Tuning nonlinear state-space models using unconstrained multiple shooting

Tuning nonlinear state-space models using unconstrained multiple shooting

J. Decuyper [*] **M. C. Runacres** [*] **J. Schoukens** [*,**] **K. Tiels** [***]

[*] *Department of Engineering Technology (INDI), Vrije Universiteit Brussel, 1050 Brussel, Belgium (e-mail: jan.decuyper@vub.be, mark.runacres@vub.be, johan.schoukens@vub.be).*
[**] *Department of Electrical Engineering, Eindhoven University of Technology, Eindhoven, The Netherlands*
[***] *Department of Mechanical Engineering, Eindhoven University of Technology, 5600 MB Eindhoven, The Netherlands (e-mail: k.tiels@tue.nl)*

Abstract: A persisting challenge in nonlinear dynamical modelling is parameter inference from data. Provided that an appropriate model structure was selected, the identification problem is profoundly affected by a choice of initialisation. A particular challenge that may arise is initialisation within a region of the parameter space where the model is not contractive. Exploring such regions is not feasible using the conventional optimisation tools for they require a bounded evaluation of the cost. This work proposes an unconstrained multiple shooting technique, able to mitigate stability issues during the optimisation of nonlinear state-space models. The technique is illustrated on simulation results of a Van der Pol oscillator and benchmark results on a Bouc-Wen hysteretic system.

Keywords: Unconstrained multiple shooting, Nonlinear state-space models, Nonlinear optimisation, Unstable initialisation.

1. INTRODUCTION

Given the advent of powerful model structures, and especially with respect to the notion of universal approximators, a persistent challenge of nonlinear system identification lies within parameter estimation. The challenges are multiple. It is well known that dynamic model structures, which are typically nonlinear in the parameters, result in non-convex objective functions. An additional challenge is crossing regions of the parameter space where the model is not contractive. The presence of an unstable fixed point will in that case drive the output of the model out of bound. Conventional optimisation methods require a bounded evaluation of the cost in order to be able to proceed to a model update. The non-contractive regions may however not always be avoided, they might even be deliberately chosen as initialisation point. Three scenarios may be envisaged:

- *Unintentional unstable initialisation:* initial model paramaters are generated following an alternative (i.e. not stability) objective. An example is provided in Section 4.2 where a reduced model is initialised starting from the large model and the reduced initialisation turns out to lie in a non-contractive region.
- *Intentional unstable initialisation:* such as when insight dictates that the system inherently behaves unstable (e.g. the Van der Pol equation in Section 3.4).
- *Crossing unstable regions:* when an optimiser is trapped while it could potentially proceed if it were allowed to step into an unstable region.

A solution, enabling to deal with unstable regions, was provided by the technique of *multiple shooting*. The fundamental idea is to split the time record into a number of subrecords thereby preventing the instability to build during simulation. The technique originates back to the 1960s where it was used to solve boundary value problems in the framework of ordinary differential equations (Morrison et al., 1962). Shooting methods turn a boundary value problem into a sequence of easier to solve initial value problems (Osborne, 1969). Splitting the solution interval into smaller subintervals was found to stabilise the solution process. Within system identification, the application of shooting methods was proposed for output error models in Ribeiro and Aguirre (2017) and more specifically for nonlinear state-space models in Van Mulders et al. (2010). The idea also appeared in some earlier works, e.g. van Domselaar and Hemker (1975); Bock (1983); Buchsbaum (2007). It is worth noting that also recurrent neural networks may suffer from non-contractive behaviour and hence may benefit from multiple shooting.

Ribeiro et al. (2019) attributes the success of multiple shooting to a smoothing of the objective function. They show that the Lipschitz constant and its derivate (β-smoothness) may blow up exponentially for non-

[⋆] This work was supported by the Fund for Scientific Research (FWO-Vlaanderen) under projects G.0280.15N and G.0901.17N, EOS Project no 30468160, the Swedish Research Council (VR) via the project NewLEADS – New Directions in Learning Dynamical Systems (contract number: 621-2016-06079), and by the Swedish Foundation for Strategic Research (SSF) via the project ASSEMBLE (contract number: RIT15-0012).

contractive models. Making the simulation length a design parameter allows to reshape the optimisation landscape.

When splitting the time record in subrecords, the original problem becomes larger. Apart from having to estimate the model parameters, also an initial value problem needs to be solved for every subrecord (illustration in Fig. 3). Depending on how this initial value problem is treated, a number of categories of 'shooting' exist:

- *Single shooting:* the conventional approach where one simulates the entire time record at once.
- *Fully constrained multiple shooting:* every output sample is assumed a function of an unknown internal state and the input. All state vectors (i.e. at every time step) are treated as unknown parameters that need to be estimated, together with the model parameters. A vector of constraints is added to ensure proper state evolution.
- *Partially constrained multiple shooting:* create subrecords of specified lengths. Only constrain the record's beginning to match the previous record's end and simulate freely within one record. This is the approach adopted in Ribeiro et al. (2019).

Both the fully constrained and the partially constrained multiple shooting method have proven their worth in escaping non-contractive regions. They, however, both involve solving a constrained optimisation problem which is considerably more expensive than single shooting and requires a dedicated solver.

