2023 62nd IEEE Conference on Decision and Control (CDC)
December 13-15, 2023. Marina Bay Sands, Singapore

Adaptive Estimation of Time-Varying Parameters using DREM

Emil Lykke Diget and Christoffer Sloth

***Abstract*— In this paper we present a method for estimating time-varying parameters in a linear regression equation. We combine local polynomial regression with dynamic regressor extension and mixing to independently estimate the parameters. During local polynomial regression, a time-varying parameter is approximated by locally constant polynomial coefficients. We propose to use the Bernstein basis instead of the commonly used monomial basis to improve numerical conditioning. A simulation example shows that our proposed estimator has improved performance compared to a similar method and allows a higher polynomial order.**

I. Introduction

Parameter estimation plays an important role in adaptive control, as many practical applications involve time-varying parameters. In safety-critical applications such as the remote maintenance of fusion reactors, it is pivotal to estimate the parameters to guarantee overall system safety.

A common assumption of gradient-based and least-squares parameter estimators is that the parameters are quasi-static [1]. This can be a limitation in practice when dealing with time-varying parameters. To overcome this issue, one can fit a local polynomial to the parameter estimate. Thus, the parameter goes from being time-varying to locally constant. Locally weighted polynomial regression was carried out in [2] for different scenarios, where different types of regression, e.g., polynomials of different orders, were applied to a local neighbourhood. Local polynomial regression was combined with the recursive least-squares (RLS) algorithm in the parameter space [3] and in the time domain [4]. In the papers [5], [6] time was divided into small sections, and in each window the parameters were fitted by a local polynomial with constant parameters.

A further limitation of the standard gradient-based parameter estimator is that it requires the restrictive condition of persistence of excitation (PE) to converge. In [7] and [8] linear filters were used to modify the stability and convergence properties of the linear regression equation. A new method of designing parameter estimators called dynamic regressor extension and mixing (DREM) was introduced in [9]. A key feature of DREM is the guaranteed performance increase compared with the gradient-based parameter estimator and the least-squared parameter estimator, [1]. Ortega explains in [10] that DREM can be interpreted as two different Luenberger observers with a gradient descent. The dynamics of DREM depends on the choice of the filters. I-DREM, introduced in [11], relaxes the square integrability condition of DREM by using an integral-based filter. The same authors used the method to estimate time-varying parameters using a first order approximation in [12].

An issue with most polynomial regression methods [3], [4], [5], [6], [12] is that they rely on the monomial polynomial basis, which can often be numerically ill-conditioned in practical applications [13]. An alternative to the monomial basis is the Bernstein basis [14], that has had great use in the area of computer graphics, see e.g., [15] for details. The Bernstein basis was compared with the monomial basis in [13], and it was found that if one has the choice between the two, one should always use the Bernstein basis, due to it being more numerically robust against floating point errors.

In this paper we show how to combine local polynomial regression and DREM to enable estimation of time-varying parameters, and that the use of the Bernstein basis, in comparison with the monomial basis, increases the numerical stability when the polynomial order is increased.

The rest of the paper is organised as follows. Section II describes two methods for estimation of constant parameters, while Section III describes two methods for estimation of time-varying parameters. Our contributions are presented in Section IV and Section V. Numerical simulations are presented in Section VI and the conclusion is given in Section VII.

II. Estimation of Constant Parameters

In this section, two methods for estimating constant parameters in a linear regression equation (LRE) is presented. The LRE is defined as:

$$y(t) = \phi^{\mathsf{T}}(t)\theta + w(t), \tag{1}$$

where $y : \mathbb{R}_+ \to \mathbb{R}$ and $\phi : \mathbb{R}_+ \to \mathbb{R}^p$ are the known bounded output and input functions of time, respectively, $\theta \in \mathbb{R}^p$ is the unknown constant parameter vector, $w : \mathbb{R}_+ \to \mathbb{R}$ is the measurement noise, and p is the number of parameters.

The standard gradient-based parameter estimator is described in Section II-A, and DREM is described in Section II-B.

A. Gradient-Based Parameter Estimation

This section presents the gradient-based parameter estimator defined in [1]. The parameter θ can be estimated using:

$$\dot{\hat{\theta}}(t) = \gamma\phi(t)\left[y(t) - \phi^{\mathsf{T}}(t)\hat{\theta}(t)\right], \tag{2}$$

where $\hat{\theta}(t)$ is the estimate of θ, and $\gamma > 0$ is a tuning parameter. The dynamics of the parameter estimation error is given as:

$$\dot{\tilde{\theta}}(t) = -\gamma\phi(t)\phi^{\mathsf{T}}(t)\tilde{\theta}(t), \tag{3}$$

Emil Lykke Diget and Christoffer Sloth are with the Maersk Mc-Kinney Moller Institute, University of Southern Denmark, Denmark {eld,chsl}@mmmi.sdu.dk.

where $\tilde{\theta}(t) := \theta - \hat{\theta}(t)$. It is seen that γ affects the rate of convergence, and that $\phi(t)\phi^{\mathsf{T}}(t)$ also affects the error dynamics; in a way that can be quantified via persistence of excitation defined as follows:

Definition 1 (Persistency of Excitation (PE), Ch. 2 in [1])**.** A function $\phi : \mathbb{R}_+ \to \mathbb{R}^p$ is *persistently exciting* (PE) if there exists μ_1, μ_2, $\delta > 0$ such that

$$\mu_2 I_p \geq \int_t^{t+\delta} \phi(\tau)\phi^{\mathsf{T}}(\tau)\,\mathrm{d}\tau \geq \mu_1 I_p \tag{4}$$

for all $t \geq 0$, where I_p is the $p \times p$ identity matrix.

