Autoregressive models for biomedical signal processing

Jonas F. Haderlein, Andre D. H. Peterson, Anthony N. Burkitt, *Senior Member, IEEE*,
Iven M. Y. Mareels, *Fellow, IEEE* and David B. Grayden, *Senior Member, IEEE*

Abstract— **Autoregressive models are ubiquitous tools for the analysis of time series in many domains such as computational neuroscience and biomedical engineering. In these domains, data is, for example, collected from measurements of brain activity. Crucially, this data is subject to measurement errors as well as uncertainties in the underlying system model. As a result, standard signal processing using autoregressive model estimators may be biased. We present a framework for autoregressive modelling that incorporates these uncertainties explicitly via an overparameterised loss function. To optimise this loss, we derive an algorithm that alternates between state and parameter estimation. Our work shows that the procedure is able to successfully denoise time series and successfully reconstruct system parameters.**

Clinical relevance— **This new paradigm can be used in a multitude of applications in neuroscience such as brain-computer interface data analysis and better understanding of brain dynamics in diseases such as epilepsy.**

I. INTRODUCTION

Biomedical signal processing necessitates a principled treatment of uncertainties in the data generating process and measurement acquisition. Recent applications in this field are the analyses of electroencephalogram (EEG) or other brain-computer interface (BCI) measurements. Careful data analysis is crucial for a study of brain function and dysfunction, brain signal decoding, and BCI-based control of external devices in fields like bionic prostheses and robotics. However, inherent sources of uncertainty are, inevitably, measurement noise as well as the underlying system model. The latter source is a key challenge in the biomedical domain, since system models (usually in the forms of dynamical systems) for the brain or the observed neuronal populations are unknown, and the system state is inherently immeasurable with current technology.

Instead of postulating data-generating models for the brain, many approaches focus on black-box analysis with autoregressive (AR) models. Due to the time-varying and non-autonomous nature of the true dynamics, AR models are commonly fitted over short finite horizons of measurements that are oscillatory in nature. In the neuroscience domain,

AR models are used to infer neuronal network structures [1]–[3] and for intervention or control mechanisms [4]. In the BCI domain, AR models are often used to estimate the frequency spectrum of brain activity [5]–[7]. We observe that it is the reconstruction of parameters and denoised states that are often of interest here, rather than time series prediction. As we show in the following, classical estimators for AR models can be biased under noise, and may perform sub-optimally in such tasks.

AR modelling is an area within system identification, a well-studied and broad field [8]–[10]. The general problem is to model the dynamics of a partially observed system from measurements only. Denote the measurements $y_t \in R^p$ with finite time index $t = 1, \cdots, N$. In AR modelling, a function g is proposed that models the evolution of y_t as an approximation of the true dynamics f with an unknown state dimension. The existence of such models g follows from the Takens-Aeyels-Sauer theorem [11] (having observed noise-less measurements of sufficient length N) with a delay embedding of the form

$$y_{t+1} \quad = \quad g_\Theta(y_t, y_{t-1}, \cdots, y_{t-r+1})^T, \quad 0 < r < N. \tag{1}$$

Here, Θ is a parameter vector specifying the model and r reflects our prior assumption of the true system state dimension.

Linear or nonlinear autoregressive (NAR) models g_Θ can be constructed with basis functions of various form, such as linear or polynomial functions or via (recurrent) neural networks [9]. Modelling, i.e., system identification, is then a search for optimal model parameters Θ.

In the following, we assume $p = 1$ for simplicity, but the analysis can easily be generalised for $p > 1$, as shown in Example III-C.

The following sections provide an overview of the two main modelling frameworks [12].

A. Prediction error framework

The prediction error framework assumes a model

$$y_{t+1} \quad = \quad g_\Theta(y_t, y_{t-1}, \cdots, y_{t-r+1}) + \nu_t, \tag{2}$$

where ν_t is a transition noise. One optimises the one-step-ahead prediction loss $\ell_P(r, \Theta, Y)$ over parameters $\{r, \Theta\}$:

$$\min_{r, \Theta} \quad \sum_{t=r}^{N-1} (y_{t+1} - g_\Theta(x_t))^2, \tag{3}$$

with delay embedding $x_t = (y_t, y_{t-1}, \cdots, y_{t-r+1})^T \in R^r$ and measurements $Y = (y_1, \cdots, y_N)^T$.

By directly using Y as input and target variables, this approach assigns a high degree of trust in the measurements and, as a result, reconstructs biased Θ in the presence of measurement noise [12]–[14].

The linear AR case can be solved directly with least squares techniques, which can be readily extended to any linear-in-the-parameters model class [15]–[17]. We note that Koopman theory also provides a way to model nonlinear behaviour in terms of measurements only [18]–[20]. Alternative methods that reconstruct linear state-space models from measurements are known as sub-space estimation [8], [21].

Examples of nonlinear models based on the program in (3) have been presented in [22]–[24].