In this work we will show that lots is to be gained by introducing a so called *unconstrained multiple shooting* method. It will turn out that even without constraints, an optimiser can be built that navigates through non-contractive regions of the parameter space. An additional advantage of the unconstrained approach is that only minor changes are needed to the optimisation algorithm, which is generally already in place for single shooting, in order to be able to tackle the multiple shooting case.

2. PROBLEM STATEMENT

We will focus here on discrete-time polynomial nonlinear state-space models (PNLSS) (Paduart et al., 2010), although all statements will hold for the entire set of output error models. As generic set of equations we consider

$$\begin{cases} \mathbf{x}(k+1) = \mathbf{A}\mathbf{x}(k) + \mathbf{B}\mathbf{u}(k) + \mathbf{f}(\boldsymbol{\theta}_f, \mathbf{x}(k), \mathbf{u}(k)) & \text{(1a)} \\ \hat{\mathbf{y}}(k) = \mathbf{C}\mathbf{x}(k) + \mathbf{D}\mathbf{u}(k), & \text{(1b)} \end{cases}$$

where $k = {}^{t}/_{T_s}$ is the time index with T_s the sampling period and the matrices have the following dimensions: $\mathbf{A} \in \mathbb{R}^{n\times n}$ with n the number of state variables, $\mathbf{B} \in \mathbb{R}^{n\times m}$ with m the number of input variables, $\mathbf{C} \in \mathbb{R}^{p\times n}$ with p the number of outputs, $\mathbf{D} \in \mathbb{R}^{p\times m}$ and $\boldsymbol{\theta}_f \in \mathbb{R}^{n_{\theta_f}}$ containing the parameters of the nonlinear function $\mathbf{f}$. Grouping all unknowns in a single vector $\boldsymbol{\theta} \in \mathbb{R}^{n_\theta}$ using the vec-operator to stack matrix elements we have

$$\boldsymbol{\theta} = [\mathrm{vec}(\mathbf{A})^T \ \mathrm{vec}(\mathbf{B})^T \ \mathrm{vec}(\mathbf{C})^T \ \mathrm{vec}(\mathbf{D})^T \ \boldsymbol{\theta}_f^T]^T. \quad (2)$$

Problem statement: *given an input vector $\boldsymbol{u} \in \mathbb{R}^{m\times N}$ and an output vector $\boldsymbol{y} \in \mathbb{R}^{p\times N}$, find the minimising $\boldsymbol{\theta}$*

$$\arg\min_{\boldsymbol{\theta}} \sum_{k=1}^{N} \|\mathbf{y}(k) - \hat{\mathbf{y}}(\boldsymbol{\theta}, k)\|_2^2 \quad (3)$$

starting from an initialisation $\boldsymbol{\theta}_0$, which includes the initial state $\boldsymbol{x}(0)$, and for which

$$\lim_{k\to N} \|\mathbf{y}(k) - \hat{\mathbf{y}}(\boldsymbol{\theta}_0, k)\|_2^2 < L, \quad (4)$$

with L some practical upper bound.

This can alternatively be interpreted as parameter estimation, initialised from a non-contractive model.

Additional requirement: *formulate the objective as an unconstrained problem.*

3. UNCONSTRAINED MULTIPLE SHOOTING

The idea is to leverage the smoothness of the objective function from multiple shooting while loosing the computationally expensive constraints (Van Mulders, 2012).

3.1 Approach

Consider an input-output data record of finite length N. Following the principle of multiple shooting the record is split into M parts, each of length $\Delta m = {}^{N}/_{M}$. Every record is attributed an unknown initial condition vector, $\mathbf{x}_{0i}$ (i-th record). The problem is then formulated into two objective functions, one pursuing the model parameters $\boldsymbol{\theta}$ and one leading to the unknown $\mathbf{x}_{0i}$. Introducing the sample error of a subrecord as

$$e_i(k) = \|\mathbf{y}(\delta_i + k) - \hat{\mathbf{y}}(\mathbf{x}_{0i}, \boldsymbol{\theta}, \delta_i + k)\|_2, \quad (5)$$

where $k \in \{1, 2, ..., \Delta m\}$ and $\delta_i = (i-1)\Delta m$. An average cost over all subrecords that produce bounded errors is then given by

$$V_{\boldsymbol{\theta}} = \frac{1}{M_b \Delta m}\sum_{i=1}^{M} b_i \mathbf{e}_i^T \mathbf{e}_i \qquad \begin{cases} b_i = 1 & \mathbf{e}_i^T\mathbf{e}_i \le L \\ b_i = 0 & \mathbf{e}_i^T\mathbf{e}_i > L \\ M_b = \sum_i b_i & \end{cases} \quad (6)$$

Notice that one can tune the model parameters without requiring bounded errors on all subrecords. This will prevent M from becoming needlessly large when some subrecords are more sensitive than others. Eq. (6) can alternatively be used to handle parameter estimation from multiple experiments, e.g. when the underlying system is unstable and experiments have to be limited in time (Shafique et al., 2017).

Each $\mathbf{x}_{0i}$ only affects the course of the i-th subrecord. Hence it is updated following a subrecord specific cost function

$$V_{0i} = \frac{1}{\Delta m} b_i \mathbf{e}_i^T \mathbf{e}_i. \quad (7)$$

This is where the unconstrained approach distinguishes itself from the constrained approach: all sub records are treated individually, i.e. without linking them with constraints, although they jointly contribute to the parameter update via Eq. (6).

