Robust Gradient-based Iterative Learning Control

D.H. Owens, J. Hätönen, S. Daley
Department of Automatic Control and Systems Engineering,
University of Sheffield,
Mappin Street, Sheffield S1 3JD, United Kingdom
Email: D.H.Owens@sheffield.ac.uk

Abstract—This paper considers the robustness of a gradient-based Iterative Learning Control (ILC) algorithm to ensure monotonic convergence with respect to the mean square value of the error time series. The paper provides necessary and sufficient conditions for robust monotonic convergence and sufficient frequency domain conditions for robust monotonic convergence on finite time intervals.

Keywords: Iterative learning control, robust control, parameter optimization, positive-real systems

I. INTRODUCTION

Iterative Learning Control (abbreviated to ILC in the sequel) is concerned with the performance of systems that operate in a repetitive manner where the task is to follow some specified output trajectory in a specified time interval with high precision. ILC uses information from previous executions of the task in an attempt to improve performance from repetition to repetition in the sense that the tracking error (between the output and the specified reference trajectory) is sequentially reduced to zero (see [1] and [6]). Note that repetitions are often called trials, passes or iterations in the literature.

This paper considers the idea of gradient-based ILC algorithms for discrete-time systems. The paper analyses the behaviour and robustness of these algorithms in a rigorous manner. Note that the analysis of continuous-time gradient based algorithms have been carried out in [2] and [5]. In this paper, robustness is defined in terms of a new concept of Robust Monotone convergence introduced by the authors in [3]:

Definition: An ILC algorithm has the property of robust monotone convergence with respect to a vector norm \( \| \cdot \| \) in the presence of a defined set of model uncertainties if, and only if, for every choice of control on the first trial (and hence for every choice of initial error) and for any choice of model uncertainty within the defined set, the resulting sequence of iteration error time signals converges to zero with a strictly monotonically decreasing norm.

The requirement of monotonicity is representative of a practical requirement to improve tracking from trial to trial. The mean square value of the error time series is used as a norm as it will be seen that it has useful analytical properties in generating checkable design conditions.

A companion paper [3] uses the idea of an inverse model-based algorithm with learning gain \( \beta \in (0, 1) \) with excellent results if the plant model mismatch is zero but, in the presence of a multiplicative uncertainty (with transfer function \( U(z) \)), robust monotone convergence is ensured if

\[
\frac{1}{\beta} - U(z) < \frac{1}{\beta}, \quad \forall |z| = 1
\] (1)

A simple analysis of this expression indicates that:

1) significant high frequency errors such as high frequency parasitic resonant modes will require small values of learning gain \( \beta \) and hence slow convergence of the algorithm.

2) The phase of the uncertainty must lie in the open range \((-\pi, \pi)\), constraining the form of uncertainty that can be tolerated. It is equivalent to \( U(z) \) being strictly positive real i.e. \( \text{Re}[U(z)] > 0, |z| = 1 \).

3) If \( U(z) \) is not known but is known to belong to the set characterized by an inequality of the form with \( \beta \) replaced by \( \beta^* \) then robust monotone convergence is guaranteed for all choice of gains in the range \( 0 < \beta < \beta^* \) (see [7] for a more extensive review of this topic).

In contrast, for a process with transfer function \( G(z) = G_0(z)U(z) \) where \( G_0(z) \) is a nominal model used for control purposes, this paper will show that the proposed gradient-based algorithm is robust monotone convergent if

\[
\frac{1}{\beta} - |G_0(z)|^2 U(z) < \frac{1}{\beta}, \quad \forall |z| = 1
\] (2)

This does not remove the need for a strictly positive real \( U(z) \). It can however remove the destabilizing
effect of high frequency errors as, in practice, both
\( G(z) \) and \( G_0(z) \) are low pass filters and hence \( G_0(z) \)
will be small at high frequencies.