B. Simulation error framework

The simulation error framework asks for a model

$$
\begin{aligned}
\hat{y}_{t+1} &= g_\Theta(\hat{y}_t, \hat{y}_{t-1}, \cdots, \hat{y}_{t-r+1}), \\
y_t &= \hat{y}_t + \mu_t,
\end{aligned} \tag{4}
$$

where μ_t is measurement noise. We optimise a loss $\ell_S(r, \Theta, \hat{x}_r, Y)$ over parameters $\{r, \Theta, \hat{x}_r\}$, including the initial condition $\hat{x}_r = (\hat{y}_r, \cdots, \hat{y}_1)^T$:

$$
\begin{aligned}
\min_{r, \Theta, \hat{x}_r} \quad & \sum_{t=r+1}^{N} (y_t - \hat{y}_t)^2 \\
\text{s.t.} \quad & \hat{y}_{t+1} = g_\Theta(\hat{x}_t), \quad t = r, \cdots, N-1,
\end{aligned} \tag{5}
$$

where $\hat{y}_t$ is the simulation estimate of the model at time t and $\hat{x}_t = (\hat{y}_t, \cdots, \hat{y}_{t-r+1})^T$ is the delay embedding.

This optimisation provides a denoised state estimate $\hat{y}_t \; \forall t$, but, at the same time, assumes no uncertainty in the model. While in principle appealing, it is often complex to optimise the non-convex ℓ_S in practice, even for very simple g_Θ [25].

The complexity of the loss in (5) motivates the use of advanced optimisation programs such as multiple shooting or evolutionary algorithms [14], [26], [27]. Alternatively, prediction error-based models can be further pruned by minimising a simulation error loss [13]. Decomposing the trajectory Y and optimising only over multi-step predictions is also employed in order to relax the problem [28], [29].

Recent work includes the use of neural delay differential equations, a class of neural network basis functions that can be tuned via simulation error optimisation [30].

C. Outline

Neither prediction error nor simulation error frameworks incorporate both measurement and model errors separately. Alternative approaches that do so have been proposed for known models [31], [32] and black-box models [33]. Similarly, model estimation in Bayesian frameworks deals with both kinds of uncertainties and is widely adopted [34]–[36]. One approach for these methods is expectation-maximisation, which requires iterative state and parameter estimation, e.g., via Kalman filtering [37]–[39].

Our main contribution in this study is to extend the classical frameworks by incorporating these ideas into AR modelling. We first derive an overparameterised loss function so that we neither only search for a denoised state with total

certainty in the model (simulation error framework) nor totally rely on the measurements (prediction error framework). We then propose a novel optimisation algorithm akin to expectation-maximisation. This recipe solves both parameter and state estimation in batch mode, as opposed to recursive techniques such as Kalman filtering and smoothing [8].

II. AR MODELS FOR STATE AND MEASUREMENT ERRORS

With an AR model in state-space form,

$$
\begin{aligned}
\hat{x}_{t+1} &= g_\Theta(\hat{x}_t) + \nu_t, \\
y_t &= C\hat{x}_t + \mu_t,
\end{aligned} \tag{6}
$$

with row vector $C \in R^{1 \times r}$, we minimise a loss, over the set $\{r, \hat{\Theta}, \hat{X}\}$ derived from the log-likelihood of state transitions [39],

$$
\begin{aligned}
\ell(r, \Theta, \hat{X}, Y) = \sum_{t=r}^{N-1} (\hat{x}_{t+1} - g_\Theta(\hat{x}_t))^2 + \\
\rho \sum_{t=r}^{N} (y_t - C\hat{x}_t)^2,
\end{aligned} \tag{7}
$$

with $\rho > 0$ as a variable expressing our belief in the measurements vs. the model and $\hat{x}_t$ as a state estimate of respective delays.

The loss in (7) incorporates both denoised states $\hat{X} = (\hat{x}_r, \cdots \hat{x}_N)^T$ and parameters Θ that need to be optimised. It is thus, in some sense, a relaxation of problem (5) via the auxiliary state variables.

We discuss the linear and nonlinear case in the following.

A. The linear finite horizon case

We consider the linear map

$$
\begin{aligned}
\hat{x}_{t+1} &= A\hat{x}_t + \nu_t, \\
y_t &= C\hat{x}_t + \mu_t,
\end{aligned} \tag{8}
$$

with a priori unknown A in companion matrix form with the autoregressive coefficients $\Theta \in R^r$ in the first column and trivial $C = (1, 0, \cdots, 0)$.

In the prediction error framework, Θ is approximated by a (regularised) least squares estimate,

$$
\hat{\Theta} = (\Gamma^T \Gamma + \lambda I)^{-1} \Gamma^T Y_+, \tag{9}
$$

where $\Gamma = (\hat{x}_r, \hat{x}_{r+1}, \cdots, \hat{x}_{N-1})^T$ is a matrix of the respective delay embeddings, $Y_+ = (\hat{y}_{r+1}, \hat{y}_{r+2}, \cdots, \hat{y}_N)^T$, and $\hat{Y} = Y$, $\hat{X} = X$. $\lambda \geq 0$ is the regularisation strength and I is an identity matrix of appropriate size.

We assume throughout that the above inverse exists, for which we require sufficiently large N.

Given C as well as estimates for $\hat{A}, \hat{Y}$, we can then map all available information onto an updated state estimate,

$$
\hat{X} = \mathcal{O}^{-1} \rho \mathcal{C}^T \hat{Y}, \tag{10}
$$

where $\hat{X} = (\hat{x}_1^T \cdots \hat{x}_N^T)^T$ and matrices $\mathcal{O}, \mathcal{C}$ for state reconstruction as defined in [32].

