Optimality and Flexibility in Iterative Learning Control for Varying Tasks

Jurgen van Zundert [a], Joost Bolder [a], Tom Oomen [a]

[a] *Eindhoven University of Technology, Department of Mechanical Engineering, Control Systems Technology group P.O. Box 513, 5600 MB Eindhoven, The Netherlands*

Abstract

Iterative Learning Control (ILC) can significantly enhance the performance of systems that perform repeating tasks. However, small variations in the performed task may lead to a large performance deterioration. The aim of this paper is to develop a novel ILC approach, by exploiting rational basis functions, that enables performance enhancement through iterative learning while providing flexibility with respect to task variations. The proposed approach involves an iterative optimization procedure after each task, that exploits recent developments in instrumental variable-based system identification. Enhanced performance compared to pre-existing results is proven theoretically and illustrated through simulation examples.

Key words: Iterative Learning Control, Basis functions, Varying reference, Instrumental variable

1 Introduction

ILC enables a significant performance enhancement of batch-repetitive processes. In ILC the command signal is iteratively updated from one experiment (trial) to the next. Typical ILC algorithms generate a control signal that exactly compensates for the trial-invariant exogenous disturbances during a specific task. ILC has been thoroughly researched, including convergence analysis [21,20], and robustness to model uncertainty [1,7] and disturbances [10,25]. In addition, many successful applications have been reported, including wafer scanners [19,24] and printing systems [6].

ILC can perfectly compensate for non-varying disturbances, but is consequently very sensitive to varying disturbances. These varying disturbances include measurement noise and also changing reference trajectories. As a result, a learned signal corresponds to a specific reference signal and a change in this signal potentially leads to performance deterioration [9,23,14,17]. To overcome this drawback, several solutions to enhance the extrapolation properties of ILC have been developed. In [17], the extrapolation properties are enhanced by constructing the task such that it consists of a set of basis tasks. This provides extrapolation to tasks consisting of a finite set of elementary tasks. A more general approach is to parameterize the command signal in a set of basis functions [23,22]. Such an approach allows for arbitrary tasks. Examples include polynomial basis functions [29,6,14,9] for which the associated optimization problem has an explicit analytic solution [12]. These polynomial approaches have clear advantages from an optimization perspective, since global optimality can be guaranteed and the implementation and computation is generally inexpensive and fast.

Recently, rational basis functions have been introduced in ILC in [5]. These rational basis functions are more general than polynomial basis functions since the latter are recovered as a special case. In the rational case, an analytic solution can be retained if the poles are pre-specified [16]. To enable enhanced performance the poles are also optimized in [5], where the non-convex optimization problem is solved using a similar algorithm as in [26]. In [5], fast convergence to a stationary point and increased performance is reported. In addition, the algorithm is reported to be less sensitive to local minima when compared to a Gauss-Newton type of algorithm as shown in, for example, [4]. However, in the present paper both a theoretical and numerical analysis are presented that reveal that the stationary point of the iteration is not necessarily a minimum of the objective function, which in fact has also been observed in related system identification algorithms [28].

Although important contributions have been made to enhance extrapolation capabilities of ILC through basis functions, presently available optimization algorithms suffer from the problem of non-optimality or poor convergence

Email addresses: j.c.d.v.zundert@tue.nl (Jurgen van Zundert), j.j.bolder@tue.nl (Joost Bolder), t.a.e.oomen@tue.nl (Tom Oomen).

properties. The aim of this paper is to develop a new approach that guarantees that the stationary point of the iterative solution is always an optimum. As a consequence, increased performance is achieved compared to pre-existing approaches. The proposed approach is related to instrumental variable system identification. Note that the instrumental variable approach in [3] is essentially different in that it deals with an estimation problem and not an ILC problem.

The contributions of this paper are threefold. First, a new iterative solution algorithm for rational basis functions in ILC is proposed which constitutes the main contribution of this paper. Second, non-optimality of the pre-existing approach for rational basis functions in ILC is established. Third, it is shown by two simulation examples that i) the proposed approach outperforms the pre-existing approach, and ii) ILC with basis functions outperforms standard ILC for varying reference tasks. Since the proposed approach has close connections to instrumental variable-based system identification, the simulation study may be of interest to instrumental variable based system identification.

In [5] a different iterative solution for rational basis functions in ILC is provided. In this paper it is theoretically proven and illustrated through simulation examples that this pre-existing approach is non-optimal by construction and is outperformed by the proposed approach. Related to the present paper, preliminary results can be found in [30]. This paper significantly extends this earlier research on several aspects. First of all, a more general basis parameterization is considered. Second, a proof for the optimal proposed approach is presented. Third, solutions of ILC with polynomial basis functions and standard norm-optimal ILC are recovered as special cases. Fourth, the non-optimality of the pre-existing approach is mathematically proven, motivating the use of the novel approach. Fifth, a numerical simulation example on convergence is presented to provide insight into both approaches. Finally, a simulation example is presented demonstrating the enhanced extrapolation properties with respect to norm-optimal ILC and ILC with polynomial basis functions.

The outline of this paper is as follows. In section 2, the problem considered in this paper is introduced. The proposed approach is presented in section 3. In section 4, the proposed approach is compared with the pre-existing approach [5]. Moreover, non-optimality of the pre-existing approach is established. The two iterative approaches are compared by use of a simulation example in section 5, demonstrating that the proposed approach outperforms the pre-existing approach on a complex industrial system. In section 6, a simulation example is presented revealing the benefit of using basis functions in ILC. Section 7 contains conclusions.

Notation In this paper, systems are discrete-time, linear, time-invariant (LTI), single-input, single-output (SISO). Systems are generally rational in complex indeterminate z and indicated in boldface with the argument z, for example $\mathbf{H}(z)$. Let $x(k)$ denote a signal x at time k. Let

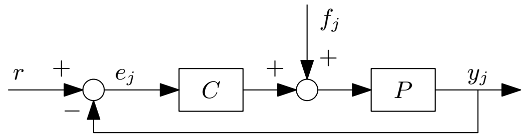

Fig. 1. Block diagram of closed-loop system under consideration.

$h(l)$ be the impulse response of the system $\mathbf{H}(z)$. The output $y(k)$ of the response of $\mathbf{H}(z)$ to input u is given by $y(k) = \sum_{l=-\infty}^{\infty} h(l)u(k-l)$. Let $N \in \mathbb{Z}^+$ denote the trial length, i.e. the number of samples per trial. Assuming $u(k) = 0$ for $k < 0$ and $k > N-1$, then the input-output relation can be recast as

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

with $u, y \in \mathbb{R}^N$ the input and output, respectively. Let $\|x\|_W := x^\top W x$, where $x \in \mathbb{R}^N$ and $W = W^\top \in \mathbb{R}^{N \times N}$. W is positive definite ($W \succ 0$) iff $x^\top W x > 0$, $\forall x \neq 0$ and positive semi-definite ($W \succeq 0$) iff $x^\top W x \geq 0$, $\forall x$.

To facilitate presentation, occasionally transfer functions are assumed causal to enable a direct relation between infinite and finite time. This is standard in ILC [21] and not a restriction on the presented results. For instance, the approach in [3, Appendix A] may be adopted.

2 Problem formulation

In this section the considered problem is defined by describing the system, introducing norm-optimal ILC, and highlighting the limitations of standard norm-optimal ILC. Finally, the contributions are listed explicitly.

