15th Triennial World Congress, Barcelona, Spain

CONVEX MODIFICATIONS TO AN ITERATIVE LEARNING CONTROL LAW

D. H. Owens [*] **J. Hätönen** [**,*]

[*] *Department of Automatic Control and Systems Engineering, University of Sheffield, Mappin Street, Sheffield S1 3JD*
[**] *Systems Engineering Laboratory, University of Oulu, P.O.BOX 4300, FIN-90014 University of Oulu*

Abstract: In this paper a predictive norm-optimal iterative learning control algorithm from (Amann *et al.*, 1998) is analyzed. As a main new result in this paper it is shown that if also the predictive inputs from the algorithm are used in the control of the plant, a faster convergence can be achieved than with the approach in (Amann *et al.*, 1998). Furthermore, the nature of the convergence of this new scheme is analyzed in detail in terms of the free parameters of the algorithm. *Copyright © 2002 IFAC*

Keywords: Iterative learning control, optimal control, predictive control

1. INTRODUCTION

Iterative learning control is a technique to control systems operating in a repetitive mode with the additional requirement that a specified output trajectory $r(t)$ in an interval $[0, T]$ is to be followed with high precision. Examples of such systems are robot manipulators that are required to repeat a given task to high precision, chemical batch processes, or more generally, the class of tracking systems. Motivated by human learning, the basic idea of iterative learning control is to use information from previous executions of the task in order to improve performance from trial to trial in the sense that the tracking error is sequentially reduced (Arimoto *et al.*, 1984), (Moore, 1993). Typical iterative learning control algorithms construct the input to the plant on a given trial from the input used on the last trial plus an additive incremental which is typically a function of the past values of the observed output error, that is, the difference between achieved output and desired output. The objective of constructing a sequence of input functions $\{u_k(t)\}$, $t \in [0, T]$, such that the performance is gradually improving as the task is repeated, can be refined to a convergence condition on the input and error

$$\lim_{k\to\infty} \|e_k\| \to 0 \quad \lim_{k\to\infty} \|u_k - u_\infty\| = 0 \qquad (1)$$

where $e_k(t)$ is the difference between $r(t)$ and the system output $y_k(t)$ at the kth trial and $u_k(t)$ is the input to the system. This definition of convergent learning is a stability problem on an infinite-dimensional product space, typically of the form $\mathbb{N} \times L_2[0, T]$. Hence it is clear that the analysis of iterative learning systems lie firmly outside the realm of traditional control theory.

In (Amann *et al.*, 1996) an optimality based iterative learning control law was proposed. The idea was to solve the following optimization problem

$$\min_{u_{k+1}} J(u_{k+1}) = \|e_{k+1}\|^2 + \|u_{k+1} - u_k\|^2 \qquad (2)$$

where the tracking error $e_{k+1}(t) = r(t) - G[u_{k+1}](t)$ at trial $k + 1$ is defined as the difference between the desired output trajectory (or reference) and the actual output of the plant G to the input $u_{k+1}(t)$. The system is assumed to be described by the input-output mapping $y_{k+1}(t) = G[u_{k+1}](t)$ where G is a linear operator defined by the system dynamics. The norms $\|\cdot\|^2$ are induced norms from the inner products $< \cdot, \cdot >$ of the chosen input and output Hilbert-spaces U and Y respectively. For a continuous time system the inner-product spaces Y, U are typically chosen as finite product of $L_2[0, T]$. The solution of the optimization problem (2) is

$$u_{k+1} = u_k + G^* e_{k+1} \tag{3}$$

where G^* is the non-causal adjoint operator of G. The corresponding optimal error evolution equation is given by

$$e_{k+1} = (I + H)^{-1} e_k \tag{4}$$

where $H := GG^*$. In (Amann *et al.*, 1996) it is shown that the algorithm gives guaranteed convergence with monotonically decreasing error for all linear, possibly time-varying plants, and the convergence can be fast, if the system is minimum phase. Furthermore, if the system is represented in discrete-time, the convergence is *geometric*. In addition, if the plant G has an equivalent state-space representation, the non-causal solution (3) has an equivalent causal form, which consists of a Riccati state-feedback term and a so-called predictive term.