II. PROBLEM DEFINITION

Consider a standard discrete-time, linear, time-
invariant single-input, single-output state-space re-
presentation defined over a finite, discrete time interval,
\( t \in [0, N] \). The system is assumed to be repetitive
and, at the end of each repetition, the state is reset to
a specified repetition-independent initial condition. A
reference signal \( r(t) \) is assumed to be specified and the
ultimate control objective is to find an input function
\( u^*(t) \) so that the resultant output function \( y(t) \) tracks
this reference signal \( r(t) \) exactly on \([0, N]\). The process
model is written in the form:

\[
\begin{align*}
x(t + 1) &= Ax(t) + Bu(t) \quad x(0) = x_0 \\
y(t) &= Cx(t) + Du(t)
\end{align*}
\]

where \( t \) is the sample number, the state \( x(t) \in \mathbb{R}^n \),
output \( y(t) \in \mathbb{R} \) and input \( u(t) \in \mathbb{R} \). From now
on it will be assumed that either \( D \neq 0 \) or that
\( CA^{k-1}B = 0 \) and \( CA^k-1B \neq 0 \) for some \( k^* \geq 1 \) (trivially satisfied in practice). If \( D \neq 0 \),
then take \( k^* = 0 \). By construction, \( k^* \) is then
the relative degree of the transfer function \( G(z) \) of the
system. The notation \( f_k(t) \) will denote the value of a
signal \( f \) at sample interval \( t \) on iteration \( k \).

The repetitive nature of the problem opens up possi-
bilities for modifying iteratively the input function
\( u(t) \) so that, as the number of repetitions increases,
the system asymptotically learns the input function that
gives perfect tracking i.e. the control objective is to find
a causal recursive control law with the properties that,
independent of the control input time series chosen for
the first trial, the resultant sequence of error and input
signals satisfy

\[
\lim_{k \to \infty} \| e_k(t) \| = 0 \quad \lim_{k \to \infty} \| u_k(t) - u^*(t) \| = 0
\]

where \( \| \cdot \| \) denotes any norm for the time series. In
what follows, this norm is taken to be the Euclidean
\( \| f \| = \sqrt{\int f^T f} \) in \( \mathbb{R}^p \) which is related to the
mean square error of the time series by the multiplier
\( \sqrt{\text{p}} \).

III. MATRIX REPRESENTATIONS OF PLANT
DYNAMICS

The idea of matrix models is not new (see for example [3]) but their use in analysis has been limited
to computation. To construct this matrix model in
\( \mathbb{R}^{N+1} \), define the time series “super-vectors” on the

\( k^{th} \) trial via

\[
\begin{align*}
u_k &= [u_k(0), \ldots, u_k(N)]^T \\
y_k &= [y_k(0), \ldots, y_k(N)]^T \\
r &= [r(0), \ldots, r(N)]^T \\
e_k &= [e_k(0), \ldots, e_k(N)]^T = r - y_k
\end{align*}
\]

Furthermore, let \( u^* \) be the input sequence (in time
series or supervector form) that gives \( r(t) = [G, u^*](t) \)
where \( G_e \) is process model convolution map (3).

The input-output response of the system can be
written in the form, \( k \geq 0 \),

\[
y_k = G_e u_k + d_0
\]

where \( G_e \) has dimension \((N+1) \times (N+1)\) and
the lower triangular band structure \( G_e = (G_e)_{ij} =
\underbrace{(G_e)_{i}(i+1)(j+1)}_{(i,j)} \) required by causality and time invari-
ance i.e.

\[
G_e = \begin{bmatrix}
D & 0 & 0 & \ldots & 0 \\
CB & D & 0 & \ldots & 0 \\
CAB & CB & D & \ldots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
CA^{N-1}B & CA^{N-2}B & \ldots & \ldots & D
\end{bmatrix}
\]

Also \( d_0 = [Cx_0, CAx_0, \ldots, CA^N x_0]^T \).