Observe that $\nabla_{\mathbf{x}_{0i}} V_{\boldsymbol{\theta}} = \frac{1}{M_b}\nabla_{\mathbf{x}_{0i}} V_{0i}$, therefore there is essentially only one cost function, i.e. Eq. (6).

The approach can be summarised in the following steps:

(1) Select Δm, and accordingly also M, on the basis of the time constant of a linearisation of the model. Use 3 to 5 time constants as rule of thumb. In practice, monitor the ${}^{M_b}/_{M}$-ratio since it will give an indication of the fraction of subrecords from which is learned. Avoid training (and potentially overfitting) to only a handful of subrecords.
(2) Tune the model by descending Eq. (6) and Eq. (7) simultaneously. An optimisation strategy is provided in Section 3.3. When a stopping criterion is met, go back to step (1) and increase Δm.
(3) Iterating between steps (1) and (2) will guide the optimiser out of the non-contractive region. Once successful, continue tuning the model using *single shooting*.

The unconstrained multiple shooting approach can hence be seen as an initialisation for the classical single shooting method. The final estimate will therefore inherit the properties of the single shooting estimator.

3.2 On the equivalence with single shooting

It is easy to see that for the limiting case of $M = 1$, Eq. (6) boils down to *single shooting*.

We can moreover show that, in the absence of model errors and output noise, the global minimum of the single shooting cost coincides with the global minimum of the unconstrained multiple shooting cost.

Lemma 1. Let $V_{\boldsymbol{\theta}}$ with $M = 1$ be called V_s (single shooting) and let $\boldsymbol{\theta}_s^*$ be the global minimiser of V_s. In the absence of model errors and output noise, $V_s(\boldsymbol{\theta}_s^*) = 0$. The claim is that $V_\theta(\boldsymbol{\theta}_s^*) = 0$ for any $M \geq 1$ (single shooting and multiple shooting).

Proof: Let $\mathbf{x}(\boldsymbol{\theta}_s^*)$ be the simulated state vector found following single shooting and choose the initial condition vector $\mathbf{x}_{0_i}$ of the i-th data record (single or multiple shooting, $i = 1, \ldots, M$) equal to $\mathbf{x}(\boldsymbol{\theta}_s^*)$ at the corresponding time step. Then the state vector following (single or) multiple shooting with M data records is the same as $\mathbf{x}(\boldsymbol{\theta}_s^*)$ when $\boldsymbol{\theta} = \boldsymbol{\theta}_s^*$, and thus $V_\theta(\boldsymbol{\theta}_s^*) = 0$. $\square$

Note that step (3) of the approach consists of single shooting. Hence, the estimation of $\mathbf{x}_{0_i}$, and potentially the introduced errors on the initial conditions, become obsolete.

3.3 Optimisation scheme

Minimising Eq. (6) and Eq. (7) can be tackled using any nonlinear optimisation algorithm. We will discuss one particular strategy, based primarily on the classical Levenberg-Marquardt (LM) algorithm. LM balances gradient descent and Gauss-Newton in order to benefit from a large region of convergence (gradient descent) as well as from fast convergence (for quadratic cost). The parameter update is computed by solving

$$(J^T J + \lambda^2 I_{n_\theta})\Delta\boldsymbol{\theta} = \underbrace{-J^T \mathbf{e}}_{\frac{1}{2}\nabla_{\boldsymbol{\theta}} V}, \tag{8}$$

with λ the damping factor determining the weight between both methods. Large values of λ favour gradient descent, small values lean towards Gauss-Newton. When used to descend Eq. (6), $\mathbf{e} = [\mathbf{e}_1^T \ \ldots \ \mathbf{e}_M^T]^T$, is a concatenation of the subrecords of errors and $J(k) = \frac{\partial \mathbf{e}(k)}{\partial \boldsymbol{\theta}}$ (model parameters are updated on the basis of all subrecords), while when used for Eq. (7) $\mathbf{e} = \mathbf{e}_i$ and $J(k) = \frac{\partial \mathbf{e}_i(k)}{\partial \mathbf{x}_{0_i}}$.

Notice that λ plays a double role. Apart from balancing both methods it also governs the step size. This turns out to be unpractical, especially in those cases where the gradient blows up due to high errors, i.e. within non-contractive regions. We propose to add an alternative optimisation loop, along side the Levenberg-Marquardt loop. This alternative loop optimises along the normalised gradient and includes a scan over the step size (line search). Hence, even for large gradients (due to high errors) an appropriate step size can be found. The update is then given by

$$\Delta\boldsymbol{\theta} = -\alpha \frac{J^T \mathbf{e}}{\|J^T \mathbf{e}\|_2}, \tag{9}$$

with α the learning rate. Switching between normalised gradient descent and Levenberg-Marquardt is done on the basis of the size of the error. As quantifier the relative root-mean-squared error is used,

$$e_{\text{rel}} = \frac{\text{rms}(\mathbf{e})}{\text{rms}(\mathbf{y})}. \tag{10}$$

Whenever $e_{\text{rel}} > 10$ normalised gradient descent is used, for $e_{\text{rel}} < 10$ LM is used. The optimisation algorithm is schematised in Fig. 1. It illustrates the update of both the model parameters and the $\mathbf{x}_{0_i}$ vectors by collecting them together in $\boldsymbol{\theta}$. Correspondingly also the Jacobians are collected in $\mathbf{J}$. A typical flow of operation could be as follows:

- Evaluate whether $e_{\text{rel}} > 10$, if true continue with normalised gradient descent.
- Evaluate Eq. (9), update the parameters and recompute the cost.
- If the update is successful, double α, if not, $\alpha = 0.5\alpha$.
- Continue until $e_{\text{rel}} > 10$ is false, then enter the LM loop.
- Evaluate Eq. (8), update the parameters, and recompute the cost.
- If the update is successful, lean towards Gauss-Newton: $\lambda = 0.5\lambda$, if not lean towards gradient descent, $\lambda = \sqrt{10}\lambda$.
- Iterate until the maximum number of iterations is reached.

