Optimality of Linear Policies in Distributionally Robust Linear Quadratic Control

Bahar Taşkesen
University of Chicago
bahar.taskesen@chicagobooth.edu

Dan A. Iancu
Stanford University
daniancu@stanford.edu

Çağıl Koçyiğit
University of Luxembourg
cagil.kocyigit@uni.lu

Daniel Kuhn
EPFL
daniel.kuhn@epfl.ch

Abstract

We study a generalization of the classical discrete-time, Linear-Quadratic-Gaussian (LQG) control problem where the distributions of the noise terms affecting the states and observations are unknown and chosen adversarially from divergence-based ambiguity sets centered around a known nominal distribution. The noise terms are allowed to have nonzero mean but are required to have finite second moments that satisfy an orthogonality condition, which is equivalent to uncorrelatedness if noise means are zero. For a finite horizon model with zero-mean Gaussian nominal noise and a structural assumption on the divergence that is satisfied by many examples – including 2-Wasserstein distance, Kullback-Leibler divergence, moment-based divergences, entropy-regularized optimal transport, or Fisher (score-matching) divergence – we prove that a control policy that is *affine* in the observations is optimal and the adversary's corresponding worst-case optimal distribution is Gaussian. Under a weak condition satisfied by all these examples, we then prove that the adversary should optimally set the distribution's mean to zero and the optimal control policy becomes *linear*. Moreover, the adversary should optimally "inflate" the noise by choosing covariance matrices that dominate the nominal covariance in Loewner order. Exploiting these structural properties, we develop a Frank-Wolfe algorithm whose inner step solves standard LQG subproblems via Kalman filtering and dynamic programming and show that the implementation consistently outperforms semidefinite-programming reformulations of the problem. All structural and algorithmic results extend to an infinite-horizon, average-cost formulation, yielding stationary linear policies and a time-invariant Gaussian distribution for the adversary. Lastly, we show that when the divergence is 2-Wasserstein, the entire framework remains valid when the nominal distributions are elliptical rather than Gaussian.

Problem. The Linear Quadratic Gaussian (LQG) control problem has served as a fundamental building block for a wide range of applications in engineering and computer science (Auger et al., 2013; Chen, 2012), management (Bensoussan et al., 2007), economics (Hansen and Sargent, 2005), finance (Abeille et al., 2016), or medicine (Patek et al., 2007; Todorov and Jordan, 2002).

Motivated by practical settings where noise distributions may not be readily available or may not be Gaussian, we consider a generalization of the discrete-time LQG model where an adversary chooses

39th Conference on Neural Information Processing Systems (NeurIPS 2025) Workshop: MLxOR: Mathematical Foundations and Operational Integration of Machine Learning for Uncertainty-Aware Decision-Making.

the noise distributions from an ambiguity set and the decision maker's goal is to minimize the costs incurred under the worst-case distribution, for which we prove structural results on the optimal policies and design tractable solution procedures. Specifically, we consider linear dynamical systems that evolve over a finite number of periods $t \in \{0, 1, \ldots, T-1\}$ according to the equations:

$$x_{t+1} = A_t x_t + B_t u_t + w_t \quad \forall t \in \{0, 1, \ldots, T-1\}, \tag{1}$$

where $x_t \in \mathbb{R}^n$ denotes the system states, $u_t \in \mathbb{R}^m$ denotes the control inputs, $w_t \in \mathbb{R}^n$ denotes an exogenous noise process, and the system matrices $A_t \in \mathbb{R}^{n \times n}$ and $B_t \in \mathbb{R}^{n \times m}$ are known. The decision maker only has access to imperfect state measurements

$$y_t = C_t x_t + v_t \quad \forall t \in \{0, 1, \ldots, T-1\}, \tag{2}$$

corrupted by exogenous observation noise $v_t \in \mathbb{R}^p$, where $C_t \in \mathbb{R}^{p \times n}$ (and usually $p \leq n$). The control inputs u_t are *causal*, i.e., depend on the past observations $y_0, \ldots, y_t$ but not on the future observations $y_{t+1}, \ldots, y_{T-1}$, so that the set of feasible control inputs $\mathcal{U}_y$ is the set of random vectors $u = (u_0, u_1, \ldots, u_{T-1})$ such that $u_t = \varphi_t(y_0, \ldots, y_t)$ for every t, where $\varphi_t : \mathbb{R}^{p(t+1)} \to \mathbb{R}^m$ is a measurable control policy. Controlling the system generates quadratic costs:

$$J(u) = \sum_{t=0}^{T-1} (x_t^\top Q_t x_t + u_t^\top R_t u_t) + x_T^\top Q_T x_T, \tag{3}$$

where $Q_t \in \mathbb{R}^{n \times n}$ are positive semidefinite matrices governing the state costs, and $R_t \in \mathbb{R}^{m \times m}$ are positive definite matrices governing the input costs.

Different from the classical LQG formulation, which assumes that the joint probability distribution $\mathbb{P}$ for the noise terms is known, we adopt a distributionally robust model. We assume that the initial state x_0 and the noise terms $\{w_t\}_{t=0}^{T-1}$ and $\{v_t\}_{t=0}^{T-1}$ are exogenously determined and governed by an unknown probability distribution. Because all random vectors appearing in our model are functions of these exogenous uncertainties, we set the sample space without loss of generality as $\Omega = \mathbb{R}^n \times \mathbb{R}^{n \times T} \times \mathbb{R}^{p \times T}$. We use $\mathcal{F}$ to denote the Borel σ-algebra on Ω and $\mathbb{P}$ to denote the joint probability distribution of these random vectors. The joint distribution $\mathbb{P}$ is only known to belong to an ambiguity set $\mathcal{B}$. To model $\mathcal{B}$, we consider a known *nominal* distribution $\hat{\mathbb{P}}$ with marginals $\hat{\mathbb{P}}_z$ for every $z \in \mathcal{Z} = \{x_0, w_0, \ldots, w_{T-1}, v_0, \ldots, v_{T-1}\}$ and a *divergence* $\mathbb{D} : \mathcal{P}(\mathbb{R}^{d_z}) \times \mathcal{P}(\mathbb{R}^{d_z}) \to [0, +\infty]$, and we construct the ambiguity set $\mathcal{B}$ as:

$$\mathcal{B} = \left\{ \mathbb{P} \in \mathcal{P}(\mathbb{R}^{n+T(n+p)}) : \mathbb{P}_z \in \mathcal{B}_z,\ \mathbb{E}_{\mathbb{P}}[z' z^\top] = 0\ \forall z \neq z' \in \mathcal{Z} \right\}, \tag{4}$$

where, for all $z \in \mathcal{Z}$ and for finite $\rho_z \geq 0$, $\mathcal{B}_z = \{\mathbb{P}_z \in \mathcal{P}(\mathbb{R}^{d_z}) : \mathbb{D}(\mathbb{P}_z, \hat{\mathbb{P}}_z) \leq \rho_z\}$. We refer to the requirement that $\mathbb{E}_{\mathbb{P}}[z' z^\top] = 0$ for any $z \neq z' \in \mathcal{Z}$ as the *second moment orthogonality* (SMO) condition. Note that if $\mathbb{E}_{\mathbb{P}}[z] = \mathbb{E}_{\mathbb{P}}[z'] = 0$, this is equivalent to requiring that z, z' are uncorrelated. The decision maker seeks a causal control policy $u \in \mathcal{U}_y$ that minimizes the costs incurred under the worst-case distribution, $\sup_{\mathbb{P} \in \mathcal{B}} \mathbb{E}_{\mathbb{P}}[J(u)]$.

This problem – which we refer to as the distributionally-robust linear quadratic (DRLQ) problem – is challenging: both the decision maker and nature optimize over infinite-dimensional spaces, and the ambiguity set $\mathcal{B}$ contains many non-Gaussian distributions, so it is unclear which structural results from the classical LQG model hold or how one could compute an optimal control policy.

Notation. Subsequently, for $z \in \mathcal{Z}$, we use μ_z, M_z, and Σ_z to denote the mean, second moment, and covariance matrix, respectively, under a generic distribution $\mathbb{P} \in \mathcal{B}$, and use $\hat{\mu}_z, \hat{M}_z$, and $\hat{\Sigma}_z$ to denote these quantities, respectively, under the nominal distribution $\hat{\mathbb{P}}$. We let $\mathcal{M}_2^d = \{(\mu, M) : \mu \in \mathbb{R}^d,\ M \in \mathbb{S}_+^d,\ M \succeq \mu\mu^\top\}$ denote the set of valid pairs of first and second moments for a d-dimensional random vector. For $(\mu, M) \in \mathcal{M}_2^d$, we use $\mathcal{N}(\mu, M)$ to denote a Gaussian distribution with mean μ and second-moment matrix M.

Main Contributions. We first consider a finite-horizon formulation, where the nominal distribution $\hat{\mathbb{P}}$ is Gaussian and the nominal covariance matrices $\hat{\Sigma}_{v_t}$ of all observation noise terms are positive definite. (Both requirements are consistent with the classical LQG setting and some of our results relax these; see §A details.) We require the divergence $\mathbb{D}$ to satisfy the following key assumption.

Assumption. *The divergence $\mathbb{D}$ satisfies the following properties:*
(i) For every $(\mu_z, M_z) \in \mathcal{M}_2^{d_z}$, a Gaussian distribution minimizes the divergence $\mathbb{D}$ from $\hat{\mathbb{P}}_z$ among all distributions with mean μ_z and second moment matrix M_z, that is,

$$\mathbb{D}(\mathcal{N}(\mu_z, M_z), \hat{\mathbb{P}}_z) = \inf_{\mathbb{P}_z \in \mathcal{P}(\mathbb{R}^{d_z})} \Big\{ \mathbb{D}(\mathbb{P}_z, \hat{\mathbb{P}}_z) \ : \ \mathbb{E}_{\mathbb{P}_z}[z] = \mu_z, \ \ \mathbb{E}_{\mathbb{P}_z}[zz^\top] = M_z \Big\}.$$

(ii) The set $\mathcal{M}_{(\mu_z, M_z)} = \{(\mu_z, M_z) \in \mathcal{M}_2^{d_z} : \mathbb{D}(\mathcal{N}(\mu_z, M_z), \hat{\mathbb{P}}_z) \leq \rho_z\}$ is convex and compact.

The assumption admits an intuitive interpretation and holds in several important cases discussed below. (§A.2 discusses this more extensively, and the formal results and proofs are in Appendix §G.)

Wasserstein Ambiguity Sets. The assumption holds if the divergence $\mathbb{D}$ between the distributions $\mathbb{P}_z, \hat{\mathbb{P}}_z \in \mathcal{P}(\mathbb{R}^{d_z})$ corresponds to the 2-Wasserstein distance $\mathbb{W}$, defined as: $\mathbb{W}(\mathbb{P}_z, \hat{\mathbb{P}}_z) = (\inf_{\pi \in \Pi(\mathbb{P}_z, \hat{\mathbb{P}}_z)} \int_{\mathbb{R}^{d_z} \times \mathbb{R}^{d_z}} \|z - \hat{z}\|_2^2 \mathrm{d}\pi(z, \hat{z}))^{\frac{1}{2}}$, where $\Pi(\mathbb{P}_z, \hat{\mathbb{P}}_z)$ denotes the set of all couplings of $\mathbb{P}$ and $\hat{\mathbb{P}}$, that is, all joint distributions of the random vectors z and $\hat{z}$ with marginal distributions $\mathbb{P}_z$ and $\hat{\mathbb{P}}_z$, respectively. Ambiguity sets based on Wasserstein distance have become popular in the DRO literature due to their advantageous statistical and computational properties (Kuhn et al., 2019; Blanchet et al., 2021).

Kullback-Leibler Ambiguity Sets. The assumption holds if the divergence $\mathbb{D}$ corresponds to the Kullback-Leibler (KL) divergence $\mathbb{K}$, defined as $\mathbb{K}(\mathbb{P}_z, \hat{\mathbb{P}}_z) = \int_{\mathbb{R}^{d_z}} \log(\frac{\mathrm{d}\mathbb{P}_z}{\mathrm{d}\hat{\mathbb{P}}_z}(z)) \mathrm{d}\mathbb{P}_z(z)$ if $\mathbb{P}_z$ is absolutely continuous with respect to $\hat{\mathbb{P}}_z$, and $\mathbb{K}(\mathbb{P}_z, \hat{\mathbb{P}}_z) = \infty$ otherwise. The KL divergence is a well-established measure of discrepancy between probability distributions that has found applications in statistics, information theory, computer science, economics, and many other fields, and ambiguity sets based on KL divergence have been frequently considered in distributionally robust optimization (Whittle, 1981; El Ghaoui et al., 2003; Hu and Hong, 2013; Hansen and Sargent, 2005).

Moment Ambiguity Sets. The assumption holds if the divergence $\mathbb{D}$ between two probability distributions relies only on the first two moments of the distributions. Specifically, for $\mathbb{P}_z, \hat{\mathbb{P}}_z \in \mathcal{P}(\mathbb{R}^{d_z})$ with first and second moments (μ_z, M_z) and $(\hat{\mu}_z, \hat{M}_z)$, respectively, we take $\mathbb{D}(\mathbb{P}_z, \hat{\mathbb{P}}_z) = \mathbb{M}((\mu_z, M_z), (\hat{\mu}_z, \hat{M}_z))$, where $\mathbb{M} : \mathcal{M}_2^{d_z} \times \mathcal{M}_2^{d_z} \to [0, +\infty]$ is any divergence between pairs of first two moments satisfying $\mathbb{M}(m_z, m_z) = 0$ for all $m_z \in \mathcal{M}_2^{d_z}$. The assumption would then hold if the sublevel sets of the function $\mathbb{M}\big(\cdot, \hat{m}_z\big)$ restricted to its first variable are convex and compact.

In addition to these examples, Appendices §K and §L show that the assumption is also satisfied when the divergence $\mathbb{D}$ is chosen as the *entropy-regularized optimal transport divergence* or as the *Fisher divergence* (also known as *score-matching distance*), respectively. Ambiguity sets based on entropy-regularized discrepancies have appeared in the DRO literature before (Wang et al., 2021; Azizian et al., 2023), but have not been used in a control context. The Fisher divergence has been used recently in machine learning and computer vision (Hyvärinen, 2005; Song and Ermon, 2019), but to the best of our knowledge, it has never been considered in the DRO or control literature before.

For any ambiguity set consistent with this assumption, we prove that an optimal control policy exists that is *affine* in the observations, $u_t^\star = q_t + \sum_{\tau=0}^{t} U_{t,\tau} y_\tau$ for $q_t \in \mathbb{R}^m$ and $U_{t,\tau} \in \mathbb{R}^{m \times p}$, and that the associated worst-case optimal distribution (i.e., nature's optimal choice) $\mathbb{P}^\star$ is *Gaussian*. Our proof is novel and does not rely on traditional recursive dynamic programming arguments. Instead, we re-parameterize the control policy using purified observations and derive an upper bound for the resulting minimax formulation by relaxing the ambiguity set (to an outer approximation determined by the first two moments) while simultaneously restricting the decision maker to affine policies. We then use convex duality to show that the upper bound matches a lower bound obtained by restricting the ambiguity set (to Gaussian distributions) in the dual of the minimax formulation. The matching bounds then certify the optimality of affine output-feedback policies for the decision maker and of Gaussian distributions for the adversary.

Under two mild and intuitive assumptions that hold for every divergence we consider, we derive additional structural results that yield sharp managerial insights and facilitate computation.

The first result concerns nature's optimal choices of means $\mu_z^\star$ for the noise terms. We prove that when the nominal distribution has zero means ($\hat{\mu}_z = 0, \forall z \in \mathcal{Z}$), nature's optimal distribution $\mathbb{P}^\star$ also sets the mean to zero, $\mu_z^\star = 0$. Whereas the vast majority of papers formulating robust LQG models restrict attention to zero-mean noise for simplicity or in keeping with classical assumptions, our findings provide a different justification: this assumption/choice is *conservative*, because allowing the adversary to use zero means gives the adversary more power and results in the worst-case costs for the decision maker. Moreover, we prove that when noise is zero-mean, the optimal robust control policy is purely *linear* in the outputs, $u_t^\star = \sum_{\tau=0}^{t} U_{t,\tau} y_\tau$, which generalizes classical LQG results.

The second result concerns nature's optimal choice of covariances, $\Sigma_z^\star$. Restricting attention to zero-mean distributions ($\mu_z = 0, \forall z \in \mathcal{Z}, \forall \mathbb{P}_z \in \mathcal{B}_z$) and linear control policies $u_t^\star = \sum_{\tau=0}^{t} U_{t,\tau} y_\tau$, we prove that nature's optimal choice of covariance matrices dominate the nominal covariance matrices in Loewner order, $\Sigma_z^\star \succeq \hat{\Sigma}_z, \forall z \in \mathcal{Z}$. This shows that when model misspecification is a concern, a simple yet effective safeguard (even against adversarial distributional choices) is to "inflate" the nominal covariance matrix and solve a nominal model under the resulting noisier Gaussian distribution.

We leverage the structural results to design efficient algorithms for finding optimal control policies in the DRLQ problem. We propose an algorithm based on a Frank-Wolfe first-order method that solves sub-problems corresponding to classical LQG control problems, using Kalman filtering and dynamic programming. We show that our algorithm has sublinear convergence rate and is susceptible to parallelization. Our PyTorch implementation using automatic differentiation yields uniformly lower runtime than a more direct (semidefinite programming) method. We also find that the optimal robust policy significantly reduces worst-case costs, with virtually no loss under the nominal distribution.