To improve the convergence speed particularly for non-minimum phase systems, the following cost criterion was introduced in (Amann *et al.*, 1998):

$$J_{k+1,n}(\boldsymbol{u}_{k+1}) = \sum_{i=1}^{n} \lambda^{i-1} (\|e_{k+i}\|^2 + \|u_{k+i} - u_{k+i-1}\|^2) \tag{5}$$

where $\boldsymbol{u}_{k+1} = [u_{k+1} \; u_{k+2} \; \dots \; u_{k+n}]^T$. In the following material $u_{k+1,i}$ refers to the i'th element of $\boldsymbol{u}_{k+1}$ and $e_{k+1,i}$ refers to the i'th element of the vector $\boldsymbol{e}_{k+1} = [e_{k+1} \; e_{k+2} \; \dots \; e_{k+n}]^T$.

The proposed criterion (5) includes the error not only of the current trial but also the predicted error in the next n trials (i.e. n is the prediction horizon) as well as the corresponding changes in the input. The weight parameter $\lambda > 0$ determines the importance of more distant (future) errors and incremental inputs. Furthermore, like in General Predictive Control (Camacho and Bordons, 1998), a receding horizon principle is proposed in (Amann *et al.*, 1998) where at trial $k+1$ only the input $u_{k+1,1}$ is used as an input into the plant, and the optimization is repeated again at the next trial.

By including more future signals into the performance criterion (5), it was argued that the algorithm should become less 'short sighted', and faster convergence should be obtained when compared to (2). This was rigorously proved in (Amann *et al.*, 1998) when the input $u_{k+1,1}$ is used as the 'true' input. In addition, the resulting algorithm from (5) has a causal implementation, if the original plant can be described with a state-space representation.

In this paper it is noted that we can use the receding horizon principle for any other input $u_{k+1,j}$, or more generally, we could take as a 'true' input a positive convex combination of the inputs $u_{k+1,j}$. It is then an important question whether or not a higher convergence rate can be achieved with these choices when compared to the choice where $u_{k+1,1}$ is used as the input. The answer is positive, and will be proved rigorously in the following sections.

2. DERIVATION OF THE ERROR EVOLUTION EQUATIONS

To answer the question proposed above, we establish the error evolution equation between $e_{k+1,j}$ and e_k when using the input $u_{k+1,j}$ for $j = 1, 2, \dots n$. It is a straightforward task (see (Amann *et al.*, 1998)) to show that in matrix form the solution of (5) is given by

$$N_{n,e}(\lambda)\boldsymbol{u}_{k+1} = \tilde{u}_k + G_e^* \boldsymbol{e}_{k+1} \tag{6}$$

where $\tilde{u}_k = [0 \; 0 \; \dots \; 0 \; u_k]^T$ and $N_{n,e}(\lambda)$ is given by

$$N_{n,e}(\lambda) = \begin{bmatrix} I & -I & 0 & \dots & 0 \\ -\lambda I & I+\lambda I & -I & \dots & 0 \\ 0 & -\lambda I & I+\lambda I & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & I+\lambda I \end{bmatrix} \tag{7}$$

and G_e^* is given by

$$G_e^* = \begin{bmatrix} G^* & 0 & 0 & \dots & 0 \\ 0 & G^* & 0 & \dots & 0 \\ 0 & 0 & G^* & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & G^* \end{bmatrix} \tag{8}$$

Based on (6) the error evolution equation can be written as

$$L_{n,e}(H,\lambda) \begin{bmatrix} e_{k+1,n} \\ e_{k+1,n-1} \\ e_{k+1,n-2} \\ \vdots \\ e_{k+1,2} \\ e_{k+1,1} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 0 \\ e_k \end{bmatrix} \tag{9}$$

where

$$L_{n,e}(H,\lambda) = \begin{bmatrix} I+H & -I & 0 & \dots & 0 \\ -\lambda I & S(\lambda,H) & -I & \dots & 0 \\ 0 & -\lambda I & S(\lambda,H) & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & -I \\ 0 & 0 & 0 & \dots & S(\lambda,H) \end{bmatrix} \tag{10}$$

where $S(\lambda, H) := I + \lambda I + H$. Because the elements in (10) commute with each other, and they are linear, bounded, and $I+H$ and $I+\lambda I+H$ are strictly positive operators and each operator has as its range dense in Y (showing that we can calculate with operators as we would calculate with elements of a field) the inverse of $L_{n,e}(H,\lambda)$ can be calculated with the Cramer's rule (MacLane and Birkoff, 1971) as

