\mathbf{e} = [e(0), \ldots, e(K-1)]^T$ be the corresponding vector of noise terms. Then, a model for the data can be formulated as

$$\mathbf{y} = \begin{bmatrix} \mathbf{\Psi} & \mathbf{I}_K \end{bmatrix} \begin{bmatrix} \mathbf{c} \\ \mathbf{e} \end{bmatrix} = \mathbf{B}\beta, \tag{9}$$

where $\mathbf{\Psi} = \begin{bmatrix} \psi_0 & \psi_1 & \ldots & \psi_0^2 & \psi_0\psi_1 & \cdots \end{bmatrix}$ is the regression matrix constructed from the Laguerre filter outputs $\psi_j = \begin{bmatrix} \psi_j(0) & \ldots & \psi_j(K-1) \end{bmatrix}$, and $\beta = \begin{bmatrix} \beta_1 & \ldots & \beta_{N_c+K} \end{bmatrix}^T$. The parameters of the VL model may be estimated using ordinary least squares (LS). However, with a higher Laguerre order follows a higher number of estimands, which in turn results in a higher variance of the parameter estimates. This, logically, demands a higher degree of excitation in the system input $u(t)$, ultimately making it white noise. While being convenient in theoretical development, white noise inputs are infeasible in most practical system identification problems.

For instance, in the identification of SPS, white visual stimuli would fail to invoke smooth pursuit in the test subject. Thus, it is of interest to reduce the number of model parameters as much as possible, e.g. by using sparse estimation.

E. SPICE

One method for sparse parameter estimation is *SParse Iterative Covariance-based Estimation* (SPICE) [21]. While other methods for sparse estimation, e.g. LASSO [23], require proper tuning of some hyperparameters, the SPICE algorithm does not.

The SPICE algorithm can be formulated as follows. Introduce the quantities $w_k = \frac{\|\mathbf{b}_k\|_2}{\|\mathbf{y}\|_2}$, where $\mathbf{b}_k$ denotes the k:th column of $\mathbf{B}$ in (9) and $\|\cdot\|_2$ is the Euclidean vector norm. The SPICE estimate of the coefficients is then found by solving the linear program

$$\begin{aligned} &\min_{\alpha,\beta} \sum_{i=1}^{N_c+K} w_i\alpha_i \\ &\text{s.t.} -\alpha_i \leq \beta_i \leq \alpha_i, \\ &\quad \alpha_i \geq 0,\ i = 1,\ldots,N_c+K \\ &\quad \mathbf{y} = \mathbf{B}\beta. \end{aligned} \tag{10}$$

a) Weighted SPICE estimation: Since the Volterra series converge, the coefficient values of higher order kernels decrease in magnitude. When relevant system information is communicated by higher-order terms, it becomes necessary to weigh the higher order coefficients when using the SPICE algorithm, so that these are not "lost" in pursuit of sparsity. The weight ω_i for the individual coefficient β_i is introduced in the constraint $-\alpha_i \leq \omega_i\beta_i \leq \alpha_i$. Thus, to weigh only the quadratic part with some uniform value ω, let $\omega_i = \omega$ for coefficients corresponding to the quadratic kernels, and $\omega_i = 1$ otherwise.

b) Constrained SPICE: If some function is known beforehand to be absent from a Volterra kernel, a constraint may be introduced in the SPICE algorithm, so that the corresponding coefficient is set to zero. Similarly, additional linear constraints in the estimated coefficients, such as positivity of certain terms in the Volterra series, can be added to the linear program to be solved.

F. Principal Component Analysis

Consider M measurements $\{\mathbf{x}_i, i = 1,\ldots,M\} \in \mathbb{R}^{N_c}$. Principal component analysis (PCA) can be thought of as fitting an ellipsoid to the measurements, and projecting these onto the axes of the ellipsoid, i.e. the principal components. If the data variance along some principal component is small, this component may be omitted, thus reducing the dimensionality of the problem. The PCA can also be used as an exploratory tool for finding functional relationships between data components.

