2014

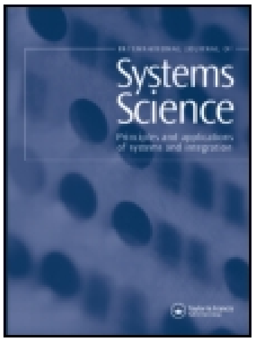
Robust noise attenuation under stochastic noises and worst-case unmodelled dynamics

Araz Hashemi

Ben G. Fitzpatrick
Loyola Marymount University, ben.fitzpatrick@lmu.edu

Le Yi Wang

George Yin



International Journal of Systems Science

ISSN: 0020-7721 (Print) 1464-5319 (Online) Journal homepage: https://www.tandfonline.com/loi/tsys20

Robust noise attenuation under stochastic noises and worst-case unmodelled dynamics

Araz Hashemi, Ben Fitzpatrick, Le Yi Wang & George Yin

International Journal of Systems Science, 2014
Vol. 45, No. 7, 1563–1578, http://dx.doi.org/10.1080/00207721.2014.916481

Robust noise attenuation under stochastic noises and worst-case unmodelled dynamics

Araz Hashemi[a], Ben Fitzpatrick[b,c], Le Yi Wang[d] and George Yin[e,*]

[a]*Department of Mathematics, Wayne State University, Detroit, MI, USA;* [b]*Department of Mathematics, Loyola Marymount University, Los Angeles, CA, USA;* [c]*Tempest Technologies, Los Angeles, CA, USA;* [d]*Department of Electrical and Computer Engineering, Wayne State University, Detroit, MI, USA;* [e]*Department of Mathematics, Wayne State University, Detroit, MI, USA*

(*Received 20 June 2013; accepted 23 September 2013*)

This paper investigates noise attenuation problems for systems with unmodelled dynamics and unknown noise characteristics. A unique methodology is introduced that employs signal estimation in one phase, followed by control design for noise rejection. The methodology enjoys certain advantages in its simple control design process, accommodation of unmodelled dynamics, and non-conservative noise rejection performance. Under mild information on unmodelled dynamics, we first derive robust performance bounds on noise attenuation with respect to unmodelled dynamics without noise estimation errors. Then more general results are presented for systems that are subject to both stochastic signal estimation errors and unmodelled dynamics. Examples are also presented to demonstrate our findings.

Keywords: noise attenuation; unmodelled dynamics; signal estimation

1. Introduction

This paper investigates noise attenuation problems for systems with unmodelled dynamics, which are central in many feedback control problems. For instance, in regulation problems, output deviations from the set values must be controlled to ensure control qualities. This is evidenced by motor speed control, voltage regulation for high-precision sensing devices, thickness control in pulp-paper industry, satellite position control for global monitoring systems, doping control in semiconductor industry, etc.

Fundamental research on noise attenuation problems can be traced back to the 1940s when extensive research was conducted and fundamental progress was made, especially on the well-known Wiener filters (see Wiener, 1949) in the continuous-time domain and Kolmogorov filters in the discrete-time domain. Later, such filters were further extended to Kalman filters by employing state space model structures (see Anderson & Moore, 1979), and to more general recursive structures in adaptive filters and stochastic approximation algorithms (Kushner & Yin, 2003). Noise attenuation in a feedback control setting is inherently a nonlinear structure since operator inverses are involved. However, by employing the Youla parametrisation (see Youla, Jabri, & Bongiorno, 1976a), all stabilising controllers in linear time invariant (LTI) systems are parametrised by a stable operator in an affine structure; as such the design problem is reduced to solutions to an integral equation. Consequently, the design problem can be solved by using the Wiener–Hopf method, especially in the frequency domain (see Youla et al., 1976a; Youla, Jabri, & Bongiorno, 1976b); see also related works of Nguyen, Veselya, and Rosinova (2013)and Zhao, Liu, and Wang (2013)). Usually, these classical filtering theories do not consider model uncertainties. In that sense, they are not robust.

Robust noise attenuation problems drew intensive attention in the 1980s in the framework of robust control and H^∞ sensitivity minimisation (Zames, 1981). It is in that framework the unmodelled dynamics have become a key component in noise attenuation problems. H^∞ sensitivity optimisation relies on noise characterisation information. That information is described as a weighting function in its minimisation problem. When this information is not available, modified methods must be developed. In addition, computational complexity in H^∞ is higher than the traditional Wiener filters and Kalman filters, which are projection operators and easy to compute. Also, due to its worst-case design foundation, H^∞ designs are often conservative since they select controllers that attenuate all possible worst-case noises in a large class, rather than a special noise as in Wiener and Kalman filters.

This paper introduces a methodology to address some essential issues from these approaches. First, the process employs noise estimation so that noise characteristics do not need to be predetermined. The estimates are then used in control design to minimise output mean-square errors. Since estimated noises are used in the design process, its algorithms are of least-squares (LS) structures and, hence, are computationally very efficient. On the other hand, to

*Corresponding author. Email: gyin@math.wayne.edu

accommodate unmodelled dynamics, we analyse and derive error bounds due to unmodelled dynamics and noise estimation errors. As such, robustness is embedded in our methodology.

In some limited aspects, this paper is related to signal estimation, system identification, and adaptive controls, which have received considerable attention in the control and systems theory literature. There is a vast literature; see for example, Chen and Guo (1991), He, Wang, and Yin (2013), Ljung (1987), Milanese and Belforte (1982), Milanese and Vicino (1991), Milanese and Vicino (1993), Poolla and Tikku (1994), Sayeed (2003), Wiener (1949), Wang, Yin, Zhang, and Zhao (2010), Zames (1981), and references therein. Consistency of parameter estimates and optimality of adaptive design have been well studied. Many important issues have been investigated. In reference to these developments, this paper examines the problems from a different angle. We focus on noise attenuation for systems involving stochastic noise and unmodelled dynamics. Our approach includes both worst-case analysis that is of truly deterministic feature as well as stochastic analysis for noisy systems.

The rest of the paper is arranged as follows. Section 2 describes the control design process for optimal noise attenuation. The well-known Youla parametrisation (Youla et al., 1976a) is employed to transform the feedback design problem into a signal matching problem. A two-phase design process is introduced in which noise estimation is followed by a control design that is much simplified from H^∞ minimisation. Section 3 focuses on noise-free systems that involve unmodelled dynamics. Here the worst-case analysis techniques are used to derive explicit bounds on the impact of unmodelled dynamics. These results are of value for selecting model orders to ensure noise attenuation performance requirements. Section 4 expands to more complex system environments that involve both unmodelled dynamics and signal estimation errors. Error bounds are derived. Simulation examples are presented to demonstrate the issues discussed in this paper. For conciseness and clarity, more extensive proofs are placed in an appendix at the end of the paper. Finally, Section 5 summarises the main findings and highlights a few open issues.

2. Optimal noise attenuation: controller parametrisation and optimisation

We begin with a regulation problem under the LTI plant P and controller F in Figure 1. For convenience of algorithm development, we will work in the discrete-time domain. The output x is to be controlled to follow the constant reference value x_r. The system output is subjected to stochastic disturbance d. Since the system is LTI, the output can be expressed in its transfer function form as

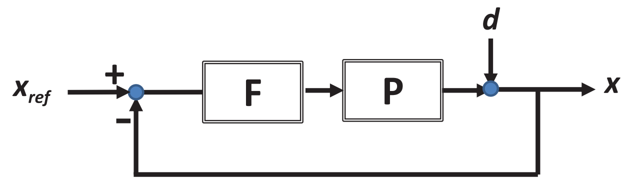


Figure 1. The original regulation problem.

$$X(z) = \frac{F(z)P(z)}{1+F(z)P(z)}X_r(z) + \frac{1}{1+F(z)P(z)}D(z)$$
$$= U(z) + \frac{1}{1+F(z)P(z)}D(z),$$

where the systems are represented by their z-transfer functions and the signals by their z transforms, and $U(z) = \frac{F(z)P(z)}{1+F(z)P(z)}X_r(z)$. x is measured. Denote $y_k = x_k - x_r$. Since x_k is measured and x_r is known, y_k is also a measured signal. Then,

$$Y(z) = (U(z) - X_r(z)) + \frac{1}{1+F(z)P(z)}D(z).$$

The first term $U(z) - X_r(z)$ is deterministic and the second term is stochastic.

If the controller F is stabilising and the system is at least of type 1 (including at least one integrator in the forward path), then the first term converges to zero exponentially fast. Since this is a very fast transient and our interest here is in noise rejection in a persistent sense, we will mandate a stabilising controller in our design and then ignore this term in our analysis on noise attenuation. As a result, we will focus on

$$Y(z) = \frac{1}{1+F(z)P(z)}D(z),$$

which can be represented by the diagram in Figure 2. Our goal is to attenuate the impact of the noise d on the output y. For simplicity, assume that P is an exponentially stable system.

Since the transfer function $P(z)$ is exponentially stable, we may represent it by a finite impulse response (FIR) filter

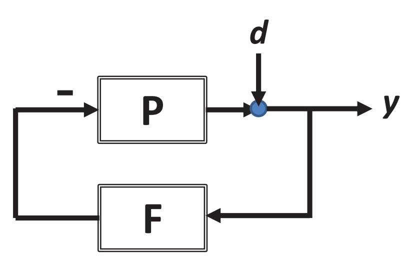


Figure 2. A basic feedback configuration for noise attenuation.

$P_0(z)$ (the modelled part), plus an unmodelled dynamics δ,

$$P(z) = P_0(z) + \delta(z). \tag{1}$$

More precisely,

$$P(z) = p_0 + p_1 z^{-1} + \cdots + p_n z^{-n} + \delta(z), \tag{2}$$

where $\delta(z) = \sum_{j=n+1}^{\infty} p_j z^{-j}$ and $\sum_{j=n+1}^{\infty} |p_j| \le \varepsilon_n$. Due to exponential stability, $|\varepsilon_n| \le \kappa \lambda^n$ for some $\kappa > 0$ and $0 < \lambda < 1$, namely, it is an exponentially decaying function with respect to n. An immediate implication of this is that for a given required bound ε on the modelling error, a model order n (model complexity) can be predetermined such that $\varepsilon_n \le \varepsilon$. In our subsequent results, all bounds due to unmodelled dynamics should be interpreted as a function of model complexity n.

2.1. Parametrisation of stabilising controllers

The following parametrisation of stabilising controllers is the Youla parametrisation (Youla et al., 1976a). In the special case of stable plants, it is called Q parametrisation (Francis, 1987; Francis & Zames, 1984).

