Multivariable nonparametric learning: A robust iterative inversion-based control approach

Robin de Rozario | Tom Oomen

Control Systems Technology, Department of Mechanical Engineering, University of Technology Eindhoven, Eindhoven, The Netherlands

Correspondence
Tom Oomen, Control Systems Technology, Department of Mechanical Engineering, University of Technology Eindhoven, Eindhoven, The Netherlands. Email: t.a.e.oomen@tue.nl

Funding information
Nederlandse Organisatie voor Wetenschappelijk Onderzoek, Grant/Award Number: 15698

Summary
Learning control enables significant performance improvement for systems by utilizing past data. Typical design methods aim to achieve fast convergence by using prior system knowledge in the form of a parametric model. To ensure that the learning process converges in the presence of model uncertainties, it is essential that robustness is appropriately introduced, which is particularly challenging for multivariable systems. The aim of the present article is to develop an optimization-based design framework for fast and robust learning control for multivariable systems. This is achieved by connecting robust control and nonparametric frequency response function identification, which results in a design approach that enables the synthesis of learning and robustness parameters on a frequency-by-frequency basis. Application to a multivariable benchmark motion system confirms the potential of the developed framework.

KEYWORDS
frequency response methods, learning control, multivariable control systems, robust control, System identification

1 INTRODUCTION

Learning control methods can increase the performance of systems by using past operational data. Iterative learning control (ILC),1,2 repetitive control (RC),3 and iterative inversion-based control (IIC)4 have been developed to increase the tracking performance of systems that start each task under identical conditions or that operate continuously in a periodic fashion.5,6 Examples of high-precision applications whose performance has been enhanced through learning control include nanopositioning devices,7,8 additive manufacturing machines,9,10 and industrial printing systems.11,12

Convergence, usually in terms of signal norms, is an essential aspect of learning and is typically subject to the following requirements.1,13§5.1.1

(R1) The learning process converges in a way that is free of poor learning transients and converges relatively fast.
(R2) The performance upon convergence is improved.

These requirements should hold in the presence of uncertain dynamics and unknown disturbances,14 since learning control is typically employed to increase the performance in the presence of these unmodeled phenomena.
In ILC, these requirements are typically addressed by designing a learning and robustness filter as in the following archetypal approach. Requirement R1 satisfied if the input $u_t$ converges monotonically to a fixed point $u_{\infty}$, which is achieved if there exists a $\kappa \in (0, 1]$ such that
\[
\|u_{i+1} - u_{\infty}\| < \kappa \|u_i - u_{\infty}\|, \quad \forall i \geq 0.
\]

The monotonic convergence condition ensures the absence of poor learning transients and $\kappa$ determines the rate of convergence in the signal norm $\|\cdot\|$. A small $\kappa$ can be achieved by designing a learning filter that closely approximates the inverse plant dynamics. A typical approach is to first estimate a parametric plant model and subsequently invert this model using dedicated techniques in case the model is nonminimum phase. Both steps commonly introduce errors which may hamper convergence at specific frequencies. Establishing convergence under model uncertainties can be achieved by incorporating a robustness filter. Introducing this filter changes the fixed point $u_{\infty}$, which in turn affects the asymptotic performance such that perfect tracking can no longer be achieved. Satisfying requirement R2 for a large range of frequencies thus requires careful design of the robustness filter.

Methods to design robustness filters are intuitive and well developed for single-input single-output (SISO) systems, whereas designing such filters is substantially more involved for multiinput multioutput (MIMO) systems. For SISO systems, the monotonic convergence condition establishes a bound on the magnitude of the inversion error transfer function, which enables an intuitive manual tuning approach. For MIMO systems, the same condition establishes a bound on the maximum singular value of the matrix transfer function, which makes it challenging to manually design a robustness filter in a way that addresses the performance-robustness trade-off in a nonconservative manner. Robust control methods offer a formal framework for MIMO robustness filter design but require accurate parametric models of the plant and of the uncertainties, which may be costly and challenging to obtain. In Reference 29, the modeling requirements are reduced by developing a framework to design diagonal robustness filters based on convergence criteria that can be evaluated using nonparametric frequency response function (FRF) models, yet the performance and convergence speed are not explicitly taking into consideration. Existing results indicate a trade-off between modeling-effort and performance, and the aim of this article is to go beyond this trade-off by enabling performance optimization using nonparametric FRF models.

ILC avoids the use of parametric models and instead relies on nonparametric FRF models. In IIC, input updating is performed by explicitly transforming the signals to the frequency domain, such that the filtering operations become multiplications between the signal spectra and the FRF representations of the learning and robustness filters, which are complex coefficient matrices. The learning coefficients are typically obtained by inverting an FRF estimate of the plant, and the robustness coefficients are tuned according to an uncertainty model of that FRF. These FRF models are generally considered to be accurate and require less user intervention to obtain compared to parametric models, see for example for a systematic procedure. However, the preexisting IIC approach for MIMO systems does not include a robustness matrix and leaves the associated design freedom unexplored.

Although important developments have been made to enable high-performance control through fast and robust learning, presently available techniques do not scale well to complex large-scale multivariable systems. This article aims to overcome the limitations that are imposed by the requirement for parametric multivariable plant and uncertainties models in learning control. This is achieved by developing an FRF-based IIC framework for fast and robust learning. The main challenge in this framework is the optimal design of the frequency-dependent robustness and learning matrices $Q(\omega_k)$ and $L(\omega_k)$.

The main contribution of this article is a design framework for these matrices, which is developed in two parts. In the first part, a deterministic robust control approach is taken to optimally design $Q(\omega_k)$ and $L(\omega_k)$, based on a nominal FRF model with additive norm-bounded uncertainties. In the second part, an approach is developed to identify the required FRF model set in the presence of stochastic disturbances. The combination of these approaches results in a learning method that meets requirements R1 and R2 with a probability that is chosen by the user. Specifically, the outline and contributions of this article are as follows. In Section 2, the iterative tracking control problem is introduced and a MIMO FRF-based IIC method is proposed as a solution to this problem. To design the $Q(\omega_k)$ and $L(\omega_k)$ matrices that play a central role in this approach, a deterministic convergence and performance analysis is developed in Section 3. These results enable an optimization-based approach to $Q(\omega_k)$ and $L(\omega_k)$ design, resulting in the following contribution.
(C1) The $Q(\omega_k)$-matrix is optimized through a semidefinite program (SDP) to guarantee robust convergence and to achieve maximum nominal performance, for the case where $L(\omega_k)$ is taken as the inverse of the system FRF (Section 5).

This approach is the main contribution of this article, and the remaining sections aim to motivate and support the particular design choices. The choice to take $L(\omega_k)$ equal to the inverse of the system FRF is motivated by the second contribution.

(C2) It is shown that if $L(\omega_k)$ equals the inverse system FRF, then the robust convergence rate is optimal (Section 6).

Furthermore, the choice to optimize $Q(\omega_k)$ for nominal performance, instead of robust performance, is motivated by the third contribution.

(C3) It is shown that optimizing $Q(\omega_k)$ for robust performance can be prohibitively conservative, since it results either in perfect tracking, that is, $Q(\omega_k) = I$, or in no learning at all, that is, $Q(\omega_k) = 0$ (Section 7).

The following contributions support the application the developed design method in practice.

(C4) An identification method is developed to estimate the nominal model and the set of norm-bounded uncertainties (Section 8) and the combination with the $L(\omega_k)$ and $Q(\omega_k)$ design approach is formulated as a stepwise procedure (Section 9).

(C5) The potential of the developed framework is confirmed by application to a benchmark motion system (Section 10).

Contributions C1 to C3 take a deterministic robust control approach to learning as in References 26, 32. Contribution C1 provides a means to optimally design $Q(\omega_k)$ in a way that is numerically efficient, and that can readily be parallelized. The results developed in C2 and C3 establish conditions under which the design choices articulated in C1 are optimal. Contribution C4 establishes a connection between stochastic FRF identification and deterministic robust control, which closely connects to References 33, 34. In contribution C5, it is illustrated that the performance can be significantly improved by using fully parametrized robustness matrices instead of diagonal matrices, thereby considerably extending the results in References 29, 31.

In relation to the initial results presented in Reference 35, contribution C1 is established in a rigorous mathematical manner, and C3 is new. Furthermore, contributions C2 and C3 are extended by providing detailed proofs, and C4 is enhanced by explicitly treating real uncertainties that arise at specific frequencies, whereas C5 is extended with additional simulation results.

1.1 | Preliminaries

The sets of real numbers, complex numbers, and integers are denoted by $\mathbb{R}$, $\mathbb{C}$, and $\mathbb{Z}$. For a matrix $A \in \mathbb{C}^{n \times m}$, $[A]_{ij}$ denotes the $ij$th element, $A^H$ is the conjugate transpose, and $\lambda(A)$ and $\sigma(A)$ represent the largest eigenvalue and largest singular value. For $a \in \mathbb{C}^n$, $\|a\|$ denotes the 2-norm, whereas for $A \in \mathbb{C}^{n \times m}$, $\|A\|_{2,2}$ denotes the induced 2,2-norm. The discrete Fourier transform (DFT) and its inverse are defined as $S(\omega_k) \triangleq \frac{1}{\sqrt{N}} \sum_{t=0}^{N-1} s(t) e^{-j\omega_k t}$, $s(t) \triangleq \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} S(\omega_k) e^{j\omega_k t}$, $\omega_k \in \Omega_N$, $\Omega_N \triangleq \left\{ \omega_k \in \mathbb{R} \mid \omega_k = \frac{2\pi k}{N}, \ k = 0, \ldots, N-1 \right\}$, MIMO LTI DT systems $G_o : u \mapsto y$ are considered, which are represented by their transfer function matrices $G_o(z) \in \mathbb{C}^{n \times m}$, $z \in \mathbb{C}$, and $G_o(e^{\omega_k})$ is the FRF of $G_o$ on the DFT grid $\Omega_N$.

2 | PROBLEM FORMULATION

In this section, the iterative tracking problem is introduced and a MIMO FRF-based IIC approach is presented. The first challenge considered this article is to design the associated robustness and learning matrices to ensure that requirements R1 and R2 are satisfied in the presence of uncertainties in the FRF model. The second challenge is to identify such FRF models from data.
2.1 The iterative tracking problem

The control objective is to achieve accurate tracking and disturbance rejection for the following class of systems.

**Assumption 1.** The system \( G_o \) is stable, square, and its FRF is invertible on the DFT grid, that is, \( \det(G_o(e^{\omega_k})) \neq 0 \), \( \forall \omega_k \in \Omega_N \).

This assumption is standard in system inversion and ensures that the response of \( G_o \) is bounded, the transient response decays exponentially, and the inverse response is well defined. When the system is nonsquare, squaring down\(^{37}\) or optimization-based methods\(^{38}\) can be used to obtain a square system. Furthermore, it is assumed that the system performs identical tasks, while the output is corrupted by trial-varying disturbances and trial-invariant disturbances.

**Assumption 2.** The output as measured during the \( i \)th trial is disturbed by \( v_i(t) \) and \( d(t) \), such that \( y_i(t) = G_o(z)u_i(t) + d(t) + v_i(t) \), as shown in Figure 1. Here, \( d(t) \) represents the disturbances that are identical for each trial and \( v_i(t) \) represents the trial-varying disturbances. Furthermore, \( r(t) \) is a trial-invariant bounded reference signal.

Since \( \|X(\omega_k)\|_2 = \|x(t)\|_2 \) by Parseval’s theorem,\(^{39}\) the iterative tracking problem can be formulated in the frequency domain, aided by the following definition of monotonic convergence for DFT vectors.

**Definition 1.** A sequence \( \{X_i\}_{i=0}^{\infty}, X \in \mathbb{C}^n \), is said to convergence monotonically in the 2-norm to a fixed point \( X_\infty \in \mathbb{C}^n \) if there exists a \( \kappa \in (0,1] \) such that \( \|X_{i+1} - X_\infty\|_2 < \kappa \|X_i - X_\infty\|_2, \forall i \), where \( \kappa \) is referred to as the convergence rate.