In practice, the principal components are found as follows. Let $\overline{x}_i$ denote the mean of $\mathbf{x}_i$, and let

$$\widetilde{\mathbf{X}} = \begin{bmatrix} \mathbf{x}_1 - \overline{x}_1\mathbf{1} & \ldots & \mathbf{x}_M - \overline{x}_M\mathbf{1} \end{bmatrix} \tag{11}$$

be the mean subtracted data matrix. Consider now the singular value decomposition of $\widetilde{\mathbf{X}}$: $\widetilde{\mathbf{X}} = \mathbf{U}\mathbf{\Sigma}\mathbf{W}^\mathbf{T}$, where $\mathbf{\Sigma}$ contains the singular values of $\widetilde{\mathbf{X}}$ and $\mathbf{W}$ contains the eigenvectors of $\widetilde{\mathbf{X}}^T\widetilde{\mathbf{X}}$. The transformation

$$\mathbf{T} = \widetilde{\mathbf{X}}\mathbf{W} = \mathbf{U}\mathbf{\Sigma}\mathbf{W}^\mathbf{T}\mathbf{W} = \mathbf{U}\mathbf{\Sigma} \tag{12}$$

then gives the projection onto the principal components. By deleting rows and columns from $\mathbf{U}$ and $\mathbf{\Sigma}$ and keeping only the ones corresponding to the d largest singular values, the dimensionality of the data set is reduced. At the same time, the error $\|\widetilde{\mathbf{X}} - \widetilde{\mathbf{X}}_d\|_2^2$ is minimized. Here, $\widetilde{\mathbf{X}}_d$ denotes the reduced dimension data matrix.

III. Experiments and Data Collection

The experimental setup consisted of a computer screen and a video-based eye tracker from SmartEye AB, Sweden. The eye tracker records the *gaze position* of the test subject, i.e. the point on the screen where the subject is looking. Test subjects were placed about 50 cm from the computer screen, and a stimulus was displayed. The stimulus consisted of a colored dot moving on a black background. The stimuli, acting as input signals to the SPS were generated using the method described in [10], and the gaze position was sampled at 60 Hz. The measured gaze position is then considered as the system output. Notice that as nonlinear identification is performed, the stimuli have to provide excitation both in amplitude and frequency. For this application, it implies that the dot has to visit all the areas within the stimuli window and move within a suitable for tracking range of acceleration.

Two different groups of test subjects are considered. One control group consisting of 22 healthy individuals between 50 and 76 years old, and one group of 7 individuals diagnosed with Parkinson's disease, between 61 and 80 years old. In total 499 measurements from the control group and 144 measurements from the patient group have been collected and used in the present study.

The tests were conducted at CTC (Clinical Trials Consultants AB) Center at the University Hospital in Uppsala, Sweden, between May and August 2015. The study was performed as part of the project *"MuSyQ: Multimodal motor symptoms quantification platform for individualized Parkinson's disease treatment"* and is described in detail in [17].

It is important to note that, prior to the first eye-tracking test, each patient received a dose of an anti-PD drug, after an 8 hours long washout. The administered dose was set to 150% of the patient's usual morning dose of medication. This experimental protocol allows to track the patient's symptoms, as they transition from off-state, to normal mobililty and/or dyskinesia, and back. Thus, the measurements from the patient group exhibit a range of behaviors, including lack of visible PD symptoms. Because of this, a larger domain of coefficients is expected to be needed to describe the SPS dynamics of the patient group, compared to that of the control group. Further, the distribution of model coefficients estimated for the patients is anticipated to partially overlap with those for the control group.

IV. VL MODEL REDUCTION USING SPICE AND PCA

It has been shown that a VL model of Volterra order 2, with Volterra kernels parametrized in the three first Laguerre functions is a feasible model structure for modeling the SPS dynamics [12]. This model may be written as

$$y(k) = y_0 + \sum_{j=0}^{2} \gamma_1(j)\psi_j(k) + \\ + \sum_{j_1=0}^{2}\sum_{j_2=j_1}^{2} \gamma_2(j_1,j_2)\psi_{j_1}(k)\psi_{j_2}(k), \tag{13}$$

and is parametrized by altogether 10 coefficients. The output $y(k)$ is in this case the gaze position and $\psi(k)$ are the Laguerre filter outputs, using the stimulus position as input signal. The model thus predics gaze position from a known visual stimuli. As shown in [12], model (13) is significantly overparametrized and the SPICE algorithm can be used to reduce the number of nonzero parameters. Model (13) is therefore chosen as a starting point for describing the SPS dynamics of the test subjects described above. The steps outlined next describe how to systematically reduce the number of parameters until a minimal model is acquired.

A. Finding redundant parameters

Using the model structure described above and estimating the parameters for test subjects by SPICE gives a sparse representation of the dynamics. Yet, due to inter- and intra-individual variability, different sparse representations are obtained for different experimental datasets. To enable effective comparison between healthy controls and PD patients, and to quantify the treatment effect in the latter, a parsimonious model representation that captures the dynamics of all the data sets is sought.