We then extend our structural results to an infinite-horizon formulation of the DRLQ problem with time-invariant system matrices, time-invariant nominal distribution $\hat{\mathbb{P}}$, and average-cost objective. Specifically, we consider $T = \infty$ and linear *time-invariant* systems of the form (1) and (2) where $A_t = A_0, B_t = B_0, C_t = C_0, Q_t = Q_0, R_t = R_0$ for all $t \in \mathbb{N}$, and where $Q_0 \in \mathbb{S}_+^n$ and $R_0 \in \mathbb{S}_{++}^m$. In keeping with standard LQG models, we assume that (A_0, B_0) is stabilizable (i.e., there exists $K \in \mathbb{R}^{m \times n}$ such that $A_0 + B_0 K$ is Schur stable[1]), (A_0, C_0) is detectable (i.e., there exists $L \in \mathbb{R}^{n \times p}$ such that $A_0 - LC_0$ is Schur stable), and $Q_0 \succ 0$. We slightly strengthen our assumptions concerning the *nominal* distribution $\hat{\mathbb{P}}$ by requiring it to be zero-mean Gaussian and *time-invariant*, i.e., we assume there exist $\hat{\Sigma}_w \in \mathbb{S}_+^n, \hat{\Sigma}_v \in \mathbb{S}_+^p$ such that $\mathbb{E}_{\hat{\mathbb{P}}}[x_0 x_0^\top] = \hat{\Sigma}_w$ and for all $t \in \mathbb{N}$, $\mathbb{E}_{\hat{\mathbb{P}}}[w_t w_t^\top] = \hat{\Sigma}_w$ and $\mathbb{E}_{\hat{\mathbb{P}}}[v_t v_t^\top] = \hat{\Sigma}_v$. We then define the ambiguity set $\mathcal{B}^\infty$ exactly as the ambiguity set $\mathcal{B}$ from §A. Importantly, we do *not* require the control policies or all distributions in the ambiguity set to be stationary. For the objective, we consider minimizing the worst-case, long-run average cost:

$$\sup_{\mathbb{P} \in \mathcal{B}^\infty} \limsup_{T \to \infty} \frac{1}{T} \sum_{t=0}^{T-1} \mathbb{E}_{\mathbb{P}} \left[x_t^\top Q_0 x_t + u_t^\top R_0 u_t \right].$$

For this setting (which turns out to be technically more challenging), we prove that a *stationary*, linear control policy – that is, a policy of the form $u^\star = Uy$ where the matrix U is an infinite-dimensional, block lower triangular Toeplitz matrix and y is the stacked vector with all system observations – is optimal for the decision maker. Moreover, we prove that nature's optimal distribution $\mathbb{P}^\star$ is a *time-invariant* Gaussian distribution. The result not only generalizes our finite-horizon findings but also aligns seamlessly with the structural insights long-observed in traditional infinite-horizon LQG settings with known distributions.

The Appendix elaborates on several extensions of the framework. §K and §L confirm that all our structural results hold when the divergence $\mathbb{D}$ is chosen as the entropy-regularized optimal transport divergence or as the Fisher divergence (also known as score-matching distance), respectively. §M then replaces the nominal Gaussian distribution with an *elliptical* nominal distribution $\hat{\mathbb{P}}$ with finite second moments – a family that includes many non-Gaussian laws such as the Laplace, logistic, or hyperbolic distributions. Focusing on the 2-Wasserstein distance, we show that all our structural results hold and our scalable Frank-Wolfe algorithm is applicable.

[1] Recall that a matrix $A \in \mathbb{R}^{n \times n}$ is *Schur* stable if its eigenvalues are strictly within the unit complex circle.

Appendix

A Detailed Ambiguity Model, Assumptions, and Examples

We briefly reiterate the notation used throughout the rest of the paper. Subsequently, all random objects are defined on a measurable space $(\Omega, \mathcal{F})$. If this measurable space is equipped with a probability measure $\mathbb{P}$, then the distribution of any random vector $z : \Omega \to \mathbb{R}^{d_z}$ is given by the pushforward distribution $\mathbb{P}_z = \mathbb{P} \circ z^{-1}$ of $\mathbb{P}$ with respect to z. The expectation under $\mathbb{P}$ is denoted by $\mathbb{E}_{\mathbb{P}}[\cdot]$. We use $\mathcal{P}(\mathbb{R}^d)$ to denote the set of probability measures supported on $\mathbb{R}^d$ with finite second moments. The sets of all natural numbers with and without 0 are denoted by $\mathbb{N}$ and $\mathbb{N}_+$, respectively. For any $t \in \mathbb{N}$, we set $[t] = \{0, \ldots, t\}$. We use $\|A\|_F$ to denote the Frobenius norm and $\|A\|_2$ to denote the spectral norm of a matrix $A \in \mathbb{R}^{n \times m}$. We let $\mathcal{M}_2^d = \{(\mu, M) : \mu \in \mathbb{R}^d,\ M \in \mathbb{S}_+^d,\ M \succeq \mu\mu^\top\}$ denote the set of valid pairs of first and second moments for a d-dimensional random vector. For $(\mu, M) \in \mathcal{M}_2^d$, we use $\mathcal{N}(\mu, M)$ to denote a Gaussian distribution with mean μ and second-moment matrix M. (Occasionally and when no confusion can arise, we also refer to the Gaussian distribution in terms of its mean μ and covariance matrix $\Sigma = M - \mu\mu^\top$.) Subsequently, for every $z \in \mathcal{Z}$, we use μ_z, M_z, and Σ_z to denote the mean, second moment, and covariance matrix, respectively, under a generic distribution $\mathbb{P} \in \mathcal{B}$, and use $\hat{\mu}_z$, $\hat{M}_z$, and $\hat{\Sigma}_z$ to denote these quantities, respectively, under the nominal distribution $\hat{\mathbb{P}}$.

A.1 Assumptions for Tractability

To maintain tractability and rule out uninteresting cases, we impose a few assumptions on the nominal distribution $\hat{\mathbb{P}}$ and on the structure of the ambiguity set $\mathcal{B}$.

Assumption 1. *$\hat{\mathbb{P}}$ is a Gaussian distribution satisfying $\hat{\mu}_z = 0$ for all $z \in \mathcal{Z}$ and $\hat{\Sigma}_{v_t} \succ 0$ for all $t \in [0, T-1]$.*

Requiring $\hat{\mathbb{P}}$ to be Gaussian renders our model computationally tractable for several cases of practical interest and is consistent with the assumptions in the classical LQG model. Appendix §M shows that when the divergence $\mathbb{D}$ corresponds to a 2-Wasserstein distance, our results also hold for any *elliptical* nominal distribution $\hat{\mathbb{P}}$ with finite second moments – a class that includes many *non*-Gaussian distributions such as the Laplace, logistic, or hyperbolic distributions. In general, computing the optimal control policy for an *arbitrary* distribution $\mathbb{P}$ would be very difficult because even computing the state estimator $\hat{x}_t$ is hard in that case, as formalized in the following result.

Theorem A (Computational complexity of state estimation)**.** *Computing the state estimator $\hat{x}_t = \mathbb{E}_{\mathbb{P}}[x_t | y_0, \ldots, y_t]$ is #P-hard even if $t = 0$, $A_0 = B_0 = C_0 = 0$, v_0 and w_0 are mutually independent, and $\mathbb{P}_{w_0}$ is a uniform distribution on some nonempty polytope $P \subseteq [0,1]^n$ given as an intersection of finitely many half-spaces.*

Requiring the observation noise terms v_t to have positive definite covariance matrices, $\hat{\Sigma}_{v_t} \succ 0$, is consistent with the LQG framework and allows using the Kalman filter recursions, which compute inverses of these matrices (see Appendix §F). All our structural results in §B.1-§B.5 hold without this assumption, so this is only required in §B.6. In practice, even if the assumption is not satisfied, one can construct a positive definite covariance matrix that provides an ϵ-approximation to the optimal value, so the assumption does not introduce significant optimality gaps.

Regarding the ambiguity set $\mathcal{B}$, note that our model already includes two requirements: under all valid distributions $\mathbb{P}$, the exogenous noise terms $z \in \mathcal{Z}$ should have finite second moments (by the definition of $\mathcal{P}(\mathbb{R}^{d_z})$) and should satisfy the pairwise second-moment orthogonality condition by (4). Requiring finite second moments (which is encoded in the definition of $\mathcal{P}$) ensures that the DRLQ problem has a finite objective value; without this assumption, the adversary could choose a distribution $\mathbb{P}$ resulting in an unbounded objective under any finite control policy, because the LQG cost depends quadratically on the noise terms $z \in \mathcal{Z}$. We require second-moment orthogonality (SMO) rather than uncorrelatedness for tractability. When noise means are allowed to be nonzero, a

constraint that would require uncorrelatedness would be bilinear in means and covariances, making the feasible set of first and second moments non-convex. In contrast, the SMO condition is linear in the second moments, which gives rise to convex optimization problems. Because the SMO condition is equivalent to uncorrelatedness for zero-mean distributions, once we establish that the worst-case means are zero, we also indirectly establish that the worst-case distribution is uncorrelated, which recovers the classical LQG structure ex post, while maintaining tractability ex ante.

We also include two assumptions on the divergence $\mathbb{D}$ characterizing the ambiguity sets $\mathcal{B}_z$.

Assumption 2. *The divergence $\mathbb{D}$ satisfies the following properties:*

(i) For every $(\mu_z, M_z) \in \mathcal{M}_2^{d_z}$, a Gaussian distribution minimizes the divergence $\mathbb{D}$ from $\hat{\mathbb{P}}_z$ among all distributions with mean μ_z and second moment matrix M_z, that is,

$$\mathbb{D}(\mathcal{N}(\mu_z, M_z), \hat{\mathbb{P}}_z) = \left\{ \begin{array}{ll} \inf\limits_{\mathbb{P}_z \in \mathcal{P}(\mathbb{R}^{d_z})} & \mathbb{D}(\mathbb{P}_z, \hat{\mathbb{P}}_z) \\ \text{s.t.} & \mathbb{E}_{\mathbb{P}_z}[z] = \mu_z, \quad \mathbb{E}_{\mathbb{P}_z}[zz^\top] = M_z. \end{array} \right.$$

(ii) The set $\mathcal{M}_{(\mu_z, M_z)} = \{(\mu_z, M_z) \in \mathcal{M}_2^{d_z} : \mathbb{D}(\mathcal{N}(\mu_z, M_z), \hat{\mathbb{P}}_z) \leq \rho_z\}$ is convex and compact.

Assumption 2 holds in several important cases (see §A.2) and admits an intuitive interpretation. Requirement (i) readily holds if the divergence $\mathbb{D}(\mathbb{P}_z, \hat{\mathbb{P}}_z)$ depends only on the first two moments of the distributions $\mathbb{P}_z, \hat{\mathbb{P}}_z$. If the divergence is based on an information-theoretic principle related to uncertainty, the Gaussian distribution may satisfy requirement (i) because it is the "most uncertain" (maximum-entropy) distribution for a given set of first and second moments; this happens in two of our examples, corresponding to the Kullback-Leibler and Fisher divergences. More broadly, (i) can be thought of as restricting attention to ambiguity sets $\mathcal{B}_z$ that are "dense in Gaussians": the requirement is satisfied if for any $0 \leq \rho \leq \rho_z$, the set of distributions in $\mathcal{B}_z$ with "distance" of at most ρ from the nominal distribution – if nonempty – contains at least one Gaussian distribution. Part (ii) requires that the set of first and second moments characterizing all Gaussian distributions in the ambiguity set $\mathcal{B}_z$ is "well behaved," i.e., it is convex and compact. This enables us to evaluate worst-case expectations of *quadratic* functions of z (prominent in the LQG model) over the ambiguity set $\mathcal{B}_z$ by solving finite-dimensional, convex optimization problems.

A.2 Examples

Assumption 2 holds in several important instances, which we describe below. The formal results and proofs that help verify these properties are all included in Appendix §G.

Wasserstein Ambiguity Sets

Consider an ambiguity set where $\mathbb{D}$ corresponds to the 2-Wasserstein distance $\mathbb{W}$, defined as follows.

Definition 1 (2-Wasserstein distance)**.** *The 2-Wasserstein distance between two distributions $\mathbb{P}_z, \hat{\mathbb{P}}_z \in \mathcal{P}(\mathbb{R}^{d_z})$ is given by*

$$\mathbb{W}(\mathbb{P}_z, \hat{\mathbb{P}}_z) = \left(\inf_{\pi \in \Pi(\mathbb{P}_z, \hat{\mathbb{P}}_z)} \int_{\mathbb{R}^{d_z} \times \mathbb{R}^{d_z}} \|z - \hat{z}\|_2^2 \, \mathrm{d}\pi(z, \hat{z}) \right)^{\frac{1}{2}},$$

where $\Pi(\mathbb{P}_z, \hat{\mathbb{P}}_z)$ denotes the set of all couplings of $\mathbb{P}$ and $\hat{\mathbb{P}}$, that is, all joint distributions of the random vectors z and $\hat{z}$ with marginal distributions $\mathbb{P}_z$ and $\hat{\mathbb{P}}_z$, respectively.

Ambiguity sets based on Wasserstein distance have become popular in the DRO literature due to their advantageous statistical and computational properties (Kuhn et al., 2019; Blanchet et al., 2021).

Appendix §G.1.1 shows that the Wasserstein distance $\mathbb{W}$ satisfies Assumption 2. Leveraging known results in the literature, Proposition 8 proves that for any two distributions $\mathbb{P}_z, \hat{\mathbb{P}}_z \in \mathcal{P}(\mathbb{R}^{d_z})$ with first and second moment pairs given by $(\mu_z, M_z) \in \mathcal{M}_2^{d_z}$ and $(\hat{\mu}_z, \hat{M}_z) \in \mathcal{M}_2^{d_z}$, respectively,

$$\mathbb{W}(\mathbb{P}_z, \hat{\mathbb{P}}_z) \geq \mathbb{G}\left((\mu_z, M_z - \mu_z \mu_z^\top), (\hat{\mu}_z, \hat{M}_z - \hat{\mu}_z \hat{\mu}_z^\top) \right), \tag{A.5}$$

with equality holding if $\mathbb{P}_z$ and $\hat{\mathbb{P}}_z$ are Gaussian. Here, $\mathbb{G}((\mu_z, \Sigma_z), (\hat{\mu}_z, \hat{\Sigma}_z))$ is the Gelbrich distance between two mean-covariance pairs,

$$\mathbb{G}\big((\mu_z, \Sigma_z), (\hat{\mu}_z, \hat{\Sigma}_z)\big) = \sqrt{\|\mu_z - \hat{\mu}_z\|^2 + \mathrm{Tr}\left(\Sigma_z + \hat{\Sigma}_z - 2\left(\hat{\Sigma}_z^{1/2}\Sigma_z\hat{\Sigma}_z^{1/2}\right)^{1/2}\right)}. \tag{A.6}$$

This implies that Assumption 2-(i) is satisfied. Assumption 2-(ii) is also satisfied because the set $\mathcal{M}_{(\mu_z, M_z)}$ can be written as $\{(\mu_z, M_z) \in \mathcal{M}_2^{d_z} : \mathbb{G}((\mu_z, M_z - \mu_z\mu_z^\top), (\hat{\mu}_z, \hat{\Sigma}_z))^2 \leq \rho_z^2\}$, which is known to be convex and compact (Nguyen, 2019, Proposition 3.17).

Kullback-Leibler Ambiguity Sets

Next, consider an ambiguity set where the divergence $\mathbb{D}$ corresponds to the Kullback-Leibler (KL) divergence $\mathbb{K}$, defined as follows.

Definition 2 (KL divergence)**.** *The KL divergence from distribution $\mathbb{P}_z \in \mathcal{P}(\mathbb{R}^{d_z})$ to distribution $\hat{\mathbb{P}}_z \in \mathcal{P}(\mathbb{R}^{d_z})$ is defined as*

$$\mathbb{K}(\mathbb{P}_z, \hat{\mathbb{P}}_z) = \int_{\mathbb{R}^{d_z}} \log\left(\frac{\mathrm{d}\mathbb{P}_z}{\mathrm{d}\hat{\mathbb{P}}_z}(z)\right) \mathrm{d}\mathbb{P}_z(z)$$

if $\mathbb{P}_z$ is absolutely continuous with respect to $\hat{\mathbb{P}}_z$, and $\mathbb{K}(\mathbb{P}_z, \hat{\mathbb{P}}_z) = \infty$ otherwise.

The KL divergence is a well-established measure of discrepancy between probability distributions that has found applications in statistics, information theory, computer science, economics, and many other fields, and ambiguity sets based on KL divergence have been frequently considered in distributionally robust optimization and in robust control – see, e.g., Whittle (1981); El Ghaoui et al. (2003); Hu and Hong (2013); Hansen and Sargent (2005).

Appendix §G.1.2 leverages known results in the literature to argue that the KL divergence satisfies Assumption 2. Proposition 9 formalizes the key result that for any two distributions $\mathbb{P}_z, \hat{\mathbb{P}}_z \in \mathcal{P}(\mathbb{R}^{d_z})$ with first and second moments $(\mu_z, M_z) \in \mathcal{M}_2^{d_z}$ and $(\hat{\mu}_z, \hat{M}_z) \in \mathcal{M}_2^{d_z}$, respectively,

$$\mathbb{K}(\mathbb{P}_z, \hat{\mathbb{P}}_z) \geq \mathbb{T}\left((\mu_z, M_z - \mu_z\mu_z^\top), (\hat{\mu}_z, \hat{M}_z - \hat{\mu}_z\hat{\mu}_z^\top)\right), \tag{A.7}$$

with equality holding if $\mathbb{P}_z$ and $\hat{\mathbb{P}}_z$ are Gaussian. Here, $\mathbb{T}\big((\mu_z, \Sigma_z), (\hat{\mu}_z, \hat{\Sigma}_z)\big)$ is a (KL-type) divergence between two mean-covariance pairs,

$$\mathbb{T}\big((\mu_z, \Sigma_z), (\hat{\mu}_z, \hat{\Sigma}_z)\big) = \frac{1}{2}\left((\mu_z - \hat{\mu}_z)^\top\hat{\Sigma}^{-1}(\mu_z - \hat{\mu}_z) + \mathrm{Tr}\left(\Sigma_z\hat{\Sigma}_z^{-1}\right) - \log\det\left(\Sigma_z\hat{\Sigma}_z^{-1}\right) - d\right).$$

Result (A.7) implies that Assumption 2-(i) is satisfied. Assumption 2-(ii) is also satisfied because the set $\mathcal{M}_{(\mu_z, M_z)}$ can be expressed as $\big\{(\mu_z, M_z) \in \mathcal{M}_2^{d_z} : M_z - \mu_z\mu_z^\top \in \mathbb{S}_{++}^{d_z},\ \mathbb{T}\big((\mu_z, \Sigma_z), (\hat{\mu}_z, \hat{\Sigma}_z)\big) \leq \rho_z\big\}$. This representation follows from (A.7) and because we are interested in ρ_z finite (which implies that $\mathbb{T}$ must be finite, so any relevant $\mathbb{P}_z \in \mathcal{B}_z$ must satisfy $\Sigma_z = M_z - \mu_z\mu_z^\top \succ 0$ due to the $-\log\det\Sigma_z$ term in $\mathbb{T}$). The latter set is known to be convex and compact (Taşkesen et al., 2021, Lemma A.3).