Thus, we propose the alternating minimisation of the loss in (7) outlined in Algorithm 1, which does not require the use of first-order or second-order optimisation methods.

Algorithm 1 AR model identification (given r, Y, C, λ, ρ)

1: initialise $\hat{Y} = Y$ and $\hat{Y}_+ = Y_+$
2: **while** $\hat{A}, \hat{Y}$ not converged **do**
3:     Based on the current estimate $\hat{Y}$, find a new $\hat{A}$ by:
$$\hat{\Theta} = (\Gamma^T \Gamma + \lambda I)^{-1} \Gamma^T \hat{Y}_+$$
4:     Based on $\hat{A}, C$ calculate $\mathcal{O}$
5:     Retrieve a new state estimate based on Y:
$$\hat{X} = \mathcal{O}^{-1} \rho \mathcal{C}^T Y$$
6:     Retrieve updated $\hat{Y}, \hat{Y}_+$ from $\hat{X}$:
$$\hat{Y} = \mathcal{C} \hat{X}$$

The convergence of Algorithm 1 (here for $\lambda = 0$) can be argued as follows. We can associate a loss with Algorithm 1,

$$\ell(\hat{A}, \hat{X}, Y) = \sum_{t=r}^{N-1} \left(\hat{x}_{t+1} - \hat{A}\hat{x}_t \right)^2 + \rho \sum_{t=r}^{N} (y_t - C\hat{x}_t)^2, \tag{11}$$

which is bounded from below by zero. The parameter estimation step results in a new estimate $\hat{\Theta}$ at fixed $\hat{Y}, \hat{X}$. Being the least-squares estimate, it hence minimises the contribution of the first term to the loss, while keeping the second term unchanged. The state estimation step, resulting in a new $\hat{X}$, finds exactly the $\hat{X}$ that minimises ℓ for a given $\hat{A}$. Loss ℓ in (11) is thus a Lyapunov function for the algorithm, which has to converge as a result.

B. The nonlinear finite horizon case

We consider further the case of nonlinear AR models that are nonetheless linear-in-the-parameters. These are models with basis functions such as polynomials [17], [40] and path signatures from rough path theory that have been used in time series applications [41]–[44]. With a vector of basis functions s_t of the embedding x_t, we postulate the state-space map

$$\begin{aligned} s_{t+1} &= As_t + \nu_t, \\ y_t &= Cs_t + \mu_t. \end{aligned} \tag{12}$$

s_t could, for example, be the vector of all polynomials up to a certain order (of the delays $y_r, \cdots, y_{t-r+1}$) that evolves under the dynamics A.

Here, much like in the linear dynamics case, we can find estimates $\hat{A}$ and $\hat{C}$ by least squares for any $\hat{Y}$ and auxiliary matrices $\Gamma_- = (s_r, s_{r+1}, \cdots, s_{N-1})^T$ and $\Gamma_+ = (s_{r+1}, s_{r+2}, \cdots, s_N)^T$ of stacked basis functions. Note that, here, C is not necessarily trivial but defined by a vector that maps the state s_t back to the measurement y_t.

We present a NAR modelling approach with the signature up to a certain depth defining a state-space embedding. A vector of signatures $s_t = s_t^d$ of depth d is extracted from a geometric path of $\hat{x}_t$ in $R^{r \times k}, k > 1$, i.e.,

$$s_t^d = \text{Sig}^d(\hat{x}_t) \in R^{(k^{d+1}-1)/(k-1)}. \tag{13}$$

We can approximate a nonlinear g with a linear function in signature space, provided large enough d, so that

$$\|g_\Theta(\hat{x}_t) - A(s_t^d)\| \leq \epsilon, \tag{14}$$

for any $\epsilon > 0$ [42]. Thus, the linear-in-the-parameters signature model in (12) can represent the nonlinear case (6).

The NAR procedure is outlined in Algorithm 2.

Algorithm 2 NAR model identification (given r, d, Y, λ, ρ)

1: initialise $\hat{Y} = Y$ and $\hat{Y}_+ = Y_+$
2: **while** $\hat{A}, \hat{C}, \hat{Y}$ not converged **do**
3:     Compute all s_t^d from $\hat{Y}$ and form Γ_-, Γ_+
4:     Based on the current estimate $\hat{Y}$, find a new $\hat{A}$:
$$\hat{A} = (\Gamma_-^T \Gamma_- + \lambda I)^{-1} \Gamma_-^T \Gamma_+$$
5:     Based on the current estimate $\hat{Y}$, calculate:
$$\hat{C} = (\Gamma_+^T \Gamma_+ + \lambda I)^{-1} \Gamma_+^T \hat{Y}_+$$
6:     Based on updated $\hat{A}, \hat{C}$ calculate $\mathcal{O}$ and $\mathcal{C}$
7:     Retrieve a new state estimate based on Y:
$$\hat{X} = (\mathcal{O} + \lambda I)^{-1} \rho \mathcal{C}^T Y$$
8:     Retrieve updated $\hat{Y} = \mathcal{C} \hat{X}$

III. EXAMPLES

A. Finding the state dimension r

Many standard techniques exist for the estimation of an approximate state dimension $\hat{r}$. We aim to show that, using our methodology, the following measures provide a good estimate for marginally stable systems like EEG: a) the value of the normalised loss function (11), i.e., $\frac{1}{N}\ell$, and b) the smallest eigenvalue (in magnitude) of the reconstructed system matrix $\hat{A}$, since, for modelling, all eigenvalues are ideally close to the unit disc.