Proposition 1 (According to [1])**.** The estimator (2) has the following properties:

1) The error function given by (3).
2) The Euclidean norm of $\tilde{\theta}(t)$ is monotonically non-increasing.
3) The following equivalence holds, [1, Theorem 2.5.1]:

$$\lim_{t\to\infty} \left\|\tilde{\theta}(t)\right\| = 0 \;\Leftarrow\; \phi(t) \in \text{PE}, \tag{5}$$

where $\|\cdot\|$ is the Euclidean norm.
4) If $\phi(t) \in$ PE, then the estimated parameter $\hat{\theta}(t)$ converges exponentially to the nominal parameter θ, [1, Theorem 2.5.3].

B. Dynamic Regressor Extension and Mixing

In this section a method of relaxing the restrictive condition of PE is presented. The method is called DREM [9], [16]. DREM converts (1) to p, one-dimensional LREs to independently estimate each of the parameters. A linear, single-input p-output, bounded-input bounded-output operator (BIBO) $\mathcal{H}$ is introduced to reformulate (1) as:

$$Y(t) = \varphi(t)\theta, \tag{6}$$

where

$$Y(t) := \mathcal{H}[y(t)] \in \mathbb{R}^p, \qquad \varphi(t) := \mathcal{H}[\phi^{\mathsf{T}}(t)] \in \mathbb{R}^{p\times p}. \tag{7}$$

Different choices can be made on the operator $\mathcal{H}$ [10], [17], [16]. Recall that for any matrix $M \in \mathbb{R}^{m\times m}$ we have $\operatorname{adj}\{M\}M = \det\{M\}I_m$, where $\operatorname{adj}\{\cdot\}$ is the adjugate, $\det\{\cdot\}$ is the determinant, and $I_m \in \mathbb{R}^{m\times m}$ is the identity matrix. Left-multiplying (6) with the adjugate of $\varphi(t)$ results in the following scalar equations:

$$\mathcal{Y}_i(t) = \Delta(t)\theta_i, \quad i = 1, 2, \ldots, p, \tag{8}$$

where

$$\begin{aligned} \Delta(t) &:= \det\{\varphi(t)\} \in \mathbb{R}, \\ \mathcal{Y}(t) &:= \operatorname{adj}\{\varphi(t)\}Y(t) \in \mathbb{R}^p. \end{aligned} \tag{9}$$

Then, each parameter θ_i can be estimated individually by a gradient-based parameter estimator:

$$\dot{\hat{\theta}}_i(t) = \gamma_i \Delta(t)\left[\mathcal{Y}_i(t) - \Delta(t)\hat{\theta}_i(t)\right], \tag{10}$$

where $\gamma_i > 0$ is a tuning parameter.

Proposition 2 (We reformulate from [16])**.** The estimator (10) has the following properties:

1) The error function given by:

$$\dot{\tilde{\theta}}_i(t) = -\gamma_i \Delta^2(t)\tilde{\theta}_i(t), \tag{11}$$

where $\tilde{\theta}_i(t) := \theta_i - \hat{\theta}_i(t)$.
2) The absolute value of the *individual* parameter errors $\tilde{\theta}_i(t)$ are monotonically non-increasing.
3) The following equivalence holds:

$$\lim_{t\to\infty} \left|\tilde{\theta}_i(t)\right| = 0 \;\Leftrightarrow\; \Delta(t) \notin \mathcal{L}_2, \tag{12}$$

where $|\cdot|$ is the absolute value.
4) If $\Delta(t) \in$ PE, then the convergence is exponential.

III. Estimation of Time-Varying Parameters

In this section two methods for estimating time-varying parameters are presented. The LRE is defined as:

$$y(t) = \phi^{\mathsf{T}}(t)\theta(t) + w(t), \tag{13}$$

where $y : \mathbb{R}_+ \to \mathbb{R}$ and $\phi : \mathbb{R}_+ \to \mathbb{R}^p$ are the known bounded output and input functions of time, respectively, $\theta : \mathbb{R}_+ \to \mathbb{R}^p$ is the unknown time-varying parameter vector, $w : \mathbb{R}_+ \to \mathbb{R}$ is the measurement noise, and p is the number of parameters.

A method for local polynomial regression using the monomial polynomial basis is described in Section III-A, and a method from literature combining I-DREM and first order local polynomial regression is briefly described in Section III-B.

A. Local Polynomial Regression

The gradient-based parameter estimator and DREM assumes that the parameter vector θ is constant, though in many practical applications, this is not the case. The idea of local polynomial regression is to approximate the parameter locally using a polynomial. Thus, the parameter estimation problem goes from being time-varying to locally constant [4], [5], [6].

