Sampled-Data and Discrete-Time H2 Optimal Control

Trentelman, H.L.; Stoorvogel, A.A.

Published in:
SIAM Journal on Control and Optimization

Publication date:
1995

SIAM J. CONTROL AND OPTIMIZATION
Vol. 33, No. 3, pp. 834–862, May 1995

009

SAMPLED-DATA AND DISCRETE-TIME H_2 OPTIMAL CONTROL*

H. L. TRENTELMAN† AND A. A. STOORVOGEL‡

Abstract. This paper deals with the sampled-data H_2 optimal control problem. Given a linear time-invariant continuous-time system, the problem of minimizing the H_2 performance over all sampled-data controllers with a fixed sampling period can be reduced to a pure discrete-time H_2 optimal control problem. This discrete-time H_2 problem is always singular. Motivated by this, in this paper we give a treatment of the discrete-time H_2 optimal control problem in its full generality. The results we obtain are then applied to the singular discrete-time H_2 problem arising from the sampled-data H_2 problem. In particular, we give conditions for the existence of optimal sampled data controllers. We also show that the H_2 performance of a continuous-time controller can always be recovered asymptotically by choosing the sampling period sufficiently small. Finally, we show that the optimal sampled-data H_2 performance converges to the continuous-time optimal H_2 performance as the sampling period converges to zero.

Key words. sampled-data, lifting technique, discrete-time, H_2 optimal control, algebraic Riccati equation, small sampling periods

AMS subject classifications. 93C05, 93C35, 93C60

1. Introduction. Recently, much attention has been paid to H_2 and H_∞ optimal control of linear systems using sampled-data control (see [6], [7], [12], [2], [4] and [5], [11], [10], [1], [3], [17], [21]). For a given a continuous-time plant, a sampled-data controller consists of the cascade connection of an A/D converter, a discrete-time controller, and a D/A converter. The A/D device converts the continuous-time measured plant output into a discrete-time signal, which is used as an input for the discrete-time controller. The discrete-time controller generates a discrete-time output signal, which, in turn, is converted into a continuous-time signal that is used as a control input for the continuous-time plant.

Apart from a control input and a measurement output, the plant under consideration has an exogenous input and an output to be controlled. The quality of a controller is given by the performance of the corresponding closed-loop system. This performance measures the influence of the exogenous input on the output to be controlled. In the present paper, we will take the H_2 performance of the closed-loop system as performance measure.

In contrast to the H_∞ performance of a sampled-data control system, which in analogy with the pure continuous-time context can simply be defined as the norm of the input/output operator between the exogenous inputs and the outputs to be controlled, it is not clear from the outset how one should define the H_2 performance of a sampled-data control system. One definition was proposed in [6]: the H_2 performance of the closed-loop system is the number obtained by applying at each input channel a Dirac distribution and by taking the sum of integral squares of the resulting outputs. Of course, this definition exactly mimics the one that is common in the pure continuous-time context.

* Received by the editors December 28, 1992; accepted for publication (in revised form) November 24, 1993.

† Mathematics Institute, University of Groningen, P.O. Box 800, 9700 AV Groningen, The Netherlands (h.l.trentelman@math.rug.nl).

‡ Faculty of Mathematics and Computing Science, Eindhoven University of Technology, P.O. Box 513, 5600 MB Eindhoven, The Netherlands (wscoas@win.tue.nl). This research was supported by a fellowship of the Royal Netherlands Academy of Sciences and Arts.

In our opinion, a more natural definition was given independently in [12] and [2]. In these references, the crucial observation is that the closed-loop system resulting from a sampled data controller, albeit time-varying, is in fact a periodic system, with period equal to the sampling period. It is then argued that, instead of applying impulsive inputs at time $t = 0$, one should in fact apply these inputs at all time instances between 0 and the sampling period and take the mean of the integral squares of the resulting outputs. This leads to an H_2 performance measure that captures the essential features of a sampled-data closed-loop system more satisfactorily. For a given continuous-time plant, the sampled-data H_2 optimal control problem is then to minimize the H_2 performance of the closed-loop system over all internally stabilizing sampled-data controllers with a fixed sampling period. It is the latter problem that will be studied in this paper.

It was shown in [12] and [2] (see also [4]) that the sampled-data H_2 optimal control problem can be reduced to a pure discrete-time H_2 optimal control problem in the following way. First one defines an auxiliary time-invariant discrete-time system (involving the parameters of the original continuous-time plant and the given sampling period). Next, one expresses the sampled-data H_2 performance in terms of the 'normal' H_2 performance of the closed-loop system obtained by interconnecting the auxiliary discrete-time system and the discrete-time controller defining the sampled-data controller. Thus, the sampled-data H_2 optimal control problem under consideration is completely resolved once the auxiliary discrete-time H_2 problem is. This procedure makes use of the so-called *lifting technique* (see [20], [1], [3])

Now it turns out that the auxiliary discrete-time H_2 problem obtained in this way *is always a singular problem*: the direct feedthrough matrix from the exogenous input to the measurement output is always equal to 0. Apart from this, in the auxiliary discrete-time system the direct feedthrough matrix from the control input to the output to be controlled is in general not injective. (Note that, in general, an H_2 optimal control problem is called *regular* if the direct feedthrough matrix from the control input to the output to be controlled is injective, and the direct feedthrough matrix from the exogenous input to the measurement output is surjective. If the problem is not regular it is called *singular.*) In [12], this difficulty is partly removed by introducing an additional noise on the sampled measured output signal and by assuming the corresponding feedthrough matrix to be surjective.

In the present paper we want to consider the completely general formulation of the sampled-data H_2 problem. As a starting point we will take the auxiliary discrete-time H_2 problem derived in [12] and [2]. As noted, this problem is inherently singular. To our best knowledge, no resolution of the discrete-time singular H_2 optimal is known in the literature. Therefore, a substantial part of this paper is devoted to a study of the completely general discrete-time H_2 problem (no assumptions on the direct feedthrough matrices, no assumptions on the absence of zeros on the unit circle). We will describe a complete resolution to this problem, including a characterization of the optimal performance, and necessary *and* sufficient conditions for the existence of optimal controllers. The expression for the optimal performance is different from the one that might be expected in analogy with the continuous-time case (see [15]). Due to the fact that the role of the imaginary axis is taken over by the unit circle, for the discrete-time H_2 performance to be finite it is no longer required that the closed-loop transfer matrix is *strictly proper*. Intuitively, this enlarges the class of admissible controllers and yields a smaller optimal performance.

We will apply our results on the discrete-time H_2 optimal control problem to the sampled-data H_2 problem by simply applying them to the auxiliary discrete-time system derived in [12] and [2]. Our expression for the optimal sampled-data H_2 performance will be an immediate consequence of these results. We will, however, also be interested in conditions guaranteeing the existence of optimal sampled-data controllers. Our results on the general discrete-time H_2 problem give such conditions in terms of the auxiliary discrete-time system, but we will reformulate these conditions *in terms of the original continuous-time plant.* Preliminary results in that direction were also found in [12].

Obviously, the sampled-data H_2 optimal performance is a function of the sampling period. An important question is: what happens if the sampling period tends to zero. In particular, we will answer the following two questions. First, if we control the original continuous-time plant by a "normal" continuous-time compensator, is it then possible to recover this performance asymptotically by using a sampled-data controller with sufficiently small sampling period? This question was also studied for the H_∞ performance and for the H_2 performance à la Chen and Francis in [6]. A second, related, question that we will answer is: does the optimal sampled-data H_2 performance converge to the optimal continuous-time H_2 performance as the sampling period decreases to zero?

The outline of this paper is as follows. In §2 we will define the sampled-data H_2 optimal control problem and recall the main results of [12] and [2]. We will also introduce some notation and recall the notions of left-invertibility and right-invertibility of linear systems, zeros, and their most important state space interpretations. In §3 we deal with the discrete-time H_2 optimal control problem. In this section we will not yet treat the completely general case but make some assumptions on the absence of zeros on the unit circle. In §4, the results of §3 will be extended to derive a resolution of the general discrete-time H_2 optimal control problem. Then, in §5, we return to the sampled-data context and apply the results of §§3 and 4 to the sampled-data H_2 optimal control problem. In particular, we will derive conditions in terms of the original continuous-time plant that guarantee the existence of optimal controllers for the sampled-data H_2 problem. Finally, in §6 we study the aforementioned questions regarding the behavior of the (optimal) performance as the sampling period tends to zero.

2. Problem formulation. Consider a continuous-time, linear, time-invariant, finite-dimensional plant Σ. Let Σ have inputs d and u and outputs z and y, where d is an exogenous input, u is a control input, z is an output to be controlled, and y is a measured output. We want to control Σ by means of sampled-data feedback control. We take a fixed $\Delta > 0$, called the *sampling period.* From the measured output y we obtain a discrete-time signal $\bar{y} = \{y_k\}$ defined by $y_k := (S_\Delta y)_k$, where S_Δ denotes the sampling operator defined by $(S_\Delta y)_k := y(k\Delta)$. This discrete-time signal is taken as input for a discrete-time, linear, time-invariant, finite-dimensional compensator Γ_{dis}. The latter compensator generates a discrete-time signal $\bar{u} = \{u_k\}$, which, in turn, yields a (piecewise constant) continuous-time input signal u for the plant by defining $u(t) := (H_\Delta \bar{u})(t)$, where H_Δ is the hold operator defined by $(H_\Delta \bar{u})(t) := u_k$ $(t \in [k\Delta, (k+1)\Delta))$. This type of feedback control is depicted in Fig. 1.

If we control the system Σ by means of a sampled-data controller with sampling period Δ, then the resulting closed-loop system will no longer be time-invariant. In [12] and [2] the following definition of H_2 performance in the context of sampled-data control is proposed. First, it is observed that the closed-loop system resulting from a sampled-data controller with sampling period Δ is always a time-varying, Δ-periodic

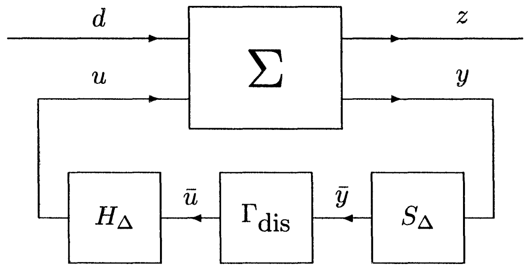

FIG. 1.

system. Then, for Δ-periodic systems the notion of H_2 performance is defined as follows. Suppose we have a finite-dimensional, time-varying, Δ-periodic system Σ_{per} described by

$$z(t) = \int_0^t G(t,s)d(s)ds. \tag{2.1}$$

It is argued in [12] and [2] that a natural way to define the H_2 performance of (2.1) is

$$\|\Sigma_{\text{per}}\|_2^2 := \frac{1}{\Delta}\int_0^\Delta \operatorname{tr} \int_s^\infty G^{\mathrm{T}}(t,s)G(t,s)dt\,ds. \tag{2.2}$$

Next, if Γ is a sampled-data controller with sampling period Δ, the associated performance is defined as $J_{\Sigma,\Delta}(\Gamma) := \|\Sigma \times \Gamma\|_2^2$, the H_2 performance of the (Δ-periodic) closed-loop system $\Sigma \times \Gamma$. The sampled-data H_2 problem is then to minimize, for a fixed sampling period Δ, the performance criterion $J_{\Sigma,\Delta}(\Gamma)$ over all internally stabilizing sampled-data controllers Γ with sampling period Δ. It was shown in [12] and [2] that this problem can be reduced to a discrete-time 'normal' H_2 optimal control problem. To be specific, let the plant Σ be given by the equations

$$\begin{aligned} \dot{x}(t) &= Ax(t) + Bu(t) + Ed(t)\ , \\ y(t) &= C_1x(t)\ , \\ z(t) &= C_2x(t) + D_2u(t)\ , \end{aligned} \tag{2.3}$$

with $x(t) \in \mathbb{R}^n$, $u(t) \in \mathbb{R}^m$, $d(t) \in \mathbb{R}^r$, $y(t) \in \mathbb{R}^p$, and $z(t) \in \mathbb{R}^q$. It will be a standing assumption in this paper that (A,B) is stabilizable and that (C_1,A) is detectable, both with respect to $\mathcal{C}^- := \{s \in \mathcal{C} \mid \Re e\ s < 0\}$. Introduce a finite-dimensional linear time-invariant discrete-time system Σ_Δ:

$$\begin{aligned} x_{k+1} &= A_\Delta x_k + B_\Delta u_k + E_\Delta d_k\ , \\ y_k &= C_1 x_k\ , \\ z_k &= C_{2,\Delta}x_k + D_{2,\Delta}u_k\ , \end{aligned} \tag{2.4}$$

where we define

$$A_\Delta := e^{\Delta A}, \quad B_\Delta := \int_0^\Delta e^{tA}dtB,$$

where E_Δ is any matrix satisfying

$$E_\Delta E_\Delta^{\mathrm{T}} = \int_0^\Delta e^{tA} E E^{\mathrm{T}} e^{tA^{\mathrm{T}}} \, dt, \tag{2.5}$$

and where $C_{2,\Delta}$ and $D_{2,\Delta}$ are matrices satisfying

$$(C_{2,\Delta} \quad D_{2,\Delta})^{\mathrm{T}} (C_{2,\Delta} \quad D_{2,\Delta}) = \int_0^\Delta e^{t\underline{A}^{\mathrm{T}}} (C_2 \quad D_2)^{\mathrm{T}} (C_2 \quad D_2) \, e^{t\underline{A}} \, dt. \tag{2.6}$$

Here we have denoted

$$\underline{A} := \begin{pmatrix} A & B \\ 0 & 0 \end{pmatrix}. \tag{2.7}$$

Let $\mathbf{\Delta}$ denote the set of sampling periods for which either (A_Δ, B_Δ) is not stabilizable or (C_1, A_Δ) is not detectable, both with respect to the open unit disc $\{z \in \mathcal{C} \mid |z| < 1\}$. It is well known [13], [8] that if (A, B) is stabilizable and (C_1, A) is detectable, then every bounded subset of $\mathbb{R}^+$ contains only finitely many elements of $\mathbf{\Delta}$. We will restrict ourselves to sampling periods that are not in $\mathbf{\Delta}$. The plant Σ is controlled using sampled-data controllers $\Gamma := H_\Delta \Gamma_{\mathrm{dis}} S_\Delta$, with Γ_{dis} given by the equations

$$\begin{aligned} w_{k+1} &= K w_k + L y_k \,, \\ u_k &= M w_k + N y_k \,. \end{aligned} \tag{2.8}$$

Let us denote by $J_{\Sigma_\Delta}(\Gamma_{\mathrm{dis}})$ the discrete-time H_2 performance of the closed-loop system $\Sigma_\Delta \times \Gamma_{\mathrm{dis}}$, i.e., the value $\sum_k \mathrm{tr}\ (G_k G_k^{\mathrm{T}})$, where $\{G_k\}$ denotes the pulse response of the closed-loop system. The main result of [12] and [2] is the following:

THEOREM 2.1. *Assume that $\Delta \notin \mathbf{\Delta}$. Then there exists a sampled-data controller Γ with sampling period Δ such that the closed-loop system $\Sigma \times \Gamma$ is internally stable. The sampled-data controller $\Gamma = H_\Delta \Gamma_{\mathrm{dis}} S_\Delta$ internally stabilizes Σ if and only if the discrete-time controller Γ_{dis} internally stabilizes Σ_Δ. Furthermore, for every such controller we have*

$$J_{\Sigma,\Delta}(\Gamma) = \frac{1}{\Delta} \int_0^\Delta \int_0^{\Delta - s} \mathrm{tr}\ \left(C_2 e^{tA} E E^{\mathrm{T}} e^{tA^{\mathrm{T}}} C_2^{\mathrm{T}} \right) dt\, ds + \frac{1}{\Delta} J_{\Sigma_\Delta}(\Gamma_{\mathrm{dis}}).$$

We shall use this theorem as a starting point and study in this paper the discrete-time H_2 optimal control problem for the discrete-time system Σ_Δ given by (2.4). This H_2 problem is inherently singular, due to the fact that the direct feedthrough matrix from the disturbance input to the measured output is always equal to zero.