**Problem 1.** For each relevant frequency \( \omega_k \), determine a sequence of inputs \( \{U_i(\omega_k)\}_{i=0}^{\infty} \) such that

1. \( \{U_i(\omega_k)\}_{i=0}^{\infty} \) converges monotonically in the 2-norm to a fixed point \( U_\infty(\omega_k) \in \mathbb{C}^n \) at a given rate \( \kappa(\omega_k) \in (0,1] \), and
2. \( \lim_{i \to \infty} \|E_i(\omega_k)\|_2 \) is minimal, where \( E_i(\omega_k) \equiv R(\omega_k) - Y_i(\omega_k) \).

Note that Problem 1.(a) and 1.(b) directly reflect requirements R1 and R2.

2.2 Robust finite-time IIC

IIC algorithms, including References 4,6,31, provide an approach to Problem 1 by using the following update law

\[
U_{i+1}(\omega_k) = Q(\omega_k)U_i(\omega_k) + L(\omega_k)E_i(\omega_k),
\]

where \( L(\omega_k), Q(\omega_k) \in \mathbb{C}^{nxn} \) are the learning matrix and robustness matrix, respectively. Note that in (1), \( U_{i+1}(\omega_k) \) is determined by multiplication of the DFT coefficient vectors \( U_i(\omega_k) \) and \( E_i(\omega_k) \) by the matrices \( L(\omega_k) \) and \( Q(\omega_k) \) for each frequency \( \omega_k \). This enables a frequency-by-frequency approach to design \( L(\omega_k) \) and \( Q(\omega_k) \). By assuming that the transient response is negligible, a decoupled approach can be formulated that significantly reduces the design complexity and enables parallelization. To see this, consider the DFT of \( y_i(t) \),

\[
Y_i(\omega_k) = G_o(e^{\omega_k})U_i(\omega_k) + T_{U_i}(\omega_k) + V_i(\omega_k) + D(\omega_k).
\]

Here, \( T_{U_i}(\omega_k) \) represents the transient response, which vanishes exponentially for stable \( G_o \).\(^{40}\) In steady state, that is, for \( T_{U_i}(\omega_k) = 0 \), the relation between \( Y_i(\omega_k) \) and \( U_i(\omega_k) \) is decoupled for each \( \omega_k \), such that \( Q(\omega_k) \) and \( L(\omega_k) \) can be designed for each \( \omega_k \) individually. Practically, negligible transients can be achieved by ensuring that the task is long enough or by using a periodic implementation, as is presented in Reference 6. The following algorithm encompasses many IIC approaches including References 4,12.
Algorithm 1. Robust finite-time IIC

Select an initial input $u_0(t)$.

1. Apply the input $u_i(t)$ and record the tracking error $e_i(t) = r(t) - y_i(t)$ when $y_i$ is in steady state.
2. (a) Obtain $U_i(\omega_k)$ and $E_i(\omega_k)$ by applying the DFT.
   (b) Update $U_i(\omega_k)\psi(\omega_k)$ by means of (1).
   (c) Obtain $u_{i+1}(t)$ by applying the IDFT.
3. Set $i ← i + 1$ and repeat from step (1.) until $e_i$ has converged.

The main objective of this article is to solve Problem 1 using Algorithm 1 by suitable design of $Q(\omega_k)$ and $L(\omega_k)$. This is achieved using a two-step approach where (i) $L(\omega_k)$ and $Q(\omega_k)$ are optimized based on a deterministic FRF model set containing $G_0(e^{j\omega})$ and (ii) the deterministic model set is estimated from probabilistic data.

2.3 A deterministic FRF-based design approach for $L$ and $Q$

The frequency-by-frequency updating approach in IIC enables the design of $L(\omega_k)$ and $Q(\omega_k)$ based on an FRF model of $G_n$. This is in contrast to filter-based ILC approaches which typically require a MIMO parametric plant model instead. To facilitate the FRF-based design approach, it is assumed that the following FRF model with additive uncertainty is given, where the argument $\omega_k$ is omitted in the forthcoming sections to facilitate a transparent exposition.

Definition 2. Let $G(\Delta) ∈ C^{n×n}$ be an FRF model set

$$G(\Delta) = G_n + W_y \Delta W_u, \quad \Delta ∈ \Delta, \quad \Delta = \{\Delta ∈ C^{n×n} | \sigma(\Delta) ≤ \gamma\}. \quad (3)$$

where $W_y, W_u ∈ C^{n×n}$ are weighting matrices.

Here, $G_n$ is referred to as the nominal model and $\Delta$ is a set of unstructured, norm-bounded perturbations.

Assumption 3. $G_n(\omega_k)$ is invertible $\forall \omega_k ∈ Ω_N$.

This assumption ensures that $G_n$ corresponds to an FRF of a system whose inverse response is well defined. In the next sections, $L(\omega_k)$ and $Q(\omega_k)$ are designed for optimal performance, and to ensure convergence for all possible models, that is, $\forall \Delta ∈ \Delta$.

2.4 Identifying a deterministic FRF model set from probabilistic data

In Section 8, a method is developed to estimate the model set (3). It is assumed that the disturbance $v(t)$ is Gaussian during the identification experiments, and an FRF matrix estimator $\hat{G}$ is formulated that inherits the Gaussian probabilistic properties. By explicit analysis of the probability, an approach is developed to select the matrices $W_u, W_y$, and the uncertainty bound $\gamma$, such that the true FRF $G_n$ is captured by the model $G(\Delta)$ with probability $\alpha$. This approach is illustrated for the SISO case in Figure 2. Let the estimate be given by $\hat{G} = G_n + V$, where $V$ is a complex Gaussian variable. Clearly, for $G_n = \hat{G}$ and $W_n = W_y = 1$, $G_n = G(\Delta_n) = V$ if $\Delta_n = V$. This requires that $\gamma$ is large enough such that $\Delta_n = V ∈ \Delta$.

The left plot shows the probability density function (PDF) of $V$, where $V$ is circularly complex Gaussian distributed, and the right plot shows the PDF of $|V|$, denoted by $f(|V|)$, which is a Rayleigh distribution. If the true system should be captured with probability $\alpha$, then $\gamma$ must be taken such that $\text{Prob}(|V| ≤ \gamma) = \int_0^\gamma f(|V|)d|V| = \alpha$, which corresponds to the blue area in Figure 2. This approach enables the construction of norm-bounded model sets $\Delta$, and the outspoken unimodal PDF of $|V|$ motivates an $L$ and $Q$ design approach that optimizes the performance associated to the nominal model $G_n$, while ensuring convergence $\forall \Delta ∈ \Delta$. 
In this section, a robust MIMO IIC approach is presented as a solution to the iterative tracking problem, and the two-step $Q$ and $L$ design approach is outlined. This approach builds on the analysis results regarding robust convergence and asymptotic performance that are developed in the next section.

3 | ROBUST CONVERGENCE AND ASYMPTOTIC PERFORMANCE ANALYSIS

In this section, the convergence properties of the update law (1) are analyzed. The analysis provides fundamental relations that enable the deterministic optimization results developed in the subsequent sections. A deterministic robustness analysis is facilitated by assuming that the true model is captured by the model set.

Assumption 4. Let $\mathcal{G}(\Delta)$ be given by Definition 2. There exists a $\Delta_0 \in \Delta$ such that $\mathcal{G}(\Delta_0) = G_0$.

The following assumption enables a deterministic analysis of convergence.\textsuperscript{6,12} For the case with stochastic disturbances during the learning procedure, see Reference 21.

Assumption 5. In (2), $T_{U_i} = 0$ and $V_i = 0$.

This assumption formalizes that the transient response and the trial-varying disturbances are zero during learning. Under these assumptions, it follows readily that Problem 1.(a) is solved if (1) is robustly monotonically convergent, that is, if (1) yields monotonically converging sequences $\{U_i\}_{i=0}^\infty \forall \Delta \in \Delta$. The following lemma presents a sufficient condition for robust monotonic convergence and it is emphasized that $\Delta$ is independent of $i$.

Lemma 1. Consider (1) under Assumptions 1 to 5 and let $\mathcal{G}(\Delta)$ be given by Definition 2. If for a given $\kappa \in (0, 1]$ it holds that

$$\|H(L, Q, \Delta)\|_2 < \kappa \quad \forall \Delta \in \Delta,$$

then for any $R, D \in \mathbb{C}^n$, the sequences $\{U_i\}_{i=0}^\infty$ converge monotonically to $U_\infty$ in the 2-norm, and the sequences $\{E_i\}_{i=0}^\infty$ converge to $E_\infty$, where

$$U_\infty = (I - Q(I - LG_\Delta))^{-1}QL(R - D), \quad \text{and} \quad E_\infty = S_\delta(L, Q)(R - D),$$

with

$$S_\delta(L, Q) \triangleq I - G_\delta(I - Q(I - LG_\Delta))^{-1}QL.$$

A proof of this lemma is provided in Appendix A. This lemma shows that the matrix $S_\delta(L, Q)$ determines the asymptotic performance and the matrix $H(L, Q, \Delta_\ast)$ determines if the IIC input converges monotonically. This result enables the formulation of a constrained optimization problem to design $L$ and $Q$.

4 | OPTIMIZATION-BASED $L$ AND $Q$ DESIGN

Problem 1, as formulated in Section 2, can be reformulated as an optimization problem in $L$ and $Q$. In particular, Problem 1.(a) is solved if $L$ and $Q$ are such that (4) holds for a desired convergence rate $\kappa \in (0, 1]$. Problem 1.(b) is solved
by minimizing $\|S_0\|_{2,2}$, with $S_0$ as given by (7). However, $\|S_0\|_{2,2}$ cannot be minimized directly, since $G_0$ is unknown. Instead, the model set $G(\Delta)$ is used to determine a model set of $S_0$, \[ S(L, Q, \Delta) \triangleq I - G(\Delta)(I - Q(I - LG(\Delta)))^{-1}QL. \] (8)

Note that if the convergence condition is satisfied, then (8) is well-posed, that is, the inverse in (8) exists.

**Lemma 2.** If (4) holds for any $\kappa \in (0, 1]$, then $\det(I - Q(I - LG(\Delta))) \neq 0$.

A proof to this lemma is provided in Appendix B. The uncertain matrices $H(L, Q, \Delta)$ and $S(L, Q, \Delta)$ naturally enable an optimization-based $L$ and $Q$ design approach. More specifically, the performance can be optimized with respect to the nominal model, or with respect to the worst-case model, under the constraint of robust convergence. Particularly, consider the following general optimization problems for a given $\kappa \in (0, 1]$.

- Robust convergence and nominal performance
  \[ \inf_{L,Q \in C_{v_{\kappa}}} \|S(L,Q,0)\|_{2,2}, \text{ subject to: } \|H(L,Q,\Delta)\|_{2,2} < \kappa, \quad \forall \Delta \in \Delta. \] (9)

- Robust convergence and robust performance
  \[ \inf_{L,Q \in C_{v_{\kappa}}} \sup_{\Delta \in \Delta} \|S(L,Q,\Delta)\|_{2,2}, \text{ subject to: } \|H(L,Q,\Delta)\|_{2,2} < \kappa, \quad \forall \Delta \in \Delta. \] (10)

These optimization problems are generally nonconvex since $H(L,Q,\Delta)$ and $S(L,Q,\Delta)$ are nonlinear in the matrix-parameters $L$ and $Q$. This makes determination of the global optimizer practically impossible, and an alternative approach is proposed.

### 4.1 The proposed optimization approach

The following approach is proposed to obtain a numerically efficient design method. Consider (9) and fix $L$ as $L = G_n^{-1}$, which results in the following problem.

