Iterative learning control for varying tasks: achieving optimality for rational basis functions

DOI:
10.1109/ACC.2015.7171884

Document status and date:
Published: 01/07/2015

Document Version:
Accepted manuscript including changes made at the peer-review stage

Iterative Learning Control for Varying Tasks: Achieving Optimality for Rational Basis Functions

Jurgen van Zundert, Joost Bolder and Tom Oomen - `ACC2015_IV_v99(31/05/2018)`

***Abstract*— Iterative Learning Control (ILC) can achieve superior tracking performance for systems that perform repeating tasks. However, the performance of standard ILC deteriorates dramatically when the task is varied. In this paper ILC is extended with rational basis functions to obtain excellent extrapolation properties. A new approach for rational basis functions is proposed where the iterative solution algorithm is of the form used in instrumental variable system identification algorithms. The optimal solution is expressed in terms of learning filters similar as in standard ILC. The proposed approach is shown to be superior over existing approaches in terms of performance by a simulation example.**

I. Introduction

For systems performing the same task over and over, the performance can be optimized by learning from previous executions. In Iterative Learning Control (ILC) [1], [2], [3], [4], the repetitive behavior is exploited by updating the command signal using data from previous executions.

ILC achieves optimal performance for a specific task only since the command signal is learned. For varying tasks the performance may severely deteriorate [5]. The main reason is that the command signal is not a function of the reference trajectory. In [6], the extrapolation properties are enhanced by constructing the task such that it comprises basis tasks that are learned in a training routine. Consequently, this approach is limited to tasks which can be constructed by a finite set of basis tasks. A more general approach is to parameterize the command signal in a set of basis functions [7], [8]. Examples of basis functions include polynomial basis functions [9], [10], [11], [12], and more recently rational basis functions [13]. Rational basis functions are more general than polynomial basis functions, since the latter are recovered as a special case.

The optimization associated with polynomial basis functions in ILC has an explicit analytic solution [3], whereas this is generally not the case for rational basis functions. In [13], an iterative solution based on Steiglitz-McBride is presented. The approach achieves fast convergence and is insensitive to local optima [14], however, the stationary point is not necessarily an optimum, as is well known in related system identification algorithms [15].

The authors are with the Eindhoven University of Technology, Department. of Mechanical Engineering, Control Systems Technology group, P.O. Box 513, 5600 MB Eindhoven, The Netherlands. E-mail: `j.c.d.v.zundert@tue.nl`. This research is supported by Océ Technologies, P.O. Box 101, 5900 MA Venlo, The Netherlands; and the Innovational Research Incentives Scheme under the VENI grant "Precision Motion: Beyond the Nanometer" (no. 13073) awarded by NWO (The Netherlands Organization for Scientific Research) and STW (Dutch Science Foundation).

Although important contributions have been made by developing general rational basis functions in ILC, presently available optimization algorithms suffer from non-optimality or poor convergence properties.

The aim of this paper is to show non-optimality of the pre-existing approach and develop a new solution algorithm for which the stationary point is always an optimum. The proposed approach has strong connections to instrumental variable system identification [16]. The contributions of this paper are threefold:

- I Non-optimality of the pre-existing approach in [13] for rational basis functions in ILC is illustrated.
- II A new iterative solution algorithm for rational basis functions in ILC is proposed and its optimality is shown which constitutes the main contribution of this paper.
- III It is shown that the proposed approach outperforms the pre-existing approach by a simulation example.

The outline of this paper is as follows. In section II, the problem considered in this paper is introduced. The non-optimality of the pre-existing approach [13] is illustrated in section III. The proposed approach is presented in section IV. The iterative approaches are compared by use of a simulation example in section V, demonstrating that the proposed approach outperforms the pre-existing approach. Section VI contains conclusions.

II. Problem formulation

In this section the considered problem is defined.

A. Notation

In this paper discrete-time, linear, time-invariant (LTI), singe-input, single-output (SISO) systems are considered to facilitate the presentation.

Let $z \in \mathbb{C}$ be the complex indeterminate and $x(k)$ the signal x evaluated at time index k. Let $h(l)$ be the impulse response of the system $H(z)$. The output $y(k)$ of the response of $H(z)$ to input u is given by the [17] $y(k) = \sum_{l=-\infty}^{\infty} h(l)u(k-l)$. Assuming $u(k) = 0$ for $k < 0$ and $k > N-1$, then

$$\underbrace{\begin{bmatrix} y[0] \\ y[1] \\ \vdots \\ y[N-1] \end{bmatrix}}_{y} = \underbrace{\begin{bmatrix} h(0) & h(-1) & \dots & h(1-N) \\ h(1) & h(0) & \dots & h(2-N) \\ \vdots & \vdots & \ddots & \vdots \\ h(N-1) & h(N-2) & \dots & h(0) \end{bmatrix}}_{H} \underbrace{\begin{bmatrix} u[0] \\ u[1] \\ \vdots \\ u[N-1] \end{bmatrix}}_{u}, \quad (1)$$

with H the finite-time matrix representation of $H(z)$, and $u, y \in \mathbb{R}^N$ the input and output, respectively, with $N \in \mathbb{Z}^+$ the trial length. Let j be the trial index and $\|x_j\|_W :=$

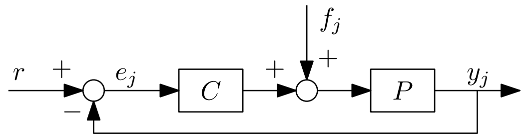


Fig. 1. Closed-loop system under consideration.