2.1 System description

The control setup is shown in Fig. 1. Here $P = \frac{B_0}{A_0}$, $B_0, A_0 \in \mathbb{R}[z]$, is the rational system and C an internally stabilizing feedback controller. The closed-loop system is assumed to operate batch-repetitive, i.e. the same process of fixed length N is repeated over and over. A single execution is referred to as a trial. The aim is to determine the feedforward f_{j+1} for trial $j+1$ such that the output y_{j+1} follows the trial-invariant reference r, i.e. minimizes the error $e_{j+1} = r - y_{j+1}$.

The error for trial j is given by

$$e_j = Sr - SPf_j \tag{1}$$
$$= \tilde{r} - Jf_j, \tag{2}$$

with sensitivity $S := (I + PC)^{-1}$, process sensitivity $J := SP$, and $\tilde{r} := Sr$. The error for trial $j+1$ is given by

$$e_{j+1} = \tilde{r} - Jf_{j+1}. \tag{3}$$

Eliminating $\tilde{r}$ from (3) by using (2) yields the trial-to-trial dynamics

$$e_{j+1} = e_j + J\left(f_j - f_{j+1}\right), \tag{4}$$

which are optimized in norm-optimal ILC.

2.2 Norm-optimal ILC

Norm-optimal ILC is an important class of ILC in which the feedforward signal f_{j+1} for the next trial is determined by minimizing a performance criterion as in Definition 1.

Definition 1 *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 $W_e, W_f, W_{\Delta f} \succeq 0$ *and* e_{j+1} *given by* (4).

Since $\mathcal{J}(f_{j+1})$ is quadratic in f_{j+1}, the optimal feedforward signal f^*_{j+1} can be computed analytically [12] from

$$\left.\frac{d\mathcal{J}(f_{j+1})}{df_{j+1}}\right|_{f_{j+1}=f^*_{j+1}} = 0 \tag{6}$$

and is provided by Theorem 2.

Theorem 2 *Given* $J^\top W_e J + W_f + W_{\Delta f} \succ 0$*, model* J*, and measurement data* r, f_j, e_j*, optimal* f^*_{j+1} *for norm-optimal ILC with the performance criterion of Definition 1 is*

$$\begin{aligned} f^*_{j+1} &= Qf_j + Le_j, \\ Q &= \left(J^\top W_e J + W_f + W_{\Delta f}\right)^{-1}\left(J^\top W_e J + W_{\Delta f}\right), \\ L &= \left(J^\top W_e J + W_f + W_{\Delta f}\right)^{-1} J^\top W_e. \end{aligned} \tag{7}$$

Proof. Substitute (4) in (5) and solve (6) for f^*_{j+1}. □

With norm-optimal ILC excellent performance is achieved for exactly repeating tasks. Indeed, if P is invertible, $W_e \succ 0$, and $W_f = W_{\Delta f} = 0$, convergent ILC with a perfect model results in $f^*_{j+1} = P^{-1}r$ and hence $e_{j+1} = Sr - SPf^*_{j+1} = 0$, $\forall r$. However, this result does not hold when r varies as shown next. Let the reference signal at trail j be denoted by r_j. Then, under the same conditions as before, $f^*_{j+1} = P^{-1}r_j$ and hence $e_{j+1} = Sr_{j+1} - SPf^*_{j+1} = S\left(r_{j+1} - r_j\right)$. Consequently, $e_{j+1} \neq 0$ if $r_{j+1} \neq r_j$, i.e. for a trial-varying reference signal.

Ideally $f^*_{j+1} = P^{-1}r_{j+1}$, which corresponds to inverse model feedforward and shows that the optimal feedforward signal is a function of the applied reference signal. A key observation for norm-optimal ILC is that only information of previous trials is exploited. Hence, the learned signal will only be optimal for one specific constant reference signal and non-optimal for varying reference signals, i.e. extrapolation properties are poor. To enhance extrapolation properties, basis functions are exploited in this paper.

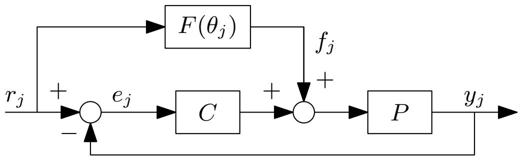

Fig. 2. Implementation of basis functions in $f_j = F(\theta_j)r_j$.

2.3 Problem formulation

Inspired by inverse model feedforward, extrapolation properties are introduced in ILC by use of basis functions as

$$f_j = F(\theta_j)r_j, \tag{8}$$

where $F(\theta_j) \in \mathbb{R}^{N\times N}$ denotes the matrix notation of the feedforward filter parameterized in parameters $\theta_j \in \mathbb{R}^m$. Similar parameterizations are used in, for example, [13,29]. The implementation in the control scheme of Fig. 1 is depicted in Fig. 2. Note that f_{j+1} is a function of r_{j+1}, which is in sharp contrast to standard norm-optimal ILC where it is implicitly only a function of r_j. Substitution of (8) in (1) yields $e_j = Sr_j - SPF(\theta_j)r_j = S(I - PF(\theta_j))r_j = 0$, $\forall r_j$ if $F(\theta_j) = P^{-1}$. Hence, by proper selection of $F(\theta_j)$, and learning θ_j, zero error may be achieved for arbitrary reference signals.

The basic idea is that if instead of learning f_j, the ILC algorithm optimizes θ_j, then e_j is invariant under r if $F(\theta_j) = P^{-1} = \frac{A_0}{B_0}$. Note that a standard model-based feedforward is recovered from (8) if θ_j is pre-specified. In this paper rational basis functions are used for parameterizing the feedforward filter $F(\theta_j)$, see Definition 3, which enables optimization of both zeros and poles.

Definition 3 *Rational basis functions in the parameters* $\theta_j \in \mathbb{R}^m$ *with reference* r_j *as basis are defined as in* (8) *with* $F(\theta_j)$ *the matrix representation of* $\mathbf{F}(\theta_j, z) \in \mathcal{F}$,

$$\begin{aligned} \mathcal{F} &= \left\{ \left.\frac{\mathbf{A}(\theta_j, z)}{\mathbf{B}(\theta_j, z)} \right| \theta_j \in \mathbb{R}^m \right\}, \\ \mathbf{A}(\theta_j, z) &= \boldsymbol{\xi}^{\mathbf{A}}_{\mathbf{0}}(z) + \sum_{n=1}^{m} \boldsymbol{\xi}^{\mathbf{A}}_{\mathbf{n}}(z)\theta_j[n], \\ \mathbf{B}(\theta_j, z) &= \boldsymbol{\xi}^{\mathbf{B}}_{\mathbf{0}}(z) + \sum_{n=1}^{m} \boldsymbol{\xi}^{\mathbf{B}}_{\mathbf{n}}(z)\theta_j[n], \end{aligned}$$

where $\boldsymbol{\xi}^{\mathbf{A}}_{\mathbf{n}}(z), \boldsymbol{\xi}^{\mathbf{B}}_{\mathbf{n}}(z) \in \mathbb{R}[z]$, $n = 0, 1, \ldots, m$ *are polynomial in* z.

Note that for $\mathbf{B}(\theta_j, z) = \boldsymbol{\xi}^{\mathbf{B}}_{\mathbf{0}}$ the pole locations of $\mathbf{F}(\theta_j, z)$ are pre-specified. Polynomial basis functions are the special case of rational basis functions with $\mathbf{B}(\theta_j, z) = 1$.