Moment Ambiguity Sets

Lastly, we consider moment ambiguity sets where the divergence $\mathbb{D}$ between two probability distributions relies only on the first two moments of the distributions. Specifically, for $\mathbb{P}_z, \hat{\mathbb{P}}_z \in \mathcal{P}(\mathbb{R}^{d_z})$ with first and second moments (μ_z, M_z) and $(\hat{\mu}_z, \hat{M}_z)$, respectively, we take

$$\mathbb{D}(\mathbb{P}_z, \hat{\mathbb{P}}_z) = \mathbb{M}\big((\mu_z, M_z), (\hat{\mu}_z, \hat{M}_z)\big),$$

where $\mathbb{M} : \mathcal{M}_2^{d_z} \times \mathcal{M}_2^{d_z} \to [0, +\infty]$ is any divergence between pairs of first two moments satisfying $\mathbb{M}(m_z, m_z) = 0$ for all $m_z \in \mathcal{M}_2^{d_z}$. The ambiguity set $\mathcal{B}_z$ for the random variable z with nominal distribution $\hat{\mathbb{P}}_z$ (with mean $\hat{\mu}_z$ and second moment matrix $\hat{M}_z$) can therefore be expressed as:

$$\mathcal{B}_z = \left\{\mathbb{P}_z \in \mathcal{P}(\mathbb{R}^d) : \mathbb{E}_{\mathbb{P}_z}[z] = \mu_z,\ \mathbb{E}_{\mathbb{P}_z}[zz^\top] = M_z,\ \mathbb{M}\left((\mu_z, M_z), (\hat{\mu}_z, \hat{M}_z)\right) \leq \rho_z\right\}.$$

Assumption 2-(i) holds because every distribution (including a Gaussian distribution) with the same mean and second moment would yield the same divergence from the nominal distribution $\hat{\mathbb{P}}_z$ and would therefore minimize the divergence from $\hat{\mathbb{P}}_z$. Assumption 2-(ii) holds if the sublevel sets of the function $\mathbb{M}(\cdot, \hat{m}_z)$ restricted to its first variable are convex and compact (for the given $\hat{m}_z = (\hat{\mu}_z, \hat{M}_z)$). Any restriction of $\mathbb{M}$ that is quasiconvex and coercive would satisfy the requirement.

B Nash Equilibrium and Optimality of Linear Policies and Gaussians

This section proves our main structural results concerning the DRLQ problem. We view this problem as a game between the decision maker, who chooses causal control inputs, and nature, which chooses a distribution $\mathbb{P} \in \mathcal{B}$. We show that this game admits a Nash equilibrium under our standing assumptions, wherein nature's strategy is a *Gaussian* distribution, $\mathbb{P}_z^\star = \mathcal{N}(\mu_z^\star, M_z^\star)$ for any $z \in \mathcal{Z}$, and the decision maker's strategy is an *affine* output feedback policy. Under mild additional assumptions, we also prove that nature's optimal distribution has a mean of zero, in which case the decision maker's optimal strategy becomes *linear* and nature's optimal strategy entails choosing a covariance matrix for the noise terms $\Sigma_z^\star$ that dominates the nominal covariance matrix $\hat{\Sigma}_z$ in Loewner order, $\Sigma_z^\star \succeq \hat{\Sigma}_z$.

B.1 Reformulation with Purified Observations

We first simplify the problem formulation by re-parametrizing the control inputs in a more convenient form. (This follows ideas similar to Ben-Tal et al., 2005, 2006; Hadjiyiannis et al., 2011.) Note that the control inputs in the formulation are subject to cyclic dependencies, as u depends on y, while y depends on x through (2), and x depends again on u through (1), etc. Because these dependencies make the problem hard to analyze, it is preferable to instead consider the controls as functions of a new set of so-called *purified* observations instead of the actual observations y_t.

Specifically, we first introduce a fictitious *noise-free* system

$$x'_{t+1} = A_t x'_t + B_t u_t \quad \forall t \in [T-1] \quad \text{and} \quad y'_t = C_t x'_t \quad \forall t \in [T-1]$$

with states $x'_t \in \mathbb{R}^n$ and outputs $y'_t \in \mathbb{R}^p$, which is initialized with $x'_0 = 0$ and controlled by the *same* inputs u_t as the original system (1). We then define the purified observation at time t as $\eta_t = y_t - y'_t$ and we use $\eta = (\eta_0, \dots, \eta_{T-1})$ to denote the trajectory of *all* purified observations.

Because the inputs u_t are causal, the decision maker can compute the fictitious state x'_t and output y'_t from the observations $y_0, \dots, y_t$. Thus, η_t is representable as a function of $y_0, \dots, y_t$. Conversely, one can show by induction that y_t can also be represented as a function of $\eta_0, \dots, \eta_t$. Moreover, any measurable function of $y_0, \dots, y_t$ can be expressed as a measurable function of $\eta_0, \dots, \eta_t$ and vice-versa (Hadjiyiannis et al., 2011, Proposition II.1). So if we define $\mathcal{U}_\eta$ as the set of all control inputs $(u_0, u_1, \dots, u_{T-1})$ so that $u_t = \phi_t(\eta_0, \dots, \eta_t)$ for some measurable function $\phi_t : \mathbb{R}^{p(t+1)} \to \mathbb{R}^m$ for every $t \in [T-1]$, the above reasoning implies that $\mathcal{U}_\eta = \mathcal{U}_y$. Moreover, the class of causal control policies that are affine (respectively, linear) in η is equivalent to the class of causal control policies that are affine (respectively, linear) in y (see Ben-Tal et al. (2005); Skaf and Boyd (2010) and Lemma E in §F.3 for a concise proof). Therefore, in interpreting all our results, the existence of optimal affine (linear) policies $u^\star \in \mathcal{U}_\eta$ is equivalent to the existence of optimal affine (respectively, linear) policies $u^\star \in \mathcal{U}_y$.

In view of this, we can rewrite the DRLQ problem equivalently as:

$$\begin{aligned} p^\star &= \left\{ \begin{array}{ll} \min\limits_{x,u,y} & \max\limits_{\mathbb{P}\in\mathcal{B}} \mathbb{E}_{\mathbb{P}}\left[u^\top R u + x^\top Q x\right] \\ \text{s.t.} & u \in \mathcal{U}_y, x = Hu + Gw, y = Cx + v \end{array} \right. \\ &= \left\{ \begin{array}{ll} \min\limits_{x,u} & \max\limits_{\mathbb{P}\in\mathcal{B}} \mathbb{E}_{\mathbb{P}}\left[u^\top R u + x^\top Q x\right] \\ \text{s.t.} & u \in \mathcal{U}_\eta,\ x = Hu + Gw, \end{array} \right. \end{aligned} \tag{A.8}$$

where $x = (x_0, \ldots, x_T)$, $u = (u_0, \ldots, u_{T-1})$, $y = (y_0, \ldots, y_{T-1})$, $w = (x_0, w_0, \ldots, w_{T-1})$, $v = (v_0, \ldots, v_{T-1})$, $\eta = (\eta_0, \ldots, \eta_{T-1})$, and R, Q, H, G and C are suitable block matrices (see Appendix §F.2 for their precise definitions).

The latter reformulation involving the purified observations η is useful for two reasons. First, the purified outputs are *independent* of the inputs u. Indeed, by recursively combining the equations of the original and the noise-free systems, one can show that $\eta = Dw + v$ for some block triangular matrix D (see Appendix §F.2 for the construction). So the purified observations depend (affinely) on the exogenous uncertainties but do *not* depend on the control inputs u, and hence, the cyclic dependencies complicating the original system are eliminated in (A.8). Second and more importantly, the purified observations will allow us to reformulate the non-convex DRLQ problem as a *convex* optimization problem, as will become obvious subsequently.

Our results also rely on the dual of (A.8), defined as

$$d^\star = \left\{ \begin{array}{lll} \max\limits_{\mathbb{P}\in\mathcal{B}} & \min\limits_{x,u} & \mathbb{E}_{\mathbb{P}}\left[u^\top R u + x^\top Q x\right] \\ & \text{s.t.} & u \in \mathcal{U}_\eta, x = Hu + Gw. \end{array} \right. \tag{A.9}$$

We prove that $p^\star = d^\star$ and that a Nash equilibrium for our game exists wherein $\mathbb{P}^\star$ is Gaussian and $u^\star$ is affine. The classical minimax inequality implies that $p^\star \geq d^\star$. To prove that $p^\star = d^\star$, we construct an upper bound for $p^\star$ and a lower bound for $d^\star$, and then argue that these bounds match.

B.2 Upper Bound for Primal

We obtain an upper bound for $p^\star$ by suitably *enlarging* the ambiguity set $\mathcal{B}$ and *restricting* the control policies u_t to be affine. We first define the following ambiguity set for the noise terms:

$$\overline{\mathcal{B}} = \{\mathbb{P} \in \mathcal{P}(\mathbb{R}^{n+T(n+p)}) : \mathbb{P}_z \in \overline{\mathcal{B}}_z,\ \mathbb{E}_{\mathbb{P}}[z'z^\top] = 0\ \forall z \in \mathcal{Z}, \forall z' \neq z \in \mathcal{Z}\},$$

where, for all $z \in \mathcal{Z}$,

$$\overline{\mathcal{B}}_z = \left\{\mathbb{P}_z \in \mathcal{P}(\mathbb{R}^{d_z}) : \exists (\mu_z, M_z) \in \mathcal{M}_2^{d_z} \text{ with } \mathbb{E}_{\mathbb{P}_z}[z] = \mu_z, \mathbb{E}_{\mathbb{P}_z}[zz^\top] = M_z, \mathbb{D}\big(\mathcal{N}(\mu_z, M_z), \hat{\mathbb{P}}_z\big) \leq \rho_z\right\}.$$

To see that $\overline{\mathcal{B}}$ constitutes an outer approximation for the ambiguity set $\mathcal{B}$, note first that the random vectors x_0, $\{w_t\}_{t=0}^{T-1}$ and $\{v_t\}_{t=0}^{T-1}$ satisfy the SMO condition and have finite second moments under any $\mathbb{P} \in \overline{\mathcal{B}}$. Moreover, any $\mathbb{P}_z \in \mathcal{B}_z$ satisfies $\mathbb{D}(\mathbb{P}_z, \hat{\mathbb{P}}_z) \leq \rho_z$ and because $\mathbb{D}\big(\mathcal{N}(\mu_z, M_z), \hat{\mathbb{P}}_z\big) \leq \mathbb{D}(\mathbb{P}_z, \hat{\mathbb{P}}_z)$ by Assumption 2-(i), we have that $\mathbb{P}_z \in \overline{\mathcal{B}}_z$ and therefore $\mathcal{B} \subseteq \overline{\mathcal{B}}$.

To finalize our construction of the upper bound on $p^\star$, we focus on affine policies of the form $u = q + U\eta = q + U(Dw + v)$, where $q = (q_0, \ldots, q_{T-1})$, and U is a block lower triangular matrix

$$U = \begin{bmatrix} U_{0,0} & & & \\ U_{1,0} & U_{1,1} & & \\ \vdots & & \ddots & \\ U_{T-1,0} & \ldots & \ldots & U_{T-1,T-1} \end{bmatrix}. \tag{A.10}$$

The block lower triangularity of U ensures that the corresponding control policy is causal, which in turn ensures that $u \in \mathcal{U}_\eta$. In the following, we denote by $\mathcal{U}$ the set of all block lower triangular matrices of the form (A.10).

An upper bound on problem (A.8) can now be obtained by *restricting* the decision maker's feasible set to causal control policies that are *affine* in the purified observations η and by *relaxing* nature's feasible set to the outer approximation $\overline{\mathcal{B}}$ of $\mathcal{B}$. The resulting upper bound is given by:

$$\overline{p}^\star = \left\{ \begin{array}{lll} \min\limits_{U,q,x,u} & \max\limits_{\mathbb{P}\in\overline{\mathcal{B}}} & \mathbb{E}_{\mathbb{P}}\left[u^\top R u + x^\top Q x\right] \\ \text{s.t.} & & U \in \mathcal{U}, u = q + U(Dw + v), x = Hu + Gw. \end{array} \right. \tag{A.11}$$

Because we obtained (A.11) by restricting the feasible set of the outer minimization problem and relaxing the feasible set of the inner maximization problem in (A.8), it is clear that $\overline{p}^\star \geq p^\star$.

Although problem (A.8) is still an infinite-dimensional, zero-sum game because nature's choices are over distributions $\mathbb{P}$, important simplifications are possible. Specifically, note that for any fixed U, q, the control policies u and induced states $x = Hu + Gw$ are affine functions on the noise terms w, v, and therefore the expected value of the objective, $\mathbb{E}_{\mathbb{P}}\left[u^\top Ru + x^\top Qx\right]$, only depends on the first two moments of the random vector (w, v) under distribution $\mathbb{P}$. This implies that problem (A.11) can be rewritten as a finite-dimensional zero-sum game, as formalized in the following result.

Proposition 1. *Problem* (A.11) *has the same optimal value as the optimization problem:*

$$\overline{p}^\star = \begin{cases} \min\limits_{\substack{q\in\mathbb{R}^{pT}\\ U\in\mathcal{U}}} \max\limits_{\substack{(\mu_w,M_w)\in\mathcal{M}_{(\mu_w,M_w)}\\ (\mu_v,M_v)\in\mathcal{M}_{(\mu_v,M_v)}}} \operatorname{Tr}\Big(\big((UD)^\top RUD + (G+HUD)^\top Q(G+HUD)\big)M_w + U^\top \bar{R}UM_v\Big) \\ \qquad\qquad + 2q^\top(\bar{R}UD + G^\top QH)\mu_w + 2q^\top \bar{R}U\mu_v + q^\top \bar{R}q, \end{cases} \tag{A.12}$$

where $\bar{R} = R + H^\top QH$ *and*

$$\mathcal{M}_{(\mu_w,M_w)} = \mathcal{M}_{(\mu_{x_0},M_{x_0})} \times \prod_{t=0}^{T-1} \mathcal{M}_{(\mu_{w_t},M_{w_t})}, \qquad \mathcal{M}_{(\mu_v,M_v)} = \prod_{t=0}^{T-1} \mathcal{M}_{(\mu_{v_t},M_{v_t})}.$$

For this and subsequent results with omitted proofs, we refer the reader to §H.

B.3 Lower Bound for Dual

To derive a tractable lower bound on $d^\star$, we restrict nature's feasible set to the family $\mathcal{B}_\mathcal{N}$ of all *Gaussian* distributions in the ambiguity set $\mathcal{B}$. The resulting bounding problem is thus given by

$$\underline{d}^\star = \begin{cases} \max\limits_{\mathbb{P}\in\mathcal{B}_\mathcal{N}} \quad \min\limits_{x,u} & \mathbb{E}_{\mathbb{P}}\left[u^\top Ru + x^\top Qx\right] \\ \qquad\qquad \text{s.t.} & u \in \mathcal{U}_\eta, x = Hu + Gw. \end{cases} \tag{A.13}$$

As we obtained (A.13) by restricting the feasible set of the outer maximization problem in (A.9), it is clear that $\underline{d}^\star \leq d^\star$. Next, by leveraging the fact that the inner minimization problem in (A.13) is solved by an affine control policy for any fixed Gaussian distribution in $\mathcal{B}_\mathcal{N}$, we show that (A.13) can be recast as a finite-dimensional zero-sum game.

Proposition 2. *Problem* (A.13) *has the same optimal value as the optimization problem:*

$$\underline{d}^\star = \begin{cases} \max\limits_{\substack{(\mu_w,M_w)\in\mathcal{M}_{(\mu_w,M_w)}\\ (\mu_v,M_v)\in\mathcal{M}_{(\mu_v,M_v)}}} \min\limits_{\substack{q\in\mathbb{R}^{pT}\\ U\in\mathcal{U}}} \operatorname{Tr}\Big(\big((UD)^\top RUD + (G+HUD)^\top Q(G+HUD)\big)M_w + U^\top \bar{R}UM_v\Big) \\ \qquad\qquad + 2q^\top(\bar{R}UD + G^\top QH)\mu_w + 2q^\top \bar{R}U\mu_v + q^\top \bar{R}q, \end{cases} \tag{A.14}$$

where $\bar{R}$, $\mathcal{M}_{(\mu_w,M_w)}$ *and* $\mathcal{M}_{(\mu_v,\Sigma_v)}$ *are defined exactly as in Proposition 1.*

B.4 Optimality of Affine Policies and Gaussian Distributions

The next result leverages the primal and dual relaxations to prove our main result that the primal DRLQ problem in (A.8) and its dual in (A.9) actually have the same optimal value.

Theorem A (Strong duality)**.** *The optimal value in problem* (A.8) *equals the optimal values in problems* (A.12)*,* (A.14) *and* (A.9)*, i.e.,* $p^\star = \overline{p}^\star = \underline{d}^\star = d^\star$*, and all optimal values are attained.*

Proof. By weak duality and the construction of the bounding problems (A.12) and (A.14), we readily have that $\underline{d}^\star \leq d^\star \leq p^\star \leq \overline{p}^\star$. To complete the argument, we prove that $\underline{d}^\star = \overline{p}^\star$. Consider problem (A.12), with optimal value $\overline{p}^\star$, and problem (A.14), with optimal value $\underline{d}^\star$. Note that these problems are dual to each other, that is, they can be transformed into one another by interchanging minimization and maximization. In both problems, the set $\mathcal{U}$ appearing in the minimization is convex and closed, and the feasible sets $\mathcal{M}_{(\mu_w,M_w)}$ and $\mathcal{M}_{(\mu_v,M_v)}$ in the maximization are convex and compact under Assumption 2 (ii). Moreover, the objective in both problems is convex quadratic in the minimization variables (U, q) and is linear in the maximization variables (μ_w, M_w) and (μ_v, M_v). Therefore, the conditions of Sion's minimax theorem (Sion, 1958) are satisfied and we conclude that

$\underline{d}^\star = \overline{p}^\star$. Moreover, the optimal values in (A.8) and (A.12) are attained because of our assumption that $R \succ 0$, whereas the optimal values in (A.14) and (A.9) are attained because the feasible sets $\mathcal{M}_{\mu_w, M_w}$ and $\mathcal{M}_{\mu_v, M_v}$ are compact. □

Theorem A has several important implications. From a mathematical perspective, it establishes strong duality between two *infinite-dimensional* zero-sum games, which is not a priori expected to hold. From a practical perspective, it implies that a decision maker faced with solving the DRLQ problem (A.8) can restrict attention to policies that depend *affinely* on the purified outputs η (or, equivalently, on the original outputs y).

Corollary 1 (Decision maker's Nash strategy is affine)**.** *The primal DRLQ problem* (A.8) *admits an optimal* affine *policy of the form* $u^\star = q^\star + U^\star \eta$ *for some* $U^\star \in \mathcal{U}$ *and* $q^\star \in \mathbb{R}^m$.

