

Benign Geometry and Distributional Robustness of $\mathcal{H}_2$ Synthesis

Arman Sharifi Kolarijani^{1,2}, Peyman Mohajerin Esfahani³, Tamás Keviczky²,
and Mohamad Amin Sharifi Kolarijani²

¹Alpha Brain Technologies, The Netherlands

²Delft University of Technology, The Netherlands

³University of Toronto, Canada

ABSTRACT. In this paper, we study standard and distributionally robust $\mathcal{H}_2$ synthesis problem of a stabilizing state-feedback controller for discrete-time linear time-invariant systems. Without requiring a nonsingular disturbance controllability Gramian or a positive-definite control penalty, we establish that stationarity of a stabilizing gain in standard $\mathcal{H}_2$ synthesis is equivalent to global optimality and derive an exact Pythagorean identity for the performance difference. We then generalize the $\mathcal{H}_2$ synthesis problem by considering i.i.d. disturbances with zero mean and uniformly bounded second moments, and extend the stationarity-optimality equivalence to the corresponding distributionally robust $\mathcal{H}_2$ synthesis. Moreover, we show that every standard $\mathcal{H}_2$ -optimal gain is simultaneously optimal for every such distributionally robust problem. Finally, we specialize the framework to Frobenius, Kullback–Leibler, and Wasserstein-2 ambiguity sets and obtain explicit characterizations of their worst-case covariances, which, in turn, provide performance certificates for a common optimal controller.

KEYWORDS: $\mathcal{H}_2$ synthesis; benign non-convexity; distributionally robust control; covariance ambiguity.

1. Introduction

Reliable feedback design in the presence of uncertainty and exogenous disturbances is a fundamental challenge across control applications. Robust control provides a systematic framework for addressing this challenge, with influence extending beyond classical engineering to economics [23], networked systems [10], biological systems [28], and optimization [2].

Among the principal robust-performance frameworks are $\mathcal{H}_2$, $\mathcal{H}_\infty$, and mixed $\mathcal{H}_2/\mathcal{H}_\infty$ synthesis [13, 50]. The $\mathcal{H}_2$ criterion in particular minimizes the closed-loop impulse-response energy, or, equivalently, the steady-state performance output power under zero-mean white noise with unit covariance. The main difficulty is of course that the $\mathcal{H}_2$ synthesis leads to a nonconvex problem, even in the case of static state-feedback design. However, the state-feedback $\mathcal{H}_2$ synthesis problem admits a convex linear matrix inequality (LMI) formulation using Lyapunov inequalities and trace

Date: September 24, 2026.

Correspondence to: Mohamad Amin Sharifi Kolarijani <m.a.sharifikolarijani@tudelft.nl>.

This work was partially supported by the Horizon Europe Pathfinder Open project RELIEVE-101099481.

bounds on closed-loop Gramians [4, 7], after suitable changes of variables [34]. Another classical synthesis method uses the algebraic Riccati equation (ARE) [50]. Indeed, state-feedback $\mathcal{H}_2$ synthesis is closely related to linear quadratic regulation (LQR) problem by interpreting the squared performance output as a quadratic state-and-control cost [26, 42]. Under the usual regular LQR assumptions, the stabilizing Riccati solution yields an $\mathcal{H}_2$ -optimal gain.

The Riccati characterization also connects classical $\mathcal{H}_2$ synthesis to optimality conditions expressed directly in the feedback gain for the LQR problem. To be precise, gradient-dominance results under regular LQR assumptions establish that first-order stationarity is necessary and *sufficient* for global optimality [18, 6, 46]. Such bounds also underpin global convergence guarantees for suitable policy-gradient methods. However, these results require certain assumptions on the problem data that, among others, include positive-definite control penalty and positive definiteness of the accumulated state covariance [46, Sec. 3.3]. The latter condition, in particular, translates to nonsingularity of the disturbance controllability Gramian in the context of $\mathcal{H}_2$ synthesis. This motivates examining stationarity and global optimality in $\mathcal{H}_2$ synthesis without requiring these conditions.

The standard formulation of $\mathcal{H}_2$ synthesis prescribes the disturbance covariance. However, in many applications, the underlying distribution is often estimated from finite data, affected by distribution shift, or only partially known. Distributionally robust optimization (DRO) [12, 48, 37] addresses such misspecification by optimizing against the worst-case probability distribution in an ambiguity set. In its generic form, DRO seeks a decision x solving $\inf_{x \in \mathcal{X}} \sup_{\mathbf{P} \in \mathcal{D}} \mathbb{E}_{\xi \sim \mathbf{P}}[\ell(x, \xi)]$, where ℓ is a loss function and $\mathcal{D}$ encodes the available distributional information. Different ambiguity sets induce different notions of robustness. Moment-based sets constrain low-order statistics and are attractive when only means and covariances can be estimated reliably [12, 48]. Kullback–Leibler (KL) divergence produces relative-entropy neighborhoods and exponential-tilting representations [24]. Wasserstein ambiguity sets compare distributions through optimal transport, accommodate support mismatch, and often yield finite-sample guarantees and tractable reformulations in data-driven settings [35, 21]. In control, moment-based ambiguity has been used to enforce distributionally robust chance and conditional value-at-risk constraints [43] and to address LQR problems with multiplicative disturbance [11]. Another line of research formulates minimax state-feedback control with adversarial conditional disturbance distributions, using Wasserstein ambiguity or penalties [49, 27] and stagewise relative-entropy constraints [16]. Relative-entropy duality also connects such formulations to exponential criteria in risk-sensitive control [25, 47, 36]. For partially observed systems controlled through output feedback, optimality of linear policies has been established under Wasserstein and KL ambiguity on temporally independent disturbances [41, 20]. Related studies impose a common disturbance distribution across time [40] or consider infinite-horizon constrained control with temporal dependence [5]. More directly, Gramlich et al. [22] connect Wasserstein distributionally robust control to $\mathcal{H}_2$ synthesis and derive convex LMI formulations for independent and temporally correlated disturbances. These formulations differ in their information patterns, temporal assumptions, and admissible disturbance distributions. For state-feedback with zero-mean i.i.d. additive disturbances, the regular LQR gain is already independent of the disturbance covariance and consequently optimal across the corresponding distributional ambiguity sets, as also

noted in [41]. We note, however, that uncertainty in disturbance statistics changes the performance guarantee associated with a controller, even when the same gain remains optimal for all admissible covariances. These observations motivate examining the persistence of nominal optimality under distributional ambiguity beyond the regular LQR setting.

In this paper, we consider a DR framework for state-feedback $\mathcal{H}_2$ synthesis for discrete-time linear time-invariant (LTI) systems with full state observation. The ambiguity concerns the common marginal distribution of an i.i.d. zero-mean disturbance sequence that belongs to a set $\mathcal{D}$ that is independent of the gain and has uniformly bounded second moments. Following the direct gain-optimization viewpoint originating with [26] and extended to output feedback in [31], we consider

$$\inf_{K \in \mathcal{K}_s} \sup_{P \in \mathcal{D}} v(K, P),$$

where $\mathcal{K}_s$ is the set of stabilizing gains and $v(K, P)$ is the generalized $\mathcal{H}_2$ cost accounting for a generic disturbance distribution P . In particular, $v(K, P)$ reduces to the standard $\mathcal{H}_2$ cost for $P = \mathcal{N}(0, I)$, that is, normal distribution with zero mean and unit covariance. Our analysis only assumes attainment of the optimal value in the standard $\mathcal{H}_2$ synthesis problem and does not impose any extra restrictions on the problem data such as nonsingularity of disturbance controllability Gramian or positive-definiteness of control penalty. Our **main contributions** are as follows:

- **Sufficiency of stationarity for $\mathcal{H}_2$ -optimality:** We show that first-order stationarity is necessary and sufficient for global optimality for the DR $\mathcal{H}_2$ synthesis (Theorem 3.6). In particular, for the standard $\mathcal{H}_2$ synthesis this result follows from an exact Pythagorean performance-difference identity (Theorem 2.3).
- **Distributionally robustness of standard $\mathcal{H}_2$ -optimal gain:** Exploiting the Loewner minimality of their *compressed* observability Gramian (Corollary 2.4), we show that *every* standard $\mathcal{H}_2$ -optimal gain is simultaneously optimal for the considered DR $\mathcal{H}_2$ synthesis problems (Theorem 3.7).
- **Performance certificates for specific ambiguity sets:** We obtain explicit worst-case covariances/distributions and provide robustness certifications for the DR $\mathcal{H}_2$ synthesis problems with ambiguity sets defined based on the Frobenius, KL, and Wasserstein-2 (W2) balls around a nominal covariance/distribution (Propositions 4.2, 4.4, and 4.6).

The remainder of the paper is organized as follows. In Section 2, we revisit standard $\mathcal{H}_2$ synthesis and analyze the regularity properties of the corresponding cost and optimal solution. In Section 3, we generalize $\mathcal{H}_2$ cost by considering i.i.d., zero-mean disturbances with arbitrary covariance, and analyze the regularity properties of the corresponding DR cost and optimal solution. In Section 4, we specialize the general results of Section 3 to ambiguity sets defined by Frobenius-norm, KL-divergence, and W2 balls. Section 5 concludes the paper with some final remarks on the limitations of the present paper.

Notations. The set of real numbers is denoted by $\mathbf{R}$. We also denote $\mathbf{N}_0 := \{0, 1, \dots\}$. For a vector $x \in \mathbf{R}^n$, the 2-norm is denoted by $\|x\|_2 = \sqrt{x^\top x}$. For a matrix $M \in \mathbf{R}^{n \times n}$, $\text{tr}(M)$ is the trace of M (i.e., the sum of diagonal entries of M), and $\rho(M)$ is spectral radius of M (i.e., the

largest absolute value among its eigenvalues). For a symmetric matrix $M \in \mathbf{R}^{n \times n}$, we use $M \succeq 0$ (respectively, $M \succ 0$) to denote M is positive semi-definite (respectively, positive definite). We use I to denote the identity matrix. For two matrices $M, N \in \mathbf{R}^{n \times m}$, $M^\top$ is the transpose of M , $\langle M, N \rangle = \text{tr}(M^\top N)$ is the (Frobenius) inner product of M and N , $\|M\|_F = \sqrt{\langle M, M \rangle}$ is the Frobenius norm of M , and $\text{range}(M)$ is the range space of M . $\mathcal{B}(\mathbf{R}^p)$ is the set of Borel probability measures on $\mathbf{R}^p$. We use $w \sim \mathbf{P} \in \mathcal{B}(\mathbf{R}^p)$ to denote that the random vector $w \in \mathbf{R}^p$ has the distribution $\mathbf{P}$. $\mathbb{E}[\cdot]$ denotes the expectation operation; we occasionally use the subscript as in $\mathbb{E}_{\mathbf{P}}[\cdot]$ to emphasize that the expectation is with respect to a random variable with distribution $\mathbf{P}$. For a random vector $w \sim \mathbf{P} \in \mathcal{B}(\mathbf{R}^p)$, we use $\mu_{\mathbf{P}} := \mathbb{E}_{\mathbf{P}}[w]$ and $\Sigma_{\mathbf{P}} := \mathbb{E}_{\mathbf{P}}[ww^\top]$ to denote its first and second moments, respectively. For distributions with zero mean then $\Sigma_{\mathbf{P}}$ denotes also the covariance. We use $\mathcal{N}(\mu, \Sigma)$ to denote the Gaussian distribution with mean μ and covariance Σ . For a linear subspace $\mathcal{L} \subset \mathbf{R}^{m \times n}$, we use $\mathcal{L}^\perp$ to denote the orthogonal complement of $\mathcal{L}$ with respect to the Frobenius inner product. Throughout the paper, we use K to denote the static state-feedback gain. To avoid clutter, we exclusively use K in the subscript to show the dependence on K ; e.g., A_K denotes the dependence of the variable A_K on the gain K .

2. Standard $\mathcal{H}_2$ synthesis

In this section, we revisit the standard full-state $\mathcal{H}_2$ optimal control problem. We analyze the regularity properties of the corresponding cost, and, in particular, show that first-order stationarity is a *sufficient* condition for optimality of a stabilizing gain.

2.1. Standard $\mathcal{H}_2$ cost

Consider the discrete-time LTI system

$$\begin{cases} x_{t+1} &= Ax_t + Bu_t + Ew_t, \\ z_t &= Cx_t + Du_t, \end{cases} \quad t \in \mathbf{N}_0, \quad (1)$$

where $x_t \in \mathbf{R}^n$ is the system state, $u_t \in \mathbf{R}^m$ is the control input, $w_t \in \mathbf{R}^p$ is a disturbance input, and $z_t \in \mathbf{R}^q$ is the performance signal. A , B , E , C , and D are matrices of compatible dimensions that describe the dynamics of the system. Under linear state-feedback law $u_t = Kx_t$, where $K \in \mathbf{R}^{m \times n}$ is the gain, the closed-loop system becomes

$$\begin{cases} x_{t+1} &= A_K x_t + Ew_t, \\ z_t &= C_K x_t, \end{cases} \quad (2a)$$

where

$$A_K := A + BK \quad \text{and} \quad C_K := C + DK. \quad (2b)$$

In this paper, we are concerned with the $\mathcal{H}_2$ synthesis of a stabilizing state-feedback controller. Let us define

$$\mathcal{K}_s := \{K \in \mathbf{R}^{m \times n} : \rho(A_K) < 1\},$$

to be the set of all stabilizing controllers for which A_K is Schur. Notice that $\mathcal{K}_s$ non-empty if and only if the pair (A, B) is stabilizable. The standard $\mathcal{H}_2$ synthesis is then formulated as

$$\inf_{K \in \mathcal{K}_s} \{v^s(K) := \|G_K\|_{\mathcal{H}_2}^2\}, \quad (3)$$

with the cost function being the squared $\mathcal{H}_2$ -norm of the closed-loop transfer function

$$G_K(z) := C_K(zI - A_K)^{-1}E,$$

from w to z [50, Sec. 21.5]. This cost can be expressed in the frequency domain or, equivalently, in the time domain as the sum of the squared norms of the impulse response of the system. This equivalence is a consequence of Parseval's theorem [39, Thm. 8.16]. In particular, we have

$$v^s(K) = \|G_K\|_{\mathcal{H}_2}^2 := \frac{1}{2\pi} \int_0^{2\pi} \text{tr} (G_K(e^{j\omega})^* G_K(e^{j\omega})) d\omega = \sum_{i=0}^{\infty} \|C_K A_K^i E\|_F^2. \quad (4)$$

The standard $\mathcal{H}_2$ cost can be alternatively expressed using the *observability* Gramian of the pair (A_K, C_K) , given by [8, Ch. 6.6]

$$X_K := \sum_{t=0}^{\infty} (A_K^t)^\top C_K^\top C_K A_K^t. \quad (5)$$

Indeed, we have (see, e.g., [50, Rem. 21.6])

$$v^s(K) = \langle I, M_K \rangle, \quad \forall K \in \mathcal{K}_s, \quad (6a)$$

with the *compressed* observability Gramian defined as

$$M_K := E^\top X_K E. \quad (6b)$$

The preceding representation follows directly from the time-domain expression; see also Lemma 3.1. We note that, for each $K \in \mathcal{K}_s$, the Gramian X_K is the unique positive semi-definite solution of the Lyapunov equation [50, Lem. 21.2]

$$\mathbb{L}_{A_K}(X_K) + C_K^\top C_K = 0 \quad \text{where} \quad \mathbb{L}_A(X) := A^\top X A - X. \quad (7)$$

We next establish smoothness and compute the gradient of the standard cost v^s . Our derivations are direct applications of differential calculus to matrices; see, e.g., [32]. Similar results are provided for the LQR problem in, e.g., [18, 17]. For the purpose of generality, given a matrix $\Sigma \succeq 0$ and any stabilizing gain $K \in \mathcal{K}_s$, let us define the corresponding *controllability* Gramian of the pair $(A_K, E\Sigma^{1/2})$ by [50, Lem. 21.2]

$$Y_K(\Sigma) := \sum_{t=0}^{\infty} A_K^t E \Sigma E^\top (A_K^t)^\top. \quad (8)$$

We notice that $Y_K(\Sigma)$ is the unique positive semi-definite solution of the Lyapunov equation [50, Lem. 21.2]

$$\mathbb{L}_{A_K}^*(Y_K(\Sigma)) + E \Sigma E^\top = 0 \quad \text{where} \quad \mathbb{L}_A^*(Y) := A Y A^\top - Y. \quad (9)$$

In particular, let

$$Y_K^s := Y_K(I),$$

denote the controllability Gramian of the pair (A_K, E) .