We simulate a stable AR model $x_{t+1} = Ax_t + \nu_t$, $y_t = Cx_t + \mu_t$ with $r_{\text{true}} = 5$, where A is a companion matrix with five eigenvalues of magnitude 1 (the angles of the eigenvalues left of the imaginary axis are $3\pi/5 + i\,\pi/5$, $i = 0, 1, \cdots, 4$), $C = (1, 0, \cdots, 0)$, $x_1 = (1, 0, \cdots, 0)^T$, $N = 200$, and both state and measurement noise i.i.d., $\nu_t \sim N(0, 0.01)$ and $\mu_t \sim N(0, 1)$. We simulate 100 trajectories, and evaluate the above criteria for $\hat{r} = 1, \cdots, 10$ using Algorithm 1 for 100 iterations each, with $\rho = 0.1$ and $\lambda = 0$.

Fig. 1 confirms that the loss ℓ decreases with increasing $\hat{r}$, but plateaus for $\hat{r} > r_{\text{true}}$. We observe additionally that the minimum eigenvalue of the resulting estimate $\hat{A}$ has a peak at the correct embedding size $\hat{r} = r_{\text{true}}$. The latter property holds only after multiple iterations of Algorithm 1: in this example, the standard least squares estimate (i.e., Algorithm 1 after one iteration) gives, in the median, a smallest $\hat{A}$ eigenvalue of 0.514, whereas, after 100 iterations, the median of the smallest eigenvalue is 0.987 (see Fig. 1 at $\hat{r} = 5$). This shows that the algorithm converges towards stable dynamics, as in the simulation, with its eigenvalues closer to the unit disc. This may help the estimation of model order $\hat{r}$, a crucial step in AR modelling.
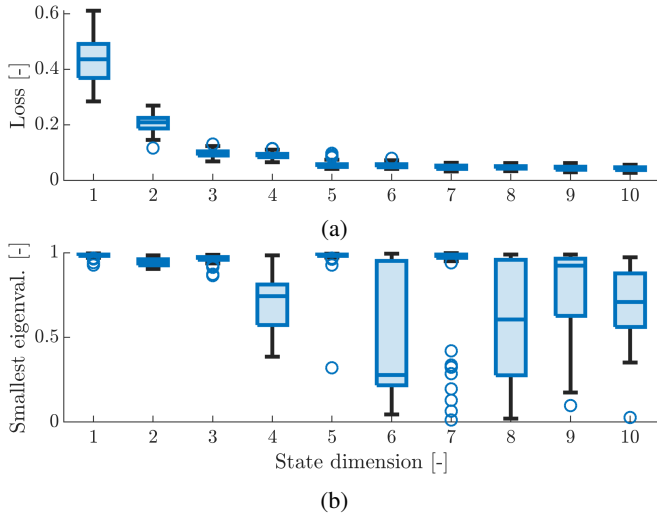
Fig. 1: Boxplots of (a) normalised loss, and (b) smallest $\hat{A}$ eigenvalue over state dimensions $\hat{r} = 1, \cdots, 10$, after 100 iterations of Algorithm 1, for 100 simulations with $r = 5$.

B. Convergence under noise

AR modelling has, in general, no unique solution. Algorithm 1 may thus not always converge to a single global optimum, even in controlled cases with simulated data.

In order to analyse the convergence properties of Algorithm 1, we test the alternating minimisation towards ground truth values from the same model, $x_{t+1} = Ax_t + \nu_t$, $y_t = Cx_t + \mu_t$ with $r = 5$ (A, C, x_1, and N same as above), and same random i.i.d. noise, $\nu_t \sim N(0, 0.01)$ and $\mu_t \sim N(0, 1)$. We simulate 100 trajectories, and run Algorithm 1 for 100 iterations with $\rho = 0.1$ and $\lambda = 0$.

We analyse the error norm $e_{\|\Theta\|} = \|\hat{\Theta} - \Theta\|/\|\Theta\|$ and relative deviation $e_\Theta = (\hat{\Theta} - \Theta)/\|\Theta\|$ in reconstructing an estimate $\hat{\Theta}$, as well as the error $e_x = \|\hat{Y} - x\|/\|x\|$ of the denoised estimate $\hat{Y}$ vs. the true noiseless trajectory $x = (Cx_1, \cdots, Cx_t)^T$. Results are depicted in Fig. 2, with 6 out of 100 iterations not converging to the ground truth A (despite small loss, compare with Fig. 1), but a different local minimum solution.

C. AR EEG connectivity

We employ a version of Algorithm 1 for the estimation of linear EEG inter-channel connectivity (like cross-correlation). To this end, we use seizure 2 and (adjacent) channels $1 - 4$ from intracranial EEG data available online [45]. We take only the first-order finite difference in EEG value of subsequent time steps in order to remove the high autocorrelation, low frequency part of the signal. Take a time horizon of the first 400 time steps, corresponding to one second of EEG. Denote this vector $Y = (y_1, \cdots, y_{400})^T \in R^{400 \times 4}$. See Figure 3 for a plot of these measurements.