$$L_{n,e}(H,\lambda)^{-1} = (\det(L_{n,e}(H,\lambda)))^{-1} \text{adj}\,(L_{n,e}(H,\lambda)) \tag{11}$$

where we can prove that $(\det(L_{n,e}(H,\lambda)))^{-1}$ exists. However, because we are only interested in finding the

relationship between $e_{k+1,j}$ and e_k it is easily seen that the operator mapping e_k to $e_{k+1,j}$ is given by

$$\det(L_{n,e}(H,\lambda))^{-1}(-1)^{2n-j+1}M(H,\lambda) \qquad (12)$$

for $j = 1, 2, \ldots, n$ where

$$M(H,\lambda) = |L_{n,e}(H,\lambda)^{(n,n-j+1)}| \qquad (13)$$

and $|L_{n,e}(H,\lambda)^{(n,n-j+1)}|$ is the minor of $L_{n,e}(H,\lambda)$ at position $(n, n-j+1)$. As a preliminary remark it is observed that $L_{n,e}(H,\lambda)$ can be written recursively as

$$L_{n,e}(H,\lambda) = \begin{bmatrix} & & 0 \\ & & 0 \\ & L_{n-1,e}(H,\lambda) & \vdots \\ & & -I \\ 0\ 0 & \ldots & -\lambda I\ \ S(\lambda,H) \end{bmatrix} \qquad (14)$$

or even as

$$L_{n,e}(H,\lambda) = \begin{bmatrix} & & 0 & 0 \\ & & 0 & 0 \\ & L_{n-2,e}(H,\lambda) & \vdots & \vdots \\ & & -I & 0 \\ 0\ 0 & \ldots & -\lambda I\ \ S(\lambda,H) & -I \\ 0\ 0 & \ldots & 0 \quad -\lambda I & S(\lambda,H) \end{bmatrix} \qquad (15)$$

From (14) and (15) it is seen that

$$\det(L_{n,e}(H,\lambda)) = (I + \lambda I + H)\det(L_{n-1,e}(H,\lambda)) - \lambda \det(L_{n-2,e}(H,\lambda)) \qquad (16)$$

with $\det(L_{0,e}(H,\lambda)) = I$ and $\det(L_{1,e}(H,\lambda)) = I + H$. To shorten the notation, we call the n'th operator obtained from the iteration (16) as $p_n(H,\lambda)$, i.e. $p_n(H,\lambda) := \det(L_{n,e}(H,\lambda))$. We know from (Amann *et al.*, 1998) that $p_n(H,\lambda)$ for $n = 1, 2, 3, \ldots$ are bounded self-adjoint strictly positive linear operators and $\mathrm{Rg}\,(p_n(H,\lambda))$ is dense in Y and thus it is mathematically correct to use (11) to calculate the inverse of $L_{n,e}(H,\lambda)$.

To calculate the functional dependence from e_k to $e_{k+1,1}$ we need the cofactor of $L_{n,e}(H,\lambda)$ at position (n,n). But (16) points out that it is $\det(L_{n-1,e}(H,\lambda)) = p_{n-1}(H,\lambda)$. Hence

$$e_{k+1,1} = p_n(H,\lambda)^{-1}p_{n-1}(H,\lambda)e_k \qquad (17)$$

A similar argument shows that for $e_{k+1,j}$, $j \in \{1, 2, \ldots, n\}$

$$e_{k+1,j} = p_n(H,\lambda)^{-1}p_{n-j}(H,\lambda)e_k \qquad (18)$$

As we now know the error evolution equation between $e_{k+1,j}$ and e_k we can answer the question proposed in the first section, i.e. do the inputs $u_{k+1,j}$ for $j = 2, 3, \ldots n$ or, more generally, convex combinations of the inputs, result in faster convergence than that observed when $u_{k+1,1}$ alone is used? This is done in the next section by carefully analyzing the properties of the family of learning operators given in (18).