To further investigate the sparse structure of the models, a PCA was performed. For the linear part, 96% of the data variance is described by the two largest principal components. Similarly, for the quadratic part, 93% of the data variance is described by the three largest principal components.

a) Linear part: Fig. 1 shows the linear coefficients projected onto the two largest principal components. A symmetric structure is evident in the data, suggesting redundancy in the model parameters. This redundancy can be removed by setting one coefficient in the linear part to zero. Table I presents the median normalized (with respect to signal norm) mean square error (NMSE) of the model output, in the case when all linear coefficients are estimated, as well as in the three different cases when one linear coefficient in set to zero. The results are similar in all three cases, although for the control group, the NMSE remains approximately unchanged when the coefficient $\gamma_1(2)$ is excluded. Therefore, the third coefficient was decided to be set to zero. Note however, that for the patient group, $\gamma_1(1)$ seems redundant as removing this gives the smallest increase in error. This may stem from differences in the dynamics between healthy individuals and individuals with Parkinson's disease. Fig. 2 shows the coefficient values of the linear part, with this constraint, and

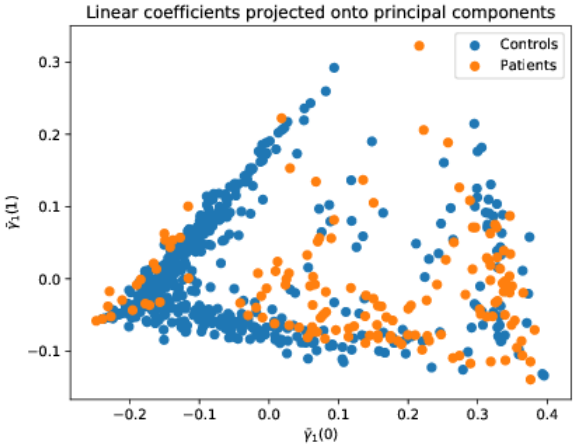

Fig. 1. Linear coefficients of VL model, estimated by SPICE, projected onto the two largest principal components. There is a pronounced symmetric, triangular structure in the data, indicating redundancy in the parametrization.

TABLE I
MEDIAN ERROR FOR DIFFERENT MODEL STRUCTURES.

	Controls	Patients
No constraints	**0.454**	**0.886**
$\gamma_1(0) = 0$	0.458	0.890
$\gamma_1(1) = 0$	0.462	**0.887**
$\gamma_1(2) = 0$	**0.454**	0.920

demonstrates that the symmetric structure has disappeared with the reduced kernel representation.

b) Nonlinear part: In the quadratic part, symmetries similar to those in the linear part were not observed. However, only three functions occur often. These functions correspond to the components of the quadratic kernel functions ϕ_0^2, $\phi_0\phi_2$ and ϕ_2^2. Three stimuli were used in the eye tracking experiments. With dissimilar excitation properties of the stimuli, they may excite different dynamics of the SPS and in turn result in different sets of Laguerre functions in the estimated kernels of the sparse model. However, as the data in Table II indicate, the choice of stimuli set does not seem to have had any impact on the structure of the quadratic term of the VL model. That is, there is no significant difference in excitation ability between the different stimuli.

Furthermore, Fig. 3 shows that when a weight ω is introduced on the quadratic kernel in the SPICE algorithm, the probability of a function being present in the quadratic kernel increases as ω decreases. The weight acts as a threshold on the quadratic kernel functions and, when the threshold is lowered, the relation between the functions is intact. When the weight is lowered even further, to $\omega = 10^{-4}$, all functions are present in the quadratic kernel, every time. This strategy of using weights in the SPICE algorithm may be applied to get an idea of which functions are significant to the model. In the present case, the three kernel functions ϕ_0^2, $\phi_0\phi_2$, and ϕ_2^2 are the most significant, followed by ϕ_1^2. This is in line with the results presented in Table II. Consequently, the model can be reduced further, to contain only the functions ϕ_0^2, $\phi_0\phi_2$