We conclude this section by introducing some notation and recalling some basic concepts. In this paper, any given continuous-time system $\dot{x} = Ax + Bu, y = Cx + Du$ or discrete-time system $x_{k+1} = Ax_k + Bu_k, y_k = Cx_k + Du_k$ will be denoted simply by (A, B, C, D). It will be clear from the context which interpretation we have in mind. For any such system, the *system matrix* is defined as the first-order polynomial matrix

$$P(s) = \begin{pmatrix} sI - A & -B \\ C & D \end{pmatrix}.$$

If the underlying system is discrete-time, we will rather use the indeterminate z instead of s. For a real rational matrix R, its *normal rank*, normrank R, is defined as the rank of R as a matrix with entries in the field of real rational functions. It is well known that normrank $R = \max_\sigma \operatorname{rank} R(\sigma)$. A *zero* of the system (A, B, C, D) is any complex number λ with the property that rank $P(\lambda) <$ normrank P. The system (A, B, C, D) is called left-invertible (right-invertible) if its transfer matrix $G(s) = C(sI - A)^{-1}B + D$ is a left-invertible (right-invertible) rational matrix. Assuming that $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{m\times n}$, and $C \in \mathbb{R}^{p\times n}$ we have that (A, B, C, D) is left-invertible (right-invertible) if and only if its system matrix has normal rank $n + m$ $(n + p)$.

If $M \in \mathbb{R}^{n\times n}$ and $\mathcal{L}$ is a subspace of $\mathbb{R}^n$, then $\langle M \mid \mathcal{L}\rangle$ will denote the smallest M-invariant subspace containing $\mathcal{L}$. The largest M-invariant subspace contained in $\mathcal{L}$ will be denoted by $\langle \mathcal{L} \mid M\rangle$. In particular, given (A, B, C, D), the reachable subspace is equal to $\langle A \mid \operatorname{im} B\rangle$ and the unobservable subspace is equal to $\langle \ker C \mid A\rangle$.

Given the system (A, B, C, D), we define *the weakly unobservable subspace* $\mathcal{V}$ to be the smallest subspace $\mathcal{L}$ of $\mathbb{R}^n$ with the property that there exists $F \in \mathbb{R}^{m\times n}$ such that $(A + BF)\mathcal{L} \subset \mathcal{L}$ and $(C + DF)\mathcal{L} = 0$ (see [14]). In addition, *the controllability subspace* $\mathcal{R}$ of (A, B, C, D) is defined as follows:

$$\mathcal{R} := \langle A + BF \mid \mathcal{V} \cap B \ker D\rangle,$$

for any F such that $(A + BF)\mathcal{V} \subset \mathcal{V}$ and $(C + DF)\mathcal{V} = 0$ (any such F yields the same $\mathcal{R}$). It was shown in [14] that the system (A, B, C, D) is left-invertible if and only if $\ker B \cap \ker D = 0$ and $\mathcal{V} \cap B \ker D = 0$. Note that $\mathcal{V} \cap B \ker D = 0$ if and only if $\mathcal{R} = 0$.

Finally, the set of zeros of (A, B, C, D) can be shown to be equal to $\sigma(A + BF \mid \mathcal{V}/\mathcal{R})$, for any F such that $(A + BF)\mathcal{V} \subset \mathcal{V}$ and $(C + DF)\mathcal{V} = 0$. Here, $A + BF \mid \mathcal{V}/\mathcal{R}$ is the quotient map of $A + BF \mid \mathcal{V}$ modulo $\mathcal{R}$ (see, e.g., [19]).

3. The discrete-time H_2 problem: No zeros on the unit circle. In this section we shall consider the discrete-time H_2 problem. Consider the finite-dimensional, linear, time-invariant, discrete-time system Σ_{dis} given by the equations

$$\begin{aligned} x_{k+1} &= Ax_k + Bu_k + Ed_k\,, \\ y_k &= C_1x_k + D_1d_k\,, \\ z_k &= C_2x_k + D_2u_k\,. \end{aligned} \tag{3.1}$$

There will be no assumptions on the direct feedthrough matrices D_1 and D_2. In the present section, however, we will have assumptions on the absence of system zeros on the unit circle in the complex plane: it will be assumed that (A, B, C_2, D_2) and (A, E, C_1, D_1) do not have zeros on the unit circle $|z| = 1$. In the next section we will drop these assumptions and treat the completely general case. Of course, it will be a standing assumption that (A, B) is stabilizable and that (C_1, A) is detectable, both with respect to the open unit disc.

We will consider discrete-time controllers Γ_{dis} given by (2.8). For any internally stabilizing controller Γ_{dis}, let $J_{\Sigma_{\text{dis}}}(\Gamma_{\text{dis}})$ be its H_2 performance. Denote by J^* the optimal performance, i.e., the infimum over all internally stabilizing controllers Γ_{dis}.

For a given matrix M, we will denote by M^+ its Moore–Penrose inverse. The solution of the discrete-time H_2 optimal control problem centers around the following two algebraic Riccati equations:

$$P = A^{\mathrm{T}}PA + C_2^{\mathrm{T}}C_2 - (C_2^{\mathrm{T}}D_2 + A^{\mathrm{T}}PB)(D_2^{\mathrm{T}}D_2 + B^{\mathrm{T}}PB)^+(D_2^{\mathrm{T}}C_2 + B^{\mathrm{T}}PA), \tag{3.2}$$

$$Q = AQA^{\mathrm{T}} + EE^{\mathrm{T}} - (AQC_1^{\mathrm{T}} + ED_1^{\mathrm{T}})(D_1D_1^{\mathrm{T}} + C_1QC_1^{\mathrm{T}})^+(D_1E^{\mathrm{T}} + C_1QA^{\mathrm{T}}). \tag{3.3}$$

For any real symmetric matrix P, we shall denote

$$D_P := (D_2^{\mathrm{T}} D_2 + B^{\mathrm{T}} P B)^{\frac{1}{2}}, \tag{3.4}$$

$$C_P := D_P^{+}(D_2^{\mathrm{T}} C_2 + B^{\mathrm{T}} P A). \tag{3.5}$$

Note that, since for any matrix $M \geq 0$ we have $(M^{\frac{1}{2}})^{+} = (M^{+})^{\frac{1}{2}}$, we have $D_P^{+} C_P = (D_2^{\mathrm{T}} D_2 + B^{\mathrm{T}} P B)^{+}(D_2^{\mathrm{T}} C_2 + B^{\mathrm{T}} P A)$. If, in addition, P is a real symmetric solution of (3.2), then $C_P^{\mathrm{T}} C_P = A^{\mathrm{T}} P A - P + C_2^{\mathrm{T}} C_2$. Note also that D_P is symmetric by definition. Finally, since im $(D_2^{\mathrm{T}} C_2 + B^{\mathrm{T}} P A) \subset$ im D_P, we have $D_P C_P = D_2^{\mathrm{T}} C_2 + B^{\mathrm{T}} P A$. (Note that it is a property of the Moore–Penrose inverse that MM^{+} is the orthogonal projection onto im M.)

The following is a corrected and slightly extended version of a theorem from [14]. A proof can be given along the lines of the proof of [14, Thm. 18].

THEOREM 3.1. *Consider the system* (A, B, C_2, D_2) *together with the algebraic Riccati equation* (3.2). *The following two statements are equivalent :*

(i) (A, B) *is stabilizable and* (A, B, C_2, D_2) *has no zeros on the unit circle* $|z| = 1$,

(ii) *Equation* (3.2) *has a real symmetric solution* P *with the following property: there exists a matrix* F_1 *such that*

$$|\sigma(A - BD_P^{+} C_P + B(I - D_P^{+} D_P) F_1)| < 1. \tag{3.6}$$

Furthermore, if P *satisfies this condition, it is the unique real symmetric solution of* (3.2) *for which this condition holds. In addition,* P *is positive semidefinite and is in fact the largest real symmetric solution of* (3.2).

Next we consider the dual algebraic Riccati equation (3.3). For any real symmetric matrix Q, denote

$$D_Q := (D_1 D_1^{\mathrm{T}} + C_1 Q C_1^{\mathrm{T}})^{\frac{1}{2}}, \tag{3.7}$$

$$E_Q := (AQC_1^{\mathrm{T}} + ED_1^{\mathrm{T}}) D_Q^{+}. \tag{3.8}$$

By dualizing the previous theorem, the corresponding result on the Riccati equation (3.3) can be found:

THEOREM 3.2. *Consider the system* (A, E, C_1, D_1) *together with the algebraic Riccati equation* (3.3). *The following two statements are equivalent :*

(i) (C_1, A) *is detectable and* (A, E, C_1, D_1) *has no zeros on the unit circle* $|z| = 1$.

(ii) *Equation* (3.3) *has a real symmetric solution* Q *with the following property: there exists a matrix* K_1 *such that*

$$|\sigma(A - E_Q D_Q^{+} C_1 + K_1(I - D_Q D_Q^{+}) C_1)| < 1. \tag{3.9}$$

Furthermore, if Q *satisfies this condition, it is the unique real symmetric solution of* (3.3) *for which this condition holds. In addition,* Q *is positive semidefinite and is in fact the largest real symmetric solution of* (3.3).

In the remainder of this section we will always denote by P and Q the largest real symmetric solution of (3.2) and (3.3), respectively. Now we will state the main result of this section:

THEOREM 3.3. *Consider the system* (3.1). *Assume that* (A, B) *is stabilizable and* (C_1, A) *is detectable. Assume that* (A, B, C_2, D_2) *and* (A, E, C_1, D_1) *have no zeros on the unit circle. Then we have the following:*

(i)

$$J^* = \text{tr}\,(E^{\mathrm{T}}PE) + \text{tr}\,(C_PQC_P^{\mathrm{T}}) - \text{tr}\,((D_PN^*D_Q)(D_PN^*D_Q)^{\mathrm{T}}), \tag{3.10}$$

where N^* *is defined by*

$$N^* := -(D_P^+)^2(D_PC_PQC_1^{\mathrm{T}} + B^{\mathrm{T}}PED_1^{\mathrm{T}})(D_Q^+)^2. \tag{3.11}$$

(ii) *There exists an optimal controller, i.e., an internally stabilizing controller* Γ^*_{dis} *such that* $J_{\Sigma_{\text{dis}}}(\Gamma^*_{\text{dis}}) = J^*$. *One such optimal controller is given by the following "construction":*

(a) *Choose a state feedback matrix* F *such that* $|\sigma(A+BF)| < 1$ *and* $C_P + D_PF = 0$.
(b) *Choose an output injection matrix* G *such that* $|\sigma(A+GC_1)| < 1$ *and* $E_Q + GD_Q = 0$.
(c) *Define* $\Gamma^*_{\text{dis}} = (K^*, L^*, M^*, N^*)$ *by choosing* N^* *given by* (3.11), *and by choosing* $K^* := A + BF + GC_1 - BN^*C_1$, $L^* := BN^* - G$, *and* $M^* := F - N^*C_1$.

In the remainder of this section we shall prove this theorem. In addition to the system Σ_{dis}, consider the system $\Sigma_{\text{dis},P}$ given by the equations

$$\begin{aligned} x_{k+1} &= Ax_k + Bu_k + Ed_k\,, \\ y_k &= C_1x_k + D_1d_k, \\ z_k &= C_Px_k + D_Pu_k\,, \end{aligned} \tag{3.12}$$

with P the largest real symmetric solution of the algebraic Riccati equation (3.2). The following basic lemma can be proven by a standard completion-of-the-squares argument:

LEMMA 3.4. *For every compensator* $\Gamma_{\text{dis}} = (K, L, M, N)$ *we have* Γ_{dis} *internally stabilizes* Σ_{dis} *if and only if* Γ_{dis} *internally stabilizes* $\Sigma_{\text{dis},P}$. *For any such compensator we have*

$$J_{\Sigma_{\text{dis}}}(\Gamma_{\text{dis}}) = \text{tr}\,(E^{\mathrm{T}}PE) + 2\text{tr}\,(D_1^{\mathrm{T}}N^{\mathrm{T}}B^{\mathrm{T}}PE) + J_{\Sigma_{\text{dis},P}}(\Gamma_{\text{dis}}). \tag{3.13}$$

In addition to $\Sigma_{\text{dis},P}$ we consider the system $\Sigma_{\text{dis},P,Q}$ defined by

$$\begin{aligned} x_{k+1} &= Ax_k + Bu_k + E_Qd_k\,, \\ y_k &= C_1x_k + D_Qd_k, \\ z_k &= C_Px_k + D_Pu_k\,, \end{aligned} \tag{3.14}$$

with Q the largest real symmetric solution of the dual algebraic Riccati equation (3.3). It is clear that the H_2 performance of a given compensator Γ_{dis} applied to Σ_{dis} is equal to the H_2 performance of the dual compensator $\Gamma^{\mathrm{T}}_{\text{dis}} := (K^{\mathrm{T}}, M^{\mathrm{T}}, L^{\mathrm{T}}, N^{\mathrm{T}})$ applied to the dual system $\Sigma^{\mathrm{T}}_{\text{dis}}$. By applying Lemma 3.4 to the dual system $\Sigma^{\mathrm{T}}_{\text{dis},P}$ and the dual compensator $\Gamma^{\mathrm{T}}_{\text{dis}}$ we thus arrive at the following theorem:

THEOREM 3.5. *For every compensator* $\Gamma_{\text{dis}} = (K, L, M, N)$ *we have :* Γ_{dis} *internally stabilizes* Σ_{dis} *if and only if* Γ_{dis} *internally stabilizes* $\Sigma_{\text{dis},P,Q}$. *For any such compensator we have*

$$\begin{aligned} J_{\Sigma_{\text{dis}}}(\Gamma_{\text{dis}}) = \text{tr}\,(E^{\mathrm{T}}PE) + \text{tr}\,(C_PQC_P^{\mathrm{T}}) + 2\text{tr}\,(D_1^{\mathrm{T}}N^{\mathrm{T}}B^{\mathrm{T}}PE) \\ + 2\text{tr}\,(C_PQC_1^{\mathrm{T}}N^{\mathrm{T}}D_P^{\mathrm{T}}) + J_{\Sigma_{\text{dis},P,Q}}(\Gamma_{\text{dis}}). \end{aligned}$$

Now note that in the above formula the first two terms do not depend on the compensator Γ_{dis}. The remaining three terms do depend on the compensator. Also note that in the closed-loop system $\Sigma_{\mathrm{dis},P,Q} \times \Gamma_{\mathrm{dis}}$ the direct feedthrough matrix from the disturbance input to the output to be controlled is equal to $D_P N D_Q$. As a consequence, $J_{\Sigma_{\mathrm{dis},P,Q}}(\Gamma_{\mathrm{dis}}) \geq \mathrm{tr}\ ((D_P N D_Q)(D_P N D_Q)^{\mathrm{T}})$, with equality if and only if the transfer matrix $G_{P,Q,\Gamma_{\mathrm{dis}}}(z)$ of the closed-loop system $\Sigma_{\mathrm{dis},P,Q} \times \Gamma_{\mathrm{dis}}$ is equal to the constant matrix $D_P N D_Q$. It thus follows immediately from Theorem 3.5 that

LEMMA 3.6. *For every internally stabilizing compensator* $\Gamma_{\mathrm{dis}} = (K, L, M, N)$ *we have*

$$J_{\Sigma_{\mathrm{dis}}}(\Gamma_{\mathrm{dis}}) \geq \mathrm{tr}\ (E^{\mathrm{T}} P E) + \mathrm{tr}\ (C_P Q C_P^{\mathrm{T}}) + 2\mathrm{tr}\ (D_1^{\mathrm{T}} N^{\mathrm{T}} B^{\mathrm{T}} P E)$$

$$+ 2\mathrm{tr}\ (C_P Q C_1^{\mathrm{T}} N^{\mathrm{T}} D_P^{\mathrm{T}}) + \mathrm{tr}\ ((D_P N D_Q)(D_P N D_Q)^{\mathrm{T}}),$$

with equality if and only if $G_{P,Q,\Gamma_{\mathrm{dis}}}(z) = D_P N D_Q$.

This lemma shows that, in order to minimize $J_{\Sigma_{\mathrm{dis}}}(\Gamma_{\mathrm{dis}})$ over all internally stabilizing compensators, we should do the following:

(i) First minimize the quadratic matrix function

$$\Phi(N) := 2\mathrm{tr}\ (D_1^{\mathrm{T}} N^{\mathrm{T}} B^{\mathrm{T}} P E) + 2\mathrm{tr}\ (C_P Q C_1^{\mathrm{T}} N^{\mathrm{T}} D_P^{\mathrm{T}}) + \mathrm{tr}\ ((D_P N D_Q)(D_P N D_Q)^{\mathrm{T}}), \tag{3.15}$$

yielding an optimal N^*.

(ii) Next find a compensator Γ^*_{dis}, described by the quadruple (K^*, L^*, M^*, N^*), that is internally stabilizing and yields $G_{P,Q,\Gamma^*_{\mathrm{dis}}}(z) = D_P N^* D_Q$, i.e., the closed-loop system $\Sigma_{\mathrm{dis},P,Q} \times \Gamma^*_{\mathrm{dis}}$ has the constant transfer matrix $D_P N^* D_Q$.