Substitution of (8) in (5) yields Definition 4, which reveals that $\mathcal{J}(f_{j+1})$ is a function of θ_{j+1} by the fixed structure of (8). Instead of determining f^*_{j+1}, θ^*_{j+1} is to be determined.

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

$$\begin{aligned}\mathcal{J}(\theta_{j+1}) := &\|e_{j+1}(\theta_{j+1})\|_{W_e} + \|f_{j+1}(\theta_{j+1})\|_{W_f} \\ &+ \|f_{j+1}(\theta_{j+1}) - f_j\|_{W_{\Delta f}},\end{aligned} \tag{9}$$

with $W_e, W_f, W_{\Delta f} \succeq 0$, and using Definition 3 and (4)*,*

$$\begin{aligned} f_{j+1} &= B^{-1}(\theta_{j+1})A(\theta_{j+1})r_j, \\ e_{j+1} &= e_j + Jf_j - JB^{-1}(\theta_{j+1})A(\theta_{j+1})r_j. \end{aligned} \tag{10}$$

The goal in this paper is to solve the following problem.

Problem 5 *Given Definition 3, a model of J, parameters θ_j, and measurement data r_j, f_j, e_j, determine*

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

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

The contributions of this paper are as follows.

I An iterative approach to solve Problem 5 is proposed and its optimality is shown.
II By analysis of the approach it is shown that the solution in [5] solves Problem 5 non-optimally.
III The results of I and II are confirmed by use of simulation examples. In particular, it is validated that the proposed approach outperforms the pre-existing approach, and the benefits of ILC with basis functions in terms of extrapolation properties are demonstrated.

3 Proposed approach

In this section, Problem 5 is analyzed and the proposed approach is introduced, forming contribution I. Finally, polynomial basis functions and standard norm-optimal ILC are recovered as special cases.

3.1 Analysis

In standard norm-optimal ILC (subsection 2.2) the optimal feedforward is found through (6). Similarly, θ^*_{j+1} satisfies

$$\left.\frac{d\mathcal{J}(\theta_{j+1})}{d\theta_{j+1}}\right|_{\theta_{j+1}=\theta^*_{j+1}} = 0, \tag{11}$$

with the gradient given by Lemma 6.

Lemma 6 *Given Definition 3, the gradient of $\mathcal{J}(\theta_{j+1})$ with respect to θ_{j+1} is given by*

$$\begin{aligned}\left(\frac{d\mathcal{J}(\theta_{j+1})}{d\theta_{j+1}}\right)^\top = 2&\left(\frac{df_{j+1}}{d\theta_{j+1}}\right)^\top \times \\ &\left[\left(-J^\top W_e J - W_{\Delta f}\right) f_j + J^\top W_e e_j\right. \\ &\left.+ \left(J^\top W_e J + W_f + W_{\Delta f}\right) B^{-1}(\theta_{j+1})A(\theta_{j+1})r_j\right].\end{aligned} \tag{12}$$

Proof. Follows from substituting (4) in (9) and using that for $x, b \in \mathbb{R}^N$, $A \in \mathbb{R}^{N\times N}$, and $W = W^\top \in \mathbb{R}^{N\times N}$,

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

The stated result is found by substitution of (10). □

Lemma 6 reveals that $\frac{d\mathcal{J}(\theta_{j+1})}{d\theta_{j+1}}$ is nonlinear in θ_{j+1} because of the terms $\frac{df_{j+1}}{d\theta_{j+1}}$ and $B^{-1}(\theta_{j+1})$. Consequently, there is no general analytic solution available and there may exist multiple optima. To solve this optimization problem, an iterative solution is proposed in the following subsection.

3.2 Optimal solution

In this subsection the proposed approach to solve Problem 5 is introduced which forms the main contribution of this paper. The idea is to iteratively solve a sequence of convex optimization problems which are also a solution of the non-convex optimization problem if the iterative procedure converges. A weighted version of (11) is considered which is affine in the parameters and enables an analytic solution. Upon convergence, the gradient in Lemma 6 is recovered and a solution to Problem 5 is obtained.

First, an auxiliary iteration index $^{\langle q \rangle}$ is introduced and a weighting is applied to (12) as given by Definition 7.

Definition 7 *The weighted gradient of the performance criterion is defined as*

$$\begin{aligned}\left(\overline{\frac{d\mathcal{J}(\theta^{\langle q\rangle}_{j+1})}{\theta^{\langle q\rangle}_{j+1}}}\right)^\top &= 2\zeta^{\langle q\rangle} \times \\ &\Big[-\left(J^\top W_e J + W_{\Delta f}\right) B(\theta^{\langle q\rangle}_{j+1}) f_j - J^\top W_e B(\theta^{\langle q\rangle}_{j+1}) e_j \\ &+ \left(J^\top W_e J + W_f + W_{\Delta f}\right) A(\theta^{\langle q\rangle}_{j+1}) r_j\Big] \in \mathbb{R}^m, \\ \textit{with}\quad \zeta^{\langle q\rangle} &= \left(\overline{\frac{df^{\langle q-1\rangle}_{j+1}}{d\theta^{\langle q-1\rangle}_{j+1}}}\right)^\top B^{-1}(\theta^{\langle q-1\rangle}_{j+1}) \in \mathbb{R}^{m\times N}.\end{aligned} \tag{13}$$

Note that the expression in Lemma 6 is recovered from Definition 7 upon convergence, i.e. $\theta^{\langle q\rangle}_{j+1} = \theta^{\langle q-1\rangle}_{j+1} = \theta_{j+1}$.

Second, it is observed that (13) is affine in $A(\theta_{j+1}^{\langle q\rangle})$ and $B(\theta_{j+1}^{\langle q\rangle})$. Since both $A(\theta_{j+1}^{\langle q\rangle})$ and $B(\theta_{j+1}^{\langle q\rangle})$ are affine in $\theta_{j+1}^{\langle q\rangle}$ (see Definition 3), (13) is affine in $\theta_{j+1}^{\langle q\rangle}$. This is exploited in Theorem 8.

Theorem 8 *Given $r_j, f_j, e_j, \theta_{j+1}^{\langle q-1\rangle}$, the solution to $\overline{\frac{d\mathcal{J}(\theta_{j+1}^{\langle q\rangle})}{\theta_{j+1}^{\langle q\rangle}}} = 0$ is given by*

$$
\begin{aligned}
\theta_{j+1}^{\langle q\rangle *} &= -\left(\zeta^{\langle q\rangle} Q\right)^{-1} \zeta^{\langle q\rangle} R, \qquad (14)\\
Q &= \left(J^\top W_e J + W_f + W_{\Delta f}\right) \Psi_{r_j}^A \\
&\quad - J^\top W_e \Psi_{e_j}^B - \left(J^\top W_e J + W_{\Delta f}\right) \Psi_{f_j}^B,\\
R &= \left(J^\top W_e J + W_f + W_{\Delta f}\right) \xi_0^A r_j \\
&\quad - J^\top W_e \xi_0^B e_j - \left(J^\top W_e J + W_{\Delta f}\right) \xi_0^B f_j.
\end{aligned}
$$

Proof. Using the notation defined in section 1:

$$
\begin{aligned}
A(\theta_{j+1}^{\langle q\rangle}) r_j &= \xi_0^A r_j + \Psi_{r_j}^A \theta_{j+1}^{\langle q\rangle},\\
B(\theta_{j+1}^{\langle q\rangle}) e_j &= \xi_0^B e_j + \Psi_{e_j}^B \theta_{j+1}^{\langle q\rangle},\\
B(\theta_{j+1}^{\langle q\rangle}) f_j &= \xi_0^B f_j + \Psi_{f_j}^B \theta_{j+1}^{\langle q\rangle}.
\end{aligned}
$$

Substituting these expressions into (13), equating to zero, and solving for $\theta_{j+1}^{\langle q\rangle} = \theta_{j+1}^{\langle q\rangle *}$ completes the proof. □

Note that (14) only depends on data of trial j (r_j, f_j, e_j) and the previous parameter estimate $\theta_{j+1}^{\langle q-1\rangle}$.