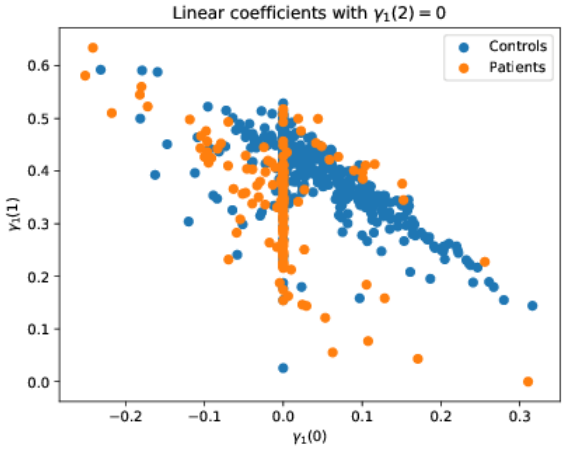

Fig. 2. Linear coefficients of VL model, estimated by SPICE, with the constraint that $\gamma_1(2) = 0$. The symmetric structure from before has disappeared. Because of the sparsity-promoting nature of the SPICE algorithm, the first linear coefficient, $\gamma_1(0)$, has been set to zero a number of times, which is visible in the data.

TABLE II

FREQUENCY WITH WHICH A KERNEL FUNCTION IS PRESENT IN THE QUADRATIC KERNEL FOR A CERTAIN STIMULUS.

	Stimulus #		
Function	1	2	3
ϕ_0^2	0.53	0.43	0.27
$\phi_0\phi_1$	0	0	0
$\phi_0\phi_2$	0.15	0.21	0.18
ϕ_1^2	0.15	0.07	0
$\phi_1\phi_2$	0.08	0.14	0
ϕ_2^2	0.69	0.43	0.45

and ϕ_2^2 in the quadratic kernel.

B. Finding and exploiting relationships between coefficients

While only three functions are present in the quadratic part, there is a linear relationship between the coefficient values, as seen in Fig. 4. This relation is estimated using linear regression to be

$$\begin{aligned}\gamma_2(0,2) &= -1.81468 \cdot \gamma(0,0),\\ \gamma_2(2,2) &= 0.938454 \cdot \gamma(0,0),\end{aligned} \tag{14}$$

which can also be seen in the figure. Using this relation further reduces the number of coefficients that need to be estimated, while not introducing much modeling error.

C. A minimal model

The final minimal model contains only four coefficients to be estimated, $\{y_0, \gamma_1(0), \gamma_1(1), \gamma_2(0,0)\}$, and can be written as

$$\begin{aligned}y(k) = y_0 &+ \gamma_1(0)\psi_0(k) + \gamma_1(1)\psi_1(k) + \\ &+ \gamma_2(0,0)\left(\psi_0^2(k) + \theta_{02}\psi_0(k)\psi_2(k) + \theta_{22}\psi_2^2(k)\right),\end{aligned} \tag{15}$$

where $\theta_{02} = -1.81468$ and $\theta_{22} = .938454$, for both the patient and control group.

The distributions of the estimated coefficients of the linear part, $\gamma_1(0)$ and $\gamma_1(1)$, as well as the distribution of the parameter in the quadratic part, $\gamma_2(0,0)$, are shown in Fig. 5. The distributions for the control and patient groups have approximately the same mode, but the patient group distribution is wider. This could be due to the wider range of smooth pursuit dynamics being present in the patient group because of the design of the study. In a similar way, the distributions of the linear parameters are different, though, as expected, largely overlapping.

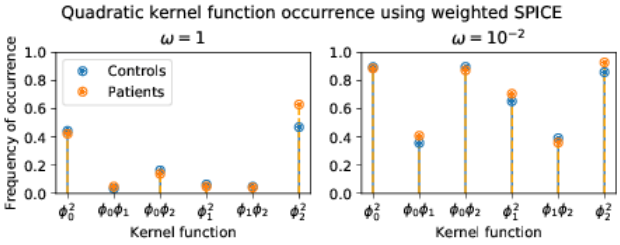

Fig. 3. Occurrence of functions in the quadratic VL kernel, using different weights in SPICE. The same functions are present in the control group and in the patient group, and the relations between functions are intact across weights. The most commonly occurring functions are ϕ_0^2, $\phi_0\phi_2$ and ϕ_2^2.

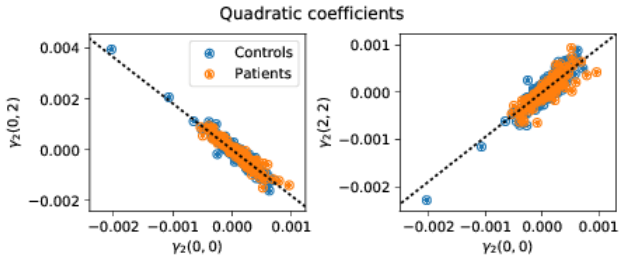

Fig. 4. Coefficients of the quadratic part of the reduced VL model. There is a clear linear relationship between the coefficient values, indicated by the dotted black line.