Indeed, if N^* minimizes $\Phi(N)$ and if $G_{P,Q,\Gamma^*_{\mathrm{dis}}}(z) = D_P N^* D_Q$, then we have

$$J_{\Sigma_{\mathrm{dis}}}(\Gamma^*_{\mathrm{dis}}) = \mathrm{tr}\ (E^{\mathrm{T}} P E) + \mathrm{tr}\ (C_P Q C_P^{\mathrm{T}}) + \Phi(N^*),$$

while for any internally stabilizing compensator $\Gamma_{\mathrm{dis}} = (K, L, M, N)$ we have

$$J_{\Sigma_{\mathrm{dis}}}(\Gamma_{\mathrm{dis}}) \geq \mathrm{tr}\ (E^{\mathrm{T}} P E) + \mathrm{tr}\ (C_P Q C_P^{\mathrm{T}}) + \Phi(N) \geq \mathrm{tr}\ (E^{\mathrm{T}} P E) + \mathrm{tr}\ (C_P Q C_P^{\mathrm{T}}) + \Phi(N^*).$$

This clearly implies that

$$J^* = \mathrm{tr}\ (E^{\mathrm{T}} P E) + \mathrm{tr}\ (C_P Q C_P^{\mathrm{T}}) + \Phi(N^*)$$

and that

$$J_{\Sigma_{\mathrm{dis}}}(\Gamma^*_{\mathrm{dis}}) = J^*.$$

We will first study the minimization of $\Phi(N)$.

LEMMA 3.7. *Let* $\Phi(N)$ *be defined by* (3.15). *Define*

$$R^* := D_P^+ (D_P C_P Q C_1^{\mathrm{T}} + B^{\mathrm{T}} P E D_1^{\mathrm{T}}) D_Q^+.$$

Then

$$\Phi^* := \min\{\Phi(N) \mid N \in \mathbb{R}^{m \times p}\} = -\mathrm{tr}\ (R^* R^{*T}).$$

N minimizes Φ, *i.e.*, $\Phi(N) = \Phi^*$, *if and only if N is a solution to the linear equation* $D_P N D_Q = -R^*$. *One particular solution of this linear equation is given by* $N^* = -D_P^+ R^* D_Q^+$. *We have* $\Phi^* = -\mathrm{tr}\ ((D_P N^* D_Q)(D_P N^* D_Q)^{\mathrm{T}})$.

Proof. Using the facts that

$$\ker D_Q \subset \ker (D_P C_P Q C_1^{\mathrm{T}} + B^{\mathrm{T}} P E D_1^{\mathrm{T}}),$$
$$\mathrm{im}\ D_P \supset \mathrm{im}\ (D_P C_P Q C_1^{\mathrm{T}} + B^{\mathrm{T}} P E D_1^{\mathrm{T}}),$$

it can be shown by straightforward calculation that

$$\Phi(N) = -\mathrm{tr}\ (R^* R^{*T}) + \mathrm{tr}\ ((D_P N D_Q + R^*)(D_P N D_Q + R^*)^{\mathrm{T}}).$$

The equation $D_P N D_Q = -R^*$ has a solution since $\ker D_Q = \ker D_Q^{\mathrm{T}} = \ker D_Q^+ \subset \ker R^*$ and $\mathrm{im}\ D_P = \mathrm{im}\ D_P^{\mathrm{T}} = \mathrm{im}\ D_P^+ \supset \mathrm{im}\ R^*$. Clearly, one particular solution is then given by $N^* = -D_P^+ R^* D_Q^+$. Finally, the expression for Φ^* can be checked in a straightforward manner. □

Next we study the question whether, starting with N^* above, it is possible to find K^*, L^*, M^* such that the resulting compensator $\Gamma_{\mathrm{dis}}^* = (K^*, L^*, M^*, N^*)$ yields a closed-loop system $\Sigma_{\mathrm{dis},P,Q} \times \Gamma_{\mathrm{dis}}^*$ with constant transfer matrix $D_P N^* D_Q$. We will first prove the following lemma:

LEMMA 3.8. *Assume that (A, B) is stabilizable and that (A, B, C_2, D_2) has no zeros on the unit circle. Let P be the largest real symmetric solution of the algebraic Riccati equation* (3.2). *There exists a matrix F such that*

(i) $|\sigma(A + BF)| < 1$,

(ii) $C_P + D_P F = 0$.

Proof. Let F_1 be such that (3.6) holds, and define $F := -D_P^+ C_P + (I - D_P^+ D_P) F_1$. Then (i) is satisfied. To prove (ii), note that $\mathrm{im}\ C_P \subset \mathrm{im}\ D_P^+ = \mathrm{im}\ D_P$. Consequently, $-D_P D_P^+ C_P = -C_P$, which proves (ii). □

We will also need the dual of this lemma, which reads as follows:

LEMMA 3.9. *Assume that (C_1, A) is detectable and that (A, E, C_1, D_1) has no zeros on the unit circle. Let Q be the largest real symmetric solution of the dual algebraic Riccati equation* (3.3). *There exists a matrix G such that*

(i) $|\sigma(A + GC_1)| < 1$,

(ii) $E_Q + G D_Q = 0$.

Now we show that by suitable choice of compensator Γ_{dis}, the transfer matrix of $\Sigma_{\mathrm{dis},P,Q} \times \Gamma_{\mathrm{dis}}$ can be made equal to any constant matrix product $M_1 M_2$, as long as $\mathrm{im}\ D_P \subset \mathrm{im}\ M_1$ and $\ker D_Q \subset \ker M_2$.

LEMMA 3.10. *Consider the system* (3.1). *Assume that (A, B) is stabilizable and (C_1, A) is detectable. Assume that (A, B, C_2, D_2) and (A, E, C_1, D_1) have no zeros on the unit circle. Let P and Q be the largest real symmetric solution of the algebraic Riccati equation* (3.2) *and* (3.3), *respectively. Then for any pair of matrices M_1, M_2 such that the product $M_1 M_2$ is defined and such that* $\mathrm{im}\ D_P \subset \mathrm{im}\ M_1$ *and* $\ker D_Q \subset \ker M_2$ *there exists an internally stabilizing compensator Γ_{dis} such that the transfer matrix of $\Sigma_{\mathrm{dis},P,Q} \times \Gamma_{\mathrm{dis}}$ is equal to the constant $M_1 M_2$.*

Specifically, for given M_1 and M_2 let F_2 be a solution of $M_1 = D_P F_2$ and G_2 be a solution of $M_2 = -G_2 D_Q$ and take F such that the conditions in Lemma 3.8 *are satisfied and G such that the conditions of Lemma* 3.9 *are satisfied. Then the compensator $\Gamma_{\mathrm{dis}} := (K, L, M, N)$ with $K := A + BF + GC_1 + BF_2 G_2 C_1$, $L := -BF_2 G_2 - G$, $M := F + F_2 G_2 C_1$, and $N := -F_2 G_2$ satisfies the requirements.*

Proof. The equations of the compensator are given by (2.8). Using the specifications of K, L, M, and N given above, we find that the error $e_k := w_k - x_k$ satisfies $e_{k+1} = (A + GC_1) e_k$. Thus, if $w_0 = 0$ and $x_0 = 0$, we have $x_k = w_k$ for all k. In particular, this implies that $u_k = F x_k + F_2 M_2 w_k$. The output of the closed-loop system is then equal to $z_k = C_P x_k + D_P u_k = M_1 M_2 w_k$. This implies that the closed-loop transfer matrix is equal to the constant matrix $M_1 M_2$. Finally, the spectrum of the closed-loop system matrix A_e is easily shown to be equal to $\sigma(A+BF) \cup \sigma(A+GC_1)$. This implies that the closed-loop system is internally stable. □

Clearly, if in this lemma we take $M_1 = D_P$ and $M_2 = N^* D_Q$, we arrive at an internally stabilizing compensator Γ_{dis} such that the closed-loop transfer matrix is equal to the constant matrix $D_P N^* D_Q$. In the formulas for the compensator as given in the lemma, we should then take $F_2 = I$ and $G_2 = -N^*$. The result of Theorem 3.3 follows immediately by combining the above lemmas.

Remark 3.11. For later use we note that Lemma 3.8 also provides a resolution of the discrete-time *linear quadratic problem* for the case that (A, B, C_2, D_2) has no zeros on the unit circle (see also [14]). Given $x_{k+1} = Ax_k + Bu_k$, the problem is to minimize the cost-functional $J(x_0, u) := \sum_k \|(C_2 x_k + D_2 u_k\|^2$ over all inputs $u = \{u_k\}$ such that $x_k \to 0$. It was pointed out in [14] that for each such input u we have the completion-of-the-squares formula $J(x_0, u) = x_0^{\mathrm{T}} P x_0 + J_P(x_0, u)$, with $J_P(x_0, u) := \sum_k \|C_P x_k + D_P u_k\|^2$. Thus, if we take F satisfying (i) and (ii) of Lemma 3.8, then the input $u_k = F x_k$ leads to the optimal cost $J^*(x_0) = x_0^{\mathrm{T}} P x_0$. Note that we could also formulate the linear quadratic problem as a minimization over all internally stabilizing feedback laws: minimize the cost-functional $J(x_0, F) := \sum_k \|(C_P + D_P F) x_k\|^2$ over all $F \in \mathbb{R}^{m \times n}$ such that $|\sigma(A+BF)| < 1$. By the above argument, any F satisfying (i) and (ii) of Lemma 3.8 is then optimal and the optimal cost is again given by $x_0^{\mathrm{T}} P x_0$.

Remark 3.12. An interesting question is under what conditions the Moore–Penrose inverse $(D_2^{\mathrm{T}} D_2 + B^{\mathrm{T}} P B)^+$ reduces to the inverse $(D_2^{\mathrm{T}} D_2 + B^{\mathrm{T}} P B)^{-1}$, equivalently, under what conditions $D_2^{\mathrm{T}} D_2 + B^{\mathrm{T}} P B$ is positive definite. Using the ideas from [14] it can be shown that if P is a positive semidefinite solution of the algebraic Riccati equation (3.2), then $D_2^{\mathrm{T}} D_2 + B^{\mathrm{T}} P B > 0$ if and only if (A, B, C_2, D_2) is a left-invertible system. Of course, dually, if Q is a positive semidefinite solution of the algebraic Riccati equation (3.3), then $D_1 D_1^{\mathrm{T}} + C_1 Q C_1^{\mathrm{T}} > 0$ if and only if the system (A, E, C_1, D_1) is right-invertible. In view of this, it is perhaps more natural to call the discrete-time H_2 problem regular if (A, B, C_2, D_2) is a left-invertible system and (A, E, C_1, D_1) is a right-invertible system.

4. The discrete-time H_2 problem: The general case. In this section we will extend the results of the previous section and treat the discrete-time H_2 problem in its full generality. This means that we will drop the assumption on the absence of zeros on the unit circle that was made in the previous section. First we will prove that also without the assumption that (A, B, C_2, D_2) has no zeros on the unit circle, the Riccati equation (3.2) has a largest real symmetric solution. We will prove that this solution can be obtained as the limit of solutions of algebraic Riccati equations associated with suitable perturbations of the system (A, B, C_2, D_2).

THEOREM 4.1. *If (A, B) is stabilizable, then the Riccati equation* (3.2) *has a largest real symmetric solution, say P. P is positive semidefinite. We have $P = \lim_{\varepsilon \downarrow 0} P_\varepsilon$, where for $\varepsilon > 0$ P_ε is the largest real symmetric solution of the algebraic Riccati equation*

$$\begin{aligned} &A^{\mathrm{T}} P_\varepsilon A - P_\varepsilon + C_2^{\mathrm{T}} C_2 + \varepsilon^2 I \\ &\quad -(A^{\mathrm{T}} P_\varepsilon B + C_2^{\mathrm{T}} D_2)(D_2^{\mathrm{T}} D_2 + B^{\mathrm{T}} P_\varepsilon B)^+ (B^{\mathrm{T}} P_\varepsilon A + D_2^{\mathrm{T}} C_2) = 0. \end{aligned} \tag{4.1}$$

Remark 4.2. Note that (4.1) is the Riccati equation associated with the perturbed system $(A, B, \binom{C_2}{\varepsilon I}, \binom{D_2}{0})$. (Here, I denotes the $n \times n$ identity matrix, and 0 denotes the $n \times m$ zero matrix). For $\varepsilon > 0$, the perturbed system has no zeros. Consequently, the existence of P_ε follows from Theorem 3.1.

The idea of the proof of Theorem 4.1 is to show first that the P_ε indeed converge to some matrix P and next to show that P satisfies (3.2). The difficulty is that in the general case we are considering, the term $D_2^{\mathrm{T}}D_2 + B^{\mathrm{T}}PB$ need not be invertible, so that we cannot conclude that $(D_2^{\mathrm{T}}D_2 + B^{\mathrm{T}}P_\varepsilon B)^+$ converges to $(D_2^{\mathrm{T}}D_2 + B^{\mathrm{T}}PB)^+$. We will show, however, that we can get around this difficulty by considering the so-called linear matrix inequality. Our proof is split up in three lemmas. In the following, let $J(x_0, u)$ be the cost-functional of the linear quadratic problem, and let $J^*(x_0)$ be the optimal cost (see Remark 3.11).

LEMMA 4.3. *Let P_ε be the largest real symmetric solution of* (4.1). *There exists a real positive semidefinite matrix P such that $P_\varepsilon \downarrow P$ $(\varepsilon \downarrow 0)$. For all $x_0 \in \mathbb{R}^n$ we have $J^*(x_0) = x_0^{\mathrm{T}}Px_0$.*

Proof. Let $J_\varepsilon(x_0, u) := \sum_k \|C_P x_k + D_P u_k\|^2 + \varepsilon^2\|x_k\|^2$, and let $J_\varepsilon^*(x_0)$ be the infimum of $J_\varepsilon(x_0, u)$ over all u such that $x_k \to 0$. According to Remark 3.11 we have $J_\varepsilon^*(x_0) = x_0^{\mathrm{T}}P_\varepsilon x_0$. From this interpretation it follows that P_ε is monotonically non-increasing as $\varepsilon \downarrow 0$. Being bounded from below by 0, this yields the existence of a limit P. Obviously, for all $\varepsilon > 0$ we have $J^*(x_0) \le J_\varepsilon^*(x_0) = x_0^{\mathrm{T}}P_\varepsilon x_0$, so $J^*(x_0) \le x_0^{\mathrm{T}}Px_0$. Conversely, for all $\varepsilon > 0$ and for all u we have $J_\varepsilon(x_0, u) \ge x_0^{\mathrm{T}}P_\varepsilon x_0$. Taking the limit on both sides this yields $J(x_0, u) \ge x_0^{\mathrm{T}}Px_0$ for all u. Taking the infimum over u then yields the converse inequality. □

LEMMA 4.4. *P is the largest real symmetric solution of the linear matrix inequality*

$$M(P) := \begin{pmatrix} A^{\mathrm{T}}PA - P + C_2^{\mathrm{T}}C_2 & C_2^{\mathrm{T}}D_2 + A^{\mathrm{T}}PB \\ D_2^{\mathrm{T}}C_2 + B^{\mathrm{T}}PA & D_2^{\mathrm{T}}D_2 + B^{\mathrm{T}}PB \end{pmatrix} \ge 0.$$

Proof. Denote the left-hand side of (4.1) by $R_\varepsilon(P_\varepsilon)$. Also consider the linear matrix inequality associated with the perturbed system:

$$M_\varepsilon(P_\varepsilon) := \begin{pmatrix} A^{\mathrm{T}}P_\varepsilon A - P_\varepsilon + C_2^{\mathrm{T}}C_2 + \varepsilon^2 I & C_2^{\mathrm{T}}D_2 + A^{\mathrm{T}}P_\varepsilon B \\ D_2^{\mathrm{T}}C_2 + B^{\mathrm{T}}P_\varepsilon A & D_2^{\mathrm{T}}D_2 + B^{\mathrm{T}}P_\varepsilon B \end{pmatrix} \ge 0.$$

We have $M_\varepsilon(P_\varepsilon) \ge 0$ if and only if $R_\varepsilon(P_\varepsilon) \ge 0$. This follows from the fact that the latter is equal to the Schur complement of $D_2^{\mathrm{T}}D_2 + B^{\mathrm{T}}P_\varepsilon B$ in $M_\varepsilon(P_\varepsilon)$. The Schur complement is defined here with matrix inverse replaced by Moore–Penrose inverse. This can be done because of the fact that

$$\ker(D_2^{\mathrm{T}}D_2 + B^{\mathrm{T}}P_\varepsilon B) \subset \ker(C_2^{\mathrm{T}}D_2 + A^{\mathrm{T}}P_\varepsilon B).$$

Since $R_\varepsilon(P_\varepsilon) = 0$, we indeed have $M_\varepsilon(P_\varepsilon) \ge 0$. Taking the limit $\varepsilon \downarrow 0$ then yields $M(P) \ge 0$. To show that P is the largest real symmetric solution, let P_1 be any real symmetric solution of the linear matrix inequality. Using a standard completion-of-the-squares argument then yields $J(x_0, u) \ge x_0^{\mathrm{T}}P_1x_0$ for any x_0 and any u such that $x_k \to 0$. Taking the infimum over all such u yields $x_0^{\mathrm{T}}Px_0 = J^*(x_0) \ge x_0^{\mathrm{T}}P_1x_0$. □

Now we will show that P in fact satisfies the algebraic Riccati equation (3.2). Denote

$$R(P) := A^{\mathrm{T}}PA - P + C_2^{\mathrm{T}}C_2 - (C_2^{\mathrm{T}}D_2 + A^{\mathrm{T}}PB)(D_2^{\mathrm{T}}D_2 + B^{\mathrm{T}}PB)^+(D_2^{\mathrm{T}}C_2 + B^{\mathrm{T}}P).$$

Again, by the fact that $\ker(D_2^{\mathrm{T}}D_2 + B^{\mathrm{T}}PB) \subset \ker(C_2^{\mathrm{T}}D_2 + A^{\mathrm{T}}PB)$, $R(P)$ is equal to the Schur complement of $D_2^{\mathrm{T}}D_2 + B^{\mathrm{T}}PB$ in $M(P)$. In particular this implies that

$$\operatorname{rank} M(P) = \operatorname{rank}\,(D_2^{\mathrm{T}}D_2 + B^{\mathrm{T}}PB) + \operatorname{rank} R(P).$$

In order to prove that $R(P) = 0$ we should therefore prove that P has the property expressed in the following lemma:

LEMMA 4.5. $\operatorname{rank} M(P) = \operatorname{rank}\,(D_2^{\mathrm{T}}D_2 + B^{\mathrm{T}}PB)$.