To represent time-varying parameters, time is divided into segments of width T which begins at t_0 given by [6]:

$$t_0(t) = T \cdot \left\lfloor \frac{t}{T} \right\rfloor, \tag{14}$$

where $\lfloor\cdot\rfloor$ is the floor-function, which rounds down to the nearest integer. Equation (14) defines a non-decreasing sequence of time instants $t_0 = [t_{0,j}], j = 0, 1, \ldots,$ where the difference between each t_0 instant is T. In the following the function argument t is omitted for brevity.

It is well known that a smooth function can be represented by a Taylor series around a point t_0. Assuming that the parameter elements $\theta_i(t)$ are smooth, the approximation can be applied in each window $t \in [t_0, t_0 + T)$:

$$\theta_i(t) = L_i(t, t_0)\alpha_i(t_0) + \epsilon_i, \tag{15}$$

where ϵ_i is the Lagrange error, and

$$L_i(t, t_0) = \begin{bmatrix} 1 & t - t_0 & (t-t_0)^2 & \ldots & (t-t_0)^{k_i} \end{bmatrix}, \tag{16}$$

$$\alpha_i(t_0) = \begin{bmatrix} a_{i,0}(t_0) & a_{i,1}(t_0) & \ldots & a_{i,k_i}(t_0) \end{bmatrix}^{\mathsf{T}}, \tag{17}$$

and for all parameters, where ϵ_i is disregarded:

$$\theta(t) = L(t, t_0)\alpha(t_0), \tag{18}$$

where $K = \sum_{i=1}^{p} k_i$ and

$$L(t, t_0) = \begin{bmatrix} L_1(t, t_0) & & \\ & \ddots & \\ & & L_p(t, t_0) \end{bmatrix} \in \mathbb{R}^{p\times(p+K)}, \tag{19}$$

$$\alpha(t_0) = \begin{bmatrix} \alpha_0^\intercal(t_0) & \alpha_1^\intercal(t_0) & \dots & \alpha_{k_i}^\intercal(t_0) \end{bmatrix}^\intercal \in \mathbb{R}^{p+K}. \tag{20}$$

By inserting (18) into (13):

$$y(t) = \phi^\intercal(t)\theta(t) = \underbrace{\phi^\intercal(t)L(t, t_0)}_{:=\Phi^\intercal(t,t_0)} \alpha(t_0), \tag{21}$$

where $\alpha(t_0)$ can be estimated by $\hat{\alpha}(t)$. The vector of parameters $\alpha(t_0)$ is constant in each time window $t \in [t_0, t_0 + T)$, and $\theta(t)$ is varying with time in the same window. The parameter estimate $\hat{\theta}$ can be found from $\hat{\alpha}$ using (18):

$$\hat{\theta}(t) = L(t, t_0)\hat{\alpha}(t). \tag{22}$$

To make sure the parameter estimate $\hat{\theta}$ is continuous across time windows, $\hat{\alpha}$ should be reset in the beginning of each time window:

$$\hat{\alpha}(t_0 + T) = X(t_0 + T, t_0)\hat{\alpha}((t_0 + T)^-), \tag{23}$$

for $t = t_0 + T$, where $(t_0 + T)^-$ is just before the resetting point and $X(t_0 + T, t_0) \in \mathbb{R}^{(p+K)\times(p+K)}$ is the resetting matrix, see [5], [6] for details.

B. Integral Dynamic Regressor Extension and Mixing

In this section the integral dynamic regressor extension and mixing (I-DREM) is introduced in short. It was recently introduced in [11] to relax the square integrability condition ($\Delta(t) \notin \mathcal{L}_2$) present in ordinary DREM. The same authors used I-DREM to estimate time-varying parameters in [12]. Here, the authors combine a first order local polynomial regression, like Section III-A, with I-DREM.

The update rule in [12] is conditional on the value of $\Delta(t)$; it switches between an I-DREM estimator and a gradient-based parameter estimator. With the symbols of this paper, their update rule is:

$$\dot{\hat{\theta}}(t) = \begin{cases} -\frac{\gamma_0}{\Delta^2(t)}\Delta(t)\left[\Delta(t)\hat{\theta}(t) - \Upsilon(t)\right] & \text{if } \Delta(t) \geq \kappa, \\ -\Gamma\phi(t)\left[\hat{\theta}(t)\phi(t) - y(t)\right]^\intercal - & \\ \quad \sigma\Gamma\hat{\theta}(t) & \text{otherwise}, \end{cases} \tag{24}$$

where $\Upsilon(t) = L(0, 0)\mathcal{Y}(t)$, Γ is a gain matrix, κ is a threshold, and σ is a forgetting factor, see [12] for more details.

The main point of concern is the conditional nature of the update rule, where the system behaviour depends heavily on κ. Based on experience, the value of Δ can be quite small, see Section VI, so κ can be difficult to tune.

In this paper we aim at providing another update rule, such that switching is not required, and the order of the local polynomial approximation can be increased beyond one.