The output NMSE, for the final model as well as for the full model with parameters estimated using ordinary LS and SPICE, are approximately log-normal distributed. This can be seen in the normal probability plots for the logarithm of the errors shown in Fig. 6. A comparison of the fitted log-normal error distributions is provided in Fig. 7. For the control group, the difference in error is very small; the mode of the fitted distributions for the full, LS estimated, model and the reduced model are 0.26 and 0.29. For the patient group the difference is slightly larger, with corresponding modes at 0.53 and 0.75.

V. CONCLUSIONS

While the black-box Volterra-Laguerre model structure is known to be suitable for modeling of nonlinear systems, it suffers from over-parametrization. Sparse estimation introduces a way of reducing the number of model parameters allowing at the same time for weights and constraints on the parameters. The resulting model structure can be scrutinized for redundancy and functional dependance between the parameter estimates by e.g. principal component analysis.

The efficacy of the approach is illustrated on a real world example, the human smooth pursuit system, where

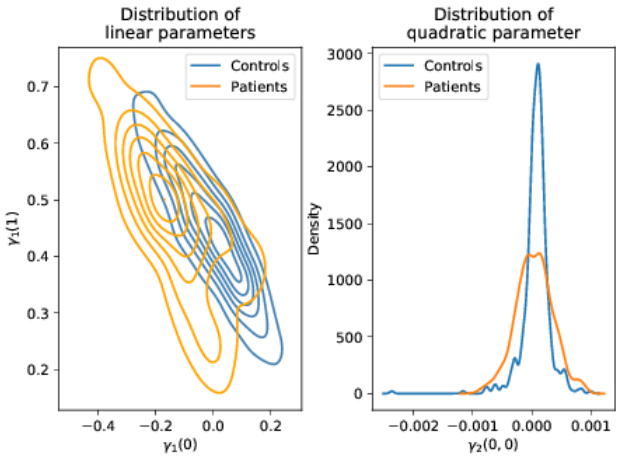

Fig. 5. Left: Distribution of coefficient values for the linear part of the final model, for both the control and patient group. Right: Distribution of the quadratic part coefficient for both control and patient group.

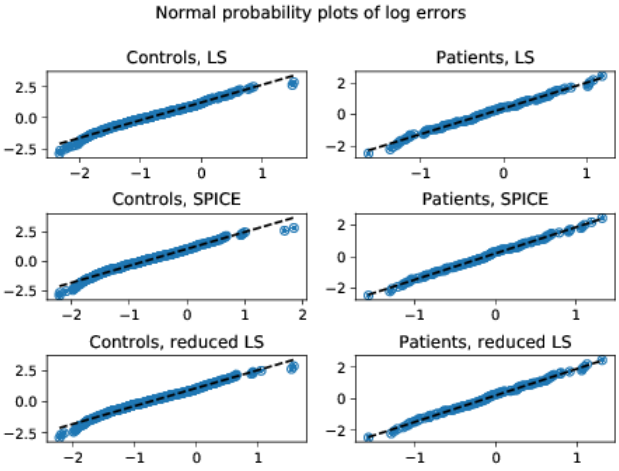

Fig. 6. Normal probability plots of logarithm of errors, for three different models, for both the control group and the patient group. In all cases, the distribution of the errors is approximately log-normal.

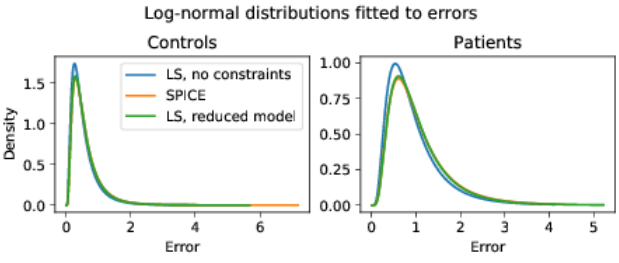

Fig. 7. Distribution of errors using three different model structures and estimation schemes. *LS, no constraints* indicates the case where all 10 parameters were estimated using least squares; *SPICE* indicates the case where the SPICE algorithm was used to get a sparse estimate of the parameters, without constraints, and *LS, reduced model* indicates the case where least squares was used to estimate only 4 parameters of the model.

significant simplifications have been made in terms of model complexity, reducing the number of parameters from 10 to 4. Meanwhile, the increase in error is small, indicating that the informative parts of the model are intact and only superfluous terms due to overparametrization are eliminated. A log-normal distribution of the modelling error indicates that the model quality cannot be improved given the data.