Proof. Let $\tilde{C}$ and $\tilde{D}$ be matrices such that

$$M(P) = \begin{pmatrix} \tilde{C} & \tilde{C} \end{pmatrix}^{\mathrm{T}} \begin{pmatrix} \tilde{C} & \tilde{D} \end{pmatrix}.$$

Again using a standard completion-of-the-squares argument, for any initial state x_0 and for any input sequence u such that $x_k \to 0$ we have

$$J(x_0, u) = x_0^{\mathrm{T}}Px_0 + \sum_k \|\tilde{C}x_k + \tilde{D}u_k\|^2 \geq x_0^{\mathrm{T}}x_0 + \|\tilde{C}Px_0 + \tilde{D}u_0\|^2 \tag{4.2}$$

From Lemma 4.3 we have that $J^*(x_0) = x_0^{\mathrm{T}}Px_0$. In particular this implies that the infimum of $\|\tilde{C}x_0 + \tilde{D}u_0\|^2$ over all $u_0 \in \mathbb{R}^m$ is equal to 0. Consequently, for all x_0 there exists $u_0 \in \mathbb{R}^m$ such that $\tilde{C}x_0 + \tilde{D}u_0 = 0$. This implies $\operatorname{im} \tilde{C} \subset \operatorname{im} \tilde{D}$ so

$$\operatorname{rank} M(P) = \operatorname{rank} \begin{pmatrix} \tilde{C} & \tilde{D} \end{pmatrix} = \operatorname{rank} \tilde{D} = \operatorname{rank}\,(D_2^{\mathrm{T}}D_2 + B^{\mathrm{T}}PB). \qquad \square$$

Clearly, the proof of Theorem 4.1 follows by combining these three lemmas. The fact that P is the largest real symmetric solution of the algebraic Riccati equation follows by noting that any real symmetric solution is also a solution of the linear matrix inequality and by applying Lemma 4.4.

Remark 4.6. For later use, note that by combining the above results with Remark 3.11 we obtain that also for the general case the optimal cost $J^*(x_0)$ of the discrete-time linear quadratic problem associated with the system (A, B, C_2, D_2) is given by $J^*(x_0) = x_0^{\mathrm{T}}Px_0$, with P the largest real symmetric solution of the Riccati equation (3.2). We will also need the dual result of Theorem 4.1, which is stated below:

THEOREM 4.7. *If* (C_1, A) *is detectable, then the Riccati equation* (3.3) *has a largest real symmetric solution, say* Q. Q *is positive semidefinite. We have* $Q = \lim_{\varepsilon \downarrow 0} Q_\varepsilon$, *where for* $\varepsilon > 0$ Q_ε *is the largest real symmetric solution of the algebraic Riccati equation*

$$\begin{aligned} &AQ_\varepsilon A^{\mathrm{T}} - Q_\varepsilon + EE^{\mathrm{T}} + \varepsilon^2 I \\ &\quad -(AQ_\varepsilon C_1^{\mathrm{T}} + ED_1^{\mathrm{T}})(D_1D_1^{\mathrm{T}} + C_1Q_\varepsilon C_1^{\mathrm{T}})^+(C_1Q_\varepsilon A^{\mathrm{T}} + D_1E^{\mathrm{T}}) = 0. \end{aligned} \tag{4.3}$$

Now we are in a position to state the main results of this section. It turns out that also for the discrete-time H_2 problem in its full generality, so without any assumptions on the zeros, the optimal performance J^* is given by (3.10), with P and Q the largest real symmetric solutions of the respective Riccati equations. However, in general no optimal controller will exist. We will, however, derive necessary and sufficient conditions for the existence of an optimal controller. Our first main result deals with the optimal performance.

THEOREM 4.8. *Consider the system* (3.1). *Assume that* (A, B) *is stabilizable and* (C_1, A) *is detectable. Then the optimal performance* J^* *is given by* (3.10), *where* P *and* Q *are the largest real symmetric solutions of* (3.2) *and* (3.3), *respectively.*

Proof. In addition to the system (3.1), consider its perturbation $\Sigma^{\varepsilon}_{\text{dis}}$:

$$
\begin{aligned}
x_{k+1} &= \quad Ax_k + \quad Bu_k + (E \;\; \varepsilon I)v_k \;, \\
y_k &= \quad C_1 x_k \qquad\qquad + (D_1 \;\; 0)v_k \;, \\
z_k &= \begin{pmatrix} C_2 \\ \varepsilon I \end{pmatrix} x_k + \begin{pmatrix} D_2 \\ 0 \end{pmatrix} u_k \,.
\end{aligned}
\tag{4.4}
$$

Let $J_{\Sigma^{\varepsilon}_{\text{dis}}}(\Gamma_{\text{dis}})$ denote the H_2 performance, and let J^*_ε denote the optimal H_2 performance. Since, for $\varepsilon > 0$, neither $(A, B, \left(\begin{smallmatrix} C_2 \\ \varepsilon I \end{smallmatrix}\right), \left(\begin{smallmatrix} D_2 \\ 0 \end{smallmatrix}\right))$ nor $(A, (E \;\; \varepsilon I), C_1, (D_1 \;\; 0))$ have zeros; we can apply Theorem 3.3 to obtain

$$
\begin{aligned}
J^*_\varepsilon = \operatorname{tr}\,((EE^{\mathrm{T}} + \varepsilon^2 I)P_\varepsilon) + \operatorname{tr}\,(A^{\mathrm{T}}P_\varepsilon A - P_\varepsilon + C_2^{\mathrm{T}}C_2 + \varepsilon^2 I)Q_\varepsilon) \\
-\operatorname{tr}\,((D_{P_\varepsilon} N^*_\varepsilon D_{Q_\varepsilon})(D_{P_\varepsilon} N^*_\varepsilon D_{Q_\varepsilon})^{\mathrm{T}}),
\end{aligned}
$$

where P_ε and Q_ε are the largest real symmetric solutions of (4.1) and (4.3), respectively, and where D_{P_ε}, N^*_ε, and D_{Q_ε} are defined by (3.4), (3.11), and (3.7), with P and Q replaced by P_ε and Q_ε. From Lemma 3.7, recall that

$$
-\operatorname{tr}\,((D_{P_\varepsilon} N^*_\varepsilon D_{Q_\varepsilon})(D_{P_\varepsilon} N^*_\varepsilon D_{Q_\varepsilon})^{\mathrm{T}}) = \Phi_\varepsilon(N^*_\varepsilon) = \min_N \Phi_\varepsilon(N),
$$

with

$$
\begin{aligned}
\Phi_\varepsilon(N) :=\ & 2\operatorname{tr}\,\left(\begin{pmatrix} D_1 \\ 0 \end{pmatrix}^{\mathrm{T}} N^{\mathrm{T}} B^{\mathrm{T}} P_\varepsilon (E \;\; \varepsilon I) \right) + 2\operatorname{tr}\,(C_{P_\varepsilon} Q_\varepsilon C_1^{\mathrm{T}} N^{\mathrm{T}} D_{P_\varepsilon}) \\
& + \operatorname{tr}\,((D_{P_\varepsilon} N D_{Q_\varepsilon})(D_{P_\varepsilon} N D_{Q_\varepsilon})^{\mathrm{T}}) \\
=\ & 2\operatorname{tr}\,(D_1^{\mathrm{T}} N^{\mathrm{T}} B^{\mathrm{T}} P_\varepsilon E) + 2\operatorname{tr}\,(Q_\varepsilon C_1^{\mathrm{T}} N^{\mathrm{T}} (D_2^{\mathrm{T}} C_2 + B^{\mathrm{T}} P_\varepsilon A)) \\
& + \operatorname{tr}\,((D_{P_\varepsilon} N D_{Q_\varepsilon})(D_{P_\varepsilon} N D_{Q_\varepsilon})^{\mathrm{T}}).
\end{aligned}
$$

Since $P_\varepsilon \to P$ and $Q_\varepsilon \to Q$, we see that for every N we have $\Phi_\varepsilon(N) \to \Phi(N)$ ($\varepsilon \downarrow 0$), where $\Phi(N)$ is defined by (3.15). Since of course for all $\varepsilon > 0$ we have $J^* \leq J^*_\varepsilon$ we see that for all $\varepsilon > 0$, for all N we have

$$
J^* \leq \operatorname{tr}\,((EE^{\mathrm{T}} + \varepsilon^2 I)P_\varepsilon) + \operatorname{tr}\,((A^{\mathrm{T}}P_\varepsilon A - P_\varepsilon + C_2^{\mathrm{T}}C_2 + \varepsilon^2 I)Q_\varepsilon) + \Phi_\varepsilon(N).
$$

Now, letting $\varepsilon \downarrow 0$ on the left in this inequality, we find that for all N

$$
J^* \leq \operatorname{tr}\,(EE^{\mathrm{T}}P) + \operatorname{tr}\,((A^{\mathrm{T}}PA - P + C_2^{\mathrm{T}}C_2)Q) + \Phi(N).
$$

Finally, taking the minimum over all N, this yields

$$
J^* \leq \operatorname{tr}\,(EE^{\mathrm{T}}P) + \operatorname{tr}\,(C_P^{\mathrm{T}}C_P Q) - \operatorname{tr}\,((D_P N^* D_Q)(D_P N^* D_Q)^{\mathrm{T}}).
$$

To prove the converse inequality note that by using the fact that P and Q satisfy (3.2) and (3.3) we can apply a repeated completion-of-the-squares argument as in §3 to obtain that for any internally stabilizing compensator Γ_{dis} we have

$$
J_{\Sigma_{\text{dis}}}(\Gamma_{\text{dis}}) \geq \operatorname{tr}\,(E^{\mathrm{T}}PE) + \operatorname{tr}\,(C_P Q C_P^{\mathrm{T}}) + \Phi(N^*).
\tag{4.5}
$$

Taking the infimum over all such Γ_{dis} yields the desired inequality. □

Next we will study the question: Under what conditions does there exist an optimal controller? Again, let P and Q be the largest real symmetric solutions of the respective Riccati equations. Define a system $\Sigma_{\text{dis},P,Q}$ by (3.14). Again, for any internally stabilizing compensator $\Gamma_{\text{dis}} = (K, L, M, N)$ we have the inequality (4.5).

As noted in §3, we have equality *if* $N = N^*$ and Γ_{dis} has the property that the closed loop system $\Sigma_{\text{dis},P,Q} \times \Gamma_{\text{dis}}$ has the constant transfer matrix $D_P N^* D_Q$. Of course, the latter statement only gives a sufficient condition for a compensator to be optimal. In the following theorem we will give necessary *and* sufficient conditions for optimality. Let R^* be as defined in Lemma 3.7.

THEOREM 4.9. *A controller* Γ_{dis} *is optimal if and only if* $\Sigma_{\text{dis},P,Q} \times \Gamma_{\text{dis}}$ *is internally stable and has constant transfer matrix* R^*.

Proof. If $\Gamma_{\text{dis}} = (K, L, M, N)$ is optimal, then we have

$$J_{\Sigma_{\text{dis}}}(\Gamma_{\text{dis}}) = \text{tr}\,(E^{\mathrm{T}} P E) + \text{tr}\,(C_P^{\mathrm{T}} Q C_P) + \Phi^*.$$

By Lemma 3.6 we also have

$$J_{\Sigma_{\text{dis}}}(\Gamma_{\text{dis}}) \geq \text{tr}\,(E^{\mathrm{T}} P E) + \text{tr}\,(C_P^{\mathrm{T}} Q C_P) + \Phi(N).$$

This clearly yields $\Phi(N) = \Phi^*$, i.e., N minimizes the function Φ. Again by Lemma 3.6 this implies that $\Sigma_{\text{dis},P,Q} \times \Gamma_{\text{dis}}$ has the constant transfer matrix $D_P N D_Q$. However, since N minimizes Φ, by Lemma 3.7 we have $D_P N D_Q = -R^*$. The converse statement is also an immediate consequence of Lemma 3.6. □

Our aim is to reformulate these conditions in terms of the original system Σ_{dis}. For any given matrix $N \in \mathbb{R}^{m \times p}$, consider the system $\Sigma^N_{\text{dis},P,Q}$ that is obtained by applying to $\Sigma_{\text{dis},P,Q}$ the static output feedback $u = Ny + v$. This system $\Sigma^N_{\text{dis},P,Q}$ is described by

$$\begin{array}{lll} x_{k+1} = & (A + BNC_1)x_k + & Bv_k + (BND_Q + E_Q)d_k\ , \\ y_k \quad = & C_1 x_k & + \qquad D_Q d_k\ , \\ z_k \quad = & (C_P + D_P N C_1)x_k + D_P v_k\ , & \end{array} \tag{4.6}$$

Also, for a given compensator $\Gamma_{\text{dis}} = (K, L, M, N)$, let $\Gamma^0_{\text{dis}} := (K, L, M, 0)$ be the compensator with direct feedthrough matrix N replaced by 0. It is clear that the closed-loop system $\Sigma_{\text{dis},P,Q} \times \Gamma_{\text{dis}}$ has constant transfer matrix $D_P N D_Q$ if and only if $\Sigma^N_{\text{dis},P,Q} \times \Gamma^0_{\text{dis}}$ has transfer matrix equal to 0. Consequently, an internally stabilizing compensator $\Gamma_{\text{dis}} = (K, L, M, N)$ is optimal if and only if $D_P N D_Q = -R^*$ and $\Sigma^N_{\text{dis},P,Q} \times \Gamma^0_{\text{dis}}$ has transfer matrix 0. In other words, in order to find necessary and sufficient conditions for the existence of an optimal controller, we should study the problem of disturbance decoupling with internal stability. This problem has been studied extensively in [16]. One of the main results of [16] gives necessary and sufficient conditions for the existence of an internally stabilizing strictly proper compensator Γ^0_{dis} for the system Σ_{dis} given by (3.1). We will briefly recall this result here. Given Σ_{dis}, let $\mathcal{V}_g$ denote the largest subspace of $\mathbb{R}^n$ for which there exists $F \in \mathbb{R}^{m \times n}$ such that $(A + BF)\mathcal{V}_g \subset \mathcal{V}_g$, $|\sigma(A + BF \mid \mathcal{V}_g)| < 1$, and $(C_2 + D_2 F)\mathcal{V}_g = 0$. Dually, let $\mathcal{S}_g$ be the smallest subspace of $\mathbb{R}^n$ for which there exists a matrix $G \in \mathbb{R}^{n \times p}$ such that $(A + GC_1)\mathcal{S}_g \subset \mathcal{S}_g$, $|\sigma(A + GC_1 \mid \mathbb{R}^n/\mathcal{S}_g)| < 1$, and im $(E + GD_1) \subset \mathcal{S}_g$. It was shown in [16, Thm. 2.4] that there exists an internally stabilizing compensator $\Gamma^0_{\text{dis}} = (K, L, M, 0)$ such that $\Sigma_{\text{dis}} \times \Gamma^0_{\text{dis}}$ has transfer matrix 0 if and only if the following conditions hold: (i) (A, B) is stabilizable and (C_1, A) is detectable, (ii) the following four subspace inclusions hold: im $E \subset \mathcal{V}_g$, $\mathcal{S}_g \subset \ker C_2$, $\mathcal{S}_g \subset \mathcal{V}_g$, and $A\mathcal{S}_g \subset \mathcal{V}_g$.