In the previous steps two key elements are derived which are briefly summarized as follows. First, a weighted gradient is introduced in Definition 7, from which the gradient in Lemma 6 is recovered for $\theta_{j+1}^{\langle q\rangle} = \theta_{j+1}^{\langle q-1\rangle} = \theta_{j+1}$. Second, an analytic solution for $\theta_{j+1}^{\langle q\rangle *} = \theta_{j+1}^{\langle q\rangle}$ is obtained for which the weighted gradient in Definition 7 is zero. Hence, upon convergence, the actual gradient also converges to zero and optimal performance is achieved. Combining these two elements, there is an optimal 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 1.

Algorithm 1 *The proposed algorithm for solving Problem 5 is given by the following sequence of steps.*

(1) Given r_j, f_j, e_j, θ_j, set $q = 1$, initialize $\theta_{j+1}^{\langle q-1\rangle} = \theta_j$.
(2) Compute $\theta_{j+1}^{\langle q\rangle *}$ in (14).
(3) Set $q \to q+1$ and go back to (2) until an appropriate stopping criterion is satisfied.

Note that the iteration in $^{\langle q\rangle}$ is performed off-line and hence does not require additional experiments in the usual ILC sense.

The convergence of similar type of algorithms is experienced to be good in well-established related algorithms in instrumental variable system identification [11,2], yet at present global convergence has only been proved under certain assumptions [27].

3.3 Recovering pre-existing results as special cases

For polynomial basis functions the optimization problem is convex and convergence is achieved in a single step. Indeed, $\zeta^{\langle q\rangle} = \left(\overline{\frac{dA(\theta_{j+1}^{\langle q-1\rangle}) r_j}{d\theta_{j+1}^{\langle q-1\rangle}}}\right)^\top = \left(\Psi_{r_j}^A\right)^\top$ in Algorithm 1 is constant for all $^{\langle q\rangle}$ and hence the procedure yields the single-step solution provided by Corollary 9, which recovers the well-known results in [29] and [5]. Moreover, for standard norm-optimal ILC, i.e., $f_j = \theta_j$, Theorem 2 is recovered from Theorem 8, see Corollary 9.

Corollary 9 *For polynomial basis functions, the solution in* (14) *reduces to the analytic solution*

$$
\begin{aligned}
\theta_{j+1}^* = &\left(\left(\Psi_{r_j}^A\right)^\top \left(J^\top W_e J + W_f + W_{\Delta f}\right) \Psi_{r_j}^A\right)^{-1} \times \\
&\left(\Psi_{r_j}^A\right)^\top \left(-\left(J^\top W_e J + W_f + W_{\Delta f}\right) \xi_0^A r_j \right.\\
&\left. + J^\top W_e e_j + \left(J^\top W_e J + W_{\Delta f}\right) f_j\right).
\end{aligned}
$$

If in addition $\boldsymbol{\xi}_\mathbf{0}^\mathbf{A}(z) = 0$, then

$$
\begin{aligned}
\theta_{j+1}^* &= Q f_j + L e_j, \qquad (15)\\
Q &= \left(\left(\Psi_{r_j}^A\right)^\top \left(J^\top W_e J + W_f + W_{\Delta f}\right) \Psi_{r_j}^A\right)^{-1} \left(\Psi_{r_j}^A\right)^\top \left(J^\top W_e J + W_{\Delta f}\right)\\
L &= \left(\left(\Psi_{r_j}^A\right)^\top \left(J^\top W_e J + W_f + W_{\Delta f}\right) \Psi_{r_j}^A\right)^{-1} \left(\Psi_{r_j}^A\right)^\top J^\top W_e.
\end{aligned}
$$

If $f_j = \theta_j$, then Theorem 2 is recovered.

Proof. Substitute $\zeta^{\langle q\rangle} = \left(\Psi_{r_j}^A\right)^\top$, $\xi_0^B = I$, and $\Psi_{e_j}^B = \Psi_{f_j}^B = 0$ in (14). If in addition $\boldsymbol{\xi}_\mathbf{0}^\mathbf{A}(z) = 0$, then $\xi_0^A = 0$ and $f_j = \Psi_{r_j}^A \theta_j$. For $f_j = \theta_j$, $\Psi_{r_j}^A = I$ which substituted in (15) yields (7). □

4 Non-optimality of pre-existing approach

In [5] an alternative solution to Problem 5 is proposed. In this section it is demonstrated that this iterative procedure generally converges to a non-optimal stationary point. As a result, the proposed approach potentially yields better performance. This section forms contribution II.

4.1 Pre-existing approach

In the proposed approach the gradient of the performance criterion is weighted. In contrast, in the pre-existing approach of [5] the performance criterion is weighted, see Definition 10.

Definition 10 *The weighted performance criterion is defined as*

$$\hat{\mathcal{J}}(\theta_{j+1}^{\langle q\rangle}) := \left\| B^{-1}(\theta_{j+1}^{\langle q-1\rangle}) B(\theta_{j+1}^{\langle q\rangle}) e_{j+1}^{\langle q\rangle} \right\|_{W_e} + \left\| B^{-1}(\theta_{j+1}^{\langle q-1\rangle}) B(\theta_{j+1}^{\langle q\rangle}) f_{j+1}^{\langle q\rangle} \right\|_{W_f} + \left\| B^{-1}(\theta_{j+1}^{\langle q-1\rangle}) B(\theta_{j+1}^{\langle q\rangle}) f_{j+1}^{\langle q\rangle} - f_j \right\|_{W_{\Delta f}}. \tag{16}$$

Note that if $\theta_{j+1}^{\langle q\rangle} = \theta_{j+1}^{\langle q-1\rangle} = \theta_{j+1}$, then $\hat{\mathcal{J}}(\theta_{j+1}^{\langle q\rangle}) = \mathcal{J}(\theta_{j+1})$, i.e., the unweighted performance criterion is recovered.