$x_j^\top W x_j$, where $x_j \in \mathbb{R}^N$ and $W \in \mathbb{R}^{N\times N}$. W is positive definite if $x^\top W x > 0$, $\forall x \neq 0$ and positive semi-definite if $x^\top W x \geq 0$, $\forall x$.

Let a system $G(\theta, z)$ be linearly parameterized as

$$G(\theta, z) = \sum_{n=0}^{m-1} x_n(z)\theta[n],$$

with parameters $\theta \in \mathbb{R}^m$ and basis functions $\xi_n(z)$, which is equivalent to

$$y = \Psi_{Gu}\theta, \tag{2}$$

with $\Psi_{Gu} = \begin{bmatrix}\xi_0 u, & \xi_1 u, & \ldots, & \xi_{m-1}u\end{bmatrix} \in \mathbb{R}^{N\times m}$ where $\xi_n \in \mathbb{R}^{N\times N}$ are finite-time matrix representations of $\xi_n(z)$ according to (1).

Example 1. *Consider $G(\theta, z)$ with $N = 3$ and $\xi_0(z) = z$, $\xi_1(z) = 1$, $\xi_2(z) = z^{-2}$, i.e., $G(z) = \theta[0]z + \theta[1] + \theta[2]z^{-2}$. Using* (1)*:*

$$\underbrace{\begin{bmatrix} y[0] \\ y[1] \\ y[2] \end{bmatrix}}_{y} = \underbrace{\begin{bmatrix} \theta[1] & \theta[0] & 0 \\ 0 & \theta[1] & \theta[0] \\ \theta[2] & 0 & \theta[1] \end{bmatrix}}_{G} \underbrace{\begin{bmatrix} u[0] \\ u[1] \\ u[2] \end{bmatrix}}_{u} = \underbrace{\begin{bmatrix} u[1] & u[0] & 0 \\ u[2] & u[1] & 0 \\ 0 & u[2] & u[0] \end{bmatrix}}_{\Psi_{Gu}} \underbrace{\begin{bmatrix} \theta[0] \\ \theta[1] \\ \theta[2] \end{bmatrix}}_{\theta},$$

where the latter is in the form of (2)*.*

B. System description

The considered ILC setup is shown in Fig. 1. Here is $P \in \mathbb{R}^{N\times N}$ a single-input, single output system and $C \in \mathbb{R}^{N\times N}$ a stabilizing feedback controller. The aim is to determine the feedforward $f_j \in \mathbb{R}^N$ such that the output $y_j \in \mathbb{R}^N$ follows the reference $r \in \mathbb{R}^N$ as accurately as possible, i.e., such that the tracking error $e_j = r - y_j$ is minimized. The tracking error for trial j is given by

$$e_j = Sr - SPf_j, \tag{3}$$

with sensitivity $S := (I + PC)^{-1}$. The tracking error evaluated at trial $j + 1$ is given by

$$e_{j+1} = Sr - SPf_{j+1}. \tag{4}$$

Eliminating Sr from (4) by (3) yields the tracking error propagation from trial j to trial $j + 1$:

$$e_{j+1} = e_j + SP\left(f_j - f_{j+1}\right).$$

C. Norm-optimal ILC

The goal is to minimize the error signal e_{j+1} at the next trial using f_{j+1}. ILC [1] enables the determination of f_{j+1} through an iterative procedure that only requires approximate model knowledge.

Norm-optimal ILC is an important class of ILC in which the feedforward signal f_{j+1} for the next trial is selected as the signal that minimizes the performance criterion in Definition 2.

Definition 2 (Performance criterion norm-optimal ILC)**.** *The performance criterion for norm-optimal ILC is given by*

$$\mathcal{J}(f_{j+1}) := \|e_{j+1}\|_{W_e} + \|f_{j+1}\|_{W_f} + \|f_{j+1} - f_j\|_{W_{\Delta f}}, \tag{5}$$

with $W_e \in \mathbb{R}^{N\times N}$ a symmetric, positive definite weighting matrix, and $W_f, W_{\Delta f} \in \mathbb{R}^{N\times N}$ symmetric, positive semi-definite weighting matrices.

Norm-optimal ILC is well-known and can be found in, for example, [1], [18]. In norm-optimal ILC the signal f_j is learned over the trials. However, the optimal feedforward signal yielding $e_j = 0$ in (3) is given by $f_j = P^{-1}r$, under the assumption that P is invertible. This case corresponds to inverse model feedforward and shows that the optimal feedforward signal is a function of the reference signal. Hence, the learned signal in norm-optimal ILC will only be optimal for one specific reference signal and non-optimal for different reference signals. In order to introduce extrapolation properties, basis functions are exploited in the next section.