Notice that an update is only considered successful when (1) the cost has dropped and (2) $M_b' \geq M_b$, i.e. the number of subrecords producing bounded errors may not drop.

3.4 A Van der Pol example

The Van der Pol equation is a second order ODE with a position dependent damping term,

$$\frac{d^2 y(t)}{dt^2} - \beta \frac{dy(t)}{dt} + \gamma y(t) + \varepsilon y^2(t) \frac{dy(t)}{dt} = u(t). \tag{11}$$

It is particularly well suited to illustrate the unconstrained multiple shooting approach since it results in an unstable fixed point at the origin whenever $\beta > 0$. For a typical application, ε would need to be tuned in order to counter

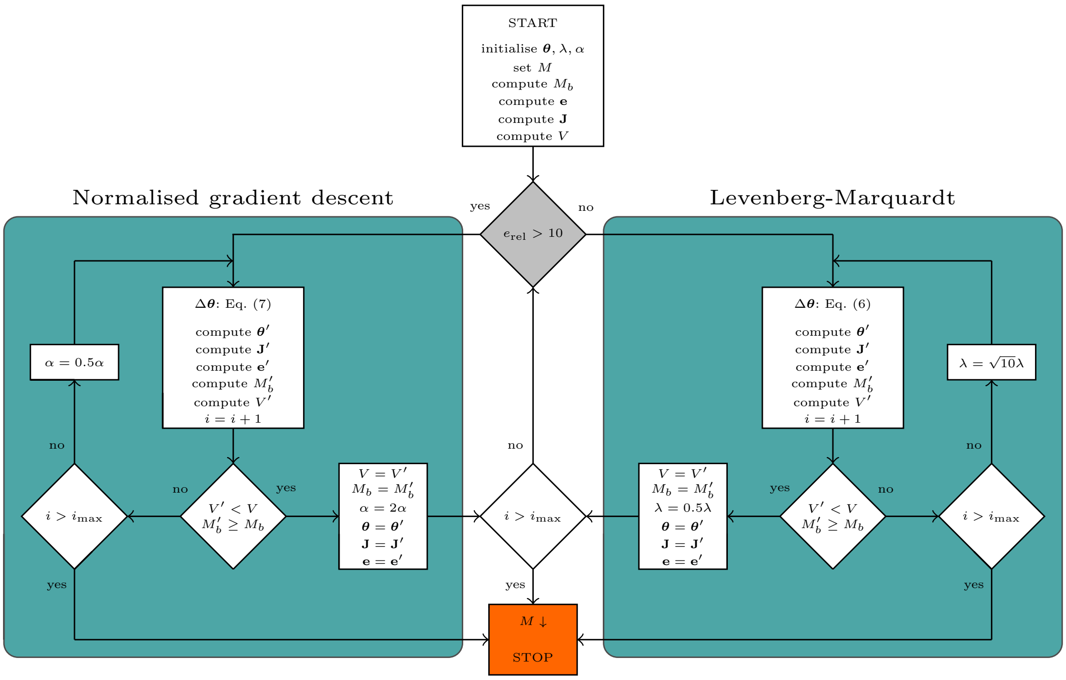


Fig. 1. Schematisation of the double optimisation loop. When $e_{\text{rel}} > 10$, steps are computed using normalised gradient descent, otherwise Levenberg-Marquardt is used. Primes denote updated values.

the negative damping and create a periodic orbit (stable limit cycle).

We will study the scenario where ε is to be estimated from input-output data while initialised from $\varepsilon_0 = 0$, evidently lying in a non-contractive region of the parameter space.

To comply with the discrete time PNLSS structure (see Section 2) Eq. (11) is discritised using a first-order Euler approximation. With the state vector being $\mathbf{x} = [y(t) \quad \frac{dy(t)}{dt}]^T$, and a sample period T_s we have the nonlinear state-space model

$$\begin{cases} \mathbf{x}(k+1) = \begin{bmatrix} 1 & T_s \\ -\gamma T_s & \beta T_s + 1 \end{bmatrix} \mathbf{x}(k) + \begin{bmatrix} 0 \\ T_s \end{bmatrix} u(k) \qquad (12a) \\ \qquad\qquad + \begin{bmatrix} 0 \\ -\varepsilon T_s \end{bmatrix} x_1^2(k) x_2(k) \\ \hat{\mathbf{y}}(k) = [1 \quad 0]\, \mathbf{x}(k). \qquad (12b) \end{cases}$$

For small T_s Eq. (12) approximates Eq. (11). In this context we are not concerned with approximation errors and will consider the discretised model (Eq. (12)) as the true underlying model. The following data are generated for training:

- $\beta = 9.4248$, $\gamma = (2\pi 5)^2$, $\varepsilon = 9.4248$, $T_s = 0.001$.
- $u(k)$: random-phase multisine realisation (Pintelon and Schoukens, 2012), $\text{rms}(u) = 500$, excitation band 0 Hz - 40 Hz in steps of $2f_0 = 0.4$ Hz (odd-multisine), $N = 5000$ samples. Notice that the frequency band contains the eigen-frequency of the linear part, $f_n = \frac{1}{2\pi}\sqrt{\gamma} = 5$ Hz and multiples of f_n.