References

[1] S. A. Billings. Nonlinear System Identification: NARMAX Methods in the Time, Frequency, and Spatio-Temporal Domains 2013

[2] J. M. Gibson, R. Pimlott and C. Kennard. Ocular motor and manual tracking in Parkinson's disease and the effect of treatment. Journal of Neurology Neurosurgery and Psychiatry 50(7), pp. 853-860. 1987.

[3] F. Giri and E. Bai. Block-Oriented Nonlinear System Identification. 2010

[4] V. Kekatos, D. Angelosante and G. B. Giannakis. Sparsity-aware estimation of nonlinear Volterra kernels. 2009 3rd IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP), Aruba, Dutch Antilles, pp. 129-132. 2009.

[5] V. Kekatos and G. B. Giannakis. Sparse Volterra and polynomial regression models: Recoverability and estimation. IEEE Transactions on Signal Processing 59(12), pp. 5907-5920. 2011.

[6] S. Marino et al. Quantitative analysis of pursuit ocular movements in Parkinson's disease by using a video-based eye tracking system. European Neurology 58(4), pp. 193-197. 2007.

[7] V. Z. Marmarelis. Identification of nonlinear biological systems using Laguerre expansions of kernels. Annals of Biomedical Engineering 21(6), pp. 573-589. 1993.

[8] D. Jansson et al. Mathematical modeling and grey-box identification of the human smooth pursuit mechanism. IEEE CCA, 2010.

[9] D. Jansson and A. Medvedev. Dynamic smooth pursuit gain estimation from eye tracking data. IEEE CDC and ECC, Orlando, 2011.

[10] D. Jansson and A. Medvedev. Visual stimulus design in parameter estimation of the human smooth pursuit system from eye-tracking data. American Control Conference (ACC), Washington, 2013.

[11] D. Jansson, O. Rosen and A. Medvedev. Non-parametric analysis of eye-tracking data by anomaly detection. ECC, Zurich, 2013.

[12] D. Jansson and A. Medvedev. Volterra modeling of the Smooth Pursuit System with application to motor symptoms characterization in Parkinson's disease. 2014 European Control Conference (ECC), Strasbourg. pp. 1856-1861. 2014.

[13] G. A. O'Driscoll and B. L. Callahan. Smooth pursuit in schizophrenia: A meta-analytic review of research since 1993. Brain and Cognition 68(3), pp. 359-370. 2008.

[14] J. Paduart et al. Identification of nonlinear systems using polynomial nonlinear state space models. Automatica 46(4), pp. 647-656. 2010.

[15] M. Schetzen. The Volterra and Wiener Theories of Nonlinear Systems 1980.

[16] T. B. Schön et al. System identification of nonlinear state-space models. Automatica 47(1), pp. 39-49. 2011.

[17] M. Senek et al. Levodopa/carbidopa microtablets in Parkinson's disease: A study of pharmacokinetics and blinded motor assessment. European Journal of Clinical Pharmacology 2017

[18] K. Shi and P. Shi. Adaptive sparse Volterra system identification with l(0)-norm penalty. Signal Processing 91(10), pp. 2432-2436. 2011.

[19] M. R. Silvestre and L. Luan Ling. Statistical evaluation of pruning methods applied in hidden neurons of the MLP neural network. IEEE Latin America Transactions 4(4), pp. 249-256. 2006.

[20] J. Sjöberg et al. Nonlinear black-box modeling in system identification: A unified overview. Automatica 31(12), pp. 1691. 1995.

[21] P. Stoica et al. SPICE and LIKES: Two hyperparameter-free methods for sparse-parameter estimation. Signal Processing 92(7), pp. 1580-1590. 2012.

[22] Peter Thier and Uwe J Ilg. The neural basis of smooth-pursuit eye movements. Current Opinion in Neurobiology 15(6), pp. 645-652. 2005.

[23] R. Tibshirani. Regression shrinkage and selection via the LASSO. Journal of the Royal Statistical Society. Series B (Methodological) 58(1), pp. 267-288. 1996.

[24] I. M. S. Wilkinson. The influence of drugs and alcohol upon human eye movement. Proceedings of the Royal Society of Medicine 69(7), pp. 479-480. 1976.