Lastly, Theorem A implies that the worst-case distribution in the DRLQ problem is Gaussian.

Corollary 2 (Nature's Nash strategy is a Gaussian distribution)**.** *The dual DRLQ problem* (A.9) *admits an optimal solution that is a Gaussian distribution,* $\mathbb{P}^\star \in \mathcal{B}_\mathcal{N}$.

Corollary 2 follows from the equality $\underline{d}^\star = d^\star$. Note that the optimal Gaussian distribution $\mathbb{P}^\star$ is uniquely determined by the first and second moments $(\mu_w^\star, M_w^\star)$ and $(\mu_v^\star, M_v^\star)$ of the exogenous uncertain parameters, which can be computed by solving problem (A.14). That the worst-case distribution is actually Gaussian is not a-priori expected and is surprising given that the ambiguity set $\mathcal{B}$ contains many non-Gaussian distributions.

At an intuitive level, Theorem A delivers two key insights. First, it confirms the merits of affine output feedback policies, which extend from the classical LQG setting to the distributionally robust LQG setting: even when the noise distribution is unknown, a more elaborate control policy cannot outperform an affine one. Second, the result supplies a novel justification for the standard focus on Gaussian noise in the LQG model: beyond analytical tractability, this assumption is also *conservative* because it allows capturing the worst-case costs that the decision maker might face within the ambiguity set (provided the nominal distribution remains Gaussian).

B.5 Optimality of Linear Policies and Zero-Mean Distributions

Under some additional mild assumptions on the ambiguity sets, we can further refine the structural results concerning the decision maker's and nature's Nash strategies. We first state an additional assumption on the set of first two moments $\mathcal{M}_{(\mu_z, M_z)}$ defined in Assumption 2.

Assumption 3. *For any* $z \in \mathcal{Z}$*, the set* $\mathcal{M}_{(\mu_z, M_z)} = \{(\mu_z, M_z) \in \mathcal{M}_2^{d_z} : \mathbb{D}(\mathcal{N}(\mu_z, M_z), \hat{\mathbb{P}}_z) \leq \rho_z\}$ *defined in Assumption 2 is such that* $(\mu_z, M_z) \in \mathcal{M}_{(\mu_z, M_z)}$ *implies that* $(0, M_z) \in \mathcal{M}_{(\mu_z, M_z)}$.

The assumption bears an intuitive interpretation. Formulated as a feasibility condition on the ambiguity set $\mathcal{B}_z$, it states that if a Gaussian distribution $\mathbb{P}_z = \mathcal{N}(\mu_z, M_z)$ belongs to $\mathcal{B}_z$, then the "centered" Gaussian $\mathbb{P}'_z = \mathcal{N}(0, M_z)$ – obtained by setting the mean to zero while keeping the second moment unchanged – should also be feasible, $\mathbb{P}'_z \in \mathcal{B}_z$. In other words, the adversary may always transfer deterministic bias into additional variance without leaving $\mathcal{B}_z$. Because this transformation raises the covariance from $M_z - \mu_z \mu_z^\top$ to M_z and the latter dominates the former in the Loewner (positive semidefinite) order, the key intuition behind the assumption is to allow the adversary to "inflate" uncertainty (while holding the second moment fixed).

The validity of Assumption 3 depends on three key inputs: the divergence $\mathbb{D}$, the nominal distribution $\hat{\mathbb{P}}_z = \mathcal{N}(\hat{\mu}_z, \hat{M}_z)$, and the radius ρ_z of the ambiguity set. The following result shows that all our examples from §A.2 satisfy Assumption 3 under very mild conditions if the nominal distribution has zero mean.

Proposition 3. *If* $\hat{\mu}_z = 0$*, the divergences* $\mathbb{W}, \mathbb{K}$ *defined in §A.2 satisfy Assumption 3 and the moment-based ambiguity set based on divergence* $\mathbb{M}$ *also satisfies the assumption if*

$$\mathbb{M}\left((0, M_z), (0, \hat{M}_z)\right) \leq \mathbb{M}\left((\mu_z, M_z), (0, \hat{M}_z)\right), \ \forall (\mu_z, M_z) \in \mathcal{M}_2^{d_z}. \tag{A.15}$$

To see that the conditions are mild, note that requiring the nominal mean to be zero is very common in LQR/LQG models. Textbook treatments of the classical LQG model restrict attention to zero-mean noise without loss of generality (Bertsekas, 2017), and the vast majority of distributionally-robust LQR/LQG formulations require that *all* distributions in the ambiguity set be zero-mean (see, for instance, Taşkesen et al. (2023); Kim and Yang (2023); Han (2023)). For the moment-based ambiguity set, the condition holds if the divergence $\mathbb{M}$ penalizes deviations from the nominal mean $\hat{\mu}_z = 0$, which is a sensible requirement (for instance, this holds if $\mathbb{M}$ penalizes separately the distance between the first moments and the second moments). Assumption 3 allows refining our structural results on the solution to the DRLQ problem.

Theorem B (Worst-case nominal mean and linear controls)**.** *Under Assumptions 1-3, problem* (A.9) *admits an optimal solution $\mathbb{P}^\star$ that is Gaussian and has zero mean, $\mathbb{E}_{\mathbb{P}^\star_z}[z] = 0$, for all $z \in \mathcal{Z}$. Moreover, under such $\mathbb{P}^\star$, the inner minimization problem in* (A.9) *admits an optimal* linear *control policy, $u^\star = U^\star \eta$ for some $U \in \mathcal{U}$.*

Theorem B extends the classical LQG insight – that *linear* output feedback policies are optimal – to our DRLQ model and *proves* that zero-mean (Gaussian) distributions are optimal for nature.

To appreciate the intuition behind this result, recall that under nature's optimal choice of Gaussian distribution $\mathbb{P}^\star$, the decision maker can restrict attention to an affine policy, $u = q + U\eta$. The proof of Theorem B shows that under the *optimal* choice of intercept q, the objective becomes a quadratic function of the means μ_z that is maximized by setting $\mu_z = 0$. This arises because nature can increase the costs achieved with a (Gaussian) distribution with nonzero mean μ_z by instead choosing a zero-mean Gaussian distribution and increasing the covariance matrix by $\mu_z \mu_z^\top$. This change – which leads to feasible distributions, by Assumption 3 – magnifies the noise and increases the decision maker's costs. In equilibrium, when nature uses zero-mean policies, the decision maker can restrict attention to *linear* – rather than *affine* – policies. This conveys a simple intuition: nature cannot gain by adding a predictable offset/bias to its distribution (through the mean), because that could be corrected by the decision maker through an appropriate choice of intercept q, so at optimality, neither nature nor the decision maker add predictable offsets to their actions.

The result in Theorem B is also important because it provides a novel rationale for considering zero-mean noise distributions. Whereas the vast majority of papers formulating robust LQG models restrict attention to zero-mean noise for simplicity or in keeping with the classical LQG setting, the result in Theorem B provides a different justification: this assumption/choice is *conservative*, because allowing the adversary to use zero means gives the adversary more power and results in the worst-case costs for the decision maker.

In view of the structural result in §B.5 and to simplify exposition, we assume throughout the rest of the paper that the mean of the noise under any distribution in the ambiguity set is 0, i.e., $\mathbb{E}_{\mathbb{P}_z}[z] = \mu_z = 0$ for all $z \in \mathcal{Z}$ and all $\mathbb{P}_z \in \mathcal{B}_z$.

B.6 Worst-Case Covariance Matrix

Our final structural result further develops the intuition above and shows that nature's optimal distribution $\mathbb{P}^\star$ entails suitably "inflating" the covariance matrix of the nominal distribution $\hat{\mathbb{P}}$. In view of the structural result in §B.5 and to simplify exposition, we assume throughout the rest of this section that the mean of the noise under any distribution in the ambiguity set is 0, i.e., $\mathbb{E}_{\mathbb{P}_z}[z] = \mu_z = 0$ for all $z \in \mathcal{Z}$ and all $\mathbb{P}_z \in \mathcal{B}_z$. Because any zero-mean Gaussian distribution $\mathbb{P}_z$ is fully specified through the second moment/covariance matrix $M_z^\star = \Sigma_z^\star$, we can shorten notation by using $\mathcal{M}_{\Sigma_z}$ to denote the sets $\mathcal{M}_{(\mu_z=0, M_z)}$ defined in Assumption 2.

Our final result requires a mild condition on the sets $\mathcal{M}_{\Sigma_z}$ that further refines Assumption 3.

Assumption 4. *Fix $\mu_z = \hat{\mu}_z = 0$ for all $z \in \mathcal{Z}$. There exists $g : \mathbb{S}_+^{d_z} \to \mathbb{R}$ such that the sets defined in Assumption 2 can be represented as $\mathcal{M}_{\Sigma_z} = \{\Sigma_z \in \mathbb{S}_+^{d_z} \,:\, g(\Sigma_z) \leq 0\}$ for any $z \in \mathcal{Z}$, where g is convex on $\mathbb{S}_+^{d_z}$, differentiable on $\mathbb{S}_{++}^{d_z}$, and satisfies: (i) $g(\hat{\Sigma}_z) < 0$, (ii) $\hat{\Sigma}_z \in \operatorname{argmin}_{\Sigma_z \succeq 0} g(\Sigma_z)$, and (iii) $\nabla g(\Sigma_1) \succeq \nabla g(\Sigma_2)$ implies $\Sigma_1 \succeq \Sigma_2$, for any $\Sigma_1, \Sigma_2 \in \mathbb{S}_+^{d_z}$.*

Proposition 10 in Appendix §H.4.1 shows that Assumption 4 is satisfied by all examples in §A.2 if $\rho_z > 0$ (and, for the case of the moment-based ambiguity, if the divergence $\mathbb{M}$ satisfies a mild condition). The key requirements are intuitive if one takes $g(\Sigma_z)$ as a (convex, increasing) transformation of the distance $\mathbb{D}(\mathbb{P}_z, \hat{\mathbb{P}}_z)$ between a distribution $\mathbb{P}_z = \mathcal{N}(0, \Sigma_z)$ and the nominal $\hat{\mathbb{P}}_z = \mathcal{N}(0, \hat{\Sigma}_z)$. Requirement (i) simply asks that $\hat{\Sigma}_z$ is an interior point of the set of valid covariances $\mathcal{M}_z$, which holds in all our examples provided there is ambiguity, $\rho_z > 0$. Requirement (ii) is readily satisfied because $\mathbb{D}(\mathbb{P}_z, \hat{\mathbb{P}}_z) \geq \mathbb{D}(\hat{\mathbb{P}}_z, \hat{\mathbb{P}}_z)$ for any divergence $\mathbb{D}$. Lastly, (iii) asks that the gradient map ∇g be order-reflecting, i.e., that gradients that dominate in the Loewner (positive semidefinite) order should correspond to (distributions whose) covariances also dominate in the Loewner order. Put more intuitively, this means that "noisier" gradient maps should come from "noisier" distributions.

Theorem C. *Under Assumptions 1-4, problem* (A.9) *admits an optimal solution* $\mathbb{P}^\star$ *such that* $\mathbb{P}^\star_z = \mathcal{N}(0, \Sigma^\star_z)$ *and the optimal covariance satisfies* $\Sigma^\star_z \succeq \hat{\Sigma}_z$*, for every* $z \in \mathcal{Z}$*.*

Theorem C offers a clear and powerful insight: when nature selects a zero-mean Gaussian distribution that maximizes the cost, its optimal strategy is to *inflate* the nominal covariance matrix $\hat{\Sigma}$, so that the optimal covariance $\Sigma^\star$ dominates $\hat{\Sigma}$ in the Loewner (positive semidefinite) ordering. This result generalizes a familiar principle – that increasing a random variable's variance creates "more uncertainty" – to the significantly more complex, dynamic environment provided by the LQG. This also suggests an important practical take-away: in LQG problems where model ambiguity is a concern, a simple yet effective rule of thumb is to inflate the covariance matrix of the noise, which will provide additional protection against uncertainty.

C Efficient Numerical Solution of DRLQ Problems

Our duality results in Theorem A lead to immediate algorithms for computing optimal strategies: the optimal value in the DRLQ problem is the same as the optimal value in problems (A.12) and (A.14), and the latter problems are finite-dimensional, convex-concave problems with smooth objectives, which are amenable to saddle-point methods (Juditsky and Nemirovski, 2022; Schiele et al., 2024). However, such approaches would fail to exploit the temporal structure of the original control problem and would generally result in large-dimensional optimization problems.

We next leverage our results from §B to develop a set of more efficient algorithms to solve the DRLQ problem via Kalman filtering and DP techniques. Without loss of generality, we consider a zero-mean nominal distribution with $\hat{\mu}_z = 0$ for any $z \in \mathcal{Z}$. Our algorithms will rely on all structural results from §B, as formalized in the next result.

Corollary 3 (Solving DRLQ via Kalman Filtering)**.** *Under Assumptions 1-4, the solution to the DRLQ problem* (A.8) *can be computed using the Kalman filtering techniques in Appendix §F applied to a classical LQG model with distribution* $\mathbb{P}^\star$*. Specifically, the optimal control policy is* $u^\star_t = K_t \hat{x}_t$ *for every* $t \in [T-1]$*, where* K_t *is the optimal state-feedback gain given by* (A.32b) *and* $\hat{x}_t$ *is obtained using the Kalman filter recursions* (A.34) *corresponding to distribution* $\mathbb{P}^\star$*.*

This follows from the theorems in §B, but it is worth emphasizing that *all* those structural results are needed. Specifically, Theorem B implies that the optimal linear policy $u^\star$ can be found by solving a *classical* LQG problem corresponding to the (unknown, optimal) zero-mean Gaussian distribution $\mathbb{P}^\star$. However, to solve this problem with the Kalman filter recursions requires the covariance matrices of all noise terms v_t under $\mathbb{P}^\star$ to be positive *definite*. Assumption 1 only requires this for the nominal distribution, $\hat{\Sigma}_{v_t} \succ 0$, but it is Theorem C that ensures that $\Sigma^\star_{v_t} \succeq \hat{\Sigma}_{v_t} \succ 0$.

We design an iterative algorithm to compute $\mathbb{P}^\star$. $\mathbb{P}^\star$ is uniquely determined by the covariance matrices $\Sigma^\star_w \in \mathcal{M}_{\Sigma_w}$ and $\Sigma^\star_v \in \mathcal{M}_{\Sigma_v}$, chosen to maximize the objective in (A.14). Moreover, we can replace $\mathcal{M}_{\Sigma_v}$ with $\mathcal{M}^+_{\Sigma_v} = \{\Sigma_{v_t} \in \mathcal{M}_{\Sigma_{v_t}} : \Sigma_{v_t} \succeq \lambda I\}$[2] for a small $\lambda > 0$, which is without loss of optimality due to Theorem C.

[2]For instance, λ can be set as the minimum eigenvalue of $\hat{\Sigma}_v$.

We first reformulate (A.14) as

$$\max_{\Sigma_w \in \mathcal{M}_{\Sigma_w}, \Sigma_v \in \mathcal{M}^+_{\Sigma_v}} f(\Sigma_w, \Sigma_v), \tag{A.16}$$

where $f(\Sigma_w, \Sigma_v)$ denotes the optimal value of the classical LQG problem corresponding to the Gaussian distribution $\mathbb{P}$ with covariance matrices Σ_w and Σ_v. We propose a Frank-Wolfe algorithm for solving problem (A.16) (see Frank and Wolfe, 1956; Levitin and Polyak, 1966). Because this requires certain smoothness properties for f (Jaggi, 2013), we first provide a structural result.

Proposition 4. *Under Assumptions 1-C, f is concave and β-smooth on $\mathcal{M}_{\Sigma_w} \times \mathcal{M}^+_{\Sigma_v}$.*

For a proof, see Appendix §I.1. This allows using a Frank-Wolfe algorithm to solve problem (A.16). Each iteration of this algorithm solves a direction-finding subproblem, that is, a variant of problem (A.16) that maximizes the first-order Taylor expansion of f around the current iterates $(\Sigma_w^{(k)}, \Sigma_v^{(k)})$:

$$\max_{\Sigma_w \in \mathcal{M}_{\Sigma_w}, \Sigma_v \in \mathcal{M}^+_{\Sigma_v}} \left\langle \nabla_{\Sigma_w} f(\Sigma_w^{(k)}, \Sigma_v^{(k)}), \Sigma_w - \Sigma_w^{(k)} \right\rangle + \left\langle \nabla_{\Sigma_v} f(\Sigma_w^{(k)}, \Sigma_v^{(k)}), \Sigma_v - \Sigma_v^{(k)} \right\rangle. \tag{A.17}$$

The next iterates are obtained by moving towards a maximizer $(\Sigma_w^\star, \Sigma_v^\star)$ of (A.17), i.e., we update

$$(\Sigma_w^{(k+1)}, \Sigma_v^{(k+1)}) \leftarrow (\Sigma_w^{(k)}, \Sigma_v^{(k)}) + \alpha \cdot (\Sigma_w^\star - \Sigma_w^{(k)}, \Sigma_v^\star - \Sigma_v^{(k)}),$$

where α is an appropriate step size. The proposed Frank-Wolfe algorithm enjoys a very low per-iteration complexity because problem (A.17) is separable. To see this, we reformulate (A.17) as:

$$\begin{aligned} \max_{\{\Sigma_z\}_{z \in \mathcal{Z}}} \quad & \sum_{z \in \mathcal{Z}} \left\langle \nabla_{\Sigma_z} f\big(\Sigma_w^{(k)}, \Sigma_v^{(k)}\big), \Sigma_z - \Sigma_z^{(k)} \right\rangle \\ \text{s.t.} \quad & \Sigma_z \in \mathcal{M}_{\Sigma_z} \ \forall z \in \mathcal{Z} \\ & \Sigma_z \in \mathcal{M}^+_{\Sigma_z} \ \forall z \in \{v_0, \ldots, v_{T-1}\}. \end{aligned} \tag{A.18}$$

Hence, (A.17) decomposes into $|\mathcal{Z}| = 2T + 1$ separate subproblems that can be solved in parallel.