The weighted signals are affine in $\theta_{j+1}^{\langle q\rangle}$ since the term $B^{-1}(\theta_{j+1}^{\langle q\rangle})$ is canceled. As a result, $\hat{\mathcal{J}}(\theta_{j+1}^{\langle q\rangle})$ is quadratic in $\theta_{j+1}^{\langle q\rangle}$ and there is a unique solution for $\theta_{j+1}^{\langle q\rangle *}$, which can be determined analytically from

$$\left(\frac{d\hat{\mathcal{J}}(\theta_{j+1}^{\langle q\rangle})}{d\theta_{j+1}^{\langle q\rangle}} \right)^\top \Bigg|_{\theta_{j+1}^{\langle q\rangle} = \theta_{j+1}^{\langle q\rangle *}} = 0. \tag{17}$$

The idea is to iteratively determine $\theta_{j+1}^{\langle q\rangle *}$ for $\hat{\mathcal{J}}(\theta_{j+1}^{\langle q\rangle})$ in Definition 10 using (17). The reasoning is that upon convergence of the parameters, i.e., $\theta_{j+1}^{\langle q\rangle} \to \theta_{j+1}^{\langle q-1\rangle}$, $\theta_{j+1}^{\langle q\rangle *}$ are also the optimal parameters for Problem 5, because $\mathcal{J}(\theta_{j+1})$ is recovered from $\hat{\mathcal{J}}(\theta_{j+1}^{\langle q\rangle})$ for $\theta_{j+1}^{\langle q\rangle} = \theta_{j+1}^{\langle q-1\rangle} = \theta_{j+1}$. However, in the next section it demonstrated that this reasoning is incorrect, i.e., the stationary point of the iteration is not necessarily a minimum of $\mathcal{J}(\theta_{j+1})$.

4.2 Non-optimality

In the approach outlined in section 4.1, (17) is solved which yields the minimum of $\hat{\mathcal{J}}(\theta_{j+1}^{\langle q\rangle})$. However, this is not necessarily a minimum of $\mathcal{J}(\theta_{j+1})$. As a consequence, the parameters do not necessarily provide the solution to Problem 5. The non-optimality of this approach is highlighted in Theorem 11.

Theorem 11 *For the purpose of exposition let $W_f = W_{\Delta f} = 0$ and assume that the pre-existing iterative procedure described in subsection 4.1 converges to a stationary point, and let $\theta_{j+1}^{pre} = \lim_{q\to\infty} \theta_{j+1}^{\langle q\rangle *}$ with $\theta_{j+1}^{\langle q\rangle *}$ the solution to* (17). *Then non-optimal performance is achieved if*

$$\left(B^{-1}(\theta_{j+1}^{pre}) \frac{dB(\theta_{j+1}^{pre})}{d\theta_{j+1}^{pre}} e_{j+1}^{pre} \right)^\top W_e e_{j+1}^{pre} \neq 0.$$

Proof. Substitution of $W_f = W_{\Delta f} = 0$ in (12) and using $\frac{df_{j+1}}{d\theta_{j+1}} = \frac{dF_{j+1}}{d\theta_{j+1}} r_j$ it follows

$$\left(\frac{d\mathcal{J}(\theta_{j+1})}{d\theta_{j+1}} \right)^\top = -2 \left(J \frac{dF(\theta_{j+1})}{d\theta_{j+1}} r_j \right)^\top W_e e_{j+1}, \tag{18}$$

with e_{j+1} given by (4). The gradient of (16) for $W_f = W_{\Delta f} = 0$ is given by

$$\left(\frac{d\hat{\mathcal{J}}(\theta_{j+1}^{\langle q\rangle})}{d\theta_{j+1}^{\langle q\rangle}} \right)^\top = 2 \left(B^{-1}(\theta_{j+1}^{\langle q-1\rangle}) \frac{dB(\theta_{j+1}^{\langle q\rangle}) e_{j+1}^{\langle q\rangle}}{d\theta_{j+1}} \right)^\top \times W_e B^{-1}(\theta_{j+1}^{\langle q-1\rangle}) B(\theta_{j+1}^{\langle q\rangle}) e_{j+1}^{\langle q\rangle},$$

$$\frac{dB(\theta_{j+1}^{\langle q\rangle}) e_{j+1}^{\langle q\rangle}}{d\theta_{j+1}} = \frac{dB(\theta_{j+1}^{\langle q\rangle})}{d\theta_{j+1}^{\langle q\rangle}} e_{j+1}^{\langle q\rangle} - B(\theta_{j+1}^{\langle q\rangle}) J \frac{dF(\theta_{j+1}^{\langle q\rangle})}{d\theta_{j+1}^{\langle q\rangle}} r_j.$$

Evaluating this gradient after convergence (i.e., $\theta_{j+1}^{\langle q-1\rangle} = \theta_{j+1}^{\langle q\rangle} = \theta_{j+1}^{pre}$) yields

$$2 \left(B^{-1}(\theta_{j+1}^{pre}) \frac{dB(\theta_{j+1}^{pre})}{d\theta_{j+1}^{pre}} e_{j+1}^{pre} - J \frac{dF(\theta_{j+1}^{pre})}{d\theta_{j+1}^{pre}} r_j \right)^\top W_e e_{j+1}^{pre}$$

which is zero by definition of θ_{j+1}^{pre}. Hence, if

$$\left(B^{-1}(\theta_{j+1}^{pre}) \frac{dB(\theta_{j+1})}{d\theta_{j+1}^{pre}} e_{j+1}^{pre} \right)^\top W_e e_{j+1}^{pre} \neq 0,$$

then by (18), $\frac{d\mathcal{J}(\theta_{j+1})}{d\theta_{j+1}} \Big|_{\theta_{j+1} = \theta_{j+1}^{pre}} \neq 0$, indicating non-optimality which concludes the proof. □

Theorem 11 implies that the pre-existing approach solves Problem 5 if $e_{j+1} = 0$. This requires there exists θ_{j+1} such that $F(\theta_{j+1}) = P^{-1}$, see section 2.2. This is, however, generally not the case due to unmodeled dynamics and therefore non-optimal performance is achieved. Moreover, the non-optimality of the pre-existing approach may be more severe under stochastic disturbances, see for example [28].

The key difference between the two approaches is the level at which the weight is applied. In the pre-existing approach this is at the level of the performance criterion, whereas with the proposed approach this is at the level of the gradient of the performance criterion. Since the parameters are determined at the level of the gradient, the proposed approach is optimal, whereas this is generally not the case for the pre-existing approach. The non-optimality of the pre-existing approach is illustrated in a simulation example in section 5.

5 Example: convergence

In this section a simulation example is provided to demonstrate the optimality of the proposed approach (see section 3) and the non-optimality of the pre-existing approach (see section 4). In addition, the convergence behavior of both approaches is analyzed and compared to the Gauss-Newton algorithm. This section forms the first part of contribution III.

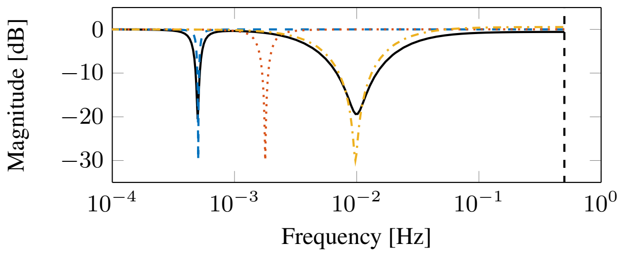

Fig. 3. The inverse system P^{-1} (——) and feedforward filters $F(\theta = 10^{-5})$ (- - -), $F(\theta = 10^{-3})$ (······), and $F(\theta = 10^{-1})$ (- · -). By design, $F(\theta)$ is only able to (partially) compensate one of the resonances.