IV. Dynamic Regressor Extension and Mixing with Local Polynomial Regression

We propose to combine local polynomial regression, Section III-A with the DREM procedure, Section II-B, to obtain a method that benefits from both. Following [18], we apply Kreisselmeier's regressor extension, [8], as the operator $\mathcal{H}$ of the form

$$\dot{x}(t) = -\ell x(t) + \Phi^\intercal(t, t_0)u(t),$$
$$y_u(t) = x(t),$$

where $x(t) \in \mathbb{R}^n$ is the internal state vector, and $\ell > 0$ is a tuning parameter. Applying Kreisselmeier's regressor extension to $\Phi(t, t_0)$ and $y(t)$ from (21) as $\mathcal{H}$ in (7), results in

$$\begin{aligned} \dot{\varphi}_{\text{pol}} &= -\ell\varphi_{\text{pol}} + \Phi^\intercal(t, t_0)\Phi(t, t_0), & \varphi_{\text{pol}}(0) &= 0, \\ \dot{Y}_{\text{pol}} &= -\ell Y_{\text{pol}} + \Phi^\intercal(t, t_0)y(t), & Y_{\text{pol}}(0) &= 0, \end{aligned} \tag{25}$$

giving a polynomial version of (6):

$$Y_{\text{pol}}(t) = \varphi_{\text{pol}}\alpha(t_0). \tag{26}$$

The DREM procedure is applied to the above:

$$\mathcal{Y}_{\text{pol,i}}(t) = \Delta_{\text{pol}}(t)\alpha_i(t_0), \tag{27}$$

where

$$\begin{aligned} \Delta_{\text{pol}}(t) &:= \det\{\varphi_{\text{pol}}(t)\} \in \mathbb{R}, \\ \mathcal{Y}_{\text{pol}}(t) &:= \text{adj}\{\varphi_{\text{pol}}(t)\}Y_{\text{pol}}(t) \in \mathbb{R}^{p+K}, \end{aligned} \tag{28}$$

which can be individually estimated using the following least squares parameter adaption scheme [1], [18]:

$$\begin{aligned} \dot{\hat{\alpha}}_i(t) &= \gamma_i\Delta_{\text{pol}}(t)p_i(t)\left[\mathcal{Y}_{\text{pol},i}(t) - \Delta_{\text{pol}}(t)\hat{\alpha}_i(t)\right], \\ \dot{p}_i(t) &= \gamma_i(\lambda_i p_i(t) - p_i^2(t)\Delta_{\text{pol}}^2(t)), \end{aligned} \tag{29}$$

where $p_i(0) > 0$, and $\gamma_i > 0$ are the design parameters, and λ_i is the forgetting factor. The parameter estimator (29) yields the following error dynamics:

$$\dot{\tilde{\alpha}}_i(t) = -\gamma_i\Delta_{\text{pol}}^2(t)p_i(t)\tilde{\alpha}_i(t), \tag{30}$$

where $\tilde{\alpha}_i(t) = \alpha_i(t_0) - \hat{\alpha}_i(t)$.

The estimated parameter vector $\hat{\theta}$ can be extracted as:

$$\hat{\theta}(t) = L(t, t_0)\hat{\alpha}(t), \qquad t \in [t_0, t_0 + T), \tag{31}$$

and $\hat{\alpha}$ has to be reset in the beginning of each time window using the reset matrix X, [5], [6]:

$$\hat{\alpha}(t_0 + T) = X(t_0 + T, t_0)\hat{\alpha}((t_0 + T)^-). \tag{32}$$

To compensate for the resetting of $\hat{\alpha}$, the filtered dynamics (26) should be reset accordingly at the beginning of each time window:

$$\varphi_{\text{pol}}(t_0 + T) = \varphi_{\text{pol}}((t_0 + T)^-)X^{-1}(t_0 + T, t_0), \tag{33}$$

which is not something that is considered in [6].

Proposition 3 (Proposition 2 in [18])**.** Let i be the parameter index. Consider the estimation algorithm (29) with $p_i(0) > 0$ and $\gamma_i > 0$.

1) If $\lambda_i = 0$ then

a) if $\Delta_{\text{pol}} \in \mathcal{L}_2$ then for all nonzero $\tilde{\alpha}_i(0)$ the signal $\tilde{\alpha}_i$ does not converge to zero;
b) if $\Delta_{\text{pol}} \notin \mathcal{L}_2$ then $\tilde{\alpha}_i$ is monotonic and converges to zero asymptotically;
c) if $\Delta_{\text{pol}} \in$ PE, it does not imply exponential convergence.

2) If $\lambda_i > 0$ the
a) p_i is bounded from below;
b) if $\Delta_{\text{pol}} \in \mathcal{L}_2$ or

$$\Delta_{\text{pol}} \notin \mathcal{L}_2 \text{ and } \Delta_{\text{pol}} \to 0, \tag{34}$$

then the estimator is unstable and p_i tends to infinity;
c) if $\Delta_{\text{pol}} \in$ PE the p_i is bounded, $\tilde{\alpha}_i$ is monotonic and converges to zero exponentially fast.

The proof is given in [18].