For the sake of illustration all parameters, except for ε, will be assumed fixed. Hence a 1-dimensional cost function can be drawn. This is depicted in Fig. 2. The hatched area indicates the non-contractive region. In order to estimate ε, initialised from $\varepsilon_0 = 0$, we can use multiple

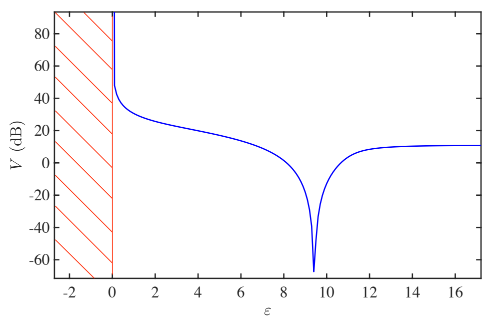
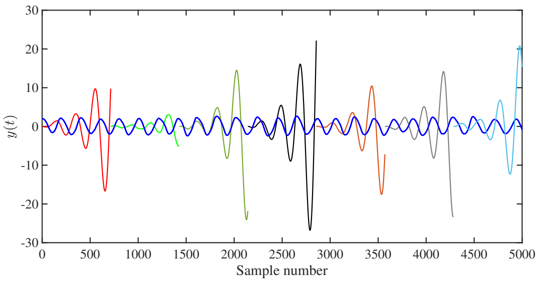


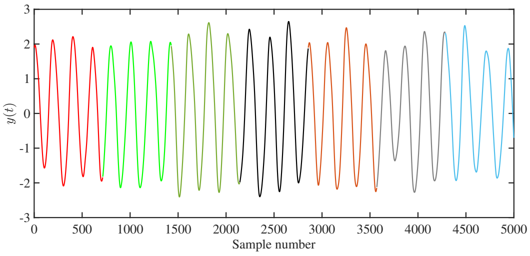
Fig. 2. Cost as a function of ε, computed from Eq. (12) on the basis of the training data. The hatched area indicates the non-contractive region.

shooting. Following the scheme in Fig. 1 we first select an appropriate length for the subrecords. In this case $M = 7$ seems adequate since $\Delta m T_s = 3.75\tau$ with $\tau = 1/f_n$ (the time constant). A simulation of the subrecords is shown together with the original signal in Fig. 3a. The initial conditions at the start of each subrecord are initialised with zeros, i.e. $\mathbf{x}_{0_i} = \mathbf{0}$.

Fig. 3b shows again the simulated subrecords, this time after tuning ε and all $\mathbf{x}_{0_i}$. Close inspection of the signal reveals small gaps between the subrecords. This indicates minor errors on $\mathbf{x}_{0_i}$ which may still exist given the unconstrained nature of the approach. The initial conditions $\mathbf{x}_{0_i}$ become however obsolete when the model is used in normal operation (single shooting). At that point only the accuracy of the model parameters matter.



(a) Simulation with $\varepsilon = \varepsilon_0$.



(b) Simulation after tuning ε.

Fig. 3. Unconstrained multiple shooting applied to the estimation of ε (Eq. (12)), initialised from $\varepsilon_0 = 0$ in the non-contractive region. The record is split into $M = 7$ subrecords, preventing the instability to build.

4. BENCHMARK PROBLEMS

We will discuss two scenarios in which unconstrained multiple shooting can provide a solution, illustrated on the basis of the Bouc-Wen system identification benchmark (Schoukens and Noël, 2017): Section 4.1 discusses *intentional unstable initialisation* of the linear part of a PNLSS model and Section 4.2 is on *unintentional unstable initialisation* based on a *decoupled polynomial model* (Decuyper et al., 2019b).

The Bouc-Wen system has been intensively used to represent hysteretic effects in mechanical engineering (Morrison et al., 2001; Bertotti, 1998; Mueller, 1985). The hysteresis loop is modelled using a nonlinear memory-dependent restoring force (f_H). The dynamics of a single-degree-of-freedom Bouc-Wen oscillator are governed by the second order nonlinear ODE,

$$m\ddot{y}(t) + c\dot{y}(t) + ky(t) + f_H(y(t), \dot{y}(t)) = u(t), \qquad (13)$$

where k and c are the linear stiffness and viscous damping coefficients, respectively. The hysteretic force f_H obeys the first order ODE,

$$\dot{f}_H(t) = \alpha\dot{y}(t) - \left(\gamma|\dot{y}(t)||f_H(t)|^{\nu-1}f_H(t) + \delta\dot{y}(t)|f_H(t)|^{\nu}\right), \qquad (14)$$

with the Bouc-Wen parameters α, β, γ, δ and ν. We will consider a particular realisation with the following setting of parameters:

Parameter	m	c	k	α	β	γ	δ	ν
in SI unit	2	10	5 10^4	5 10^4	1 10^4	0.8	-1.1	1

The benchmark dataset consists of 4 realisations of a random-phase multisine exciting the frequency band 5 – 150 Hz. The input level corresponds to rms($\mathbf{u}$) = 50 N. The signal-to-noise ratio on the output is roughly 40 dB.