Lemma 2.1 (Regularity of v^s). *The standard cost v^s is real-analytic on $\mathcal{K}_s$, and*

$$\nabla_K v^s(K) = 2S_K Y_K^s \quad \text{where} \quad S_K := B^\top X_K A_K + D^\top C_K, \quad \forall K \in \mathcal{K}_s. \quad (10)$$

Proof. Recall that $v^s(K) = \langle I, M_K \rangle$ for $K \in \mathcal{K}_s$. We start with showing v^s is real-analytic on $\mathcal{K}_s$. Observe that the mapping $A_K \mapsto X_K$ admits the convergent series representation (5) for all Schur A_K . Therefore, X_K depends real-analytically on the entries of $A_K = A + BK$ and hence on K in the constraint set $\mathcal{K}_s$. Now, observe that the maps $X_K \mapsto M_K = E^\top X_K E$ and $M_K \mapsto v^s(K) = \langle I, M_K \rangle$ are both linear. Thus, the map $K \mapsto v^s(K)$ is real-analytic for all $K \in \mathcal{K}_s$.

We next derive the gradient of v^s . First, observe that

$$dv^s = \langle I, dM_K \rangle = \langle I, E^\top (dX_K) E \rangle = \langle EE^\top, dX_K \rangle. \quad (11)$$

Recall now the Lyapunov operator $\mathbb{L}_{A_K}$ and its adjoint operator $\mathbb{L}_{A_K}^*$ on the space of n -by- n , real, symmetric matrices and notice the corresponding adjoint identity

$$\langle Y, \mathbb{L}_{A_K}(X) \rangle = \langle \mathbb{L}_{A_K}^*(Y), X \rangle. \quad (12)$$

By assumption, we have $\mathbb{L}_{A_K}^*(Y_K^s) + EE^\top = 0$. Then, plugging in $EE^\top = -\mathbb{L}_{A_K}^*(Y_K^s)$ to (11) and using the identity (12), we obtain

$$dv^s = \langle -\mathbb{L}_{A_K}^*(Y_K^s), dX_K \rangle = \langle -Y_K^s, \mathbb{L}_{A_K}(dX_K) \rangle. \quad (13)$$

Now, observe that, using the Lyapunov equation (7), that is,

$$\mathbb{L}_{A_K}(X_K) + C_K^\top C_K = A_K^\top X_K A_K - X_K + C_K^\top C_K = 0,$$

we have

$$(dA_K)^\top X_K A_K + A_K^\top (dX_K) A_K + A_K^\top X_K (dA_K) - dX_K + (dC_K)^\top C_K + C_K^\top (dC_K) = 0,$$

and hence

$$\mathbb{L}_{A_K}(dX_K) = -(dA_K)^\top X_K A_K - A_K^\top X_K (dA_K) - (dC_K)^\top C_K - C_K^\top (dC_K).$$

Plugging in the latter equality in (13) and using the symmetry of Y_K^s , we arrive at

$$dv^s = 2\langle Y_K^s, A_K^\top X_K (dA_K) + C_K^\top (dC_K) \rangle.$$

Finally, we use the identities $dA_K = B(dK)$ and $dC_K = D(dK)$ to get

$$dv^s = \langle Y_K^s, 2(A_K^\top X_K B + C_K^\top D)(dK) \rangle = \langle 2S_K Y_K^s, dK \rangle.$$

By definition, we have $dv^s = \langle \nabla_K v^s(K), dK \rangle$ and the claim follows. $\square$

2.2. Properties of standard $\mathcal{H}_2$ -optimal solution

In this subsection, we discuss two important properties of solution of the standard $\mathcal{H}_2$ problem (3). In this regard, let us notice that non-emptiness of $\mathcal{K}_s$ is a sufficient condition for problem (3) to have a finite optimal value [42, Thm. 4.8]. Indeed, there exists a stabilizing LQ-optimal value matrix $X_{\text{LQ}} \succeq 0$, that is, the largest real symmetric solution of the generalized discrete algebraic Riccati equation (DARE) associated with (A, B, C, D) , such that [42, Sec. 3 and 4]

$$\langle I, E^\top X_{\text{LQ}} E \rangle = \inf_{K \in \mathcal{K}_s} v^s(K). \quad (14)$$

However, for the optimal value to be attained by a stabilizing gain, extra conditions are required; we refer the reader to [42, Sec. 4] for the detailed discussion on the necessary and sufficient conditions for the existence of a stabilizing optimal solution for the standard $\mathcal{H}_2$ synthesis. The following remark provides a set of sufficient conditions for which the regular LQ-optimal gain is also a solution of the standard- $\mathcal{H}_2$ synthesis problem.

Remark 2.2 (Regular LQ-optimal and standard $\mathcal{H}_2$ -optimal gains). *Assume that the pair (A, B) is stabilizable, the pair (A, C) is detectable, $C^\top D = 0$, and $D^\top D \succ 0$. In this case the solution to the regular LQR problem is also a solution of the standard- $\mathcal{H}_2$ synthesis problem [42, Sec. 3]. To be precise, the LQ-optimal feedback gain*

$$K_{\text{LQ}} = -(D^\top D + B^\top X_{K_{\text{LQ}}} B)^{-1} B^\top X_{K_{\text{LQ}}} A,$$

is a standard $\mathcal{H}_2$ -optimal stabilizing gain, where $X_{K_{\text{LQ}}} = X_{\text{LQ}}$ is the unique positive semi-definite solution of the DARE

$$\mathbb{L}_A(X) + C^\top C - A^\top X B (D^\top D + B^\top X B)^{-1} B^\top X A = 0.$$

Observe that the stabilizing LQ-optimal gain K_{LQ} is independent of the disturbance matrix E .

Our next results establishes that stationarity is not only necessary, but also sufficient for optimality of a stabilizing gain for the standard $\mathcal{H}_2$ problem.

Theorem 2.3 (Stationarity and optimality for v^s). *A stabilizing gain $K_\star \in \mathcal{K}_s$ is a solution of the standard $\mathcal{H}_2$ problem (3) if and only if $\nabla_K v^s(K_\star) = 0$. In fact, if $K_\star \in \mathcal{K}_s$ and $\nabla_K v^s(K_\star) = 0$, then*

$$\|G_K\|_{\mathcal{H}_2}^2 - \|G_{K_\star}\|_{\mathcal{H}_2}^2 = \|G_K - G_{K_\star}\|_{\mathcal{H}_2}^2, \quad \forall K \in \mathcal{K}_s. \quad (15)$$

Proof. By Lemma 2.1, v^s is real-analytic and hence continuously differentiable on $\mathcal{K}_s$. The necessity of stationarity then simply follows first-order condition. We next focus on the sufficiency of first-order condition for optimality. Hence, let $K_\star \in \mathcal{K}_s$ be such that $\nabla_K v^s(K_\star) = 0$. By Lemma 2.1, we thus have

$$S_{K_\star} Y_{K_\star}^s = 0. \quad (16)$$

Then, since $A_{K_\star}$ is Schur, we can use the convergent series representation (8) of the Gramian $Y_{K_\star}^s$, to write

$$0 = S_{K_\star} Y_{K_\star}^s S_{K_\star}^\top = \sum_{t=0}^{\infty} S_{K_\star} A_{K_\star}^t E E^\top (A_{K_\star}^\top)^t S_{K_\star}^\top = \sum_{t=0}^{\infty} (S_{K_\star} A_{K_\star}^t E) (S_{K_\star} A_{K_\star}^t E)^\top.$$

Every summand on the right-hand side of the preceding equation is positive semidefinite. Hence, every summand must be zero, which yields

$$S_{K_\star} A_{K_\star}^t E = 0, \quad \forall t \in \mathbf{N}_0. \quad (17)$$

Thus, although $S_{K_\star}$ need not vanish on the entire state space, it vanishes on the closed-loop disturbance-reachable subspace.

Now, let $K \in \mathcal{K}_s$ be arbitrary. Also, let $e_j \in \mathbf{R}^p$, $j \in \{1, \dots, p\}$, denote the canonical basis vectors of the disturbance space. We now consider the disturbance impulse-response with state-feedback gains $K_\star$ and K , one at a time for a fixed j . To that end, we consider the deterministic trajectory (i.e., with $w_t = 0$ for all $t \in \mathbf{N}_0$) produced by $K_\star$ and K with initial state $x_0 = E e_j$, by defining the two state, input, and performance sequences

$$\begin{aligned} x_{t,j}^\star &:= A_{K_\star}^t E e_j, & u_{t,j}^\star &:= K_\star x_{t,j}^\star, & z_{t,j}^\star &:= C_{K_\star} x_{t,j}^\star, \\ x_{t,j} &:= A_K^t E e_j, & u_{t,j} &:= K x_{t,j}, & z_{t,j} &:= C_K x_{t,j}. \end{aligned}$$

We also introduce the trajectory differences

$$\delta x_{t,j} := x_{t,j} - x_{t,j}^\star, \quad \delta u_{t,j} := u_{t,j} - u_{t,j}^\star, \quad \delta z_{t,j} := z_{t,j} - z_{t,j}^\star.$$

Since both trajectories have the same initial condition, we have $\delta x_{0,j} = 0$. Moreover, since both trajectories satisfy the original dynamics, we get

$$\delta x_{t+1,j} = A \delta x_{t,j} + B \delta u_{t,j}, \quad \delta z_{t,j} = C \delta x_{t,j} + D \delta u_{t,j}. \quad (18)$$

We next construct the costate along the trajectory produced by $K_\star$. Define

$$p_{t,j}^\star := X_{K_\star} x_{t,j}^\star.$$

We now derive two identities satisfied by this sequence. Multiplying the Lyapunov equation (7) (with $K = K_\star$) by $x_{t,j}^\star$ gives

$$p_{t,j}^\star = X_{K_\star} x_{t,j}^\star = A_{K_\star}^\top X_{K_\star} A_{K_\star} x_{t,j}^\star + C_{K_\star}^\top C_{K_\star} x_{t,j}^\star.$$

Then, using $x_{t+1,j}^\star = A_{K_\star} x_{t,j}^\star$ and $z_{t,j}^\star = C_{K_\star} x_{t,j}^\star$, we arrive at

$$p_{t,j}^\star = A_{K_\star}^\top p_{t+1,j}^\star + C_{K_\star}^\top z_{t,j}^\star.$$

Expanding $A_{K_\star} = A + B K_\star$ and $C_{K_\star} = C + D K_\star$, we then have

$$p_{t,j}^\star = A^\top p_{t+1,j}^\star + C^\top z_{t,j}^\star + K_\star^\top (B^\top p_{t+1,j}^\star + D^\top z_{t,j}^\star).$$

On the other hand, using again $x_{t+1,j}^\star = A_{K_\star} x_{t,j}^\star$ and $z_{t,j}^\star = C_{K_\star} x_{t,j}^\star$, we obtain

$$\begin{aligned} B^\top p_{t+1,j}^\star + D^\top z_{t,j}^\star &= B^\top X_{K_\star} A_{K_\star} x_{t,j}^\star + D^\top C_{K_\star} x_{t,j}^\star = (B^\top X_{K_\star} A_{K_\star} + D^\top C_{K_\star}) x_{t,j}^\star \\ &= S_{K_\star} x_{t,j}^\star = S_{K_\star} A_{K_\star}^t E e_j = 0. \end{aligned}$$

By (17), it then follows that

$$B^\top p_{t+1,j}^\star + D^\top z_{t,j}^\star = 0, \quad (19)$$

and as a result,

$$p_{t,j}^* = A^\top p_{t+1,j}^* + C^\top z_{t,j}^*. \quad (20)$$

With the construction above, we can now show that the stationary performance sequence $(z_{t,j}^*)_{t=0}^\infty$ and the feasible performance variation sequence $(\delta z_{t,j}^*)_{t=0}^\infty$ are orthogonal. Indeed, using (18), we have

$$(z_{t,j}^*)^\top \delta z_{t,j} = (z_{t,j}^*)^\top (C \delta x_{t,j} + D \delta u_{t,j}) = (C^\top z_{t,j}^*)^\top \delta x_{t,j} + (D^\top z_{t,j}^*)^\top \delta u_{t,j}.$$

Then, using (19) and (20), we obtain

$$\begin{aligned} (z_{t,j}^*)^\top \delta z_{t,j} &= (p_{t,j}^* - A^\top p_{t+1,j}^*)^\top \delta x_{t,j} - (B^\top p_{t+1,j}^*)^\top \delta u_{t,j} \\ &= (p_{t,j}^*)^\top \delta x_{t,j} - (p_{t+1,j}^*)^\top (A \delta x_{t,j} + B \delta u_{t,j}). \end{aligned}$$

Using the difference dynamics (18), we hence arrive at

$$(z_{t,j}^*)^\top \delta z_{t,j} = (p_{t,j}^*)^\top \delta x_{t,j} - (p_{t+1,j}^*)^\top \delta x_{t+1,j}.$$

Summing the preceding equation from $t = 0$ to $t = T$ gives

$$\sum_{t=0}^T (z_{t,j}^*)^\top \delta z_{t,j} = (p_{0,j}^*)^\top \delta x_{0,j} - (p_{T+1,j}^*)^\top \delta x_{T+1,j}. \quad (21)$$

The first term on the right-hand side is zero since $\delta x_{0,j} = 0$. Also, since both $A_{K_\star}$ and A_K are Schur stable, $x_{t,j}^*$ and $x_{t,j}$ converge exponentially to zero as $t \rightarrow \infty$. This, in turn, implies that $\delta x_{t,j} = x_{t,j} - x_{t,j}^*$ and $p_{t,j}^* = X_{K_\star} x_{t,j}^*$ converges exponentially to zero as $t \rightarrow \infty$. As a result, we have

$$\lim_{T \rightarrow \infty} (p_{T+1,j}^*)^\top \delta x_{T+1,j} = 0.$$

Taking the limit in (21) therefore gives

$$\sum_{t=0}^{\infty} (z_{t,j}^*)^\top \delta z_{t,j} = 0. \quad (22)$$

That is, the performance trajectory generated by $K_\star$ is orthogonal to its difference from the trajectory generated by any other stabilizing K . Now, observe that since $z_{t,j} = z_{t,j}^* + \delta z_{t,j}$, we have

$$\|z_{t,j}\|_2^2 = \|z_{t,j}^*\|_2^2 + 2(z_{t,j}^*)^\top \delta z_{t,j} + \|\delta z_{t,j}\|_2^2.$$

Then, summing over $t \in \mathbf{N}_0$ and $j \in \{1, \dots, p\}$ and using (22), we arrive at

$$\sum_{j=1}^p \sum_{t=0}^{\infty} \|z_{t,j}\|_2^2 - \sum_{j=1}^p \sum_{t=0}^{\infty} \|z_{t,j}^*\|_2^2 = \sum_{j=1}^p \sum_{t=0}^{\infty} \|z_{t,j} - z_{t,j}^*\|_2^2. \quad (23)$$

Now, observe that

$$\sum_{j=1}^p \sum_{t=0}^{\infty} \|z_{t,j}\|_2^2 = \sum_{j=1}^p \sum_{t=0}^{\infty} \|C_K A_K^t E e_j\|_2^2 = \sum_{t=0}^{\infty} \|C_K A_K^t E\|_F^2.$$

Therefore, using the representation (4) of the standard $\mathcal{H}_2$ cost, the equality (23) can be written as

$$\|G_K\|_{\mathcal{H}_2}^2 - \|G_{K_\star}\|_{\mathcal{H}_2}^2 = \|G_K - G_{K_\star}\|_{\mathcal{H}_2}^2 \geq 0.$$

Then, since $K \in \mathcal{K}_s$ was arbitrary, $K_\star$ is globally optimal. This concludes the proof. $\square$

The proof of the preceding result is inspired by the completion-of-squares treatment of singular discrete-time LQR and $\mathcal{H}_2$ problems in [42], combined with the Lyapunov policy-gradient representation used for modern LQR policy optimization in, e.g., [18]. Let us note that the existing gradient-domination bounds in [18, Lem. 3 and Cor. 4] and [46, Thm. 3] imply global optimality of stationary policies under the regular LQR conditions in Remark 2.2 and when the disturbance controllability Gramian $Y_{K_\star}^s$ is *nonsingular*. Theorem 2.3 however does not impose any conditions and, in particular, allows for $Y_{K_\star}^s$ to be singular. The main additional observation is that stationarity of $K_\star$ implies that the residual $S_{K_\star}$ vanishes on the closed-loop disturbance-reachable subspace (that is, $\text{range}(Y_{K_\star}^s)$), even when it need not vanish on the entire state space. This leads to the exact Pythagorean performance-difference identity (15) without assuming a positive-definite controllability Gramian. Theorem 2.3 also provides us with the following instrumental property of standard $\mathcal{H}_2$ -optimal solution.