Let $\mathbb{S}$ represents the space of exponentially stable systems. For internal stability, the closed-loop systems,

$$\frac{1}{1+FP}, \quad \frac{F}{1+FP}, \quad \frac{P}{1+FP}, \text{ and } \quad \frac{FP}{1+FP}$$

must all be (exponentially) stable, that is, belong to $\mathbb{S}$. Denote

$$Q = \frac{F}{1+FP} \in \mathbb{S}. \tag{3}$$

Since $P \in \mathbb{S}$, if $Q \in \mathbb{S}$, we have

$$\begin{aligned} PQ &= \frac{FP}{1+FP} \in \mathbb{S} \\ &\Rightarrow \frac{1}{1+FP} = 1 - \frac{FP}{1+FP} \in \mathbb{S} \\ &\Rightarrow \frac{P}{1+FP} \in \mathbb{S}. \end{aligned} \tag{4}$$

Thus, the stability requirement is satisfied if we choose $Q \in \mathbb{S}$, and design

$$F = \frac{Q}{1-QP}. \tag{5}$$

This implies that F in Figure 2 can be implemented by using this Q parametrisation as shown in Figure 3. Note that a positive feedback is used due to the presence of $1 - QP$ in the expression for F.

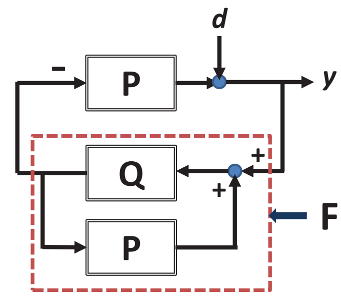


Figure 3. Feedback controller using the Q parametrisation.

Let $Y(z)$ and $D(z)$ be the Laplace transforms of the output y and disturbance d, respectively. Then we have

$$\begin{aligned} Y(z) &= \frac{1}{1+FP} D(z) = \left(1 - \frac{FP}{1+FP}\right) D(z) \\ &= (1 - QP)\, D(z). \end{aligned}$$

Let $W(z) = P(z)D(z)$. Then $Y(z) = D(z) - QW(z)$. From

$$D(z) = d_0 + d_1 z^{-1} + \cdots; \quad W(z) = w_0 + w_1 z^{-1} + \cdots,$$

we obtain the recursive representation,

$$y_k = d_k - Q * w_k.$$

Suppose that Q is an FIR of order m. Then

$$\begin{aligned} y_k &= d_k - (q_0 w_k + q_1 w_{k-1} + \cdots + q_m w_{k-m}) \\ &= d_k - [w_k, w_{k-1}, \ldots, w_{k-m}][q_0, q_1, \ldots, q_m]' \\ &= d_k - \phi_k' \theta, \end{aligned} \tag{6}$$

with $\phi_k' = [w_k, w_{k-1}, \ldots, w_{k-m}]$. Note that

$$\begin{aligned} w_k &= \sum_{j=0}^{\infty} p_j d_{k-j} \\ &= \sum_{j=0}^{n} p_j d_{k-j} + \sum_{j=n+1}^{\infty} p_j d_{k-j} \\ &= [d_k, d_{k-1}, \ldots, d_{k-n}]p + [d_{k-(n+1)}, \ldots]p^* \\ &= \psi_k' p + \widetilde{\psi}_k' p^*, \end{aligned} \tag{7}$$

where $p = [p_0, \ldots, p_n]'$ represents the modelled part of the plant and $p^* = [p_{n+1}, p_{n+2}, \ldots]'$ represents the unmodelled dynamics, and $\psi_k' = [d_k, d_{k-1}, \ldots, d_{k-n}]$, $\widetilde{\psi}_k' = [d_{k-(n+1)}, \ldots]$.

Assumption 1: (1) *d_k is estimated by $\widehat{d_k} = d_k + e_k$. e_k is stationary, $Ee_k = 0$, $Ee_k^2 \le \sigma^2 < \infty$.* (2) *The modelled part p is known. The unmodelled dynamics p^* has a uniform norm bound ε_n.*

Under Assumption 1, we have

$$\widehat{\psi}'_k = [\widehat{d}_k, \widehat{d}_{k-1}, \ldots, \widehat{d}_{k-n}] = [d_k + e_k, d_{k-1} + e_{k-1}, \ldots, d_{k-n} + e_{k-n}] = \psi'_k + \xi'_k,$$

where $\xi'_k = [e_k, e_{k-1}, \ldots, e_{k-n}]$. It follows that

$$w_k = \psi'_k p + \widetilde{\psi}'_k p^* = \widehat{\psi}'_k p - \xi'_k p + \widetilde{\psi}'_k p^* = \widehat{w}_k + \widetilde{\varepsilon}_k, \tag{8}$$

where

$$\widehat{w}_k = \widehat{\psi}'_k p, \quad \widetilde{\varepsilon}_k = -\xi'_k p + \widetilde{\psi}'_k p^*.$$

Consequently,

$$\begin{aligned} y_k &= d_k - [w_k, w_{k-1}, \ldots, w_{k-m}][q_0, q_1, \ldots, q_m]' \\ &= \widehat{d}_k - e_k - [\widehat{w}_k + \widetilde{\varepsilon}_k, \widehat{w}_{k-1} + \widetilde{\varepsilon}_{k-1}, \ldots, \widehat{w}_{k-m} + \widetilde{\varepsilon}_{k-m}] \\ &\quad \times [q_0, q_1, \ldots, q_m]' \\ &= \widehat{d}_k - e_k - \widehat{\phi}'_k\theta - \zeta'_k\theta, \end{aligned}$$

where $\widehat{\phi}'_k = [\widehat{w}_k, \widehat{w}_{k-1}, \ldots, \widehat{w}_{k-m}]$ and $\zeta'_k = [\widetilde{\varepsilon}_k, \widetilde{\varepsilon}_{k-1}, \ldots, \widetilde{\varepsilon}_{k-m}]$. It is observed that ζ_k is affected by both the unmodelled dynamics and signal estimation errors.

For estimation, after N observations the available regression data are

$$\widehat{D}_N = \begin{bmatrix} \widehat{d}_1 \\ \vdots \\ \widehat{d}_N \end{bmatrix}; \quad \widehat{\Phi}_N = \begin{bmatrix} \widehat{\phi}'_1 \\ \vdots \\ \widehat{\phi}'_N \end{bmatrix}.$$

In a nominal-system-based design procedure, the control parameter Q is then designed by

$$\theta_N = \left(\widehat{\Phi}'_N \widehat{\Phi}_N\right)^{-1} \widehat{\Phi}'_N \widehat{D}_N.$$

If we define

$$\Phi_N = \begin{bmatrix} \phi'_1 \\ \vdots \\ \phi'_N \end{bmatrix}; \quad \Xi_N = \begin{bmatrix} \zeta'_1 \\ \vdots \\ \zeta'_N \end{bmatrix}; \quad E_N = \begin{bmatrix} e_1 \\ \vdots \\ e_N \end{bmatrix},$$

then

$$\Phi_N = \widehat{\Phi}_N + \Xi_N \tag{9}$$

and

$$Y_N = \widehat{D}_N - E_N - \widehat{\Phi}_N\theta - \Xi_N\theta.$$

These will be useful for error analysis.

2.2. *Two-phase mechanism for signal estimation and noise rejection*

In this section, we discuss the signal estimation aspect of our noise attenuation scheme, which is based on sample paths. This idea differs from the classical Wiener filter design (Wiener, 1949) or H^∞ sensitivity minimisation (Doyle, Francis, & Tannenbaum, 1990; Zames, 1981). We first elaborate on some aspects of available signals for control design. Although y_k is measured, usually d_k is not available. However, the following observations will explain why certain signals can be approximately extracted for control design.

We note that after a controller F is designed and implemented, if the design is successful the output y_k will be small due to the rejection of disturbance by the feedback system. Under this environment, y_k contains nearly no information that can be used for control design. This phase will be called 'noise rejection phase'.

However, suppose that the disturbance d_k is stationary and its power spectrum density is limited in certain frequency bands. As a result, there exists an open-loop causal and stable filter $H(z)$ such that $H(z)D(z) \approx 0$ in a certain sense. $H(z)$ is an annihilating filter for d_k. It follows that if such a filter is inserted into the feedback loop in Figure 2 for a period of time, shown in Figure 4, the plant output v_k will be

$$V(z) = \frac{FP}{1 + FPH} HD \approx 0.$$

It follows that during that period, the signal d_k can be estimated by

$$y_k \approx d_k.$$

Consequently, in the following control design, we should use the available signal,

$$y_k = \widehat{d}_k = d_k + e_k, \tag{10}$$

where $\widehat{d}_k$ is an estimate of d_k with estimation error e_k. This phase will be called 'signal estimation phase'.

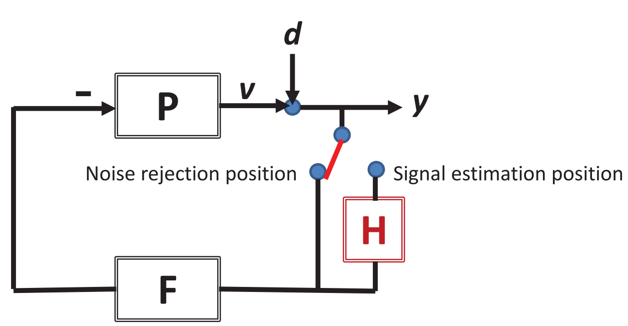


Figure 4. Modified feedback configuration by using annihilating filters for signal estimation.

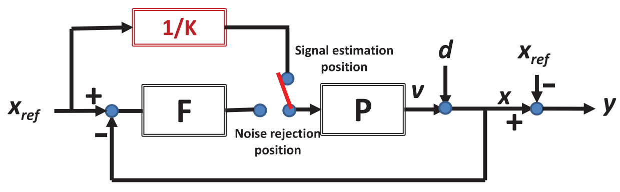


Figure 5. Using open-loop control for signal estimation when the plant is stable.

The approach of using annihilating filters can potentially work for unstable plants. On the other hand, for stable plants, a simple and general open-loop scheme works for the signal estimation phase. Suppose that $P(z)$ is stable. Let the direct current (DC) gain of the plant be $K = P(1)$. Then, by switching to the open-loop control illustrated in Figure 5, the actual output v_k of the plant (which is, however, not known) is a deterministic signal and converges to x_r exponentially. Consequently, the measured y_k becomes d_k after an exponentially fast convergent transient. This approach does not require any prior information on d_k. In this case, we also have $y_k = \widehat{d}_k = d_k + e_k$.

In this two-phase approach illustrated in Figure 6, control design is performed during the signal estimation phase. As a result, in the following algorithms, $\widehat{d}_k$ will be available in control design. The impact of signal estimation error e_k on noise rejection will be analysed.