Suppose that the plant transfer function \( G(z) =
(CzI - A)^{-1} B + D \) has relative degree (pole-zero
excess) \( k^* \geq 0 \). Assume also that the reference signal
\( r(t) \) satisfies \( r(j) = CA^j x_0 \) for \( 0 \leq j < k^* \)
(or, alternatively, that tracking in this interval is not
important). Then (in a similar manner to [4]) it is
sufficient to analyse a 'lifted' plant equation that is
just the above if \( k^* = 0 \) or, if \( k^* \geq 1 \),

\[
y_{k,l} = G_{e,l} u_{k,l} + d_1
\]

where the signals \( u, y, e, r \) etc are modified to
\( u_{k,l} = [u_k(0), \ldots, u_k(N - k^*)]^T, y_{k,l} =
y_k(k^*) y_k(2) \ldots y_k(N)]^T \) etc and

\[
G_{e,l} = \begin{bmatrix}
CA^{k-1}B & 0 & \ldots & 0 & 0 \\
CA^k B & CA^{k-1}B & \ldots & 0 & 0 \\
CA^{k+1}B & \ldots & \ldots & 0 & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
CA^{N-1}B & \ldots & \ldots & CA^{N-1}B
\end{bmatrix}
\]

with \( d_1 = [CA^{k*} x_0, \ldots, CA^N x_0]^T \). For notational
convenience, the subscripts \( e, l \) are dropped and the
model is written in all cases \( k^* \geq 0 \) in the simplified
notational form

\[
y_k = Gu_k + d
\]

where

1) \( G \) is invertible by construction
2) A comparison of $G$ with $G_e$ indicates that $G$ can be identified with a plant with transfer function $G^*(z) = z^k G(z)$.

3) An examination of $G_e$ or $G$ indicates that higher order Markov parameters do not appear in the matrix model. As a consequence, the system is indistinguishable from any of the Finite Impulse Response (FIR) models with transfer function $G^*(z) = z^k G(z)$.

The gradient-based ILC algorithm evolves from its initial error $e_0$ as follows

$$k+1 = (I - \beta GG^T)k, \quad k \geq 0 \quad (24)$$

From now on this lifted plant model will be used as a starting point for analysis and the identification of the matrix $G$ with the transfer function $G^*(z)$ will be used as required.

Let $F$ be the (right-shift) matrix with elements $F_{ij} = \delta_{i,j+1}$

$$F = \begin{bmatrix} 0 & 0 & 0 & \ldots & 0 \\ 1 & 0 & 0 & \ldots & 0 \\ 0 & 1 & 0 & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & 1 & 0 \end{bmatrix} \quad (12)$$

so that

$$F^j \neq 0, \quad 0 \leq j \leq N - k^*, \quad F^j = 0 \quad \forall \quad j \geq N + 1 - k^* \quad (13)$$

A simple calculation then indicates that

$$G = \Sigma_{j=1}^{N+1-k^*} g_j F^j \quad (14)$$

for suitable choice of scalars $\{g_j\}$. It is also true that all such matrices can be identified (non-uniquely) with linear time invariant systems. Let

$$\mathcal{L}_l = \{G \in \mathcal{R}^{l \times 1} : G \in \text{span}\{F^j\}_{1 \leq j \leq l}\} \quad (15)$$

Then the following statements are easily proven:

$$\{G_1 \in \mathcal{L}_l \quad \& \quad G_2 \in \mathcal{L}_l\} \implies \{G_1 + G_2 \in \mathcal{L}_l\} \quad (16)$$

$$\{G_1 \in \mathcal{L}_l \quad \& \quad G_2 \in \mathcal{L}_l\} \implies \{G_1G_2 \in \mathcal{L}_l\} \quad (17)$$

$$\{G_1 \in \mathcal{L}_l \quad \& \quad G_2 \in \mathcal{L}_l\} \implies \{G_1G_2 = G_2G_1\} \quad (18)$$

$$\{G \in \mathcal{L}_l \quad \& \quad |G| \neq 0\} \implies \{G^{-1} \in \mathcal{L}_l\} \quad (19)$$

Matrix representations obey all of the normal rules of transfer functions in series and parallel connections (provided that they operate on the same underlying time series).