We fit a four-dimensional AR model of order 1, i.e., $r = 1$, $p = 4$. Thus, our model for this data is $x_{t+1} = Ax_t + \nu_t$, $y_t = Cx_t + \mu_t$, $t = 1, \cdots, 400$, where $x_t = y_t \in R^4$ are
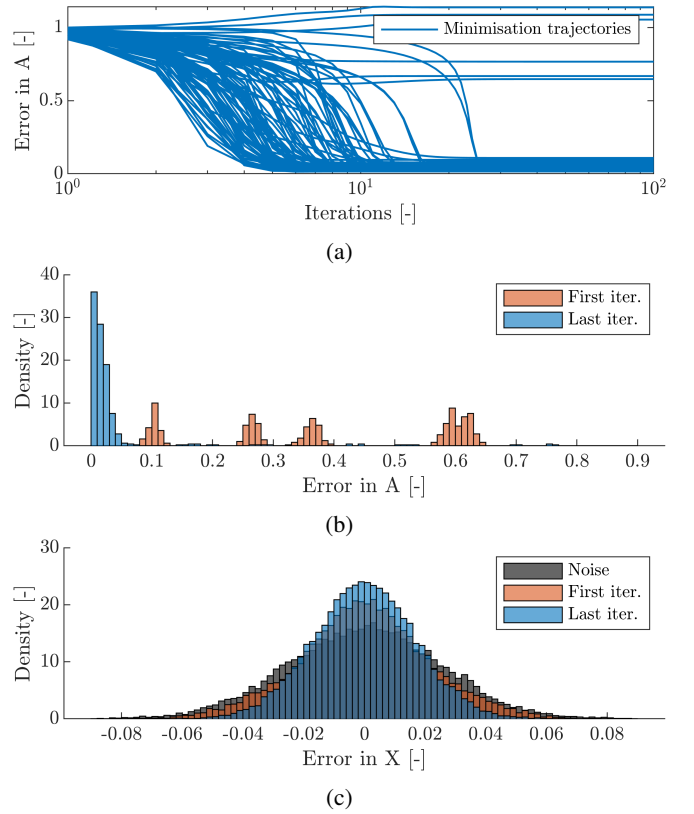

Fig. 2: Algorithm 1 over 100 iterations: a) Convergence to ground truth A as per error norm $e_{\|\Theta\|} = \|\hat{\Theta} - \Theta\|/\|\Theta\|$; b) Error distribution in AR coefficients $e_\Theta = (\hat{\Theta} - \Theta)/\|\Theta\|$ for first and last iteration; c) Error distribution in state reconstruction $e_X = (\hat{Y} - x)/\|x\|$, for measurements $\hat{Y} = Y$ (black), $\hat{Y}$ after 1 iteration (red), and $\hat{Y}$ after 100 iterations (blue). Iteration 1 is the least squares estimate for A.

measurements at time step t only. $\hat{A}$ then contains the 4×4 connectivity terms and $C \in R^{4 \times 4}$ is an identity matrix.

We use $\rho = 1$ and $\lambda = 0$ and iterate 100 times through Algorithm 1. Results are shown in Figs. 3 and 4: The algorithm not only denoises the time series, but also finds a higher connectivity estimate than the standard, least squares estimate without denoising (again, this is effectively the resulting $\hat{A}$ matrix after one iteration).

D. NAR EEG modelling

We demonstrate a NAR case by modelling the EEG from Example III-C with the signature approach in Algorithm 2 and an artificially introduced artefact in the data. To this end, we *replace* the EEG in channel 1 of seizure 2 from time steps 150 to 250 (0.25 seconds) with random white noise with standard deviation of 20 μV. We then use the same first-order finite difference of the signal, Y_{train}, for modelling, corresponding to the first 400 time steps with artefact. We use Y_{test}, corresponding to the second 400 time steps without artificial artefacts, for a test of the model. See Figs. 5 and 6. The task is to find a good model from Y_{train}, despite the artefact, and measure its performance on predictions in Y_{test}.
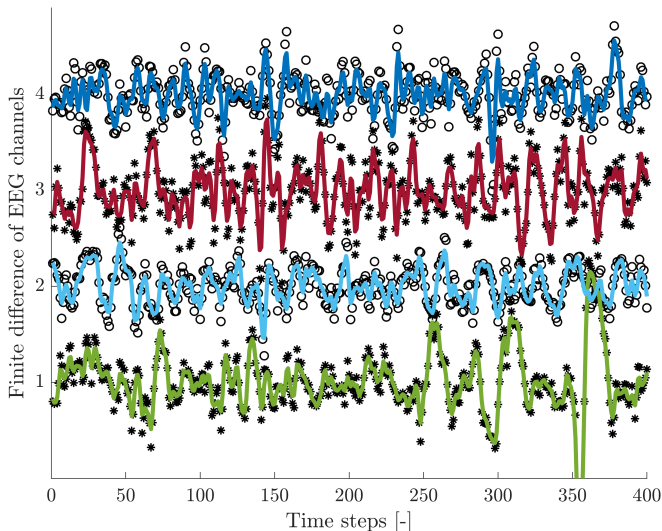
Fig. 3: Qualitative plot of the first-order difference in EEG channels, Y, (black) and the denoised estimate $\hat{Y}$ for each channel (coloured lines) after 100 iterations of Algorithm 1.
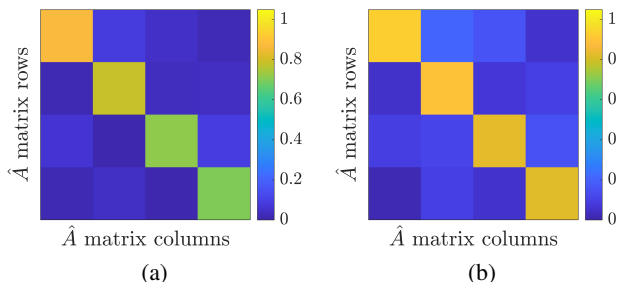