3. PARTIAL ORDERINGS FOR THE LEARNING OPERATORS

In this section we introduce partial orderings for the operators $L_{n,j}(H,\lambda) = p_n(H,\lambda)^{-1}p_{n-j}(H,\lambda)$ for $j \in \{1, \ldots n\}$ in the following way (Kreyszig, 1978): Let T_1 and T_2 be bounded linear self-adjoint operators on a complex Hilbert space Y. Then $T_1 \geq T_2$ if and only if $\langle T_1 e, e\rangle \geq \langle T_2 e, e\rangle$ (or $T_1 > T_2$ if and only if $\langle T_1 e, e\rangle > \langle T_2 e, e\rangle$) for an arbitrary non-zero element $e \in Y$. Another important concept from functional analysis in this section is a positive operator (Kreyszig, 1978): a bounded self-adjoint linear operator T is positive if and only if $T \geq 0$ (or T is strictly positive if and only if $T > 0$). This implies that $T_1 \geq T_2$ if and only if $T_1 - T_2 \geq 0$ (or $T_1 > T_2$ if and only if $T_1 - T_2 > 0$).

In the end of the subsection we show how the partial orderings are related to convergence rate, and find out that by using a positive convex combination of the inputs $\{u_{k+1,1} \ldots u_{k+1,n}\}$ (i.e. using the input $u_{k+1} := \sum_{i=1}^{n} \alpha_i u_{k+1,i}$, where $\sum_{i=1}^{n} \alpha_i = 1$ and $\alpha_i \geq 0$) as the true input for the plant, a quicker convergence rate can be obtained than using merely $u_{k+1,1}$ as the true input. Furthermore, the convergence is geometric for a positive convex combination if for $\forall e \in Y, e \neq 0$, $\langle He, e\rangle \geq \sigma^2\|e\|^2$ where $\sigma^2 > 0$ and almost geometric if $\sigma^2 = 0$.

Most of the proofs are based on a standard result in functional analysis (Kreyszig, 1978), namely that if two bounded self-adjoint linear operators S and T on a complex Hilbert space Y are positive and commute, then their product $ST = TS$ is positive.

We start by collecting from (Amann *et al.*, 1998) the following useful properties of the family of learning operator $L_n(H,\lambda) = L_{n,1}(H,\lambda)$:

Property 1. Assume that $H \geq 0$, $\|H\| < \infty$, $0 < \lambda < \infty$. Then $L_n(\lambda, H) > 0,\ n = 0, 1, 2, \ldots$.

From this on we can assume always that $\|H\| < \infty$, because G is supposed to be a linear dynamical system defined over a *finite-time*, and hence it is bounded, implying $H = GG^*$ is bounded.

Property 2. $0 < L_{n+1}(H,\lambda) < L_n(H,\lambda) < I$ for $H > 0$, $\lambda > 0$, $n = 0, 1, 2, \ldots$, and the operators $L_n(H,\lambda)$ are self-adjoint and commute with each other for $n = 0, 1, 2, \ldots$.

Property 3. Assume that $H > 0$, $0 < \lambda < \lambda'$. Then $L_n(H,\lambda) > L_n(H,\lambda'),\ n = 1, 2, 3, \ldots$.

The first step is to show that a similar partial ordering exists as in the Property 2 for the operators $L_{n,i}(H,\lambda)$ as a function of i.

Proposition 1. Assume that H>0. Then $L_n = L_{n,1} > L_{n,2} > \cdots > L_{n,n-1} > L_{n,n}$

Proof. First note that for $j \in \{1, \ldots n\}$

$$\begin{aligned} &L_{n,j} = p_n^{-1} p_{n-j} \\ &= p_n^{-1} p_{n-1} p_{n-1}^{-1} \cdots p_{n-j+1} p_{n-j+1}^{-1} p_{n-j} \\ &= L_n L_{n-1} \ldots L_{n-j+1} \end{aligned} \tag{19}$$

Furthermore $L_{n,k} = L_n L_{n-1} \ldots L_{n-k+1} > 0$ and $I - L_{n-k} > 0$ based on Property 2. Thus $L_{n,k} > L_n L_{n-1} \ldots L_{n-k+1} L_{n-k} = L_{n,k+1}$ for $k \in \{1, \ldots n-1\}$.