For the purposes of this paper, $L_l$ has additional useful structure. Let $F_0$ be defined to be the (time-reversal) matrix with elements $F_{ij} = \delta_{i,N-k^*+j}$ i.e.

$$F_0 = F_0^T = \begin{bmatrix} 0 & \ldots & 0 & 1 \\ 0 & \ldots & 1 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & \ldots & 0 \end{bmatrix}, \quad F_0^2 = I \quad (20)$$

If $s \in \mathcal{R}^l$ is the column vector of a time series of length $l$, then $F_0 s$ is a column vector of the same time series but reversed in time i.e. $(F_0 s)_j = s_{l+1-j}$ for $1 \leq j \leq l$. After a little manipulation,

$$G \in \mathcal{L}_l \implies F_0 G F_0 = G^T \quad (21)$$

These definitions enable the interpretation of $G^T$ as a dynamical system.

$$\{\tilde{y} = G^T \tilde{u}\} \Leftrightarrow \{(F_0 \tilde{u}) = G(F_0 \tilde{u})\} \quad (22)$$

i.e. the time series $\tilde{y} = G^T \tilde{u}$ is simply the time reversed response of the linear system $G$ (with zero initial conditions) to the time reversal of $\tilde{u}$.

IV. A GRADIENT-BASED ILC ALGORITHM

The purpose of this section is to provide necessary and sufficient conditions for monotonic convergence of the mean square error to zero in the presence of a specific multiplicative modelling error. These conditions take the form of matrix inequalities that will be transformed into more useful frequency domain conditions in the following sections.

Consider the matrix model

$$y_k = Gu_k + d, \quad k \geq 0.$$

The resultant error is $e_k = r - d - Gu_k$. A simple analysis of $||e_k||^2 = e_k^T e_k$ indicates that the steepest descent direction for the error is just $G^T e_k$ and hence that the feedforward ILC algorithm

$$u_{k+1} = u_k + \beta G^T e_k \quad (23)$$

may be capable of ensuring a monotonic sequence of Euclidean error norms if the learning gain $\beta > 0$ is sufficiently small.

In the following sections, an analysis is undertaken of the choice of learning gain $\beta$. It generates an estimate of an appropriate range in both the case of zero and non-zero modelling errors. Initially, the analysis is in the form of matrix inequalities. Subsequently these will be converted into conditions in the frequency domain.

V. THE GRADIENT ALGORITHM: THE CASE OF NO MODELLING ERROR

A simple calculation reveals that the gradient-based ILC algorithm evolves from its initial error $e_0$ as follows

$$e_{k+1} = (I - \beta GG^T) e_k, \quad k \geq 0 \quad (24)$$
Noting that $\beta > 0$ by assumption and that
\begin{equation}
||e_{k+1}||^2 = ||e_k||^2 - 2\beta e_k^T G G^T e_k + \beta^2 e_k^T G G^T G G^T e_k
\tag{25}
\end{equation}
it follows that, as $G$ is nonsingular by construction,

\textbf{Theorem:} Suppose that $\beta > 0$. A necessary and suffi-
cient condition for the gradient-based ILC algorithm to have the
monotonicity and convergence properties

1) $||e_{k+1}|| < ||e_k||$, $\forall k \geq 0$ $\forall e_0 \in \mathbb{R}^{N+1-k^*}$
2) $\lim_{k \to \infty} e_k = 0$ $\forall e_0 \in \mathbb{R}^{N+1-k^*}$
in some range $0 < \beta < \beta'$ is that
\begin{equation}
2I > \beta G^T G > 0
\tag{26}
\end{equation}

\textbf{Proof:} $2I > \beta G^T G$ implies the existence of $\epsilon > 0$ such that $\beta G^T G - 2I < -\epsilon I$. Monotonicity follows from the preceding discussion. To prove convergence to zero, note that
\begin{equation}
||e_{k+1}||^2 \leq ||e_k||^2(1 - \beta \epsilon) \forall k \geq 0
\tag{27}
\end{equation}
Hence $||e_k||$ goes to zero faster than $(1 - \beta \epsilon)^{\frac{k}{T}}$. \square

\textbf{Corollary:} Under the conditions of the theorem above, monotone convergence to zero is achieved if $0 < \beta \sigma(G) < 2$ where $\sigma(G)$ is the largest singular value of $G$.