**Problem 2.** Let $G(\Delta)$ be given by Definition 2. For a given $\kappa \in (0, 1]$, determine

\[ Q^* = \arg \inf_{Q \in C_{v_{\kappa}}} \|S(G_n^{-1}, Q, 0)\|_{2,2}, \text{ subject to: } \|H(G_n^{-1}, Q, \Delta)\|_{2,2} < \kappa \quad \forall \Delta \in \Delta, \] (11)

\[ S(G_n^{-1}, Q, 0) = G_n(I - Q)G_n^{-1}, \quad H(G_n^{-1}, Q, \Delta) = QG_n^{-1}W_y\Delta W_u. \] (12)

This approach is motivated in the remainder of the article in 5-fold as follows.

1. It is shown in Section 5 that Problem 2 can be reformulated as an SDP, whose global optimum can be obtained in a numerically efficient manner.
2. It is shown in Section 6 that $L = G_n^{-1}$ optimizes the robust convergence rate $\kappa$ in case of perfect tracking, that is, for $Q = I$.
3. It is shown in Section 7 that optimizing $Q$ with respect to the worst-case performance can be relatively conservative since (10) results in either $Q = I$ or $Q = 0$, when $L = G_n^{-1}$.
4. It is shown in Section 7 that taking $\kappa < 1$ can be used to bound the worst-case performance, which confirmed by demonstration in Section 10.
5. It is similarly demonstrated in Section 10 that the nominal performance closely approximates the true performance, such that proposed approach provides an effective way to solve the iterative tracking problem presented in Section 2.

Furthermore, the framework to identify the required FRF model set is developed in Section 8, and the resulting overall design framework is summarized in Section 9. It is remarked that Sections 6 and 7 provide the theoretical motivations for the proposed design approach and can be skipped by the practically oriented reader.
5 | SYNTHESIZING Q FOR OPTIMAL NOMINAL PERFORMANCE

Motivated by the results summarized in the previous section, Q is optimized for nominal performance and robust convergence for \( L = G_n^{-1} \). More specifically, Problem 2 is reformulated as an SDP, which constitutes contribution C1.

The following theorem enables a reformulation of Problem 2 as an SDP.

**Theorem 1.** Let \( \mathcal{G}(\Delta) \) be given by Definition 2 and satisfy Assumptions 3 and 6. The optimization problem specified by (11) and (12) in Problem 2 is identical to

\[
Q^* = \arg \inf_{Q \in \mathbb{C}^{n \times n}, \quad x, \beta \in \mathbb{R}} \beta, \quad \text{subject to:}
\]

\[
\begin{bmatrix}
-\beta I & (G_n(I - Q)G_n^{-1})^H \\
G_n(I - Q)G_n^{-1} & -I
\end{bmatrix} < 0, \quad \beta > 0, \tag{13a}
\]

\[
\begin{bmatrix}
xW_u^H W_u - I & 0 & 0 \\
0 & \frac{-xI}{\kappa} & \frac{(\frac{x}{\kappa} QG_n^{-1} W_y)^H}{I}
\end{bmatrix} < 0, \quad x > 0. \tag{13c}
\]

This theorem enables a numerically efficient way to design \( Q \), since (13) can be solved by using various algorithms that are known to converge globally in polynomial time.\(^{41}\) Furthermore, (13) is solved for each frequency \( \omega_k \) independently, such that the optimization of the entire \( Q(\omega_k) \) is readily parallelized. The performance associated to the resulting \( Q^* \) can be analyzed by noting that if \( Q^* \neq I \) and \( Q^* \neq 0 \), then \( \mathcal{G}(\Delta) \) is a nonconstant matrix function of \( \Delta \). In this case, the worst-case performance \( \sup_{\Delta \in \Delta} \| S(G_n^{-1}, Q^*, \Delta) \|_{2,2} \) can be determined numerically as is described in Appendix F. A proof to Theorem 1 requires the following result.

**Lemma 3.** Let \( \Delta \) be given by (3), and \( A, B \in \mathbb{C}^{n \times n} \), then it holds that

\[
\sup_{\Delta \in \Delta} \| A \Delta B \|_{2,2} < \kappa, \tag{14}
\]

if and only if there exists a real \( x > 0 \) such that

\[
\begin{bmatrix}
xB^H B - I & 0 \\
0 & (\frac{x}{\kappa} A^H A - xI)
\end{bmatrix} < 0. \tag{15}
\]

A proof to this lemma is provided in Appendix C. A proof to Theorem 1 is the following.

**Proof.** Consider the following implications

\[
\| A \|_{2,2} < \sqrt{\beta} \iff A^H A - \beta I < 0 \iff \begin{bmatrix}
-\beta I & A^H \\
A & -I
\end{bmatrix} < 0, \tag{16}
\]

where the second implication follows by the Schur complement.\(^{42(A.5.5)}\) Applying this to \( \| G_n(I - Q)G_n^{-1} \|_{2,2} < \sqrt{\beta}, \beta > 0 \), results in (13b). Clearly, minimizing \( \| G_n(I - Q)G_n^{-1} \|_{2,2} \) is identical to minimizing \( \beta \). Next, consider \( \| H(G_n^{-1}, Q, \Delta) \|_{2,2} < \kappa \). By Lemma 3, this condition is identical to

\[
\begin{bmatrix}
xW_u^H W_u - I & 0 \\
0 & (\frac{x}{\kappa} QG_n^{-1} W_y)^H(\frac{x}{\kappa} QG_n^{-1} W_y) - xI
\end{bmatrix} < 0. \tag{17}
\]

Applying the Schur complement\(^{42(A.5.5)}\) to the lower right matrix entry results in (13c), which concludes the proof. \( \blacksquare \)

In this section, a numerically efficient approach is developed to optimize \( Q \) for nominal performance when \( L = G_n^{-1} \). In the next section, it analyzed under what conditions perfect tracking can be achieved, and conditions are established under which the design choice \( L = G_n^{-1} \) is optimal.
FIGURE 3 (1.a, 2.a) \( G(\Delta) \) is a circular set centered around \( G_n \) with radius \( \gamma |W_y W_u| \). If \( G(\Delta) \) contains 0, that is, if \( \gamma |G_n^{-1}W_y W_u| \geq 1 \), as is the case for (1.a), then the set that determines convergence, \( H(\rho G_n^{-1}, 1, \Delta) \), will always contain the unit circle, that is, \( |H(\rho G_n^{-1}, 1, \Delta)| \geq 1 \), since it is a circular set centered around \( (1 - \rho) \) with radius \( \gamma \rho |G_n^{-1}W_y W_u| \). Consequently, convergence cannot be achieved and the worst-case minimum is obtained for \( \rho^* = 0 \), which implies that \( L^* = 0 \) is the optimal solution (1.c). If \( G(\Delta) \) does not contain 0, as is the case for (2.a), then convergence can be achieved, and the worst-case minimum is obtained for \( \rho^* = 1 \), which implies that \( L^* = G_n^{-1} \) is the optimal solution (2.c) [Colour figure can be viewed at wileyonlinelibrary.com]

6 | OPTIMIZING \( L \) FOR FAST CONVERGENCE AND PERFECT TRACKING

This section provides the main motivation for taking \( L = G_n^{-1} \) in the approach proposed in Section 4.1. Namely, it is shown that perfect tracking can only be achieved if the uncertainties are small enough, and in this case \( L = G_n^{-1} \) achieves the best possible robust convergence rate. This result constitutes contribution C2.

Perfect tracking is obtained if \( Q = I \) is feasible, that is, if it satisfies the convergence constraint as given by (4), since \( Q = I \), \( S(L, I, I) = S_c = 0 \). If \( Q \) is taken as \( Q = I \) in the optimization problems posed by (9) and (10), then only the convergence constraint (4) needs to be satisfied. Hence, (9) and (10) reduce to a feasibility problem in \( L \). If a set of feasible \( L \) exists, then a unique solution can be determined by minimizing \( \kappa \) in (4). This problem is formalized as follows.

**Problem 3.** Let \( G(\Delta) \) be given by Definition 2. Determine

\[
L^* = \arg \inf_{L \in \mathbb{C}} \sup_{\Delta \in \Delta} \| H(L, I, \Delta) \|_{2,2}, \quad \text{where} \quad H(L, I, \Delta) = I - L(G_n + W_y \Delta W_u). \tag{18}
\]

The solution is first illustrated for the SISO case and is followed by a formal treatment for MIMO systems in Section 6.2.

6.1 | Optimizing \( L \): Illustration of the SISO case

Consider (18) for the SISO case and parametrize \( L \) identically as \( L = \rho G_n^{-1} \), where \( \rho \) can take any value, that is, \( \rho \in \mathbb{C} \),

\[
\rho^* = \arg \inf_{\rho \in \mathbb{C}} \sup_{\Delta \in \Delta} |H(\rho G_n^{-1}, 1, \Delta)|, \quad H(\rho G_n^{-1}, 1, \Delta) = (1 - \rho) - \rho(G_n^{-1}W_y \Delta^* W_u). \tag{19}
\]

Note that \( H(\rho G_n^{-1}, 1, \Delta) \) represents a circular set in \( \mathbb{C} \) with center \((1 - \rho)\) and radius \( |\rho G_n^{-1}W_y W_u| \), as is illustrated in Figure 3. The worst-case magnitude \( \sup_{\Delta \in \Delta, \|\Delta\| \leq \gamma} |H(\rho G_n^{-1}, 1, \Delta)| \) is obtained for \( \Delta = e^{i \phi} \), where \( \phi \) equals the phase of \( (1 - \rho)/\rho G_n^{-1}W_y W_u \). Clearly, the worst-case magnitude equals the distance to the center point plus the radius of the circle, as is shown in Figure 3. Consequently,

\[
\rho^* = \arg \inf_{\rho \in \mathbb{C}} \sup_{\Delta \in \Delta, \|\Delta\| \leq \gamma} |(1 - \rho) - \rho(G_n^{-1}W_y \Delta^* W_u)| = \arg \inf_{\rho \in \mathbb{C}} (|1 - \rho| + |\rho||G_n^{-1}W_y W_u|). \tag{20}
\]

The solution to (20) follows from the following lemma.
Lemma 4. For a given $\alpha \geq 0$, consider $ho^* = \arg \inf_{\rho \in \mathbb{C}} (|1 - \rho| + |\rho|\alpha)$, then $ho^* = 0$ if $\alpha \geq 1$ and $ho^* = 1$ if $\alpha < 1$.

A proof to this lemma is provided in Appendix D. Lemma 4 implies that the solution to (19) depends on the radius $\gamma |G_n^{-1}W_yW_u|$, which can be recognized as the relative size of the uncertainty. More specifically, if $\gamma |G_n^{-1}W_yW_u| \geq 1$, then it cannot be fitted inside the unit circle. In this case, $\rho^* = 0$ is optimal, since it shrinks the circular set to 0 at the point 1, resulting in $|H(1, 0, \Delta)| = 1, \forall \Delta$. Hence, if the relative uncertainty is too large, then setting the learning parameters to 0 is optimal in the worst-case sense. An alternative interpretation follows by noting that $\gamma |G_n^{-1}W_yW_u| \geq 1$ implies that 0 is an element of the model set $G(\Delta)$, for which it is immediate that learning cannot be achieved. Conversely, if $\gamma |G_n^{-1}W_yW_u| < 1$ then smallest distance to the origin is obtained by centering the circular set at the origin, such that $\rho^* = 1$, with $|H(1, 0, \Delta)| = |G_n^{-1}W_yW_u| < 1$. To conclude, $L^* = G_n^{-1}$ if $\gamma |G_n^{-1}W_yW_u| < 1$ and $L^* = 0$ if $\gamma |G_n^{-1}W_yW_u| \geq 1$. Next, these results are generalized to the MIMO case.

### 6.2 Optimizing $L$: The MIMO case

The following theorem generalizes the solution to Problem 3 to the MIMO case under the following assumption.