This result suggests that by including the information from the inputs $u_{k+1,j}$, $j = \{2, \ldots n\}$ in the calculation of the current input signal quicker convergence can be achieved when compared to the case where merely u_{k+1} is used (see the end of this subsection how the partial order is related to convergence speed). One interesting approach to do this inclusion would be to select the current input as a positive convex combination of the input signals, i.e. $u_{k+1} = \sum_{j=1}^{n} \alpha_i u_{k+1,j}$, where $\alpha_j \geq 0$, $\sum_{j=1}^{n} \alpha_i = 1$. In this case the error evolution is governed by the following equation

$$\begin{aligned} &e_{k+1} = \\ &p_n^{-1}(\alpha_1 p_{n-1} + \alpha_2 p_{n-2} + \cdots + \alpha_n p_0) e_k \\ &= (\alpha_1 L_{n,1} + \alpha_2 L_{n,2} + \cdots + \alpha_n L_{n,n}) e_k \\ &= L_{n,c}(H, \lambda, \boldsymbol{\alpha}) e_k \end{aligned} \tag{20}$$

where $L_{n,c}(H, \lambda, \boldsymbol{\alpha})$ reflects the fact that the properties of the new learning operator $L_{n,c}(H, \lambda, \boldsymbol{\alpha})$ is also dependent on the choice of $\boldsymbol{\alpha} := \{\alpha_1, \ldots \alpha_n\}$. The next proposition shows that by taking an arbitrary convex combination results gives a learning operator $L_{n,c}(H, \lambda, \boldsymbol{\alpha})$ which lies between $L_{n,1}(H, \lambda)$ and $L_{n,n}(H, \lambda)$ in the sense of the partial order.

Proposition 2. Assume that $H > 0$. Then $L_n > L_{n,c} > L_{n,n}$ if $L_{n,c} \neq L_n$ and $L_{n,c} \neq L_{n,n}$.

Proof.

$$\begin{aligned} &L_{n,n} = \sum_{i=1}^{n} \alpha_i L_{n,n} < \sum_{i=1}^{n} \alpha_i L_{n,i} = L_{n,c} \\ &< \sum_{i=1}^{n} \alpha_i L_{n,1} = L_{n,1} = L_n \end{aligned} \tag{21}$$

The next issue is to analyze whether or not λ introduces a partial ordering for the learning operator. This question is answered in the following proposition:

Proposition 3. Let $\lambda' > \lambda$, $\lambda', \lambda > 0$ and $H > 0$. Then $L_{n,c}(H, \lambda', \boldsymbol{\alpha}) < L_{n,c}(H, \lambda, \boldsymbol{\alpha})$.

Proof. Take an arbitrary $j \in \{1, \ldots n-1\}$. From Property 3 we now that $L_{n-j}(H, \lambda') < L_{n-j}(H, \lambda)$. Thus $L_{n-j+1}(H, \lambda) L_{n-j}(H, \lambda) > L_{n-j+1}(H, \lambda) L_{n-j}(H, \lambda')$. However, because $L_{n-j+1}(\lambda, H) - L_{n-j+1}(H, \lambda') > 0$, it implies that

$$\begin{aligned} &L_{n-j}(H, \lambda') L_{n-j+1}(H, \lambda) > \\ &L_{n-j}(H, \lambda') L_{n-j+1}(\lambda', H) \end{aligned} \tag{22}$$

Thus

$$\begin{aligned} &L_{n-j+1}(H, \lambda) L_{n-j}(H, \lambda) > \\ &L_{n-j+1}(H, \lambda') L_{n-j+1}(H, \lambda') \end{aligned} \tag{23}$$

We can repeat this argument inductively and thus

$$\begin{aligned} &L_n(H, \lambda) L_{n-1}(H, \lambda) \ldots L_{n-j}(H, \lambda) > \\ &L_n(H, \lambda') L_{n-1}(H, \lambda') \ldots L_{n-j}(H, \lambda') \end{aligned} \tag{24}$$

and consequently

$$\begin{aligned} &L_{n,c}(H, \lambda, \boldsymbol{\alpha}) = \sum_{i=1}^{n} \alpha_i L_{n,i}(H, \lambda) > \\ &\sum_{i=1}^{n} \alpha_i L_{n,i}(H, \lambda') = L_{n,c}(H, \lambda', \boldsymbol{\alpha}) \end{aligned} \tag{25}$$

which completes the proof.