Example 1: For an example of this approach, we consider a plant with transfer function, before sampling, $\frac{1}{s+2}$. The DC gain of this plant is 0.5. As a result, the open-loop controller is $K = 2$. Suppose that the disturbance is $d_k = a_k \sin(200\tau k)$, where τ is the sampling interval and $\tau = 0.001$. a_k is independent and identically distributed (i.i.d.) and uniformly distributed in $[-5, 5]$. Now, in the first two seconds, we run this system open loop. Then in $t \in [2, 10]$, we switch on the feedback controller which is a high-gain feedback $F = 20, 000$. The trajectories of the disturbance d_k and the targeted y_k are shown in Figure 7 .

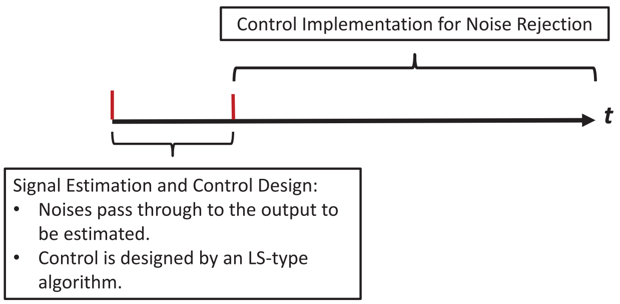


Figure 6. The two-phase design.

2.3. *Noise attenuation problems*

We will introduce the following two design procedures.

(1) **Nominal design procedure**: Suppose that d_k is unknown but can be estimated by $\widehat{d}_k$ with error e_k.

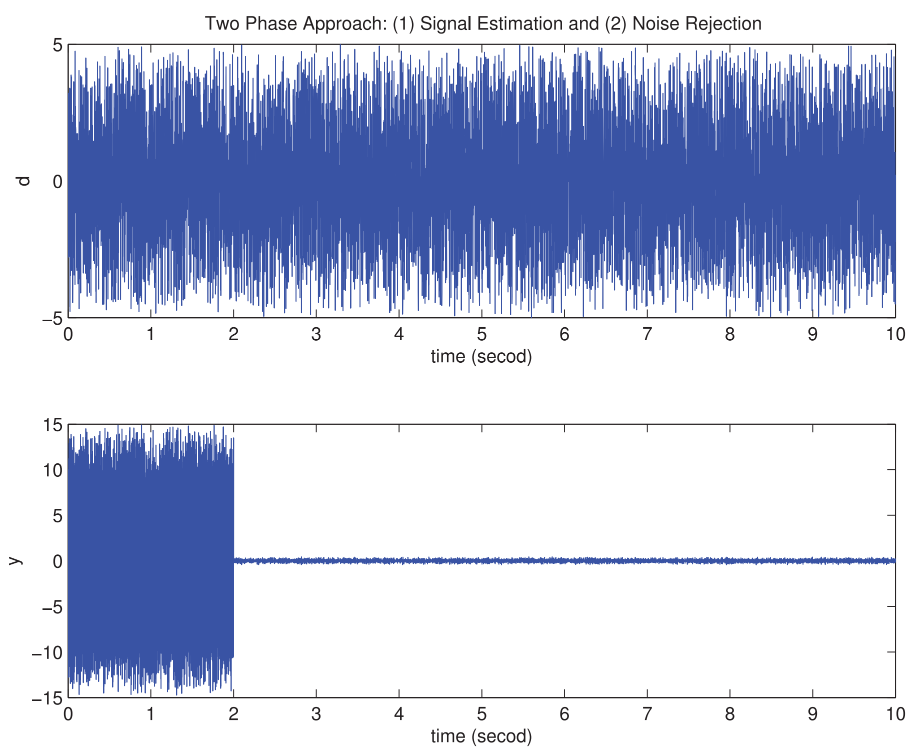


Figure 7. The example on the two-phase design.

The nominal plant P_0 is known. The unmodelled dynamics $\Delta \neq 0$ and its norm bound is ε_n. The control parameter Q is designed by

$$\theta_N = (\widehat{\Phi}'_N \widehat{\Phi}_N)^{-1} \widehat{\Phi}'_N \widehat{D}_N. \tag{11}$$

(2) **Robust design procedure**: As an alternative design, we will also consider a modified design for Q. In Equation (11), Q is designed on the basis of

$$\min_Q (\widehat{D}_N - \widehat{\Phi}_N \theta)'(\widehat{D}_N - \widehat{\Phi}_N \theta).$$

Here, the design criterion is modified to

$$\min_Q \max_{\Delta \in \Gamma} \left(\widehat{D}_N - (\widehat{\Phi}_N + \Delta)\theta\right)' \left(\widehat{D}_N - (\widehat{\Phi}_N + \Delta)\theta\right),$$

where Γ is the bounded set for Ξ_N.

Remark 1: We now comment on our approaches.

(1) This problem involves both worst-case and stochastic uncertainties and, hence, is of hybrid type. The error bounds derived in the paper reflect this nature with both worst-case components and probability bounds.
(2) Noise attenuation and sensitivity minimisation problems have been very extensively studied in the past, most notably in classical filter designs of Wiener–Hopf types and more advanced robust control in H^∞ and other related areas. To understand our approach, we point out that Wiener filters are based on a fixed and known noise spectrum, and the filter is optimally designed on the basis of this prior information. To overcome this drawback, the H^∞ sensitivity minimisation aims to reject robustly a class of noises with the prior knowledge of their spectrum bounds. While the Wiener filter is an orthogonal projection and hence easy to design, H^∞ design is more involved. Also, by targeting a class of noises, the H^∞ design becomes naturally more conservative.
Our approach does not require prior knowledge on the noise spectrum. By using an open-loop control for a small period of time, we observe the noise partially and temporarily. Then the controller is designed based on the actual noise. Since the noise is stationary, such designed controller can achieve targeted noise rejection for the specific noise class and, therefore, less conservative than the H^∞ design. By using the signal-based least-squares design, it is computationally much easier than the H^∞ design. In comparison to the Wiener filter, our approach does not require prior knowledge on the noise spectrum. Rather, the noise information is acquired in real time and used immediately in control design.
(3) Naturally, one may suggest the following two-step design. When the noise is partially observed, one can first estimate the noise spectrum. Then in the second step, one designs a Wiener filter on the basis of the estimated spectrum. This approach can be viewed as an 'indirect' design to borrow the term from adaptive control. This approach is, however, more cumbersome than our approach. Our approach may be viewed as a 'direct' design in which the available data are used directly in control design.
(4) Our design is a least-squares type of algorithm. As a result, it is simple and easily recursified. Furthermore, a rich literature on LS algorithms will support further extension of our approach to accommodate other practical issues in noise rejection problems.
(5) For performance analysis, we distinguish the following two stages. In the learning stage (in the signal estimation phase), the signal is estimated and controller is designed. In this case, $\widehat{d}_k$ is known approximately and used in design. In the controller execution phase, y_k is the output and the disturbance will be different from the first case but will have the same stochastic properties as in the first phase. Consequently, the designed controller will achieve noise attenuation with desirable performance.

3. Unmodelled dynamics and robust noise attenuation

In this section, we analyse impact of unmodelled dynamics and investigate suitable control design that can attenuate noise effects on the system output. For clarity, we will concentrate on unmodelled dynamics only in this section. Hence, $e_k \equiv 0$ in this section. As a result, some previous expressions are simplified to

$$\widetilde{d}_k = d_k,\ \widehat{\psi}'_k = \psi'_k = [d_k, \ldots, d_{k-n}],\ \xi'_k = 0,$$
$$w_k = \psi'_k p + \widetilde{\psi}'_k p^*.$$

The observation equation is simplified to

$$Y_N = D_N - (\widehat{\Phi}_N + \Xi_N)\theta,$$

where

$$\widehat{\Phi}_N = \begin{bmatrix} \widehat{\phi}'_1 \\ \vdots \\ \widehat{\phi}'_N \end{bmatrix},\ \Xi_N = \begin{bmatrix} \zeta'_1 \\ \vdots \\ \zeta'_N \end{bmatrix}$$

and $\widehat{\phi}'_k = [\psi'_k p, \psi'_{k-1} p, \ldots, \psi'_{k-m} p]$ and $\zeta'_k = [\widetilde{\psi}'_k p^*, \widetilde{\psi}'_{k-1} p^*, \ldots, \widetilde{\psi}'_{k-m} p^*]$.

The set of uncertainty for Ξ_N will be denoted by Γ, which is the set that includes all possible unmodelled dynamics p^* whose norm is bounded by ε_n.

3.1. Nominal design

Without signal estimation errors, in the design phase, d_k is correctly estimated. The nominal plant P_0 is known, so p is known. The nominal design ignores the unmodelled dynamics in the design consideration. Hence, it minimises

$$\min_{\theta_N}(D_N - \widehat{\Phi}_N\theta_N)'(D_N - \widehat{\Phi}_N\theta_N),$$

and the resulting control parameter θ is

$$\theta_N = \left(\widehat{\Phi}'_N\widehat{\Phi}_N\right)^{-1}\widehat{\Phi}'_N D_N. \quad (12)$$

For performance analysis, we consider the residual of noise attenuation when θ_N from Equation (12) is used,

$$\begin{aligned}
&\mu_N(\Xi_N, D_N)\\
&= \frac{1}{N}(D_N - (\widehat{\Phi}_N + \Xi_N)\theta_N)'(D_N - (\widehat{\Phi}_N + \Xi_N)\theta_N)\\
&= \frac{1}{N}(D_N - (\widehat{\Phi}_N + \Xi_N)\left(\widehat{\Phi}'_N\widehat{\Phi}_N\right)^{-1}\widehat{\Phi}'_N D_N)'\\
&\quad \times (D_N - (\widehat{\Phi}_N + \Xi_N)\left(\widehat{\Phi}'_N\widehat{\Phi}_N\right)^{-1}\widehat{\Phi}'_N D_N)\\
&= \frac{1}{N}D'_N(I - (\widehat{\Phi}_N + \Xi_N)\left(\widehat{\Phi}'_N\widehat{\Phi}_N\right)^{-1}\widehat{\Phi}'_N)'\\
&\quad \times (I - (\widehat{\Phi}_N + \Xi_N)\left(\widehat{\Phi}'_N\widehat{\Phi}_N\right)^{-1}\widehat{\Phi}'_N)D_N\\
&= \frac{1}{N}D'_N\Pi'(D_N, \Xi_N)\Pi(D_N, \Xi_N)D_N,
\end{aligned}$$

where

$$\Pi(D_N, \Xi_N) = I - (\widehat{\Phi}_N + \Xi_N)\left(\widehat{\Phi}'_N\widehat{\Phi}_N\right)^{-1}\widehat{\Phi}'_N,$$

whose dependence on D_N stems from the fact that $\widehat{\Phi}_N$ depends on D_N. Then, the worst-case performance is

$$\mu_N(D_N) = \max_{\Xi_N\in\Gamma}\mu(\Xi_N, D_N).$$

The dependence of $\widehat{\Phi}_N$ on D_N can be explicitly derived. First, the vector $\widehat{w} = [\widehat{w}_1, \ldots, \widehat{w}_N]'$ can be expressed as

$$\widehat{w} = H_p D_N,$$

where H_p is the Hankel matrix,

$$H_p = \begin{bmatrix} p_0 & 0 & \cdots & 0 & 0\\ p_1 & p_0 & \cdots & 0 & 0\\ \vdots & & & & \vdots\\ p_n & p_{n-1} & \cdots & 0 & 0\\ 0 & p_n & \cdots & 0 & 0\\ \vdots & & & & \vdots\\ 0 & 0 & \cdots & p_0 & 0\\ 0 & 0 & \cdots & p_1 & p_0 \end{bmatrix}.$$

Now, define the $p \times p$ shifting matrix,

$$G\begin{bmatrix} x_1\\ x_2\\ \vdots\\ x_{p-1}\\ x_p \end{bmatrix} = \begin{bmatrix} 0\\ x_1\\ \vdots\\ x_{p-2}\\ x_{p-1} \end{bmatrix},$$

namely

$$G = \begin{bmatrix} 0 & 0 & \cdots & 0\\ 1 & 0 & \cdots & 0\\ \vdots & & & \vdots\\ 0 & \cdots & 1 & 0 \end{bmatrix}.$$

Then,

$$\widehat{\Phi}_N(D_N) = [I, G, \ldots, G^{m-1}](I_m \otimes H_p D_n),$$

where $\otimes$ is the Kronecker product.