**Assumption 6.** $W_u$ is unitary, that is, $W_uW_u^H = W_u^H W_u = I$.

This assumption is readily satisfied by using the identification method as is developed in Section 8. The essential aspect of this assumption is to render $\Delta W_u$ unstructured, such that any input vector $U_t$ can be arbitrarily rotated without affecting its norm.

**Theorem 2.** Let $G(\Delta)$ be given by Definition 2 and satisfy Assumptions 3 and 6. The solution to (18) in Problem 3 satisfies

$$L^* = \begin{cases} 0 & \text{if } \gamma \|G_n^{-1}W_y\|_{2,2} \geq 1, \text{ such that } \sup_{\Delta \in \Delta} \|H(I, 0, \Delta)\|_{2,2} = 1, \\ G_n^{-1} & \text{if } \gamma \|G_n^{-1}W_y\|_{2,2} < 1, \text{ such that } \sup_{\Delta \in \Delta} \|H(I, G_n^{-1}, \Delta)\|_{2,2} = \gamma \|G_n^{-1}W_y\|_{2,2}. \end{cases}$$  \hspace{1cm} (21a)

$$L^* = \begin{cases} 0 & \text{if } \gamma \|G_n^{-1}W_y\|_{2,2} \geq 1, \text{ such that } \sup_{\Delta \in \Delta} \|H(I, 0, \Delta)\|_{2,2} = 1, \\ G_n^{-1} & \text{if } \gamma \|G_n^{-1}W_y\|_{2,2} < 1, \text{ such that } \sup_{\Delta \in \Delta} \|H(I, G_n^{-1}, \Delta)\|_{2,2} = \gamma \|G_n^{-1}W_y\|_{2,2}. \end{cases}$$  \hspace{1cm} (21b)

Similar to the SISO case, Theorem 2 reveals that the solution depends on the relative uncertainty size, $\gamma \|G_n^{-1}W_y\|_{2,2}$. More specifically, (21a) states that if the uncertainty is too large, then there does not exist an $L$ that achieves convergence $\forall \Delta \in \Delta$ and $Q = I$, such that perfect tracking cannot be achieved. Conversely, if the uncertainty is sufficiently bounded, then perfect tracking is achieved at a convergence rate that is optimal in a robust sense if $L = G_n^{-1}$. This insight motivates the design choice of setting $L = G_n^{-1}$, even though this may be a suboptimal solution when $Q \neq I$ is required to achieve convergence. A proof of Theorem 2 is the following.

**Proof.** Parametrize $L$ identically as $L^* = \rho G_n^{-1}$, where $\rho$ is any complex matrix $\rho \in \mathbb{C}^{n \times n}$, and consider (18), which yields

$$\rho^* = \arg \inf_{\rho \in \mathbb{C}^{n \times n}} \sup_{\Delta \in \Delta} \|I - \rho - \rho G_n^{-1}W_y\Delta W_u\|_{2,2}. \hspace{1cm} (22)$$

This problem can be solved by reducing it to the SISO case. This is achieved by considering the direction along which the effect of uncertainty is maximal, followed by considering the specific unity input along this direction to evaluate the 2,2-induced matrix norm. The worst-case uncertainty is determined by taking the singular value decomposition (SVD) of $G_n^{-1}W_y$, that is, $G_n^{-1}W_y = UV^H$, which results in

$$\|I - \rho - \rho UV^H\Delta W_u\|_{2,2} = \|U(I - \tilde{\rho} - \tilde{\rho} \Sigma \tilde{\Delta})V^H\|_{2,2} = \|I - \tilde{\rho} - \tilde{\rho} \Sigma \tilde{\Delta}\|_{2,2}. \hspace{1cm} (23)$$

where $\tilde{\rho} \triangleq U^H \rho U$ and $\tilde{\Delta} \triangleq V^H \Delta W_u U$, where it is used that $U$ is a unitary matrix. Furthermore, since $V$ and $W_u$ are also unitary, $\tilde{\Delta} \in \Delta$, such that the transformed decision variables $\tilde{\Delta}$ and $\tilde{\rho}$ can be identically considered. Next, consider the particular perturbation $\Delta^* = -\gamma \rho^* I$, which is admissible since $\|\Delta^*\|_{2,2} \leq \gamma$. This perturbation provides a lower bound to the worst case

$$\sup_{\tilde{\Delta} \in \Delta} \|I - \tilde{\rho} - \tilde{\rho} \Sigma \tilde{\Delta}\|_{2,2} \geq \|I - \tilde{\rho} - \tilde{\rho} \Sigma \Delta^*\|_{2,2}. \hspace{1cm} (24)$$
Next, by definition of the induced 2,2-norm, it holds that

$$\|I - \hat{\rho} - \rho\Sigma\Delta\|_{2,2} \geq \|(I - \hat{\rho} - \rho\Sigma\Delta)z\|_2,$$

(25)

for any $z$ such that $\|z\|_2 = 1$. Taking $z = [1 \cdots 0]^T$, and noting that by convention $[\Sigma]_{11} = \tilde{\sigma}(G^{-1}_nW_y)$, the right-hand side of (25) can be written as

$$\left(1 - [\hat{\rho}]_{11} + |\hat{\rho}|_{11}\gamma\varepsilon\|ho\Sigma\Delta\|_2 + \sum_{i=2}^n |\hat{\rho}|_{1i}\right)^{1/2} \geq |1 - [\hat{\rho}]_{11} + |\hat{\rho}|_{11}\gamma\tilde{\sigma}(G^{-1}_nW_y)e^{j\phi}|,$$

(26)

which is similar to the SISO case discussed in Section 6.1. By similarly taking $\phi$ equal to the phase of $\frac{1 - |\hat{\rho}|_{11}}{|\hat{\rho}|_{11}\gamma\tilde{\sigma}(G^{-1}_nW_y)}$, the latter equals $|1 - [\hat{\rho}]_{11} + |\hat{\rho}|_{11}\gamma\tilde{\sigma}(G^{-1}_nW_y)$. In conclusion, with respect to (22), it holds that

$$\inf_{\rho \in \mathbb{C}^+} \sup_{\Delta \in \Delta} \|I - \rho G^{-1}_nW_y\Delta W_u\|_2 \geq \inf_{|\hat{\rho}|_{11} \in \mathbb{C}} |1 - [\hat{\rho}]_{11} + |\hat{\rho}|_{11}\gamma\tilde{\sigma}(G^{-1}_nW_y),$$

(27)

which enables a proof of Theorem 2 by contradiction. First, consider (21a), that is, let $\gamma\tilde{\sigma}(G^{-1}_nW_y) \geq 1$, and note that $\|H(I, 0, \Delta)\|_{2,2} = 1$. Assume that $\exists \rho, \rho \neq 0$, that results in a strictly lower costs than $\rho = 0$. In this case, it must hold that

$$1 > \inf_{\rho \in \mathbb{C}^+} \sup_{\Delta \in \Delta} \|I - \rho G^{-1}_nW_y\Delta W_u\|_2 \geq \inf_{|\hat{\rho}|_{11} \in \mathbb{C}} |1 - [\hat{\rho}]_{11} + |\hat{\rho}|_{11}\gamma\tilde{\sigma}(G^{-1}_nW_y).$$

(28)

By Lemma 4, the right-hand side equals 1. This contradicts the above statement, implying that $\rho = 0$ is optimal. Similarly, consider (21b), that is, let $\gamma\tilde{\sigma}(G^{-1}_nW_y) < 1$, and note that $\|H(I, G^{-1}_n, \Delta)\|_{2,2} = \gamma\tilde{\sigma}(G^{-1}_nW_y)$. Assume that $\exists \rho, \rho \neq I$, that results in a smaller costs than $\rho = I$. In this case, it must hold that

$$\gamma\tilde{\sigma}(G^{-1}_nW_y) > \inf_{\rho \in \mathbb{C}^+} \sup_{\Delta \in \Delta} \|I - \rho G^{-1}_nW_y\Delta W_u\|_2 \geq \inf_{|\hat{\rho}|_{11} \in \mathbb{C}} |1 - [\hat{\rho}]_{11} + |\hat{\rho}|_{11}\gamma\tilde{\sigma}(G^{-1}_nW_y).$$

(29)

By Lemma 4, the right-hand side equals $\gamma\tilde{\sigma}(G^{-1}_nW_y)$. This contradicts the above statement, implying that $\rho = I$ is optimal.

In this section, it is shown that perfect tracking can be obtained if the relative uncertainty is sufficiently bounded, and in that case, $L = G^{-1}_n$ optimizes the robust convergence rate. The next section considers the optimization of $Q$ for robust performance.

7 | OPTIMIZING Q FOR ROBUST PERFORMANCE

This section provides the main motivation for optimizing $Q$ for nominal performance as is advocated by the approach proposed in Section 4.1. Namely, it is shown that optimizing $Q$ for robust performance can be relatively conservative, since it results either in perfect tracking, or in not applying learning, that is, $Q = I$ or $Q = 0$, when $L = G^{-1}_n$. It is also shown that taking $\kappa < 1$ can be used to bound the worst-case performance. These insights constitute contribution C3.

Consider (10) for $L = G^{-1}_n$, which is formulated as follows.

**Problem 4.** Let $G(\Delta)$ be given by Definition 2. For a given $\kappa \in (0, 1]$, determine

$$Q^* = \arg \inf_{Q \in \mathbb{C}^{n \times n}} \sup_{\Delta \in \Delta} \|S(G^{-1}_n, Q, \Delta)\|_{2,2}, \quad \text{subject to: } \|H(G^{-1}_n, Q, \Delta)\|_{2,2} < \kappa \quad \forall \Delta \in \Delta,$$

(30)

$$S(G^{-1}_n, Q, \Delta) = I - (G_n + W_y\Delta W_u)(I + QG^{-1}_nW_y\Delta W_u)^{-1}QG^{-1}_n, \quad H(G^{-1}_n, Q, \Delta) = QG^{-1}_nW_y\Delta W_u.$$
Consider the optimization of $Q$ for robust performance and robust convergence for the SISO case, such that (30) and (31) yield

$$Q^* = \arg \inf_{Q \in C_n^{-1}} \sup_{\Delta \in \Delta} \left| \frac{1 - (G_n + W_1^2 \Delta W_n)QG_n^{-1}}{1 + QG_n^{-2}W_1^2 \Delta W_n} \right|, \quad \text{subject to: } |QG_n^{-1}W_1^2 \Delta W_n| < \kappa \quad \forall \Delta \in \Delta. \quad (32)$$

Note that the performance function is identical to $S(G_{n}^{-1}, Q, \Delta) = (1 - Q)(1 + QG_{n}^{-1}W_1^2 \Delta W_n)^{-1}$, which directly shows that $Q = 1$ is optimal if it feasible, that is, if it satisfies the convergence constraint. If the constraint is not satisfied for $Q = 1$ and $\kappa = 1$, then $\sup_{\Delta} |S(G_{n}^{-1}, Q, \Delta)| \geq 1$ for all feasible $Q$, implying that the worst-case performance cannot be improved. To see this, note that if $\exists \Delta$ such that $|G_{n}^{-1}W_1^2 \Delta W_n| \geq 1$, then $|QG_{n}^{-1}W_1^2 \Delta W_n| < 1$ does not hold for $Q = 1$. This is equivalent to stating that there exists a $\Delta^*$ such that $(G_n + W_1^2 \Delta W_n) = 0$, which implies $S(G_{n}^{-1}, Q, \Delta^*) = 1$. Consequently, $\sup_{\Delta} |S(G_{n}^{-1}, Q, \Delta)| \geq 1 \forall \Delta$. Hence 1 presents a lower bound, and this lower bound is attained $\forall \Delta \in \Delta$ for $Q = 0$, implying that $Q^* = 0$. Now consider the case where a certain $Q$ is such that the convergence constraint is active for $\kappa = 1$, that is, $|QG_{n}^{-1}W_1^2 \Delta W_n| < 1, \forall \Delta \in \Delta$. For such a $Q$, $\Delta^*$ such that the denominator in $S(G_{n}^{-1}, Q^*, \Delta^*)$ approaches 0, resulting in an unbounded worst-case performance. By taking $\kappa < 1$ this situation can be relieved. These mechanisms are illustrated in Figure 4. Next, these results are generalized to the MIMO case.