Corollary 2.4 (Minimality of $M_{K_\star}$). *Let the stabilizing gain $K_\star \in \mathcal{K}_s$ be a solution of the standard $\mathcal{H}_2$ problem (3). Then,*

$$M_{K_\star} \preceq M_K, \quad \forall K \in \mathcal{K}_s.$$

Proof. Optimality of $K_\star$ for the standard $\mathcal{H}_2$ problem (3) implies that (see (17) in the proof of Theorem 2.3)

$$S_{K_\star} A_{K_\star}^t E = 0, \quad t \in \mathbf{N}_0.$$

For any $y \in \mathbb{R}^p$, this also implies stationarity of $K_\star$ for the standard $\mathcal{H}_2$ cost with the modified disturbance matrix $\tilde{E} = Ey$. Applying Theorem 2.3 to this modified system, we obtain

$$y^\top (M_K - M_{K_\star}) y = \|(G_K - G_{K_\star})y\|_{\mathcal{H}_2}^2 \geq 0, \quad \forall K \in \mathcal{K}_s.$$

Since y is arbitrary, $M_{K_\star} \preceq M_K$ for every $K \in \mathcal{K}_s$. $\square$

The preceding result can be viewed as a matrix-valued strengthening of standard $\mathcal{H}_2$ optimality. Indeed, the latter directly gives only $\text{tr}(M_{K_\star}) \leq \text{tr}(M_K)$, whereas Corollary 2.4 establishes the stronger Loewner-order relation $M_{K_\star} \preceq M_K$ for all $K \in \mathcal{K}_s$. In this regard, we note that the LQ-optimal value matrix X_{LQ} in (14) provides the common lower bound $X_{\text{LQ}} \preceq X_K$, which, after compression through the disturbance matrix E , gives $M_{\text{LQ}} := E^\top X_{\text{LQ}} E \preceq M_K$ for all $K \in \mathcal{K}_s$. The attainment of the standard $\mathcal{H}_2$ optimum then implies $M_{K_\star} = M_{\text{LQ}} = E^\top X_{\text{LQ}} E$, however, it does not require $X_{K_\star} = X_{\text{LQ}}$. When the disturbance matrix E is rank deficient, the standard $\mathcal{H}_2$ cost only measures the closed-loop performance along state directions excited through E . Hence, an $\mathcal{H}_2$ -optimal gain need not be LQ-optimal for every initial state and may therefore satisfy $X_{K_\star} \neq X_{\text{LQ}}$. Let us illustrate this through an example.

Example 2.5 (Singular disturbance controllability Gramian and nonunique optima). *Consider*

$$A = \begin{bmatrix} 0 & 0 \\ 0 & 2 \end{bmatrix}, \quad B = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \quad E = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad C = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad D = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

Let $K = [k_1 \ k_2]$ and set $\beta = 2 + k_2$. Then, $K \in \mathcal{K}_s$ if and only if $|\beta| < 1$, and $v^s(K) = 1 + \frac{k_1^2}{1-\beta^2}$. Thus, every $K_\beta = [0 \ \beta - 2]$ with $|\beta| < 1$ is stationary and globally optimal for the standard $\mathcal{H}_2$

problem. At these gains, $Y_{K_\beta}^s = \text{diag}(1, 0)$, $X_{K_\beta} = \text{diag}\left(1, \frac{1}{1-\beta^2}\right)$, and $S_{K_\beta} = \begin{bmatrix} 0 & \frac{\beta}{1-\beta^2} \end{bmatrix}$. In particular, $S_{K_\beta} \neq 0$ for $\beta \neq 0$. The LQ-optimal value matrix is $X_{\text{LQ}} = I_2$, attained by the gain $K_{\text{LQ}} = \begin{bmatrix} 0 & -2 \end{bmatrix}$. Observe that, for $\beta \neq 0$, $X_{K_\beta} \neq X_{\text{LQ}}$, however, $M_{K_\beta} = M_{K_{\text{LQ}}} = 1$. This illustrates both stationarity sufficiency with a singular disturbance Gramian and optimality only after compression.

3. Distributionally robust $\mathcal{H}_2$ synthesis

In this section, we formulate and study distributionally robust (DR) $\mathcal{H}_2$ synthesis by considering i.i.d. zero-mean disturbances with a generic covariance. We again analyze the regularity properties of the corresponding cost and show the *sufficiency* of the first-order stationarity condition for optimality. More importantly, we show that the standard $\mathcal{H}_2$ synthesis is inherently distributionally robust against this class of disturbances. The analysis is in particular enabled by the covariance representation of the performance cost of $\mathcal{H}_2$ synthesis, combined with the Loewner minimality in Corollary 2.4.

3.1. Generalized $\mathcal{H}_2$ cost with arbitrary disturbance covariance

The standard $\mathcal{H}_2$ cost can be alternatively expressed as the steady-state power of the performance signal under zero initial state and i.i.d. white noise disturbance with zero mean and unit covariance, that is,

$$v^s(K) = \lim_{t \rightarrow \infty} \mathbb{E} \left[\|z_t\|_2^2 \mid x_0 = 0, w_i \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, I) \right], \quad \forall K \in \mathcal{K}_s.$$

We now generalize the standard cost above by allowing the disturbance to have a generic zero-mean distribution with finite covariance. To that end, let us define the corresponding set of distributions

$$\mathcal{P} := \{P \in \mathcal{B}(\mathbf{R}^p) : \mu_P = 0, \|\Sigma_P\|_F < \infty\},$$

and the generalized $\mathcal{H}_2$ cost functional

$$v(K, P) := \lim_{t \rightarrow \infty} \mathbb{E} \left[\|z_t\|_2^2 \mid x_0 = 0, w_i \stackrel{\text{i.i.d.}}{\sim} P \right], \quad \forall K \in \mathcal{K}_s, \forall P \in \mathcal{P}. \quad (24)$$

Similar to the standard cost, the generalized cost can be computed algebraically using the observability Gramian.

Lemma 3.1 (Generalized cost v). *We have*

$$v(K, P) = \langle \Sigma_P, M_K \rangle, \quad \forall K \in \mathcal{K}_s, \forall P \in \mathcal{P}.$$

Proof. For the closed-loop system (2) with $x_0 = 0$ and driven by the disturbance w_t , we have

$$z_{t+1} = \sum_{i=0}^t C_K A_K^i E w_i, \quad \forall t \in \mathbf{N}_0.$$

Then, we can write for all $t \geq 1$,

$$\mathbb{E} \left[\|z_t\|_2^2 \mid x_0 = 0 \right] = \mathbb{E} \left[\left\| \sum_{i=0}^{t-1} C_K A_K^i E w_i \right\|_2^2 \right].$$

Now, we use the properties of the disturbance. First, using linearity of expectation and the fact that w_i 's are independent and zero-mean, we arrive at

$$\mathbb{E} \left[\left\| \sum_{i=0}^{t-1} C_K A_K^i E w_i \right\|_2^2 \right] = \sum_{i=0}^{t-1} \mathbb{E} \left[\|C_K A_K^i E w_i\|_2^2 \right].$$

Next, using the identity $y^\top M y = \langle y y^\top, M \rangle$ for vector y and matrix M , and the fact that w_i 's are identically distributed with distribution $\mathbf{P}$, we have $(H_{K,i} := E^\top (A_K^i)^\top C_K^\top C_K A_K^i E)$

$$\sum_{i=0}^{t-1} \mathbb{E} \left[\|C_K A_K^i E w_i\|_2^2 \right] = \sum_{i=0}^{t-1} \mathbb{E}_{\mathbf{P}} \left[w_i^\top H_{K,i} w_i \right] = \sum_{i=0}^{t-1} \mathbb{E}_{\mathbf{P}} \left[\langle w_i w_i^\top, H_{K,i} \rangle \right] = \left\langle \Sigma_{\mathbf{P}}, \sum_{i=0}^{t-1} H_{K,i} \right\rangle.$$

Finally, using the definition of the cost in (24) and the observability Gramian in (5), we obtain

$$v(K, \mathbf{P}) = \lim_{t \rightarrow \infty} \left\langle \Sigma_{\mathbf{P}}, E^\top \left(\sum_{i=0}^{t-1} (A_K^i)^\top C_K^\top C_K A_K^i E \right) E \right\rangle = \langle \Sigma_{\mathbf{P}}, E^\top X_K E \rangle.$$

This concludes the proof. $\square$

Lemma 3.1 reveals several key insights. Most importantly, by expressing the cost as the Frobenius inner product $\langle \Sigma_{\mathbf{P}}, M_K \rangle$, it separates the effect of disturbance statistics $\Sigma_{\mathbf{P}}$ from the dynamics encoded in M_K , which is a key element in the subsequent distributional robustness analysis. Moreover, it shows that the cost $v(\cdot, \mathbf{P})$ generalizes the standard cost v^s synthesis by allowing a generic covariance $\Sigma_{\mathbf{P}}$ for the disturbance. In particular, given $\mathbf{P} \in \mathcal{P}$, let $\tilde{v}^s$ be the standard cost for the dynamics matrices $(A, B, C, D, \tilde{E})$ with $\tilde{E} = E \Sigma_{\mathbf{P}}^{1/2}$, that is,

$$\tilde{v}^s(K) = \sum_{i=0}^{\infty} \|C_K A_K^i E \Sigma_{\mathbf{P}}^{1/2}\|_F^2.$$

Then, using the representation (6) for the standard cost, we have

$$v(K, \mathbf{P}) = \langle \Sigma_{\mathbf{P}}, M_K \rangle = \langle \Sigma_{\mathbf{P}}, E^\top X_K E \rangle = \langle I, (\Sigma_{\mathbf{P}}^{1/2} E)^\top X_K (E \Sigma_{\mathbf{P}}^{1/2}) \rangle = \tilde{v}^s(K), \quad \forall K \in \mathcal{K}_s. \quad (25)$$

This, in turn, implies that the results of the previous section for the standard $\mathcal{H}_2$ cost and problem can be extended to the generalized $\mathcal{H}_2$ cost (24) and the corresponding problem $\inf_{K \in \mathcal{K}_s} v(K, \mathbf{P})$ for every $\mathbf{P} \in \mathcal{P}$, by replacing the disturbance matrix E with $E \Sigma_{\mathbf{P}}^{1/2}$. In particular, we have the following results on smoothness and gradient of the generalized cost and the sufficiency of stationarity for optimality of stabilizing gains for the generalized cost.

Corollary 3.2 (Regularity of v). *Let $\mathbf{P} \in \mathcal{P}$. The nominal cost $v(\cdot, \mathbf{P})$ is real-analytic on $\mathcal{K}_s$. Also,*

$$\nabla_K v(K, \mathbf{P}) = 2 S_K Y_K^{\mathbf{P}}, \quad \forall K \in \mathcal{K}_s,$$

where $Y_K^{\mathbf{P}} := Y_K(\Sigma_{\mathbf{P}})$ is the solution to the Lyapunov equation (9) with $\Sigma = \Sigma_{\mathbf{P}}$.

Proof. The result follows from Lemma 2.1 and the equivalence (25). $\square$

Corollary 3.3 (Stationarity and optimality for v). *Let $\mathbf{P} \in \mathcal{P}$. We have $K_\star \in \operatorname{argmin}_{K \in \mathcal{K}_s} v(K, \mathbf{P})$ if and only if $\nabla_K v(K_\star, \mathbf{P}) = 0$. In fact, if $\nabla_K v(K_\star, \mathbf{P}) = 0$, then*

$$v(K, \mathbf{P}) - v(K_\star, \mathbf{P}) = \left\| (G_K - G_{K_\star}) \Sigma_{\mathbf{P}}^{1/2} \right\|_{\mathcal{H}_2}^2, \quad \forall K \in \mathcal{K}_s.$$

Proof. The result follows from Theorem 2.3 and the equivalence (25). $\square$

3.2. DR $\mathcal{H}_2$ cost

With the generalized $\mathcal{H}_2$ cost (24) at hand, we can now formulate the DR version of $\mathcal{H}_2$ problem. Given an ambiguity set $\mathcal{D}$, we define the DR $\mathcal{H}_2$ synthesis by

$$v_\star^{\mathcal{D}} := \inf_{K \in \mathcal{K}_s} \left\{ v^{\mathcal{D}}(K) := \sup_{\mathbf{P} \in \mathcal{D}} v(K, \mathbf{P}) \right\}. \quad (26)$$

Throughout this section, we assume that the ambiguity set $\mathcal{D}$ has the following properties.

Assumption 3.4 (Ambiguity). *$\mathcal{D}$ is a non-empty subset of $\mathcal{P}$ and independent of the gain K . Moreover, $\sup_{\mathbf{P} \in \mathcal{D}} \|\Sigma_{\mathbf{P}}\|_F < \infty$.*

We next consider the regularity properties of the DR cost $v^{\mathcal{D}}$. To characterize the regularity of the DR cost, it is useful to introduce the set of covariance matrices generated by the ambiguity set,

$$\mathfrak{C} := \{\Sigma_{\mathbf{P}} : \mathbf{P} \in \mathcal{D}\},$$

and its closed convex hull

$$\mathcal{C} := \overline{\text{conv}}(\mathfrak{C}).$$

Under Assumption 3.4, the set $\mathcal{C}$ is a nonempty compact subset of the set of p -by- p positive semi-definite matrices. Moreover, since the generalized cost is linear in the covariance matrix, Lemma 3.1 gives

$$v^{\mathcal{D}}(K) = \sup_{\mathbf{P} \in \mathcal{D}} \langle \Sigma_{\mathbf{P}}, M_K \rangle = \sup_{\Sigma \in \mathfrak{C}} \langle \Sigma, M_K \rangle = \max_{\Sigma \in \mathcal{C}} \langle \Sigma, M_K \rangle, \quad \forall K \in \mathcal{K}_s. \quad (27)$$

Thus, although the supremum in the definition of $v^{\mathcal{D}}$ need not be attained by a distribution in $\mathcal{D}$, the equivalent maximization problem (27) is always attained at the covariance level in $\mathcal{C}$. For $K \in \mathcal{K}_s$, define the corresponding set of active worst-case covariances by

$$\mathcal{C}_K := \operatorname{argmax}_{\Sigma \in \mathcal{C}} \langle \Sigma, M_K \rangle.$$

The set $\mathcal{C}_K$ is a nonempty compact convex set, namely, an exposed face of $\mathcal{C}$. We can now characterize the regularity and the Clarke subdifferential of the DR cost. We recall that $Y_K(\Sigma)$ is the solution to the Lyapunov equation (9).

Proposition 3.5 (Regularity of $v^{\mathcal{D}}$). *Let Assumption 3.4 hold. The DR cost $v^{\mathcal{D}}$ is finite, locally Lipschitz continuous, and Clarke regular on $\mathcal{K}_s$. Moreover, the Clarke subdifferential of $v^{\mathcal{D}}$ is*

$$\partial_{\mathcal{C}} v^{\mathcal{D}}(K) = \{2 S_K Y_K(\Sigma_K) : \Sigma_K \in \mathcal{C}_K\}, \quad \forall K \in \mathcal{K}_s. \quad (28)$$

In particular, if $\mathcal{C}_K = \{\Sigma_K\}$ is a singleton, then $v^{\mathcal{D}}$ is differentiable at K and

$$\nabla_K v^{\mathcal{D}}(K) = 2 S_K Y_K(\Sigma_K). \quad (29)$$

Proof. We first establish the representation of the DR cost as a pointwise maximum of smooth functions. For every $\Sigma \in \mathcal{C}$, define

$$v_\Sigma(K) := v(K, \mathcal{N}(0, \Sigma)) = \langle \Sigma, M_K \rangle, \quad K \in \mathcal{K}_s.$$

By Corollary 3.2, v_Σ is real-analytic on $\mathcal{K}_s$ and $\nabla_K v_\Sigma(K) = 2S_K Y_K(\Sigma)$. Moreover, notice that the index set $\mathcal{C}$ is compact and independent of K , and $(K, \Sigma) \mapsto v_\Sigma(K)$ and $(K, \Sigma) \mapsto \nabla_K v_\Sigma(K)$ are continuous on $\mathcal{K}_s \times \mathcal{C}$. By the equality (27), we have

$$v^D(K) = \max_{\Sigma \in \mathcal{C}} v_\Sigma(K). \quad (30)$$

That is, the DR cost is a pointwise maximum of smooth functions.