D. Introducing extrapolation properties with basis functions

Inspired by inverse model feedforward, extrapolation properties are introduced in ILC by use of basis functions [9]:

$$f_j = F(\theta_j)r, \quad \theta_j \in \mathbb{R}^m. \tag{6}$$

Substitution of (6) in (3) yields

$$e_j = S(I - PF(\theta_j))r.$$

Hence, if $F(\theta_j) = P^{-1}$, $e_j = 0$ $\forall r$.

In this paper rational basis functions are chosen for the feedforward filter $F(\theta_j)$ (see Definition 3 and Fig. 2) since it allows to fully describe the plant inverse P^{-1} and thereby obtain perfect tracking.

Definition 3 (Rational basis functions)**.** *Rational basis functions in the parameters $\theta_j \in \mathbb{R}^m$ with reference r as basis are defined as in* (6) *with $F(\theta_j)$ the matrix representation of $F(\theta_j, z) \in \mathcal{F}$,*

$$\mathcal{F} = \left\{ \frac{A(\theta_j, z)}{B(\theta_j, z)} \middle| \, \theta_j \in \mathbb{R}^{m_a+m_b} \right\},$$

with

$$A(\theta_j, z) = \sum_{n=0}^{m_a-1} \xi_n^A(z)\theta_j[n],$$

$$B(\theta_j, z) = 1 + \sum_{n=0}^{m_b-1} \xi_n^B(z)\theta_j[m_a + n],$$

where $\xi_n^A(z)$, $n = 0, 1, 2, \ldots, m_a-1$, and $\xi_n^B(z)$, $n = 0, 1, 2, \ldots, m_b-1$ are basis functions.

The polynomial basis functions in [9], [10], [11], [12] are recovered by setting $m_b = 0$. Interestingly, an analytic solution for the case $m_b = 0$ is presented in [9]. Such analytic solution does not exist for the general case $m_b > 0$.

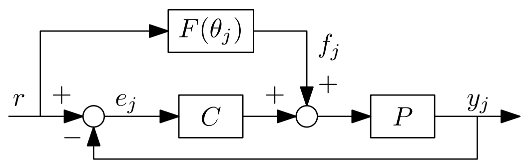


Fig. 2. Implementation of rational basis functions (Definition 3) in Fig. 1.

Evaluating (5) for (6) yields the performance criterion stated in Definition 4, which is a function of the parameters θ_{j+1}. Instead of determining the optimal feedforward signal, the optimal parameters are to be determined. Since $m = \dim(\theta_{j+1}) \ll \dim(f_{j+1}) = N$, additional robustness against model uncertainties is introduced.

Definition 4 (Performance criterion ILC with basis functions)**.** *The performance criterion for norm-optimal ILC with basis functions is given by*

$$\begin{aligned}\mathcal{J}(\theta_{j+1}) := &\|e_{j+1}(\theta_{j+1})\|_{W_e} + \\ &\|f_{j+1}(\theta_{j+1})\|_{W_f} + \\ &\|f_{j+1}(\theta_{j+1}) - f_j(\theta_j)\|_{W_{\Delta f}},\end{aligned} \tag{7}$$

with $W_e \in \mathbb{R}^{N\times N}$ a symmetric, positive definite weighting matrix, and $W_f, W_{\Delta f} \in \mathbb{R}^{N\times N}$ symmetric, positive semi-definite weighting matrices.

E. Problem formulation

The goal in this paper is to solve Problem 5.

Problem 5 (Main problem)**.** *Given Definition 3 and θ_j, determine*

$$\theta^*_{j+1} = \arg\min_{\theta_{j+1}} \mathcal{J}(\theta_{j+1}),$$

with $\mathcal{J}(\theta_{j+1})$ in Definition 4.

III. Non-optimality pre-existing approach

In this section the non-optimality of the pre-existing approach of [13] is demonstrated, forming contribution I.

A. Pre-existing approach

Similar to standard norm-optimal ILC [1], there is an analytic solution to Problem 5 if both e_{j+1} and f_{j+1} are linear in θ_{j+1}, which is not the case for $m_b > 0$. In the pre-existing approach, which is closely related to Steiglitz-McBride system identification, the performance criterion in Definition 6 is considered. Note that if $\theta^{\langle q\rangle}_{j+1} = \theta^{\langle q-1\rangle}_{j+1} = \theta_{j+1}$, then $\hat{\mathcal{J}}(\theta^{\langle q\rangle}_{j+1}) = \mathcal{J}(\theta_{j+1})$.