### 7.2 Optimizing $Q$: The MIMO case

The following theorem presents the general solution to Problem 4 for MIMO systems.

**Theorem 3.** Let $G(\Delta)$ be given by Definition 2 and satisfy Assumptions 3 and 6. The solution to (30) and (31) in Problem 4 satisfies

$$Q^* = \begin{cases} 0 & \text{if } \gamma \|G_n^{-1}W_1\|_{2,2} \geq 1, \text{ such that } S(G_n^{-1}, 0, \Delta) = S_n(G_n^{-1}, 0) = I, \quad (33a) \\ I & \text{if } \gamma \|G_n^{-1}W_1\|_{2,2} < \kappa, \text{ such that } S(G_n^{-1}, I, \Delta) = S_n(G_n^{-1}, I) = 0. \quad (33b) \end{cases}$$

Similar to the SISO case, (33a) states that if the relative uncertainty is too large, then there does not exist a $Q$ that increases the performance $\forall \Delta \in \Delta$, such that the optimal solution is to not apply IIIC, that is, $Q = 0$. Conversely by (33b), if the relative uncertainty small compared $\kappa$, then the optimal robust performance is attained for $Q = I$, which implies...
perfect tracking. This all-or-nothing solution is indicative of conservatism since it leaves no room to partially improve the performance when \( \kappa = 1 \).

Any \( Q \neq I \) that optimizes the nominal performance, that is, is a solution to the SDP posed by (13), is such that the constraint \( \| H(G_n^{-1}, Q, \Delta) \|_{2,2} < \kappa \) is active. For \( \kappa = 1 \), this implies that \( \exists \Delta^* \in \Delta \) for which the factor \( I + Q^* G_n^{-1} W_\Delta^* W_u \) in \( S(G_n^{-1}, Q, \Delta) \), given by (31), approaches singularity, yielding unbounded worst-case performance. This observation indicates that the worst-case performance can be moderated by taking \( \kappa < 1 \), as is demonstrated in Section 10. A proof of Theorem 3 requires the following auxiliary result.

**Lemma 5.** Let \( \Delta \) be given by (3), and \( A, B \in \mathbb{C}^{n \times n} \), with \( B \) unitary, then it holds that

\[
\sup_{\Delta \in \Delta} \lambda(A \Delta B) = \gamma \| A \|_{2,2}.
\]  

A proof of this lemma is provided in Appendix E. A proof to Theorem 3 is the following.

*Proof.* Consider (33b) and note that \( \gamma \| G_n^{-1} W_\gamma \|_{2,2} < \kappa \) is identical to \( \| H(G_n^{-1}, I, \Delta) \|_{2,2} < \kappa \ \forall \Delta \in \Delta \), which implies that \( Q = I \) satisfies the convergence constraint and that \( S(G_n^{-1}, I, \Delta) \) is well-posed by virtue of Lemma 2. Evaluating \( S(G_n^{-1}, I, \Delta) \) results in

\[
S(G_n^{-1}, I, \Delta) = I - (G_n + W_\gamma \Delta W_u)(I + G_n^{-1} W_\gamma \Delta W_u)^{-1} G_n^{-1} = I - (G_n + W_\gamma \Delta W_u) G_n^{-1} (I + W_\gamma \Delta W_u G_n^{-1})^{-1} = I - (I + W_\gamma \Delta W_u G_n^{-1})(I + W_\gamma \Delta W_u G_n^{-1})^{-1} = 0,
\]

where the push-through rule is used, that is, \( (I + AB)^{-1} A = A(I + BA)^{-1} \). Hence, for \( Q = I \), \( S(G_n^{-1}, I, \Delta) = 0 \ \forall \Delta \), which is the optimal solution. Consider (33a) and note that \( \gamma \| W_\gamma \delta_n^{-1} W_\gamma \|_{2,2} \geq 1 \) is identical to \( \sup_{\Delta \in \Delta} \tilde{\lambda}(G_n^{-1} W_\gamma \Delta W_u) \geq 1 \) by virtue of Lemma 5. Consequently, \( \exists \Delta^* \in \Delta \) such that \( \tilde{\lambda}(G_n^{-1} W_\gamma \Delta W_u) = 1 \), since \( \tilde{\lambda}(.) \) is continuous. This implies that \( \det(I + \lambda G_n^{-1} W_\gamma \Delta W_u) = 0 \) for \( \lambda = 1 \), which implies that \( \det(G_n) \det(I + G_n^{-1} W_\gamma \Delta^* W_u) = \det(G_n + W_\gamma \Delta^* W_u) = 0 \), since \( \det(G_n) \neq 0 \). This in turn implies that \( \exists \nu \) such that \( (G_n + W_\gamma \Delta^* W_u)\nu = 0 \). Define \( r \triangleq Z^{-1} v \), with \( Z \triangleq (I + G_n^{-1} W_\gamma \Delta W_u)^{-1} G_n^{-1} \) where \( Z^{-1} \) exists by virtue of Lemma 2. Next, using the definition of the 2,2-norm, it follows that

\[
\| S(G_n^{-1}, Q, \Delta^*) \|_{2,2} \geq \| S(G_n^{-1}, Q, \Delta^*) \|_2 \| r \|_2^{-1} = \| \nu \|_2^{-1} \|
u - (G_n + W_\gamma \Delta W_u) Z Z^{-1} v \|_2 \| r \|_2^{-1} = \| \nu \|_2^{-1} \|
u \|_2^{-1} = 1.
\]

Hence, \( \sup_{\Delta \in \Delta} \| S(Q, \Delta) \|_{2,2} \geq 1 \), and it follows directly from (31) that this lower bound is attained \( \forall \Delta \in \Delta \) for \( Q = 0 \), which is therefore the optimal solution.

In this section, \( Q \) is optimized for robust performance. It is shown that this approach can be relatively conservative, and that the worst-case performance can be bounded by taking \( \kappa < 1 \). This concludes the first part of this article, in which the approach proposed in Section 4 is motivated. In the next section, this approach is complemented by the development of an identification method to identify the model set.

# 8 Identification of the Model Set

In this section, a method is developed to identify the model set \( \mathcal{G}(\Delta) \), as given by (3), which constitutes contribution C4. A connection is established between the stochastic disturbances that corrupt the output during the identification process and the resulting uncertainty of the nominal model. This connection results in an approach to determine the uncertainty bound \( \gamma \).

## 8.1 MIMO FRF identification

Consider the configuration in Figure 1. The MIMO FRF \( \hat{G}(\omega_k) \) can be estimated from \( n \) input-output data pairs as

\[
\hat{G}(\omega_k) = \mathcal{Y}(\omega_k) \mathcal{U}^{-1}(\omega_k),
\]

\[
\mathcal{U}(\omega_k) = \begin{bmatrix} U_1(\omega_k) & \cdots & U_n(\omega_k) \end{bmatrix} \in \mathbb{C}^{n \times n}, \quad \mathcal{Y}(\omega_k) = \begin{bmatrix} Y_1(\omega_k) & \cdots & Y_n(\omega_k) \end{bmatrix} \in \mathbb{C}^{n \times n}.
\]
where it is assumed that $\cup(\omega_k)$ is regular $\forall \omega_k$, which is achieved appropriate input signal selection. To quantify the uncertainty of $\hat{G}(\omega_k)$, it is assumed that the output is corrupted by the disturbances as follows, where a distinction is made between the DFT coefficients that are real and complex by property.

**Definition 3.** The real DFT coefficients correspond to the frequency set $\Omega_R \triangleq \{ \omega_k \in \Omega_N \mid k = 0 \land \text{if } N \text{ is even } k = \frac{N}{2} \}$, and the complex DFT coefficients correspond to the remaining set of frequencies, $\Omega_\mathbb{C} \triangleq \Omega_N \setminus \Omega_R$.

**Assumption 7.** The output $Y_i(\omega_k)$ is given by (2), where $T_U(\omega_k) = 0$, $D(\omega_k) = 0$. Furthermore, $V_i(\omega_k)$ is circularly symmetric complex Gaussian distributed $\forall \omega_k \in \Omega_\mathbb{C}$ and $V_i(\omega_k)$ is real Gaussian distributed $\forall \omega_k \in \Omega_R$. The covariance matrix $C_V(\omega_k)$ is Hermitian positive definite, and different realizations $V_i(\omega_k)$ are mutually independent.

**Remark 1.** Assumption 7 can be approximately satisfied in practice as follows.

- For a stable LTI system, $T_U(\omega_k) \rightarrow 0$ as $N \rightarrow \infty$. Hence, the transient response becomes negligible by taking a long measurement interval, or by applying periodic excitations \cite{DeRo} and excluding the first periods from the measurement.
- The trial-invariant deterministic disturbance can be eliminated, that is, $D(\omega_k) = 0$, by taking the difference between two measurements with different input signals. Specifically, let $Y_1 = GU_1 + D + V_1$, and $Y_2 = GU_2 + D + V_2$. Then, $Y_1 - Y_2 = G(U_1 - U_2) + (V_1 - V_2)$, which can be redefined as $Y_\delta = GU_\delta + V_\delta$, which is a measurement with $D = 0$.
- The trial-variant stochastic disturbance $V_i(\omega_k)$ satisfies Assumption 7 asymptotically ($N \rightarrow \infty$) if $v_i(t) = H(z) e_i(t)$ with $H(z)$ a stable filter and $e_i(t)$ independent, identically distributed noise with existing moments of any order. This condition is considered mild since it does not assume a specific distribution in the time domain.

Assumption 7 motivates the following model set that establishes a direct connection between the disturbances and $\Delta$.

**Proposition 1.** Under Assumption 7, set $G_n = \hat{G}$, with $\hat{G}$ given by (36), set $W_u = \sigma_u \cup^{-1}$, where $\sigma_u \triangleq \sigma(\cup)$ and where $\cup$ is given by (37), and set $W_y = \frac{1}{\sigma_u} C_{1/2} V$ where $C_{1/2} V (C_{1/2} V)^H = C_V$ is the Cholesky decomposition, which exists since $C_V$ is Hermitian positive definite. For these settings, (3) results in

$$G(\Delta) = G_n + C_{1/2} V (\Delta) \cup^{-1}, \quad \Delta \in \Delta,$$

(38)

where $V \triangleq C_{1/2} V$ is such that $V V^H$ a complex Wishart matrix, as is described in Reference \cite{DeRo} and Appendix G.

**Proof.** Under Assumption 7, (36) can be rewritten as

$$\hat{G} = Y \cup^{-1} = G_o + V \cup^{-1}, \quad V = \left[ \begin{array}{c} V_1 \cdots V_n \end{array} \right] \in \mathbb{C}^{n \times n},$$

(39)

where $V$ is a random matrix whose columns $V_i$ satisfy Assumption 7, such that $V \triangleq C_{1/2} V$ is a Wishart matrix. For the proposed settings, it holds that $G(\Delta) = G_n + W_y \Delta W_u = G_o + V \cup^{-1} + \frac{1}{\sigma_u} C_{1/2} V \Delta \cup^{-1}$, which is identical to (38).

Considering (38), Proposition 1 implies that if there exists $\Delta^* \in \Delta$ such that $\Delta^* = -\overline{V}$, then the true system $G_o$ is contained in the model set $G(\Delta)$. Consequently, $\gamma$ can be used to tune the probability that $\overline{V} \in \Delta$, as is considered in the next section.

**Remark 2.** By setting $W_u = \sigma_u \cup^{-1}$, Assumption 6 is satisfied if $\sigma_u^{-1} \cup$ is unitary, which can be achieved by using orthogonal multisine excitation \cite{DeRo}.