Let us first verify local Lipschitz continuity. Set $r_{\mathcal{D}} := \max_{\Sigma \in \mathcal{C}} \|\Sigma\|_F < \infty$. Then, for every $K_1, K_2 \in \mathcal{K}_s$, we have

$$|v^D(K_1) - v^D(K_2)| \leq \max_{\Sigma \in \mathcal{C}} |\langle \Sigma, M_{K_1} - M_{K_2} \rangle| \leq r_{\mathcal{D}} \|M_{K_1} - M_{K_2}\|_F.$$

Recall that $K \mapsto M_K$ is real-analytic and hence locally Lipschitz on $\mathcal{K}_s$. Thus, v^D is locally Lipschitz on $\mathcal{K}_s$.

We next apply the classical result on pointwise maxima of smooth functions; see, e.g., [9, Thm. 2.8.2] and [38, Thm. 10.31]. Since the maximizing set in (30) is compact and the parametrization is continuously differentiable in K , v^D is Clarke regular and

$$\partial_C v^D(K) = \text{conv} \{ \nabla_K v_{\Sigma_K}(K) = 2 S_K Y_K(\Sigma_K) : \Sigma_K \in \mathcal{C}_K \}.$$

Notice that the map $\Sigma_K \mapsto Y_K(\Sigma_K)$ is linear, which follows from the convergent series expression (8) of the controllability Gramian. Therefore, the map $\Sigma_K \mapsto 2S_K Y_K(\Sigma_K)$ is also linear. Then, since $\mathcal{C}_K$ is convex, its image under a linear map is already convex. Therefore, the convex hull in the preceding expression is redundant, and we obtain (28). Finally, if the active covariance is unique, $\mathcal{C}_K = \{\Sigma_K\}$, the Clarke subdifferential is the singleton $\partial_C v^D(K) = \{2S_K Y_K(\Sigma_K)\}$. The pointwise-maximum differentiation result then gives (29), completing the proof. $\square$

3.3. Properties of DR $\mathcal{H}_2$ -optimal solution

In this subsection, we extend the result of Theorem 2.3 for standard $\mathcal{H}_2$ problem to the DR $\mathcal{H}_2$ problem (26). To that end, observe that Proposition 3.5 allows us to define stationarity for the possibly nonsmooth DR cost in the standard sense of Clarke. Namely, a stabilizing gain $K \in \mathcal{K}_s$ is Clarke stationary if $0 \in \partial_C v^D(K)$. In general nonsmooth and nonconvex optimization, Clarke stationarity is only a necessary condition for local optimality and is not sufficient for optimality. However, similar to the standard $\mathcal{H}_2$ problem, stationarity is also sufficient for global optimality for the DR $\mathcal{H}_2$ problem considered here.

Theorem 3.6 (Stationarity and optimality for v^D). *Let Assumption 3.4 hold. Then, a stabilizing gain $K_\star \in \mathcal{K}_s$ is a solution of the DR $\mathcal{H}_2$ problem (26) if and only if $K_\star$ is Clarke stationary, that*

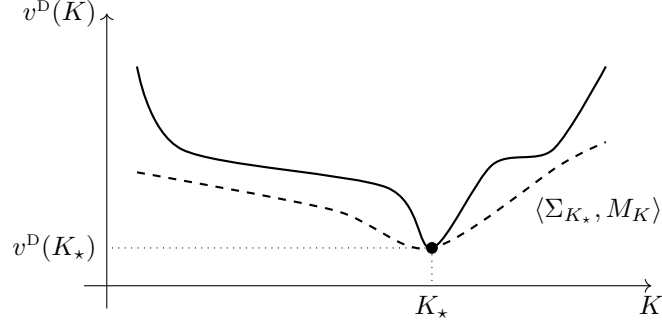


FIGURE 1. An intuitive presentation of Theorem 3.6. $K_* \in \mathcal{K}_s$ is a Clarke stationary point of the DR $\mathcal{H}_2$ cost $v^D(K) = \max_{\Sigma \in \mathcal{C}} \langle \Sigma, M_K \rangle$, and $\Sigma_{K_*} \in \operatorname{argmax}_{\Sigma \in \mathcal{C}} \langle \Sigma, M_{K_*} \rangle$ is a corresponding active covariance.

is, $0 \in \partial_C v^D(K_*)$. In fact, if $K_* \in \mathcal{K}_s$ is Clarke stationary, then there exists an active covariance $\Sigma_{K_*} \in \mathcal{C}_{K_*}$ such that

$$v^D(K) - v^D(K_*) \geq \left\| (G_K - G_{K_*}) \Sigma_{K_*}^{1/2} \right\|_{\mathcal{H}_2}^2, \quad \forall K \in \mathcal{K}_s. \quad (31)$$

Proof. We prove the two implications separately. Suppose first that $v_*^D = v^D(K_*)$. Since the stabilizing set $\mathcal{K}_s$ is open, K_* is an unconstrained local minimizer of v^D on an open neighborhood contained in $\mathcal{K}_s$. By Proposition 3.5, v^D is locally Lipschitz on $\mathcal{K}_s$. The Clarke necessary optimality condition [9, Prop. 2.3.2] therefore yields Clarke stationarity, that is, $0 \in \partial_C v^D(K_*)$. This proves necessity.

We next prove sufficiency. Let $K_* \in \mathcal{K}_s$ satisfy $0 \in \partial_C v^D(K_*)$. By the exact subdifferential representation (28), there exists $\Sigma_{K_*} \in \mathcal{C}_{K_*}$ such that $2S_{K_*} Y_{K_*}(\Sigma_{K_*}) = 0$. Now, let $P_{K_*} := \mathcal{N}(0, \Sigma_{K_*}) \in \mathcal{P}$ and consider the generalized $\mathcal{H}_2$ cost associated with P_{K_*} , that is, $v(K, P_{K_*}) = \langle \Sigma_{K_*}, M_K \rangle$. By Corollary 3.2, $\nabla_K v(K_*, P_{K_*}) = 2S_{K_*} Y_{K_*}(\Sigma_{K_*}) = 0$. We can therefore apply Corollary 3.3 to get

$$v(K, P_{K_*}) - v(K_*, P_{K_*}) = \left\| (G_K - G_{K_*}) \Sigma_{K_*}^{1/2} \right\|_{\mathcal{H}_2}^2, \quad \forall K \in \mathcal{K}_s. \quad (32)$$

On the other hand, since $\Sigma_{K_*} \in \mathcal{C}_{K_*}$, we have

$$v^D(K_*) = \langle \Sigma_{K_*}, M_{K_*} \rangle = v(K_*, P_{K_*}). \quad (33)$$

Furthermore, for an arbitrary $K \in \mathcal{K}_s$, it follows that

$$v^D(K) = \max_{\Sigma \in \mathcal{C}} \langle \Sigma, M_K \rangle \geq \langle \Sigma_{K_*}, M_K \rangle = v(K, P_{K_*}). \quad (34)$$

Combining (32), (33), and (34), we obtain (31), which also proves global optimality of K_* . $\square$

Theorem 3.6 provides an interpretation of Clarke stationarity in terms of an active worst-case covariance. Indeed, $0 \in \partial_C v^D(K_*)$ if and only if there exists $\Sigma_{K_*} \in \mathcal{C}_{K_*}$ such that K_* is stationary for the generalized cost $v(\cdot, \mathcal{N}(0, \Sigma_{K_*}))$. The special landscape of the $\mathcal{H}_2$ problem then implies that K_* is globally optimal for this generalized cost. Since Σ_{K_*} is simultaneously active for the DR cost at K_* , this nominal global optimality transfers directly to the DR problem. We can construct a *supporting-plane-type* lower bound for the DR cost in the form of $v^D(K) \geq \langle \Sigma_{K_*}, M_K \rangle$ for all $K \in \mathcal{K}_s$, with the bound being tight at $K = K_*$; see Figure 1 for a graphical illustration.

It is important that $\Sigma_{K_\star}$ need not, under Assumption 3.4 alone, be the covariance of an actual worst-case distribution in $\mathcal{D}$. Rather, it belongs to the closed convex hull of the covariance matrices generated by $\mathcal{D}$. No attainment assumption on the original distributional maximization is therefore required. If, in addition, $\mathfrak{C}$ is closed and convex, then $\mathcal{C} = \mathfrak{C}$ and the active covariance $\Sigma_{K_\star}$ is the covariance of an actual distribution $\mathbf{P}_{K_\star} \in \mathcal{D}$. In that case, $\mathbf{P}_{K_\star}$ can be chosen to be a worst-case distribution at $K_\star$, and $\nabla_K v(K_\star, \mathbf{P}_{K_\star}) = 0$.

3.4. Distributional robustness of standard $\mathcal{H}_2$ synthesis

We now focus on the connection between the solutions of the standard and DR $\mathcal{H}_2$ problems by showing the inherent distributionally robustness of standard $\mathcal{H}_2$ synthesis. The following result is a direct consequence of the characterization of the generalized cost in Lemma 3.1 and the minimality of the compressed observability Gramian $M_{K_\star}$ in Loewner order established in Corollary 2.4.

Theorem 3.7 (From standard to DR $\mathcal{H}_2$ optimality). *Let $K_\star \in \mathcal{K}_s$ be a solution of the standard $\mathcal{H}_2$ problem (3). Then,*

$$v(K_\star, \mathbf{P}) \leq v(K, \mathbf{P}), \quad \forall K \in \mathcal{K}_s, \forall \mathbf{P} \in \mathcal{P}. \quad (35)$$

In particular, under Assumption 3.4, we have $v_\star^\mathbf{D} = v^\mathbf{D}(K_\star)$, that is, the gain $K_\star$ is also a solution of the DR $\mathcal{H}_2$ problem (26).

Proof. By Corollary 2.4, for the standard $\mathcal{H}_2$ -optimal gain $K_\star$, we have $M_K - M_{K_\star} \succeq 0$ for all $K \in \mathcal{K}_s$. On the other hand, every distribution $\mathbf{P} \in \mathcal{P}$ has a positive semi-definite covariance matrix $\Sigma_{\mathbf{P}} \succeq 0$. As a result,

$$\langle \Sigma_{\mathbf{P}}, M_K - M_{K_\star} \rangle = \langle \Sigma_{\mathbf{P}}, M_K \rangle - \langle \Sigma_{\mathbf{P}}, M_{K_\star} \rangle \geq 0, \quad \forall K \in \mathcal{K}_s, \forall \mathbf{P} \in \mathcal{P}.$$

By Lemma 3.1, $v(K, \mathbf{P}) = \langle \Sigma_{\mathbf{P}}, M_K \rangle$ and hence we arrive at the pointwise inequality (35). Since (35) holds for every $\mathbf{P} \in \mathcal{D} \subset \mathcal{P}$, taking the supremum over the same ambiguity set $\mathcal{D}$ on both sides yields

$$v^\mathbf{D}(K_\star) = \sup_{\mathbf{P} \in \mathcal{D}} v(K_\star, \mathbf{P}) \leq \sup_{\mathbf{P} \in \mathcal{D}} v(K, \mathbf{P}) = v^\mathbf{D}(K), \quad \forall K \in \mathcal{K}_s.$$

The DR cost $v^\mathbf{D}$ is finite under Assumption 3.4. Therefore, the preceding inequality shows that $K_\star \in \operatorname{argmin}_{K \in \mathcal{K}_s} v^\mathbf{D}(K)$, which completes the proof. $\square$

The preceding result indicates that a solution $K_\star \in \mathcal{K}_s$ to the standard $\mathcal{H}_2$ problem (3) can be used for all the ambiguity sets satisfying Assumption 3.4. The ambiguity set changes its certified worst-case performance, even though a standard $\mathcal{H}_2$ -optimal gain remains optimal. The theorem gives an inclusion of optimizer sets; it does not, without additional assumptions, assert uniqueness or equality of the two optimizer sets. In Appendix A.2, we describe cost-preserving gain perturbations that can explain nonuniqueness of standard $\mathcal{H}_2$ -optimal gains. Moreover, in Appendix A.3, we provide a sufficient condition for a DR-optimal gain to also be standard $\mathcal{H}_2$ -optimal.

Recall that under the regular assumptions of Remark 2.2, a common optimal controller is already provided by classical LQR theory. In particular, the stabilizing LQ-optimal gain K_{LQ} depends only on (A, B, C, D) and not on E or the disturbance covariance. Most importantly, its value matrix satisfies $X_{K_{\text{LQ}}} = X_{\text{LQ}} \preceq X_K$ for all $K \in \mathcal{K}_s$, and hence $\langle \Sigma_{\mathbf{P}}, E^\top X_{K_{\text{LQ}}} E \rangle \leq \langle \Sigma_{\mathbf{P}}, E^\top X_K E \rangle$ for

every $\Sigma_P \succeq 0$. That is, K_{LQ} is optimal for every covariance, and consequently for every ambiguity set considered here, as also noted in [41]. This classic inherent distributional robustness of K_{LQ} can also be established using the gradient dominance property of the LQR objective studied in [18, 6, 46]. Theorem 3.7 however applies more generally to *every* attained standard $\mathcal{H}_2$ optimum, without imposing any particular assumption on the system matrices and, in particular, without requiring a nonsingular disturbance controllability Gramian $Y_{K_\star}^S$. As discussed above, a standard $\mathcal{H}_2$ -optimal gain $K_\star$ can satisfy $X_{K_\star} \neq X_{LQ}$, where X_{LQ} is the LQ-optimal value matrix in (14), while $E^\top X_{K_\star} E = E^\top X_{LQ} E$. Theorem 3.7 shows that this compressed equality, rather than equality of the full value matrices, is sufficient for the DR optimality.

We finish this section with a simple application of the preceding result for the “spectral” ambiguity set characterized by an upper bound $\Sigma_0 \succeq 0$ on the disturbance covariance.

Proposition 3.8 (Spectral DR $\mathcal{H}_2$). *Let $\Sigma_0 \succeq 0$ and consider the DR $\mathcal{H}_2$ problem (26) with*

$$\mathcal{D} = \{P \in \mathcal{P} : \Sigma_P \preceq \Sigma_0\}. \quad (36)$$

Then,

$$v^D(K) = \langle \Sigma_0, M_K \rangle, \quad \forall K \in \mathcal{K}_s.$$

Moreover, if $K_\star \in \mathcal{K}_s$ is a standard $\mathcal{H}_2$ -optimal solution, then $v_\star^D = \langle \Sigma_0, M_{K_\star} \rangle$ and any $P_\star \in \mathcal{D}$ with covariance $\Sigma_{P_\star} = \Sigma_0$ is a worst-case distribution in (26).

Proof. Fix $K \in \mathcal{K}_s$ and recall that $M_K \succeq 0$. By Lemma 3.1, we have

$$v^D(K) = \sup_{P \in \mathcal{D}} v(K, P) = \max_{0 \preceq \Sigma_P \preceq \Sigma_0} \langle \Sigma_P, M_K \rangle.$$

Let Σ_P be any feasible covariance and define $\Delta := \Sigma_0 - \Sigma_P \succeq 0$. Observe that

$$\langle \Sigma_0, M_K \rangle - \langle \Sigma_P, M_K \rangle = \langle \Sigma_0 - \Sigma_P, M_K \rangle = \langle \Delta, M_K \rangle \geq 0,$$

where we used the fact that $\Delta \succeq 0$ and $M_K \succeq 0$ for the last inequality. Hence, $\langle \Sigma_P, M_K \rangle \leq \langle \Sigma_0, M_K \rangle$ for every feasible Σ_P , and the maximum is attained at $\Sigma_{P_\star} = \Sigma_0$. Therefore,

$$v^D(K) = \langle \Sigma_0, M_K \rangle, \quad \forall K \in \mathcal{K}_s.$$

The result then follows from Theorem 3.7. $\square$

The preceding result indicates that the worst-case distribution is independent of the feedback gain for the spectral ambiguity characterized by (36). This is because $M_K \succeq 0$ for any gain K and hence the functional $\Sigma_P \mapsto v(K, \Sigma_P) = \langle \Sigma_P, M_K \rangle$ is monotone with respect to the Loewner order. As a result, the inner maximization in DR problem (26) always selects the extreme point $\Sigma_{P_\star} = \Sigma_0$, which is independent of K ; see Figure 2 for a graphical illustration. The simplification observed here is not universal and may break down for other ambiguity sets (or uncertainty models). In such cases, the adversary reacts to the controller through M_K and the worst-case distribution depends on K . This dependence affects performance evaluation, while Theorem 3.7 still provides a common optimal gain. The next section derives explicit worst-case performance certificates and identifies the corresponding covariances for three cases of ambiguity sets.