5.1 Setup

The parameter update over a single trial is considered and therefore the subscript j is omitted throughout this section. An open-loop system is considered (i.e., $C = 0$) with the system defined as

$$P = \frac{(1 + 2\beta_{1,1} + \omega_1)\, z^2 - 2(1 - \beta_{1,1})z + 1}{(1 + 2\beta_{1,2} + \omega_1)\, z^2 - 2(1 - \beta_{1,2})z + 1} \times \frac{(1 + 2\beta_{2,1} + \omega_2)\, z^2 - 2(1 - \beta_{2,1})z + 1}{(1 + 2\beta_{2,2} + \omega_2)\, z^2 - 2(1 - \beta_{2,2})z + 1},$$

where $\beta_{1,1} = 0.12$, $\beta_{1,2} = 0.01$, $\omega_1 = 0.0005 \cdot 2\pi$, $\beta_{2,1} = 1.2$, $\beta_{2,2} = 0.1$, and $\omega_2 = 0.01 \cdot 2\pi$.

The reference signal is defined as

$$r(k) = 40 \sin(\omega_1 k) + \sin(\omega_2 k), \quad k = 0, 1, \dots N - 1,$$

with trial length $N = 1000$. The weighting filters in Definition 4 are set to $W_e = 10^{-4} I$ and $W_f = W_{\Delta f} = 0$ in order to only weigh the error.

The feedforward filter is parameterized as

$$F(\theta) = \frac{(z - 1)^2 + (2\beta_1(z - 1) + \omega_n)\, \theta}{(z - 1)^2 + (2\beta_2(z - 1) + \omega_n)\, \theta},$$

with $\beta_1 = 0.001$, $\beta_2 = 2$, $\omega_n = 0.02 \cdot 2\pi$. Note that $F(\theta)$ can be written in the form of Definition 3 with $m = 1$ and

$$\begin{aligned} \boldsymbol{\xi}_0^{\mathbf{A}}(z) &= (z - 1)^2, & \boldsymbol{\xi}_1^{\mathbf{A}}(z) &= 2\beta_1(z - 1) + \omega_n, \\ \boldsymbol{\xi}_0^{\mathbf{B}}(z) &= (z - 1)^2, & \boldsymbol{\xi}_1^{\mathbf{B}}(z) &= 2\beta_2(z - 1) + \omega_n. \end{aligned}$$

The Bode plot of $F(\theta)$ is depicted in Fig. 3 for various values θ, together with the system inverse. The figure shows that there exist θ such that $F(\theta)$ resembles (part of) the system inverse. However, by design, $F(\theta)$ is only able to (partially) compensate one of the two system resonances. Assuming minor influence of transient behavior, it is to be expected that there are two optima: compensation of either the resonance at 0.0005 Hz or the resonance at 0.01 Hz.

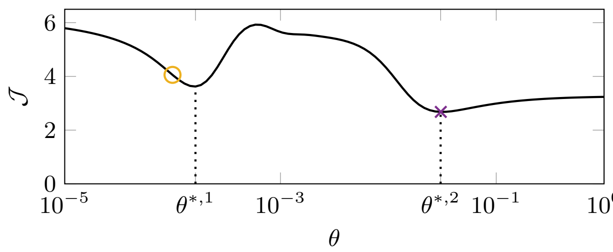

Fig. 4. $\mathcal{J}$ for a grid in θ and the corresponding optimal parameters. For $\theta^{\langle 0 \rangle} = 10^{-3}$, the pre-existing approach (○) converges to the point near the local minimum, whereas the proposed approach (×) converges to the global minimum.

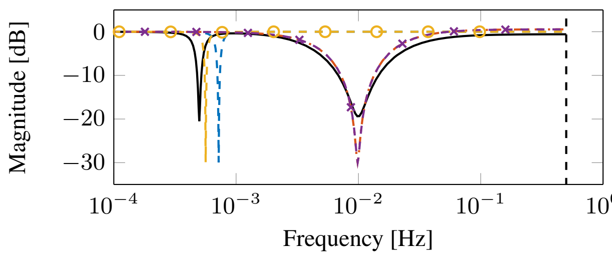

Fig. 5. $F(\theta)$ for $\theta^{*,1}$ (- - -), $\theta^{*,2}$ (- · -), and pre-existing (· ⊖ ·) and proposed (· ⨯ ·) approach (both after ten iterations) partially compensate different resonances of P^{-1} (——).

	θ^*	$\mathcal{J}$	$\left\lvert \frac{d\mathcal{J}}{d\theta} \right\rvert$
$\theta^{*,1}$	1.5950×10^{-4}	3.6226	−0.5151
$\theta^{*,2}$	0.0315	2.6776	0.0001
pre-existing	9.9856×10^{-5}	4.0573	14556
proposed	0.0307	2.6781	1.3158

Table 1
Convergence properties of the pre-existing and proposed approach after ten iterations.

5.2 Results

The results are shown in Fig. 4, Fig. 5, Table 1, and Table 2.

Performance criterion analysis
In Fig. 4, $\mathcal{J}(\theta)$ is depicted for a grid of values for θ. As expected, there are two minima: $\theta^{*,1} = 1.5959 \times 10^{-4}$ and $\theta^{*,2} = 0.0315$, of which the feedforward filters are depicted in Fig. 5. Visual inspection reveals that for $\theta = \theta^{*,1}$ the first resonance of the system is (partially) compensated, whereas for $\theta = \theta^{*,2}$ the second system resonance is (partially) compensated. Note that the difference is resonance frequency between $F(\theta^{*,1})$ and the first resonance frequency of P appears to be large due to the logarithmic scale, but is small and in the same order as the difference between the resonance frequency of $F(\theta^{*,2})$ and the second resonance frequency of P.

Non-optimality pre-existing approach
First, the pre-existing approach is discussed. In Fig. 4 the stationary point of the pre-existing approach $\theta = 9.7249 \times 10^{-5}$ for an initial value $\theta^{\langle 0 \rangle} = 10^{-3}$ is indicated after ten iter-

initial value $\theta^{\langle 0\rangle}$	Gauss-Newton	pre-existing	pro-posed
$10^{-5} < \theta^{\langle 0\rangle} \leq 6\times 10^{-4}$	$\theta^{*,1}$	near $\theta^{*,1}$	$\theta^{*,2}$
$6\times 10^{-4} < \theta^{\langle 0\rangle} \leq 10^{0}$	$\theta^{*,2}$	near $\theta^{*,1}$	$\theta^{*,2}$

Table 2
Converged parameters. For the given interval, the stationary point of the pre-existing approach is always near $\theta^{*,1}$, whereas the stationary point of Gauss-Newton depends on the initial parameter.

ations. The stationary point is located near local minimum $\theta^{*,1}$. Clearly, this stationary point yields non-optimal performance since there exist θ for which $\mathcal{J}(\theta)$ is lower. Indeed, since $\tilde{\mathcal{J}}$ instead of $\mathcal{J}$ is minimized, non-optimal performance is obtained in terms of $\mathcal{J}$, see also Table 1. The results confirm Theorem 11 since the stationary point of the pre-existing approach is not a local minimum of $\mathcal{J}$.

Optimality proposed approach
In contrast to the pre-existing approach, the gradient of $\mathcal{J}$ converges to zero for the proposed approach. Indeed, Table 1 shows a small value $\frac{d\mathcal{J}}{d\theta}$ for the proposed approach. As a result, optimal performance is achieved and the value of $\mathcal{J}$ is the lowest for the proposed approach.

Convergence behavior
In Table 2 the stationary point of the approaches is indicated for a range of $\theta^{\langle 0\rangle}$. From the table it is observed that the pre-existing approach always converges to the same stationary point located closely to the local minimum, independent of the initial parameter. In contrast, for this case the stationary point of the proposed approach is always the global optimum. Hence, the proposed approach outperforms both the pre-existing and Gauss-Newton approach.