(a)   (b)

Fig. 4: Absolute values $|\hat{A}|$ as per colorbar: a) least-squares solution after 1 iteration; b) after 100 iterations of Algorithm 1. Off-diagonal terms depict inter-channel connectivity.

We use the signature implementation *esig* [43] in Python, and parameters $r = 4, k = 2$, and depth $d = 2$. Note that higher embedding dimensions r at fixed d lead here to a more conservative denoising. To calculate $s_t^d \in R^7$, we use the geometric path in $R^{4 \times 2}$ with $\hat{x}_t$ in the first column, and its cumulative sum $(\hat{y}_{t-r+1}, \hat{y}_{t-r+1} + \hat{y}_{t-r+2}, \cdots)^T$ in the second. We apply Algorithm 2 for 20 steps with $\rho = 0.1$ and a small regularisation $\lambda = 0.001$ on states and parameters.

First, we show that Algorithm 2 denoises the training data, i.e., ignores the artefact successfully (Fig. 5). Second, to elucidate the ability of the algorithm to find a reasonable model $\hat{A}, \hat{C}$, we run the model on the test set Y_{test} in a one-step-ahead prediction: $\hat{y}_{t+1} = \hat{C}\hat{A}\hat{s}_t, \ t = 400, \cdots, 799$. We illustrate the differences in $\hat{A}$ by plotting these predictions against the true Y_{test}, for the first and last iteration of Algorithm 2 (Fig. 6). The first iteration, i.e., the classical least squares estimate for $\hat{A}, \hat{C}$ retrieved from signatures of the original signal Y_{train} with artefact, has a mean-squared-error of $e_1 = 82 \ \mu V$. The model in the last iteration, i.e., the model from signatures calculated on the denoised signal shown in Fig. 5, has a mean-squared-error of $e_{20} = 31 \ \mu V$.
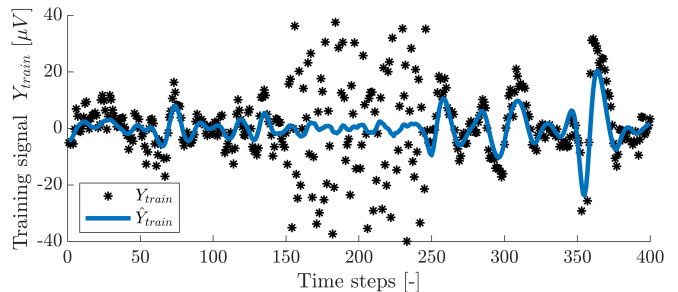

Fig. 5: First-order difference in EEG channel 1 with random artefact, Y_{train} (black stars), and the denoised time series estimate, $\hat{Y}_{\text{train}}$ (blue line), after 20 iterations of Algorithm 2.
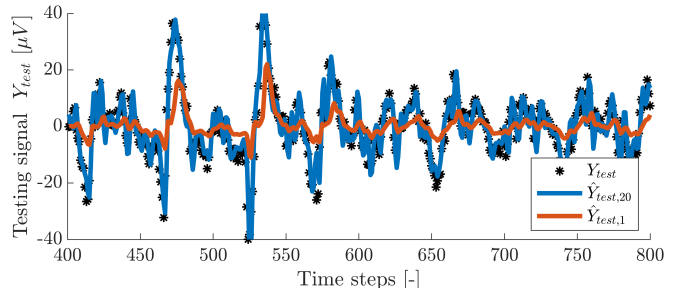

Fig. 6: Plot of measurements $\hat{Y}_{\text{test}}$ (black stars) against the one-step-ahead predictions $\hat{Y}_{\text{test},1}$ and $\hat{Y}_{\text{test},20}$, after 1 (red) and 20 (blue) iterations of Algorithm 2, respectively.

IV. DISCUSSION AND CONCLUSION

We have identified a procedure for estimating autoregressive models under measurement and state transition noise, as is typically encountered in biomedical signal processing. Assumptions on both these sources of uncertainty lead to a loss function that can be optimised via an iterative algorithm alternating between estimating model parameters and model states. The described method is essentially an incarnation of the expectation-maximisation algorithm. Our assumptions on the model type, i.e., autoregressive models, made it possible to derive analytic expressions for both the parameter and state estimation steps in a novel way. This alleviates substantial complexity, since this batch estimation requires no gradient-based optimisation, filtering or sampling techniques.