V. Bernstein Polynomial Basis

It is observed in Section VI that using the polynomial DREM estimator with the monomial basis has numerical issue; therefore, the more numerically robust Bernstein basis is presented in this section. The Bernstein basis has always a better numerical condition than the conventional monomial basis [13]. This leads to better numerical stability against arithmetic round-off error when the method is implemented. The Bernstein basis was first introduced in [14], and is given by:

$$\mathfrak{B}_{\nu,n}(t) = \binom{n}{\nu}\frac{(b-t)^{n-\nu}(t-a)^{\nu}}{(b-a)^n}, \quad \nu = 0,\dots,n, \tag{35}$$

where n is the order, and $t \in [a, b]$. In our case where time is divided into windows defined by (14), $a = t_0$ and $b = t_0 + T$.

The Bernstein basis $\mathfrak{B}_{\nu,n}(t)$ fulfills the properties of positivity and partition of unity [13]:

$$\begin{aligned} \mathfrak{B}_{\nu,n}(t) \geq 0, \qquad & \forall t \in [a, b] \\ \text{and } \sum_{\nu=0}^{n} \mathfrak{B}_{\nu,n}(t) = 1, \qquad & \nu = 0,\dots,n. \end{aligned} \tag{36}$$

The time-varying parameter $\theta(t)$ can be approximated as:

$$\theta_i(t) = L_i^{\mathsf{T}}(t,t_0)\alpha_i(t_0) + \epsilon_i(t), \tag{37}$$

where

$$\begin{aligned} L_i(t,t_0) &= \begin{bmatrix}\mathfrak{B}_{0,k_i}(t) & \mathfrak{B}_{1,k_i}(t) & \dots & \mathfrak{B}_{k_i,k_i}(t)\end{bmatrix}, \\ \alpha_i(t_0) &= \begin{bmatrix}a_{i,0}(t_0) & a_{i,1}(t_0) & \dots & a_{i,k_i}(t_0)\end{bmatrix}^{\mathsf{T}}, \end{aligned} \tag{38}$$

and for all parameters, disregarding $\epsilon_i(t)$:

$$\theta(t) = L(t,t_0)\alpha(t_0), \tag{39}$$

where

$$L(t,t_0) = \begin{bmatrix} L_1(t,t_0) & & \\ & \ddots & \\ & & L_p(t,t_0)\end{bmatrix} \in \mathbb{R}^{p\times(p+K)}, \tag{40}$$

$$\alpha(t_0) = \begin{bmatrix}\alpha_0^{\mathsf{T}}(t_0) & \alpha_1^{\mathsf{T}}(t_0) & \dots & \alpha_{k_i}^{\mathsf{T}}(t_0)\end{bmatrix}^{\mathsf{T}} \in \mathbb{R}^{p+K}, \tag{41}$$

and the parameter estimate is thus:

$$\hat{\theta}(t) = L(t,t_0)\hat{\alpha}(t). \tag{42}$$

As for the monomial basis we have to make sure that the estimate $\hat{\theta}$ is continuous across time windows. The reset matrix is constructed using the relationship between Bernstein coefficients $\bar{\alpha}$ on the interval $[\bar{a}, \bar{b}]$ and α on $[a, b]$, [13], [15, Sec. 11.4] given by:

$$\bar{\alpha}_l = \sum_{k=0}^{k_i} A_{i,lk}\alpha_k, \tag{43}$$

where

$$A_{i,lk} = \sum_{\ell=\max(0,\,l+k-k_i)}^{\min(l,\,k)} \mathfrak{B}_{k-\ell,k_i-l}(\bar{a})\mathfrak{B}_{\ell,l}(\bar{b}), \tag{44}$$

for $l, k = 0, 1, \dots, k_i$, that are the row and column, respectively. The reset matrix should be formulated to be on the same form as (23):

$$\hat{\alpha}(t_0+T) = X(t_0+T,t_0)\hat{\alpha}((t_0+T)^-), \tag{45}$$

where $X(t_{0,j+1}, t_{0,j})$ is the resetting matrix given by:

$$X(t_0+T,t_0) = \begin{bmatrix} A_1(t_0+T,t_0) & & \\ & \ddots & \\ & & A_{k_i}(t_0+T,t_0)\end{bmatrix}, \tag{46}$$

where the elements $A_{i,lk}$ are given by (44). Since the time is divided into time windows, $\bar{a} = t_0 + T$, $\bar{b} = t_0 + 2T$. With these, the basis function (35) calculated at $\bar{a}$ simplifies to:

$$\mathfrak{B}_{\nu,n}(\bar{a}) = \begin{cases} 1 & \text{if } \nu = n, \\ 0 & \text{otherwise}, \end{cases} \tag{47}$$

i.e., only the last iteration of the sum in (44) is included, thus the i-th reset matrix can be written as

$$A_i(t_0+T,t_0) = \begin{bmatrix} 0 & 0 & \dots & 0 & \mathfrak{B}_{0,0}(\bar{b}) \\ 0 & 0 & \dots & \mathfrak{B}_{1,0}(\bar{b}) & \mathfrak{B}_{1,1}(\bar{b}) \\ \vdots & \vdots & \iddots & \vdots & \vdots \\ 0 & \mathfrak{B}_{k_i-1,0}(\bar{b}) & \dots & \mathfrak{B}_{k_i-1,k_i-2}(\bar{b}) & \mathfrak{B}_{k_i-1,k_i-1}(\bar{b}) \\ \mathfrak{B}_{k_i,0}(\bar{b}) & \mathfrak{B}_{k_i,1}(\bar{b}) & \dots & \mathfrak{B}_{k_i,k_i-1}(\bar{b}) & \mathfrak{B}_{k_i,k_i}(\bar{b}) \end{bmatrix}. \tag{48}$$

It can be seen that (48) does not depend on the window size T as the monomial reset matrix [5], [6] does.