6 Example: performance

In this section the pre-existing and proposed approach for ILC with rational basis functions are compared with standard ILC and ILC with polynomial basis functions, for varying reference signals using a model of a complex industrial printer. The simulation demonstrates the excellent extrapolation properties of ILC with basis functions compared to standard ILC. Moreover, the simulation highlights the difference in performance for the three variants of ILC with basis functions. This section forms the second part of contribution III.

6.1 System description

The system is a model of the carriage position of an Océ Arizona 550 GT flatbed printer of which the Bode magnitude plot is depicted in Fig. 6. The system operates in closed-loop with a bandwidth of 25 Hz.

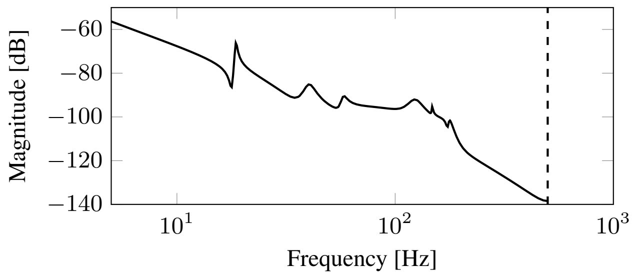

Fig. 6. Bode magnitude plot of the printer model.

6.2 Simulation setup

Nine trials are considered with a trial-varying reference signal according to Fig. 7. All trials have a length of $N = 4000$ samples.

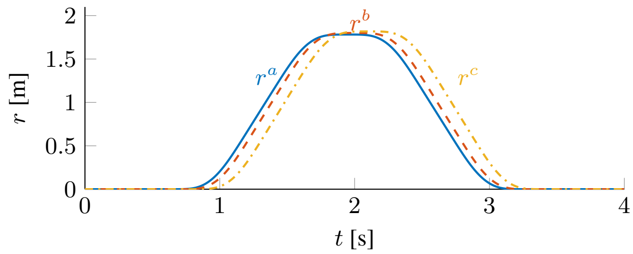

Fig. 7. A trial-varying reference signal is considered. Reference signal r^a (——) is active during trial $j = 0, 1, 2$; r^b (- - -) during $j = 3, 4, 5$; and r^c (- · -) during $j = 6, 7, 8$.

Rational basis functions with $\theta_j \in \mathbb{R}^4$ are defined according to Definition 3 and [18] as follows

$$\mathbf{A}(\theta_j, z) = \left(1 + \left(\tfrac{z-1}{zT_s}\right)\theta_j[1] + \left(\tfrac{z-1}{zT_s}\right)^2\theta_j[2]\right)\left(\tfrac{z-1}{zT_s}\right)^2,$$
$$\mathbf{B}(\theta_j, z) = \left(1 + \left(\tfrac{z-1}{zT_s}\right)\theta_j[3] + \left(\tfrac{z-1}{zT_s}\right)^2\theta_j[4]\right),$$

where the choice of basis functions is in part based on [18]. Recall that polynomial basis functions are the special case of rational basis functions with $\mathbf{B}(\theta_j, z) = 1$, i.e., $\theta_j[3] = \theta_j[4] = 0$, $\forall j$. The initial parameters θ_0 are set to zero. To simulate noise, white noise with a variance of 10^{-4} μm^2 is injected on the error signal. The weighting filters in Definition 4 are set to $W_e = I$, $W_f = W_{\Delta f} = 0$ in order to only penalize the error.

For polynomial basis functions $\mathbf{B}(\theta_j, z) = 1$, i.e., $\theta_j[3] = \theta_j[4] = 0$, $\forall j$. The error is contaminated by normally distributed zero mean white noise with variance 10^{-4} μm^2 and $\theta_0 = 0$. The weighting filters in Definition 4 are set to $W_e = I$, $W_f = W_{\Delta f} = 0$ in order to only penalize the error.

6.3 Simulation results

In Fig. 8, $\mathcal{J}_j$ is shown as function of the trial index j. The following observations are made. First, the proposed approach outperforms the pre-existing approach for rational basis functions by a factor 40. Second, rational basis functions outperform polynomial basis functions; even the pre-existing approach is a factor 480 better. Third, ILC with basis functions outperforms standard ILC after a change in reference signal by at least a factor 9. The error signal after such a change is depicted in Fig. 9. The results clearly highlight the problem encountered in standard ILC with respect to extrapolation, demonstrate the excellent extrapolation properties of ILC with basis functions, and confirms the analysis in section 2.2.

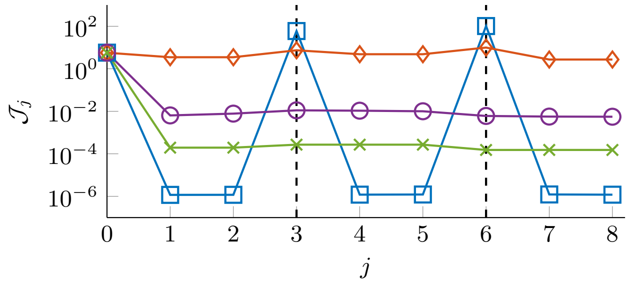

Fig. 8. For rational basis functions the proposed approach (×) outperforms the pre-existing approach (○), which both outperform polynomial basis functions (◇). At reference signal changes (trial $j = 3$ and $j = 6$) basis functions outperform standard ILC (□).

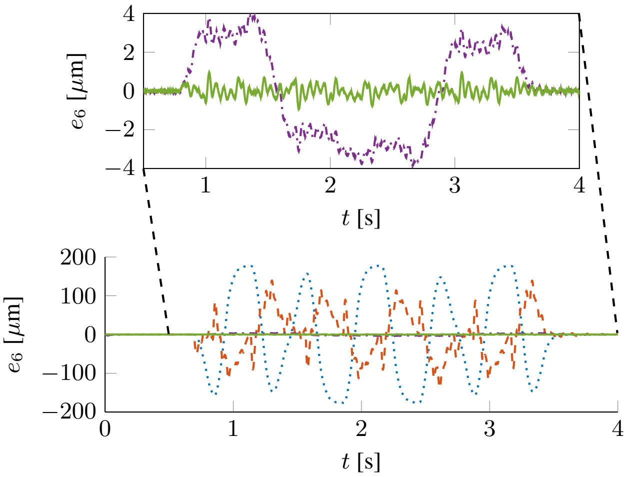

Fig. 9. Time domain error signals after a change in reference signal ($j = 6$) for standard ILC (⋯⋯), polynomial basis functions (- - -), and rational basis functions using the pre-existing (-·-·) and proposed (——) approach. The change in reference signal has significant impact on the performance of standard ILC, whereas the effect on ILC with basis functions is significantly smaller, especially for rational basis functions. Also here the performance difference between the pre-existing and proposed approach is significant as can be seen in the zoom on these two signals in the top.

7 Conclusion

For systems operating repetitively, high performance can be achieved by use of ILC. However, standard norm-optimal ILC has poor extrapolation properties with respect to varying reference signals. In this paper, rational basis functions are used to enhance these extrapolation properties. The associated optimization problem is significantly more complex than for polynomial basis functions, but can be solved iteratively. The main contribution of this paper is a new iterative algorithm for rational basis functions. The algorithm relates to techniques from system identification and is guaranteed to converge to a minimum. The algorithm is compared with the pre-existing iterative algorithm in [5] which is shown to be non-optimal. As a consequence, the proposed approach outperforms the pre-existing results.