We will impose the following assumption on D_N.

Assumption 2:

$$D_N \in M_D = \{\|D_N/\sqrt{N}\|_2 \le \sigma^2\}.$$

This is a sample-path version of disturbances variances being bounded by σ^2.

The disturbance attenuation performance is defined in the worst-case sense as

$$\mu = \max_{D_N\in M_D}\mu_N(D_N).$$

To obtain bounds on μ, we first normalise the signal. Let $\|D_N/\sqrt{N}\|_2 = \lambda$ and define $v_N = \frac{D_N/\sqrt{N}}{\lambda}$ with $\|v_N\|_2 = 1$. For $D_N \in M_D$, $\lambda \le \sigma^2$. Then,

$$\begin{aligned}
\widehat{\Phi}_N(D_N) &= [I, G, \ldots, G^{m-1}](I_m \otimes H_p D_n)\\
&= \sqrt{N}\lambda[I, G, \ldots, G^{m-1}](I_m \otimes H_p v_n)\\
&= \sqrt{N}\lambda\widehat{\Phi}_N(v_N).
\end{aligned}$$

Denote $\sigma_{\min}$ as the smallest singular value of a matrix, and

$$b_{\min} = \min_{\|v_N\|_2=1} \sigma_{\min}(\widehat{\Phi}_N(v_N)).$$

Due to normalisation, $b_{\min}$ is independent of the size of D_N. Also, denote

$$f(\varepsilon_N) = \max_{\Xi_N \in \Gamma} \frac{\|\Xi_N\|}{\sqrt{N}},$$

where ε_N is the bound on unmodelled dynamics.

Theorem 1:

$$\mu \leq \frac{f(\varepsilon_N)}{b_{\min}}.$$

Proof:

$$\begin{aligned}
&\frac{1}{N} D'_N \Pi'(D_N, \Xi_N)\Pi(D_N, \Xi_N)D_N \\
&\quad = v'_N(\lambda\Pi(D_N, \Xi_N))'(\lambda\Pi(D_N, \Xi_N))v_N,
\end{aligned}$$

in which

$$\begin{aligned}
\lambda\Pi(D_N, \Xi_N) &= \lambda(I - (\widehat{\Phi}_N(D_N) + \Xi_N) \\
&\quad \times \left(\widehat{\Phi}'_N(D_N)\widehat{\Phi}_N(D_N)\right)^{-1} \widehat{\Phi}'_N(D_N)) \\
&= \lambda(I - (\sqrt{N}\lambda\widehat{\Phi}_N(v_N) + \Xi_N) \\
&\quad \times \frac{1}{\sqrt{N}\lambda} \left(\widehat{\Phi}'_N(v_N)\widehat{\Phi}_N(v_N)\right)^{-1} \widehat{\Phi}'_N(v_N)) \\
&= \lambda\Bigg(I - \left(\widehat{\Phi}_N(v_N) + \frac{\Xi_N}{\sqrt{N}\lambda}\right) \\
&\quad \times \left(\widehat{\Phi}'_N(v_N)\widehat{\Phi}_N(v_N)\right)^{-1} \widehat{\Phi}'_N(v_N)\Bigg) \\
&:= \widehat{\Pi}(v_n, \Xi_N).
\end{aligned}$$

Now,

$$\begin{aligned}
\mu &= \max_{D_N \in M_D} \mu_N(D_N) \\
&= \max_{\|v_N\|_2=1} \max_{\Xi_N \in \Gamma} v'_N \widehat{\Pi}(v_n, \Xi_N)' \widehat{\Pi}(v_n, \Xi_N) v_N.
\end{aligned}$$

It follows that

$$\mu \leq \max_{\|v_N\|_2=1} \max_{\Xi_N \in \Gamma} \|\widehat{\Pi}(v_n, \Xi_N)\|,$$

where $\|\cdot\|$ is the largest singular value.

Since

$$\begin{aligned}
\widehat{\Pi}(v_N, \Xi_N)\widehat{\Phi}_N(v_N) &= \lambda\Bigg(I - \left(\widehat{\Phi}_N(v_N) + \frac{\Xi_N}{\sqrt{N}\lambda}\right) \\
&\quad \times \left(\widehat{\Phi}'_N(v_N)\widehat{\Phi}_N(v_N)\right)^{-1} \widehat{\Phi}'_N(v_N)\Bigg) \\
&\quad \times \widehat{\Phi}_N(v_N) \\
&= \lambda\left(\widehat{\Phi}_N - \left(\widehat{\Phi}_N + \frac{\Xi_N}{\sqrt{N}\lambda}\right)\right) \\
&= -\frac{\Xi_N}{\sqrt{N}},
\end{aligned}$$

we have

$$\|\widehat{\Pi}(v_N, \Xi_N)\widehat{\Phi}_N(v_N)\| = \frac{\|\Xi_N\|}{\sqrt{N}} \leq \max_{\Xi_N \in \Gamma} \frac{\|\Xi_N\|}{\sqrt{N}} = f(\varepsilon_N). \tag{13}$$

Now, from Equation (13) and

$$\begin{aligned}
\|\widehat{\Pi}(v_N, \Xi_N)\widehat{\Phi}_N(v_N)\| &\geq \|\widehat{\Pi}(v_N, \Xi_N)\|\sigma_{\min}(\widehat{\Phi}_N(v_N)) \\
&\geq \|\widehat{\Pi}(v_N, \Xi_N)\|b_{\min},
\end{aligned}$$

we obtain

$$\|\widehat{\Pi}(v_N, \Xi_N)\| \leq \frac{f(\varepsilon_N)}{b_{\min}}.$$

Therefore,

$$\mu \leq \frac{f(\varepsilon_N)}{b_{\min}}. \qquad \square$$

3.2. *Robust design*

Theoretically, robust noise attenuation for systems with unmodelled dynamics employs the performance index,

$$\begin{aligned}
\eta_N(D_N, \theta_N) &= \frac{1}{N} \max_{\Xi_N \in \Gamma} (D_N - (\widehat{\Phi}_N + \Xi_N)\theta_N)' \\
&\quad \times (D_N - (\widehat{\Phi}_N + \Xi_N)\theta_N),
\end{aligned}$$

and seeks

$$\eta_N(D_N) = \min_{\theta_N} \eta_N(D_N, \theta_N),$$

and the optimal θ_N is denoted by θ_N^*.

The difference between the nominal design and robust design is that the former is a 'max–min' design in which the design is done first; and the latter is a 'min—max' design. As a consequence,

$$\eta_N(D_N) \leq \mu_N(D_N),$$

indicating a potential performance improvement in the worst-case sense. It is well known that the 'min–max' often leads to non-linear and non-quadratic optimisation problems and is usually more complicated. Often only numerical solutions are feasible. We now introduce some numerical algorithms.

The gradient of $\eta_N(D_N, \theta_N)$ with respect to θ_N is

$$G(D_N, \theta_N) = \frac{\partial \eta_N(D_N, \theta_N)}{\partial \theta_N} = \frac{2}{N} \max_{\Xi_N \in \Gamma} (\widehat{\Phi}_N + \Xi_N)'(D_N - (\widehat{\Phi}_N + \Xi_N)\theta_N).$$

The following gradient-based searching algorithm is used.

Searching Algorithm for θ_N^*:

- *Initial value.*
  The initial value θ^0 is given by the nominal design,
  $$\theta^0 = \left(\widehat{\Phi}'_N \widehat{\Phi}_N\right)^{-1} \widehat{\Phi}'_N D_N.$$
- *Iteration steps.*
  For $k = 0, 1, 2, \ldots,$
  $$\theta^{k+1} = \theta^k - \beta_k \widehat{G}(D_N, \theta^k),$$
  where β_k is the step size at the kth iteration, $\widehat{G}(D_N, \theta^k)$ is an approximate gradient. Typically, these approximate values can be obtained by using Monte Carlo methods or grid calculation in place of the uncertainty set Γ.

3.3. Examples

We now use a simulation example to demonstrate performance on noise attenuation.

Example 2: The system to be controlled is a seventh order system $P(z) = p_0 + p_1 z^{-1} + \ldots + p_7 z^{-7}$. However, a lower order model is used to represent this system: $P_0(z) = p_0 + p_1 z^{-1} + p_1 z^{-2} + p_3 z^{-3}$, leaving the higher order terms as unmodelled dynamics. Hence, the modelled part has order $n = 3$ with four parameters, and the true values are $p_0 = 1, p_1 = 0.2, p_2 = 2,$ and $p_3 = 0.5$. The unmodelled dynamics represent higher order terms which are excluded in the model, and in this example they are $p_4, p_5, p_6,$ and p_7. So, $p^* = [p_4, p_5, p_6, p_7]'$. We do not have information on the unmodelled dynamics, except for the bound $\varepsilon = |p_4| + p|p_5| + |p_6| + |p_7|$. In this example, we first use $\varepsilon = 0.6$.

The noise sequence $\{d_k\}$ is i.i.d., uniformly distributed in [− 1, 1]. As explained in the previous sections, without estimation errors, d_k are known in our design process. The data length is $N = 1000$.

The uncertainty set from unmodelled dynamics is generated by the Monte Carlo method. We randomly generate 200 values of p^*, and then normalised them so that they all satisfy $|p_4| + p|p_5| + |p_6| + |p_7| = 0.6$. The corresponding set of Ξ_N matrices is used as the uncertainty set Γ.