Definition 6 (Weighted performance criterion)**.**

$$\begin{aligned}\hat{\mathcal{J}}(\theta^{\langle q\rangle}_{j+1}) := &\left\|B^{-1}(\theta^{\langle q-1\rangle}_{j+1})B(\theta^{\langle q\rangle}_{j+1})e^{\langle q\rangle}_{j+1}\right\|_{W_e} + \\ &\left\|B^{-1}(\theta^{\langle q-1\rangle}_{j+1})B(\theta^{\langle q\rangle}_{j+1})f^{\langle q\rangle}_{j+1}\right\|_{W_f} + \\ &\left\|B^{-1}(\theta^{\langle q-1\rangle}_{j+1})B(\theta^{\langle q\rangle}_{j+1})f^{\langle q\rangle}_{j+1} - f_j\right\|_{W_{\Delta f}}.\end{aligned}$$

Note that $\hat{\mathcal{J}}(\theta^{\langle q\rangle}_{j+1})$ is quadratic in $\theta^{\langle q\rangle}_{j+1}$. Hence, there is a unique solution for the optimal parameters $\theta^{\langle q\rangle *}_{j+1}$, which can be determined analytically from:

$$\left(\frac{d\hat{\mathcal{J}}(\theta^{\langle q\rangle}_{j+1})}{d\theta^{\langle q\rangle}_{j+1}}\right)^{\top}\Bigg|_{\theta^{\langle q\rangle}_{j+1} = \theta^{\langle q\rangle *}_{j+1}} = 0. \tag{8}$$

The idea is to iteratively determine the optimal parameters $\theta^{\langle q\rangle *}_{j+1}$ for $\hat{\mathcal{J}}(\theta^{\langle q\rangle}_{j+1})$ in Definition 6, using (8). The idea is that upon convergence of the parameters, i.e., $\theta^{\langle q\rangle}_{j+1} \to \theta^{\langle q-1\rangle}_{j+1}$, $\theta^{\langle q\rangle *}_{j+1}$ are also the optimal parameters for Problem 5, because $\mathcal{J}(\theta_{j+1})$ is recovered from $\hat{\mathcal{J}}(\theta^{\langle q\rangle}_{j+1})$ for $\theta^{\langle q\rangle}_{j+1} = \theta^{\langle q-1\rangle}_{j+1} = \theta_{j+1}$. There is thus an analytic solution which after convergence provides the solution to Problem 5. The procedure of the iterative algorithm is formulated in Algorithm 7.

Algorithm 7 (Pre-existing algorithm)**.** *The pre-existing algorithm [13] for solving Problem 5 is given by the following sequence of steps.*

1) *Let r_j, f_j, e_j be given, set $q = 1$, and initialize $\theta^{\langle q-1\rangle}_{j+1} = \theta_j$.*
2) *Compute $\theta^{\langle q\rangle *}_{j+1}$ from (8).*
3) *Set $q \to q+1$ and go back to 2) until an appropriate stopping criterion is satisfied.*

The advantage of Algorithm 7 over Gauss-Newton iteration is that Algorithm 7 is insensitive for local optima and typically converges in a few iterations [14].

B. Non-optimality pre-existing approach

In the pre-existing approach (8) is solved which yields the minimum of $\hat{\mathcal{J}}(\theta^{\langle q\rangle}_{j+1})$. However,

$$\frac{d\hat{\mathcal{J}}(\theta^{\langle q\rangle}_{j+1})}{d\theta^{\langle q\rangle}_{j+1}} = 0 \quad \not\Rightarrow \quad \frac{d\mathcal{J}(\theta_{j+1})}{d\theta_{j+1}} = 0.$$

Consequently, there is no guarantee that the found parameters are also the optimal parameters for $\mathcal{J}(\theta_{j+1})$ and therefore the performance is non-optimal. A detailed proof is beyond the scope of the present paper and will be published elsewhere. The non-optimality of the pre-existing approach is confirmed by a simulation example in section V.

IV. Proposed approach

In this section a new iterative solution algorithm is proposed and its optimality is shown which constitutes contribution II.

The starting point of the proposed approach is

$$\frac{d\mathcal{J}(\theta_{j+1})}{d\theta_{j+1}} = 0, \tag{9}$$

with the gradient given by Lemma 9, exploiting the auxiliary result in Lemma 8. Note that the gradient is not linear in θ_{j+1} for $m_b > 0$. Hence, there will in general be no analytic solution to (9).

Lemma 8 (Gradient quadratic matrix function)**.** *Let $x, b \in \mathbb{R}^N$, $A \in \mathbb{R}^{N\times N}$, and $W^\top = W \in \mathbb{R}^{N\times N}$, then*

$$\frac{d\left(\|Ax+b\|_W\right)}{dx} = 2(Ax+b)^\top WA.$$

Proof. Generalization of Proposition 10.7.1 in [19]. ☐

Lemma 9 (Gradient performance criterion)**.** *The gradient of the performance criterion $\mathcal{J}(\theta_{j+1})$ in Definition 4 with respect to the parameters θ_{j+1} is given by*

$$\begin{aligned}\left(\frac{d\mathcal{J}(\theta_{j+1})}{d\theta_{j+1}}\right)^\top &= \\ &-2\left(\frac{df_{j+1}}{d\theta_{j+1}}\right)^\top \left((SP)^\top W_e SP + W_{\Delta f}\right) f_j \\ &-2\left(\frac{df_{j+1}}{d\theta_{j+1}}\right)^\top (SP)^\top W_e e_j \\ &+2\left(\frac{df_{j+1}}{d\theta_{j+1}}\right)^\top \left((SP)^\top W_e SP + W_f + W_{\Delta f}\right)\cdot \\ &\quad B^{-1}(\theta_{j+1})A(\theta_{j+1})r_j.\end{aligned}$$