### 8.2 Determining the uncertainty bound $\gamma$

The section treats the problem of selecting $\gamma$ such that $G_o$ is contained in $G(\Delta)$ with probability $\alpha$, as is illustrated for the one-dimensional case in Section 2.3. By definition of $\Delta$, as given by (3), a $\Delta \in \Delta$ can attain $-\overline{V}$ if $\overline{\sigma(\overline{V})} \leq \gamma$. Hence, the probability that $G_o$ is contained in $G(\Delta)$ equals the cumulative distribution function (CDF) $\Pr(\overline{\sigma(\overline{V})} \leq \gamma)$. The following lemma provides explicit expressions of this CDF.
Lemma 6. Let $\Gamma(x)$ denote the gamma function and let $\psi(a, x)$ denote the regularized lower incomplete gamma function, that is, $\psi(a, x) = \frac{1}{\Gamma(a)} \int_x^\infty z^{a-1} e^{-z} dz$, $\Gamma(a) = \int_0^\infty z^{a-1} e^{-z} dz$. If $\gamma$, as given by (39), satisfies Assumption 7, then the CDF of $\tilde{\sigma}(\tilde{V}(\omega_k))$, with $\tilde{V}(\omega_k) = C_V^{-1/2}(\omega_k)\overline{V}(\omega_k)$, is given by

$$
F(\gamma, \omega_k) = \Pr \left( \tilde{\sigma}(\tilde{V}(\omega_k)) \leq \gamma \right) = \begin{cases} F_\Theta(\gamma) & \text{if } \omega_k \in \Omega_\Theta, \\ F_R(\gamma) & \text{if } \omega_k \in \Omega_R, \end{cases}
$$

where $\Psi(x) \in \mathbb{R}^{d\times n}$ is a Hankel matrix function of $x \in (0, \infty)$, with $[\Psi(x)]_{ij} = \Gamma(i + 1 - j)\psi(i + j - 1, x)$, and related to $F_R(\gamma)$, $n_m = n$ if $n$ is even, and $n_m = n + 1$ if $n$ is odd. When $n$ is even, then elements of the skew symmetric matrix function satisfy $[\Phi(x)]_{ij} = I(a_i, a_j, x) - I(a_i, a_j, x), i, j = 1, \ldots, n$, where $a_i = i - \frac{1}{2}$ and $I(a, b, x) \triangleq \frac{1}{\Gamma(a)} \int_0^x t^{a-1} e^{-t} \psi(b, t) dt$. If $n$ is odd, $\Phi$ is an $n_m \times n_m$ matrix function with additional elements

$$
[\Phi(x)]_{i,n} = \frac{2^{n-1/2}}{\Gamma \left( n - \frac{1}{2} \right)} \psi \left( a_i, \frac{n}{2} \right), \quad i = 1, \ldots, n, \quad [\Phi(x)]_{n,n} = -[\Phi(x)]_{1,n}, \quad [\Phi(x)]_{n,n} = 0.
$$

A proof of this lemma is provided in Appendix G. Lemma 6 enables the following key result, which establishes a connection between FRF identification and the developed $L$ and $Q$ design approach.

Theorem 4. Consider $G(\Delta)$ as given by (3), under Assumptions 2 and 7, and let $G_n(\omega_k), W_y(\omega_k)$ and $W_u(\omega_k)$ be set as in Proposition 1. If $\gamma$ is such that $F(\gamma, \omega_k) \geq \alpha$, with $F(\gamma)$ as presented in Lemma 6, then $\Pr(G_n(\omega_k) \in G(\Delta, \omega_k)) \geq \alpha$.

Proof. Considering $G(\Delta)$ for $G_n = \hat{G}$, $W_y = \sigma(\overline{U})^{-1}C_V^{1/2}$ and $W_u = \sigma(\overline{U})U^{-1}$ yields (38). If $\tilde{\sigma}(\tilde{V}) \leq \gamma$, then $\exists \Delta_n \in \Delta$ such that $\tilde{V} + \Delta_n = 0 \Leftrightarrow G(\Delta_n) = G_n$. Hence, if $\gamma$ is such that $\Pr(\tilde{\sigma}(\tilde{V}) \leq \gamma) \geq \alpha$, then $\Pr(G_n \in G(\Delta)) \geq \alpha$.

This theorem establishes the implicit relation between $\gamma$ and the probability that $G_n$ is contained in $G(\Delta)$. The inverse relation exists since the CDF is a nondecreasing function of $\gamma$, and hence, $\gamma$ can be determined numerically by means of a root-finding algorithm to achieve a desired probability $\alpha$. Note that Theorem 4 considers a single frequency $\omega_k$. Consequently, the probability that $G_n(\omega_k)$ is contained in $G(\Delta)$ for a set of frequencies equals $\alpha^r$, with $r = N_\Re + \frac{1}{2}N_\Im$, where $N_\Re$ and $N_\Im$ are the numbers of frequencies corresponding to real and imaginary DFT coefficients.

Remark 3. In this article, a bound on $\tilde{\sigma}(\Delta) = \|\Delta\|_2$ is formulated, which is directly compatible with classical robust control. This is in contrast to traditional identification approaches, which typically formulate a bound on $\|\text{vec}(\Delta)\|_2^{18\{2.7.2\}}$.

Remark 4. Practically, the covariance matrix $C_V$ may be unknown and can be replaced by the sample covariance $C_T$ of the sample means $\overline{Y}_T$ that can be obtained from additional observations, for example by using periodic repetitions. This reduces the uncertainty by averaging, although technically (41) and (42) no longer hold, since the elements of $\overline{V}$ satisfy a Student $t$-distribution.

In this section, a method is developed to identify the model set $G(\Delta)$, which includes an approach to determine the norm bound $\gamma$ such that the true system is contained in the model set with a desired probability $\alpha$. In conjunction with the results from the previous section, a complete design approach is obtained that is concisely formulated in the next section.

## 9 DESIGN FRAMEWORK

In this section, the $L$ and $Q$ design framework summarized.
### Initialization

Given a system $G_n$ that satisfies Assumption 2.

1. Determine the desired probability $\alpha \in (0, 1)$ at which $U(\omega_k)$ should converge at an individual frequency $\omega_k$.
2. Determine the desired upper bound on the convergence rate $\kappa \in (0, 1]$, as is given by Definition 1.

Perform the following steps for each relevant frequency $\omega_k$.

#### Identifying the model set

1. Obtain input-output data $\{Y_i(\omega_k), U_i(\omega_k)\}_{i=1}^n$ by performing $n$ experiments. Construct $\mathbb{U}(\omega_k)$ and $\mathbb{V}(\omega_k)$, as given by (37). Ensure the experimental conditions satisfy Assumption 7 and that $\mathbb{F}(\mathbb{U}(\omega_k))^{-1}\mathbb{U}(\omega_k)$ is unitary, as described in Remark 2. Optionally, estimate $C_V(\omega_k)$ as described in Remark 4.

2. Compute $\hat{G}(\omega_k)$ as given by (36), and set $G_n(\omega_k) = \hat{G}(\omega_k)$, $W_f(\omega_k) = \mathbb{F}(\mathbb{U}(\omega_k))^{-1}C_F(\omega_k)^{1/2}$, and $W_u(\omega_k) = \mathbb{F}(\mathbb{U}(\omega_k))^{-1}(\omega_k)$.

3. Compute $\gamma_3$ such that $F_3(\gamma_3) = \alpha$, as given by (41), and compute $\gamma_\mathcal{R}$ such that $F_\mathcal{R}(\gamma_\mathcal{R}) = \alpha$, as given by (42).

#### Determining $L$ and $Q$

Let $\gamma = \gamma_3$ if $\omega_k \in \Omega_3$, or $\gamma = \gamma_\mathcal{R}$ if $\omega_k \in \Omega_\mathcal{R}$, where $\Omega_3$ and $\Omega_\mathcal{R}$ are specified in Definition 3.

1. Invert the nominal model and set $L(\omega_k) = G_n^{-1}(\omega_k)$.

2. If $\gamma \| G_n^{-1} W_f \|_{2,2} < \kappa$, then set $Q = I$, else, compute $Q$ by solving (13). Exploit the symmetry of the DFT, that is, $Q(\omega_k) = \overline{Q}(\omega_{-k})$, $Q(\omega_{k+N}) = Q(\omega_k)$, $p \in \mathbb{Z}$, to reduce the number of SDPs.

Recall that the worst-case performance $\sup_{\Delta \in \mathbb{A}} \| S(G_n^{-1} Q^*, \Delta) \|_{2,2}$ can be computed as is discussed in Appendix F. In case the nominal performance is only slightly smaller 1, while the worst-case performance far exceeds 1, consider setting $Q = 0$ or recompute $Q$ for a smaller $\kappa$.

In this section, the results from the previous sections are combined into a complete design approach. In the next section, this approach is illustrated by application.

## 10 APPLICATION TO A FLEXIBLE MOTION SYSTEM

In this section, the developed design framework, as formulated in Section 9, is illustrated by designing $L$ and $Q$ matrices for a benchmark flexible beam system. Both a diagonal and full $Q$ matrix are synthesized and used to apply MIMO IIC, as presented in Section 2. The results indicate high tracking performance for both cases, yet superior performance is achieved for the fully parametrized $Q$ matrix, thereby confirming the potential of the developed framework, which constitutes contribution C5.

### 10.1 The plant and initialization

The plant is simulated by 16th-order DT LTI model with input-output dimension $n = 2$ with a sample rate of 1 kHz, whose FRF $G_n(e^jo\tau)$ is shown in Figure 5. This system represents the benchmark flexible beam as discussed in Reference 47.

The frequency grid of the FRF model is made to match the DFT grid of a reference signal consisting of $N = 2^{10}$ samples. It is desired that the IIC algorithm converges with a probability $\geq 0.95$, that is, the probability that convergence occurs at all frequency bins simultaneously must be at least 0.95. This is achieved if an individual frequency converges with probability $\alpha = 0.95^\tau$, where $\tau$ is the number of independent frequency bins, that is $\tau = \frac{1}{2}N_3 + N_\mathcal{R} = \frac{1}{2}(2^{10} - 2) + 2$. Furthermore, the upper bound on the convergence rate is set to $\kappa = 0.2$.

### 10.2 Identifying the model set

The input-output data $\{Y_i(\omega_k), U_i(\omega_k)\}_{i=1}^2$ is obtained by performing two experiments, during which $v(t)$ is taken such that $V(\omega_k)$ asymptotically satisfies Assumption 7 with $C_V = \sigma^2 \begin{bmatrix} 50 & \sigma \\ 5 & 10 \end{bmatrix}$, $\sigma = 10^{-7}$. The matrix $U$ satisfies $\| U(\omega_k) \| = 1$.
\textbf{FIGURE 5} FRF of the true plant $G_n(\theta^{\text{true}})$ (---) and the estimate $\hat{G}(\omega_k)$ (●), which is taken as the nominal model $G_n(\omega_k)$ [Colour figure can be viewed at wileyonlinelibrary.com]

\forall \omega_k$, where $|A|$ denotes the elementwise absolute value of $A \in \mathbb{C}^{n \times m}$. This structure is achieved by taking the $i$th input during the $i$th experiment as random-phase multisine with unit spectrum. This makes $\mathbb{U}$ is unitary, such that the results in Section 6 hold. The nominal model $\hat{G}_n$ is determined as (36), and is shown in Figure 5. The norm bounds $\gamma_\text{I}$ and $\gamma_\text{R}$ are determined by solving $F_\text{I}(\gamma) - \alpha = 0$ and $F_\text{R}(\gamma) - \alpha = 0$ using the MATLAB-routine \texttt{fzero.m}.