Remark 1. *Although the subproblems above can be reformulated as tractable SDPs that are amenable to off-the-shelf solvers, for specific divergences $\mathbb{D}$ one may be able to further simplify this computation. For instance, for the Wasserstein and KL ambiguity sets from §A.2 (corresponding to divergences $\mathbb{W}$ and $\mathbb{K}$, respectively), the optimization problem in (A.18) for a given $z \in \mathcal{Z}$ can be reduced to solving a univariate algebraic equation, which can be done to any desired accuracy $\delta > 0$ by an efficient bisection algorithm. Appendix §I.2 provides details for this construction and a proof of all relevant results, that leverage existing results in the literature.*

Remark 2 (Automatic differentiation)**.** *Recall that $f(\Sigma_w, \Sigma_v)$ is the optimal value of the LQG problem corresponding to the Gaussian distribution $\mathbb{P}$ with the covariance matrices Σ_w and Σ_v. By using the underlying dynamic programming equations, $f(\Sigma_w, \Sigma_v)$ can thus be expressed in closed form as a serial composition of $\mathcal{O}(T)$ rational functions (see Appendix §F for details). Hence, $\nabla_{\Sigma_z} f(\Sigma_w, \Sigma_v)$ can be calculated symbolically for any $z \in \mathcal{Z}$ by repeatedly applying the chain and product rules. However, the resulting formulas are lengthy and cumbersome. We thus compute the gradients numerically using backpropagation. The cost of evaluating $\nabla_{\Sigma_z} f(\Sigma_w, \Sigma_v)$ is then of the same order of magnitude as the cost of evaluating $f(\Sigma_w, \Sigma_v)$.*

A detailed description of the proposed Frank-Wolfe method is given in Algorithm A.1 below.

By Theorem 1 and Lemma 7 in Jaggi (2013), which apply in view of Proposition 4, Algorithm A.1 attains a suboptimality gap of ϵ within $\mathcal{O}(1/\epsilon)$ iterations. Its precise computational complexity is critically dependent on the tractability of Line 6 (this is very efficient for Wasserstein and KL ambiguity, by Remark 1).

D Numerical Experiments

We conduct numerical experiments to assess the merits of the DRLQ model and the effectiveness of the algorithms introduced in §C. We consider a class of dynamical systems with $n = m = p = d$

Algorithm A.1 Frank-Wolfe algorithm for solving (A.16)

Input: initial iterates $\Sigma_w^{(0)}, \Sigma_v^{(0)}$, nominal covariance matrices $\hat{\Sigma}_w, \hat{\Sigma}_v$, oracle precision $\delta \in (0, 1)$
set initial iteration counter $k = 0$
while stopping criterion is not met **do**
    for $z \in \mathcal{Z}$ **do in parallel**
        compute $\nabla_{\Sigma_z} f(\Sigma_w^{(k)}, \Sigma_v^{(k)})$
        compute $\Sigma_z^\star$ that solves (A.18) to precision δ
    end
    $(\Sigma_w^{(k+1)}, \Sigma_v^{(k+1)}) \leftarrow (\Sigma_w^{(k)}, \Sigma_v^{(k)}) + 2/(2+k) \cdot (\Sigma_w^\star - \Sigma_w, \Sigma_v^\star - \Sigma_v)$
    $k \leftarrow k + 1$
end while
Output: $\Sigma_w^{(k)}$ and $\Sigma_v^{(k)}$

with $d \in \mathbb{N}_+$. We set $A_t = A$ where A has 0.1 on the main diagonal and the super-diagonal and zeroes elsewhere ($A_{i,j} = 0.1$ if $i = j$ or $i = j - 1$ and $A_{i,j} = 0$ otherwise), and the other matrices to $B_t = C_t = Q_t = R_t = I_d$. The nominal covariance matrices $\hat{\Sigma}_z$ are constructed randomly and with eigenvalues in the interval $[1, 2]$ (to ensure they are positive definite). We construct ambiguity sets based on either Wasserstein or KL divergence.

All experiments were conducted on an Apple M3 Max machine equipped with 64 GB of RAM. All linear SDP problems were formulated in Python (v3.8.6) using CVXPY (v1.6.1) (Agrawal et al., 2018; Diamond and Boyd, 2016) and solved with MOSEK (MOSEK ApS, 2019) (v11.0.8). Additionally, the gradients of $f(\Sigma_w, \Sigma_v)$ were computed using Pymanopt (v2.2.1) (Townsend et al., 2016) in combination with PyTorch's automatic differentiation module (v2.6.0) (Paszke et al., 2017, 2019). The code is publicly available in the Github repository `https://github.com/BaharTaskesen/DRLQG`.

D.1 Computational Efficiency of Algorithm A.1

We compare two approaches for finding the optimal value of the DRLQ problem (A.8): directly solving the SDP reformulation of (A.14) with MOSEK and using our custom Frank-Wolfe method discussed in Algorithm A.1. For these experiments, we set $d = 10$ and the corresponding radii of the ambiguity sets as $\rho_{x_0} = \rho_{w_t} = \rho_{v_t} = 10^{-1}$. We compare the two approaches in 10 problem instances (generated randomly and independently) and we compare performance as a function of the problem horizon T, which we vary. We set a stopping criterion corresponding to an optimality gap below 10^{-3} and we run the Frank-Wolfe method with $\delta = 0.95$.

Figure A.1 and Figure A.2, which correspond to the Wasserstein and KL ambiguity sets, respectively, depict the results. In both figures, the left panel illustrates the execution time for both approaches as a function of the planning horizon T (omitting runs where MOSEK exceeds 100s) and the right panel visualizes the empirical convergence behavior of the Frank-Wolfe algorithm. The results highlight that the Frank-Wolfe algorithm achieves running times that are uniformly lower than MOSEK across all problem horizons and is able to find highly accurate solutions already after a small number of iterations (50 iterations for problem instances with horizon $T = 10$).

D.2 Worst-case Performance

We next evaluate the benefits of the robust approach. Let $u^\star$ denote the *robustly optimal* policy, i.e., the optimal policy in the DRLQ model, obtained by solving problem (A.8), and let $\hat{u}$ denote the *nominally optimal* policy, i.e., the policy that minimizes the expected cost under the nominal distribution, $\mathbb{E}_{\hat{\mathbb{P}}}[J(u)]$. To gauge the robustness and conservativeness of the policies $u^\star$ and $\hat{u}$, we compare them under the nominal distribution and under their respective worst-case distributions. Specifically, letting $\mathbb{P}^\star(u)$ denote the adversary's optimal choice of distribution $\mathbb{P}$ corresponding to a control policy u and letting $J(u)$ denote the dependency of the cost on the control policy u, we

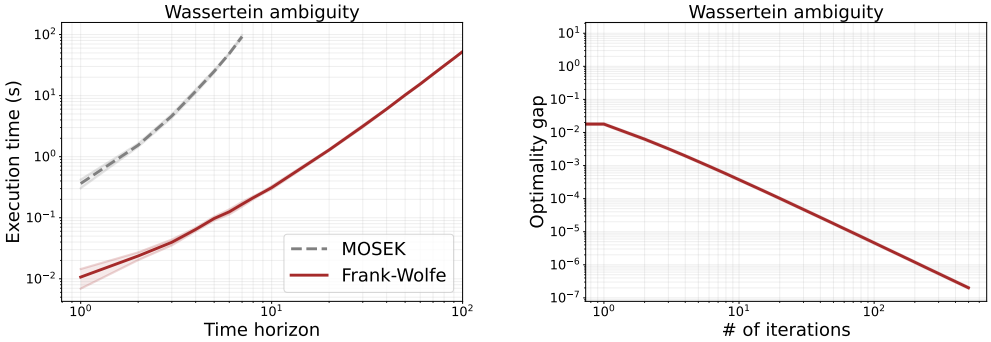

Figure A.1: Computation time (Left) and optimality gap (Right) of the Frank-Wolfe algorithm for Wasserstein ambiguity.

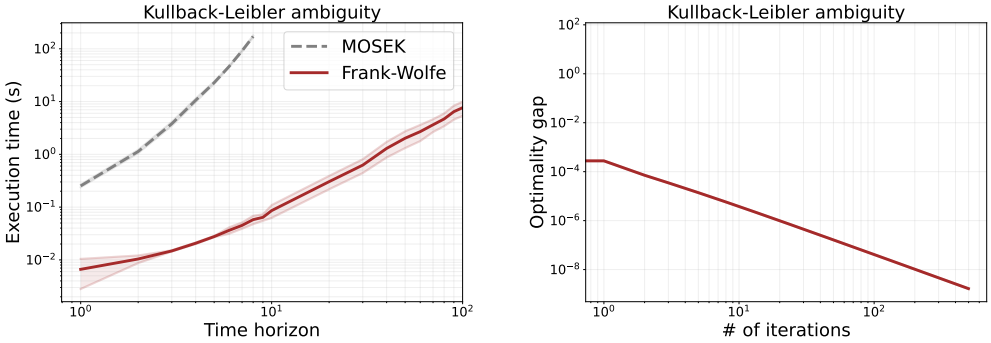

Figure A.2: Computation time (Left) and optimality gap (Right) of the Frank-Wolfe algorithm for Kullback-Leibler ambiguity.

calculate the following two performance gaps:

$$\text{Worst-case-gap} = \mathbb{E}_{\mathbb{P}^\star(\hat{u})}[J(\hat{u})] - \mathbb{E}_{\mathbb{P}^\star(u^\star)}[J(u^\star)], \qquad \text{Nominal-gap} = \mathbb{E}_{\hat{\mathbb{P}}}[J(u^\star)] - \mathbb{E}_{\hat{\mathbb{P}}}[J(\hat{u})].$$

In our experiments, we set $d = 2$, $T = 2$, $\rho_{x_0} = \rho_{w_t} = \rho_{v_t} = \rho$, and we vary the common radius ρ from 0 to 10, calculating the "Worst-case-gap" and "Nominal-gap" for each value of ρ in 10 independently generated random problem instances. The results are depicted in Figure A.3, with the left panel corresponding to the Wasserstein ambiguity set and the right panel corresponding to the KL ambiguity set. The results indicate that using the optimal policy from the DRLQ model, $u^\star$, leads to dramatically lower worst-case costs, particularly as the ambiguity radius ρ increases. Surprisingly, this improvement does not impact performance in the nominal scenario, where using $u^\star$ does not substantially increase costs relative to using the nominally optimal policy $\hat{u}$.

E Infinite-Horizon DRLQ Problems

We now extend the results of §B to infinite-horizon control problems with an average cost criterion. Throughout this section, we restrict attention to linear *time-invariant* systems of the form (1) and (2) where $T = \infty$ and $A_t = A_0$, $B_t = B_0$ and $C_t = C_0$ for all $t \in \mathbb{N}$. All random variables emerging in our model are functions of the initial state x_0 and the noise terms $\{w_t\}_{t=0}^\infty$ and $\{v_t\}_{t=0}^\infty$. Therefore, we set the sample space to $\Omega = \mathbb{R}^n \times (\times_{t=0}^\infty(\mathbb{R}^n \times \mathbb{R}^p))$ and equip it with its product Borel σ-algebra $\mathcal{F}$. In analogy to the finite-horizon theory, we define $x = (x_t)_{t=0}^\infty$, $u = (u_t)_{t=0}^\infty$, $y = (y_t)_{t=0}^\infty$, $w = (x_0, (w_t)_{t=0}^\infty)$ and $v = (v_t)_{t=0}^\infty$ as well as infinite-dimensional block matrices H, G and C (whose definitions mirror those for the finite horizon and are omitted for brevity). With these conventions, the input, state and output processes are subject to the usual system equations

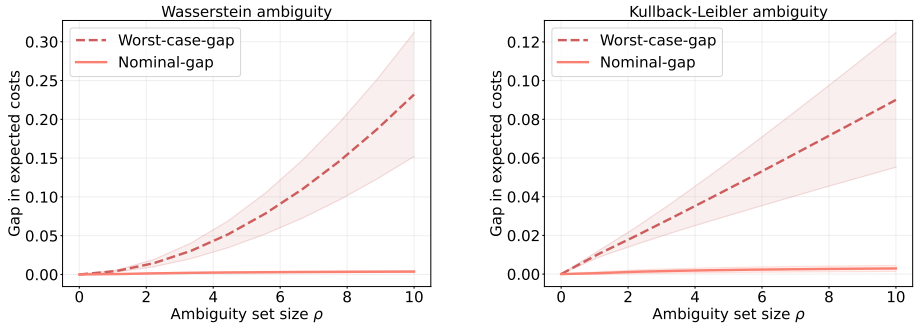

Figure A.3: The difference in worst-case expected costs when using the nominally optimal policy $\hat{u}$ instead of the robustly optimal policy $u^\star$ ("Worst-case-gap") and the difference in expected costs under the nominal distribution $\hat{\mathbb{P}}$ when using the robustly optimal policy $u^\star$ instead of the nominally optimal policy $\hat{u}$ ("Nominal-gap"), as a function of the radius of the ambiguity sets ρ, for Wasserstein ambiguity (Left) and KL ambiguity (Right). The solid line is the mean and the confidence bands correspond to one standard deviation, estimated from 10 independent runs.

$x = Hu + Gw$ and $y = Cx + v$. As before, we denote by $\mathcal{U}_y$ the set of control inputs u such that $u_t = \varphi_t(y_0, \ldots, y_t)$ for every $t \in \mathbb{N}$, where $\varphi_t : \mathbb{R}^{p(t+1)} \to \mathbb{R}^m$ is a measurable control policy. To describe the distribution of the exogenous uncertainties, it is again convenient to set $\mathcal{Z} = \{x_0, w_0, v_0, w_1, v_1, \ldots\}$. We then define $\mathcal{B}^\infty$ exactly as the ambiguity set $\mathcal{B}$ from §A.

For costs, we take $Q_t = Q_0$ and $R_t = R_0$ for all $t \in \mathbb{N}$, where $Q_0 \in \mathbb{S}^n_+$ and $R_0 \in \mathbb{S}^m_{++}$. We study an infinite-horizon DRLQ problem that minimizes the worst-case, long-run average cost:

$$J_{\mathbb{P}}(x, u) = \limsup_{T \to \infty} \frac{1}{T} \sum_{t=0}^{T-1} \mathbb{E}_{\mathbb{P}} \left[x_t^\top Q_0 x_t + u_t^\top R_0 u_t \right],$$

where the worst case is taken over all $\mathbb{P} \in \mathcal{B}^\infty$.

E.1 Assumptions

To ensure that the problem is well posed under an infinite-horizon setting and average-cost criterion, we must slightly strengthen the assumptions in §A.1 and make a few additional mild assumptions.

First, we mirror the classical infinite-horizon LQG setting by stating an assumption on system and cost matrices. Recall that we already focus on time-invariant systems and costs. To state our additional requirement, we recall the following definition of a Schur-stable matrix.

Definition 3. *A square matrix $A \in \mathbb{R}^{n \times n}$ is said to be* Schur *stable if its eigenvalues are strictly within the unit circle in $\mathbb{C}$.*

If A is a Schur stable matrix, then $A^t \to 0$ as $t \to \infty$ and the system defined through the dynamics $z_{t+1} = Az_t$ for $t \in \mathbb{N}_+$ is stable, that is, $z_t \to 0$ as $t \to \infty$ (Bertsekas, 2017, Page 135).

Subsequently, we impose the following assumption on the system and cost matrices.

Assumption 5. *The matrices A_0, B_0, C_0 and Q_0 satisfy the following properties.*
(i) (A_0, B_0) is stabilizable, i.e., there exists $K \in \mathbb{R}^{m \times n}$ such that $A_0 + B_0 K$ is Schur stable.
(ii) (A_0, C_0) is detectable, i.e., there exists $L \in \mathbb{R}^{n \times p}$ such that $A_0 - LC_0$ is Schur stable.
(iii) $Q_0 \succ 0$.

These assumptions are standard in the context of infinite-horizon control of linear dynamical systems (see Lancaster and Rodman, 1995; Bertsekas, 2017, and our brief review in §F.4). *Stabilizability* guarantees the existence of a stationary state-feedback control policy that makes the states of a noise-free system converge and allows deriving an optimal state-feedback control policy of the form

$u_t^\star = K\hat{x}_t$, where $\hat{x}_t$ denotes the MMSE state estimator (see §F.4). However, even if u_t stabilizes the noise-free system, the states under linear (purified) output feedback could diverge for large t without the *detectability* assumption and without a strictly positive definite cost matrix Q_0.[3]

We also strengthen slightly our assumptions concerning the nominal distribution $\hat{\mathbb{P}}$ by requiring this to be zero-mean Gaussian and also time-invariant.

Definition 4. *We call a probability distribution* $\mathbb{P} \in \mathcal{B}^\infty$ time-invariant *if there exist* $\Sigma_w \in \mathbb{S}_+^n$ *and* $\Sigma_v \in \mathbb{S}_+^p$ *such that* $\mathbb{E}_\mathbb{P}[x_0 x_0^\top] = \Sigma_w$ *and* $\mathbb{E}_\mathbb{P}[w_t w_t^\top] = \Sigma_w$ *and* $\mathbb{E}_\mathbb{P}[v_t v_t^\top] = \Sigma_v$ *for all* $t \in \mathbb{N}$.

Subsequently, we use $\mathcal{B}_\mathcal{N}^\infty$ to denote the subset of all time-invariant Gaussian distributions in $\mathcal{B}^\infty$. Throughout this section, we therefore replace Assumption 1 with the following assumption.

Assumption 6. *The nominal distribution* $\hat{\mathbb{P}}$ *and the ambiguity set* $\mathcal{B}^\infty$ *satisfy the requirements:*
(i) $\hat{\mathbb{P}}$ *is a time-invariant Gaussian distribution,* $\hat{\mathbb{P}} \in \mathcal{B}_\mathcal{N}^\infty$*, with* $\hat{\mu}_z = 0$ *and* $\hat{\Sigma}_z \succ 0$ *for all* $z \in \mathcal{Z}$*;*
(ii) $\rho_z = \rho_w$ *for all* $z \in \{x_0\} \cup \{w_t\}_{t=0}^\infty$ *and* $\rho_z = \rho_v$ *for all* $z \in \{v_s\}_{t=0}^\infty$*;*
(iii) all distributions in $\mathcal{B}^\infty$ *have zero mean,* $\mathbb{E}_{\mathbb{P}_z}[z] = 0$ *for every* $z \in \mathcal{Z}$ *and every* $\mathbb{P} \in \mathcal{B}^\infty$.

The requirements in (i) concerning the nominal distribution are standard in infinite-horizon LQG control problems (Lancaster and Rodman, 1995; Bertsekas, 2017). Under a time-invariant, Gaussian noise distribution $\hat{\mathbb{P}}$, these requirements are only slightly stronger than those in Assumption 1, by asking that the covariance matrices for the state noise be positive definite, $\hat{\Sigma}_{x_0} \succ 0$ and $\hat{\Sigma}_{w_t} \succ 0$ for all $t \in \mathbb{N}$. Requirement (ii) is also aligned with time-invariance by asking that the corresponding ambiguity sets have the same radius. Lastly, the restriction to zero-mean distributions in (iii) simplifies exposition and is driven by our results in §B.5; this requirement can be relaxed and arguments mirroring those in §B.5 can be used to prove that nature will optimally choose zero-mean distributions, but we omit details for brevity and instead simply state this as a requirement.