The next step is to analyze whether or not the prediction horizon n introduce a partial ordering for the operators. However, because for an arbitrary set of $\{\alpha_1, \ldots \alpha_n\}$ it is not clear how to select a new set of $\{\alpha'_1, \ldots \alpha'_{n+1}\}$ so that the comparison would make logical sense, we look at a special case where for a fixed $\alpha \in \mathbb{R}_+$ an arbitrary $n \in \mathbb{N}_+$ we always select $\alpha_1 = \alpha$, $\alpha_i = 0$ for $i \in \{2, \ldots n-1\}$ and $\alpha_n = 1 - \alpha$, i.e. $L_{n,c}(H, \lambda, \alpha) = \alpha L_1(H, \lambda) + (1 - \alpha) L_n(H, \lambda)$. This specific selection gives the following proposition:

Proposition 4. Assume that $H > 0$.
Then $L_{n+1,c}(H, \lambda, \alpha) < L_{n,c}(H, \lambda, \alpha)$

Proof. Because $I - L_{n+1}(H, \lambda) > 0$, this implies that

$$\begin{aligned} &L_n(H, \lambda) L_{n-1}(H, \lambda) \ldots L_1(H, \lambda) > \\ &L_{n+1}(H, \lambda) L_n(H, \lambda) \ldots L_1(H, \lambda) \end{aligned} \tag{26}$$

and hence

$$\begin{aligned} &L_{n,c}(H, \lambda, \alpha) = \alpha L_{n,1}(H, \lambda) + (1 - \alpha) L_{n,n}(H, \lambda) \\ &> \alpha L_{n+1,1}(H, \lambda) + (1 - \alpha) L_{n+1,n+1}(H, \lambda) \\ &= L_{n+1,c}(H, \lambda, \alpha) \end{aligned} \tag{27}$$

because based on Property 2, $L_{n,1}(H, \lambda) = L_n(H, \lambda) > L_{n+1}(H, \lambda) = L_{n+1,1}(H, \lambda)$.

The next important question is how these partial orderings are then linked to the convergence properties of the learning operators? To answer this question we need first the following proposition from (Amann *et al.*, 1998):

Proposition 5. Let $\sigma \geq 0$ such that $\langle e, He \rangle \geq \sigma^2 \langle e, e \rangle \; \forall e \in Y$. Then

$$L_n(H, \lambda) \leq l_n(\sigma, \lambda) \qquad \forall \lambda > 0, n = 0, 1, 2, \ldots \tag{28}$$

where

$$l_{n+1}(\sigma,\lambda)=\frac{1}{1+\sigma^2+\lambda-\lambda l_n(\sigma,\lambda)},\qquad l_0=1 \tag{29}$$

and the error sequence is bounded by $\|e_{k+1}\| \leq l_n(\sigma,\lambda)\|e_k\|$.

Note that if $\sigma > 0$ (which holds for the important case of discrete-time systems) then $0 < l_n(\sigma,\lambda) < 1$, $n \in \{1,2,\dots\}$ but if $\sigma = 0$ then $l_n(\sigma,\lambda) = 1$ for $n \in \{1,2,\dots\}$. For a bounded self-adjoint operator T on a complex Hilbert-space Y it is valid that $\|T\| = \sup_{\|x\|=1} |\langle Tx, x\rangle|$ implying that $\|L_n(H,\lambda)\| \leq l_n(\sigma,\lambda)$. Thus $L_n(H,\lambda)$ is a contraction if $\sigma > 0$. Furthermore in this case all the Propositions from 1 to 4 can be rewritten so that $L_n(H,\lambda)$ is replaced with $l_n(\sigma,\lambda)$ giving the following qualitative conclusions:

i) The lowest convergence rate is achieved by using $u_{k+1,1}$ as an true input and the quickest convegence rate is achieved bu using $u_{k+1,n}$ as the true input. By using a convex combination of the inputs the convergence rate lies between the convergence rate given by $u_{k+1,1}$ and $u_{k+1,n}$. The convergence is geometric in each case.

ii) Increasing λ increases the convergence rate for an arbitrary positive convex combination of $\{u_{k+1,1}\dots u_{k+1,n}\}$.

iii) If the coefficients α_i of $L_{n,c}(H,\lambda,\boldsymbol{\alpha})$ are selected as in Proposition 4 then $\|L_{n+1,c}(H,\lambda,\boldsymbol{\alpha})\| < \|L_{n,c}(H,\lambda,\boldsymbol{\alpha})\|$. This shows that by increasing the prediction horizon a quicker convergence rate is achieved.