Proposition 4. Resetting $\hat{\alpha}$ at the start of each time interval using (45) ensures a continuous parameter estimate $\hat{\theta}$.

The proof is omitted as it follows the proof in [6].

Denoting $\Delta_{\text{pol}}(t)$ when using the Bernstein basis as $\Delta_{\text{pol,B}}(t)$, and using the monomial basis as $\Delta_{\text{pol,m}}(t)$. Empirically, the following relationship was observed, see Section VI:

$$|\Delta_{\text{pol,B}}(t)| > |\Delta_{\text{pol,m}}(t)|, \tag{49}$$

for all $t \geq 0$, $k > 0$. This indicates that using the Bernstein basis results in a faster convergence than when using the monomial basis. A comparison of the numerical condition between the Bernstein and the monomial basis was performed in [15, Sec. 12.4.3], [13] which concludes that the monomial basis is much more sensitive to floating point arithmetic.

Farouki concludes in [15], [13] that if one has the choice between the Bernstein base and the monomial base, the Bernstein basis is preferable.

VI. Numerical Simulations

To highlight the advantages of the proposed method, two simulation examples are given in this section.

A. Estimation of Time-Varying Parameters

In this section several estimation methods are compared. The proposed method (29) using the Bernstein basis, Section V, is compared with the gradient-based parameter estimator (2) and the I-DREM-based method proposed in [12], Section III-B. We consider the LRE (13) defined as:

$$\phi(t) = \begin{bmatrix} 3\sin(4\pi t) \\ 2.5 \end{bmatrix}, \quad \theta(t) = \begin{bmatrix} 2+\sin(t) \\ 3+\cos(0.5t) \end{bmatrix}. \tag{50}$$

It can be seen, that $\phi(t) \in \mathrm{PE}$ and $\phi(t) \notin \mathcal{L}_2$. The system was simulated for $10\,\mathrm{s}$ with time step $\mathrm{dt} = 1 \times 10^{-4}\,\mathrm{s}$. The gain of the gradient method was $\gamma = 1.595$. The parameters of our method were:

$$\begin{gathered} T = 0.4\,\mathrm{s}, \quad \gamma = 1, \quad \ell = 1.95, \quad \lambda = [7\ 7], \\ p(0) = 1 \times 10^{10}, \quad \hat{\theta} = [0\ 0]^\intercal, \quad k = [2\ 2]. \end{gathered} \tag{51}$$

We compare with the numerical experiment from [12] and use the parameters given there:

$$\begin{aligned} &T = 0.25\,\mathrm{s}, \quad \gamma_0 = 100, \quad \beta = \frac{0.05}{T}, \\ &\Gamma = 0.75 I_2, \quad k = [1\ 1], \quad \kappa = 10^{-9}, \quad \sigma = 10^{-4}. \end{aligned} \tag{52}$$

The result of the simulation is on Fig. 1 without noise ($w(t) = 0$) and on Fig. 2 with noise ($w(t) \sim \mathcal{N}(0,1)$). The error, e, in the bottom subplot is the 2-norm of the estimation error.With no system noise, Fig. 1, our method performs clearly the best, with a low error. When introducing noise to the system, Fig. 2, our method still has the lowest error, but in general the three methods are robust to noise. Our method shows a slower initialisation time in comparison with the two others. This might be related to p_i in the least squares update rule (29) that has to "warm up".

B. Comparison of Polynomial Bases

We analyze the effect of using the Bernstein basis or the monomial basis in (29). In this simulation there is no noise, $w = 0$, and the parameters are as follows:

$$T = 0.25\,\mathrm{s}, \quad \gamma = 15, \quad \ell = 5, \quad \lambda = 0.8, \tag{53}$$

$$p(0) = 1 \times 10^{30}, \quad \hat{\theta} = 0, \quad k = 1, \dots, 10, \tag{54}$$

and the LRE is given by:

$$\phi(t) = 3\sin(4\pi t), \quad \theta(t) = 2 + \sin(t). \tag{55}$$

To remove the effect of filter initialisation the first second of the data was removed. The mean square error (MSE) was used to compare the methods. The simulation result can be seen on Fig. 3. For the polynomials of degree 3 or lower, the error is the same for the two bases, but from $k \geq 4$, the error on the monomial basis increases. This is related to the numerical condition of the two polynomial bases, see [13], [15]. The main difference between the polynomial bases is the value of Δ. A comparison of the first three orders can be seen on Fig. 4. The shape is the same, but numerical value is different, note the numerical difference in the limits on the y-axes between the monomial and the Bernstein basis. From the figure, notice that $|\Delta_{\text{pol,B}}(t)| > |\Delta_{\text{pol,m}}|(t) \quad \forall t > 0$, which results in a faster convergence time for the Bernstein basis version compared with the monomial basis version. From $k = 8$ and above, neither of the polynomial bases can estimate the parameter correctly.