4. Worst-case performance certificates for specific ambiguity sets

Once a standard $\mathcal{H}_2$ -optimal stabilizing gain is available, Theorem 3.7 guarantees the DR optimality of this gain for ambiguity sets satisfying Assumption 3.4. The remaining task is then to evaluate its worst-case performance. In this section, we look at three cases of ambiguity sets for which the worst-case distribution depends on the (optimal) feedback gain. To be precise, we consider ambiguity sets based on the Frobenius ball around a nominal covariance and KL and W2 balls around a nominal zero-mean Gaussian distribution. The presented results quantify the performance guarantee as the ambiguity radius changes and identify adverse disturbance directions. In order to exclude the degenerate case of

$$v(K_\star, \mathcal{P}) = 0, \quad \forall \mathcal{P} \in \mathcal{P},$$

we consider the following assumption throughout this section.

Assumption 4.1 (Non-zero $M_{K_\star}$). *The standard $\mathcal{H}_2$ problem (3) has a solution $K_\star$ such that $M_{K_\star} \neq 0$.*

Let us first remark that Assumption 4.1 implies that

$$\text{tr}(M_K) = v^s(K) \geq v^s(K_\star) = \text{tr}(M_{K_\star}) > 0,$$

and hence

$$M_K \neq 0, \quad \forall K \in \mathcal{K}_s.$$

4.1. Frobenius ambiguity set

Let us define a moment-based ambiguity set as the Frobenius ball of radius $\epsilon \geq 0$ around a nominal covariance $\Sigma_0 \succeq 0$, that is,

$$\mathcal{D}^F(\epsilon, \Sigma_0) := \{\mathcal{P} \in \mathcal{P} : \|\Sigma_{\mathcal{P}} - \Sigma_0\|_F \leq \epsilon\}, \quad (37a)$$

and the corresponding DR $\mathcal{H}_2$ synthesis problem

$$v_\star^F := \min_{K \in \mathcal{K}_s} \left\{ v^F(K) := \sup_{\mathcal{P} \in \mathcal{D}^F(\epsilon, \Sigma_0)} v(K, \mathcal{P}) \right\}. \quad (37b)$$

Our first result formalizes the Frobenius DR cost v^F and the corresponding worst-case distribution as a function of the gain K .

Proposition 4.2 (Frobenius DR cost v^F). *Let Assumption 4.1 hold. Then, v^F is real-analytic on $\mathcal{K}_s$ and given by*

$$v^F(K) = \langle \Sigma_K^F, M_K \rangle, \quad \forall K \in \mathcal{K}_s, \quad (38a)$$

where

$$\Sigma_K^F := \Sigma_0 + \frac{\epsilon}{\|M_K\|_F} M_K. \quad (38b)$$

In particular, any $\mathcal{P}_K^F \in \mathcal{D}^F(\epsilon, \Sigma_0)$ with covariance $\Sigma_{\mathcal{P}_K^F} = \Sigma_K^F$ is a worst-case distribution in (37b), corresponding to the gain $K \in \mathcal{K}_s$.

Proof. We first derive the characterization (38). Observe that the set of covariance matrices generated by the ambiguity set $\mathcal{D}^F(\epsilon, \Sigma_0)$ is the closed and convex set

$$\mathcal{C}^F(\epsilon, \Sigma_0) = \{\Sigma : \Sigma \succeq 0, \|\Sigma - \Sigma_0\|_F \leq \epsilon\}.$$

Then, using the representation (27) of the DR $\mathcal{H}_2$ cost, for any fixed $K \in \mathcal{K}_s$, we have

$$v^F(K) = \max_{\Sigma \in \mathcal{C}^F(\epsilon, \Sigma_0)} \langle \Sigma, M_K \rangle.$$

Using the Cauchy-Schwarz inequality for the Frobenius inner product, the objective function in the preceding maximization can be upper bounded as follows

$$\begin{aligned} \langle \Sigma, M_K \rangle &= \langle \Sigma + \Sigma_0 - \Sigma_0, M_K \rangle = \langle \Sigma_0, M_K \rangle + \langle \Sigma - \Sigma_0, M_K \rangle \\ &\leq \langle \Sigma_0, M_K \rangle + \|\Sigma - \Sigma_0\|_F \cdot \|M_K\|_F. \end{aligned}$$

Therefore, for any feasible covariance Σ with $\|\Sigma - \Sigma_0\|_F \leq \epsilon$, we have

$$\langle \Sigma, M_K \rangle \leq \langle \Sigma_0, M_K \rangle + \epsilon \cdot \|M_K\|_F.$$

By Assumption 4.1, we have $M_K \neq 0$ and hence Σ_K^F is well-defined. Now, observe that $\Sigma_K^F \in \mathcal{C}^F(\epsilon, \Sigma_0)$ (since $\Sigma_0, M_K \succeq 0$ and $\|\Sigma_K^F - \Sigma_0\|_F = \epsilon$) and the upper bound in the preceding equation is attained by $\Sigma = \Sigma_K^F$. This completes the first part of the proof.

We next show that v^F is real-analytic on $\mathcal{K}_s$. We have previously shown that the map $K \mapsto M_K$ is real-analytic on $\mathcal{K}_s$; see the proof of Lemma 2.1. Then, since $M_K \neq 0$ for all $K \in \mathcal{K}_s$ by Assumption 4.1, the function v^F is real-analytic on $\mathcal{K}_s$. $\square$

For the Frobenius ambiguity set (37a), there is a nonlinear coupling between the controller gain and the disturbance covariance in the DR cost v^F . In particular, the worst-case covariance Σ_K^F is now a function of the gain K , breaking down the separation enjoyed in the spectral DR $\mathcal{H}_2$ framework; see Figure 2 for a graphical illustration. A similar coupling between the controller gain and the disturbance covariance will arise in the next two cases of ambiguity sets.

4.2. KL ambiguity set

We now consider a model of distributional uncertainty based on the KL divergence. Let us first introduce the KL divergence. For two probability distributions $\mathbf{Q}$ and $\mathbf{P}$ on $\mathbf{R}^p$, provided $\mathbf{Q} \ll \mathbf{P}$, that is, $\mathbf{Q}$ is absolutely continuous with respect to $\mathbf{P}$, the KL divergence reads as

$$\text{dist}^{\text{KL}}(\mathbf{Q} \parallel \mathbf{P}) := \int_{\mathbf{R}^p} \log\left(\frac{d\mathbf{Q}}{d\mathbf{P}}(w)\right) d\mathbf{Q}(w),$$

where $d\mathbf{Q}/d\mathbf{P}$ is the Radon-Nikodym derivative; see [30] and [15, Sec. 1.4]. By convention, $\text{dist}^{\text{KL}}(\mathbf{Q} \parallel \mathbf{P}) = +\infty$ if $\mathbf{Q} \not\ll \mathbf{P}$. Now, given a nominal zero-mean Gaussian distribution $\mathbf{P}_0 = \mathcal{N}(0, \Sigma_0)$ with $\Sigma_0 \succeq 0$ and a radius $\epsilon > 0$, let us consider the *KL ball*

$$\mathcal{D}^{\text{KL}}(\epsilon, \Sigma_0) := \{\mathbf{P} \in \mathcal{P} : \text{dist}^{\text{KL}}(\mathbf{P} \parallel \mathbf{P}_0) \leq \epsilon\}, \quad (39a)$$

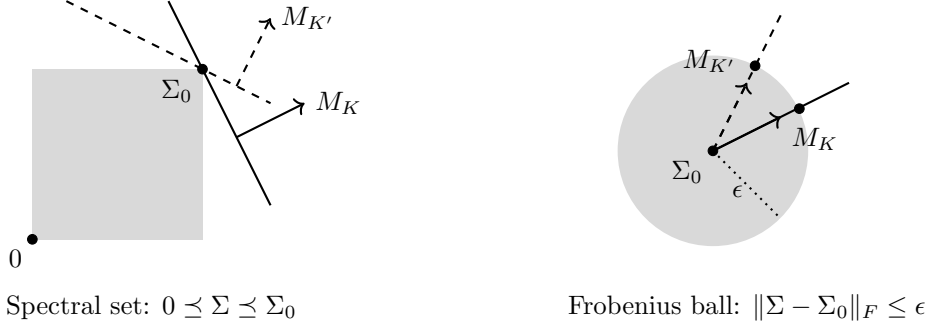


FIGURE 2. Possible dependence of the worst-case covariance on the gain. The gray areas show the feasible covariances for the two ambiguity sets. The arrows indicate the direction of the cost gradient $M_K = E^\top X_K E \succeq 0$, for two gains K and K' . (Left) For the spectral set (36), a worst-case covariance lies at the “upper right corner”, i.e., Σ_0 , regardless of the gradient (M_K vs $M_{K'}$) and hence the gain. (Right) For the Frobenius ball in (37a), the worst-case covariance lies at a point at the boundary of the ball, depending on the gradient and hence the gain.

as the ambiguity set and define the corresponding DR $\mathcal{H}_2$ synthesis problem

$$v_\star^{\text{KL}} := \min_{K \in \mathcal{K}_s} \left\{ v^{\text{KL}}(K) := \sup_{P \in \mathcal{D}^{\text{KL}}(\epsilon, \Sigma_0)} v(K, P) \right\}. \quad (39b)$$

We next provide an explicit formula for the DR cost v^{KL} assuming $\Sigma_0 \succ 0$. This is done by using the duality theory and the Donsker-Varadhan variational formula [1, Lem. 3.1] to reduce the inner maximization problem in (39b) to a *one-dimensional* convex minimization. In particular, this ambiguity set converts distributional robustness into an *entropic penalty* [1, Thm. 3.3]. To that end, let us define

$$\lambda_K^{\text{KL}} := \lambda_{\max}(\Sigma_0^{1/2} M_K \Sigma_0^{1/2}), \quad \forall K \in \mathcal{K}_s,$$

and

$$\phi^{\text{KL}}(K, \theta) := \epsilon\theta - \frac{\theta}{2} \log \det \left(I - \frac{2}{\theta} \Sigma_0^{1/2} M_K \Sigma_0^{1/2} \right), \quad \forall K \in \mathcal{K}_s, \quad \forall \theta \in (2\lambda_K^{\text{KL}}, \infty). \quad (40)$$

We have the following result for the preceding function.

Lemma 4.3 (Unique minimizer of ϕ^{KL}). *Let $\Sigma_0 \succ 0$ and Assumption 4.1 hold. For each $K \in \mathcal{K}_s$, the function $\phi^{\text{KL}}(K, \cdot)$ is strictly convex and coercive over the domain $\theta > 2\lambda_K^{\text{KL}}$, and hence has a unique minimizer*

$$\theta_K^{\text{KL}} := \underset{\theta}{\operatorname{argmin}} \{ \phi^{\text{KL}}(K, \theta) : \theta > 2\lambda_K^{\text{KL}} \}.$$

Proof. Let $\{\lambda_i\}_{i=1}^p$ be the eigenvalues of the matrix product $\Sigma_0^{1/2} M_K \Sigma_0^{1/2} \succeq 0$ (recall that $\Sigma_0 \succ 0$ by assumption and $M_K \succeq 0$ by construction). In particular, since $M_K \neq 0$ by Assumption 4.1, we have that $\lambda_K^{\text{KL}} = \max_{i \in \{1, \dots, p\}} \lambda_i > 0$. Then, it follows that

$$\phi(K, \theta) = \epsilon\theta - \frac{\theta}{2} \sum_{i=1}^p \log \left(1 - \frac{2\lambda_i}{\theta} \right).$$

and differentiating twice gives

$$\frac{\partial^2}{\partial \theta^2} \phi(K, \theta) = \sum_{i=1}^p \frac{2\lambda_i^2}{\theta(\theta - 2\lambda_i)^2} \geq \frac{2(\lambda_K^{\text{KL}})^2}{\theta(\theta - 2\lambda_K^{\text{KL}})^2} > 0, \quad \forall \theta \in (2\lambda_K^{\text{KL}}, \infty).$$

This shows that $\phi(K, \cdot)$ is strictly convex for $\theta > 2\lambda_K^{\text{KL}}$. As for coercivity, we clearly have $\phi(K, \theta) \uparrow \infty$ as $\theta \uparrow \infty$ since $\epsilon > 0$ by definition. On the other hand, observe that when $\theta \downarrow (2\lambda_K^{\text{KL}})$, the matrix $(I - \frac{2}{\theta}\Sigma_0^{1/2}M_K\Sigma_0^{1/2})$ loses positive definiteness with its log-det decreasing unboundedly and hence $\phi(K, \theta) \uparrow \infty$. This completes the proof. $\square$

With the preceding result in hand, we can now provide the explicit formula for the KL DR cost.

Proposition 4.4 (KL DR cost v^{KL}). *Let $\Sigma_0 \succ 0$ and Assumption 4.1 hold. Then, v^{KL} is real-analytic on $\mathcal{K}_s$ and given by*

$$v^{\text{KL}}(K) = \langle \Sigma_K^{\text{KL}}, M_K \rangle, \quad \forall K \in \mathcal{K}_s, \quad (41a)$$

where

$$\Sigma_K^{\text{KL}} := \left(\Sigma_0^{-1} - \frac{2}{\theta_K^{\text{KL}}} M_K \right)^{-1}. \quad (41b)$$

In particular, the worst-case distribution in (39b) corresponding to the gain $K \in \mathcal{K}_s$ is the zero-mean Gaussian $\mathbf{P}_K^{\text{KL}} = \mathcal{N}(0, \Sigma_{\mathbf{P}_K^{\text{KL}}})$ with covariance $\Sigma_{\mathbf{P}_K^{\text{KL}}} = \Sigma_K^{\text{KL}}$.

Proof. We first derive the characterization (41). Fix $K \in \mathcal{K}_s$. By Lemma 3.1, the generalized cost can be equivalently written as

$$v(K, \mathbf{P}) = \langle \Sigma_{\mathbf{P}}, M_K \rangle = \langle \mathbb{E}_{\mathbf{P}}[ww^\top], M_K \rangle = \mathbb{E}_{\mathbf{P}}[\langle ww^\top, M_K \rangle] = \mathbb{E}_{\mathbf{P}}[w^\top M_K w]. \quad (42)$$

Let us then define the performance map $g_K : \mathbb{R}^p \rightarrow \mathbb{R}$ by

$$g_K(w) := w^\top M_K w,$$

so that KL DR cost v^{KL} can be formulated as

$$v^{\text{KL}}(K) = \sup_{\mathbf{P} \in \mathcal{D}^{\text{KL}}(\epsilon, \Sigma_0)} \mathbb{E}_{\mathbf{P}}[g_K(w)]. \quad (43)$$

In what follows, we are going to expand the ambiguity set (by removing the restrictions on $\mathbf{P}$ to be zero-mean and have finite covariance) and consider the problem

$$\bar{v}^{\text{KL}}(K) := \sup_{\mathbf{P}} \{ \mathbb{E}_{\mathbf{P}}[g_K(w)] : \mathbf{P} \in \mathcal{B}(\mathbf{R}^p), \text{dist}^{\text{KL}}(\mathbf{P} \| \mathbf{P}_0) \leq \epsilon \}. \quad (44)$$

As we shall see, the optimal solution of the preceding “relaxed” problem (44) is a zero-mean Gaussian distribution belonging to the set $\mathcal{D}^{\text{KL}}(\epsilon, \Sigma_0) \subseteq \mathcal{P}$. Therefore, the optimal solution to the original problem (43) is the same zero-mean Gaussian distribution.