The controller has order $m = 20$, hence θ has 21 parameters. We consider the nominal design in this example. After generating the matrices D_N, Φ_N, we obtain

$$\theta_N = \left(\widehat{\Phi}'_N \widehat{\Phi}_N\right)^{-1} \widehat{\Phi}'_N D_N = \begin{bmatrix} 0.0289 \\ -0.1221 \\ 0.4797 \\ 0.0720 \\ -0.2364 \\ -0.0412 \\ 0.1163 \\ 0.0236 \\ -0.0576 \\ -0.0130 \\ 0.0284 \\ 0.0075 \\ -0.0143 \\ -0.0039 \\ 0.0072 \\ 0.0020 \\ -0.0035 \\ -0.0013 \\ 0.0016 \\ 0.0009 \\ -0.0010 \end{bmatrix}.$$

To evaluate performance on noise attenuation, we use the noise attenuation factor, defined as follows. The size of the noise is $\|D_N\|_2/N$ and the size of the output is $\|Y_N\|_2/N$. Then the factor is

$$\gamma = \frac{\|Y_N\|_2/N}{\|D_N\|_2/N}.$$

Consequently, $\gamma < 1$ indicates noise attenuation. The smaller the factor, the better the noise attenuation performance.

When there is no unmodelled dynamics ($\varepsilon = 0$), the nominal design delivers a performance factor $\gamma = 0.0148$, which is an excellent 98.5% noise attenuation. However, when the unmodelled dynamics are introduced with $\varepsilon = 0.6$, this factor is increased to $\gamma = 0.2943$ (70.1% noise reduction attenuation), a substantial loss of performance.

Figure 8 demonstrates noise attenuation performances. The top plot is the original unattenuated noise, whose magnitude bound is 1. The second plot shows the noise attenuation performance of the controller when the plan does not contain unmodelled dynamics. It can be seen that the output values are around 0 and have much smaller magnitudes than the original noise, indicating substantial noise reduction. However, when the plan involves unmodelled dynamics, its impact is shown in the third plot. By considering

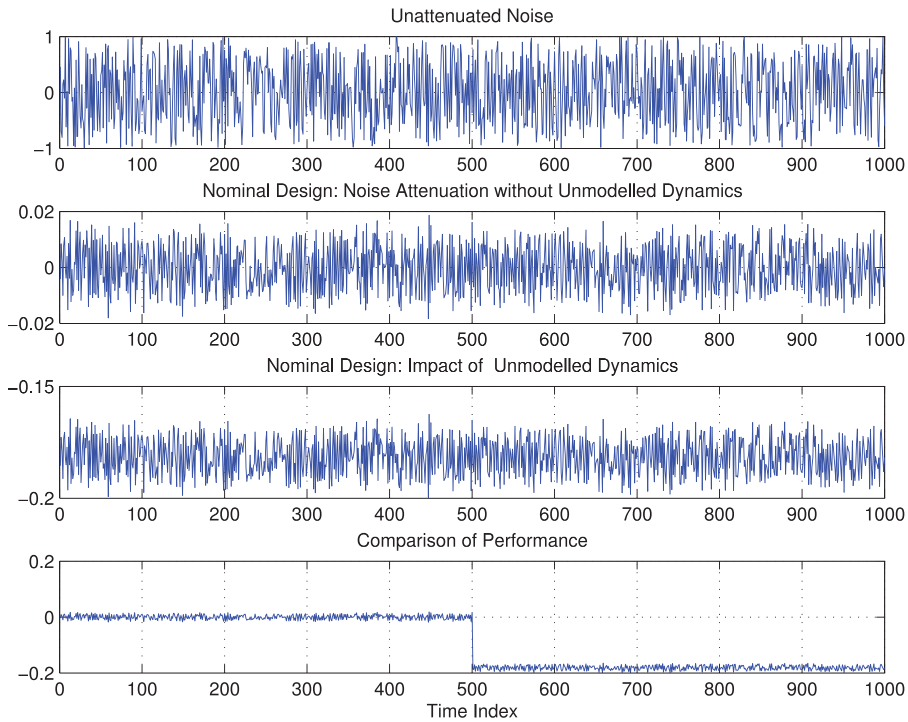


Figure 8. Noise attenuation under the nominal design.

the worst case in the uncertainty set Γ, the noise reduction capability is significantly diminished when the nominally designed controller is used. To make this point clearer, the fourth plot compares directly the performances between no unmodelled dynamics and with unmodelled dynamics. The first 500 points are the output when no unmodelled dynamics are involved; and the next 500 data points show impact of unmodelled dynamics. We should point out that this is a worst-case study. There are some incidence in Γ under which the noise attenuation performance may be much better. This is the key issue of 'robustness' of the controller which is assessed under the worst-case scenario.

Example 3: Impact of unmodelled dynamics on noise reduction performance is quite significant. To sustain acceptable noise reduction factors, one needs to use a well-representative model so that the unmodelled dynamics are not too big. To illustrate such impact, we choose different sizes ε for unmodelled dynamics for the same example as in Example 2 under the same simulation conditions. The resulting noise reduction factors and the corresponding noise reduction percentages are included in Table 1 .

Table 1. Impact of unmodelled dynamics.

Size ε of unmodelled dynamics	0.1	0.3	0.5	0.7	0.9
Reduction factor	0.0570	0.1464	0.2512	0.3459	0.4493
Reduction percentage	94.3%	85.4%	74.9%	65.4%	55.1%

4. Impact of signal estimation errors

In this section, we will analyse impact of measurement errors and unmodelled dynamics. The following assumptions are imposed.

Assumption 3: *The following conditions hold:*

(1) *$\{d_k\}$ is a sequence of i.i.d. random variables satisfying $\mathbb{E}d_k = 0$ and $\mathbb{E}d_k^2 = \sigma_d^2 < \infty$. The fourth moment of d_k is finite: $\mathbb{E}d_k^4 < \infty$.*
(2) *$\{d_k\}$ is estimated by $\widehat{d}_k = d_k + e_k$ such that $\{e_k\}$ is a sequence of i.i.d. random variables with $\mathbb{E}e_k = 0$ and $\mathbb{E}e_k^2 = \sigma_e^2 < \infty$. $\{e_k\}$ is independent of $\{d_k\}$.*
(3) *The modelled part p is known. The unmodelled dynamics p^* has a uniform norm bound ρ_n.*

4.1. Limit with measurement errors

Let

$$\begin{aligned}\theta_N^e &= (\widehat{\Phi}_N'\widehat{\Phi}_N)^{-1}\widehat{\Phi}_N'\widehat{D}_N \\ &= ((\Phi_N - \Xi_N)'(\Phi_N - \Xi_N))^{-1} \\ &\quad \times (\Phi_N' - \Xi_N')(D_N + E_N) \end{aligned} \quad (14)$$

be the estimates from the design with both measurement errors and unmodelled dynamics. We begin by showing that for this nominal design the unmodelled dynamics are cancelled out. This is done by separating the modelled and unmodelled components of Φ_N and Ξ_N as follows:

Recall that

$$\Phi_N = \begin{bmatrix} \phi_1' \\ \phi_2' \\ \vdots \\ \phi_N' \end{bmatrix} = \begin{bmatrix} w_1 & w_0 & \cdots & w_{1-n} \\ w_2 & w_1 & \cdots & w_{2-n} \\ \vdots & \vdots & \vdots & \vdots \\ w_N & w_{N-1} & \cdots & w_{N-n} \end{bmatrix}.$$

We separate the modelled and unmodelled parts of w_k by writing

$$\begin{aligned} w_k &= \psi_k' p + \widetilde{\psi}_k' p^* \\ &= \sum_{j=0}^{n} d_{k-j} p_j + \sum_{j=n+1}^{\infty} d_{k-j} p_j \\ &=: w_k^0 + \widetilde{w}_k, \end{aligned}$$

where $w_k^0 := \sum_{j=0}^{n} d_{k-j} p_j$ is a stationary, mean zero, strong mixing process (Billingsley, 1968) as d_k is i.i.d. mean zero. Thus we may represent Φ_N by

$$\Phi_N = W_N^0 + \widetilde{W}_N,$$

where W_N^0 and $\widetilde{W}_N$ are the $N \times (n+1)$ matrix collections of w_k^0 and $\widetilde{w}_k$, respectively. Also, we have

$$\Xi_N = \begin{bmatrix} \zeta_1' \\ \zeta_2' \\ \vdots \\ \zeta_N' \end{bmatrix} = \begin{bmatrix} \widetilde{\varepsilon}_1 & \widetilde{\varepsilon}_0 & \cdots & \widetilde{\varepsilon}_{1-n} \\ \widetilde{\varepsilon}_2 & \widetilde{\varepsilon}_1 & \cdots & \widetilde{\varepsilon}_{2-n} \\ \vdots & \vdots & \vdots & \vdots \\ \widetilde{\varepsilon}_N & \widetilde{\varepsilon}_{N-1} & \cdots & \widetilde{\varepsilon}_{N-n} \end{bmatrix},$$

where

$$\begin{aligned} \widetilde{\varepsilon}_k &= \widetilde{\psi}_k' p^* - \xi_k' p \\ &= \sum_{j=n+1}^{\infty} d_{k-j} p_j - \sum_{j=0}^{n} e_{k-j} p_j \\ &=: \widetilde{w}_k - \varepsilon_k^0. \end{aligned}$$

Thus we have the decomposition,

$$\Xi_N = \widetilde{W}_N - \Upsilon_N,$$

where Υ_N is the $N \times (n+1)$ matrix of $\varepsilon_k^0 = \sum_{j=0}^{n} e_{k-j} p_j$, a stationary, mean zero, ergodic process. With this new notation, we have

$$\widehat{\Phi}_N = \Phi_N - \Xi_N = W_N^0 + \Upsilon_N, \tag{15}$$

and so

$$\begin{aligned} \theta_N^e &= \left[\widehat{\Phi}_N' \widehat{\Phi}_N\right]^{-1} \widehat{\Phi}_N' \widehat{D}_N \\ &= \left[\frac{N}{N} (W_N^0 + \Upsilon_N)'(W_N^0 + \Upsilon_N)\right]^{-1} \\ &\quad \times (W_N^0 + \Upsilon_N)'(D_N + E_N) \\ &= A_N \frac{1}{N} \left(W_N^{0\prime} D_N + W_N^{0\prime} E_N + \Upsilon_N' D_N + \Upsilon_N' E_N\right), \end{aligned} \tag{16}$$

where

$$A_N := \left[\frac{1}{N}\left(W_N^{0\prime} W_N^0 + \Upsilon_N' W_N^0 + W_N^{0\prime} \Upsilon_N + \Upsilon_N' \Upsilon_N\right)\right]^{-1}. \tag{17}$$

Write

$$P_n^0 = \left[\sum_{j=0}^{n-|l_2-l_1|} p_j p_{j+|l_2-l_1|}\right]_{l_1,l_2=0,1,\dots,n}. \tag{18}$$

Then we can formulate the limit of the estimate θ_N^e in terms of P_n^0 as follows.