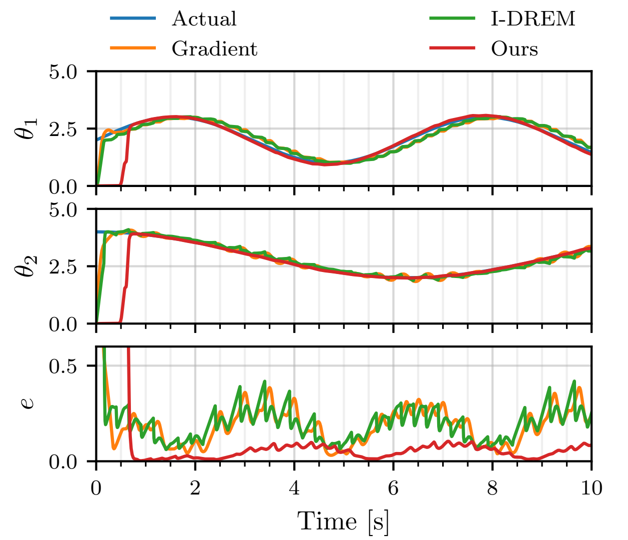

Fig. 1. Estimation of time-varying parameters comparing our method with the gradient-based parameter estimator and the I-DREM-based method in [12] without noise ($w = 0$). In θ_1, θ_2 the actual value is hard to see, as the other lines overlap it.

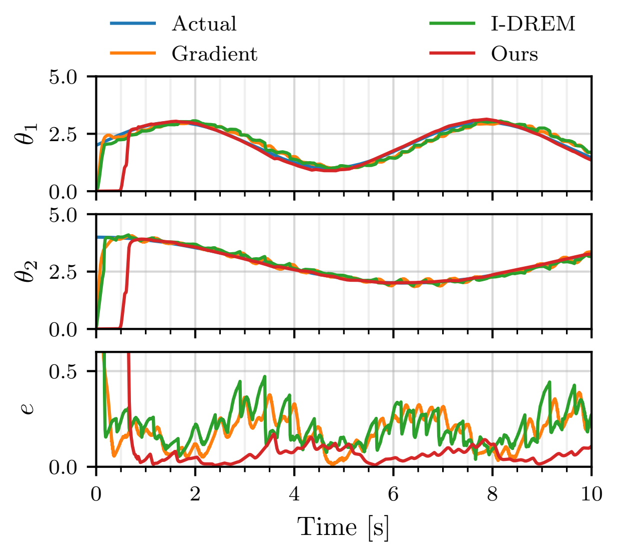

Fig. 2. Estimation of time-varying parameters comparing our method with the gradient-based parameter estimator and the I-DREM-based method in [12] with noise ($w \sim \mathcal{N}(0,1)$). In θ_1, θ_2 the actual value is hard to see, as the other lines overlap it.

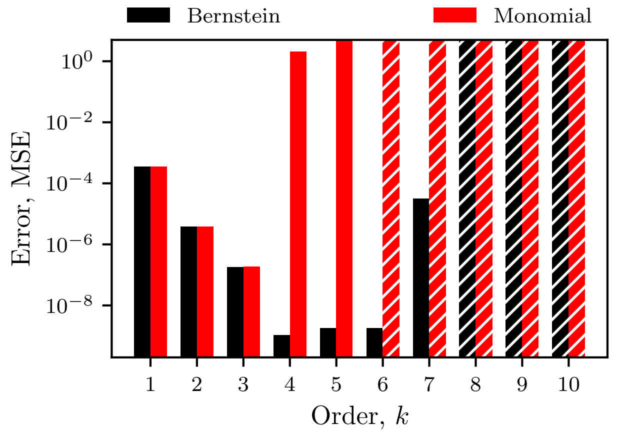

Fig. 3. Bar plot comparing the MSE when using the Bernstein and monomial base. The polynomial order ranges from 1 to 10. The hatched bars indicate when the estimator is not working, and the error is at maximum.

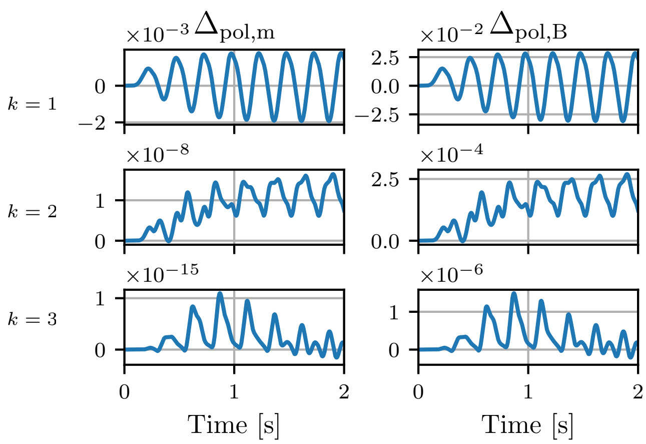

Fig. 4. Comparison of the Δ of the orders $k = 1, 2, 3$.