\section{VI. THE GRADIENT ALGORITHM: ROBUST MONOTONE CONVERGENCE CONDITIONS}

Now let $G(z)$ and $G_0(z)$ be transfer functions of the plant and nominal model respectively. The relative degree of the model $G_0$ is denoted $k^*$ and the lifted representations (and associated input and output supervectors) are based on this parameter. To ensure that the matrix representations of plant, nominal model and multiplicative perturbations are causal, it is assumed that the relative degree of the plant is $\geq$ that of the nominal model.

The gradient-based ILC algorithm naturally uses the approximation
\begin{equation}
u_{k+1} = u_k + \beta G_0^T e_k
\tag{28}
\end{equation}
where $G_0$ is the lifted matrix representation of a model of $G_0(z)$. The error evolution equation becomes
\begin{equation}
e_{k+1} = (I - \beta G_0 G_0^T) e_k
\tag{29}
\end{equation}
Suppose now that plant and model are related by the expression $G(z) = G_0(z) U(z)$ and $U(z)$ is assumed to be proper and stable. It follows that, if $U(z)$ has a matrix representation $U_e$ (without lifting), then
\begin{equation}
G = G_0 U_e = U_e G_0
\tag{30}
\end{equation}
Note that $\beta > 0$ by assumption and that
\begin{equation}
||e_{k+1}||^2 = ||e_k||^2 - 2\beta e_k^T G_0 U_e G_0^T e_k + \beta^2 e_k^T G_0 G_0^T U_e U_e G_0 G_0^T e_k
\tag{25}
\end{equation}
it follows that, as $G$ is nonsingular by construction,

\textbf{Theorem (Robust Monotone Convergence):} The gradient-based ILC algorithm is robust monotone convergent in the presence of the multiplicative modelling error $U(z)$ if, and only if,
\begin{equation}
U_e + U_e^T > \beta G_0^T U_e G_0 > 0
\tag{32}
\end{equation}

\textbf{Proof:} Monotonicity follows trivially from the above noting that $G_0$ is nonsingular by construction. The proof of convergence to zero follows in a similar way to the previous case. \square

\textbf{Corollary:} A necessary condition for monotone robust convergence is that the modelling error matrix representation $U_e$ is positive definite in the sense that $U_e + U_e^T$ is positive definite.

\textbf{Proof:} The proof follows trivially from the observation $\beta G_0^T U_e G_0 > 0$. \square

\textbf{Note:} The case of no modelling error is retrieved by choosing $U = I$ in the above.

In the next section, useful frequency domain conditions are provided to check the matrix inequalities.

\section{VII. ROBUSTNESS: FREQUENCY DOMAIN CONDITIONS}

In this section the matrix inequalities of the previous sections are converted into sufficient conditions for robust monotone convergence in terms of the transfer functions of the system, model and uncertainty. Frequency domain conditions are more easily checked and throw more light on to the benefits and issues facing the application of the gradient-based algorithm.

The approach taken is based on the analysis of matrix inequalities in $\mathbb{R}^{l \times l}$ of the form
\begin{equation}
H_1^T H_1 < H_2 + H_2^T
\tag{33}
\end{equation}
where both $H_1 \in \mathcal{L}_1$ and $H_2 \in \mathcal{L}_1$ are matrix representations of SISO linear time-invariant systems $H_1(z)$ and $H_2(z)$ on the resultant interval $0 \leq j \leq l - 1$.

The development of frequency domain conditions is based on examining dynamics on the infinite half interval $[0, \infty)$. Complex integration, positivity and causality then provide the necessary connections.