Here, we want to apply this result to the system $\Sigma^N_{\text{dis},P,Q}$, with N any solution of $D_P N D_Q = -R^*$. In the following, we will omit some of the details. Using the fact that im $(C_P + D_P N C_1) \subset$ im D_P, it can be shown that the subspace $\mathcal{V}_g$ associated with $\Sigma^N_{\mathrm{dis},P,Q}$ is given by

$$\mathcal{V}_g = \mathcal{X}_g(A - BD_P^+ C_P) + \langle A - BD_P^+ C_P \mid B \ker D_P \rangle, \tag{4.7}$$

where for a given matrix M, $\mathcal{X}_g(M)$ is the sum of the generalized eigenspaces of M associated with its eigenvalues in $|z| < 1$, and where $\langle M \mid \mathcal{L} \rangle$ is the smallest M-invariant subspace contained in $\mathcal{L}$. It can also be shown, using the fact that $\ker D_Q \subset \ker(BND_Q + E_Q)$, that

$$\mathcal{S}_g = \mathcal{X}_b(A - E_Q D_Q^+ C_1) \cap \langle C_1^{-1} \mathrm{im}\, D_Q \mid A - E_Q D_Q^+ C_1 \rangle, \tag{4.8}$$

where $\mathcal{X}_b(M)$ is the sum of the generalized eigenspaces of M associated with its eigenvalues in $|z| \geq 1$ and where $\langle \mathcal{L} \mid M \rangle$ is the largest M-invariant subspace containing $\mathcal{L}$. Using the fact that, from (4.7), $B \ker D_P \subset \mathcal{V}_g$, it can be shown that im $(BND_Q + E_Q) \subset \mathcal{V}_g$ if and only if

$$\mathrm{im}\,(E_Q - BD_P^+ R^*) \subset \mathcal{V}_g. \tag{4.9}$$

Using the fact that, by (4.8), $\mathcal{S}_g \subset C_1^{-1}\mathrm{im}\, D_Q$, it can be shown that $\mathcal{S}_g \subset \ker(C_P + D_P N C_1)$ if and only if

$$\mathcal{S}_g \subset \ker(C_P - R^* D_Q^+ C_1). \tag{4.10}$$

Finally, it can be shown that $(A + BNC_1)\mathcal{S}_g \subset \mathcal{V}_g$ if and only if

$$(A - BD_P^+ R^* D_Q^+ C_1)\mathcal{S}_g \subset \mathcal{V}_g. \tag{4.11}$$

Collecting the above facts, we then obtain the following necessary and sufficient conditions for the existence of an optimal controller for the discrete-time H_2 optimal control problem associated with the system Σ_{dis}:

THEOREM 4.10. *Consider the system* (3.1). *Assume that* (A, B) *is stabilizable and* (C_1, A) *is detectable. Let* P *and* Q *be the largest real symmetric solution of* (3.2) *and* (3.3), *respectively. Let* $\mathcal{V}_g$ *and* $\mathcal{S}_g$ *be given by* (4.7) *and* (4.8). *Then we have: there exists an optimal controller, i.e., an internally stabilizing controller* $\Gamma^*_{\mathrm{dis}} = (K^*, L^*, M^*, N^*)$ *such that* $J_{\Sigma_{\mathrm{dis}}}(\Gamma^*_{\mathrm{dis}}) = J^*$, *if and only if the four subspace inclusions* $\mathcal{S}_g \subset \mathcal{V}_g$, (4.9), (4.10), *and* (4.11) *are satisfied.*

5. The sampled-data H_2 problem. Now we return to the sampled-data H_2 problem. Consider the continuous-time system Σ given by (2.3), and let $\Delta \notin \boldsymbol{\Delta}$ be a given sampling period. Let the discrete-time system Σ_Δ be given by (2.4). According to Theorem 2.1, the optimal sampled-data H_2 performance $J^*_{\Sigma,\Delta}$ is equal to

$$J^*_{\Sigma,\Delta} = \frac{1}{\Delta}\int_0^\Delta \int_0^{\Delta - s} \mathrm{tr}\, \left(C_2 e^{tA} E E^{\mathrm{T}} e^{tA^{\mathrm{T}}} C_2^{\mathrm{T}}\right) dt\, ds + \frac{1}{\Delta} J^*_{\Sigma_\Delta}, \tag{5.1}$$

where $J^*_{\Sigma_\Delta}$ is the optimal discrete-time H_2 performance associated with Σ_Δ. According to Theorem 4.8, the optimal performance $J^*_{\Sigma_\Delta}$ can be found in terms of two algebraic Riccati equations associated with Σ_Δ. According to Theorem 4.10, an optimal compensator $\Gamma_{\mathrm{dis},\Delta}$ exists if and only if four subspace inclusions involving subspaces associated with the system Σ_Δ are satisfied. According to Theorem 3.3, if the systems $(A_\Delta, B_\Delta, C_{2,\Delta}, D_{2,\Delta})$ and $(A_\Delta, E_\Delta, C_1, 0)$ have no zeros on the unit circle, then an optimal compensator $\Gamma_{\mathrm{dis},\Delta}$ exists and can be calculated using the "construction" in the statement of Theorem 3.3. The sampled-data controller $\Gamma := H_\Delta \Gamma_{\mathrm{dis},\Delta} S_\Delta$ is then optimal for the sampled-data H_2 problem under consideration.

In this section we study the following question: what are conditions *in terms of the original system* Σ that guarantee that there exists an optimal compensator for the sampled-data H_2 problem? Instead of being completely general, we will study the following question: what are necessary and sufficient conditions in terms of the original system Σ such that $(A_\Delta, B_\Delta, C_{2,\Delta}, D_{2,\Delta})$ and $(A_\Delta, E_\Delta, C_1, 0)$ have no zeros on the unit circle? In the following, let $\mathcal{R}$ be the controllability subspace of the system (A, B, C_2, D_2) (see §2). The main results of this section are the following:

THEOREM 5.1. *Consider the system* Σ. *Let* $\Delta > 0$.

(i) *Let* λ *be a zero of* $(A_\Delta, B_\Delta, C_{2,\Delta}, D_{2,\Delta})$, $\lambda \neq 1$. *Then there exists a unobservable eigenvalue* μ *of* (C_2, A) *such that* $\lambda = e^{\mu\Delta}$.

(ii) *If* (A, B, C_2, D_2) *is left-invertible, then also the converse of* (i) *holds: if* μ *is an unobservable eigenvalue of* (C_2, A), *then* $e^{\mu\Delta}$ *is a zero of* $(A_\Delta, B_\Delta, C_{2,\Delta}, D_{2,\Delta})$.

(iii) 1 *is a zero of* $(A_\Delta, B_\Delta, C_{2,\Delta}, D_{2,\Delta})$ *if and only if at least one of the following two conditions hold:*

(a) *0 is a zero of* (A, B, C_2, D_2),

(b)

$$\mathcal{R} \not\subset \langle \ker C_2 \mid A \rangle. \tag{5.2}$$

(iv) *If* (A, B, C_2, D_2) *is left-invertible, then* 1 *is a zero of* $(A_\Delta, B_\Delta, C_{2,\Delta}, D_{2,\Delta})$ *if and only if* 0 *is a zero of* (A, B, C_2, D_2).

COROLLARY 5.2. *Consider the system* Σ. *Let* $\Delta > 0$.

(i) *If* (C_2, A) *has no unobservable eigenvalues on the imaginary axis,* 0 *is not a zero of* (A, B, C_2, D_2), *and* $\mathcal{R} \subset \langle \ker C_2 \mid A \rangle$, *then* $(A_\Delta, B_\Delta, C_{2,\Delta}, D_{2,\Delta})$ *has no zeros on the unit circle.*

(ii) *If* (A, B, C_2, D_2) *is left-invertible, then* $(A_\Delta, B_\Delta, C_{2,\Delta}, D_{2,\Delta})$ *has no zeros on the unit circle if and only if* (C_2, A) *has no unobservable eigenvalues on the imaginary axis and* 0 *is not a zero of* (A, B, C_2, D_2).

THEOREM 5.3. *Consider the system* Σ. *Let* $\Delta > 0$.

(i) *Let* λ *be a zero of* $(A_\Delta, E_\Delta, C_1, 0)$. *Then there exists an uncontrollable eigenvalue* μ *of* (A, E) *such that* $\lambda = e^{\mu\Delta}$.

(ii) *If* $(A, E, C_1, 0)$ *is right-invertible, then also the converse of* (i) *holds*; *i.e., if* μ *is an uncontrollable eigenvalue of* (A, E), *then* $e^{\mu\Delta}$ *is a zero of* $(A_\Delta, E_\Delta, C_1, 0)$.

COROLLARY 5.4. *Consider the system* Σ. *Let* $\Delta > 0$. *If* (A, E) *has no uncontrollable eigenvalues on the imaginary axis, then* $(A_\Delta, E_\Delta, C_1, 0)$ *has no zeros on the unit circle. If, in addition,* $(A, E, C_1, 0)$ *is right-invertible, then also the converse holds:* $(A_\Delta, E_\Delta, C_1, 0)$ *has no zeros on the unit circle if and only if* (A, E) *has no uncontrollable eigenvalues on the imaginary axis.*

Note that the conditions on Σ obtained in these theorems are independent of the sampling period. In the remainder of this section we shall prove these results.

In order to study the zeros of (A, B, C_2, D_2) and $(A_\Delta, B_\Delta, C_{2,\Delta}, D_{2,\Delta})$, consider the system matrices of these systems. Let

$$P_\Delta(z) := \begin{pmatrix} zI - A_\Delta & -B_\Delta \\ C_{2,\Delta} & D_{2,\Delta} \end{pmatrix}, \quad P(s) := \begin{pmatrix} sI - A & -B \\ C_2 & D_2 \end{pmatrix}.$$

Recall that λ is a zero of $(A_\Delta, B_\Delta, C_{2,\Delta}, D_{2,\Delta})$ if and only if the rank of the complex matrix $P_\Delta(\lambda)$ is less than the normal rank of P_Δ (see §2). In order to find out in which points λ this happens, we will study for $\lambda \in \mathcal{C}$ the subspace

$$\mathcal{V}_\lambda := \ker P_\Delta(\lambda) \subset \mathcal{C}^{n+m}.$$

Clearly, for all λ we have $\dim \mathcal{V}_\lambda = n + m - \text{rank } P_\Delta(\lambda)$. Consequently, for all but finitely many λ we have $\dim \mathcal{V}_\lambda = d$, where

$$d := n + m - \text{normrank } P_\Delta.$$

Hence, λ is a zero of $(A_\Delta, B_\Delta, C_{2,\Delta}, D_{2,\Delta})$ if and only if $\dim \mathcal{V}_\lambda > d$. In the following lemma we will calculate for each λ the subspace $\mathcal{V}_\lambda$, its dimension $\dim \mathcal{V}_\lambda$, and the number d. Denote the unobservable subspace $\langle \ker C_2 \mid A \rangle$ by $\mathcal{N}$. Define a subspace $\mathcal{W}$ as follows:

$$\mathcal{W} := B^{-1}\mathcal{N} \cap \ker D_2. \tag{5.3}$$

LEMMA 5.5. *For every $\lambda \in \mathcal{C}$, $\lambda \neq 1$ we have*

$$\mathcal{V}_\lambda = (\mathcal{N} \times \mathcal{W}) \cap \ker \begin{pmatrix} \lambda I - A_\Delta & B_\Delta \end{pmatrix}, \tag{5.4}$$

$$\dim \mathcal{V}_\lambda = \dim \mathcal{N} + \dim \mathcal{W} - \dim((\lambda I - A_\Delta)\mathcal{N} + B_\Delta \mathcal{W}). \tag{5.5}$$

For all but finitely many λ we have $\dim \mathcal{V}_\lambda = d = \dim \mathcal{W}$, *equivalently,* $\text{normrank } P_\Delta = n + m - \dim \mathcal{W}$. *In addition we have*

$$\mathcal{V}_1 = \ker \begin{pmatrix} -A & -B \\ C_2 & D_2 \end{pmatrix}. \tag{5.6}$$

Proof. We will first prove (5.4). We know $\binom{x_0}{u_0} \in \mathcal{V}_\lambda$ if and only if

$$A_\Delta x_0 + B_\Delta u_0 = \lambda x_0, \tag{5.7}$$

$$C_{2,\Delta} x_0 + D_{2,\Delta} u_0 = 0. \tag{5.8}$$

Consider the differential equation $\dot{x}(t) = Ax(t) + Bu_0$, $x(0) = x_0$; and define $z(t) := C_2 x(t) + Du_0$. Clearly, $x(\Delta) = A_\Delta x_0 + B_\Delta u_0$, so (5.7) is equivalent to $x(\Delta) = \lambda x_0$. In turn, this is equivalent to

$$(\lambda - 1)x_0 = \int_0^\Delta e^{At}(Ax_0 + Bu_0)dt. \tag{5.9}$$

Using the definition (2.6) of $C_{2,\Delta}$ and $D_{2,\Delta}$, we see that (5.8) is equivalent to

$$(C_2 \;\; D_2)e^{\underline{A}t}\begin{pmatrix} x_0 \\ u_0 \end{pmatrix} = 0 \;\text{ for all } t \in [0, \Delta],$$

which, in turn, is equivalent to $z(t) = 0$ for all $t \in [0, \Delta]$. Obviously,

$$z(t) = C_2 e^{At} x_0 + \left[C_2 \int_0^t e^{As} B ds + D_2 \right] u_0.$$

Since $z(t) = 0$ for all $t \in [0, \Delta]$ is satisfied if and only if $z(0) = 0$ and $\dot{z}(t) = 0$ for all $t \in [0, \Delta]$, we find that (5.8) is equivalent to

$$C_2 x_0 + D_2 u_0 = 0 \quad \text{and} \quad C_2 e^{At}(Ax_0 + Bu_0) = 0, \qquad t \in [0, \Delta].$$

In other words (5.8) is satisfied if and only if

$$(5.10) \qquad C_2x_0 + D_2u_0 = 0 \quad \text{and} \quad Ax_0 + Bu_0 \in \mathcal{N}.$$

Now assume that $\lambda \neq 1$. Then (5.9) and (5.10) imply that $x_0 \in \mathcal{N} \subset \ker C_2$, so $u_0 \in \ker D_2$. Also it follows that $Ax_0 \in \mathcal{N}$, so $Bu_0 \in \mathcal{N}$ and, in fact, $u_0 \in \mathcal{W}$. We conclude that, for $\lambda \neq 1$, $\mathcal{V}_\lambda \subset (\mathcal{N} \times \mathcal{W}) \cap \ker \begin{pmatrix} \lambda I - A_\Delta & B_\Delta \end{pmatrix}$. To prove the converse inclusion, note that $u_0 \in \mathcal{W}$ implies that $D_2u_0 = 0$ and $Bu_0 \in \mathcal{N}$. If, in addition, $x_0 \in \mathcal{N}$, then we have $C_2x_0 + D_2u_0 = 0$ and $Ax_0 + Bu_0 \in \mathcal{N}$. By the above this is equivalent to (5.8). This completes the proof of (5.4).

To prove (5.5), note that, in general, if $\mathcal{L}$ is a subspace of some finite-dimensional linear space $\mathcal{X}$ and if T is a linear map acting on $\mathcal{X}$, then we have $\dim(\mathcal{L} \cap \ker T) = \dim \mathcal{L} - \dim T\mathcal{L}$. Applying this to the situation at hand, we find that for any $\lambda \neq 1$ we have

$$\dim \mathcal{V}_\lambda = \dim(\mathcal{N} \times \mathcal{W}) - \dim(\lambda I - A_\Delta \;\; B_\Delta)(\mathcal{N} \times \mathcal{W}),$$

which immediately yields (5.5).

Next, we will prove the statement on the dimension of $\mathcal{V}_\lambda$. First note that since $\mathcal{N}$ is A-invariant, it is also e^{At}-invariant, for any t. In particular, this implies that $\mathcal{N}$ is A_Δ-invariant and invariant under $\int_0^\Delta e^{At}dt$. Now assume that $\lambda \notin \sigma(A_\Delta)$. Then we have $(\lambda I - A_\Delta)\mathcal{N} = \mathcal{N}$. Also, since $B\mathcal{W} \subset \mathcal{N}$, we have $B_\Delta \mathcal{W} \subset \mathcal{N}$. This implies that $(\lambda I - A_\Delta)\mathcal{N} + B_\Delta\mathcal{W} = \mathcal{N}$. If, in addition, we assume that $\lambda \neq 1$, then (5.5) yields $\dim \mathcal{V}_\lambda = \dim \mathcal{W}$.