where $\frac{df_{j+1}}{d\theta_{j+1}} = Kr_j$, with K the matrix representation of

$$\begin{aligned}K(z) =& \frac{1}{B(\theta_{j+1},z)}\frac{dA(\theta_{j+1},z)}{d\theta_{j+1}} \\ &- \frac{A(\theta_{j+1},z)}{B^2(\theta_{j+1},z)}\frac{dB(\theta_{j+1},z)}{d\theta_{j+1}}.\end{aligned}$$

Proof. Follows from applying Lemma 8 to (7). ☐

The second step is to apply a weighting similar as in the pre-existing approach, see Definition 10. Lemma 9 is recovered from Definition 10 for $\theta_{j+1}^{\langle q\rangle} = \theta_{j+1}^{\langle q-1\rangle} = \theta_{j+1}$.

Definition 10 (Weighted gradient of performance criterion)**.** *Let*

$$\begin{aligned}\left(\overline{\frac{d\mathcal{J}(\theta_{j+1}^{\langle q\rangle})}{d\theta_{j+1}^{\langle q\rangle}}}\right)^\top &= \\ &-2\zeta^{\langle q\rangle}\left((SP)^\top W_e SP + W_{\Delta f}\right) B(\theta_{j+1}^{\langle q\rangle}) f_j \\ &-2\zeta^{\langle q\rangle}(SP)^\top W_e B(\theta_{j+1}^{\langle q\rangle}) e_j \\ &+2\zeta^{\langle q\rangle}\left((SP)^\top W_e SP + W_f + W_{\Delta f}\right) A(\theta_{j+1}^{\langle q\rangle}) r_k,\end{aligned} \qquad (10)$$

with

$$\zeta^{\langle q\rangle} = \left(\overline{\frac{df_{j+1}^{\langle q-1\rangle}}{d\theta_{j+1}^{\langle q-1\rangle}}}\right)^\top B^{-1}(\theta_{j+1}^{\langle q-1\rangle}), \quad \overline{\frac{df_{j+1}^{\langle q-1\rangle}}{d\theta_{j+1}^{\langle q-1\rangle}}} = \bar{K}r_j$$

with $\bar{K}$ the matrix representation of

$$\begin{aligned}\bar{K}(z) =& \frac{1}{B(\theta_{j+1}^{\langle q-1\rangle},z)}\frac{dA(\theta_{j+1}^{\langle q-1\rangle},z)}{d\theta_{j+1}} \\ &- \frac{A(\theta_{j+1}^{\langle q-1\rangle},z)}{B^2(\theta_{j+1}^{\langle q-1\rangle},z)}\frac{dB(\theta_{j+1}^{\langle q-1\rangle},z)}{d\theta_{j+1}^{\langle q-1\rangle}}.\end{aligned}$$

The term $\zeta^{\langle q\rangle}$ in (10) is not a function of $\theta_{j+1}^{\langle q\rangle}$; (10) only depends on $\theta_{j+1}^{\langle q\rangle}$ through linear dependencies on $A(\theta_{j+1}^{\langle q\rangle})$ and $B(\theta_{j+1}^{\langle q\rangle})$. Since both these filters are linear in $\theta_{j+1}^{\langle q\rangle}$, (10) is also linear in $\theta_{j+1}^{\langle q\rangle}$. Therefore the solution of equating (10) to zero has an analytic solution, see Theorem 11.

Theorem 11 (Optimal parameters for weighted gradient performance criterion)**.** *The optimal parameters $\theta_{j+1}^{\langle q\rangle *}$ of* (10) *in Definition 10 are given by*

$$\theta_{j+1}^{\langle q\rangle *} = \left(\zeta^{\langle q\rangle}\Psi^{\langle q\rangle}\right)^{-1}\zeta^{\langle q\rangle}\left(Q^{\langle q\rangle}f_j + L^{\langle q\rangle}e_j\right), \qquad (11)$$

with

$$\begin{aligned}\Psi^{\langle q\rangle} &= \Big[\left((SP)^\top W_e SP + W_f + W_{\Delta f}\right)\Psi^A_{r_j}\,, \;\ldots \\ &\quad -(SP)^\top W_e \Psi^B_{e_j} - \left((SP)^\top W_e SP + W_{\Delta f}\right)\Psi^B_{f_j}\Big], \\ Q^{\langle q\rangle} &= (SP)^\top W_e SP + W_{\Delta f}, \\ L^{\langle q\rangle} &= (SP)^\top W_e.\end{aligned}$$