Let $e = [e(0), e(1), \ldots, e(l - 1)]^T$ be a time series of length $l$ and interpret $H_1 e$ as the restriction (to $0 \leq j \leq l - 1$) of the response of $H_1(z)$ (on $[0, \infty)$) to the input with $Z$-transform $e(z) = \sum_{j=0}^{l-1} e(j) z^{-j}$ i.e. to an infinite sequence $\tilde{e}$ consisting of the $l$ elements of $e$
followed by zeros. Using the fact that the mean square error on a finite interval is always less than or equal to that on the infinite interval, Parseval’s Theorem then gives
\[ e^T H_1^T e \leq \frac{1}{2\pi i} \oint_{\text{unitcircle}} |H_1(z)|^2 |e(z)|^2 \frac{dz}{z} \]
(34)
A simple calculation then indicates that
\[ \|H^{-1}\|_1 \leq \sigma(H_1) \leq \overline{\sigma}(H_1) \leq \|H\|_\infty \] (35)
where \( \sigma(H) \) and \( \overline{\sigma}(H) \) denote the smallest and largest singular values of a matrix \( H \in L_1 \) respectively and \( \|H\|_\infty \) denotes the \( H_\infty \) norm of the associated transfer function \( H(z) \) on the region \( |z| \geq 1 \).

In a similar manner, \( e^T H_2 e \) is the inner product in \( l_2 \) (the space of square summable infinite sequences) of \( e \) with the response of \( H_2(z) \) to \( e \) and hence the exact expression follows from elementary complex variable theory
\[ e^T (H_2^T + H_2) e = \frac{1}{2\pi i} \oint [H_2(z) + H_2(z^{-1})] |e(z)|^2 \frac{dz}{z} \] (36)

The matrix inequality describing robust monotone convergence hence is satisfied if, for all choices of \( e \),
\[ \frac{1}{2\pi i} \oint |H_1(z)|^2 |e(z)|^2 \frac{dz}{z} \leq \frac{1}{2\pi i} \oint [H_2(z) + H_2(z^{-1})] |e(z)|^2 \frac{dz}{z} \] (37)

It is now possible to state the following theorem:

**Theorem (Robust Monotone Convergence):** The gradient-based algorithm using the nominal model \( G_0(z) \) is robust monotone convergent in the presence of the multiplicative modelling error with transfer function \( U(z) \) if (a sufficient condition)
\[ |\beta| - |G_0(z)|^2 |U(z)| < \frac{1}{\beta} \quad \forall z \in \{|z| = 1\} \] (38)

**Proof:** The discussion preceding this result and the matrix inequality condition of the previous section indicates that a sufficient condition for robust monotone convergence is that
\[ U(z) + U(z^{-1}) > \beta |G_0(z)|^2 U(z) \quad \forall |z| = 1 \] (39)
Noting that \( G_0 \) can be replaced by \( G_0^* \) on \( |z| = 1 \), multiplying by \( \beta |G_0(z)|^2 \) and rearranging yields the required result. \( \square \)

Note: Simple calculations indicate that the frequency domain conditions have a simple and easily checked graphical interpretation, namely that:

The plot of the frequency response function \( |G_0(z)|^2 U(z) \) on the unit circle \( |z| = 1 \) lies in the interior of the circle of centre \( \frac{1}{\beta} \) and radius \( \frac{1}{\beta^2} \).

Recent work by the authors [3] using the inverse model algorithm produced the condition:
\[ |\frac{1}{\beta} - U(z)| < \frac{1}{\beta} \quad \forall z \in \{|z| = 1\} \] (40)
At its simplest level, the difference between the two results is the replacement of \( U \) by \( |G_0|^2 U \). With this in mind, the use of the gradient-based algorithm can be seen to have the following properties as compared with the inverse-model algorithm:

1) Both approaches require a strictly positive real \( U(z) \). This condition is connected with the monotonicity property of the mean square error and it is expected, as with the inverse-model-based approach, that violation may lead to lack of convergence/instability. Another possibility is that asymptotic convergence may be retained but it may also be associated with error norm sequences that can increase from trial to trial.