Finally, to prove (5.6), recall that (5.7) is equivalent to (5.9). Note that for all $\Delta > 0$, $\int_0^\Delta e^{At}dt$ is a nonsingular matrix (this can be shown using the Jordan form of A). Thus, for the case that $\lambda = 1$ (5.9) is equivalent to $Ax_0 + Bu_0 = 0$. Together with the fact that (5.8) is equivalent to (5.10), this proves (5.6). □

By applying this lemma, we are now able to prove the statements (i) and (ii) of Theorem 5.1:

Proof of Theorem 5.1 (i) *and* (ii). (i) Assume that $\lambda \neq 1$ is a zero of $(A_\Delta, B_\Delta, C_{2,\Delta}, D_{2,\Delta})$. Then we must have $\dim \mathcal{V}_\lambda > \dim \mathcal{W}$. Using (5.5) this implies

$$(5.11) \qquad \dim \mathcal{N} > \dim((\lambda I - A_\Delta)\mathcal{N} + B_\Delta \mathcal{W}).$$

As noted in the proof of Lemma 5.5, $\mathcal{N}$ is A_Δ-invariant and $B_\Delta\mathcal{W} \subset \mathcal{N}$. Consequently,

$$(\lambda I - A_\Delta)\mathcal{N} + B_\Delta\mathcal{W} \subset \mathcal{N}.$$

Together with the inequality (5.11), this implies that $(\lambda I - A_\Delta)\mathcal{N}$ is a *strict* subspace of $\mathcal{N}$. This implies that the map $(\lambda I - A_\Delta)$ restricted to $\mathcal{N}$ is singular. Thus, $\ker(\lambda I - A_\Delta) \cap \mathcal{N} \neq 0$. Clearly, this intersection is A-invariant, so the restriction of A to this intersection has an eigenvalue, say μ, with corresponding eigenvector p. This eigenvector satisfies $A_\Delta p = \lambda p$. Also, since $Ap = \mu p$, we have $A_\Delta p = e^\mu p$, so $\lambda = e^\mu$. Finally, $p \in \mathcal{N} \subset \ker C_2$, so μ is an unobservable eigenvalue of (C_2, A).

(ii) We claim that if (A, B, C_2, D_2) is left-invertible, then $\dim \mathcal{W} = 0$. Indeed, left-invertibility is equivalent to the conditions $\begin{pmatrix} B \\ D_2 \end{pmatrix}$ is injective and $\mathcal{V} \cap B \ker D_2 = 0$, where $\mathcal{V}$ denotes the weakly unobservable subspace associated with (A, B, C_2, D_2) (see §2). Assume that $u_0 \in \mathcal{W}$. Then we have $D_2u_0 = 0$ and $Bu_0 \in \mathcal{N}$. Since $\mathcal{N} \subset \mathcal{V}$, this yields $Bu_0 = 0$. Combining this with $D_2u_0 = 0$ then leads to $u_0 = 0$. This proves our claim. Now let μ be a unobservable eigenvalue of (C_2, A). There exists $x_0 \neq 0$ such that $Ax_0 = \mu x_0$ and $C_2 x_0 = 0$. This yields $A_\Delta x_0 = \lambda x_0$, with $\lambda := e^{\mu\Delta}$. From the definition of $C_{2,\Delta}$ it is also easily seen that $C_{2,\Delta} x_0 = 0$. Consequently, $\binom{x_0}{0} \in \mathcal{V}_\lambda$, so $\dim \mathcal{V}_\lambda > 0 = \dim \mathcal{W}$. This implies that λ is a zero of $(A_\Delta, B_\Delta, C_{2,\Delta}, D_{2,\Delta})$. □

In order to prove statements (iii) and (iv) of Theorem 5.1, we need the following lemma.

LEMMA 5.6. *Let $\Delta > 0$. Then we have*

$$\text{normrank } P_\Delta \geq \text{normrank } P, \tag{5.12}$$

with equality if and only if $\mathcal{R} \subset \mathcal{N}$.

Proof. For each $\lambda \notin \sigma(A)$ define a subspace $\mathcal{L}_\lambda$ by

$$\mathcal{L}_\lambda := \left\{ \begin{pmatrix} x_0 \\ u_0 \end{pmatrix} \mid u_0 \in \mathcal{W}, x_0 = (\lambda I - A)^{-1} B u_0 \right\}.$$

Clearly, $\mathcal{L}_\lambda \subset \ker P(\lambda)$ and $\dim \mathcal{L}_\lambda = \dim \mathcal{W}$. Consequently, for each $\lambda \notin \sigma(A)$ we have $\dim \mathcal{W} \leq \dim \ker P(\lambda)$. This implies normrank $P \leq n + m - \dim \mathcal{W}$. The inequality (5.12) then follows from Lemma 5.5.

Of course, normrank $P_\Delta =$ normrank P if and only if $\dim \ker P(\lambda) = \dim \mathcal{W}$ for all but finitely many λ, which, in turn, is equivalent to $\ker P(\lambda) = \mathcal{L}_\lambda$ for all but finitely many λ, $\lambda \notin \sigma(A)$. We will prove that the latter statement is equivalent to $\mathcal{R} \subset \mathcal{N}$.

Let $k := \dim \mathcal{R}$, and let $\lambda_1, \ldots, \lambda_k$ be distinct complex numbers, $\lambda_i \notin \sigma(A)$, such that $\ker P(\lambda_i) = \mathcal{L}_{\lambda_i}$. There exists $F \in \mathbb{R}^{m \times n}$ such that $(A + BF)\mathcal{R} \subset \mathcal{R}$, $(C_2 + D_2 F)\mathcal{R} = 0$, and $\sigma(A + BF \mid \mathcal{R}) = \{\lambda_1, \ldots, \lambda_k\}$. Let $x_1, \ldots, x_k \in \mathcal{R}$ be corresponding eigenvectors of $A + BF \mid \mathcal{R}$. Then $\{x_1, \ldots, x_k\}$ is a basis of $\mathcal{R}$. We will prove that $x_i \in \mathcal{N}$. Indeed, define $u_i := -Fx_i$. Then $\binom{x_i}{u_i} \in \ker P(\lambda_i) = \mathcal{L}_{\lambda_i}$. Since $u_i \in \mathcal{W}$, we have $Bu_i \in \mathcal{N}$, so $x_i = (\lambda_i I - A)^{-1} B u_i \in \mathcal{N}$ by A-invariance of $\mathcal{N}$. We conclude that $x_i \in \mathcal{N}$, so $\mathcal{R} \subset \mathcal{N}$.

Conversely, assume that $\mathcal{R} \subset \mathcal{N}$. It suffices to show that $\ker P(\lambda) \subset \mathcal{L}_\lambda$ for all but finitely many λ. Let λ be arbitrary, $\lambda \notin \sigma(A)$, and λ not a zero of (A, B, C_2, D_2). Let $\binom{x_0}{u_0} \in \ker P(\lambda)$. We will prove that $x_0 \in \mathcal{R}$, so $x_0 \in \mathcal{N}$. Assume that $x_0 \neq 0$. Let $F \in \mathbb{R}^{m \times n}$ be such that $Fx_0 = u_0$. Then we have $(A + BF)x_0 = \lambda x_0$ and $(C_2 + D_2 F)x_0 = 0$. This implies $x_0 \in \mathcal{V}$, the weakly unobservable subspace associated with the system (A, B, C_2, D_2). (Indeed, the one-dimensional subspace $\mathcal{L}$ spanned by the vector x_0 has the property that $(A + BF)\mathcal{L} \subset \mathcal{L}$ and $(C_2 + D_2 F)\mathcal{L} = 0$ and so must be contained in $\mathcal{V}$, the largest subspace for which such F exists.) By extending the linear map F to the whole subspace $\mathcal{V}$, we obtain that $(A + BF)\mathcal{V} \subset \mathcal{V}$ and $(C_2 + D_2 F)\mathcal{V} = 0$, so $\lambda \in \sigma(A + BF \mid \mathcal{V})$. We have assumed that λ is not a zero. This implies $\lambda \notin \sigma(A + BF \mid \mathcal{V}/\ R)$ (the latter spectrum is equal to the set of zeros of (A, B, C_2, D_2); see [19]). But then we must have $x_0 \in \mathcal{R}$. This implies that $x_0 \in \mathcal{N}$. Now $(\lambda I - A)x_0 - Bu_0 = 0$, so $Bu_0 \in \mathcal{N}$. This implies that $u_0 \in \mathcal{W}$. For $\lambda \notin \sigma(A)$ this then yields $x_0 \in \mathcal{L}_\lambda$. This completes the proof of the lemma. □

Proof of Theorem 5.1 (iii) *and* (iv). (iii) We will prove that 1 is *not* a zero of the system $(A_\Delta, B_\Delta, C_{2,\Delta}, D_{2,\Delta})$ if and only if 0 is not a zero of (A, B, C_2, D_2) ánd normrank $P =$ normrank P_Δ. Clearly, 1 is not a zero of $(A_\Delta, B_\Delta, C_{2,\Delta}, D_{2,\Delta})$ if and only if $\dim \mathcal{V}_1 = n + m -$normrank P_Δ. By (5.6) we have $\dim \mathcal{V}_1 = n + m -$rank $P(0) \geq n + m -$ normrank P, with strict inequality if and only if 0 is a zero of (A, B, C_2, D_2). Combining these facts proves our claim. The proof of (iii) is then completed by applying Lemma 5.6.

(iv) If (A, B, C_2, D_2) is left-invertible, then $\mathcal{R} = 0$. In that case condition (5.2) is never satisfied. □

In order to study the zeros of $(A_\Delta, E_\Delta, C_1, 0)$, consider the system matrix of this system. Let

$$Q_\Delta(z) := \begin{pmatrix} zI - A_\Delta & -E_\Delta \\ C_1 & 0 \end{pmatrix}.$$

As before, λ is a zero of $(A_\Delta, E_\Delta, C_1, 0)$ if and only if the rank of the complex matrix $Q_\Delta(\lambda)$ is less than the normal rank of Q_Δ (see §2). In order to find out in which points λ this happens, we will study for $\lambda \in \mathcal{C}$ the subspace

$$\mathcal{W}_\lambda := (\text{im } Q_\Delta(\lambda))^\perp \subset \mathcal{C}^{n+p}.$$

For all λ we have $\dim \mathcal{W}_\lambda = n + p - \text{rank } Q_\Delta(\lambda)$. Consequently, for all but finitely many λ we have $\dim \mathcal{W}_\lambda = d_1$, where

$$d_1 := n + p - \text{normrank } Q_\Delta.$$

Hence, λ is a zero of $(A_\Delta, E_\Delta, C_1, 0)$ if and only if $\dim \mathcal{W}_\lambda > d_1$. The following lemma calculates for each λ the subspace $\mathcal{W}_\lambda$, its dimension $\dim \mathcal{W}_\lambda$, and the number d_1. Let $\mathcal{M} := \langle A \mid \text{im } E\rangle$, the reachable subspace of (A, E).

LEMMA 5.7. *Let $\Delta > 0$. Then we have*

$$\mathcal{W}_\lambda = \left(M^\perp \times (C_1^{\mathrm{T}})^{-1} M^\perp\right) \cap \ker \begin{pmatrix} \lambda I - A_\Delta^{\mathrm{T}} & C_1^{\mathrm{T}} \end{pmatrix},$$

$$\begin{aligned} \dim \mathcal{W}_\lambda = {} & \dim M^\perp + \dim (C_1^{\mathrm{T}})^{-1} M^\perp \\ & - \dim((\lambda I - A_\Delta^{\mathrm{T}}) M^\perp + C_1^{\mathrm{T}} (C_1^{\mathrm{T}})^{-1} M^\perp). \end{aligned} \tag{5.13}$$

For all but finitely many λ we have $\dim \mathcal{W}_\lambda = d_1 = \dim (C_1^{\mathrm{T}})^{-1} M^\perp$, *equivalently,*

$$\text{normrank } Q_\Delta = n + p - \dim (C_1^{\mathrm{T}})^{-1} M^\perp.$$

Proof. By definition, $\binom{x_0}{y_0} \in \mathcal{W}_\lambda$ if and only if

$$(\lambda I - A_\Delta^{\mathrm{T}}) x_0 + C_1^{\mathrm{T}} y_0 = 0 \quad \text{and} \quad x_0^{\mathrm{T}} E_\Delta = 0. \tag{5.14}$$

Since, by definition, im $E_\Delta = \mathcal{M}$, we see that it suffices to show that (5.14) implies $y_0 \in (C_1^{\mathrm{T}})^{-1} M^\perp$. From the fact that $\mathcal{M}^\perp$ is A^{T}-invariant it follows that $\mathcal{M}^\perp$ is A_Δ^{T}-invariant, so $C_1^{\mathrm{T}} y_0 \in \mathcal{M}^\perp$. The statement (5.13) on the dimension of $\mathcal{W}_\lambda$ follows in the same way as the corresponding statement in the previous lemma.

Now let λ be any complex number such that $\lambda \notin \sigma(A_\Delta^{\mathrm{T}})$. Since $\mathcal{M}^\perp$ is A_Δ^{T}-invariant, we then have $(\lambda I - A_\Delta^{\mathrm{T}}) M^\perp = M^\perp$. Also we have $C_1^{\mathrm{T}} (C_1^{\mathrm{T}})^{-1} M^\perp \subset M^\perp$ (no equality!). Thus, for such λ we have $\dim \mathcal{W}_\lambda = \dim (C_1^{\mathrm{T}})^{-1} M^\perp$. □

We are now ready to prove Theorem 5.3.

Proof of Theorem 5.3. Let λ be a zero of $(A_\Delta, E_\Delta, C_1, 0)$. Then we have $\dim \mathcal{W}_\lambda > \dim (C_1^{\mathrm{T}})^{-1} M^\perp$. Consequently, by (5.13), $\dim M^\perp > \dim((\lambda I - A_\Delta^{\mathrm{T}}) M^\perp + C_1^{\mathrm{T}} (C_1^{\mathrm{T}})^{-1} M^\perp)$. In particular, this implies that $(\lambda I - A_\Delta^{\mathrm{T}}) \mathcal{M}^\perp$ is a *strict* subspace of $\mathcal{M}^\perp$, so $\ker(\lambda I - A_\Delta^{\mathrm{T}}) \cap \mathcal{M}^\perp \neq 0$. This subspace is A^{T}-invariant, so there exist μ and

$x_0 \in \mathcal{M}^\perp$, $x_0 \neq 0$, such that $A^{\mathrm{T}} x_0 = \mu x_0$, $A_\Delta^{\mathrm{T}} x_0 = \lambda x_0$, and $x_0 \in \mathcal{M}^\perp$. Obviously, this implies $\lambda = e^\mu \Delta$, and μ is an uncontrollable eigenvalue of (A, E).

Assume that $(A, E, C_1, 0)$ is right-invertible. Let

$$Q(s) := \begin{pmatrix} sI - A & -E \\ C_1 & 0 \end{pmatrix}$$

be the system matrix. We have normrank $Q = n+p$. We claim that also normrank $Q_\Delta = n+p$. Indeed, assume that $y_0 \neq 0$ is an element of $(C_1^{\mathrm{T}})^{-1} M^\perp$. For $\lambda \notin \sigma(A^{\mathrm{T}})$, define $x_0 := -(\lambda I - A^{\mathrm{T}})^{-1} C_1^{\mathrm{T}} y_0$. Then $x_0 \in \mathcal{M}^\perp$ and we have $(x_0^{\mathrm{T}} \;\; y_0^{\mathrm{T}}) Q(\lambda) = (0 \;\; 0)$. Thus, for all but finitely many λ we have rank $Q(\lambda) < n+p$, which is a contradiction. Hence we must have $(C_1^{\mathrm{T}})^{-1} M^\perp = 0$.