The results are supported using simulation examples. It is shown that, even for simple systems, the difference in performance between the pre-existing and proposed approach can be significant. The excellent extrapolation properties of ILC with basis functions are demonstrated in a simulation of a complex industrial system. Also for this simulation, the proposed approach is superior to the pre-existing approach.

Future research is aimed at experimental validation of the simulation results. Ongoing research focuses on the extension to MIMO systems, selection of basis functions, robustness analysis, numerical aspects along the lines of [15]. Finally, an interesting extension could be to investigate convergence to the global minimum along the lines of [8].

Acknowledgements

The authors would like to acknowledge fruitful discussions with and the contributions of Sjirk Koekebakker and Maarten Steinbuch. This work 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).

References

[1] Hyo-Sung Ahn, Kevin L. Moore, and YangQuan Chen. Stability analysis of discrete-time iterative learning control systems with interval uncertainty. *Automatica*, 43(5):892–902, may 2007.

[2] Rogier S. Blom and Paul M.J. Van den Hof. Multivariable Frequency Domain Identification Using IV-based Linear Regression. In *Proceedings of the 49th Conference on Decision and Control*, pages 1148–1153, Atlanta, Georgia, dec 2010. IEEE.

[3] Frank Boeren, Tom Oomen, and Maarten Steinbuch. Iterative motion feedforward tuning: A data-driven approach based on instrumental variable identification. *Control Engineering Practice*, 37:11–19, apr 2015.

[4] 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*, volume 5, pages 3189–3193, Philadelphia, Pennsylvania, jun 1998. IEEE.

[5] Joost Bolder and Tom Oomen. Rational Basis Functions in Iterative Learning Control - With Experimental Verification on a Motion System. *IEEE Transactions on Control Systems Technology*, 23(2):722–729, jun 2015.

[6] Joost Bolder, Tom Oomen, Sjirk Koekebakker, and Maarten Steinbuch. Using iterative learning control with basis functions to compensate medium deformation in a wide-format inkjet printer. *Mechatronics*, 24(8):944–953, dec 2014.

[7] Douglas A. Bristow and Andrew G. Alleyne. Monotonic Convergence of Iterative Learning Control for Uncertain Systems Using a Time-Varying Filter. *IEEE Transactions on Automatic Control*, 53(2):582–585, mar 2008.

[8] Diego Eckhard, Alexandre S. Bazanella, Cristian R. Rojas, and Håkan Hjalmarsson. On the convergence of the Prediction Error Method to its global minimum. In *Proceedings of the 16th IFAC Symposium on System Identification*, pages 698–703, Brussels, Belgium, jul 2012.

[9] Xuemei Gao and Sandipan Mishra. An Iterative Learning Control Algorithm for Portability Between Trajectories. In *Proceedings of the 2014 American Control Conference*, pages 3808–3813, Portland, Oregon, jun 2014. IEEE.

[10] Jayati Ghosh and Brad Paden. A Pseudoinverse-Based Iterative Learning Control. *IEEE Transactions on Automatic Control*, 47(5):831–837, may 2002.

[11] M. Gilson, H. Garnier, P.C. Young, and P.M.J. Van den Hof. Optimal instrumental variable method for closed-loop identification. *IET Control Theory & Applications*, 5(10):1147–1154, aug 2011.

[12] Svante Gunnarsson and Mikael Norrlöf. On the design of ILC algorithms using optimization. *Automatica*, 37(2001):2011–2016, dec 2001.

[13] J. Hätönen, D.H. Owens, and K. Feng. Basis functions and parameter optimisation in high-order iterative learning control. *Automatica*, 42(2):287–294, feb 2006.

[14] Marcel F. Heertjes and René M.J.G. van de Molengraft. Set-point variation in learning schemes with applications to wafer scanners. *Control Engineering Practice*, 17(3):345–356, mar 2009.

[15] Robbert van Herpen, Tom Oomen, and Maarten Steinbuch. Optimally Conditioned Instrumental Variable Approach for Frequency-Domain System Identification. *Automatica*, 50(9):2281–2293, sep 2013.

[16] Peter S.C. Heuberger, Paul M.J. Van den Hof, and Bo Wahlberg. *Modelling and Identification with Rational Orthogonal Basis Functions*. Springer, 2005.

[17] David J. Hoelzle, Andrew G. Alleyne, and Amy J. Wagoner Johnson. Basis Task Approach to Iterative Learning Control With Applications to Micro-Robotic Deposition. *IEEE Transactions on Control Systems Technology*, 19(5):1138–1148, sep 2011.

[18] Paul Lambrechts, Matthijs Boerlage, and Maarten Steinbuch. Trajectory planning and feedforward design for electromechanical motion systems. *Control Engineering Practice*, 13(2):145–157, feb 2005.

[19] Sandipan Mishra, Joshua Coaplen, and Masayoshi Tomizuka. Precision positioning of wafer scanners: Segmented iterative learning control for nonrepetitive disturbances. *IEEE Control Systems Magazine*, 27(August):20–25, aug 2007.

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

[21] Mikael Norrlöf and Svante Gunnarsson. Time and frequency domain convergence properties in iterative learning control. *International Journal of Control*, 75(14):1114–1126, nov 2002.

[22] Sang J. Oh, Minh Q. Phan, and Richard W. Longman. Use of Decoupling Basis Functions in Learning Control for Local Learning and Improved Transients. *Advances in the Astronautical Sciences*, 95(2):651–670, 1997.

[23] Minh Q. Phan and James A. Frueh. Learning Control for Trajectory Tracking using Basis Functions. In *Proceedings of the 35th Conference on Decision and Control*, volume 3, pages 2490–2492, Kobe, Japan, dec 1996. IEEE.

[24] Dick de Roover and Okko H. Bosgra. Synthesis of robust multivariable iterative learning controllers with application to a wafer stage motion system. *International Journal of Control*, 73(10):968–979, nov 2000.

[25] Samer S. Saab. Selection of the Learning Gain Matrix of an Iterative Learning Control Algorithm in Presence of Measurement Noise. *IEEE Transactions on Automatic Control*, 50(11):1761–1774, nov 2005.

[26] K. Steiglitz and L.E. McBride. A Technique For The Identification Of Linear Systems. *IEEE Transactions on Automatic Control*, 10(4):461–464, oct 1965.

[27] Petre Stoica and Torsten Söderström. The Steiglitz-McBride identification algorithm revisited–Convergence analysis and accuracy aspects. *IEEE Transactions on Automatic Control*, 26(3):712–717, jun 1981.

[28] A.H. Whitfield. Asymptotic behaviour of transfer function synthesis methods. *International Journal of Control*, 45(3):1083–1092, 1987.

[29] J. van de Wijdeven and O.H. Bosgra. Using basis functions in iterative learning control: analysis and design theory. *International Journal of Control*, 83(4):661–675, apr 2010.

[30] Jurgen van Zundert, Joost Bolder, and Tom Oomen. Iterative Learning Control for Varying Tasks: Achieving Optimality for Rational Basis Functions. In *Proceedings of the 2015 American Control Conference*, pages 3570–3575, Chicago, Illinois, jul 2015. IEEE.