The problem (44) is a well-studied case in DRO. Let us define the domain

$$\Theta_K := \left\{ \theta > 0 : \mathbb{E}_{\mathbf{P}_0}[e^{g_K(w)/\theta}] < \infty \right\} \subset \mathbf{R}. \quad (45a)$$

If $\Theta_K \neq \emptyset$, we then have the dual representation [24, Thm. 1]

$$\bar{v}^{\text{KL}}(K) = \inf_{\theta \in \Theta_K} \left\{ \epsilon \theta + \theta \log \mathbb{E}_{\mathbf{P}_0} [e^{g_K(w)/\theta}] \right\}. \quad (45b)$$

Now, observe that since $\mathbf{P}_0 = \mathcal{N}(0, \Sigma_0)$ with $\Sigma_0 \succ 0$ by assumption and M_K is symmetric (and positive semi-definite) for $K \in \mathcal{K}_s$, we can use the results for the moment generating function of quadratic forms of normal random variables (see, e.g., [33, Cor. 3.2a1]) to write

$$\mathbb{E}_{\mathbf{P}_0} [e^{g_K(w)/\theta}] = \begin{cases} \det \left(I - \frac{2}{\theta} \Sigma_0^{1/2} M_K \Sigma_0^{1/2} \right)^{-1/2} & \text{if } \theta > 2\lambda_K^{\text{KL}}, \\ \infty & \text{otherwise.} \end{cases}$$

Then, using the fact that

$$\log \det \left(I - \frac{2}{\theta} \Sigma_0^{1/2} M_K \Sigma_0^{1/2} \right)^{-1/2} = -\frac{1}{2} \log \det \left(I - \frac{2}{\theta} \Sigma_0^{1/2} M_K \Sigma_0^{1/2} \right),$$

we have

$$\bar{v}^{\text{KL}}(K) = \inf \left\{ \phi^{\text{KL}}(K, \theta) : \theta > 2\lambda_K^{\text{KL}} \right\}.$$

By Lemma 4.3, the preceding problem admits a unique dual minimizer θ_K^{KL} . Moreover, the corresponding worst-case distribution $\mathbf{P}_K^{\text{KL}}$ in (44) is the exponential tilt of $\mathbf{P}_0$ characterized by the Radon-Nikodym derivative (see the proof of [24, Thm. 2])

$$\frac{d\mathbf{P}_K^{\text{KL}}}{d\mathbf{P}_0}(w) = \frac{e^{g_K(w)/\theta_K^{\text{KL}}}}{\mathbb{E}_{\mathbf{P}_0} [e^{g_K(w)/\theta_K^{\text{KL}}}]}$$

For the nominal Gaussian distribution $\mathbf{P}_0 = \mathcal{N}(0, \Sigma_0)$, the corresponding density is given by

$$p_0(w) = (2\pi)^{-p/2} \det(\Sigma_0)^{-1/2} \exp \left(-\frac{1}{2} w^\top \Sigma_0^{-1} w \right).$$

Tilting by $\exp(g_K(w)/\theta_K^{\text{KL}}) = \exp(w^\top M_K w / \theta_K^{\text{KL}})$ then gives the un-normalized worst-case density

$$p_K^{\text{KL}}(w) \propto \exp(w^\top M_K w / \theta_K^{\text{KL}}) p_0(w) \propto \exp \left(-\frac{1}{2} w^\top \left(\Sigma_0^{-1} - \frac{2}{\theta_K^{\text{KL}}} M_K \right) w \right) = \exp \left(-\frac{1}{2} w^\top (\Sigma_K^{\text{KL}})^{-1} w \right).$$

In particular, by Lemma 4.3, $\theta_K^{\text{KL}} > 2\lambda_K^{\text{KL}}$, and hence

$$(\Sigma_K^{\text{KL}})^{-1} = \Sigma_0^{-1/2} \left(I - \frac{2}{\theta_K^{\text{KL}}} \Sigma_0^{1/2} M_K \Sigma_0^{1/2} \right) \Sigma_0^{-1/2} \succ 0.$$

Therefore, the normalized tilted distribution, that is, the corresponding worst-case distribution for the relaxed problem (44), is the zero-mean Gaussian distribution $\mathbf{P}_K^{\text{KL}} = \mathcal{N}(0, \Sigma_{\mathbf{P}_K^{\text{KL}}})$ with $\Sigma_{\mathbf{P}_K^{\text{KL}}} = \Sigma_K^{\text{KL}}$. This, in turn, implies that $\mathbf{P}_K^{\text{KL}}$ is also the optimal solution of the original problem (43). Then, using the primal moment formulation (42), we have

$$v^{\text{KL}}(K) = v(K, \mathbf{P}_K^{\text{KL}}) = \langle \Sigma_K^{\text{KL}}, M_K \rangle.$$

We next show that v^{KL} is real-analytic on $\mathcal{K}_s$. The main ingredient is showing that the map $K \rightarrow \theta_K^{\text{KL}}$ is real-analytic on $\mathcal{K}_s$. To that end, let us denote $R_K := \Sigma_0^{1/2} M_K \Sigma_0^{1/2}$ and define the map

$$\Phi^{\text{KL}}(K, \theta) := \frac{\partial}{\partial \theta} \phi^{\text{KL}}(K, \theta) = \epsilon - \frac{1}{2} \log \det \left(I - \frac{2}{\theta} R_K \right) - \frac{1}{\theta} \left\langle \left(I - \frac{2}{\theta} R_K \right)^{-1}, R_K \right\rangle,$$

over the domain $(K, \theta) \in \mathcal{K}_s \times (2\lambda_K^{\text{KL}}, \infty)$. In particular, observe that $(I - \frac{2}{\theta}R_K) \succ 0$ over this domain since $\lambda_K^{\text{KL}} = \lambda_{\max}(R_K)$. Also, recall that M_K is a real-analytic function over $K \in \mathcal{K}_s$; see the proof of Lemma 2.1. Then, since matrix inversion and log-det are real-analytic on the space of positive definite matrices, the map $(K, \theta) \mapsto \Phi^{\text{KL}}(K, \theta)$ is real-analytic on its domain.

Now, fix $K \in \mathcal{K}_s$. By Lemma 4.3, θ_K^{KL} is the unique minimizer of the strictly convex and coercive function $\phi^{\text{KL}}(K, \cdot)$, and hence, we have $\Phi^{\text{KL}}(K, \theta_K^{\text{KL}}) = 0$ and $\frac{\partial}{\partial \theta} \Phi^{\text{KL}}(K, \theta_K^{\text{KL}}) > 0$. Therefore, by the real-analytic implicit function theorem (see, e.g., [29, Thm. 2.4.4]), there exists a neighborhood $\mathcal{K} \subset \mathcal{K}_s$, an interval $\mathcal{I}$ containing θ_K^{KL} , and a real-analytic map $\psi_K : \mathcal{K} \rightarrow \mathcal{I}$ such that

$$\Phi^{\text{KL}}(\tilde{K}, \psi_K(\tilde{K})) = 0, \quad \forall \tilde{K} \in \mathcal{K}. \quad (46)$$

Moreover, since $\phi^{\text{KL}}(K, \cdot)$ is strictly convex and coercive and hence has unique global minimizer, the equation $\Phi^{\text{KL}}(\tilde{K}, \theta) = 0$ has exactly one solution, namely, $\theta_{\tilde{K}}$. Therefore, the implicit solution in (46) must coincide with this global minimizer, that is, $\psi_K(\tilde{K}) = \theta_{\tilde{K}}$ for all $\tilde{K} \in \mathcal{K}$. Thus, $\theta_{\tilde{K}}$ is a real-analytic function of $\tilde{K}$ in a neighborhood of every point $K \in \mathcal{K}_s$. Hence, it follows that the map $K \mapsto \theta_K^{\text{KL}}$ is real-analytic over the entire open set $\mathcal{K}_s$ of stabilizing controllers.

Finally, recall that $v^{\text{KL}}(K) = \langle \Sigma_K^{\text{KL}}, M_K \rangle$ for all $K \in \mathcal{K}_s$, where

$$\Sigma_K^{\text{KL}} = \Sigma_0^{1/2} \left(I - \frac{2}{\theta_K^{\text{KL}}} R_K \right)^{-1} \Sigma_0^{1/2}.$$

Recall that M_K and θ_K^{KL} are real-analytic on $\mathcal{K}_s$ and $(I - \frac{2}{\theta_K^{\text{KL}}} R_K) \succ 0$ for all $K \in \mathcal{K}_s$. Hence, since matrix inversion is real-analytic on the space of positive definite matrices, Σ_K^{KL} is also a real-analytic function of $K \in \mathcal{K}_s$. This, in turn, implies that v^{KL} is real-analytic on $\mathcal{K}_s$. $\square$

4.3. W2 ambiguity set

We now focus on ambiguity sets defined based on the W2 distance. Let us first recall the definition of the Wasserstein distance. For two distributions $\mathbf{Q}$ and $\mathbf{P}$ on $\mathbf{R}^p$, let $\Pi(\mathbf{Q}, \mathbf{P})$ denote the set of all probability measures π on $\mathbf{R}^p \times \mathbf{R}^p$ with marginals $\mathbf{Q}$ and $\mathbf{P}$ (a.k.a. couplings or transport plans), that is,

$$\pi(A \times \mathbf{R}^p) = \mathbf{Q}(A) \quad \text{and} \quad \pi(\mathbf{R}^p \times B) = \mathbf{P}(B), \quad \text{for all Borel } A, B \subset \mathbf{R}^p.$$

The Wasserstein-2 (W2) distance is defined by (see [44, Defs. 6.1 and 6.4])

$$\text{dist}^{\text{W}}(\mathbf{Q}, \mathbf{P}) := \left(\inf_{\pi \in \Pi(\mathbf{Q}, \mathbf{P})} \int_{\mathbf{R}^p \times \mathbf{R}^p} \|w - \tilde{w}\|_2^2 d\pi(w, \tilde{w}) \right)^{1/2}.$$

Notice that $\text{dist}^{\text{W}}(\mathbf{Q}, \mathbf{P}) < \infty$ if $\mathbf{Q}$ and $\mathbf{P}$ have a finite second moment. Given the nominal zero-mean Gaussian distribution $\mathbf{P}_0 = \mathcal{N}(0, \Sigma_0)$ with $\Sigma_0 \succeq 0$ and a radius $\epsilon > 0$, we now consider the *W2 ball*

$$\mathcal{D}^{\text{W}}(\epsilon, \Sigma_0) := \{\mathbf{P} \in \mathcal{P} : \text{dist}^{\text{W}}(\mathbf{P}, \mathbf{P}_0) \leq \epsilon\}, \quad (47a)$$

as the ambiguity set and define the corresponding DR $\mathcal{H}_2$ synthesis problem

$$v_{\star}^{\text{W}} := \min_{K \in \mathcal{K}_s} \left\{ v^{\text{W}}(K) := \sup_{\mathbf{P} \in \mathcal{D}^{\text{W}}(\epsilon, \Sigma_0)} v(K, \mathbf{P}) \right\}. \quad (47b)$$

We next provide an explicit formula for the W2 DR cost v^w . Similar to the KL case discussed above, this is done by using the duality theory (to be precise, Kantorovich duality for quadratic transport cost [45, Ch. 1]). Thus, let us define

$$\lambda_K^w := \lambda_{\max}(M_K), \quad \forall K \in \mathcal{K}_s,$$

and

$$\phi^w(K, \theta) := \epsilon^2 \theta + \text{tr}(\theta M_K (\theta I - M_K)^{-1} \Sigma_0), \quad \forall K \in \mathcal{K}_s, \forall \theta \in (\lambda_K^w, \infty).$$

We have the following result for the preceding function.

Lemma 4.5 (Unique minimizer of ϕ^w). *Let $\Sigma_0 \succ 0$ and Assumption 4.1 hold. For each $K \in \mathcal{K}_s$, the function $\phi^w(K, \cdot)$ is strictly convex and coercive over the domain $\theta > \lambda_K^w$, and hence has a unique minimizer*

$$\theta_K^w := \underset{\theta}{\operatorname{argmin}} \{ \phi^w(K, \theta) : \theta > \lambda_K^w \}.$$

Proof. Let $\{\lambda_i\}_{i=1}^p$ be the eigenvalues of $M_K \succeq 0$ (recall that M_K is a symmetric and positive semi-definite by construction). In particular, notice that since $M_K \neq 0$ by Assumption 4.1, we have that $\lambda_K^w = \max_{i \in \{1, \dots, p\}} \lambda_i > 0$. Diagonalize $M_K = U \operatorname{diag}(\lambda_1, \dots, \lambda_p) U^\top$ and denote the i -th diagonal entry of $U^\top \Sigma_0 U$ by μ_i . Notice that since $\Sigma_0 \succ 0$ by assumption, each diagonal entry μ_i is strictly positive. Also, let $\mu_K^w = \mu_j > 0$ for $j \in \operatorname{argmax}_i \lambda_i$. Then, it follows that

$$\phi^w(K, \theta) = \epsilon^2 \theta + \sum_{i=1}^p \frac{\mu_i \lambda_i \theta}{\theta - \lambda_i},$$

and differentiating twice gives

$$\frac{\partial^2}{\partial \theta^2} \phi^w(K, \theta) = \sum_{i=1}^p \frac{2\mu_i \lambda_i^2}{(\theta - \lambda_i)^3} \geq \frac{2\mu_K^w (\lambda_K^w)^2}{(\theta - \lambda_K^w)^3} > 0, \quad \forall \theta \in (\lambda_K^w, \infty).$$

Hence, $\phi^w(K, \cdot)$ is strictly convex over the domain $\theta > \lambda_K^w$. Also, $\phi^w(K, \theta) \uparrow \infty$ as $\theta \downarrow \lambda_K^w$ and as $\theta \uparrow \infty$; that is, $\phi^w(K, \cdot)$ is coercive on its domain. Thus, the minimizer θ_K^w is unique. $\square$

Using the preceding definitions, we can now provide the W2 DR cost explicitly.

Proposition 4.6 (W2 DR cost v^w). *Let $\Sigma_0 \succ 0$ and Assumption 4.1 hold. Then, v^w is real-analytic on $\mathcal{K}_s$ and given by*

$$v^w(K) = \langle \Sigma_K^w, M_K \rangle, \quad \forall K \in \mathcal{K}_s, \tag{48a}$$

where

$$\Sigma_K^w := (\theta_K^w)^2 (\theta_K^w I - M_K)^{-1} \Sigma_0 (\theta_K^w I - M_K)^{-1}. \tag{48b}$$

In particular, the worst-case distribution in (47b) corresponding to the gain $K \in \mathcal{K}_s$ is the zero-mean Gaussian $\mathbf{P}_K^w = \mathcal{N}(0, \Sigma_{\mathbf{P}_K^w})$ with covariance $\Sigma_{\mathbf{P}_K^w} = \Sigma_K^w$.

Proof. We first derive the characterization (48). Fix $K \in \mathcal{K}_s$. Recall that, using Lemma 3.1 and the quadratic performance map $g_K : \mathbb{R}^p \rightarrow \mathbb{R} : w \mapsto w^\top M_K w$, the W_2 DR cost v^w can be rewritten as

$$v^w(K) = \sup_{P \in \mathcal{D}^w(\epsilon, \Sigma_0)} \mathbb{E}_P[g_K(w)]. \quad (49)$$

In what follows, we are going to expand the ambiguity set (by removing the restrictions on P to be zero-mean) and consider the problem

$$\bar{v}^w(K) := \sup_P \{ \mathbb{E}_P[g_K(w)] : P \in \mathcal{B}(\mathbb{R}^p), \text{dist}^w(P \| P_0) \leq \epsilon \}. \quad (50)$$

As we shall see, the optimal solution of the preceding “relaxed” problem (50) is a zero-mean Gaussian distribution belonging to the set $\mathcal{D}^w(\epsilon, \Sigma_0) \subseteq \mathcal{P}$. Therefore, the optimal solution to the original problem (49) is the same zero-mean Gaussian distribution.