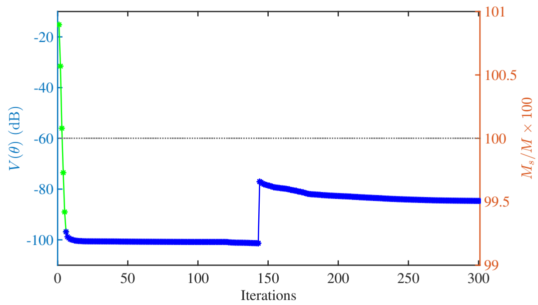


Fig. 4. Unconstrained multiple shooting applied to a non-contractive estimate of the BLA of the Bouc-Wen benchmark data. Green indicates steps computed using normalised gradient descent. Blue shows LM steps. The first blue plateau is reached for $M = 437$. From thereon single shooting is used ($M = 1$) until the second plateau is reached. Grey indicates the portion of subrecords which are used for training (the records that produce bounded errors).

4.1 Non-contractive linear initialisation of a PNLSS model

Estimating a PNLSS model of the form of Eq. (1) requires a three-step procedure (implemented in the PNLSS toolbox (Tiels, 2016)):

(1) Identify a non-parametric linear model from the data, preferably the best linear approximation (BLA) (Pintelon and Schoukens, 2012).
(2) Obtain a parametric model of the non-parametric BLA. This is typically done using linear frequency-domain subspace identification (McKelvey et al., 1996; Pintelon, 2002).
(3) Use LM to continue to tune both the linear and the nonlinear model parameters.

The explicit linear step in the process may be troublesome given the fact that this linear initialisation needs to be contractive. If not, one cannot pass on to the nonlinear stage using conventional optimisation algorithms. In some cases, a non-contractive linear model might by better suited (e.g. the Van der Pol model in Section 3.4). Moreover, many linear estimation techniques do not have stability as a direct objective. In the frequency-domain subspace algorithm of step (2) the cost is minimised on the FRF and not in an output-error sense. We will show how unconstrained multiple shooting can be used to circumvent this problem.

For the sake of illustration we will *deliberately* choose a non-contractive linear model, computed from the subspace identification algorithm of step (2), implemented in the PNLSS toolbox. The model parameters are listed in Appendix A. It is a fourth order parametric estimate of the BLA. To initiate the unconstrained multiple shooting-loop M is set to 437. This corresponds to a $\Delta m = 375$ samples which corresponds to 3 time constants. The course of the optimisation is depicted in Fig. 4. First, a plateau is reached for $M = 437$. Meanwhile the model has already become contractive. Hence M is set to 1 (single shooting) and the optimisation is reinitiated. An accurate final model is obtained yielding a figure of merit: $e_{RMS} = \sqrt{1/N\sum_{n=1}^{N}(y_{\text{val}} - \hat{y}_{\text{val}})^2} = 1.88 \times 10^{-5}$.

4.2 A decoupled PNLSS model of the Bouc-Wen system

It was shown that accurate models of the form of Eq. 1 can be obtained from the training data (Noël et al. (2017) and Section 4.1). In those cases, the nonlinear function was modelled by a generic multivariate polynomial, i.e. $\mathbf{f}(\mathbf{x}(k), \mathbf{u}(k)) = \mathbf{E}\boldsymbol{\zeta}(\mathbf{x}(k), \mathbf{u}(k))$. It was however found that a more efficient parameterisation can be obtained using so-called *decoupled polynomial* functions,

$$\mathbf{E}\boldsymbol{\zeta}(\boldsymbol{\theta}_f, \mathbf{x}(k), \mathbf{u}(k)) \approx \mathbf{W}\mathbf{g}\left(\boldsymbol{\theta}_g, \mathbf{x}(k), \mathbf{u}(k)\right), \tag{15}$$

with $\mathbf{g}$: $\mathbb{R}^r \rightarrow \mathbb{R}^r$ a vector function of r univariate polynomials and $\mathbf{W} \in \mathbb{R}^{n \times r}$ a linear transformation matrix. Decuyper et al. (2019b) show that the number of parameters needed in the decoupled form can be as low as one third of the original amount, while preserving the accuracy of the model.

Decoupling multivariate polynomials was introduced in (Dreesen et al., 2018). Decuyper et al. (2019a) discusses the approximative nature of Eq. (15). A direct consequence of the approximation is that, when plugged back into the nonlinear state-space model, the model is slightly altered. This modification may push the model into a non-contractive region, making it again unfeasible to tune it any further using conventional techniques. We will show that the unconstrained multiple shooting technique can be used in this case.