VII. Conclusions

In this paper, we presented a method for combining local polynomial regression and DREM to estimate time-varying parameters in a linear regression equation. Simulation results shows that the method performs better than the standard gradient-based parameter estimator and a state-of-the art method from [12]. We also show that using the Bernstein polynomial basis instead of the monomial polynomial basis allows the proposed method to work at a higher order. Though when the order is low enough, $k < 4$, they show similar performance.

Acknowledgement

This work was supported by the PIRAT project, funded by Innovation Fund Denmark, grant number 9069-00046B.

This work has partially been carried out within the framework of the EUROfusion Consortium, funded by the European Union via the Euratom Research and Training Programme (Grant Agreement No 101052200 — EUROfusion). Views and opinions expressed are however those of the authors only and do not necessarily reflect those of the European Union or the European Commission. Neither the European Union nor the European Commission can be held responsible for them.

References

[1] S. Sastry and M. Bodson, *Adaptive Control: Stability, Convergence, and Robustness*. USA: Prentice-Hall, Inc., 1989.

[2] W. S. Cleveland and S. J. Devlin, "Locally weighted regression: An approach to regression analysis by local fitting," *Journal of the American Statistical Association*, vol. 83, no. 403, pp. 596–610, 1988.

[3] H. A. Nielsen, T. S. Nielsen, A. K. Joensen, H. Madsen, and J. Holst, "Tracking time-varying-coefficient functions," *International Journal of Adaptive Control and Signal Processing*, vol. 14, pp. 813–828, 12 2000.

[4] A. Joensen, H. Madsen, H. A. Nielsen, and T. S. Nielsen, "Tracking time-varying parameters with local regression," *Automatica*, vol. 36, pp. 1199–1204, 8 2000.

[5] Y. Zhu and P. R. Pagilla, "Adaptive estimation of time-varying parameters in linearly parametrized systems," *Journal of Dynamic Systems, Measurement, and Control*, vol. 128, pp. 691–695, 9 2006.

[6] J. Na, J. Yang, X. Ren, and Y. Guo, "Robust adaptive estimation of nonlinear system with time-varying parameters," *International Journal of Adaptive Control and Signal Processing*, vol. 29, pp. 1055–1072, 8 2015.

[7] P. M. LION, "Rapid identification of linear and nonlinear systems.," *AIAA Journal*, vol. 5, pp. 1835–1842, 10 1967.

[8] G. Kreisselmeier, "Adaptive observers with exponential rate of convergence," *IEEE Transactions on Automatic Control*, vol. 22, pp. 2–8, 2 1977.

[9] S. Aranovskiy, A. Bobtsov, R. Ortega, and A. Pyrkin, "Performance enhancement of parameter estimators via dynamic regressor extension and mixing," *IEEE Transactions on Automatic Control*, vol. 62, pp. 3546–3550, 7 2017.

[10] R. Ortega, L. Praly, S. Aranovskiy, B. Yi, and W. Zhang, "On dynamic regressor extension and mixing parameter estimators: Two Luenberger observers interpretations," *Automatica*, vol. 95, pp. 548–551, 9 2018.

[11] A. I. Glushchenko, V. A. Petrov, and K. A. Lastochkin, "I-DREM: Relaxing the square integrability condition," *Automation and Remote Control*, vol. 82, pp. 1233–1247, 7 2021.

[12] A. Glushchenko and K. Lastochkin, "Robust time-varying parameters estimation based on I-DREM procedure," *IFAC-PapersOnLine*, vol. 55, pp. 91–96, 2022.

[13] R. Farouki and V. Rajan, "On the numerical condition of polynomials in Bernstein form," *Computer Aided Geometric Design*, vol. 4, pp. 191–216, 11 1987.

[14] S. N. Bernstein, "On the best approximation of continuous functions by polynomials of a given degree," *Comm. Soc. Math. Kharkow, Ser*, vol. 2, no. 13, pp. 49–194, 1912.

[15] R. T. Farouki, *Pythagorean-Hodograph Curves: Algebra and Geometry Inseparable*. Springer Berlin Heidelberg, 2008.

[16] R. Ortega, S. Aranovskiy, A. A. Pyrkin, A. Astolfi, and A. A. Bobtsov, "New results on parameter estimation via dynamic regressor extension and mixing: Continuous and discrete-time cases," *IEEE Transactions on Automatic Control*, vol. 66, pp. 2265–2272, 5 2021.

[17] R. Ortega, V. Nikiforov, and D. Gerasimov, "On modified parameter estimators for identification and adaptive control. a unified framework and some new schemes," *Annual Reviews in Control*, vol. 50, pp. 278–293, 2020.

[18] M. Korotina, S. Aranovskiy, R. Ushirobira, and A. Vedyakov, "On parameter tuning and convergence properties of the DREM procedure," in *2020 European Control Conference (ECC)*, IEEE, 5 2020.