Throughout this section, Assumption 2 remains unchanged. Note that in view of Assumption 6, the sets of moments defined in Assumption 2 only involve the covariance matrices Σ_z (as in §B.6 and §C) and are also time-invariant, so we define the following simpler notation:

$$\begin{aligned} \mathcal{M}_{\Sigma_w} &= \{\Sigma \in \mathbb{S}_+^n : \mathbb{D}(\mathcal{N}(0,\Sigma), \hat{\mathbb{P}}_z) \leq \rho_w\}, \quad \forall\, z \in \{x_0\} \cup \{w_t\}_{t=0}^\infty \\ \mathcal{M}_{\Sigma_v} &= \{\Sigma \in \mathbb{S}_+^p : \mathbb{D}(\mathcal{N}(0,\Sigma), \hat{\mathbb{P}}_z) \leq \rho_v\}, \quad \forall\, z \in \{v_s\}_{t=0}^\infty. \end{aligned} \tag{A.19}$$

By Assumption 2, the sets $\mathcal{M}_{\Sigma_w}$ and $\mathcal{M}_{\Sigma_v}$ are convex and compact.

Lastly, we preserve Assumption 4 suitably generalized to our infinite-horizon setting, so that

$$\mathcal{M}_{\Sigma_z} = \{\Sigma \in \mathbb{S}_+^{d_z} : g(\Sigma_z) \leq 0\},\ \forall z \in \mathcal{Z}. \tag{A.20}$$

where g is a convex, differentiable function.

E.2 Construction of Primal and Dual and Their Bounds

With these preliminaries, the DRLQ problem can be formulated as:

$$p^\star = \left\{ \begin{array}{l} \inf\limits_{x,u,y} \ \sup\limits_{\mathbb{P}\in\mathcal{B}^\infty} \ J_\mathbb{P}(x,u) \\ \text{s.t.} \ \ u \in \mathcal{U}_y, \ \ x = Hu + Gw, \ \ y = Cx + v. \end{array} \right.$$

In analogy to §B, we define $\eta = (\eta_t)_{t=0}^\infty$ as the trajectory of all purified observations that satisfies $\eta = Dw + v$, where $D = CG$, and $\mathcal{U}_\eta$ as the set of all control inputs u so that $u_t = \phi_t(\eta_0, \ldots, \eta_t)$ for some measurable function $\phi_t : \mathbb{R}^{p(t+1)} \to \mathbb{R}^m$ for every $t \in \mathbb{N}_+$. By (Hadjiyiannis et al., 2011, Proposition II.1), we can rewrite the infinite-horizon DRLQ problem equivalently as

$$p^\star = \left\{ \begin{array}{l} \inf\limits_{x,u} \ \sup\limits_{\mathbb{P}\in\mathcal{B}^\infty} \ J_\mathbb{P}(x,u) \\ \text{s.t.} \ \ u \in \mathcal{U}_\eta, \ \ x = Hu + Gw. \end{array} \right. \tag{A.21}$$

[3]This suggests that the stability properties of a system are more difficult to analyze when the control policy is parametrized in terms of the purified outputs. In contrast, the convexity properties of the system are more difficult to analyze when policies are parametrized in terms of the original outputs.

Our results also rely on the dual of (A.21) defined as

$$d^\star = \left\{ \begin{array}{lll} \sup\limits_{\mathbb{P}\in\mathcal{B}^\infty} & \inf\limits_{x,u} & J_{\mathbb{P}}(x,u) \\ & \text{s.t.} & u \in \mathcal{U}_\eta, \;\; x = Hu + Gw. \end{array} \right. \tag{A.22}$$

In the remainder of this section, we demonstrate that, under mild assumptions, the primal DRLQ problem is solved by a *stationary, linear* control policy, the dual problem is solved by a *time-invariant, Gaussian* distribution, and strong duality holds. To establish these results, we proceed as in §B: we first construct an upper bound for the primal problem (A.21) followed by a lower bound for the dual problem (A.22), and then show that the bounds coincide, which proves all three claims.

Upper Bound for Primal

Mirroring §B, we obtain an upper bound on $p^\star$ by inflating the ambiguity set $\mathcal{B}^\infty$ and restricting the control policies. We define the inflated ambiguity set $\overline{\mathcal{B}}^\infty$ exactly as the ambiguity set $\overline{\mathcal{B}}$ from Section B, but we also add the additional condition that $\mathbb{E}_{\mathbb{P}_z}[z] = 0$ for every $z \in \mathcal{Z}$ to mirror the new Assumption 6-(ii) on $\mathcal{B}^\infty$. For any divergence $\mathbb{D}$ satisfying Assumption 2-(i), we can readily verify that $\mathcal{B}^\infty$ is a subset of $\overline{\mathcal{B}}^\infty$. Restricting the control policies u requires more care in the infinite-horizon case, to ensure that long-run-average costs are finite and to enable our subsequent duality proof. To that end, we restrict attention (without loss of optimality) to a *subset* of the set of all *stationary* purified output control policies $u \in \mathcal{U}_\eta$ under which the covariance matrices of controls u_t and states x_t converge, which also guarantees that the long-run average costs converge to a finite value. To formalize these, we first define the set of block lower triangular Toeplitz matrices.

Definition 5 (Toeplitz Matrices)**.** *An infinitely long block matrix M with blocks of size $k \times l$ is called a block lower triangular Toeplitz matrix if there exist $M_t \in \mathbb{R}^{k\times l}$ for all $t \in \mathbb{N}$ so that*

$$M = \begin{bmatrix} M_0 & & & \\ M_1 & M_0 & & \\ M_2 & M_1 & M_0 & \\ \vdots & & & \ddots \end{bmatrix}.$$

All blocks of the t-th subdiagonal of M are given by the same matrix $M_t \in \mathbb{R}^{k\times l}$, so M can be uniquely identified by its blocks $\{M_t\}_{t=0}^\infty$. The family of all block lower triangular Toeplitz matrices with blocks of size $k \times l$ is denoted by $\mathcal{T}^{k\times l}$. We also define the norm of an infinitely long Toeplitz matrix $M \in \mathcal{T}^{k\times l}$ as $\|M\|_{\mathcal{T}} = (\sum_{t=0}^\infty \|M_t\|_{\mathrm{F}}^2)^{1/2}$, where $\|M_t\|_{\mathrm{F}}$ stands for the Frobenius norm of M_t.

The following lemma summarizes useful structural properties of Toeplitz matrices that will enable us to write compact expressions and study the properties of our control policies.

Lemma A. *The following properties hold for block lower triangular Toeplitz matrices:*
i) If $M \in \mathcal{T}^{k\times l}$ and $N \in \mathcal{T}^{k\times l}$, then $O = M + N \in \mathcal{T}^{k\times l}$ with $O_t = M_t + N_t \in \mathbb{R}^{k\times l}$ for all $t \in \mathbb{N}$.
(ii) If $M \in \mathcal{T}^{k\times l}$ and $N \in \mathcal{T}^{l\times m}$, then $O = MN \in \mathcal{T}^{k\times m}$ with $O_t = \sum_{s=0}^t M_s N_{t-s} \in \mathbb{R}^{k\times m}$ $\forall t \in \mathbb{N}$.
(iii) If $M \in \mathcal{T}^{k\times k}$ with M_0 invertible, then $N = M^{-1} \in \mathcal{T}^{k\times k}$, and its blocks obey the recursion

$$N_0 = M_0^{-1} \quad \textit{and} \quad N_t = -M_0^{-1} \sum_{s=1}^t M_s N_{t-s} \quad \forall t \in \mathbb{N}. \tag{A.23}$$

Subsequently, for a Toeplitz matrix M obtained by adding or multiplying Toeplitz matrices, we use the $(M)_t$ to denote the block matrix on the t-th subdiagonal of M.

With this preparation, we can formally define the class of *stationary* control policies.

Definition 6. *We call a linear output feedback policy $u \in \mathcal{U}_y$* stationary *if there exists $U' \in \mathcal{T}^{m\times p}$ such that $u = U'y$. Similarly, we call a linear* purified *output feedback policy $u \in \mathcal{U}_\eta$* stationary *if there exists $U \in \mathcal{T}^{m\times p}$ such that $u = U\eta$.*

Subsequently, we focus on *stationary, linear, purified output* feedback policies, i.e., $u \in \mathcal{U}_\eta$ such that $u = U\eta$ for some $U \in \mathcal{T}^{m\times p}$. (Through a straightforward extension of results in Lemma E, one can

verify that these are equivalent to stationary linear output feedback policies.) Under such policies, the following result – which leverages Lemma A – yields compact expressions for several important quantities.

Lemma B. *Consider $u \in \mathcal{U}_\eta$ such that $u = U\eta$ for $U \in \mathcal{T}^{m\times p}$. For any $t \in \mathbb{N}$, we have:*

$$u_t = \sum_{s=0}^{t} [(UD)_{t-s} w_s + (U)_{t-s} v_s] \quad \textit{and} \quad x_t = \sum_{s=0}^{t} [(G + HUD)_{t-s} w_s + (HU)_{t-s} v_s] \tag{A.24a}$$

$$\Sigma_{u_t} = \mathbb{E}_{\mathbb{P}}\left[u_t u_t^\top\right] = \sum_{s=0}^{t} \left((UD)_{t-s}\Sigma_{w_s}(UD)_{t-s}^\top + (U)_{t-s}\Sigma_{v_s}(U)_{t-s}^\top\right) \tag{A.24b}$$

$$\Sigma_{x_t} = \mathbb{E}_{\mathbb{P}}\left[x_t x_t^\top\right] = \sum_{s=0}^{t} \left((G+HUD)_{t-s}\Sigma_{w_s}(G+HUD)_{t-s}^\top + (HU)_{t-s}\Sigma_{v_s}(HU)_{t-s}^\top\right) \tag{A.24c}$$

$$\frac{1}{T}\sum_{t=0}^{T-1} \mathbb{E}_{\mathbb{P}}\left[x_t^\top Q_0 x_t + u_t^\top R_0 u_t\right] = \frac{1}{T}\sum_{s=0}^{T-1} \mathrm{Tr}\left(\Sigma_{w_s}\left(\sum_{t=0}^{T-1-s} M_t\right) + \Sigma_{v_s}\left(\sum_{t=0}^{T-1-s} N_t\right)\right), \tag{A.24d}$$

where

$$M_t = (UD)_t^\top R_0 (UD)_t + (G+HUD)_t^\top Q_0 (G+HUD)_t \textit{ and } N_t = (U)_t^\top R_0 (U)_t + (HU)_t^\top Q_0 (HU)_t.$$

The expressions in Lemma B for the covariances $\Sigma_{u_t}, \Sigma_{x_t}$ and the average cost reveal that additional requirements are needed on the stationary control policy u to ensure convergence and finite costs. To that end, in deriving the upper bound on $p^\star$, we further restrict the control policies in the feasible set of problem (A.21) to stationary linear policies $u = U\eta$ where U belongs to the set:

$$\mathcal{U}_\infty = \left\{U \in \mathcal{T}^{m\times p} : \|U\|_{\mathcal{T}} < \infty, \|UD\|_{\mathcal{T}} < \infty, \|G + HUD\|_{\mathcal{T}} < \infty, \|HU\|_{\mathcal{T}} < \infty\right\}.$$

Note that $\mathcal{U}_\infty$ is a convex set because $\mathcal{T}^{m\times p}$ is a linear space and the norm $\|\cdot\|_{\mathcal{T}}$ is convex.

With this, we now formalize the upper bound on $p^\star$ by inflating the ambiguity set to $\overline{\mathcal{B}}^\infty$ and using $\mathcal{U}_\infty$ to construct the stationary policies. We obtain the following upper bound:

$$\overline{p}^\star = \begin{cases} \min\limits_{U,x,u} & \max\limits_{\mathbb{P}\in\overline{\mathcal{B}}^\infty} \; J_{\mathbb{P}}(x,u) \\ \text{s.t.} & U \in \mathcal{U}_\infty, u = U(Dw+v), x = Hu + Gw. \end{cases} \tag{A.25}$$

The following result provides a refinement for the construction by showing that we can restrict attention to time-invariant, Gaussian distributions in $\overline{\mathcal{B}}^\infty$ without any optimality loss.

Proposition 5. *Under Assumptions 2, 5-6, with $u = U(Dw+v)$ for $U \in \mathcal{U}_\infty$ and $x = Hu+Gw$, the inner maximization problem in (A.25) is solved by a time-invariant Gaussian distribution $\mathbb{P}^\star \in \mathcal{B}_{\mathcal{N}}^\infty$.*

Proposition 5 implies that the feasible set of the inner maximization problem in (A.25) can be restricted to $\mathcal{B}_{\mathcal{N}}^\infty$ without any optimality loss. Because any distribution $\mathbb{P} \in \mathcal{B}_{\mathcal{N}}^\infty$ is uniquely determined by the covariance matrices Σ_w and Σ_v, the upper bounding problem (A.25) is equivalent to the simplified minimax problem:

$$\overline{p}^\star = \min_{U\in\mathcal{U}_\infty} \max_{\Sigma_w\in\mathcal{M}_{\Sigma_w}, \Sigma_v\in\mathcal{M}_{\Sigma_v}} J(U;\Sigma_w,\Sigma_v) \tag{A.26}$$

with objective function given by

$$J(U;\Sigma_w,\Sigma_v) = J_{\mathbb{P}}(HU\eta + Gw, U\eta),$$

and $\mathbb{P} = \mathbb{P}_{x_0} \otimes (\otimes_{t=0}^\infty (\mathbb{P}_{w_t} \otimes \mathbb{P}_{v_t}))$, with $\mathbb{P}_{x_0} = \mathbb{P}_{w_t} = \mathcal{N}(0,\Sigma_w)$ and $\mathbb{P}_{v_t} = \mathcal{N}(0,\Sigma_v)$ for all $t \in \mathbb{N}$.

Moreover, mirroring the developments in §B.6, we can further restrict the feasible sets in the inner maximization in (A.26) to only contain covariance matrices that dominate the nominal covariance matrices in Loewner order. More formally, if we define the sets

$$\mathcal{M}^{+}_{\Sigma_w} = \{\Sigma_w \in \mathcal{M}_{\Sigma_w} : \Sigma_w \succeq \hat{\Sigma}_w\} \text{ and } \mathcal{M}^{+}_{\Sigma_v} = \{\Sigma_v \in \mathcal{M}_{\Sigma_v} : \Sigma_v \succeq \hat{\Sigma}_v\},$$

then a straightforward adaptation of Theorem C – which apply here because Assumption 4 holds – can be used to argue that the optimal $\Sigma_w^\star$ and $\Sigma_v^\star$ in the inner maximization problem in (A.26) satisfy $\Sigma_w^\star \succeq \hat{\Sigma}_w$ and $\Sigma_v^\star \succeq \hat{\Sigma}_v$. Hence, we reformulate $\bar{p}^\star$ as

$$\bar{p}^\star = \min_{U \in \mathcal{U}_\infty} \max_{\Sigma_w \in \mathcal{M}^{+}_{\Sigma_w}, \Sigma_v \in \mathcal{M}^{+}_{\Sigma_v}} J(U; \Sigma_w, \Sigma_v). \tag{A.27}$$

Lower Bound for Dual

To derive a lower bound on $d^\star$, we restrict nature's choices to the set $\underline{\mathcal{B}}^\infty_{\mathcal{N}}$ of all time-invariant Gaussian distributions in the ambiguity set $\mathcal{B}^\infty$ constrained so their covariances belong to the sets $\mathcal{M}^{+}_{\Sigma_w}$ and $\mathcal{M}^{+}_{\Sigma_v}$, respectively; that is, $\underline{\mathcal{B}}^\infty_{\mathcal{N}} = \{\mathbb{P} \in \mathcal{B}^\infty_{\mathcal{N}} : \mathbb{E}_{\mathbb{P}}[zz^\top] \succeq \mathbb{E}_{\hat{\mathbb{P}}}[zz^\top] \text{ for all } z \in \mathcal{Z}\}$. Therefore, we propose the following lower bound on $d^\star$:

$$\underline{d}^\star = \begin{cases} \sup\limits_{\mathbb{P} \in \underline{\mathcal{B}}^\infty_{\mathcal{N}}} \inf\limits_{x,u} \; J_{\mathbb{P}}(x,u) \\ \qquad\qquad \text{s.t. } \; u \in \mathcal{U}_\eta, \;\; x = Hu + Gw. \end{cases} \tag{A.28}$$

As (A.28) is obtained by restricting the feasible set of the dual DRLQ problem (A.22), we have $\underline{d}^\star \leq d^\star$.

The following proposition, which leverages the stabilizability and detectability Assumption 5 and classical results in infinite-horizon LQG control, will allow simplifying the formulation in (A.28) by only considering stationary, linear policies.

Proposition 6. *Under Assumptions 2, 5-6 and for any $\mathbb{P} \in \underline{\mathcal{B}}^\infty_{\mathcal{N}}$, the inner minimization problem in (A.28) is solved by a policy $u = U\eta$ for some $U \in \mathcal{U}_\infty$ and a state process $x = Hu + Gw$.*

Proposition 6 implies that the feasible set of the inner minimization problem in (A.28) can be restricted to linear stationary policies of the form $u = U\eta$ for some $U \in \mathcal{U}_\infty$ without optimality loss. Thus, the lower bounding problem (A.25) is equivalent to the simplified maximin problem

$$\underline{d}^\star = \max_{\Sigma_w \in \mathcal{M}^{+}_{\Sigma_w}, \Sigma_v \in \mathcal{M}^{+}_{\Sigma_v}} \min_{U \in \mathcal{U}_\infty} \; J(U; \Sigma_w, \Sigma_v), \tag{A.29}$$

where the objective function $J(U; \Sigma_w, \Sigma_v)$ is defined as before.

Finally, we prove that $J(U; \Sigma_w, \Sigma_v)$ is a convex-concave saddle function. We first prove that the limit superior in the definition of the average cost $J_{\mathbb{P}}(x,u)$ reduces to a normal limit whenever $u = U\eta$ is a stationary policy induced by some $U \in \mathcal{U}_\infty$, from which the result on J will follow.