Proof. Due to space restrictions, the proof will be published elsewhere. ☐

The final step is the formulation of the algorithm. In the previous steps two key elements are derived. First, a weighted gradient is introduced in Definition 10, from which the gradient in Lemma 9 is recovered for $\theta_{j+1}^{\langle q\rangle} = \theta_{j+1}^{\langle q-1\rangle} = \theta_{j+1}$. Second, an analytic solution for the parameters $\theta_{j+1}^{\langle q\rangle *} = \theta_{j+1}^{\langle q\rangle}$ is obtained for which the weighted gradient in Definition 10 is zero. Hence, upon convergence, the actual gradient also converges to zero and optimal performance is achieved. Combining these two elements, there is an analytic solution to Problem 5 when there is convergence in the parameters, i.e., $\theta_{j+1}^{\langle q\rangle} \to \theta_{j+1}^{\langle q-1\rangle}$. The iterative algorithm to obtain this solution is given by Algorithm 12. The proposed algorithm is closely related to instrumental variable system identification [16].

Algorithm 12 (Proposed algorithm)**.** *The proposed algorithm for solving Problem 5 is given by the following sequence of steps.*

1) *Let r_j, f_j, e_j be given, set $q = 1$, and initialize $\theta_{j+1}^{\langle q-1\rangle} = \theta_j$.*
2) *Compute $\theta_{j+1}^{\langle q\rangle *}$ from* (11)*.*
3) *Set $q \to q+1$ and go back to 2) until an appropriate stopping criterion is satisfied.*

V. Simulation Example

In this section the pre-existing and proposed approach for solving ILC with rational basis functions are analyzed by use of simulation. This section constitutes contribution III.

A. System description

The system under consideration is an Océ Arizona 550 GT flatbed printer, see Fig. 3. In this simulation example only the movement in y-direction is considered. Given are the transfer from the input current of the carriage motor [A] to the output position of the carriage [m] as identified using

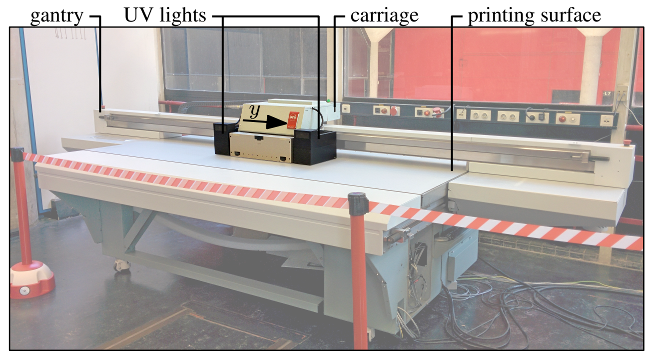


Fig. 3. Océ Arizona 550 GT at the CST Motion laboratory, Eindhoven University of Technology. The carriage moves in y-direction over the gantry which can be moved parallel to the y-direction over the printing surface.

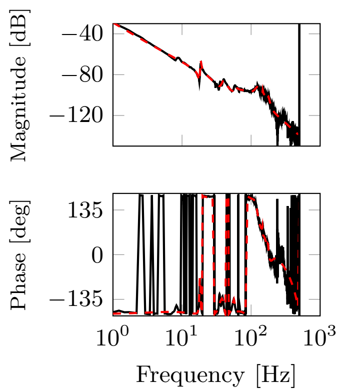


Fig. 4. Bode plot of the FRF —— and identified 20th order model - - - of the transfer of the y-position of the carriage.

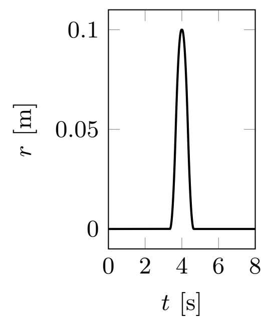


Fig. 5. The reference signal r consists of a symmetric 4th order polynomial.

a frequency response function (FRF) measurement. A 20th order model based on this FRF measurement is given and used as plant model. The Bode plots of the FRF measurement and the model are depicted in Fig. 4. The sampling time is $T_s = 0.001$ s. The feedback controller consists of a weak integrator, a lead-filter and a first-order low-pass filter, and yields a bandwidth of 25 Hz.

B. Simulation setup

Fig. 5 depicts the reference signal r which consists of a fourth order polynomial, preceded and followed by zeros to allow for pre- and post-actuation, respectively and has a trial length of $N = 8000$ samples. The weighting filters in Definition 4 are set to $W_e = 10^6 I$, $W_f = W_{\Delta f} = 0$. Rational basis functions (Definition 3) are exploited with

$$A(z) = \gamma(z-1)^2, \quad |A(z)P(z)|_{z=1} = 1,$$
$$B(\theta_j, z) = 1 + \sum_{n=0}^{1} \xi_n^B \theta_j[n], \quad \theta_j \in \mathbb{R}^2,$$

with $\xi_0^B(z) = \frac{z-1}{zT_s}$, and $\xi_1^B(z) = (\frac{z-1}{zT_s})^2$, respectively, i.e., a first- and second-order differentiator.

C. Simulation results

The results are depicted in Fig. 6 and Fig. 7. From analysis of the performance criterion over a grid of parameters, it is concluded that there is only one minimum, namely for $\theta_{j+1} \approx \begin{bmatrix} 3.13, & -0.18 \end{bmatrix}^\top$.