The preceding problem (50) is also a well-studied case in DRO. In particular, via the Kantorovich duality approach, we have (see, e.g., [3, Thm. 1 and Rem. 1])

$$\bar{v}^w(K) = \inf_{\theta \geq 0} \left\{ \epsilon^2 \theta + \mathbb{E}_{P_0} \left[\sup_{z \in \mathbb{R}^p} \{ g_K(z) - \theta \|z - w\|^2 \} \right] \right\}.$$

For fixed w , the objective of the inner maximization is finite precisely when $\theta I - M_K \succ 0$, i.e., $\theta > \lambda_K^w$. In that case, the objective is a strictly concave quadratic function in z , and the unique maximizer is $z_K(\theta, w) = \theta(\theta I - M_K)^{-1}w$. Substitution then gives the optimal value

$$\begin{aligned} \sup_z \{ z^\top M_K z - \theta \|z - w\|^2 \} &= w^\top (\theta^2(\theta I - M_K)^{-1} - \theta I) w \\ &= w^\top \theta M_K (\theta I - M_K)^{-1} w. \end{aligned}$$

Since $P_0 = \mathcal{N}(0, \Sigma_0)$, taking expectation yields

$$\bar{v}^w(K) = \min_{\theta} \{ \phi^w(K, \theta) : \theta > \lambda_K^w \}.$$

By Lemma 4.5, the preceding problem admits a unique dual minimizer θ_K^w . Moreover, the worst-case distribution can be characterized uniquely by the linear map [3, Rem. 2 and Rem. 3]

$$T_K(w) = z_K(\theta_K^w, w) = \theta_K^w(\theta_K^w I - M_K)^{-1}w.$$

That is, the worst-case distribution in the relaxed problem (50) is the pushforward $P_K^w = (T_K)_\# P_0$, which is the zero-mean Gaussian with covariance Σ_K^w as in (48b). This, in turn, implies that P_K^w is also the optimal solution of the original problem (43). Finally, using the primal moment formulation, we have

$$v^w(K) = v(K, P_K^w) = \langle \Sigma_K^w, M_K \rangle.$$

We next show that v^w is real-analytic on $\mathcal{K}_s$. The main ingredient is again showing that $K \mapsto \theta_K^w$ is real-analytic. Consider

$$\Phi^w(K, \theta) := \frac{\partial}{\partial \theta} \phi^w(K, \theta) = \epsilon^2 - \text{tr}(M_K(\theta I - M_K)^{-1} \Sigma_0 M_K(\theta I - M_K)^{-1}),$$

on the open set $(K, \theta) \in \mathcal{K}_s \times (\lambda_K^w, \infty)$. Recall that M_K is a real-analytic function on $K \in \mathcal{K}_s$; see the proof of Lemma 2.1. Then, since $\theta I - M_K \succ 0$ for $\theta > \lambda_K^w = \lambda_{\max}(M_K)$ and matrix inversion

is real-analytic on the space of positive definite matrices, Φ^w is real-analytic. By Lemma 4.5, $\Phi^w(K, \theta_K^w) = 0$ and $\frac{\partial}{\partial \theta} \Phi^w(K, \theta_K^w) > 0$. The real-analytic implicit function theorem (see, e.g., [29, Thm. 2.4.4]) therefore gives a local real-analytic representation of θ_K^w around every $K \in \mathcal{K}_s$. Since the scalar minimizer is unique, these local representations agree on overlaps; hence $K \mapsto \theta_K^w$ is real-analytic on $\mathcal{K}_s$. It follows from (48b) that $K \mapsto \Sigma_K^w$ is real-analytic. Thus, $v^w(K) = \langle \Sigma_K^w, M_K \rangle$ is real-analytic on $\mathcal{K}_s$. $\square$

4.4. DR optimality of standard $\mathcal{H}_2$ -optimal gain

The three ambiguity sets satisfy Assumption 3.4. They contain the nominal distribution and are independent of K . For the Frobenius set, $\|\Sigma_P\|_F \leq \|\Sigma_0\|_F + \epsilon$. For the KL set, we can use the relative-entropy variational inequality in [24, Sec. 2.1] with $f(w) = \|w\|_2^2/\theta$ and any fixed $\theta > 2\lambda_{\max}(\Sigma_0)$ to obtain

$$\sup_{P \in \mathcal{D}^{\text{KL}}(\epsilon, \Sigma_0)} \mathbb{E}_P[\|w\|_2^2] \leq \epsilon\theta - \frac{\theta}{2} \log \det \left(I - \frac{2}{\theta} \Sigma_0 \right) < \infty.$$

For the W2 set, the triangle inequality [44, Ch. 6], with the Dirac measure D_0 concentrated entirely at the origin, gives

$$(\mathbb{E}_P[\|w\|_2^2])^{1/2} = \text{dist}^w(P, D_0) \leq \text{dist}^w(P, P_0) + \text{dist}^w(P_0, D_0) = \epsilon + \sqrt{\text{tr}(\Sigma_0)}.$$

Since the distributions are centered, $\|\Sigma_P\|_F \leq \text{tr}(\Sigma_P) = \mathbb{E}_P[\|w\|_2^2]$, which establishes the required uniform covariance bounds. The explicit representation of the DR costs for these cases then allows us to provide the following result. Therefore, the results of the two previous sections also hold for the Frobenius DR cost and problem of this subsection. To be precise, for the DR $\mathcal{H}_2$ problems (37), (39), and (47) with the corresponding ambiguity sets, we have the following results with $a \in \{F, \text{KL}, W\}$:

- By Proposition 3.5, the gradient of the corresponding DR cost is given by

$$\nabla_K v^a(K) = 2S_K Y_K^a, \quad \forall K \in \mathcal{K}_s, \quad (51)$$

where $Y_K^a := Y_K(\Sigma_K^a)$ is the solution to the Lyapunov equation (9) with $\Sigma = \Sigma_K^a$.

- By Theorem 3.6, a stabilizing gain $K_\star \in \mathcal{K}_s$ is a solution of the corresponding DR $\mathcal{H}_2$ problem if and only if $\nabla_K v^a(K_\star) = 0$.
- By Theorem 3.7, if $K_\star \in \mathcal{K}_s$ is a standard $\mathcal{H}_2$ -optimal gain, then $K_\star$ is also a solution of the DR $\mathcal{H}_2$ problems with the corresponding worst-case covariance $\Sigma_{K_\star}^a$ and optimal value

$$v_\star^a = \langle \Sigma_{K_\star}^a, M_{K_\star} \rangle. \quad (52)$$

The last item above once again indicates that the standard $\mathcal{H}_2$ -optimal solution, if it exists, also solves the DR $\mathcal{H}_2$ problems considered in this subsection. The difference between the three cases is however in the corresponding worst-case covariance and optimal DR cost. For a discussion on the differences between the three choices of the uncertainty models considered above see Appendix A.1.

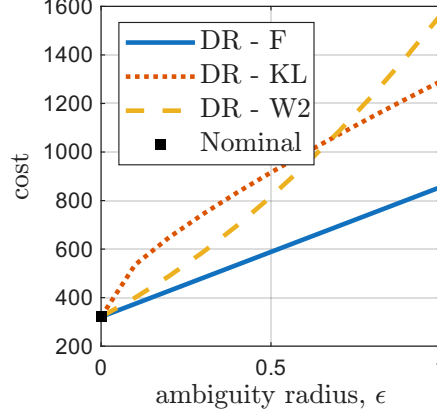


FIGURE 3. Sensitivity of the DR costs to ambiguity radius.

4.5. Numerical illustration of performance certification

In this subsection, we use a synthetic numerical example to illustrate the worst-case performance certificates of Section 4 for a controller obtained from standard $\mathcal{H}_2$ synthesis. To this end, we consider a system with $n = 10$ states, $m = 3$ inputs, and $p = 5$ disturbance channels. The state matrix A is randomly generated and normalized to be unstable with the spectral radius $\rho(A) = 2$; the input matrix B is randomly generated such that the pair (A, B) is controllable; the disturbance matrix is E randomly generated; the performance matrices C and D are set such that $z_t = [x_t^\top \ u_t^\top]^\top$. A solution to the standard $\mathcal{H}_2$ synthesis problem (3) gain for this problem is then the LQ-optimal gain K_{LQ} for the deterministic dynamics $x_{t+1} = Ax_t + Bu_t$ and state- and input-cost weighting matrices $Q = I$ and $R = I$ respectively; see Remark 2.2. We note that $\rho(A_{K_{LQ}}) = 0.53$.

Using the results of Section 4.4, we can now verify if this standard $\mathcal{H}_2$ -optimal gain solves the DR $\mathcal{H}_2$ problems corresponding the Frobenius, KL, and W2 ambiguity sets. To this end, we consider the DR $\mathcal{H}_2$ problems (37), (39), and, (47), with radius $\epsilon \in (0, 1]$ and nominal covariance $\Sigma_0 = (LL^\top)/\rho(LL^\top)$, where $L \in \mathbf{R}^{p \times p}$ is a randomly generated full-rank matrix. Computing the gradient of the corresponding standard and DR costs at K_{LQ} using (51) resulted in $\|\nabla f(K_{LQ})\|_F \leq 10^{-10}$ for all $f \in \{v^S, v^F, v^{KL}, v^W\}$. That is, the gain K_{LQ} is also a stationary point of the DR costs, and hence, by Theorem 3.6, solves the corresponding DR $\mathcal{H}_2$ problems. We note that for the KL and W2 ambiguity sets, the optimal dual variables $\theta_{K_\star}^{KL}$ and $\theta_{K_\star}^W$ in Lemmas 4.3 and 4.5, respectively, are computed using bisection. Figure 3 depicts the optimal value of these problems, computed using (52), for different values of the radius $\epsilon \in (0, 1]$. As expected, the Frobenius DR cost shows a linear sensitivity. The KL DR cost has higher sensitivity for smaller values of ϵ , while that is the case for the W2 DR cost for larger values of ϵ .

5. Final Remarks

In this paper, we considered the distributionally robust extension of the $\mathcal{H}_2$ synthesis problem for discrete-time LTI systems. We first established global landscape properties for the standard problem

and then used the covariance representation of the generalized cost to analyze ambiguity sets with zero mean and uniformly bounded covariances. The resulting DR objective may be nonsmooth, however it is locally Lipschitz and Clarke regular, and Clarke stationarity is both necessary and sufficient for global optimality. We also showed that any attained standard $\mathcal{H}_2$ optimum is simultaneously optimal for every such DR problem. Moreover, we derived explicit worst-case models for spectral, Frobenius, KL, and W2 ambiguity sets. We conclude with several remarks on limitations of these results.

Output-feedback synthesis. The analysis relies essentially on static state feedback, $u_t = Kx_t$. For static output feedback, $u_t = Ky_t$, the closed-loop matrix has the form $A + BKC_y$, and the corresponding stabilizing set is generally nonconvex and may be *disconnected* [19]. Moreover, output-feedback policy optimization can exhibit spurious stationary points, including nonglobal local minima and saddle points [14]. The trajectory-orthogonality and matrix-minimality arguments used here therefore do not transfer directly. Identifying additional structural conditions under which a componentwise or global benign-landscape result survives for static or dynamic output feedback remains an important open problem.

Other uncertainty models. Our distributional model assumes i.i.d. zero-mean disturbances with finite covariance. Under these assumptions, steady-state quadratic performance depends on the disturbance distribution only through a single covariance matrix, which is precisely what enables the universal optimality result. The Frobenius, KL, and Wasserstein models considered here represent substantially different geometries. Nonetheless, all preserve this covariance reduction. Natural extensions include nonzero or uncertain means, temporally correlated and nonstationary disturbances, ambiguity sets for entire disturbance processes, and simultaneous uncertainty in the system matrices. In these settings the cost need not reduce to $\langle \Sigma, M_K \rangle$, and a standard $\mathcal{H}_2$ optimizer need not remain distributionally robust. Another important direction is the statistical calibration of the ambiguity radius from finite trajectory data, which would connect the synthesis guarantees developed here to out-of-sample performance guarantees.

Restrictions on the controller. The common-optimizer result holds over the full stabilizing state-feedback set $\mathcal{K}_s$. Additional restrictions, such as sparsity or gain bounds, may exclude all unrestricted standard $\mathcal{H}_2$ optimizers. In this case, minimizing the nominal scalar cost over the restricted set need not minimize M_K in the Loewner order. Feasible gains may therefore trade performance across disturbance directions, and the worst-case covariance can change their ranking. Thus, the restricted nominal and DR problems need not share an optimizer, although a difference is not guaranteed. Conversely, any unrestricted standard optimizer that remains feasible retains its matrix-minimality property and hence remains optimal for both restricted problems.

Appendix A. Auxiliary results

A.1. Comparison of the three ambiguity sets

In what follows, we provide a brief discussion on the differences between the three choices of the uncertainty models discussed in Section 4.

Recall that the three worst-case covariance matrices admit the following representations:

$$\Sigma_K^F = \Sigma_0 + \frac{\epsilon}{\|M_K\|_F} M_K, \quad (53a)$$

$$\Sigma_K^{KL} = \Sigma_0^{1/2} (I - 2(\theta_K^{KL})^{-1} R_K)^{-1} \Sigma_0^{1/2}, \quad R_K := \Sigma_0^{1/2} M_K \Sigma_0^{1/2}, \quad (53b)$$

$$\Sigma_K^W = T_K \Sigma_0 T_K^\top, \quad T_K := (I - (\theta_K^W)^{-1} M_K)^{-1}, \quad (53c)$$

Let us first notice that in all three cases, the DR constraint is active and the worst-case distribution lies at the boundary of ambiguity set (under Assumption 4.1). In particular, we have

- $\|\Sigma_K^F - \Sigma_0\|_F = \epsilon$ for the Frobenius ambiguity set in Proposition 4.2,
- $\text{dist}^{KL}(\mathbf{P}_K^{KL}, \mathbf{P}_0) = \epsilon$ for the KL ambiguity set in Proposition 4.4, and,
- $\text{dist}^W(\mathbf{P}_K^W, \mathbf{P}_0) = \epsilon$ for the W2 ambiguity set in Proposition 4.6.

The Frobenius adversary performs an additive displacement of the nominal covariance Σ_0 in the direction M_K . Indeed, $M_K/\|M_K\|_F$ is the steepest-ascent direction of the linear functional $\Sigma \mapsto \langle \Sigma, M_K \rangle$ with respect to the Frobenius geometry. Notice that the direction of this displacement does not depend on Σ_0 . Since the cost depends only on the covariance, every zero-mean distribution with covariance Σ_K^F is worst case, e.g., the Gaussian representative $\mathbf{P}_K^F = \mathcal{N}(0, \Sigma_K^F)$.

For the KL ambiguity set, the worst-case distribution is uniquely obtained by exponentially tilting the nominal Gaussian distribution. The worst-case distribution remains Gaussian and satisfies $\mathbf{P}_K^{KL} = \mathcal{N}(0, \Sigma_K^{KL})$. Observe that the covariance increment in this case can be written as

$$\Sigma_K^{KL} - \Sigma_0 = \Sigma_0^{1/2} \left[(I - 2(\theta_K^{KL})^{-1} R_K)^{-1} - I \right] \Sigma_0^{1/2} \succeq 0.$$

Thus, the KL adversary produces a nonlinear multiplicative inflation in whitened coordinates. The relevant directions are the eigendirections of R_K , which jointly account for the performance sensitivity M_K and the nominal covariance Σ_0 .

For the W2 ambiguity set, the worst-case distribution is the pushforward of the nominal distribution through the linear transport map T_K , that is, $\mathbf{P}_K^W = (T_K)_\# \mathbf{P}_0 = \mathcal{N}(0, \Sigma_K^W)$. The Wasserstein perturbation is therefore fundamentally different from the Frobenius displacement and the KL exponential tilt: It acts on disturbance realizations before forming their covariance. In particular, the difference $\Sigma_K^W - \Sigma_0$ need not be positive definite when M_K and Σ_0 do not commute. Thus, Wasserstein transport may both rotate the nominal covariance ellipsoid and redistribute variance among its directions, while increasing the value $\langle \Sigma, M_K \rangle$.

The distinction between the three ambiguity sets becomes particularly transparent when M_K and Σ_0 commute. Suppose that M_K and Σ_0 are simultaneously diagonalized as

$$M_K = U \text{diag}(\lambda_1, \dots, \lambda_p) U^\top, \quad \Sigma_0 = U \text{diag}(\sigma_1, \dots, \sigma_p) U^\top.$$

Then, all three worst-case covariance matrices have the same eigenvectors with eigenvalues

$$\sigma_i^F = \sigma_i + \frac{\epsilon \lambda_i}{(\sum_{j=1}^p \lambda_j^2)^{1/2}}, \quad \sigma_i^{KL} = \frac{\sigma_i}{1 - 2\sigma_i \lambda_i / \theta_K^{KL}}, \quad \sigma_i^W = \frac{\sigma_i}{(1 - \lambda_i / \theta_K^W)^2}.$$

Hence, the Frobenius model produces an additive covariance perturbation, the KL model produces a multiplicative inflation determined jointly by σ_i and λ_i , and the Wasserstein model stretches the disturbance realization in direction i by the factor $(1 - \lambda_i/\theta_K^w)^{-1}$.

A.2. Invisible directions in standard $\mathcal{H}_2$ cost

The following result discusses the invisible directions of the standard $\mathcal{H}_2$ cost v^s using the time-domain characterization (4) via closed-loop Markov parameters. Recall that Y_K^s is the solution to the Lyapunov equation (9) associated with unit disturbance covariance, that is, the controllability Gramian of the pair (A_K, E) .

Proposition A.1 (Invisible directions in standard $\mathcal{H}_2$ cost). *Let $K \in \mathcal{K}_s$ and define*

$$\mathcal{I}_K := \left\{ \Delta \in \mathbf{R}^{m \times n} : \begin{bmatrix} B \\ D \end{bmatrix} \Delta Y_K^s = 0 \right\}.$$

Then, for every $\Delta \in \mathcal{I}_K$ such that $K + \Delta \in \mathcal{K}_s$, we have $v^s(K + \Delta) = v^s(K)$.