Examples depicting the theoretical properties as well as performance on real-world EEG data were presented and analysed. Our results indicate that the proposed methodology is able to denoise time series and recover autoregressive coefficients accurately. This might for example provide new ways to analyse the connectivity of brain regions under the inherent sources of noise in EEG data.

Another advantage of our procedure is that it works for linear and nonlinear autoregressive models alike, as long as the parameter estimation step itself is a linear problem. We note that a very broad class of autoregressive models exhibit this property, and have shown an approach using path signatures as basis functions to model arbitrary nonlinear systems. This makes the presented methodology widely applicable to tasks in biomedical signal processing and beyond.

References

[1] R. E. Greenblatt, M. E. Pflieger, and A. E. Ossadtchi, "Connectivity measures applied to human brain electrophysiological data," *Journal of Neuroscience Methods*, vol. 207, no. 1, pp. 1–16, 2012.

[2] P. Shah, A. Ashourvan, F. Mikhail, A. Pines, L. Kini, K. Oechsel, S. R. Das, J. M. Stein, R. T. Shinohara, D. S. Bassett, B. Litt, and K. A. Davis, "Characterizing the role of the structural connectome in seizure dynamics," *Brain*, vol. 142, no. 7, pp. 1955–1972, 2019.

[3] A. Li, C. Huynh, Z. Fitzgerald, I. Cajigas, D. Brusko, J. Jagid, A. O. Claudio, A. M. Kanner, J. Hopp, S. Chen, J. Haagensen, E. Johnson, W. Anderson, N. Crone, S. Inati, K. A. Zaghloul, J. Bulacio, J. Gonzalez-Martinez, and S. V. Sarma, "Neural fragility as an EEG marker of the seizure onset zone," *Nature Neuroscience*, vol. 24, no. 10, pp. 1465–1474, 2021.

[4] A. F. Brogin, J. Faber, and D. D. Bueno, "Epileptic seizure suppression: a computational approach for identification and control using real data," bioRxiv 2023.01.08.522971, 2023.

[5] J. Wolpaw and E. W. Wolpaw, *Brain–Computer InterfacesPrinciples and Practice*. Oxford University Press, 1 2012.

[6] J. Meng, S. Zhang, A. Bekyo, J. Olsoe, B. Baxter, and B. He, "Noninvasive Electroencephalogram Based Control of a Robotic Arm for Reach and Grasp Tasks," *Scientific Reports*, vol. 6, 2016.

[7] N. Padfield, J. Zabalza, H. Zhao, V. Masero, and J. Ren, "EEG-Based Brain-Computer Interfaces Using Motor-Imagery: Techniques and Challenges," *Sensors*, vol. 19, p. 1423, 3 2019.

[8] L. Ljung, *System Identification Theory for User*. Prentice Hall PTR, 1987.

[9] S. A. Billings, *Nonlinear System Identification*. Chichester, UK: John Wiley & Sons, Ltd, 7 2013.

[10] R. Pintelon and J. Schoukens, *System Identification*. Hoboken, NJ, USA: John Wiley & Sons, Inc., 3 2012.

[11] F. Takens, "Detecting strange attractors in turbulence," in *Lecture Notes in Mathematics*, vol. 898, pp. 366–381, 1981.

[12] J. Schoukens and L. Ljung, "Nonlinear System Identification: A User-Oriented Road Map," *IEEE Control Systems*, vol. 39, no. 6, pp. 28–99, 2019.

[13] L. Piroddi and W. Spinelli, "An identification algorithm for polynomial NARX models based on simulation error minimization," *International Journal of Control*, vol. 76, pp. 1767–1781, 11 2003.

[14] L. A. Aguirre, B. H. Barbosa, and A. P. Braga, "Prediction and simulation errors in parameter estimation for nonlinear systems," *Mechanical Systems and Signal Processing*, vol. 24, no. 8, pp. 2855–2867, 2010.

[15] J. Homer and I. Mareels, "LS detection guided NLMS estimation of sparse systems," *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, vol. 2, no. May, 2004.

[16] G. Tran and R. Ward, "Exact Recovery of Chaotic Systems from Highly Corrupted Data," *Multiscale Modeling & Simulation*, vol. 15, pp. 1108–1129, 1 2017.

[17] S. L. Brunton, J. L. Proctor, J. N. Kutz, and W. Bialek, "Discovering governing equations from data by sparse identification of nonlinear dynamical systems," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 113, no. 15, pp. 3932–3937, 2016.

[18] L. C. Iacob, G. I. Beintema, M. Schoukens, R. Tóth, and R. Toth, "Deep Identification of Nonlinear Systems in Koopman Form," *Proceedings of the IEEE Conference on Decision and Control*, vol. 2021-Decem, pp. 2288–2293, 2021.

[19] S. L. Brunton, M. Budišić, E. Kaiser, and J. N. Kutz, "Modern Koopman Theory for Dynamical Systems," *SIAM Review*, vol. 64, pp. 229–340, 5 2022.

[20] M. Sznaier, "A Convex Optimization Approach to Learning Koopman Operators," *arXiv:2102.03934*.

[21] K. U. Leuven, P. Van Overschee, and B. De Moor, "Subspace algorithms for the stochastic identification problem," *Automatica*, vol. 29, no. 3, pp. 649–660, 1993.