Lemma C. *Under Assumptions 2, 5-6 and for any $\mathbb{P} \in \underline{\mathcal{B}}^\infty_{\mathcal{N}}$, if $u = U\eta$ and $x = Hu + Gw$ for some $U \in \mathcal{U}_\infty$, then:*

$$J_{\mathbb{P}}(x,u) = \lim_{T \to \infty} \frac{1}{T} \sum_{t=0}^{T-1} \mathbb{E}_{\mathbb{P}} \left[x_t^\top Q_0 x_t + u_t^\top R_0 u_t \right]. \tag{A.30}$$

With this, the following result proves that $J(U; \Sigma_w, \Sigma_v)$ is convex-concave.

Proposition 7. *Under Assumptions 2,4, 5-6 and for any $\mathbb{P} \in \underline{\mathcal{B}}^\infty_{\mathcal{N}}$, the restriction of $J(U; \Sigma_w, \Sigma_v)$ to $\mathcal{U}_\infty \times (\mathcal{M}^{+}_{\Sigma_w} \times \mathcal{M}^{+}_{\Sigma_v})$ is convex in U and linear in (Σ_w, Σ_v).*

The insights of this section culminate in the following main result, which shows that all structural results shown in the finite horizon extend to the infinite horizon formulation.

Theorem A (Nash strategies and strong duality)**.** *Under Assumptions 2,4, 5-6, we have:*

(i) The optimal value in problem (A.21) equals the optimal values in problems (A.22), (A.25), and (A.28), i.e., $p^\star = d^\star = \bar{p}^\star = \underline{d}^\star$, and all optimal values are attained.

(ii) The primal DRLQ problem (A.21) *admits an optimal stationary linear control policy,* $u^\star = U\eta$ *for* $U \in U_\infty$.

(iii) The dual DRLQ problem (A.22) *admits an optimal time-invariant Gaussian distribution,* $\mathbb{P}^\star \in \mathcal{B}_{\mathcal{N}}^\infty$.

(iv) The optimal covariance matrices under $\mathbb{P}^\star$ *satisfy* $\Sigma_w^\star \succeq \hat{\Sigma}_w$ *and* $\Sigma_v^\star \succeq \hat{\Sigma}_v$.

Proof. The proof of Theorem A relies on the minimax theorem due to Fan (1953) (also see Theorem 4.2 in Sion (1958)), which states that if $\mathcal{M}$ is a non-empty convex subset of a vector space, $\mathcal{N}$ is any non-empty compact and convex subset of a topological vector space, and $f : \mathcal{M} \times \mathcal{N} \to \mathbb{R}$ is convex in its first argument and concave and upper semicontinuous in the second argument, then $\inf_{\nu\in\mathcal{N}} \max_{\mu\in\mathcal{M}} f(\mu,\nu) = \max_{\mu\in\mathcal{M}} \inf_{\nu\in\mathcal{N}} f(\mu,\nu)$.

By the construction of the upper and lower bounding problems (A.25) and (A.28), respectively,

$$\underline{d}^\star \leq d^\star \leq p^\star \leq \overline{p}^\star. \tag{A.31}$$

From Proposition 6 and the discussion immediately following it, we also have that $\underline{d}^\star$ matches the optimal value of the simplified maximin problem (A.29). Similarly, from Proposition 5 and the discussion following it, we have that $\overline{p}^\star$ matches the optimal value of the simplified minimax problem (A.26). Note that (A.26) and (A.29) differ only with respect to the order of minimization and maximization. By construction, the feasible set $\mathcal{U}_\infty$ from which policies are chosen in (A.26) and (A.29) is a non-empty subset of the infinite-dimensional linear space $\mathcal{T}^{m\times p}$ and is convex because the norm $\|\cdot\|_{\mathcal{T}}$ for Toeplitz matrices is convex. In addition, the feasible set $\mathcal{M}_{\Sigma_w}^+ \times \mathcal{M}_{\Sigma_v}^+$ from which nature chooses covariance matrices in (A.26) and (A.29) is a non-empty convex and compact subset of the finite-dimensional linear space $\mathbb{S}^n \times \mathbb{S}^p$. Finally, Proposition 7 ensures that the restriction of $J(U;\Sigma_w,\Sigma_v)$ to $\mathcal{U}_\infty \times (\mathcal{M}_{\Sigma_w} \times \mathcal{M}_{\Sigma_v})$ is convex in U and linear (and thus trivially upper semicontinuous) in (Σ_w,Σ_v). Therefore, the Fan Minimax Theorem is applicable to problems (A.26) and (A.29) and this implies that $\underline{d}^\star = \overline{p}^\star$, which in turn implies that all three inequalities in (A.31) collapse to equalities. Each of these equalities implies one of the three assertions in (i)-(iii).

Lastly, assertion (iv) follows immediately from the construction of the sets $\mathcal{M}_{\Sigma_w}^+ \times \mathcal{M}_{\Sigma_v}^+$, which implies that $\Sigma_w^\star \succeq \hat{\Sigma}_w$ and $\Sigma_v^\star \succeq \hat{\Sigma}_v$. □

F Results on the LQG Problem with Known Distributions

F.1 Finite Horizon

The finite horizon classical LQG problem assumes that all exogenous noise terms follow known Gaussian distributions with the covariance matrices for the output noise being positive definite, i.e., $\Sigma_{v_t} \succ 0$ for every $t \in [T-1]$ (Bertsekas, 2017). In this section, we restate these classical results and adapt them to the case where the exogenous noise has non-zero mean. Throughout this section, we let $\mu_{x_0}, \mu_{v_t}, \mu_{w_t}$ denote the means and $X_0 = \Sigma_{x_0}$, $V_t = \Sigma_{v_t}$, and $W_t = \Sigma_{w_t}$ to denote the covariance matrices characterizing the exogenous noise.

This problem can be solved efficiently via dynamic programming (Bertsekas, 2017). The unique optimal control inputs satisfy $u_t^\star = K_t \hat{x}_t$ for every $t \in [T-1]$, where $K_t \in \mathbb{R}^{n\times n}$ is the optimal feedback gain matrix, and $\hat{x}_t = \mathbb{E}_{\mathbb{P}}[x_t | y_0, \ldots, y_t]$ is the minimum mean-squared-error estimator of x_t given the observation history up to time t. Thanks to the celebrated separation principle, K_t can be computed by pretending that the system is deterministic and allows for perfect state observations, and $\hat{x}_t$ can be computed while ignoring the control problem.

To compute K_t, one first solves the deterministic LQR problem corresponding to the LQG problem. Its value function $x_t^\top P_t x_t$ at time t is quadratic in x_t, and P_t obeys the backward recursion

$$P_t = A_t^\top P_{t+1} A_t + Q_t - A_t^\top P_{t+1} B_t (R_t + B_t^\top P_{t+1} B_t)^{-1} B_t^\top P_{t+1} A_t \quad \forall t \in [T-1] \tag{A.32a}$$

initialized by $P_T = Q_T$. The optimal feedback gain matrix K_t can then be computed from P_{t+1} as

$$K_t = -(R_t + B_t^\top P_{t+1} B_t)^{-1} B_t^\top P_{t+1} A_t \quad \forall t \in [T-1]. \tag{A.32b}$$

Importantly, note that K_t only depends on the system matrices $\{A_\tau, B_\tau\}_{\tau \geq t}$ and on the cost matrices $\{R_\tau, Q_\tau\}_{\tau \geq t}$, but does *not depend* on the distribution of the exogenous noise terms.

Because x_t and $(y_0, \ldots, y_t)$ follow a multivariate Gaussian distribution, the minimum mean-squared-error estimator $\hat{x}_t$ can be calculated directly using the formula for the mean of a conditional Gaussian distribution. Alternatively, one can use the Kalman filter to compute $\hat{x}_t$ recursively, which is more insightful and more efficient. The Kalman filter also recursively computes the covariance matrix Σ_t of x_t conditional on $y_0, \ldots, y_t$ and the covariance matrix $\Sigma_{t+1|t}$ of x_{t+1} conditional on $y_0, \ldots, y_t$ evaluated under $\mathbb{P}$. Specifically, these covariance matrices obey the forward recursion

$$\left.\begin{aligned} \Sigma_t &= \Sigma_{t|t-1} - \Sigma_{t|t-1} C_t^\top (C_t \Sigma_{t|t-1} C_t^\top + V_t)^{-1} C_t \Sigma_{t|t-1} \\ \Sigma_{t+1|t} &= A_t \Sigma_t A_t^\top + W_t \end{aligned}\right\} \forall t \in [T-1] \tag{A.33}$$

initialized by $\Sigma_{0|-1} = X_0$. Using $\Sigma_{t|t-1}$, we then define the Kalman filter gain as

$$L_t = \Sigma_t C_t^\top V_t^{-1} \quad \forall t \in [T-1]$$

which allows us to compute the minimum mean-squared-error estimator via the forward recursion

$$\hat{x}_{t+1} = A_t \hat{x}_t + B_t K_t \hat{x}_t + \hat{\mu}_{w_t} + L_{t+1} \left(y_{t+1} - C_{t+1}(A_t \hat{x}_t + B_t K_t \hat{x}_t + \hat{\mu}_{w_t}) - \mu_{v_{t+1}} \right) \quad \forall t \in [T-1] \tag{A.34}$$

initialized by $\hat{x}_0 = L_0 y_0 - \hat{\mu}_{v_0}$.

Note that (A.32b) and (A.34) readily show that the optimal control policy $u_t^\star$ depends *affinely* on the outputs $y_0, y_1, \ldots, y_t$ and provide the explicit recursive procedure needed to compute all the relevant coefficients.

Moreover, one can verify from these expressions that the optimal value of the LQG problem is

$$\sum_{t=0}^{T-1} \mathrm{Tr}((Q_t - P_t)\Sigma_t) + \sum_{t=1}^{T} \mathrm{Tr}(P_t(A_{t-1}\Sigma_{t-1}A_{t-1}^\top + W_{t-1})) + \mathrm{Tr}(P_0 X_0). \tag{A.35}$$

F.2 Definitions of Stacked System Matrices for Finite Horizon

The stacked system matrices appearing in problem (A.8) are defined as follows. First, the stacked state and input cost matrices $Q \in \mathbb{S}^{n(T+1)}$ and $R \in \mathbb{S}^{mT}$ are set to

$$Q = \begin{bmatrix} Q_0 & & & \\ & Q_1 & & \\ & & \ddots & \\ & & & Q_T \end{bmatrix} \quad \text{and} \quad R = \begin{bmatrix} R_0 & & & \\ & R_1 & & \\ & & \ddots & \\ & & & R_{T-1} \end{bmatrix},$$

respectively. Similarly, the stacked matrices appearing in the linear dynamics and the measurement equations $C \in \mathbb{R}^{pT \times n(T+1)}$, $G \in \mathbb{R}^{n(T+1) \times n(T+1)}$ and $H \in \mathbb{R}^{n(T+1) \times mT}$ are defined as

$$C = \begin{bmatrix} C_0 & 0 & & & \\ & C_1 & 0 & & \\ & & \ddots & \ddots & \\ & & & C_{T-1} & 0 \end{bmatrix}, \quad G = \begin{bmatrix} A_0^0 & & & \\ A_0^1 & A_1^1 & & \\ \vdots & & \ddots & \\ A_0^T & A_1^T & \ldots & A_T^T \end{bmatrix}$$

and

$$H = \begin{bmatrix} 0 & & & & \\ A_1^1 B_0 & 0 & & & \\ A_1^2 B_0 & A_2^2 B_1 & 0 & & \\ \vdots & & & \ddots & \\ \vdots & & & & 0 \\ A_1^T B_0 & A_2^T B_1 & \ldots & \ldots & A_T^T B_{T-1} \end{bmatrix},$$

respectively, where $A_s^t = \prod_{k=s}^{t-1} A_k$ for every $s < t$ and $A_s^t = I$ for $s = t$.

F.3 Results on Purified Output Feedback Policies

Using the stacked system matrices, we can now express the purified observation process η as a linear function of the exogenous uncertainties w and v that is *not* impacted by u. The following result summarizes this – see also Ben-Tal et al. (2005); Skaf and Boyd (2010).

Lemma D. *We have $\eta = Dw + v$, where $D = CG$.*

Proof. The purified observation process is defined as $\eta = y - \hat{y}$. Recall now that the observations of the original system satisfy $y = Cx + v$. Similarly, one readily verifies that the observations of the fictitious noise-free system satisfy $\hat{y} = C\hat{x}$. Thus, we have $\eta = C(x - \hat{x}) + v$. Next, recall that the state of the original system satisfies $x = Hu + Gw$, and note that the state of the fictitious noise-free system satisfies $\hat{x} = Hu$. Combining all of these linear equations finally shows that u cancels out and that $\eta = CGw + v = Dw + v$. □

It is well known that every causal control policy that is linear in the original observations y can be reformulated as a causal policy that is linear in the purified observations η and vice versa (Ben-Tal et al., 2005; Skaf and Boyd, 2010). Perhaps surprisingly, however, the one-to-one transformation between the respective coefficients of y and η is *not* linear. To keep this paper self-contained, we review these insights in the next lemma.

Lemma E. *If $u = U\eta + q$ for some $U \in \mathcal{U}$ and $q \in \mathbb{R}^{pT}$, then $u = U'y + q'$ for $U' = (I + UCH)^{-1}U$ and $q' = (I + UCH)^{-1}q$. Conversely, if $u = U'y + q'$ for some $U' \in \mathcal{U}$ and $q' \in \mathbb{R}^{pT}$, then $u = U\eta + q$ for $U = (I - U'CH)^{-1}U'$ and $q = (I - U'CH)^{-1}q'$.*

Proof. If $u = U\eta + q$ for some $U \in \mathcal{U}$ and $q \in \mathbb{R}^{pT}$, then we have

$$u = U\eta + q = U(y - \hat{y}) + q = Uy - UC\hat{x} + q = Uy - UCHu + q,$$

where the second equality follows from the definition of η, the third equality holds because $y = Cx + v$, and the last equality exploits our earlier insight that $\hat{y} = C\hat{x}$. The last expression depends only on y and u. Solving for u yields $u = U'y + q'$, where $U' = (I + UCH)^{-1}U$ and $q' = (I + UCH)^{-1}q$. Note that $(I + UCH)$ is indeed invertible because $I + UCH$ is a lower triangular matrix with all diagonal entries equal to one, ensuring a determinant of one.

Similarly, if $u = U'y + q'$ for some $U' \in \mathcal{U}$ and $q' \in \mathbb{R}^{pT}$, then we have

$$u = U'y + q' = U'(\eta + \hat{y}) + q' = U'\eta + U'C\hat{x} + q' = U'\eta + U'CHu + q'.$$

Solving for u yields $u = U\eta + q$, where $U = (I - U'CH)^{-1}U'$ and $q = (I - U'CH)^{-1}q'$. Note again that $(I - U'CH)$ is indeed invertible because $(I - U'CH)$ is a lower triangular matrix with all diagonal entries equal to one. □

F.4 Infinite-Horizon Results

The infinite-horizon classical LQG problem with average cost criterion assumes that all exogenous noise terms follow known time-invariant Gaussian distributions with the covariance matrices for the noise being positive definite, *i.e.*, $\Sigma_{v_t} = \Sigma_v \succ 0$ and $\Sigma_{w_t} = \Sigma_w \succ 0$ for all $t \in \mathbb{N}$. For simplicity, we consider the case where the exogenous noise has zero mean. Furthermore, consistent with Section E, we assume that the standard assumptions outlined in Assumption 5 hold.

The infinite-horizon classical LQG problem is solved by optimal control inputs that satisfy $u_t^\star = K\hat{x}_t$ for every $t \in \mathbb{N}$, where $K \in \mathbb{R}^{n \times n}$ is the optimal steady-state feedback gain matrix, and $\hat{x}_t$ is again the minimum mean-squared-error estimator of x_t given the observation history up to time t.

To compute K, one first finds the $P \in \mathbb{S}_+^n$ that solves the discrete-time algebraic Riccati equation (DARE)

$$P = A_0^\top P A_0 + Q_0 - A_0^\top P B_0 (R_0 + B_0^\top P B_0)^{-1} B_0^\top P A_0, \tag{A.36}$$

which is guaranteed to exist and to be unique under Assumption 5 by (Lancaster and Rodman, 1995, Lemma 16.6.1). The matrix K can then be computed from P as

$$K = -(R_0 + B_0^\top P B_0)^{-1} B_0^\top P A_0.$$

As in the finite-horizon case, the Kalman filter can be used to compute $\hat{x}_t$. The steady-state covariance matrix $\tilde{\Sigma} \in \mathbb{S}_+^n$ solves

$$\tilde{\Sigma} = A_0 \tilde{\Sigma} A_0^\top + \Sigma_w - A_0 \tilde{\Sigma} C_0^\top (C_0 \tilde{\Sigma} C_0^\top + \Sigma_v)^{-1} C_0 \tilde{\Sigma} A_0^\top, \tag{A.37}$$

and is guaranteed to exist and to be unique under Assumption 5 by (Lancaster and Rodman, 1995, Theorem 17.5.3). We define the steady-state Kalman filter gain as

$$L = \tilde{\Sigma} C_0^\top (\Sigma_v + C_0 \tilde{\Sigma} C_0)^{-1}.$$

This allows state estimation via the recursion

$$\hat{x}_{t+1} = A_0 \hat{x}_t + B_0 u_t^\star + L(y_{t+1} - C_0(A_0 \hat{x}_t + B_0 u_t^\star)), \tag{A.38}$$

initialized by $\hat{x}_0 = L y_0$. If $\hat{\Sigma}_w \succ 0$, then $A(I - LC_0)$ is Schur stable by (Lancaster and Rodman, 1995, Theorem 17.5.3). Additionally, if $Q_0 \succ 0$, then $(A_0 + B_0 K)$ is Schur stable by (Lancaster and Rodman, 1995, Theorem 16.6.4). These observations will be useful for establishing that both the state x_t and its estimate $\hat{x}_t$ admit stationary covariance matrices.

Lemma F. *Suppose that Assumption 5 holds, $\hat{\Sigma}_w \succ 0$ and $\hat{\Sigma}_v \succ 0$. If $u_t^\star = K\hat{x}_t$, and $\hat{x}_t$ obeys the forward recursion*

$$\hat{x}_{t+1} = A_0 \hat{x}_t + B_0 u_t^\star + L(y_{t+1} - C_0(A_0 \hat{x}_t + B_0 u_t^\star)),$$

initialized by $\hat{x}_0 = L y_0$, then x_t and $\hat{x}_t$ admit stationary covariance matrices.