Proof. Since A_K is Schur stable, the range of its controllability Gramian is equal to the reachable subspace of the pair (A_K, E) , that is,

$$\text{range}(Y_K^s) = \text{span} \begin{bmatrix} E & A_K E & A_K^2 E & \cdots & A_K^{n-1} E \end{bmatrix}.$$

In particular,

$$\text{range}(A_K^t E) \subseteq \text{range}(Y_K^s), \quad \forall t \in \mathbf{N}_0.$$

Let $\Delta \in \mathcal{I}_K$ so that $B\Delta Y_K^s = 0$ and $D\Delta Y_K^s = 0$. Then, using the preceding inclusion, we arrive at

$$B\Delta A_K^t E = 0 \quad \text{and} \quad D\Delta A_K^t E = 0, \quad \forall t \in \mathbf{N}_0. \quad (54)$$

We now prove by induction that

$$A_{K+\Delta}^t E = A_K^t E, \quad \forall t \in \mathbf{N}_0. \quad (55)$$

The equality is immediate for $t = 0$. Suppose that it holds for some $t \in \mathbf{N}_0$. Then, using the first equality in (54), we obtain

$$A_{K+\Delta}^{t+1} E = A_{K+\Delta} A_{K+\Delta}^t E = (A_K + B\Delta) A_K^t E = A_K^{t+1} E + B\Delta A_K^t E = A_K^{t+1} E.$$

Finally, using (55) and the second equality in (54), we have

$$C_{K+\Delta} A_{K+\Delta}^t E = (C_K + D\Delta) A_K^t E = C_K A_K^t E + D\Delta A_K^t E = C_K A_K^t E, \quad \forall t \in \mathbf{N}_0.$$

That is, all closed-loop Markov parameters from the disturbance w to the performance output z remain unchanged. The claim then follows from the time-domain characterization (4) of v^s using the closed-loop Markov parameters. $\square$

In simple words, the preceding result states that any stable perturbation Δ of a stabilizing control gain K that is in the null space of $[B^\top \ D^\top]^\top$ and/or the unreachable subspace of the pair (A_K, E) will be invisible in the sense that it does not affect the value of the cost $v^s(K)$. This is indeed

expected since the closed-loop matrices $A_K = A + BK$ $C_K = C + DK$ are by construction insensitive to perturbations of K in the null space of $[B \ D]$. More importantly, $\text{range}(Y_K^s)$ defines the reachable subspace of the pair (A_K, E) , that is, the part of the state space $\mathbb{R}^n$ that the disturbance w is physically capable of exciting. By the Kalman controllability decomposition, any state in the orthogonal complement $\text{range}(Y_K^s)^\perp$ is decoupled from the disturbance. Hence, perturbations of K in $\text{range}(Y_K^s)^\perp$ act only on the unreachable modes of the pair (A_K, E) and leave the input-output behavior of the disturbance channel invariant.

A.3. From DR to standard $\mathcal{H}_2$ optimality

The following result provides a sufficient condition for a solution of the DR $\mathcal{H}_2$ problem to be also optimal for the standard $\mathcal{H}_2$ problem. Recall that $\mathcal{C}_K = \text{argmax}_{\Sigma \in \mathcal{C}} \langle \Sigma, M_K \rangle$ denotes the set of active worst-case covariances corresponding to $K \in \mathcal{K}_s$ in the DR $\mathcal{H}_2$ problem. Also, recall that $Y_K(\Sigma)$ is the solution to the Lyapunov equation (9) associated with disturbance covariance Σ .

Proposition A.2 (From DR to standard $\mathcal{H}_2$ optimality). *Let Assumption 3.4 hold and $K_\star \in \mathcal{K}_s$ be a stabilizing solution of the DR $\mathcal{H}_2$ problem (26). Also, let $\Sigma_{K_\star} \in \mathcal{C}_{K_\star}$ be an active worst-case covariance corresponding to $K_\star$ such that $S_{K_\star} Y_{K_\star}(\Sigma_{K_\star}) = 0$. If $\Sigma_{K_\star} \succ 0$, then the gain $K_\star$ is also a solution of the standard $\mathcal{H}_2$ problem (3).*

Proof. By Theorem 3.6, since $\mathcal{K}_s$ is a DR $\mathcal{H}_2$ -optimal solution, we have $0 \in \partial_{\mathcal{C}} v^D(K_\star)$. Then, by the Clarke subdifferential representation in Proposition 3.5, there exists an active worst-case covariance $\Sigma_{K_\star} \in \mathcal{C}_{K_\star}$ such that $S_{K_\star} Y_{K_\star}(\Sigma_{K_\star}) = 0$. Post-multiplying the latter equality by $S_{K_\star}^\top$ and taking the trace, we obtain

$$\text{tr}(S_{K_\star} Y_{K_\star}(\Sigma_{K_\star}) S_{K_\star}^\top) = 0.$$

Since $K_\star \in \mathcal{K}_s$, the matrix $A_{K_\star}$ is Schur stable and the controllability Gramian $Y_{K_\star}(\Sigma_{K_\star})$ admits the convergent series representation (8). Plugging in the series representation of $Y_{K_\star}(\Sigma_{K_\star})$ in the equality above, we arrive at

$$\sum_{t=0}^{\infty} \text{tr}\left(S_{K_\star} A_{K_\star}^t E \Sigma_{K_\star} E^\top (A_{K_\star}^t)^\top S_{K_\star}^\top\right) = \sum_{t=0}^{\infty} \left\| S_{K_\star} A_{K_\star}^t E \Sigma_{K_\star}^{1/2} \right\|_F^2 = 0.$$

Every summand in the preceding equality is nonnegative. Hence,

$$S_{K_\star} A_{K_\star}^t E \Sigma_{K_\star}^{1/2} = 0, \quad \forall t \in \mathbf{N}_0.$$

By assumption $\Sigma_{K_\star}$ and hence $\Sigma_{K_\star}^{1/2}$ is nonsingular. Then, pre-multiplying the preceding equality by $\Sigma_{K_\star}^{-1/2}$ gives

$$S_{K_\star} A_{K_\star}^t E = 0, \quad \forall t \in \mathbf{N}_0. \tag{56}$$

On the other hand, using the gradient formula for the standard $\mathcal{H}_2$ cost in Lemma 2.1 and the convergent series representation (8) of standard controllability Gramian $Y_{K_\star}^s = Y_{K_\star}(I)$, we have

$$\nabla_{K^s} v^s(K_\star) = 2 S_{K_\star} Y_{K_\star}^s = 2 \sum_{t=0}^{\infty} (S_{K_\star} A_{K_\star}^t E) E^\top (A_{K_\star}^t)^\top.$$

Using the equality (56), we conclude that $\nabla_K v^s(K_\star) = 0$. Then, by Theorem 2.3, $K_\star$ is a standard $\mathcal{H}_2$ -optimal solution. $\square$

We notice that the nonsingularity of $\Sigma_{K_\star}$ is used precisely in passing from $S_{K_\star} A_{K_\star}^t E \Sigma_{K_\star}^{1/2} = 0$ to $S_{K_\star} A_{K_\star}^t E = 0$ for all $t \in \mathbf{N}_0$. If $\Sigma_{K_\star}$ is singular, the DR $\mathcal{H}_2$ -optimality condition only constrains the disturbance directions contained in $\text{range}(\Sigma_{K_\star})$. Therefore, the directions contained in the null space of $\Sigma_{K_\star}$ remain invisible, and the standard $\mathcal{H}_2$ optimality does not follow.

References

- [1] Ahmadi-Javid, A. (2012). Entropic value-at-risk: A new coherent risk measure. *Journal of Optimization Theory and Applications*, 155(3):1105–1123.
- [2] Ben-Tal, A., El Ghaoui, L., and Nemirovski, A. (2009). *Robust Optimization*. Princeton Series in Applied Mathematics. Princeton University Press, Princeton, NJ.
- [3] Blanchet, J. and Murthy, K. (2019). Quantifying distributional model risk via optimal transport. *Mathematics of Operations Research*, 44(2):565–600.
- [4] Boyd, S., El Ghaoui, L., Feron, E., and Balakrishnan, V. (1994). *Linear Matrix Inequalities in System and Control Theory*, volume 15 of *SIAM Studies in Applied Mathematics*. SIAM, Philadelphia, PA.
- [5] Brouillon, J.-S., Martin, A., Lygeros, J., Dörfler, F., and Ferrari-Trecate, G. (2025). Distributionally robust infinite-horizon control: From a pool of samples to the design of dependable controllers. *IEEE Transactions on Automatic Control*, 70(10):6465–6480.
- [6] Bu, J., Mesbahi, A., Fazel, M., and Mesbahi, M. (2019). LQR through the lens of first order methods: Discrete-time case. *arXiv preprint:1907.08921*.
- [7] Caverly, R. J. and Forbes, J. R. (2019). LMI properties and applications in systems, stability, and control theory. *arXiv preprint:1903.08599*.
- [8] Chen, C.-T. (1999). *Linear System Theory and Design*. Oxford University Press, 3rd edition.
- [9] Clarke, F. H. (1990). *Optimization and Nonsmooth Analysis*, volume 5 of *Classics in Applied Mathematics*. Society for Industrial and Applied Mathematics, Philadelphia.
- [10] Cohen, R. and Havlin, S. (2010). *Complex Networks: Structure, Robustness and Function*. Cambridge University Press, Cambridge.
- [11] Coppens, P., Schuurmans, M., and Patrinos, P. (2020). Data-driven distributionally robust LQR with multiplicative noise. In *Proceedings of the 2nd Conference on Learning for Dynamics and Control*, volume 120 of *Proceedings of Machine Learning Research*, pages 521–530. PMLR.
- [12] Delage, E. and Ye, Y. (2010). Distributionally robust optimization under moment uncertainty with application to data-driven problems. *Operations Research*, 58(3):595–612.
- [13] Doyle, J. C., Francis, B. A., and Tannenbaum, A. R. (1992). *Feedback Control Theory*. Macmillan Publishing Company.
- [14] Duan, J., Cao, W., Zheng, Y., and Zhao, L. (2023). On the optimization landscape of dynamic output feedback linear quadratic control. *IEEE Transactions on Automatic Control*, 69(2):920–935.
- [15] Dupuis, P. and Ellis, R. S. (1997). *A Weak Convergence Approach to the Theory of Large Deviations*. Wiley Series in Probability and Statistics. John Wiley & Sons, New York.
- [16] Falconi, L., Ferrante, A., and Zorzi, M. (2025). Distributionally robust LQG control under distributed uncertainty. *Automatica*, 174:112128.

- [17] Fatkhullin, I. and Polyak, B. (2021). Optimizing static linear feedback: Gradient method. *SIAM Journal on Control and Optimization*, 59(5):3887–3911.
- [18] Fazel, M., Ge, R., Kakade, S., and Mesbahi, M. (2018). Global convergence of policy gradient methods for the linear quadratic regulator. In *Proceedings of the 35th International Conference on Machine Learning*, volume 80, pages 1467–1476. PMLR.
- [19] Feng, H. and Lavaei, J. (2020). Connectivity properties of the set of stabilizing static decentralized controllers. *SIAM Journal on Control and Optimization*, 58(5):2790–2820.
- [20] Fochesato, M., Falconi, L., Zorzi, M., Ferrante, A., and Lygeros, J. (2025). Distributionally robust LQG with Kullback–Leibler ambiguity sets. arXiv preprint arXiv:2505.08370.
- [21] Gao, R. and Kleywegt, A. J. (2023). Distributionally robust stochastic optimization with Wasserstein distance. *Mathematics of Operations Research*, 48(2):603–655.
- [22] Gramlich, D., Yan, S., Scherer, C. W., and Ebenbauer, C. (2025). Distributionally robust LMI synthesis for LTI systems. arXiv preprint arXiv:2509.23493.
- [23] Hansen, L. P. and Sargent, T. J. (2008). *Robustness*. Princeton University Press, Princeton, NJ.
- [24] Hu, Z. and Hong, L. J. (2013). Kullback-Leibler divergence constrained distributionally robust optimization. *Available at Optimization Online*, 1(2):9.
- [25] Jacobson, D. H. (1973). Optimal stochastic linear systems with exponential performance criteria and their relation to deterministic differential games. *IEEE Transactions on Automatic Control*, 18(2):124–131.
- [26] Kalman, R. E. (1960). Contributions to the theory of optimal control. *Boletín de la Sociedad Matemática Mexicana*, 5:102–119.
- [27] Kim, K. and Yang, I. (2023). Distributional robustness in minimax linear quadratic control with Wasserstein distance. *SIAM Journal on Control and Optimization*, 61(2):458–483.
- [28] Kitano, H. (2004). Biological robustness. *Nature Reviews Genetics*, 5(11):826–837.
- [29] Krantz, S. G. and Parks, H. R. (2002). *The implicit function theorem: history, theory, and applications*, volume 202. Springer.
- [30] Kullback, S. and Leibler, R. A. (1951). On information and sufficiency. *The Annals of Mathematical Statistics*, 22(1):79–86.
- [31] Levine, W. S. and Athans, M. (1970). On the determination of the optimal constant output feedback gains for linear multivariable systems. *IEEE Transactions on Automatic Control*, 15(1):44–48.
- [32] Magnus, J. R. and Neudecker, H. (2019). *Matrix Differential Calculus with Applications in Statistics and Econometrics*. John Wiley & Sons, Hoboken, NJ, 3rd edition.
- [33] Mathai, A. M. and Provost, S. B. (1992). *Quadratic Forms in Random Variables: Theory and Applications*, volume 126 of *Statistics: Textbooks and Monographs*. Marcel Dekker, New York.
- [34] Meilakhs, A. M. (1975). Stabilization of linear controlled systems under uncertainty conditions. *Automation and Remote Control*, 36(2):349–351.
- [35] Mohajerin Esfahani, P. and Kuhn, D. (2018). Data-driven distributionally robust optimization using the Wasserstein metric: Performance guarantees and tractable reformulations. *Mathematical Programming*, 171(1-2):115–166.
- [36] Petersen, I. R., James, M. R., and Dupuis, P. (2000). Minimax optimal control of stochastic uncertain systems with relative entropy constraints. *IEEE Transactions on Automatic Control*, 45(3):398–412.
- [37] Rahimian, H. and Mehrotra, S. (2019). Distributionally robust optimization: A review. *arXiv preprint:1908.05659*.
- [38] Rockafellar, R. T. and Wets, R. J.-B. (1998). *Variational Analysis*. Springer, Berlin.
- [39] Rudin, W. (1976). *Principles of Mathematical Analysis*. McGraw-Hill, New York, 3rd edition.

- [40] Schöbi, A., Lanzetti, N., Dörfler, F., D’Andrea, R., and Terpin, A. (2026). Distributionally robust linear quadratic Gaussian regulator with stationary distributions. *IEEE Control Systems Letters*. Early access.
- [41] Taşkesen, B., Iancu, D. A., Koçyiğit, Ç., and Kuhn, D. (2023). Distributionally robust linear quadratic control. In *Advances in Neural Information Processing Systems*, volume 36, pages 18613–18632.
- [42] Trentelman, H. L. and Stoorvogel, A. A. (1995). Sampled-data and discrete-time $\mathcal{H}_2$ optimal control. *SIAM Journal on Control and Optimization*, 33(3):834–862.
- [43] Van Parys, B. P. G., Kuhn, D., Goulart, P. J., and Morari, M. (2016). Distributionally robust control of constrained stochastic systems. *IEEE Transactions on Automatic Control*, 61(2):430–442.
- [44] Villani, C. (2009). *Optimal Transport: Old and New*, volume 338 of *Grundlehren der mathematischen Wissenschaften*. Springer, Berlin, Heidelberg.
- [45] Villani, C. (2021). *Topics in optimal transportation*, volume 58. American Mathematical Soc.
- [46] Watanabe, Y. and Zheng, Y. (2026). Gradient dominance in the linear quadratic regulator: A unified analysis for continuous-time and discrete-time systems. *arXiv preprint:2602.22577*.
- [47] Whittle, P. (1981). Risk-sensitive linear/quadratic/Gaussian control. *Advances in Applied Probability*, 13(4):764–777.
- [48] Wiesemann, W., Kuhn, D., and Sim, M. (2014). Distributionally robust convex optimization. *Operations Research*, 62(6):1358–1376.
- [49] Yang, I. (2021). Wasserstein distributionally robust stochastic control: A data-driven approach. *IEEE Transactions on Automatic Control*, 66(8):3863–3870.
- [50] Zhou, K., Doyle, J. C., and Glover, K. (1996). *Robust and Optimal Control*. Prentice Hall.