[22] L. Lu, Y. Tan, D. Oetomo, I. Mareels, E. Zhao, and S. An, "On Model-guided Neural Networks for System Identification," in *2019 IEEE Symposium Series on Computational Intelligence (SSCI)*, pp. 610–616, IEEE, 12 2019.

[23] L. Ljung, T. Chen, and B. Mu, "A shift in paradigm for system identification," *International Journal of Control*, vol. 93, no. 2, pp. 173–180, 2020.

[24] L. Ljung, C. Andersson, K. Tiels, and T. B. Schön, "Deep learning and system identification," *IFAC-PapersOnLine*, vol. 53, no. 2, pp. 1175–1181, 2020.

[25] H. U. Voss, J. Timmer, and J. Kurths, "Nonlinear dynamical system identification from uncertain and indirect measurements," *International Journal of Bifurcation and Chaos in Applied Sciences and Engineering*, vol. 14, no. 6, pp. 1905–1933, 2004.

[26] Q. Zhang, "Nonlinear system identification with output error model through stabilized simulation," *IFAC Proceedings Volumes (IFAC-PapersOnline)*, vol. 37, no. 13, pp. 501–506, 2004.

[27] J. Decuyper, M. Runacres, J. Schoukens, and K. Tiels, "Tuning nonlinear state-space models using unconstrained multiple shooting," *IFAC-PapersOnLine*, vol. 53, no. 2, pp. 334–340, 2020.

[28] M. Farina and L. Piroddi, "Simulation error minimization identification based on multi-stage prediction," *International Journal of Adaptive Control and Signal Processing*, vol. 25, pp. 389–406, 5 2011.

[29] M. Forgione and D. Piga, "Model Structures and Fitting Criteria for System Identification with Neural Networks," *14th IEEE International Conference on Application of Information and Communication Technologies, AICT 2020 - Proceedings*, pp. 1–16, 2020.

[30] A. Schlaginhaufen, P. Wenk, A. Krause, and F. Dörfler, "Learning Stable Deep Dynamics Models for Partially Observed or Delayed Dynamical Systems," *Advances in Neural Information Processing Systems*, vol. 15, no. NeurIPS, pp. 11870–11882, 2021.

[31] I. Mareels, "Recovering state trajectories from output measurements and dynamic models: A computational complexity point of view," *International Journal of Bifurcation and Chaos in Applied Sciences and Engineering*, vol. 12, no. 5, pp. 1079–1095, 2002.

[32] J. F. Haderlein, I. M. Y. Mareels, A. D. H. Peterson, P. Z. Eskikand, A. N. Burkitt, and D. B. Grayden, "On the benefit of overparameterization in state reconstruction," *2021 60th IEEE Conference on Decision and Control (CDC)*, pp. 1580–1585, 12 2021.

[33] J. Van Gorp, J. Schoukens, and R. Pintelon, "Learning neural networks with noisy inputs using the errors-in-variables approach," *IEEE Transactions on Neural Networks*, vol. 11, no. 2, pp. 402–414, 2000.

[34] M. West, "Modelling and robustness issues in Bayesian time series analysis," *Lecture Notes-Monograph Series Vol. 29, Bayesian Robustness (1996), pp. 231-251*.

[35] A. Pole, M. West, and J. Harrison, *Applied Bayesian Forecasting and Time Series Analysis*. Chapman and Hall/CRC, 1994.

[36] A. Gelman, J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin, *Bayesian Data Analysis Third edition*. No. February, Taylor & Francis, 2020.

[37] R. H. Shumway and D. S. Stoffer, "An Approach to Time Series Smoothing and Forecasting using the EM Algorithm," *Journal of Time Series Analysis*, vol. 3, pp. 253–264, 7 1982.

[38] Z. Ghahramani and S. T. Roweis, "Learning nonlinear dynamical systems using an EM algorithm," in *Advances in Neural Information Processing Systems*, pp. 431–437, 1999.

[39] S. Roweis and Z. Ghahramani, "A unifying review of linear gaussian models," *Neural Computation*, vol. 11, no. 2, pp. 305–345, 1999.

[40] J. Decuyper, K. Tiels, M. C. Runacres, and J. Schoukens, "Retrieving highly structured models starting from black-box nonlinear state-space models using polynomial decoupling," *Mechanical Systems and Signal Processing*, vol. 146, p. 106966, 2020.

[41] T. J. Lyons, "Differential equations driven by rough signals," *Revista Matematica Iberoamericana*, vol. 14, no. 2, pp. 215–310, 1998.

[42] D. Levin, T. Lyons, and H. Ni, "Learning from the past, predicting the statistics for the future, learning an evolving system," *arXiv:1309.0260*.

[43] T. Lyons, "Rough paths, signatures and the modelling of functions on streams," in *Proceeding of the International Congress of Mathematicans, ICM 2014*, vol. 4, pp. 163–184, 2014.

[44] I. Chevyrev, A. Gerasimovics, and H. Weber, "Feature Engineering with Regularity Structures," *SSRN Electronic Journal*, 8 2021.

[45] P. J. Karoly, L. Kuhlmann, D. Soudry, D. B. Grayden, M. J. Cook, and D. R. Freestone, "Seizure pathways: A model-based investigation," *PLoS Computational Biology*, vol. 14, no. 10, pp. 1–24, 2018.