\subsection*{10.3 | Determining $L$ and $Q$}

Since $G_n$ results from measured data, it is generically invertible. Hence, $L$ is readily determined as $L = G_n^{-1}$. In the following, $Q$ is designed as a full matrix, $Q_f \in \mathbb{C}^{2 \times 2}$, and as diagonal matrix, $[Q_d]_{ii} \in \mathbb{C}$, $[Q_d]_{ij} = 0$ for $i \neq j$. Comparing the performance of both matrices provides insight about the performance gap that potentially exist in preexisting approaches. For both cases, $Q$ is optimized for nominal performance, that is, steps 2(a) and 2(b) in Section 9 are performed, where (13) is readily solved using either a full or diagonal parametrization by using YALMIP\textsuperscript{48} in combination the semidefinite program (SDP) solver SeDuMi.

The resulting $Q$-matrices are shown in Figure 6, which displays the magnitude and phase. Three key observations are made. First, both $Q$ matrices satisfy $Q = I$ up to 55 Hz, which corresponds to the perfect tracking solution. Second, beyond $60 \text{ Hz}$, the freedom to assume any full matrix in $\mathbb{C}^{2 \times 2}$ is exploited by $Q_f$, that is, the off-diagonals are nonzero, $Q_f$ is clearly not Hermitian, and all entries are complex. By contrast, $Q_d$ turns out to be a real matrix, even though its parametrization is complex. This is in agreement with the well-adopted notion of phase-less robustness filters. Third, and most remarkably, beyond 55 Hz, $Q_f$ displays relatively large peaks. This goes against common SISO intuition which requires that $|Q|<1$ to increase robustness, as indeed turns out to be the optimal solution for $Q_d$. 
10.4 Validation

The performance of both $Q$-matrices is shown in Figure 7. In the top plot, the optimized nominal performance bounds are displayed, which correspond to $\beta$ in (13). These bounds indicate perfect tracking up to 55 Hz, and in the range [100, 200] Hz, a significant performance improvement is achieved, that is, $\beta < 1$. The difference between $Q_f$ and $Q_d$ is apparent in the range [150, 250] Hz, where $Q_f$ performs significantly better. In the second plot, the actually achieved performance is displayed, which shows great similarity with the first plot. This validates that optimizing for nominal performance indeed provides an effective way to optimize the true performance. In the bottom plot, the worst-case performance is displayed, which paints a relatively conservative picture in which $Q_d$ significantly outperforms $Q_f$. Before the designed $L$ and $Q$ matrices are employed to increase the tracking performance through IIC in the next section, the effect of $\kappa$ is illustrated by optimizing $Q_f$ for $\kappa \in \{0.2, 0.5, 0.7, 0.9\}$. The resulting nominal and worst-case performance are shown in Figure 8. Increasing $\kappa$ makes the convergence constraint less stringent, such that that the nominal performance improves for all frequencies. In contrast, the worst-case performance improves for those frequencies where it is $<1$, and it deteriorates significantly for those frequencies it is $>1$. The latter illustrates that in the limit $\kappa \rightarrow 1$, the worst-case performance becomes unbounded as is described in Section 7.

10.5 Applying iterative inversion-based control

The control goal is to track a high-frequency triangle wave with both outputs. To achieve this, IIC is applied with the designed $L$ and $Q$ matrices, as is presented in Algorithm 1 in Section 2.1. The error is measured after approximately 3 seconds to achieve steady state, and the disturbances are as during the identification experiments.
**Figure 7** The performance for the full matrix $Q_f$ (---) and the diagonal matrix $Q_d$ (---), where smaller values correspond to higher performance. The optimized nominal performance (top plot) is similar to the true performance (middle plot), indicating that optimizing for nominal performance is an effective strategy. Furthermore, $Q_f$ significantly outperforms $Q_d$ in the range [150, 250] Hz, whereas for the worst-case performance (bottom plot), $Q_d$ outperforms $Q_f$ for most frequencies. [Colour figure can be viewed at wileyonlinelibrary.com]

**Figure 8** The performance for the full matrix $Q_f$ as $\kappa$ is varied from 0.2 (---) to 0.9 (---). The nominal performance (top plot) improves for all frequencies, since the convergence constraint is less stringent. The worst-case performance (bottom plot) improves for those frequencies where it is <1, whereas it deteriorates significantly for those frequencies where it is >1 (note the logarithmic y-scale). [Colour figure can be viewed at wileyonlinelibrary.com]

The time domain performance is shown in Figure 9, which shows the initial tracking error, and the converged error obtained by using $Q_d$ and $Q_f$. Convergence is approximately attained after a single iteration, which results in a significant performance increase, as is shown in Figure 10. These observation confirm that Problem 1 is solved by applying robust IIC, where $L$ and $Q$ are designed by using the approach developed in this article. The performance difference, as is especially clear in in the second output, confirms there is a benefit to using full robustness matrices over diagonal matrices in IIC.
**FIGURE 9** The initial tracking error \( e(t) \) (\( -\)) and the error after one IIC iteration (Algorithm 1) for \( Q_f \) (\( -\)) and \( Q_d \) (\( -\)). This reveals that significant performance improvement is obtained, and that \( Q_f \) outperforms \( Q_d \) in output 2 since the achieved error is significantly smaller [Colour figure can be viewed at wileyonlinelibrary.com]

**FIGURE 10** The tracking error energy \( \|e(t)\|_2^2 = \sum_{t=0}^{N} e(t)^2 \) per iteration \( i \) for output 1 (left) and output 2 (right), for \( Q_f \) (\( -\)) and \( Q_d \) (\( -\)), reveals that \( Q_f \) outperforms \( Q_d \) in output 2 [Colour figure can be viewed at wileyonlinelibrary.com]

11 | CONCLUSION

An iterative inversion-based control framework is developed for multivariable systems that enables high performance and fast convergence in the presence of unknown disturbances. This is achieved by developing a design framework for the learning and robustness matrices by connecting robust control and nonparametric FRF identification. In the developed framework, the learning matrix is taken as the inverse of the system FRF and the robustness matrix is optimized to achieve optimal nominal performance while ensuring robust monotonic convergence in the presence model mismatches that are induced by stochastic disturbances that corrupt identification experiments. The resulting design process is computationally efficient and can be parallelized by virtue of the frequency-by-frequency design approach. The learning framework is developed for applications that readily allow estimation of FRF models and that perform repetitive tasks during which the system is approximately in steady state. Furthermore, since the developed approach does not require a parametric model, the dynamic order of the system does not impose any computational limitations. This makes the approach especially suitable for systems whose dynamics are predominantly described by LTI flexible dynamics, such as mechatronic motion systems. The efficacy of the developed framework is illustrated through application to a simulated benchmark flexible beam system, which shows that superior tracking performance can be achieved by exploiting the design freedom that is offered by considering full-MIMO robustness matrices. Future research is aimed at the experimental application of the developed framework to industrial systems to further motivate the adoption of the learning control framework in practice.

ACKNOWLEDGEMENT

This work is supported by Canon Production Printing and the research program VIDI (no 15698), which is (partly) financed by the Netherlands Organization for Scientific Research (NWO).

ORCID

Robin de Rozario  
https://orcid.org/0000-0002-3668-2168
REFERENCES
1. Bristow DA, Tharayil M, Alleyne AG. A survey of iterative learning control. *IEEE Control Syst Mag.* 2006;26(3):96-114.
2. Ahn HS, Chen Y, Moore KL. Iterative learning control: brief survey and categorization. *IEEE Trans Syst Man Cybern Part C (Appl Rev).* 2007;37(6):1099-1121.
3. Yu P, Liu KZ, She J, Wu M, Nakanishi Y. Robust disturbance rejection for repetitive control systems with time-varying nonlinearities. *Int J Robust Nonlinear Control.* 2019;29(5):1597-1612.
4. Tien S, Zou Q, Devasia S. Iterative control of dynamics-coupling-caused errors in piezoscanners during high-speed AFM operation. *IEEE Trans Control Syst Technol.* 2005;13(6):921-931.
5. Pipeleers G, Moore KL. Unified analysis of iterative learning and repetitive controllers in trial domain. *IEEE Trans Automat Control.* 2014;59(4):953-965.
6. de Rozario R, Fleming A, Oomen T. Finite-time learning control using frequency response data with application to a nanopositioning stage. *IEEE/ASME Trans Mechatron.* 2019;24(5):2085-2096.
7. Shan Y, Leang KK. Design and control for high-speed nanopositioning: serial-kinematic nanopositioners and repetitive control for nanofabrication. *IEEE Control Syst Mag.* 2013;33(6):86-105.
8. Xie H, Wen Y, Shen X, Zhang H, Sun L. High-speed AFM imaging of nanopositioning stages using $H_{\infty}$ and iterative learning control. *IEEE Trans Ind Electron.* 2019;67(3):2430-2439.
9. Aarnoudse L, Pannier C, Afkhami Z, Oomen T, Barton K. Multi-layer spatial iterative learning control for micro-additive manufacturing. *IFAC-PapersOnLine.* 2019;52(15):97-102.
10. Lim I, Hoelzle DJ, Barton KL. A multi-objective iterative learning control approach for additive manufacturing applications. *Control Eng Pract.* 2017;64:74-87.
11. Blanken L, Koekebakker S, Oomen T. Multivariable repetitive control: decentralized designs with application to continuous media flow printing. *IEEE/ASME Trans Mechatron.* 2019;25(1):294-304.
12. de Rozario R, Oomen T. Data-driven iterative inversion-based control: achieving robustness through nonlinear learning. *Automatica.* 2019;107:342-352.
13. Owens DH. *Iterative Learning Control: An Optimization Paradigm.* New York, NY: Springer; 2015.
14. Chen Y, Wen C. *Iterative Learning Control: Convergence, Robustness and Applications.* Berlin, Germany: Springer-Verlag; 1999.
15. Boeren F, Bareja A, Kok T, Oomen T. Frequency-domain ILC approach for repeating and varying tasks: with application to semiconductor bonding equipment. *IEEE/ASME Trans Mechatron.* 2016;21(6):2716-2727.
16. Norrlöf M, Gunnarsson S. Time and frequency domain convergence properties in iterative learning control. *Int J Control.* 2002;75(14):1114-1126.
17. Moore KL. *Iterative Learning Control for Deterministic Systems.* Berlin, Germany: Springer-Verlag; 2012.
18. Pintelon R, Schoukens J. *System Identification: A Frequency Domain Approach.* Hoboken, NJ: John Wiley & Sons; 2012.
19. Zou Q. Optimal preview-based stable-inversion for output tracking of nonminimum-phase linear systems. *Automatica.* 2009;45(1):230-237.
20. van Zundert J, Oomen T. On inversion-based approaches for feedforward and ILC. *Mechatronics.* 2018;50:282-291.
21. Oomen T, Rojas CR. Sparse iterative learning control with application to a wafer stage: achieving performance, resource efficiency, and task flexibility. *Mechatronics.* 2017;47:134-147.
22. Strijbosch N, Blanken L, Oomen T. Frequency domain design of iterative learning control and repetitive control for complex motion systems. Paper presented at: Proceedings of the IEEE International Workshop on Sensing, Actuation, Motion Control, and Optimization (SAMCON); 2018:1-2; Tokyo, Japan, IEEE.
23. Zhou K, Doyle JC, Glover K. *Robust and Optimal Control.* Vol 40. Englewood Cliffs, NJ: Prentice Hall; 1996.
24. De Roover D, Bosgra OH. Synthesis of robust multivariable iterative learning controllers with application to a wafer stage motion system. *Int J Control.* 2000;73(10):968-979.
25. Zheng M, Wang C, Sun L, Tomizuka M. Design of arbitrary-order robust iterative learning control based on robust control theory. *Mechatronics.* 2017;47:67-76.
26. van de Wijdeven J, Donkers T, Bosgra O. Iterative learning control for uncertain systems: robust monotonic convergence analysis. *Automatica.* 2009;45(10):2383-2391.
27. Ninness B, Goodwin GC. Estimation of model quality. *Automatica.* 1995;31(12):1771-1797.
28. Chen J, Gu G. *Control-Oriented System Identification: An H_{\infty} Approach.* New York, NY: John Wiley & Sons; 2000.
29. Blanken L, Oomen T. Multivariable iterative learning control design procedures: from decentralized to centralized, illustrated on an industrial printer. *IEEE Trans Control Syst Technol.* 2019;28(4):1534-1541.
30. Wu Y, Zou Q, Su C. A current cycle feedback iterative learning control approach for AFM imaging. *IEEE Trans Nanotechnol.* 2009;8(4):515-527.
31. Yan Y, Wang H, Zou Q. A decoupled inversion-based iterative control approach to multi-axis precision positioning: 3D nanopositioning example. *Automatica.* 2012;48(1):167-176.
32. Son TD, Pipeleers G, Swevers J. Robust monotonic convergent iterative learning control. *IEEE Trans Automat Control.* 2016;61(4):1063-1068.
33. Oomen T, Bosgra O. Estimating disturbances and model uncertainty in model validation for robust control. *IEEE; 2008:5513-5518.
34. Tempo R, Calafiore G, Dabbene F. *Randomized Algorithms for Analysis and Control of Uncertain Systems: With Applications.* Berlin, Germany: Springer Science & Business Media; 2012.
APPENDIX A. PROOF OF LEMMA 1