2) In both cases, the positive real requirement on \( U(z) \) will tend to require that \( G \) and \( G_0 \) have the same relative degree.

3) The gradient-based algorithm will however reduce performance limitations due to the effect of high frequency errors such as high frequency resonances in \( G \) not modelled in \( G_0 \). In such circumstances \( U(z) \) will tend to take large gain values at frequencies close to these resonances. This will then require the use of small values of learning gain \( \beta \) to satisfy the monotone convergence criterion for the inverse model algorithm. This does not occur for the gradient-based algorithm because, in practice, \( G \) is typically a low pass filter and hence both \( G(z) \) and \( G_0 \) will be small at high frequencies. The magnitude of \( |G_0|^2 U \) will then be substantially reduced (as compared with \( U \)) and permit increased learning gains leading to improved convergence rates.

4) In contrast with the beneficial high frequency effects of the gradient-based algorithm, it is possible that it could reduce performance if \( G \) (and hence \( G_0 \)) has a substantial resonance peak within its bandwidth. A similar argument to the above suggests that the learning gains permitted will be reduced (as compared with the inverse model algorithm). As a consequence, it is desirable for a feedback control to be incorporated into the plant (and hence \( G \)) before the ILC analysis is undertaken.

5) The above analysis has considered a specific uncertainty \( U \). It can easily be extended to cover sets of multiplicative uncertainties such as any subset of all proper multiplicative uncertainties satisfying an inequality of the form
\[ |\frac{1}{\beta^2} - |G_0(z)|^2 U(z)| < \frac{1}{\beta^2} \quad \forall z \in \{|z| = 1\} \] (41)
for some choice of parameter \( \beta^* \).

In conclusion, the analysis of monotone convergence...
has been seen to have elegant solutions in terms of inequalities between matrix representations of the plant and associated models. These inequalities can be converted into simple frequency domain (sufficient) conditions that indicate that the gradient-based approach has real potential for both performance and robustness.

Finally, note that, when $U(z) \equiv 1$ and hence $U_e = I$, the above results produce conditions for monotone convergence when there is no plant-model mismatch.

**Corollary:** Under the conditions of the theorem above, monotone convergence to zero is achieved in the absence of modelling errors if $0 < \beta ||G||_\infty^2 < 2$ where $||G||_\infty = \sup_{|z|=1} |G(z)|$ is the familiar $H_\infty$ norm of $G$ on $\{z: |z| \geq 1\}$.

**Proof:** Setting $U = I$, $U(z) \equiv 1$ and $G_0(z) \equiv G(z)$ in the previous result, monotone convergence follows if $\frac{1}{\beta} - |G(z)|^2 < \frac{1}{\beta} \forall z \in \{z: |z| = 1\}$. The result follows from simple complex algebra. □

In particular, the result shows that, in the absence of mismatch, monotone convergence is not dependent on the phase characteristics of the plant (an observation that links these results to the continuous-time methodology described in [8]).

**VIII. conclusions**

The paper has provided a complete analysis of the robust monotone convergence of a gradient-based Iterative Learning Control algorithm in terms of necessary and sufficient matrix inequalities and frequency domain conditions that can be easily checked in terms of plant model and modelling error transfer functions. A complete analysis of these models is provided which demonstrates that the relative degree of the plant and model are crucial parameters in the analysis of ILC dynamics and feedforward learning laws. In addition, they clearly show that the use of the "non-causal" gradient operator can be implemented using a plant model and time reversal operations.

The work parallels that published by the authors in a recent paper [3] on inverse-model-based ILC. A comparison with those results indicates that, whereas both approaches require that the multiplicative modelling error has positivity properties, the gradient approach offers considerable benefits for robustness, particularly in the presence of high frequency modelling errors.

**References**