As illustration, the decoupling technique of Decuyper et al. (2019a) is applied to a third order PNLSS model containing a multivariate polynomial function $\mathbf{f}$ of the second and third degree. The latter was identified from the previously described benchmark data. After decoupling we arrive at the following model:

$$\begin{cases} \mathbf{x}(k+1) = \mathbf{A}\mathbf{x}(k) + \mathbf{b}u(k) + \mathbf{W}\mathbf{g}\left(\mathbf{V}^T \begin{bmatrix} \mathbf{x}(k) \\ u(k) \end{bmatrix}\right) & \text{(16a)} \\ \hat{y}(k) = \mathbf{c}^T\mathbf{x}(k) + du(k), & \text{(16b)} \end{cases}$$

with $n = 3$, $p = 1$, $m = 1$, $r = 6$ and $\mathbf{V} \in \mathbb{R}^{(n+m)\times r}$. The univariate polynomials in $\mathbf{g}$ are of third order. This model turns out to be non-contractive (the parameters are listed in Appendix B), resulting in poor simulation performance.

The optimisation procedure is initiated by setting $M = 2000$. This corresponds to $\Delta m = 102$, which is approximately 3 time constants of the BLA. The course of the optimisation is depicted in Fig. 5. First a number of normalised gradient descent steps are required (indicated in green), this is followed by the LM loop (blue markers) until a plateau is reached. At this point the model has already become contractive. Hence the optimisation is continued using single shooting until the second blue plateau is reached.

When the optimisation is completed the model provides accurate simulation results on the validation data, figure of merit: $e_{RMS} = \sqrt{1/N\sum_{n=1}^{N}(y_{\text{val}} - \hat{y}_{\text{val}})^2} = 7 \times 10^{-5}$ compared to 2×10^{-5} before the decoupling was applied.

5. CONCLUSION

In this work an unconstrained multiple shooting algorithm is introduced. The method allows to tune the parameters of a nonlinear state-space model even when initialised within a non-contractive region of the parameter space.

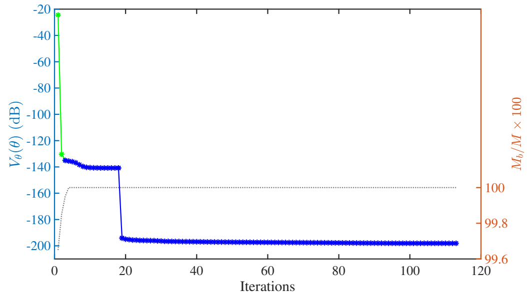


Fig. 5. Unconstrained multiple shooting applied to a non-contractive initialisation of the decoupled Bouc-Wen model (Eq. (16)). Green indicates steps computed using normalised gradient descent. Blue shows LM steps. The first blue plateau is reached for $M = 2000$. From thereon single shooting is used ($M = 1$) until the second plateau is reached. Grey indicates the portion of subrecords which are used for training (the records that produce bounded errors).

REFERENCES

Bertotti, G. (1998). *Hysteresis in Magnetism.* Academic Press, San Diego.

Bock, H.G. (1983). *Recent Advances in Parameteridentification Techniques for O.D.E. In Numerical Treatment of Inverse Problems in Differential and Integral Equations.*, volume 2 of *Progress in scientific computing.* Birkhäuser Boston.

Buchsbaum, T. (2007). Towards a winning GP strategy for continuous nonlinear dynamical system identification. *IEEE Congres on evolutionary computation*, 1269–1275.

Decuyper, J., Dreesen, P., Schoukens, J., Runacres, M., and Tiels, K. (2019a). Decoupling multivariate polynomials for nonlinear state-space models. *IEEE Control Systems Letters*, 3(3), 745–750.

Decuyper, J., Tiels, K., and Schoukens, J. (2019b). Retrieving highly structured models starting from black-box nonlinear state-space models using polynomial decoupling. *Mechanical Systems And Signal Processing*, Under review. Unpublished internal note.

Dreesen, P., De Geeter, J., and Ishteva, M. (2018). Decoupling multivariate functions using second-order information and tensors. In *Proc. 14th International Conference on Latent Variable Analysis and Signal Separation LVA/ICA 2018*, volume 10891, 79–88. Springer.

McKelvey, T., Akçay, H., and Ljung, L. (1996). Subspace-based multivariable system identification from frequency response data. *IEEE Transactions on Automatic Control*, 41, 960–979.

Morrison, D.D., Riley, J.D., and Zancanaro, J.F. (1962). Multiple shooting method for two-point boundary value problems. *Communications of the ACM*, 5(12), 613–614.

Morrison, D., Jia, Y., and Moosbrugger, J. (2001). Cyclic plasticity of nickel at low plastic strain amplitude: hysteresis loop shape analysis. *Materials Science and Engineering*, A314, 24–30.

Mueller, T. (1985). The influence of laminar separation and transition on low Reynolds number airfoil hysteresis. *AIAA Journal of Aircraft*, 22(9), 763–770.

Noël, J.P., Fakhrizadeh Esfahani, A., Kerschen, G., and Schoukens, J. (2017). A nonlinear state-space approach to hysteresis identification. *Mechanical Systems and*