Proof. Under Assumptions 2 and 5, $Y_i = G_s U_i + D$. Substituting $E_i = R - Y_i$, and $Y_i$ into (1) yields

$$U_{i+1} = Q(U_i + LE_i) = Q(I - LG_s)U_i + QL(R - D), \quad (A1)$$

The fixed point $U_{\infty}$ follows by substituting $U_{i+1} = U_i = U_{\infty}$ and solving for $U_{\infty}$, which results in (6). Furthermore, it follows readily that $U_{i+1} - U_{\infty} = Q(I - LG_s)(U_i - U_{\infty})$. Hence, $||U_{i+1} - U_{\infty}||_2 \leq ||Q(I - LG_s)||_{2,2}||U_i - U_{\infty}||_2$. Consequently, monotonic convergence is guaranteed if $||Q(I - LG_s)||_{2,2} < 1$. This is guaranteed if (4) holds, in combination with Assumption 4. Furthermore, substituting $U_{\infty}$ into $E_{\infty} = R - GU_{\infty} - D$, results $E_{\infty} = S_s(R - D)$, with $S_s$ as given by (7). □

APPENDIX B. PROOF OF LEMMA 2

Proof. Let $||H(L, Q, \Delta)||_{2,2} = ||Q(I - LG(\Delta))||_{2,2} < \kappa \forall \Delta \in \Delta$, then $\tilde{\lambda}(Q(I - LG(\Delta))) < \kappa \forall \Delta \in \Delta$, since $\tilde{\lambda}(\cdot) \leq ||\cdot||_{2,2}$. Using that det $(I + A) = \prod_{i=1} \lambda_i(A)$ with $\lambda_i(A)$ the eigenvalues of $A^{2R(A.2.1)}$, it follows det $(I - Q(I - LG(\Delta))) = \prod_{i=1} (1 + \lambda_i(Q(I - LG(\Delta)))) \neq 0$, since $|\lambda_i(Q(I - LG(\Delta)))| \leq \kappa \forall \Delta \in \Delta$.
APPENDIX C. PROOF OF LEMMA 3

The following proof of Lemma 3 employs the structured singular value, which is defined as follows.\(^{46}\)

**Definition 4.** For \( M \in \mathbb{C}^{n \times n} \) and a set of structured uncertainties \( \Delta, \mu_\Delta(M) \) is defined as
\[
\mu_\Delta(M) \triangleq \frac{1}{\inf \{ \sigma(\Delta), \ \Delta \in \Delta, \ \det (I - M \Delta) = 0 \}} . \tag{C1}
\]

In case no \( \Delta \in \Delta \) makes \( I - M \Delta \) singular, \( \mu_\Delta(M) \triangleq 0. \)

**Proof.** First, \((14)\) is reformulated a \( \mu \)-test. To this end, define \( \mathcal{M}_\Delta \triangleq \{ \Delta \in \Delta | \sigma(\Delta) \leq 1 \} \) and note that \((14)\) is equivalent to
\[
\sup_{\Delta \in \mathcal{M}_\Delta} \| \frac{\nu}{x} A \Delta B \|_{2,2} < 1 . \tag{C2}
\]

Recall the main loop theorem\(^{46}(\text{thm4.3})\)
\[
\mu_\Delta(M) < 1 \iff \begin{cases} 
\mu_\Delta(M_{22}) < 1 \\
\sup_{\Delta \in \mathcal{M}_\Delta} \mu_\Delta(\mathcal{L}(M, \Delta_2)) < 1,
\end{cases} \quad \mathcal{L}(M, \Delta) \triangleq M_{11} + M_{12} \Delta(I - M_{22} \Delta)^{-1} M_{21} . \tag{C3}
\]

and for the case under consideration, define
\[
M = \begin{bmatrix} 0 & \frac{x}{\gamma} A & B \\ I & 0 & 0 \\ 0 & 0 & I \end{bmatrix}, \quad \tilde{\Delta} \triangleq \left\{ \begin{bmatrix} \Delta_1 & 0 \\ 0 & \Delta_2 \end{bmatrix} | \Delta_1 \in \Delta_1, \Delta_2 \in \Delta_2 \right\}, \quad \Delta_1 = \mathbb{C}^{n \times n}, \quad \Delta_2 = \Delta, \tag{C4}
\]

such that the main loop theorem \((C3)\) reduces to
\[
\mu_\Delta(M) < 1 \iff \begin{cases} 
\mu_\Delta(0) < 1 \\
\sup_{\Delta \in \mathcal{M}_\Delta} \| \frac{\nu}{x} A \Delta B \|_{2,2} < 1,
\end{cases} \tag{C5}
\]

where it is used that \( \mu_\Delta(\cdot) = \sigma(\cdot) = \| \cdot \|_{2,2} \) if \( \Delta = \mathbb{C}^{n \times n}.\)\(^{46}\) Note that \( \mu_\Delta(0) < 1, \) since by definition \( \mu_\Delta(0) = 0. \) Hence, \((C5)\) provides an equivalent \( \mu \)-test to \((C2).\) In turn, an upper bound to \( \mu_\Delta(M) \) is given by
\[
\mu_\Delta(M) \leq \inf_{D \in \mathcal{D}} \| D M D^{-1} \|_{2,2}, \quad \mathcal{D} \triangleq \left\{ D \in \mathbb{C}^{2n \times 2n} | D = \begin{bmatrix} I & 0 \\ 0 & dI \end{bmatrix}, \ . \ . \ . \right\} , \quad d \neq 0 \tag{C6}
\]

where in this case, the upper bound is met with equality, since \( \tilde{\Delta} \) consist of only two complex blocks.\(^{50}(\text{thm6})\) Relation \((C6)\) is reformulated as an LMI as in Reference 51
\[
\| D M D^{-1} \|_{2,2} < 1 \iff \| D^H D M D^{-1} - I \| < 0 \iff M^H D^H D M - D^H D < 0 \iff M^H X M - X < 0 , \tag{C7}
\]

where \( X = D^H D > 0, \) that is, \( X \triangleq \begin{bmatrix} I & 0 \\ 0 & xI \end{bmatrix} > 0, \ x > 0. \) Substituting \( M \) as given by \((C4)\) yields
\[
\begin{bmatrix} xB^H B - I & 0 \\ 0 & (\frac{\nu}{x})^2 A^H A - xI \end{bmatrix} < 0 , \tag{C8}
\]

which completes the proof. \( \blacksquare \)
APPENDIX D. PROOF OF LEMMA 4

Proof. Let \( a \geq 1 \) and consider the case \( 1 - \rho \geq 0 \). Then, \( \inf_{\rho}(|1 - \rho| + |\rho|\alpha) = \inf_{\rho}(1 + (\alpha - 1)\rho) = 1 \), since \((\alpha - 1) \geq 0\), which is attained for \( \rho = 0 \). Considering the case \( 1 - \rho \leq 0 \) results in \( \inf_{\rho}(|1 - \rho| + |\rho|\alpha) = \alpha \) for \( \rho = 1 \), which is suboptimal w.r.t. to the previous, such that \( \rho^* = 0 \). Similarly, let \( a < 1 \) and consider the case \( 1 - \rho \geq 0 \). Then \( \inf_{\rho}(|1 - \rho| + |\rho|\alpha) = \inf_{\rho}(1 + (\alpha - 1)\rho) = \alpha \), since \((\alpha - 1) < 0\), which is attained for \( \rho = 1 \). Considering the case \( 1 - \rho \leq 0 \) results in \( \inf_{\rho}(|1 - \rho| + |\rho|\alpha) = 1 \), which is identical to the previous, such that \( \rho^* = 1 \).

APPENDIX E. PROOF OF LEMMA 5

Proof. For \( A, B \in \mathbb{C}^{n \times n} \) with \( B \) invertible it holds that \( \det(\lambda I - AB) = \det(B) \det(\lambda I - AB) \det(B^{-1}) = \det(\lambda I - BA) \). Consequently, \( \rho(\Delta AB) = \rho(BA\Delta B^{-1}) = \rho(BA\Delta) \). Furthermore, \( \rho(A) \leq \|A\|_{2,2} \), for any square matrix \( A \). Hence, \( \sup_{\Delta \in \Delta} \rho(BA\Delta) \leq \sup_{\Delta \in \Delta} \|BA\Delta\|_{2,2} \leq \sup_{\Delta \in \Delta} \|\Delta\|_{2,2} \|BA\|_{2,2} \). To show that this is met with equality when \( \Delta \) is unstructured take \( \Delta^* = \gamma VU^H \), where \( U\Sigma V^H \) is the SVD of \( BA \). Consequently, \( \rho(BA\Delta^*) = \rho(\gamma \Sigma) = \sup_{\Delta \in \Delta} \|\Delta\|_{2,2} \|BA\|_{2,2} = \gamma \|A\|_{2,2} \), where it is used that \( B \) is unitary.

APPENDIX F. COMPUTING THE WORST-CASE PERFORMANCE

The asymptotic sensitivity estimate \( S(G_n^{-1}, Q, \Delta) \) as given by (12) can be written as \( S(G_n^{-1}, Q, \Delta) = \mathcal{L}(M, \Delta) \), where \( \mathcal{L}(M, \Delta) \) is defined by (C3), and \( M \) equals

\[
M = \begin{bmatrix}
G_n(I - Q)G_n^{-1} & G_n(Q(I - I)G_n^{-1}W_y \\
W_yQG_n^{-1} & -W_yQG_n^{-1}W_y
\end{bmatrix},
\]

as is readily verified by direct computation. This result facilitates the computation of the worst-case performance \( \sup_{\Delta \in \Delta} \|S(G_n^{-1}, Q, \Delta)\|_{2,2} \) by using the scaled main-loop theorem \( \text{thm4.8} \) in combination with Osborne’s method to evaluate the structured singular value \( \text{thm4.8} \).

APPENDIX G. PROOF OF LEMMA 6

Proof. For \( \omega_k \in \Omega_3 \), the columns of the matrix \( \tilde{V} \triangleq C^{-1/2}_V \) are circularly symmetric complex Gaussian distributed with identity covariance matrix and zero mean. Hence, \( W \triangleq \tilde{V}V^H \) is a complex Wishart matrix and the cumulative density function of its largest eigenvalue is given by (41) \( \text{eqn8} \). Since \( \overline{\sigma}(W) = \sqrt{\rho(\tilde{V}V^H)} \), the CDF of the largest singular value of \( \tilde{V} \) satisfies (42). Similarly, for \( \omega_k \in \Omega_2 \), \( W \triangleq \tilde{V}V^H \) is a real Wishart matrix, and the cumulative density function of its largest eigenvalue is given by (42). \( \text{eqn4} \)