The underlying problem, however, is that for a strictly proper continuous-time system the lower bound in Proposition 5 is in fact zero (an operator having a strictly positive lower bound on an infinite-dimensional space Y cannot be compact but H is a compact operator). Thus in the previous propositions and properties in this section $<$ has to be replaced with $\leq$ when $H \geq 0$. Especially $\|L_{n,c}(H,\lambda,\boldsymbol{\alpha})\| \leq 1$ and thus strict geometric convergence is not possible to achieve. Consequently it would be interesting to know which error functions are left unaffected by $L_{n,c}(H,\lambda,\boldsymbol{\alpha})$. This question is answered in the following proposition for $L_n(H,\lambda)$:

Proposition 6. $L_n(H,\lambda)e = e$ if and only if $e \in$ *Ker* (H).

Proof. Suppose that $e \in \text{Ker}\,(H)$. Then $L_0(H,\lambda) = Ie = e$. Assume now that $e \in \text{Ker}\,(H)$ implies that $L_n(H,\lambda)e = e$. Then

$$L_{n+1}(H,\lambda)e = \left(I+\lambda I+H-\lambda L_n(H,\lambda)\right)^{-1}e \tag{30}$$

or (because $L_{n+1}(H,\lambda)$ is invertible and the operators commute)

$$L_{n+1}(H,\lambda)\left(I+\lambda I+H-\lambda L_n(H,\lambda)\right)e = e \tag{31}$$

but based on the induction assumption and the fact that $He = 0$

$$L_{n+1}(H,\lambda)e = e \tag{32}$$

Assume now that there exists $e \in Y$ such that $L_1(H,\lambda)e = (I+H)^{-1}e = e$. This shows that $(I+H)e = e$ implying $He = 0$ and hence $e \in \text{Ker}(H)$. Assume now that there exists $e \in Y$ such that $L_n(H,\lambda)e = e$. This implies that $\langle L_n(H,\lambda)e, e\rangle = \langle e, e\rangle = \|e\|^2$. However, Property 2 shows that $I \geq L_1(H,\lambda) \geq L_n(H,\lambda)$ showing that

$$\langle e,e\rangle \geq \langle L_1(H,\lambda)e,e\rangle \geq \langle L_n(H,\lambda)e,e\rangle = \langle e,e\rangle \tag{33}$$

which implies that $\langle L_1(H,\lambda)e,e\rangle = \langle e,e\rangle$ and hence $L_1(H,\lambda)e = e$ and $e \in \text{Ker}(H)$.

The extension to $L_{n+1,c}(H,\lambda,\boldsymbol{\alpha})$ is straightforward: from Proposition 6 it is clear that if $e \in \text{Ker}\,(H)$ then $L_{n,c}(H,\lambda,\boldsymbol{\alpha})e = e$. On the other we know from Proposition 2 that $I \geq L_1(H,\lambda) \geq L_{n,c}(H,\lambda,\boldsymbol{\alpha})$. Thus if $L_{n,c}(H,\lambda,\boldsymbol{\alpha})e = e$ a similar argument as in Proposition 6 shows that $L_1(H,\lambda)e = e$ implying $e \in \text{Ker}(H)$. Furthermore, it is easy to show that $\text{Ker}\,(H) = \text{Ker}\,(G^*) = (\text{Rg}\,(G))^{\perp}$. Thus we see that if $r \notin \text{Ker}\,(G^*) = (\text{Rg}\,(G))^{\perp}$ then the algorithm converges even when $\sigma^2 = 0$ in Proposition 5. The nature of this convergence is shown in the next section.

4. ALMOST GEOMETRIC CONVERGENCE

As it was mentioned in the previous section, for a strictly proper continuous-time plant geometric convergence cannot be achieved, and this was due to the fact that H is compact operator in an infinite-dimensional space. However, as H is a compact self-adjoint operator on a Hilbert space Y, then H has the following spectral decomposition

$$He = \sum_j \sigma_j^2 \langle e,\psi_j\rangle \psi_j \tag{34}$$

where $\{\sigma_j^2\}$ are the eigenvalues of H and $\{\psi_j\}$ are the corresponding eigenvectors. In addition, the $\sigma_i^2 \to 0$ and 0 belongs to the spectrum of H. Furthermore, if Y is separable and infinite-dimensional (for example $L_2[0,T]$) then the eigenvectors form a complete orthonormal basis for Y. This means that an arbitrary vector $e \in Y$ can be written as $e = \sum_j \langle e,\psi_j\rangle \psi_j$. Thus when using the standard algorithm (i.e. $u_{k+1,1}$ is the true input) the evolution of the error signal can be written as