Proposition 1: *Under Assumption 3, assuming P_n^0 is full rank, we have*

$$\theta_N^e = \left[\widehat{\Phi}_N' \widehat{\Phi}_N\right]^{-1} \widehat{\Phi}_N' \widehat{D}_N \xrightarrow{a.s.} \left[P_n^0\right]^{-1} \begin{bmatrix} p_0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad \text{as } N \to \infty. \tag{19}$$

The proofs of Propositions 1 and 2 are postponed to Appendix.

4.2. Limit without measurement errors

Without measurement errors, the estimates are simplified to

$$\theta_N^0 = (\Phi_N' \Phi_N)^{-1} \Phi_N' D_N = B_N \left(\frac{1}{N} \Phi_N' D_N\right), \tag{20}$$

where $B_N = \left[\frac{1}{N} \Phi_N' \Phi_N\right]^{-1}$. Denote

$$P_n = \left[\sum_{j=0}^{\infty} p_j p_{j+|l_2-l_1|}\right]_{l_1,l_2=0,1,\dots,n}. \tag{21}$$

As before, we can formulate the limit of θ_N^e in terms of P_n as follows.

Proposition 2: *Under Assumption 3 and assuming P_n is full rank, we have*

$$\theta_N^0 = \left[\Phi_N' \Phi_N\right]^{-1} \Phi_N D_N \xrightarrow{a.s.} [P_n]^{-1} \begin{bmatrix} p_0 \\ 0 \\ \vdots \\ 0 \end{bmatrix},$$

as $N \to \infty$. (22)

4.3. *Difference of estimates*

Combining Propositions 1 and 2 and assuming that $P_n^0 - P_n$ is invertible, we arrive at

$$\theta_N^e - \theta_N^0 \overset{a.s.}{\to} \left[P_n^0 - P_n\right]^{-1} \begin{bmatrix} p_0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \tag{23}$$

where

$$-\left[P_n^0 - P_n\right]_{l_1,l_2} = \sum_{j=n-|l_2-l_1|+1}^{\infty} p_j p_{j+|l_2-l_1|}.$$

Defining

$$\rho_n^{(l)} \triangleq \sum_{j=n+1}^{\infty} p_{j-l} p_j \le \sum_{j=n+1}^{\infty} |p_j| \le \rho_n \tag{24}$$

for sufficiently large n, we see that

$$\left[P_n^0 - P_n\right]^{-1} = -\begin{bmatrix} \rho_n^{(0)} & \rho_n^{(1)} & \rho_n^{(2)} & \cdots & \rho_n^{(n)} \\ \rho_n^{(1)} & \rho_n^{(0)} & \rho_n^{(1)} & \cdots & \rho_n^{(n-1)} \\ \rho_n^{(2)} & \rho_n^{(1)} & \rho_n^{(0)} & \cdots & \rho_n^{(n-2)} \\ \ddots & \ddots & \ddots & \ddots & \ddots \\ \rho_n^{(n)} & \rho_n^{(n-1)} & \cdots & \rho_n^{(1)} & \rho_n^{(0)} \end{bmatrix}^{-1} \tag{25}$$

for $l = |l_2 - l_1| \in \{0, 1, \ldots, n\}$.

Example 4: We conduct a simulation study to display the limit of the estimate differences. We take $\{d_k\} \sim \mathcal{N}(0, 1)$ and $\{e_k\} \sim \mathcal{N}(0, .1)$, both i.i.d. The plant is a stable system with infinite impulse response (IIR) coefficients $p_k = (0.5)^k$ for $k = 0, 1, \ldots$. The model order is selected as $n = 10$. We then observe the estimates θ_N^e, θ_N^0 for $N = 10, 20, \ldots, 1010$ (100 updates). Thus $\rho_n = 2 - \sum_{k=0}^{10} p_k = (.5)^{10} \approx 9.8 \times 10^{-4}$. Figure 9 shows that $||\theta_N^e - \theta_N^0||$ quickly converges to $O(\rho_n)$.

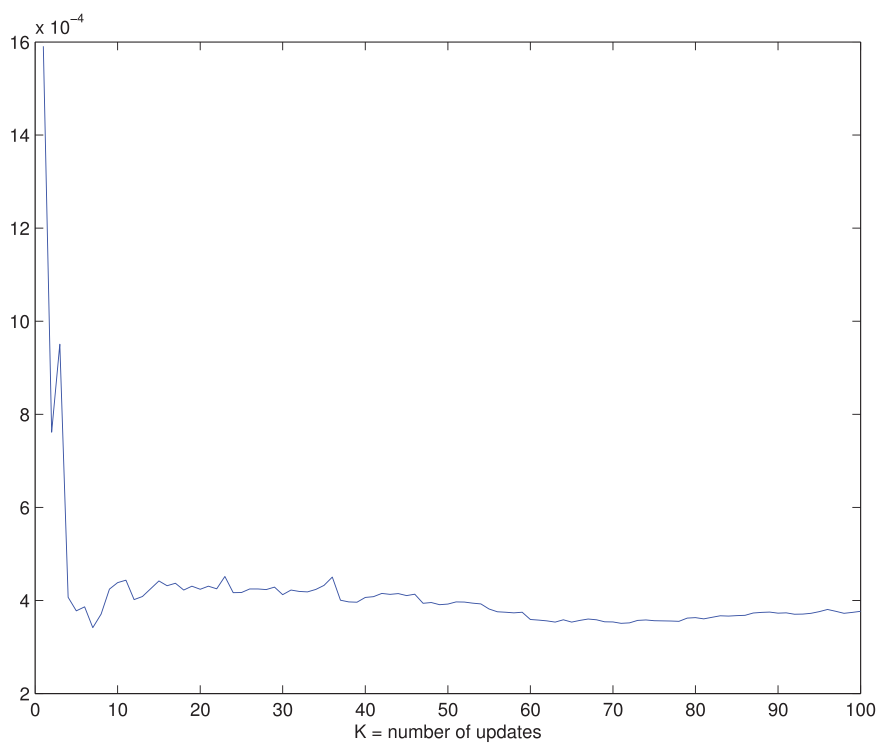


Figure 9. $||\theta_N^e - \theta_N^0||$, $N = Kn = 10, \ldots, 1010$.

5. Concluding remarks

This paper develops strategies to resolve several practical issues that arise in noise rejection problems and to overcome certain shortcomings in classical Wiener filters and more recent H^∞ sensitivity minimisation problems. By using combined signal estimation and noise rejection design, the methodology of this paper can potentially rely on highly efficient control design of least-squares types but still achieve non-conservative robust noise rejection without prior knowledge on noise characterisations.

There remain many open problems. This paper considers only linear systems. Non-linearity will introduce model mismatch and its impact needs to be carefully assessed. Also, practical systems contain time delays. It is not clear if our methodology can be extended to deal with such infinite dimensional systems without losing its computational efficiency. Furthermore, if noise characterisation or system dynamics changes with time, adaptive versions of our methodology need to be developed.

Funding

This research was supported in part by the Air Force Office of Scientific Research [grant number FA9550-12-C-0035], [grant number FA9550-09-1-0524], [grant number FA9550-10-1-0499].

Notes on contributors

Araz Hashemi received the BS degree in mathematics in 2007, the MA degree in mathematics in 2009, the MA degree in mathematical statistics in 2012, and the PhD degree in mathematics in 2014, all from Wayne State University. He will begin as a postdoctoral researcher at the University of Delaware in July, 2014. His research interests include adaptive filtering and control of stochastic systems, distributed learning and cooperative control in multi agent systems, uncertainty quantification in simulation design, and topological data analysis.

Ben Fitzpatrick, PhD, is the Clarence J. Wallen Professor of Mathematics at Loyola Marymount University (LMU) and President of Tempest Technologies. He has 30 years of experience in statistical analysis, quantitative modelling, and software development. As the holder of the Wallen Chair at LMU, Dr. Fitzpatrick directs the independent study program in biomathematics, conducts research on mathematical modelling, and statistical data analysis in ecological, genomic, and biological research projects. In his duties at Tempest, he currently manages consulting projects on planning and control of mixed initiative and semi-autonomous unmanned vehicle operations, computational non-linear control, and real-time adaptive control. Dr. Fitzpatrick holds a PhD in applied mathematics from Brown University, and masters and bachelors degrees from Auburn University. Prior to joining LMU and forming Tempest, he has held positions at Wagner and Associates, North Carolina State University, and the University of Tennessee, Knoxville, TN. He has published extensively in the areas of modelling, dynamics and control, and inverse problems, with applications to social science, biology and ecology, and engineering systems.

Le Yi Wang received the PhD degree in electrical engineering from McGill University, Montreal, Canada, in 1990. Since 1990, he has been with Wayne State University, Detroit, MI, where he is currently a professor in the Department of Electrical and Computer Engineering. His research interests are in the areas of complexity and information, system identification, robust control, H^∞ optimisation, time-varying systems, adaptive systems, hybrid and non-linear systems, information processing and learning, as well as medical, automotive, communications, power systems, and computer applications of control methodologies. He was a keynote speaker in several international conferences. He was an associate editor of the IEEE Transactions on Automatic Control and several other journals, and currently is an associate editor of the Journal of System Sciences and Complexity and Journal of Control Theory and Applications. He is a fellow of IEEE.

George Yin received the BS degree in mathematics from the University of Delaware in 1983, MS degree in electrical engineering, and PhD in applied mathematics from Brown University in 1987. He joined the Department of Mathematics, Wayne State University in 1987, and became professor in 1996. He was Co-chair of SIAM Conference on Control & Its Application (2011), and Co-chair of a couple of AMS-IMS-SIAM Summer Conferences. In addition to chairing a number of prize committees, he is one of the Board of Directors of American Automatic Control Council; he is Chair (2014–2015), Vice Chair (2012-2013), Program Director (2010-2011) of SIAM Activity Group on Control and Systems Theory; he is an associate editor of SIAM Journal on Control and Optimization, and is on the editorial board of a number of other journals. He was an Associate Editor of Automatica (2003-2011) and IEEE Transactions on Automatic Control (1994-1998). He is a fellow of IEEE.

References

Anderson, B.D.O., & Moore, J.B. (1979). *Optimal filtering*. New York, NY: Prentice Hall.

Billingsley, P. (1968). *Convergence of probability measures*. New York, NY: J. Wiley.

Chen, H.F., & Guo, L. (1991). *Identification and stochastic adaptive control*. Boston, MA: Birkhäuser.

Doyle, J., Francis, B., & Tannenbaum, A. (1990). *Feedback control theory*. New York, NY: Macmillan Publishing.