It follows that λ is a zero if and only if $\mathcal{W}_\lambda \neq 0$. Assume that μ is an uncontrollable eigenvalue of (A, E). Then there exists $x_0 \neq 0$, $x_0 \in \mathcal{M}^\perp$, such that $x_0^{\mathrm{T}} A^{\mathrm{T}} = \mu x_0$. Define $\lambda := e^{\mu\Delta}$. Then we have $x_0^{\mathrm{T}} E_\Delta^{\mathrm{T}} = 0$ and $x_0^{\mathrm{T}}(\lambda I - A_\Delta) = 0$. It follows that $\binom{x_0}{0} \in \mathcal{W}_\lambda$, so λ is a zero of $(A_\Delta, E_\Delta, C_1, 0)$. $\square$

6. Performance recovery and convergence of optimal performance. In this section we study the connection between the 'ordinary' continuous-time H_2 problem and the sampled-data H_2 problem. In particular, we are interested in the following questions:

- Suppose that we control the system Σ by means of an internally stabilizing continuous-time compensator Γ_{con}, yielding continuous-time H_2 performance $J_\Sigma(\Gamma_{\text{con}})$. Is it possible to recover this performance asymptotically by using a sampled-data controller with sufficiently small sampling period? More precisely, is it true that for all $\epsilon > 0$ there exists $\Delta > 0$ and an internally stabilizing sampled-data controller Γ with sampling-period Δ such that $|J_\Sigma(\Gamma_{\text{con}}) - J_{\Sigma,\Delta}(\Gamma)| < \epsilon$?
- Does the optimal sampled-data H_2 performance converge to the optimal continuous-time H_2 performance as the sampling period Δ decreases to zero? More precisely, suppose that $J^*_{\Sigma,\text{con}}$ is the optimal continuous-time H_2 performance associated with the system Σ and, as before, denote the optimal sampled-data H_2 performance by $J^*_{\Sigma,\Delta}$. Is it true that $\lim_{\Delta \downarrow 0} J^*_{\Sigma,\Delta} = J^*_{\Sigma,\text{con}}$?

The first question above was studied before in [6, Thm. 4] using a different definition of H_2 performance and for the H_∞ performance criterion [6, Thm. 5]. In this section we will show that both questions have an affirmative answer.

Let Σ be given by (2.2). If the system Σ is controlled by a continuous-time compensator Γ_{con} given by the equations

$$\begin{aligned} \dot{w}(t) &= \bar{K} w(t) + \bar{L} y(t) , \\ u(t) &= \bar{M} w(t) + \bar{N} y(t) , \end{aligned} \tag{6.1}$$

with $w(t) \in \mathbb{R}^\ell$, then the associated closed-loop system $\Sigma \times \Gamma_{\text{con}}$ is given by

$$\begin{aligned} \dot{x}_e(t) &= A_e x_e(t) + E_e y(t) , \\ z(t) &= C_e x_e(t) , \end{aligned}$$

with

$$A_e = \begin{pmatrix} A + B\bar{N}C_1 & B\bar{M} \\ \bar{L}C_1 & \bar{K} \end{pmatrix}, \; E_e := \begin{pmatrix} E \\ 0 \end{pmatrix}, \; C_e := \begin{pmatrix} C_2 + D_2\bar{N}C_1 & D_2\bar{M} \end{pmatrix}.$$

If Γ_{con} is internally stabilizing, i.e., $\sigma(A_e) \subset \mathcal{C}^-$, then the H_2 performance of the closed-loop system $\Sigma \times \Gamma_{\text{con}}$ is equal to

$$J_\Sigma(\Gamma_{\text{con}}) = \text{tr }(E_e P_e E_e^{\mathrm{T}}),$$

where P_e is the unique solution of the Lyapunov equation

$$A_e^{\mathrm{T}} P_e + P_e A_e + C_e^{\mathrm{T}} C_e = 0. \tag{6.2}$$

On the other hand, if the system Σ is controlled by the sampled-data controller $\Gamma = H_\Delta \Gamma_{\text{dis}} S_\Delta$, with Γ_{dis} given by (2.8), then the discrete-time closed-loop system $\Sigma_\Delta \times \Gamma_{\text{dis}}$ is given by the equations

$$\begin{aligned} x_{e,k+1} &= A_{e,\Delta} x_{e,k}\ E_{e,\Delta} y_k\ , \\ z_k &= C_{e,\Delta} x_{e,k}\ , \end{aligned}$$

with

$$A_{e,\Delta} = \begin{pmatrix} A_\Delta + B_\Delta N C_1 & B_\Delta M \\ L C_1 & K \end{pmatrix},\ E_{e,\Delta} := \begin{pmatrix} E_\Delta \\ 0 \end{pmatrix},$$

$$C_{e,\Delta} := \begin{pmatrix} C_{2,\Delta} + D_{2,\Delta} N C_1 & D_{2,\Delta} M \end{pmatrix}.$$

If Γ is internally stabilizing, equivalently $|\sigma(A_{e,\Delta})| < 1$, then the H_2 performance of the closed-loop system $\Sigma \times \Gamma$ is given by

$$J_{\Sigma,\Delta}(\Gamma) = \frac{1}{\Delta} \int_0^\Delta \int_0^{\Delta - s} \text{tr } \left(C_2 e^{tA} E E^{\mathrm{T}} e^{tA^{\mathrm{T}}} C_2^{\mathrm{T}} \right) dt\, ds + \frac{1}{\Delta} \text{tr }(E_{e,\Delta} P_{e,\Delta} E_{e,\Delta}^{\mathrm{T}}), \tag{6.3}$$

where $P_{e,\Delta}$ is the unique solution of the Lyapunov equation

$$A_{e,\Delta}^{\mathrm{T}} P_{e,\Delta} A_{e,\Delta} - P_{e,\Delta} + C_{e,\Delta}^{\mathrm{T}} C_{e,\Delta} = 0. \tag{6.4}$$

The following theorem shows that our first question above indeed has an affirmative answer:

THEOREM 6.1. *Let Γ_{con} be an internally stabilizing continuous-time compensator. For any $\Delta > 0$ define a discrete-time controller Γ_{dis} by $\Gamma_{\text{dis}} := S_\Delta \Gamma_{\text{con}} H_\Delta$, and let $\Gamma_\Delta := H_\Delta \Gamma_{\text{dis}} S_\Delta$ be the corresponding sampled-data controller with sampling period Δ. Then we have that there exists $\Delta_1 > 0$ such that for all $\Delta \notin \mathbf{\Delta}$ with $0 < \Delta < \Delta_1$, Γ_Δ is internally stabilizing. Furthermore,*

$$J_{\Sigma,\Delta}(\Gamma_\Delta) \to J_\Sigma(\Gamma_{\text{con}}) \qquad (\Delta \downarrow 0).$$

Proof. It is easily verified that $\Gamma_{\text{dis}} := S_\Delta \Gamma_{\text{con}} H_\Delta$ is described by the equations

$$\begin{aligned} w_{k+1} &= K_\Delta w_k + L_\Delta y_k\ , \\ u_k &= M w_k + N y_k\ , \end{aligned}$$

with $K_\Delta := e^{\bar{K}\Delta}$ and $L_\Delta := \int_0^\Delta e^{\bar{K}t} dt \bar{L}$. Thus we have

$$A_{e,\Delta} = \begin{pmatrix} A_\Delta + B_\Delta N C_1 & B_\Delta M \\ L_\Delta C_1 & K_\Delta \end{pmatrix}.$$

Note that $A_{e,\Delta} \to I$, the $(n+\ell)\times(n+\ell)$ identity matrix, and that $\frac{1}{\Delta}(A_{e,\Delta} - I) \to A_e$ ($\Delta \downarrow 0$). Now we will first show that for Δ sufficiently small we have $|\sigma(A_{e,\Delta})| < 1$. Since A_e is stable, there exists $Q > 0$ such that $A_e^{\mathrm{T}}Q + QA_e < 0$. Now note that

$$\frac{1}{\Delta}(A_{e,\Delta}^{\mathrm{T}} Q A_{e,\Delta} - Q) = \frac{1}{\Delta}(A_{e,\Delta}^{\mathrm{T}} - I)QA_{e,\Delta} + Q\frac{1}{\Delta}(A_{e,\Delta} - I).$$

Since the right-hand term converges to $A_e^{\mathrm{T}}Q + QA_e < 0$, for Δ sufficiently small we have $A_{e,\Delta}^{\mathrm{T}}QA_{e,\Delta} - Q < 0$. This implies that for Δ sufficiently small $A_{e,\Delta}$ is stable.

Next we show the convergence of the H_2 performance. For Δ sufficiently small we have $|\sigma(A_{e,\Delta})| < 1$, so the H_2 performance is given by (6.3), with $P_{e,\Delta}$ given by the Lyapunov equation (6.4). We shall prove that $P_{e,\Delta} \to P_e$, the unique solution of (6.2). For any Δ sufficiently small define a linear map $m_\Delta : \mathbb{R}^{n\times n} \to \mathbb{R}^{n\times n}$ by

$$m_\Delta(X) := \frac{1}{\Delta}A_{e,\Delta}^{\mathrm{T}} X A_{e,\Delta} - \frac{1}{\Delta}X.$$

Also define a linear map $m : \mathbb{R}^{n\times n} \to \mathbb{R}^{n\times n}$ by

$$m(X) := A_e^{\mathrm{T}}X + XA_e.$$

Note that m and m_Δ are all bijections. We can rewrite m_Δ as

$$m_\Delta(X) = \frac{1}{\Delta}(A_{e,\Delta}^{\mathrm{T}} - I)XA_{e,\Delta} + X\frac{1}{\Delta}(A_{e,\Delta} - I).$$

Recall that $A_{e,\Delta} \to I$ and $\frac{1}{\Delta}(A_{e,\Delta} - I) \to A_e$. Thus we see that $m_\Delta \to m$ ($\Delta \downarrow 0$). Consequently, also $m_\Delta^{-1} \to m^{-1}$ ($\Delta \downarrow 0$). Obviously, $P_{e,\Delta} = m_\Delta^{-1}(-\frac{1}{\Delta}C_{e,\Delta}^{\mathrm{T}}C_{e,\Delta})$. In addition, it follows from (2.6) that $\frac{1}{\Delta}C_{e,\Delta}^{\mathrm{T}}C_{e,\Delta} \to C_e^{\mathrm{T}}C_e$. This implies that $P_{e,\Delta} \to m^{-1}(C_e^{\mathrm{T}}C_e)$, which, in turn, is equal to P_e. By (2.5) we see that $\frac{1}{\Delta}E_{e,\Delta}E_{e,\Delta}^{\mathrm{T}} \to E_eE_e^{\mathrm{T}}$. Combining these facts we find that

$$\frac{1}{\Delta}\mathrm{tr}\,(E_{e,\Delta}E_{e,\Delta}^{\mathrm{T}}P_{e,\Delta}) \to \mathrm{tr}\,(E_eE_e^{\mathrm{T}}P_e).$$

Finally, it is immediate that

$$\frac{1}{\Delta}\int_0^\Delta \int_0^{\Delta - s} \mathrm{tr}\,\left(C_1 e^{tA}EE^{\mathrm{T}}e^{tA^{\mathrm{T}}}C_1^{\mathrm{T}}\right) dt\,ds \to 0, \qquad \Delta \downarrow 0,$$

which completes the proof of the theorem. $\square$

Now we turn to the second question posed above. In order to be able to answer this question, it is useful to consider this question first for the *linear quadratic problem.*

For this, consider the system $\dot{x}(t) = Ax(t) + Bu(t)$, $z(t) = C_2x(t) + D_2u(t)$. Assume that (A, B) is stabilizable. For a given static state feedback control law $u = Fx$ and initial state x_0, the output function is denoted by z_{F,x_0}. The linear quadratic problem is to minimize for each x_0 the cost-functional $J(x_0, F) := \int_0^\infty \|z_{F,x_0}(t)\|^2 dt$ over all $F \in \mathbb{R}^{m\times n}$ such that $\sigma(A + BF) \subset C^-$. It is well known (see [9], [18]) that for each x_0 the optimal cost

$$J^*(x_0) := \inf\{J(x_0, F) \mid F \text{ s.t. } \sigma(A + BF) \subset C^-\} = x_0^{\mathrm{T}}Px_0,$$

where P is the largest real symmetric solution of the linear matrix inequality

$$\begin{pmatrix} A^{\mathrm{T}}P + PA^{\mathrm{T}} + C_2^{\mathrm{T}}C_2 & PB + C_2^{\mathrm{T}}D_2 \\ B^{\mathrm{T}}P + D_2^{\mathrm{T}}C_2 & D_2^{\mathrm{T}}D_2 \end{pmatrix} \geq 0. \tag{6.5}$$

We want to compare this "normal" linear quadratic problem with its sampled-data version.

In the following, take a fixed sampling period $\Delta > 0$. The sampled-data version of the linear quadratic problem is to do the minimization over all stabilizing sampled-data static state feedback laws. More precisely, for a given $F \in \mathbb{R}^{m\times n}$ define the sampled-data state feedback control law $u = \mathcal{F}_\Delta x$ by $u(t) := Fx(k\Delta)$ ($t \in [k\Delta, (k+1)\Delta)$, $k = 0, 1, 2, \ldots$, or with a slight abuse of notation: $\mathcal{F}_\Delta = H_\Delta F S_\Delta$. For a given $\mathcal{F}_\Delta$ and initial state x_0, denote the output by $z_{\mathcal{F}_\Delta, x_0}$. Define the sampled-data cost functional in the obvious way, and denote it by $J(x_0, \mathcal{F}_\Delta)$. The control law $\mathcal{F}_\Delta$ is called internally stabilizing if for each initial state the controlled state trajectory $x(t)$ converges to 0 as $t \to \infty$. The sampled-data linear quadratic problem is to minimize for each x_0 $J(x_0, \mathcal{F}_\Delta)$ over all internally stabilizing control laws $\mathcal{F}_\Delta$. Let

$$J^*_\Delta(x_0) := \inf\{J(x_0, \mathcal{F}_\Delta) \mid \mathcal{F}_\Delta \text{ is internally stabilizing}\}$$

be the optimal cost. If no internally stabilizing $\mathcal{F}_\Delta$ exists, we define $J^*_\Delta(x_0) := \infty$ for all x_0. We will briefly explain here how the sampled-data linear quadratic can be resolved. First, note that for any $\mathcal{F}_\Delta = H_\Delta F S_\Delta$ we have

$$J(x_0, \mathcal{F}_\Delta) = \sum_{k=0}^{\infty} \int_{k\Delta}^{(k+1)\Delta} \|z_{\mathcal{F}_\Delta, x_0}(t)\|^2 dt.$$

Secondly, note that for all $t \in [k\Delta, (k+1)\Delta)$ we have $\dot{x}(t) = Ax(t) + Bu(t)$, $z_{\mathcal{F}_\Delta, x_0}(t) = C_2 x(t) + D_2 u(t)$, with $u(t) = Fx(k\Delta)$. Hence, on the interval $[k\Delta, (k+1)\Delta)$, x and u satisfy

$$\begin{pmatrix} \dot{x} \\ \dot{u} \end{pmatrix} = \begin{pmatrix} A & B \\ 0 & 0 \end{pmatrix} \begin{pmatrix} x \\ u \end{pmatrix},$$

with $u(k\Delta) = Fx(k\Delta)$. Consequently,

$$\begin{pmatrix} x(t) \\ u(t) \end{pmatrix} = e^{\underline{A}(t-k\Delta)} \begin{pmatrix} x(k\Delta) \\ Fx(k\Delta) \end{pmatrix}$$

for $t \in [k\Delta, (k+1)\Delta)$, with $\underline{A}$ defined by (2.7). Using this, it follows immediately from (2.6) that for $t \in [k\Delta, (k+1)\Delta)$ we have $\|z_{\mathcal{F}_\Delta, x_0}(t)\|^2 = \|C_{2,\Delta} x(k\Delta) + D_{2,\Delta} F x(k\Delta)\|^2$. Obviously, $x(k\Delta))$ evoluates according to $x((k+1)\Delta) = A_\Delta x(k\Delta) + B_\Delta F x(k\Delta)$. Hence we see that if $\mathcal{F}_\Delta = H_\Delta F S_\Delta$, then $J(x_0, \mathcal{F}_\Delta) = \sum_{k=0}^{\infty} \|(C_{2,\Delta} + D_{2,\Delta} F) x_k\|^2$, with $x_{k+1} = (A_\Delta + B_\Delta F) x_k$. It is also easily seen that $\mathcal{F}_\Delta$ is internally stabilizing if and only if $|\sigma(A_\Delta + B_\Delta F)| < 1$. Hence, $J^*_\Delta(x_0) < \infty$ for all x_0 if and only if (A_Δ, B_Δ) is stabilizable.