Signal Processing, 84, 171–184.
Osborne, M.R. (1969). On shooting methods for boundary value problems. *Journal of Mathematical Analysis and Applications*, 27(2), 417–433.
Paduart, J., Lauwers, L., Swevers, J., Smolders, K., Schoukens, J., and Pintelon, R. (2010). Identification of nonlinear systems using polynomial nonlinear state space models. *Automatica*, 46, 647–657.
Pintelon, R. (2002). Frequency-domain subspace system identification using non-parametric noise models. *Automatica*, 38, 1295–1311.
Pintelon, R. and Schoukens, J. (2012). *System Identification: A Frequency Domain Approach, 2nd Edition.* Wiley-IEEE Press.
Ribeiro, A.H. and Aguirre, L.A. (2017). Shooting methods for parameter estimation of output error models. *IFAC-PapersOnLine*, 50(13998-14003).
Ribeiro, A.H., Tiels, K., Umenberger, J., Schön, T.B., and Aguirre, L.A. (2019). On the smoothness of nonlinear system identification. *Automatica*.
Schoukens, M. and Noël, J.P. (2017). Three benchmarks addressing open challenges in nonlinear system identification. In *20th World Congress of the International Federation of Automatic Control*, 448–453. Toulouse, France.
Shafique, A.B., Joshi, R., and Tsakalis, K. (2017). Control relevant system identification using multiple short data sets. *2017 IEEE Conference on Control Technology and Applications (CCTA)*, 1728–1733.
Tiels, K. (2016). *PNLSS 1.0 A polynomial nonlinear state-space toolbox for MATLAB.* Vrije Universiteit Brussel, http://homepages.vub.ac.be/jschouk/, 1st edition.
van Domselaar, B. and Hemker, P. (1975). Nonlinear parameter estiamtion in initial value problems. Technical Report NW 18/75, Mathematical centre, Amsterdam.
Van Mulders, A., Schoukens, J., Volckaert, M., and Diehl, M. (2010). Two nonlinear optimization methods for black box identification compared. *Automatica*, 46, 1675–1681.
Van Mulders, A. (2012). *Tackling two drawbacks of polynomial nonlinear state-space models.* Ph.D. thesis, Vrije Universiteit Brussel.

Appendix A. MODEL PARAMETERS OF THE NON-CONTRACTIVE LINEARISATION OF THE BOUC-WEN MODEL

In a typical PNLSS model, the nonlinear function in the state equation is modelled using a generic multivariate polynomial: $\mathbf{f}(\mathbf{x}(k), \mathbf{u}(k)) = \mathbf{E}\boldsymbol{\zeta}(\mathbf{x}(k), \mathbf{u}(k))$ with $\mathbf{E}$ a matrix of coefficients and $\boldsymbol{\zeta}$ a vector of monomial basis functions. As initialisation $\mathbf{E} = \mathbf{0}$. The parameters of the fourth order BLA are the following:

- $\mathbf{A} = \begin{bmatrix} 1.0402 & 0 & 0 & 0 \\ 0 & 0.9202 & -0.3146 & -0.0067 \\ 0 & 0.3146 & 0.9070 & -0.0648 \\ 0 & -0.0067 & 0.0648 & 0.9843 \end{bmatrix}$.
- $\mathbf{b}^T = [0.0017 \quad -0.0013 \quad 0.0010 \quad -0.0004]$.
- $\mathbf{c}^T = [-0.00002 \quad -0.0013 \quad -0.0010 \quad -0.0004]$.
- $d = 7.655ee{-}8$.

Appendix B. MODEL PARAMETERS OF THE DECOUPLED BOUC-WEN MODEL

In a *decoupled* PNLSS model, the nonlinear function in the state equation is modelled using a vector function $\mathbf{g}$ containing a set of r-univariate polynomials. Hence $\mathbf{f}(\boldsymbol{\theta}_g, \mathbf{x}(k), \mathbf{u}(k)) = \mathbf{W}\mathbf{g}\left(\mathbf{V}^T \begin{bmatrix} \mathbf{x}(k) \\ u(k) \end{bmatrix}\right)$, with the following parameters:

- $\mathbf{A} = \begin{bmatrix} 0.9384 & 0.2969 & 0.0217 \\ -0.3027 & 0.9271 & -0.0442 \\ 0.0963 & 0.0543 & 1.0042 \end{bmatrix}$.
- $\mathbf{b}^T = [-0.0015 \quad -0.0016 \quad 0.0005]$.
- $\mathbf{c}^T = [-0.0012 \quad 0.0009 \quad 0.0008]$.
- $d = 7.5029\text{e}{-}8$.
- $\mathbf{V} = \begin{bmatrix} 0.3896 & 0.6617 & -0.1549 & -0.1600 & -0.1758 & 0.4130 \\ -0.9088 & -0.6392 & -0.8475 & -0.3561 & 0.9482 & 0.7545 \\ 0.1494 & 0.3919 & -0.5077 & -0.9206 & 0.2646 & 0.5101 \\ 0.0022 & 0.0036 & 0.0009 & -0.0008 & -0.0018 & -0.0002 \end{bmatrix}$.
- $\mathbf{W} = \begin{bmatrix} -88.6374 & 32.8868 & 47.0168 & -28.0746 & -22.4437 & -16.8661 \\ 59.8645 & -40.5707 & -125.6225 & -11.6282 & 20.4816 & 139.7866 \\ -161.6076 & 114.1107 & 156.1952 & 23.3155 & -39.8400 & -169.9802 \end{bmatrix}$.
- $\boldsymbol{\theta}_g^T = \begin{bmatrix} 42646\text{e}{-}7 & 31168\text{e}{-}7 & 18251\text{e}{-}7 & 40359\text{e}{-}8 & 76467\text{e}{-}7 & 16227\text{e}{-}7 \\ 0.00395 & 0.00225 & -0.00725 & 0.00444 & 0.02006 & 0.004150 \end{bmatrix}$.