Francis, B.A. (1987). *A course in H^∞ control theory, vol. 88 in lecture notes in Control and Information Sciences*. New York, NY: Springer-Verlag.

Francis, B.A., & Zames, G. (1984). On H^∞-optimal sensitivity theory for SISO feedback systems. *IEEE Transactions on Automatic Control, AC-29*, 9–16.

He, Q., Wang, L.Y., & Yin, G. (2013). *System identification using regular and quantized observations: Applications of large deviations principles*. New York, NY: Springer.

Karlin, S., & Taylor, H.M. (1975). *A first course in stochastic processes* (2nd ed.). New York, NY: Academic Press.

Kushner, H.J., & Yin, G. (2003). *Stochastic approximation algorithms and applications* (2nd ed.). New York, NY: Springer-Verlag.

Ljung, L. (1987). *System identification: Theory for the user*. Englewood Cliffs, NJ: Prentice-Hall.

Milanese, M., & Belforte, G. (1982). Estimation theory and uncertainty intervals evaluation in the presence of unknown but bounded errors: Linear families of models and estimators. *IEEE Transactions on Automatic Control, 27*, 408–414.

Milanese, M., & Vicino, A. (1991). Optimal estimation theory for dynamic systems with set membership uncertainty: An overview. *Automatica, 27*, 997–1009.

Milanese, M., & Vicino, A. (1993). Information-based complexity and nonparametric worst-case system identification. *Journal of Complexity, 9*, 427–446.

Nguyen, Q.T., Veselya, V., & Rosinova, D. (2013). Design of robust model predictive control with input constraints. *International Journal of Systems Science, 44*, 896–907.

Poolla, K., & Tikku, A. (1994). On the time complexity of worst-case system identification. *IEEE Transactions on Automatic Control, 39*, 944–950.

Sayeed, A.M. (2003). *A signal modeling framework for integrated design of sensor networks*. 12th IEEE Workshop Statistical Signal Processing. St. Louis, MO.

Wang, L.Y., Yin, G., Zhang, J.-F., & Zhao, Y.L. (2010). *System identification with quantized observations: Theory and applications*. Boston, MA: Birkhäuser.

Wiener, N. (1949). *Extrapolation, interpolation, and smoothing of stationary time series*. New York, NY: Wiley.

Youla, D.C., Jabri, H.A., & Bongiorno, J.J. (1976a). Modern Wiener– Hopf design of optimal controllers: Part I: The single-input-output case. *IEEE Transactions on Automatic Control, AC-21*(1), 3–13.

Youla, D.C., Jabri, H.A., & Bongiorno, J.J. (1976b). Modern Wiener– Hopf design of optimal controllers: Part II: The multivariable case. *IEEE Transactions on Automatic Control, AC-21*(3), 319–338.

Zames, G. (1981). Feedback and optimal sensitivity: Model reference transformations, multiplicative seminorms, and approximate inverses. *IEEE Transactions on Automatic Control, 26*, 301–320.

Zhao, X., Liu, H., & Wang, Z. (2013). Weighted H_∞ performance analysis of switched linear systems with mode-dependent average dwell time. *International Journal of Systems Science, 44*, 2130–2139.

Appendix: Proofs

Proof of Proposition 1: Working with the terms of A_N, we see that

$$\frac{1}{N} W_N^{0\prime} W_N^0 = \frac{1}{N}\sum_{k=1}^{N} \begin{bmatrix} w_k^0 w_k^0 & w_k^0 w_{k-1}^0 & \cdots & w_k^0 w_{k-n}^0 \\ w_{k-1}^0 w_k^0 & w_{k-1}^0 w_{k-1}^0 & \cdots & w_{k-1}^0 w_{k-n}^0 \\ \vdots & \vdots & \vdots & \vdots \\ w_{k-n}^0 w_k^0 & w_{k-n}^0 w_{k-1}^0 & \cdots & w_{k-n}^0 w_{k-n}^0 \end{bmatrix}, \tag{26}$$

with

$$\begin{aligned} \mathbb{E} w_{k-l_1}^0 w_{k-l_2}^0 &= \mathbb{E}\sum_{j_1=0}^{n}\sum_{j_2=0}^{n} d_{k-l_1-j_1} d_{k-l_2-j_2} p_{j_1} p_{j_2} \mathbb{1}_{\{l_1+j_1=l_2+j_2\}} \\ &= \sigma_d^2 \sum_{j=0}^{n-|l_2-l_1|} p_j p_{j+|l_2-l_1|}. \end{aligned} \tag{27}$$

We claim that the stationary process $\{w_{k-l_1}^0 w_{k-l_2}^0\}_k$ has mean

$$m = \mathbb{E} w_{k-l_1}^0 w_{k-l_2}^0 = \sigma_d^2 \sum_{j=0}^{n-|l_2-l_1|} p_j p_{j+|l_2-l_1|}$$

and

$$R(h) = \mathbb{E}\left\{ w_{k+h-l_1}^0 w_{k+h-l_2}^0 w_{k-l_1}^0 w_{k-l_2}^0 \right\} - m^2.$$

In fact,

$$\begin{aligned} &\mathbb{E}\left\{ w_{k+h-l_1}^0 w_{k+h-l_2}^0 w_{k-l_1}^0 w_{k-l_2}^0 \right\} \\ &\quad = \mathbb{E}\sum_{j_1,\ldots,j_4=0}^{n} d_{k+h-l_1-j_1} d_{k+h-l_1-j_2} d_{k-l_1-j_3} d_{k-l_2-j_4} p_{j_1} p_{j_2} p_{j_3} p_{j_4}. \end{aligned}$$

For $h > 2n$, $k + h - l_1 - j_1 > k + h - l_2 - j_2$ for $l, j \in \{0, \ldots, n\}$, so $d_{k+h-l_1-j_1}$ is independent of $d_{k+h-l_2-j_2}$, and thus we can reduce the terms in the sum to

$$\begin{aligned} &\mathbb{E}\left\{ w_{k+h-l_1}^0 w_{k+h-l_2}^0 w_{k-l_1}^0 w_{k-l_2}^0 \right\} \\ &\quad = \mathbb{E}\sum_{j_1,\ldots,j_4=0}^{n} d_{k+h-l_1-j_1} d_{k+h-l_1-j_2} d_{k-l_1-j_3} d_{k-l_2-j_4} \\ &\qquad \times p_{j_1} p_{j_2} p_{j_3} p_{j_4} \mathbb{1}_{\{l_1+j_1=l_2+j_2\}} \\ &\quad = \sum_{j_1=0}^{n-|l_2-l_1|}\sum_{j_3=0}^{n-|l_2-l_1|} \mathbb{E}[d_{k+h-j_1}^2]\mathbb{E}[d_{k+h-j_3}^2] p_{j_1} p_{j_1+|l_2-l_1|} \\ &\qquad \times p_{j_3} p_{j_3+|l_2-l_1|} \\ &\quad = \sigma_d^4 \left[\sum_{j=0}^{n-|l_2-l_1|} p_j p_{j+|l_2-l_1|} \right]^2 = m^2. \end{aligned} \tag{28}$$

Thus the covariance function $R(h) = 0$ for $h > 2n$. Moreover, $\sum_{h=0}^{N-1} R(h)/N \to 0$. As a result, with $X_k = w_{k-l_1}^0 w_{k-l_2}^0$, $\overline{X}_N = \frac{1}{N}\sum_{k=1}^{N} X_k \xrightarrow{L^2} m$ as $N \to \infty$ by Karlin and Taylor (1975, Theorem 9.5.1). Moreover, since $R(h) = 0$ for $h > 2n$, $\{X_k\}$ is a strong mixing process (Karlin & Taylor, 1975, p. 488). By virtue of Karlin and Taylor (1975, Theorems 9.5.6), $\{X_k\}$ is strongly ergodic, and by Karlin and Taylor (1975, Theorems 9.5.5), $\overline{X}_N \to m$ almost surely (a.s.).

Using the ergodicity obtained above and in Equation (18),

$$\frac{1}{N} W_N^{0\prime} W_N^0 \xrightarrow{a.s.} \sigma_d^2 P_n^0 \qquad \text{as } N \to \infty. \tag{29}$$

Similar arguments yield

$$\frac{1}{N} \Upsilon_N' \Upsilon_N \xrightarrow{a.s.} \sigma_e^2 P_n^0 \ \text{ and } \ \frac{1}{N} W_N^{0\prime} \Upsilon_N \xrightarrow{a.s.} 0 \ \text{ as } \ N \to \infty.$$

Thus

$$A_N \xrightarrow{a.s.} \left(\sigma_d^2 + \sigma_e^2\right)^{-1} \left[P_n^0\right]^{-1} \quad \text{as } N \to \infty.$$

Examining the terms that A_N is applied to in Equation (16),

$$\frac{1}{N}\left(W_N^{0\prime}D_N + W_N^{0\prime}E_N + \Upsilon_N' D_N + \Upsilon_N' E_N\right) = \frac{1}{N}\sum_{k=1}^{N}\begin{bmatrix} w_k^0 d_k + w_k^0 e_k + \varepsilon_k^0 d_k + \varepsilon_k^0 e_k \\ w_{k-1}^0 d_k + w_{k-1}^0 e_k + \varepsilon_{k-1}^0 d_k + \varepsilon_{k-1}^0 e_k \\ \vdots \\ w_{k-n}^0 d_k + w_{k-n}^0 e_k + \varepsilon_{k-n}^0 d_k + \varepsilon_{k-n}^0 e_k \end{bmatrix}, \tag{30}$$

where

$$\mathbb{E} w_{k-l}^0 d_k = \mathbb{E}\sum_{j=0}^{n} d_{k-l-j} d_k = \sigma_d^2 p_0 \mathbb{1}_{\{l=0\}}$$
$$\mathbb{E} w_{k-l}^0 e_k = \mathbb{E}\sum_{j=0}^{n} d_{k-l-j} e_k = 0$$
$$\mathbb{E} \varepsilon_{k-l}^0 d_k = \mathbb{E}\sum_{j=0}^{n} e_{k-l-j} d_k = 0$$
$$\mathbb{E} \varepsilon_{k-l}^0 e_k = \mathbb{E}\sum_{j=0}^{n} e_{k-l-j} e_k = \sigma_e^2 p_0 \mathbb{1}_{\{l=0\}}.$$