Next, the convergence behavior of both approaches is analyzed. The new parameters $\theta_{j+1}^{\langle q \rangle}$ only depend on the previous parameters $\theta_{j+1}^{\langle q-1 \rangle}$. Therefore, given $\theta_{j+1}^{\langle q-1 \rangle}$, an approach will always yield the same parameters $\theta_{j+1}^{\langle q \rangle}$. For both iterative approaches, the direction of $\theta_{j+1}^{\langle 0 \rangle} \rightarrow \theta_{j+1}^{\langle 1 \rangle}$ is indicated in Fig. 6a and Fig. 7a for a grid in θ_{j+1}. The result is a vector field indicating the direction the parameters $\theta_{j+1}^{\langle 1 \rangle}$ move to given $\theta_{j+1}^{\langle 0 \rangle} = \theta_j$. Moreover, the evolution of the parameters $\theta_{j+1}^{\langle q \rangle}$ for $q = 0, 1, 2, \dots, 7$ of three trajectories is displayed: starting from the upper-left (◇) and lower-right (□) corner, and from the optimum (). The corresponding performance criterion $\hat{\mathcal{J}}(\theta_{j+1}^{\langle 1 \rangle})$ is displayed in Fig. 6b and Fig. 7b. The 2-norm of the gradient with respect to the parameters is depicted in Fig. 6c and Fig. 7c. The figures indicate that both approaches converge within a couple of iterations.

Pre-existing approach: Fig. 6a indicates that the pre-existing approach always converges to a non-optimal stationary point. This can also be observed in Fig. 6c, which shows that the gradient after convergence is still considerably large. Even if the approach is initialized in the optimum (), it diverges to the non-optimal stationary point with poorer performance.

Proposed approach: In contrast, for the proposed approach the gradients for all three initial conditions converge to a value close to zero, i.e., the approach converges to the minimum. Correspondingly, the value of $\hat{\mathcal{J}}(\theta_{j+1}^{\langle q \rightarrow \infty \rangle})$ is also significantly smaller than for the pre-existing approach.

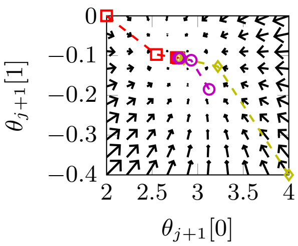
D. Conclusion on simulation example

The simulation example shows that the pre-existing approach is non-optimal. Even if initialized in the optimum, it diverges to a stationary point with worse performance. In contrast, the proposed approach converges to a stationary point close to the optimum.

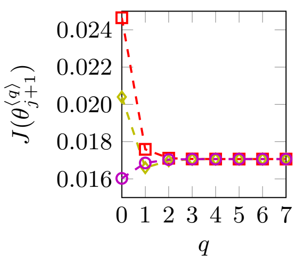
For cases with multiple optima the stationary point of gradient-based algorithms, such as the Gauss-Newton algorithm, depends on the initial parameters. In contrast, the proposed algorithm in this case converges to the global optimum, independent of the initial parameters [14].

VI. Conclusion

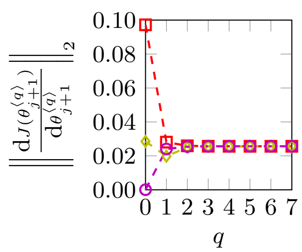
In this paper, extrapolation properties of norm-optimal ILC are enhanced through use of rational basis functions. The associated optimization problem is significantly more challenging than for polynomial basis functions. The main contribution of this paper is a new iterative solution algorithm for rational basis functions in ILC. The proposed approach has advantageous properties compared to the pre-existing approach [13], as is well-known from a related algorithm



(a) Vector field and trajectories.

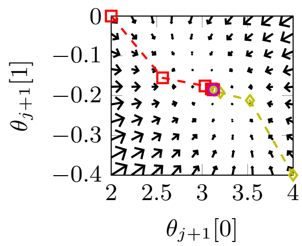


(b) Performance criterion.

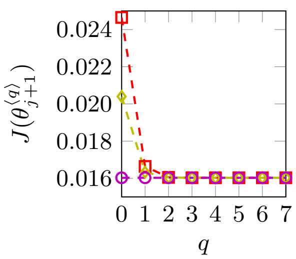


(c) Norm of gradient

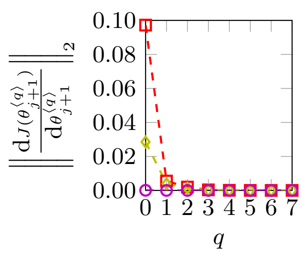
Fig. 6. The pre-existing approach converges to a stationary point that is different than the optimum, even when starting in the optimum.



(a) Vector field and trajectories.



(b) Performance criterion.



(c) Norm of gradient

Fig. 7. The proposed approach always converges to the optimum, independent of the initial parameters $\theta_{j+1}^{\langle 0 \rangle}$.

used in system identification. In particular, upon convergence the proposed approach is guaranteed to be in a minimum, whereas this is not the case for the pre-existing approach. As a result, the proposed approach outperforms the pre-existing approach.