Consequently, we can make the following conclusion: the sampled-data linear quadratic problem under consideration is equivalent to the "normal" discrete-time linear quadratic problem of minimizing, for the system $x_{k+1} = A_\Delta x_k + B_\Delta u_k$, the cost functional $J_{\mathrm{dis}}(x_0, F) := \sum_{k=0}^{\infty} \|(C_{2,\Delta} x_k + D_{2,\Delta} u_k\|^2$ over all $F \in \mathbb{R}^{m\times n}$ such that $|\sigma(A_\Delta + B_\Delta F)| < 1$. The latter problem was discussed in §3, remark (3.11) and §4, remark (4.6). By applying these results to the situation under consideration we can find a characterization of the optimal cost $J^*_\Delta(x_0)$ of the sampled-data linear quadratic problem:

LEMMA 6.2. *Let $\Delta > 0$ be such that (A_Δ, B_Δ) is stabilizable. Then for each x_0 we have*

$$J^*_\Delta(x_0) = x_0^{\mathrm{T}} P_\Delta x_0,$$

where P_Δ is the largest real symmetric solution of the algebraic Riccati equation

$$A_\Delta^{\mathrm{T}} P_\Delta A_\Delta - P_\Delta + C_{2,\Delta}^{\mathrm{T}} C_{2,\Delta} \tag{6.6}$$

$$-(C_{2,\Delta}^{\mathrm{T}} D_{2,\Delta} + A_\Delta^{\mathrm{T}} P_\Delta B_\Delta)(D_{2,\Delta}^{\mathrm{T}} D_{2,\Delta} + B_\Delta^{\mathrm{T}} P_\Delta B_\Delta)^{+}(D_{2,\Delta}^{\mathrm{T}} C_{2,\Delta} + B_\Delta^{\mathrm{T}} P_\Delta A_\Delta) = 0.$$

Now we will show that as $\Delta \downarrow 0$ the largest real symmetric solution P_Δ of (6.6) converges to P, the largest real symmetric solution of (6.5). We will prove this by proving that for each x_0 we have $J_\Delta^*(x_0) \to J^*(x_0)$. Note that if (A, B) is stabilizable, then for $\Delta > 0$ sufficiently small we have that (A_Δ, B_Δ) is stabilizable.

LEMMA 6.3. *Assume that (A, B) is stabilizable. Then there exists $\Delta_1 > 0$ such that for all $0 < \Delta < \Delta_1$, for all x_0 we have $J_\Delta^*(x_0) < \infty$. For all x_0 we have $\lim_{\Delta\downarrow 0} J_\Delta^*(x_0) = J^*(x_0)$. Also, for all $0 < \Delta < \Delta_1$, P_Δ exists and we have $\lim_{\Delta\downarrow 0} P_\Delta = P$.*

Proof. First of all note that for each sampling period Δ we have $J_\Delta^*(x_0) \geq J^*(x_0)$ for all x_0. This can be shown using that, in fact, for each x_0,

$$J^*(x_0) = \inf\left\{\int_0^\infty \|C_2 x(t) + D_2 u(t)\|^2 dt \mid u \text{ is such that } \lim_{t\to\infty} x(t) = 0\right\}.$$

Hence, by taking u to be generated by the internally stabilizing sampled-data control law $\mathcal{F}_\Delta$, it follows that $J(x_0, \mathcal{F}_\Delta) \geq J^*(x_0)$.

Now, let $\delta > 0$. Let F be such that $\sigma(A + BF) \subset \mathcal{C}^-$ and $J(x_0, F) < J^*(x_0) + \frac{\delta}{2}$. Clearly, $J(x_0, F) = x_0^{\mathrm{T}} L x_0$, where L is the unique solution of the Lyapunov equation

$$(A + BF)^{\mathrm{T}} L + L(A + BF) + (C_2 + D_2 F)^{\mathrm{T}}(C_2 + D_2 F) = 0.$$

Now consider the sampled-data control law $\mathcal{F}_\Delta = H_\Delta F S_\Delta$. By previous arguments, $J(x_0, \mathcal{F}_\Delta) = x_0^{\mathrm{T}} L_\Delta x_0$, where L_Δ is the unique solution of the Lyapunov equation

$$(A_\Delta + B_\Delta F)^{\mathrm{T}} L_\Delta (A_\Delta + B_\Delta F) - L_\Delta + (C_{2,\Delta} + D_{2,\Delta} F)^{\mathrm{T}}(C_{2,\Delta} + D_{2,\Delta} F) = 0.$$

Note that $A_\Delta + B_\Delta F \to I$, $\frac{1}{\Delta}(A_\Delta + B_\Delta F - I) \to A$, and $\frac{1}{\Delta}(C_{2,\Delta} + D_{2,\Delta} F)^{\mathrm{T}}(C_{2,\Delta} + D_{2,\Delta} F) \to (C_2 + D_2 F)^{\mathrm{T}}(C_2 + D_2 F)$ as $\Delta \downarrow 0$. Using a completely similar argument as in the proof of Theorem 6.1 we derive from this that $L_\Delta \to L$, which implies $J(x_0, \mathcal{F}_\Delta) \to J(x_0, F)$. Of course, we also have $J^*(x_0) \leq J_\Delta^*(x_0) \leq J(x_0, \mathcal{F}_\Delta)$. Combining this with $J(x_0, F) < J^*(x_0) + \frac{\delta}{2}$, we find that for δ sufficiently small we have $J^*(x_0) \leq J_\Delta^*(x_0) \leq J^*(x_0) + \delta$. Since δ was arbitrary, this proves the claim. The second statement in the formulation of the theorem is then immediate. □

Let $J_{\Sigma,\mathrm{con}}^*$ be the optimal continuous-time H_2 performance, i.e., the infimum of $J_\Sigma(\Gamma_{\mathrm{con}})$ over all internally stabilizing continuous-time compensators (6.1). It was shown in [15] that if (A, B) is stabilizable and (C_1, A) is detectable, then

$$J_{\Sigma,\mathrm{con}}^* = \mathrm{tr}\,(EE^{\mathrm{T}} P) + \mathrm{tr}\,((A^{\mathrm{T}} P + PA + C_2^{\mathrm{T}} C_2) Q), \tag{6.7}$$

where P is the largest real symmetric solution of the linear matrix inequality (6.5) and Q is the largest real symmetric solution of the dual linear matrix inequality

$$\begin{pmatrix} AQ + QA^{\mathrm{T}} + EE^{\mathrm{T}} & C_1^{\mathrm{T}} Q \\ QC_1 & 0 \end{pmatrix} \geq 0. \tag{6.8}$$

Let $J^*_{\Sigma,\Delta}$ be the optimal sampled-data H_2 performance. If $\Delta \in \boldsymbol{\Delta}$, then we define $J^*_{\Sigma,\Delta} := +\infty$. Our next theorem gives an affirmative answer to the second question posed in the introduction to this section.

THEOREM 6.4. *Let (A, B) be stabilizable and (C_1, A) be detectable. Then there exists Δ_1 such that for all $0 < \Delta < \Delta_1$, $J^*_{\Sigma,\Delta} < \infty$. We have $\lim_{\Delta \downarrow 0} J^*_{\Sigma,\Delta} = J^*_{\Sigma,\mathrm{con}}$.*

In the remainder of this section we will prove this theorem. First, recall the expression (5.1) for $J^*_{\Sigma,\Delta}$. Denote the first term in (5.1) by $I(\Delta)$. Then, under the conditions that (A, B) is stabilizable and (C_1, A) is detectable, we know that for $\Delta \notin \boldsymbol{\Delta}$

$$(6.9)\quad J^*_{\Sigma,\Delta} = I(\Delta) + \frac{1}{\Delta}\mathrm{tr}\,(E_\Delta E_\Delta^{\mathrm{T}} P_\Delta) + \frac{1}{\Delta}\mathrm{tr}\,((A_\Delta^{\mathrm{T}} P_\Delta A_\Delta - P_\Delta + C_{2,\Delta}^{\mathrm{T}} C_{2,\Delta})Q_\Delta)$$

$$- \frac{1}{\Delta}\mathrm{tr}\,((D_{P_\Delta} N_\Delta^* D_{Q_\Delta})(D_{P_\Delta} N_\Delta^* D_{Q_\Delta})^{\mathrm{T}}),$$

where P_Δ is the largest real symmetric solution of (6.6), Q_Δ is the largest real symmetric solution of the dual Riccati equation

$$(6.10)\qquad A_\Delta Q_\Delta A_\Delta^{\mathrm{T}} - Q_\Delta + E_\Delta E_\Delta^{\mathrm{T}} + A_\Delta Q_\Delta C_1^{\mathrm{T}}(C_1 Q_\Delta C_1^{\mathrm{T}})^+ C_1 Q_\Delta A_\Delta = 0,$$

and

$$N_\Delta^* = -D_{P_\Delta}(D_{P_\Delta}^+)^2 D_{P_\Delta} C_{P_\Delta} Q_\Delta C_1^{\mathrm{T}} (D_{Q_\Delta}^+)^2 D_{Q_\Delta}.$$

Here, C_{P_Δ}, D_{P_Δ}, and D_{Q_Δ} are defined by (3.5), (3.4), and (3.7), respectively, with $P = P_\Delta$ and $Q = Q_\Delta$. We will prove that $J^*_{\Sigma,\Delta} \to J^*_{\Sigma,\mathrm{con}}$ by analyzing the asymptotic behavior of the four terms appearing in (6.9) separately:

- It is immediate that the first term, $I(\Delta)$, converges to 0 as $\Delta \downarrow 0$.
- From (2.5) it follows that $\frac{1}{\Delta}E_\Delta E_\Delta^{\mathrm{T}} \to EE^{\mathrm{T}}$. Since also $P_\Delta \to P$, we conclude that the second term, $\frac{1}{\Delta}\mathrm{tr}\,(E_\Delta E_\Delta^{\mathrm{T}} P_\Delta)$, converges to $\mathrm{tr}\,(EE^{\mathrm{T}}P)$.
- To prove convergence of the third term, first note that $Q_\Delta \to Q$. This follows immediately by dualizing Lemma 6.3. Next, as before, rewrite

$$\frac{1}{\Delta}\mathrm{tr}\,(A_\Delta^{\mathrm{T}} P_\Delta A_\Delta - P_\Delta + C_{2,\Delta}^{\mathrm{T}} C_{2,\Delta})Q_\Delta)$$
$$(6.11)\qquad = \frac{1}{\Delta}(A_\Delta^{\mathrm{T}} - I)P_\Delta A_\Delta + P_\Delta \frac{1}{\Delta}(A_\Delta - I) + \frac{1}{\Delta}C_{2,\Delta}^{\mathrm{T}} C_{2,\Delta}.$$

  Since $\frac{1}{\Delta}(A_\Delta - I) \to A$, $A_\Delta \to I$, and $\frac{1}{\Delta}C_{2,\Delta}^{\mathrm{T}} C_{2,\Delta} \to C_2^{\mathrm{T}} C_2$, we conclude that the third term in (6.9) converges to $\mathrm{tr}\,(A^{\mathrm{T}}P + PA + C_2^{\mathrm{T}} C_2)$.
- In order to complete the proof of Theorem 6.4, we should hence prove that the fourth term in (6.9) converges to 0 as $\Delta \downarrow 0$. This is done in the following lemma:

LEMMA 6.5. $\frac{1}{\Delta}\mathrm{tr}\,((D_{P_\Delta} N_\Delta^* D_{Q_\Delta})(D_{P_\Delta} N_\Delta^* D_{Q_\Delta})^{\mathrm{T}}) \to 0$ *as* $\Delta \downarrow 0$.

Proof. Rewrite the fourth term in (6.9) as $\frac{1}{\Delta}\|D_{P_\Delta} N_\Delta^* D_{Q_\Delta}\|^2$, where for any matrix M, $\|M\|$ denotes the Frobenius norm $\mathrm{tr}\,(MM^{\mathrm{T}})$. Note that if M is a given matrix, then M^+M and MM^+ are orthogonal projectors, so consequently $\|MM^+\| = \|MM^+\| = \mathrm{rank}\,(M)$. In particular, this implies that if M is $n \times n$ matrix, then $\|MM^+\| = \|MM^+\| \le n$. Now make the following estimates:

$$\frac{1}{\Delta}\|D_{P_\Delta} N_\Delta^* D_{Q_\Delta}\|^2$$

$$\begin{aligned}
&\leq \frac{1}{\Delta}\|(D_{P_\Delta}D_{P_\Delta}^+)(D_{P_\Delta}^+D_{P_\Delta})C_{P_\Delta}Q_\Delta C_1^{\mathrm{T}}D_{Q_\Delta}^+(D_{Q_\Delta}^+D_{Q_\Delta})\|^2 \\
&\leq \frac{m^4p^2}{\Delta}\|C_{P_\Delta}Q_\Delta C_1^{\mathrm{T}}D_{Q_\Delta}^+\|^2 \\
&\leq \frac{m^4p^2}{\Delta}\|C_{P_\Delta}\|^2\|Q_\Delta C_1^{\mathrm{T}}D_{Q_\Delta}^+\|^2.
\end{aligned}$$

As noted before, $C_{P_\Delta}^{\mathrm{T}}C_{P_\Delta} = A_\Delta^{\mathrm{T}}P_\Delta A_\Delta - P_\Delta + C_{2,\Delta}^{\mathrm{T}}C_{2,\Delta}$, so $\frac{1}{\Delta}\|C_{P_\Delta}\|^2 \to \mathrm{tr}\,(A^{\mathrm{T}}P + PA + C_2^{\mathrm{T}}C_2)$. On the other hand, by noting that Q_Δ satisfies the Riccati equation (6.10), where $A_\Delta = e^{A\Delta}$ is invertible, we see that

$$\begin{aligned}
&\|Q_\Delta C_1^{\mathrm{T}}D_{Q_\Delta}^+\|^2 \\
&\quad = \mathrm{tr}\,(Q_\Delta C_1^{\mathrm{T}}(C_1Q_\Delta C_1)^+C_1Q_\Delta) \\
&\quad = \mathrm{tr}\,(Q_\Delta - A_\Delta^{-1}Q_\Delta A_\Delta^{-T} + A_\Delta^{-1}E_\Delta E_\Delta^{\mathrm{T}}A_\Delta^{-\mathrm{T}}).
\end{aligned}$$

Since $Q_\Delta \to Q$, $A_\Delta^{-1} \to I$ and $E_\Delta E_\Delta^{\mathrm{T}} \to 0$, the latter converges to zero as $\Delta \downarrow 0$. □

REFERENCES

[1] B. Bamieh and J. Pearson, *A general framework for linear periodic systems with application to H_∞ sampled-data control*, IEEE Trans. Autom. Control, 37 (1992), pp. 418–435.

[2] ———, *The H_2 problem for sampled data systems*, Systems and Control Letters, 19 (1992), pp. 1–12.

[3] B. Bamieh, J. Pearson, B. Francis, and A. Tannenbaum, *A lifting technique for linear periodic systems with applications to sampled-data control*, Systems Control Lett., 17 (1991), pp. 79–88.

[4] T. Chen, *A simple derivation of the H_2-optimal sampled-data controllers*, Preprint, University of Calgary, June 1992.

[5] T. Chen and B. Francis, *On the L_2 induced norm of a sampled-data system*, Systems Control Lett., 15 (1990), pp. 211–219.

[6] ———, *H_2-optimal sampled data control*, IEEE Trans. Autom. Control, AC-36 (1991), pp. 387–397.

[7] ———, *Linear time-varying H_2 optimal control of sampled-data systems*, Automatica, 27 (1991), pp. 963–974.

[8] B. Francis and T. Georgiou, *Stability theory for linear time-invariant plants with periodic digital controllers*, IEEE Trans. Autom. Control, 33 (1988), pp. 820–832.

[9] M. Hautus and L. Silverman, *System structure and singular control*, Linear Algebra Appl., 50 (1983), pp. 369–402.

[10] P. Kabamba and S. Hara, *On computing the induced norm of a sampled data system*, in Proc. American Control Conference, San Diego, CA, 1990, pp. 319–320.

[11] ———, *Worst case analysis and design of sampled data control systems*, Preprint, 1990.

[12] P. Khargonekar and N. Sivashankar, *H_2 optimal control for sampled-data systems*, Systems Control Lett., 17 (1992), pp. 425–436.

[13] Y. H. R. Kalman and K. Narendra, *Controllability of linear dynamical systems*, in Contributions to Differential Equations, vol. 1, Interscience, New York, 1963.

[14] L. Silverman, *Discrete Riccati equations*, in Control and Dynamic Systems, Advances in Theory and Applications, C. Leondes, ed., Academic Press, New York, 1976, pp. 313–386.

[15] A. Stoorvogel, *The singular H_2 control problem*, Automatica, 28 (1992), pp. 627–631.

[16] A. Stoorvogel and J. van der Woude, *The disturbance decoupling problem with measurement feedback and stability for systems with direct feedthrough matrices*, Systems Control Lett., 17 (1991), pp. 217–226.

[17] H. Toivonen, *Sampled-data control of continuous-time systems with an H_∞ optimality criterion*, Automatica, 28 (1992), pp. 45–54.

[18] J. Willems, A. Kitapci, and L. Silverman, *Singular optimal control, a geometric approach*, SIAM J. Control Optim., 24 (1986), pp. 323–337.

[19] W. M. Wonham, *Linear Multivariable Control: a Geometric Approach*, Springer-Verlag, New York, 1979.

[20] Y. YAMAMOTO, *New approach to sampled-data systems: a function space method*, in Proc. 29th CDC, 1990, pp. 1882–1887.

[21] ———, *On the state space and frequency domain characterization of H_∞-norm of sampled-data systems*, Preprint, 1991.