Inspecting the covariance function for $X_k = w_{k-l}^0 d_k$, we see that d_{k+h} is independent of d_{k+h-j} for any $j > 0$, so that

$$\begin{aligned}\mathbb{E} w_{k+h-l}^0 d_{k+h} w_{k-l}^0 d_k &= \mathbb{E}\sum_{j_1=0}^{n}\sum_{j_2=0}^{n} d_{k+h-l-j_1} d_{k+h} d_{k+h-l_2-j_2} d_k p_{j_1} p_{j_2} \mathbb{1}_{\{l=0,\, j_1=0,\, j_2=0\}} \\ &= \mathbb{E} d_{k+h}^2 d_k^2 p_0^2 = \sigma_d^4 p_0^2 \mathbb{1}_{\{l=0\}} = \left[\mathbb{E} w_{k-l}^0 d_k\right]^2 \qquad \text{if } h > 0.\end{aligned} \tag{31}$$

Thus $\frac{1}{N}\sum_{k=1}^{N} w_{k-l}^0 d_k \xrightarrow{a.s.} \mathbb{E} w_{k-l}^0 d_k$, and similarly for the other terms of Equation (30). Hence, we have

$$\frac{1}{N}\left(W_N^{0\prime}D_N + W_N^{0\prime}E_N + \Upsilon_N' D_N + \Upsilon_N' E_N\right) \xrightarrow{a.s.} \left(\sigma_d^2 + \sigma_e^2\right)\begin{bmatrix} p_0 \\ 0 \\ \vdots \\ 0\end{bmatrix}. \tag{32}$$

Using Equations (16), (17), (32), and the limits obtained thus far, we have

$$\left[\widehat{\Phi}_N' \widehat{\Phi}_N\right]^{-1} \widehat{\Phi}_N' \widehat{D}_N \xrightarrow{a.s.} \left(\sigma_d^2 + \sigma_e^2\right)^{-1}\left[P_n^0\right]^{-1}\left(\sigma_d^2 + \sigma_e^2\right)\begin{bmatrix} p_0 \\ 0 \\ \vdots \\ 0\end{bmatrix}. \tag{33}$$

Thus we have proven Proposition 1. □

Proof of Proposition 2: We have that

$$\frac{1}{N}\Phi_N'\Phi_N = \frac{1}{N}\sum_{k=1}^{N}\begin{bmatrix} w_k w_k & \cdots & w_k w_{k-n} \\ \vdots & \vdots & \vdots \\ w_{k-n} w_k & \cdots & w_{k-n} w_{k-n}\end{bmatrix}. \tag{34}$$

We observe

$$\begin{aligned}\mathbb{E} w_{k-l_1} w_{k-l_2} &= \mathbb{E}\sum_{j_1=0}^{\infty}\sum_{j_2=0}^{\infty} d_{k-l_1-j_1} d_{k-l_2-j_2} p_{j_1} p_{j_2} \\ &= \mathbb{E}\sum_{j_1=0}^{\infty}\sum_{j_2=0}^{\infty} d_{k-l_1-j_1} d_{k-l_2-j_2} p_{j_1} p_{j_2} \mathbb{1}_{\{l_1+j_1=l_2+j_2\}} \\ &= \sigma_d^2 \sum_{j=0}^{\infty} p_j p_{j+|l_2-l_1|}.\end{aligned} \tag{35}$$

Using the definition (21), we have

$$\mathbb{E}\left[w_{k-l_1} w_{k-l_2}\right]_{l_1,l_2=0,\ldots,n} = \sigma_d^2 P_n. \tag{36}$$

Establishing that the product sequences $\{w_{k-l_1} w_{k-l_2}\}_k$ are ergodic is more complicated due to the infinite sum involved in w_k. We show it for $X_k = w_k w_k$, with the shifted products done in a similar manner. Here, $[\mathbb{E} w_k w_k]^2 = [\sum_{j=0}^{\infty}\sigma_d^2 p_j^2] = \sigma_d^4 \sum_{j_1}\sum_{j_2} p_{j_1}^2 p_{j_2}^2$, and

$$\mathbb{E} w_{k+h} w_{k+h} w_k w_k = \sum_{j_1,\ldots,j_4=0}^{\infty} \mathbb{E} d_{k+h-j_1} d_{k+h-j_2} d_{k-j_3} d_{k-j_4} \times p_{j_1} p_{j_2} p_{j_3} p_{j_4}.$$

For the expectation of a term to be non-zero, every index of $d_\cdot$ must be paired with another. Writing $\mathcal{A} = \{(j_1, j_2, j_3, j_4) : j_1 = j_2, j_3 = j_4\}$, $\mathcal{B} = \{j_1 = j_3 + h, j_2 = j_4 + h\}$, and $\mathcal{C} = \{j_1 = j_4 + h, j_2 = j_3 + h\}$, we have that the non-zero terms are precisely $\mathcal{A} \cup \mathcal{B} \cup \mathcal{C}$. Furthermore, $\mathcal{A}\cap\mathcal{B} = \mathcal{A}\cap\mathcal{C} = \mathcal{B}\cap\mathcal{C} = \mathcal{A}\cap\mathcal{B}\cap\mathcal{C}$, so $\sum_{\mathcal{A}\cup\mathcal{B}\cup\mathcal{C}} = \sum_{\mathcal{A}} + \sum_{\mathcal{B}} + \sum_{\mathcal{C}} - 2\sum_{\mathcal{A}\cap\mathcal{B}\cap\mathcal{C}}$. Then we can express

$$\begin{aligned}\mathbb{E} w_{k+h} w_{k+h} w_k w_k &= \sum_{j_1=0}^{\infty}\sum_{j_3=0}^{\infty}\mathbb{E} d_{k-j_1}^2 d_{k-j_3}^2 p_{j_1}^2 p_{j_3}^2 \\ &\quad + \sum_{j_3=0}^{\infty}\sum_{j_4=0}^{\infty}\mathbb{E} d_{k-j_3}^2 d_{k-j_4}^2 p_{j_3} p_{j_3+h} p_{j_4} p_{j_4+h} \\ &\quad + \sum_{j_3=0}^{\infty}\sum_{j_4=0}^{\infty}\mathbb{E} d_{k-j_3}^2 d_{k-j_4}^2 p_{j_3+h} p_{j_3} p_{j_4+h} p_{j_4} \\ &\quad - 2\sum_{j=0}^{\infty}\mathbb{E} d_{k-j}^4 p_j^2 p_{j+h}^2 \\ &= \sum_{j_1\neq j_2+h}\sum_{j_2=0}^{\infty}\sigma_d^4 p_{j_1}^2 p_{j_2}^2 + \sum_{j=0}^{\infty}\mathbb{E} d^4 p_j^2 p_{j+h}^2 \\ &\quad + 2\sum_{j_1\neq j_2}\sum_{j_2=0}^{\infty}\sigma_d^4 p_{j_1} p_{j_1+h} p_{j_2} p_{j_2+h} \\ &\quad + 2\sum_{j=0}^{\infty}\mathbb{E} d^4 p_j^2 p_{j+h}^2 - 2\sum_{j=0}^{\infty}\mathbb{E} d^4 p_j^2 p_{j+h}^2 \\ &= \sum_{j_1\neq j_2+h}\sum_{j_2=0}^{\infty}\sigma_d^4 p_{j_1}^2 p_{j_2}^2 + \sum_{j=0}^{\infty}\mathbb{E} d^4 p_j^2 p_{j+h}^2 \\ &\quad + 2\sum_{j_1\neq j_2}\sum_{j_2=0}^{\infty}\sigma_d^4 p_{j_1} p_{j_1+h} p_{j_2} p_{j_2+h}.\end{aligned} \tag{37}$$

Thus the covariance $R(h) = \mathbb{E}\{w_{k+h}w_{k+h}w_k w_k\} - [\mathbb{E}\{w_k w_k\}]^2$ satisfies

$$\begin{aligned}
|R(h)| &= \Bigg| \sum_{j_1 \neq j_2+h} \sum_{j_2=0}^{\infty} \sigma_d^4 p_{j_1}^2 p_{j_2}^2 + \sum_{j=0}^{\infty} \mathbb{E}d^4 p_j^2 p_{j+h}^2 \\
&\quad + 2 \sum_{j_1 \neq j_2} \sum_{j_2=0}^{\infty} \sigma_d^4 p_{j_1} p_{j_1+h} p_{j_2} p_{j_2+h} - \sum_{j_1=0}^{\infty} \sum_{j_2=0}^{\infty} \sigma_d^4 p_{j_1}^2 p_{j_2}^2 \Bigg| \\
&= \Bigg| -\sigma_d^4 \sum_{j_1=j_2+h} \sum_{j_2=0}^{\infty} p_{j_1}^2 p_{j_2}^2 + \mathbb{E}d^4 \sum_{j=0}^{\infty} p_j^2 p_{j+h}^2 \\
&\quad + 2\sigma_d^4 \sum_{j_1 \neq j_2} \sum_{j_2=0}^{\infty} p_{j_1} p_{j_1+h} p_{j_2} p_{j_2+h} \Bigg| \\
&\leq |p_h| \sigma_d^4 \sum_{j=0}^{\infty} p_j^2 + |p_h| \mathbb{E}d^4 \sum_{j=0}^{\infty} p_j^2 \\
&\quad + |p_h| 2\sigma_d^4 \sum_{j_1=0}^{\infty} \sum_{j_2=0}^{\infty} p_{j_1} p_{j_2} \\
&\leq |p_h| C \to 0 \quad \text{as } h \to \infty.
\end{aligned} \tag{38}$$

Thus the process $\{w_{k+h}w_{k+h}w_k w_k\}$ is strong mixing as well. Similar argument as in the derivation of Equation (29) yields that $\frac{1}{N}\Phi_N' \Phi_N \overset{a.s.}{\to} \sigma_d^2 P_n$. Recall that P_n is full rank,

$$B_N \overset{a.s.}{\to} \sigma_d^{-2}[P_n]^{-1} \qquad \text{as } N \to \infty. \tag{39}$$

Similarly,

$$\frac{1}{N}\Phi_N' D_N = \frac{1}{N} \sum_{k=1}^{N} \begin{bmatrix} w_k d_k \\ w_{k-1} d_k \\ \vdots \\ w_{k-n} d_k \end{bmatrix}, \tag{40}$$

where

$$\mathbb{E} w_{k-l} d_k = \mathbb{E} \sum_{j=0}^{\infty} d_{k-l-j} d_k p_j = \sigma_d^2 p_0 \mathbb{1}_{\{l=0\}} \tag{41}$$

and the covariance function decays asymptotically in a manner similar to Equation (38), so that

$$\frac{1}{N}\Phi_N' D_N \overset{a.s.}{\to} \begin{bmatrix} \sigma_d^2 p_0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad \text{as } N \to \infty. \tag{42}$$

Finally, we have

$$\theta_N^0 = B_N \left(\frac{1}{N} \Phi_N' D_N \right) \overset{a.s.}{\to} \sigma_d^{-2} \{P_n\}^{-1} \begin{bmatrix} \sigma_d^2 p_0 \\ 0 \\ \vdots \\ 0 \end{bmatrix},$$

as $N \to \infty$, and thus we have proven Proposition 2. □