$$e_{k+1} = L_n(H,\lambda)e_k = \sum_j l_n(\sigma_j,\lambda)\langle e_k,\psi_j\rangle\psi_j \tag{35}$$

where $l_n(\sigma_j,\lambda)$ is obtained from the iteration (29). If a convex combination of the inputs is used as the true

input for the plant, then in this case the error evolution is governed by

$$e_{k+1} = L_{n,c}(H,\lambda)e_k = \sum_j \left(\alpha_1 l_{n,1}(\sigma_j,\lambda) + \alpha_2 l_{n,2}(\sigma_j,\lambda) + \ldots + \alpha_{n-1} l_{n,n-1}(\sigma_j,\lambda) + \alpha_n l_{n,n}(\sigma_j,\lambda)\right) \langle e_k, \psi_j \rangle \psi_j \quad (36)$$

where $l_{n,j}(\sigma_i,\lambda) = l_n(\sigma_i,\lambda) l_{n-1}(\sigma_i,\lambda) \ldots l_{n-j+1}(\sigma_i,\lambda)$ for $j \in \{1,\ldots n\}$. In the standard case where the true input is $u_{k+1,1}$ the following proposition from (Amann *et al.*, 1998) shows the 'almost geometric convergence' property.

Proposition 7. For any $\epsilon > 0$, let m be an integer such that the approximation error satisfies $\|e_0 - \sum_{j=1}^{m} \langle e_0, \psi_j \rangle \psi_j\| < \epsilon$. Then the following bound holds:

$$\|e_k\| = l_n(\sigma_m,\lambda)^k \|e_0\| + \epsilon \quad (37)$$

If the true input is a convex combination of $\{u_{k+1,1} \ldots u_{k+1,n}\}$ an easy modification of Proposition 7 shows that in case the bound in (37) can be replaced with

$$\|e_k\| \leq \left(\sum_{i=1}^{n} \alpha_i l_{n,i}(\sigma_m,\lambda) \right)^k \|e_0\| + \epsilon \quad (38)$$

Thus if $r(t) \notin R(G)^{\perp}$ then the previous propositions guarantee almost geometric convergence. Furthermore, the convergence is dependent on the projection of $r(t)$ on the eigenvectors of $L_n(H,\lambda)$ (note that $L_{n,c}(H,\lambda,\boldsymbol{\alpha})$ has the same eigenvectors as $L_n(H,\lambda)$) in the sense that if $r(t)$ can be approximated accurately with 'low-frequency' eigenvectors of $L_n(H,\lambda)$ the convergence is faster than when compared to the case where $r(t)$ contains significant 'high-frequency' components.

5. CONCLUSIONS

In this paper the mathematical properties of the predictive norm optimal iterative control proposed in (Amann *et al.*, 1998) were analyzed in detail. As a new result it was found out that by using a convex combination of the predictive inputs from the algorithm a faster convergence can be obtained than by the previously used input $u_{k+1,1}$. Furthermore, the convergence properties of the new scheme were analyzed both in terms of the weighing parameter λ and the prediction horizon n. The analysis showed that by increasing either the prediction horizon n or the weighing factor λ the convergence rate will increase.

Robustness is an issue that was not properly addressed in this paper. This is an open problem for further research.

6. ACKNOWLEDGEMENTS

J. Hätönen is supported by the EU TMR contract No FMRX-CT970137 (Nonlinear Control Network).

7. REFERENCES

Amann, N., D.H. Owens and E. Rogers (1996). Iterative learning control using optimal feedback and feedforward actions. *International Journal of Control* **65**(2), 277–293.

Amann, N., D.H. Owens and E. Rogers (1998). Predictive optimal iterative learning control. *International Journal of Control* **69**(2), 203–226.

Arimoto, S., S. Kawamura and F. Miyazaki (1984). Bettering operations of robots by learning. *Journal of Robotic Systems* **1**, 123–140.

Camacho, E.F and C. Bordons (1998). *Model Predictive Control*. Springer.

Kreyszig, E. (1978). *Introductory Functional Analysis with Applications*. John Wiley.

MacLane, S. and G. Birkoff (1971). *Algebra*. MacMillan.

Moore, K.L. (1993). *Iterative Learning Control for Deterministic Systems*. Springer-Verlag.