The convergence behavior of both approaches is analyzed using simulations of a complex industrial system. The simulations confirm that the proposed approach is superior over the pre-existing approach in terms of performance.

Currently, the simulation results are experimentally validated. Ongoing research focuses on: application to MIMO systems, selection of basis functions, robustness analysis, and numerical conditioning along the lines of [20].

Acknowledgment

The authors would like to acknowledge fruitful discussions with and the contributions of Sjirk Koekebakker and Maarten Steinbuch that have significantly contributed to the results presented in this paper.

References

[1] D. Bristow, M. Tharayil, and A. Alleyne, “A Survey of Iterative Learning Control,” *Control Systems Magazine*, vol. 26, no. 3, pp. 96–114, 2006.

[2] K. L. Moore, *Iterative Learning Control for Deterministic Systems*. Springer-Verlag, 1993.

[3] S. Gunnarsson and M. Norrlöf, “On the design of ILC algorithms using optimization,” *Automatica*, vol. 37, no. 12, pp. 2011–2016, 2001.

[4] E. Rogers, K. Galkowski, and D. H. Owens, *Control Systems Theory and Applications for Linear Repetitive Processes*. Springer, 2007, vol. 349.

[5] S. Gunnarsson and M. Norrlöf, “On the disturbance properties of high order iterative learning control algorithms,” *Automatica*, vol. 42, no. 11, pp. 2031–2034, 2006.

[6] D. J. Hoelzle, A. G. Alleyne, and A. J. W. Johnson, “Basis Task Approach to Iterative Learning Control With Applications to Micro-Robotic Deposition,” *Transactions on Control Systems Technology*, vol. 19, no. 5, pp. 1138–1148, 2011.

[7] M. Q. Phan and J. A. Frueh, “Learning Control for Trajectory Tracking using Basis Functions,” in *Proceedings of the 35th Conference on Decision and Control*, vol. 3, Kobe, Japan, 1996, pp. 2490–2492.

[8] S. J. Oh, M. Q. Phan, and R. W. Longman, “Use of Decoupling Basis Functions in Learning Control for Local Learning and Improved Transients,” *Advances in the Astronautical Sciences*, vol. 95, no. 2, pp. 651–670, 1997.

[9] J. van de Wijdeven and O. Bosgra, “Using basis functions in iterative learning control: analysis and design theory,” *International Journal of Control*, vol. 83, no. 4, pp. 661–675, 2010.

[10] J. Bolder, T. Oomen, S. Koekebakker, and M. Steinbuch, “Using iterative learning control with basis functions to compensate medium deformation in a wide-format inkjet printer,” *Mechatronics*, vol. 24, no. 8, pp. 944–953, 2014.

[11] M. F. Heertjes and R. M. van de Molengraft, “Set-point variation in learning schemes with applications to wafer scanners,” *Control Engineering Practice*, vol. 17, no. 3, pp. 345–356, 2009.

[12] X. Gao and S. Mishra, “An Iterative Learning Control Algorithm for Portability Between Trajectories,” in *Proceedings of the 2014 American Control Conference*, Portland, Oregon, 2014, pp. 3808–3813.

[13] J. Bolder and T. Oomen, “Rational Basis Functions in Iterative Learning Control - With Experimental Verification on a Motion System,” *Transactions on Control Systems Technology*, vol. 23, no. 2, pp. 722–729, 2015.

[14] C. Bohn and H. Unbehauen, “Minmax and Least Squares Multivariable Transfer Function Curve Fitting: Error Criteria, Algorithms and Comparisons,” in *Proceedings of the 1998 American Control Conference*, vol. 5, Philadelphia, Pennsylvania, 1998, pp. 3189–3193.

[15] A. H. Whitfield, “Asymptotic behaviour of transfer function synthesis methods,” *International Journal of Control*, vol. 45, no. 3, pp. 1083–1092, 1987.

[16] M. Gilson, H. Garnier, P. Young, and P. Van den Hof, “Optimal instrumental variable method for closed-loop identification,” *Control Theory & Applications*, vol. 5, no. 10, pp. 1147–1154, 2011.

[17] A. V. Oppenheim, A. S. Willsky, and S. H. Nawab, *Signals and systems*, 2nd ed. Prentice-Hall, Inc., 1997.

[18] S. Gunnarsson and M. Norrlöf, “Some Aspects of an Optimization Approach to Iterative Learning Control,” in *Proceedings of the 38th Conference on Decision and Control*, vol. 2, Phoenix, Arizona, 1999, pp. 1581–1586.

[19] D. S. Bernstein, *Matrix mathematics: theory, facts, and formulas*, 2nd ed. Princeton University Press, 2009.

[20] R. van Herpen, T. Oomen, and M. Steinbuch, “Optimally Conditioned Instrumental Variable Approach for Frequency-Domain System Identification,” *Automatica*, vol. 50, no. 9, pp. 2281–2293, 2013.