Proof. Denote by $e_t = x_t - \hat{x}_t$ the estimation error and by $z_t = [x_t^\top, e_t^\top]^\top$ the joint vector of state and estimation error under the optimal control inputs $u_t^\star = K\hat{x}_t$ for every $t \in \mathbb{N}$. The state dynamics under the optimal control inputs are then given by

$$\begin{aligned} x_{t+1} &= A_0 x_t + B_0 K \hat{x}_t + w_t = A_0(e_t + \hat{x}_t) + B_0 K \hat{x}_t + w_t \\ &= A_0 e_t + (A_0 + B_0 K)\hat{x}_t + w_t = (A_0 + B_0 K) x_t - B_0 K e_t + w_t. \end{aligned} \tag{A.39}$$

For the error dynamics, we have

$$\begin{aligned} e_{t+1} &= x_{t+1} - \hat{x}_{t+1} \\ &= A_0 x_t + B_0 K \hat{x}_t + w_t - A_0 \hat{x}_t - B_0 K \hat{x}_t - L(C_0 x_{t+1} + v_{t+1}) + LC_0(A_0 \hat{x}_t + B_0 K \hat{x}_t) \\ &= (A_0 - LC_0 A_0)(x_t - \hat{x}_t) - (I - LC_0) w_t - L v_{t+1} \\ &= (A_0 - LC_0 A_0) e_t + (I - LC_0) w_t - L v_{t+1}, \end{aligned} \tag{A.40}$$

where the second equality follows from (A.39) and $y_{t+1} = C_0 x_{t+1} + v_{t+1}$, the third equality follows from $x_{t+1} = A_0 x_t + B_0 K \hat{x}_t + w_t$ and rearranging terms, and the last equality follows from the definition of e_t. Combining the dynamics in (A.39) and (A.40), we have

$$z_{t+1} = F z_t + \Xi \xi_t, \text{ where } F = \begin{bmatrix} A_0 + B_0 K & -B_0 K \\ 0 & A_0 - LC_0 A_0 \end{bmatrix}, \ \xi_t = \begin{bmatrix} w_t \\ v_{t+1} \end{bmatrix} \text{ and } \Xi = \begin{bmatrix} I & 0 \\ I - LC_0 & -L \end{bmatrix}.$$

Note that $(\xi_t)_{t \in \mathbb{N}}$ is an i.i.d. sequence with zero mean and finite covariance in the form of

$$\Sigma_{\xi_t} = \mathbb{E}[\xi_t \xi_t^\top] = \begin{bmatrix} \Sigma_w & 0 \\ 0 & \Sigma_v \end{bmatrix} := \Sigma_\xi. \tag{A.41}$$

As ξ_t is independent of z_t, the linear recursion of z_{t+1} implies that $\Sigma_{z+1} = \mathbb{E}[z_{t+1} z_{t+1}^\top]$ follows the discrete-time Lyapunov recursion

$$\Sigma_{z_{t+1}} = F \Sigma_{z_t} F^\top + \Xi \Sigma_\xi \Xi^\top. \tag{A.42}$$

Note that F is an upper block triangular matrix, and thus its spectrum is given by $\sigma(F) = \sigma(A_0 + B_0K) \cup \sigma(A_0 - LC_0A_0)$. As both $A_0 + B_0K$ and $(I - LC_0)A_0$ are Schur stable, $\sigma(F) < 1$. Therefore, F is a Schur stable matrix. By (Kumar and Varaiya, 2015, Theorem 3.4), as F is Schur stable and $G\Sigma_\xi G^\top \succeq 0$, the matrix Σ_{z_t} converges to the unique positive semidefinite solution of

$$\Sigma_z = F\Sigma_z F^\top + G\Sigma_\xi G^\top. \tag{A.43}$$

By construction of z_t, Σ_{z_t} admits the following block from

$$\Sigma_{z_t} = \begin{bmatrix} \mathbb{E}[x_t x_t^\top] & \mathbb{E}[x_t e_t^\top] \\ \mathbb{E}[e_t x_t^\top] & \mathbb{E}[e_t e_t^\top] \end{bmatrix} \tag{A.44}$$

As Σ_{z_t} converges to a stationary matrix Σ_z solving (A.43), each of its blocks must converge as well, and thus this observation completes the first assertion of the statement that x_t admits a stationary covariance matrix.

Next, recall that the estimator satisfies $\hat{x}_t = x_t - e_t$, and thus we have

$$\mathbb{E}[\hat{x}_t\hat{x}_t^\top] = \mathbb{E}[(x_t - e_t)(x_t - e_t)^\top] = \mathbb{E}[x_t x_t^\top] + \mathbb{E}[e_t e_t^\top] - \mathbb{E}[x_t e_t^\top] - \mathbb{E}[e_t x_t^\top].$$

Because each of the components of Σ_{z_t} converges, the matrices on the right-hand-side of the expression above also converge. Hence, $\mathbb{E}[\hat{x}_t\hat{x}_t^\top]$ converges to a stationary matrix. □

G Proofs for Section A

Proof of Theorem A. Under the assumptions of the theorem we have $x_1 = w_0$ and $y_0 = v_0$. Hence, $\mathbb{E}_{\mathbb{P}}[w_0|v_0]$ is equivalent to $\mathbb{E}_{\mathbb{P}_{w_0}}[w_0]$ because v_0 and w_0 are independent. As $\mathbb{P}_{w_0}$ is the uniform distribution over polytope $\mathcal{H}$, computing $\mathbb{E}_{\mathbb{P}_{w_0}}[w_0]$ is equivalent to computing the centroid of $\mathcal{H}$, which is known to be #P-hard (Rademacher, 2007, Theorem 1). □

G.1 Verifying Assumption 2 for Examples in §A.2

G.1.1 Assumption 2 for Wasserstein Ambiguity Sets.

Our construction and proof rely on the Gelbrich distance, which we formally define next.

Definition 7 (Gelbrich distance)**.** *For any $d \in \mathbb{N}$, the Gelbrich distance between two pairs of mean vectors and covariance matrices $(\mu_z, \Sigma_z), (\hat{\mu}_z, \hat{\Sigma}_z) \in \mathcal{M}_2^{d_z}$ is given by*

$$\mathbb{G}\big((\mu_z, \Sigma_z), (\hat{\mu}_z, \hat{\Sigma}_z)\big) = \sqrt{\|\mu_z - \hat{\mu}_z\|^2 + \mathrm{Tr}\left(\Sigma_z + \hat{\Sigma}_z - 2\left(\hat{\Sigma}_z^{1/2}\Sigma_z\hat{\Sigma}_z^{1/2}\right)^{1/2}\right)}.$$

The Gelbrich distance is closely related to the 2-Wasserstein distance. Indeed, it is known that the 2-Wasserstein distance between two distributions is bounded below by the Gelbrich distance between the mean-covariance pairs of the two distributions, and when the two distributions are Gaussian, this bound tight. These results are summarized in the next proposition.

Proposition 8 (Gelbrich bound (Gelbrich, 1990, Theorem 2.1))**.** *For any two distributions $\mathbb{P}_z, \hat{\mathbb{P}}_z \in \mathcal{P}(\mathbb{R}^{d_z})$ with mean-covariance pairs $(\mu_z, \Sigma_z), (\hat{\mu}_z, \hat{\Sigma}_z) \in \mathcal{M}_2^{d_z}$, respectively, we have:*

i) $\mathbb{W}(\mathbb{P}_z, \hat{\mathbb{P}}_z) \geq \mathbb{G}((\mu_z, \Sigma_z), (\hat{\mu}_z, \hat{\Sigma}_z))$,

ii) $\mathbb{W}(\mathbb{P}_z, \hat{\mathbb{P}}_z) = \mathbb{G}((\mu_z, \Sigma_z), (\hat{\mu}_z, \hat{\Sigma}_z))$ *if $\mathbb{P}_z$ and $\hat{\mathbb{P}}_z$ are Gaussian.*[4]

[4] Gelbrich (1990) proves this equality more generally, for any two elliptical distributions with the same generator (we discuss this in Appendix §M – see Proposition 20). The equality for the Gaussian case is a special instance of that result.

Proposition 8 implies that Assumption 2 (i) is satisfied. To see that Assumption 2 (ii) is also satisfied, consider the set $\mathcal{M}_{(\mu_z, M_z)}$ for $\mathbb{D} = \mathbb{W}$, which we can rewrite as:

$$\begin{aligned}\mathcal{M}^{\mathbb{W}}_{(\mu_1, M_z)} &= \left\{(\mu_z, M_z) \in \mathcal{M}_2^{d_z} : \mathbb{W}\left(\mathcal{N}(\mu_z, M_z), \hat{\mathbb{P}}_z\right) \leq \rho_z\right\} \\ &= \left\{(\mu_z, M_z) \in \mathcal{M}_2^{d_z},\ \left(\mathbb{G}((\mu_z, M_z - \mu_z\mu_z^\top), (\hat{\mu}_z, \hat{\Sigma}_z))\right)^2 \leq \rho_z^2\right\},\end{aligned}$$

where the second equality follows from Proposition 8-(ii). The latter set is known to be convex and compact (Nguyen, 2019, Proposition 3.17), so we conclude that the 2-Wasserstein distance satisfies also Assumption 2-(ii).

G.1.2 Assumption 2 for Kullback-Leibler Ambiguity Sets.

Mirroring the Wasserstein case, we first formalize a divergence between mean-covariance pairs.

Definition 8 (KL Divergence Between Moments)**.** *The KL-type divergence from* $(\mu_z, \Sigma_z) \in \mathbb{R}^{d_z} \times \mathbb{S}_+^{d_z}$ *to* $(\hat{\mu}_z, \hat{\Sigma}_z) \in \mathbb{R}^{d_z} \times \mathbb{S}_{++}^{d_z}$ *is given by*

$$\mathbb{T}((\mu_z, \Sigma_z), (\hat{\mu}_z, \hat{\Sigma}_z)) = \frac{1}{2}\left((\mu_z - \hat{\mu}_z)^\top \hat{\Sigma}_z^{-1}(\mu_z - \hat{\mu}_z) + \mathrm{Tr}\left(\Sigma_z \hat{\Sigma}_z^{-1}\right) - \log\det\left(\Sigma_z \hat{\Sigma}_z^{-1}\right) - d_z\right).$$

In our case, the mean-covariance pair $(\hat{\mu}_z, \hat{\Sigma}_z)$ corresponds to the nominal distribution $\hat{\mathbb{P}}$. To ensure that the inverse of $\hat{\Sigma}_z$ is well-defined, we must therefore require that the nominal distribution $\hat{\mathbb{P}}_z$ of every noise term $z \in \mathcal{Z}$ be non-degenerate, that is, the nominal covariance matrix is positive definite, $\hat{\Sigma}_z \succ 0$. This slightly strengthens Assumption 1, but is without substantial practical loss.

The following proposition shows that the KL divergence from any distribution to a non-degenerate, Gaussian distribution is bounded below by the KL-type divergence between their respective mean-covariance pairs, and the bound is tight if the former distribution is Gaussian.

Proposition 9 (KL bound)**.** *For any distribution* $\mathbb{P}_z$ *on* $\mathcal{P}(\mathbb{R}^{d_z})$ *with mean-covariance pair* $(\mu_z, \Sigma_z) \in \mathcal{M}_2^{d_z}$ *and Gaussian distribution* $\hat{\mathbb{P}}_z$ *with mean-covariance pair* $(\hat{\mu}_z, \hat{\Sigma}_z) \in \mathbb{R}^{d_z} \times \mathbb{S}_{++}^{d_z}$*, we have:*

i) $\mathbb{K}(\mathbb{P}_z, \hat{\mathbb{P}}_z) \geq \mathbb{T}((\mu_z, \Sigma_z), (\hat{\mu}_z, \hat{\Sigma}_z))$

ii) $\mathbb{K}(\mathbb{P}_z, \hat{\mathbb{P}}_z) = \mathbb{T}((\mu_z, \Sigma_z), (\hat{\mu}_z, \hat{\Sigma}_z))$ *if* $\mathbb{P}_z^1$ *is Gaussian.*

Proof. Let $\hat{f}_z$ denote the density of the Gaussian distribution $\hat{\mathbb{P}}_z$. Because the KL divergence from $\mathbb{P}_z$ to $\hat{\mathbb{P}}_z$ is finite only if $\mathbb{P}_z$ is absolutely continuous with respect to $\hat{\mathbb{P}}_z$, $\mathbb{P}_z$ must admit a density on $\mathbb{R}^{d_z}$, which we denote by f_z. The KL divergence can then be written as:

$$\begin{aligned}\mathbb{K}(\mathbb{P}_z, \hat{\mathbb{P}}_z) &= \int_{\mathbb{R}^{d_z}} f_z(z)\ \ln\left(\frac{f_z(z)}{\hat{f}_z(z)}\right)\ \mathrm{d}z \\ &= h(\mathbb{P}_z) - \int_{\mathbb{R}^{d_z}} f_z(z)\ \ln \hat{f}_z(z)\mathrm{d}z,\end{aligned}$$

where $h(\mathbb{P}_z) = -\int_{\mathbb{R}^{d_z}} f_z(z)\ \ln f_z(z)\mathrm{d}z$ denotes the *differential entropy* of $\mathbb{P}_z$. Because $\hat{\mathbb{P}}_z$ is Gaussian with mean $\hat{\mu}_z$ and covariance $\hat{\Sigma}_z$, the second term above can be written as:

$$\begin{aligned}\int_{\mathbb{R}^{d_z}} f_z(z)\ \ln \hat{f}_z(z)\,\mathrm{d}z &= -\tfrac{1}{2}\,\mathbb{E}_{\mathbb{P}_z}\left[(z - \hat{\mu}_z)^\top (\hat{\Sigma}_z)^{-1}(z - \hat{\mu}_z)\right] - \tfrac{1}{2}\ \ln\big((2\pi)^d\ \det(\hat{\Sigma}_z)\big) \\ &= -\tfrac{1}{2}\,\mathbb{E}_{\mathbb{P}_z}\left[\mathrm{Tr}\Big((\hat{\Sigma}_z)^{-1}(z - \hat{\mu}_z)(z - \hat{\mu}_z)^\top\Big)\right] - \tfrac{1}{2}\ \ln\big((2\pi)^d\ \det(\hat{\Sigma}_z)\big) \\ &= -\tfrac{1}{2}\ \mathrm{Tr}\Big((\hat{\Sigma}_z)^{-1}\left(\Sigma_z + (\mu_z - \hat{\mu}_z)(\mu_z - \hat{\mu}_z)^\top\right)\Big) - \tfrac{1}{2}\ \ln\big((2\pi)^d\ \det(\hat{\Sigma}_z)\big),\end{aligned}$$

where in the last step we used that fact that the mean is μ_z and the covariance is Σ_z under $\mathbb{P}_z$. Because μ_z, Σ_z, $\hat{\mu}_z$, and $\hat{\Sigma}_z$ are fixed, the expression above is fixed (for any distribution $\mathbb{P}_z$ with mean μ_z

and covariance Σ_z), and the problem of minimizing the KL divergence reduces to the problem of *maximizing* the differential entropy $h(\mathbb{P}_z)$. It is a standard result that among all distributions on $\mathbb{R}^{d_z}$ with given mean and covariance matrix, the Gaussian distribution maximizes the differential entropy (Cover and Thomas, 2006, Theorem 9.6.5). This completes the proof. $\square$

Proposition 9 implies that Assumption 2-(i) is satisfied. To see that Assumption 2-(ii) is also satisfied, consider the set $\mathcal{M}_{(\mu_z,M_z)}$ for $\mathbb{D}=\mathbb{K}$, which can be rewritten as:

$$\begin{aligned}\mathcal{M}^{\mathbb{K}}_{(\mu_z,M_z)} &= \left\{(\mu_z,M_z)\in\mathcal{M}_2^{d_z} : \mathbb{K}\big(\mathcal{N}(\mu_z,M_z),\hat{\mathbb{P}}_z\big)\leq\rho_z\right\}\\ &= \left\{(\mu_z,M_z)\in\mathcal{M}_2^{d_z} : M_z-\mu_z\mu_z^\top\in\mathbb{S}^{d_z}_{++},\ \mathbb{T}((\mu_z,\Sigma_z),(\hat{\mu}_z,\hat{\Sigma}_z))\leq\rho_z\right\}.\end{aligned}$$

The second equality follows from Proposition 9-(ii) and because any distribution $\mathbb{P}_z$ in the set $\mathcal{M}^{\mathbb{K}}_{(\mu_z,M_z)}$ must yield a positive definite covariance matrix for z because we consider ρ_z finite.[5] The set $\mathcal{M}^{\mathbb{K}}_{(\mu_z,M_z)}$ is known to be convex and compact (Taşkesen et al., 2021, Lemma A.3).

We thus conclude that the KL divergence satisfies all premises of Assumption 2.

G.1.3 Assumption 2 for Moment Ambiguity Sets.

We finally consider moment ambiguity sets in which the divergence $\mathbb{D}$ between two probability distributions relies only on the first two moments of the respective distributions. Specifically, for distributions $\mathbb{P}_z,\hat{\mathbb{P}}_z\in\mathcal{P}(\mathbb{R}^{d_z})$ with mean-second moment matrix pairs (μ_z,M_z) and $(\hat{\mu}_z,\hat{M}_z)$, respectively, we take

$$\mathbb{D}(\mathbb{P}_z,\hat{\mathbb{P}}_z)=\mathbb{M}\big((\mu_z,M_z),(\hat{\mu}_z,\hat{M}_z)\big),$$

where $\mathbb{M}:\mathcal{M}_2^{d_z}\times\mathcal{M}_2^{d_z}\to[0,+\infty]$ is any divergence between mean-second moment matrix pairs satisfying $\mathbb{M}(m_z,m_z)=0$ for all $m_z=(\mu_z,M_z)\in\mathcal{M}_2^{d_z}$. The ambiguity set $\mathcal{B}_z$ for the random variable z with nominal distribution $\hat{\mathbb{P}}_z$ can therefore be expressed as:

$$\mathcal{B}_z=\{\mathbb{P}_z\in\mathcal{P}(\mathbb{R}^d):\mathbb{E}_{\mathbb{P}_z}[z]=\mu_z,\ \mathbb{E}_{\mathbb{P}_z}[zz^\top]=M_z,\ \mathbb{M}\left((\mu_z,M_z),(\hat{\mu}_z,\hat{M}_z)\right)\leq\rho_z\}.$$

In this case, Assumption 2-(i) is readily satisfied for any $\mathbb{M}$ because every distribution – including a Gaussian distribution – with a given mean and second moment matrix would yield the same divergence from $\hat{\mathbb{P}}_z$ and would therefore minimize the divergence from the nominal $\hat{\mathbb{P}}_z$. Moreover, Assumption 2-(ii) is satisfied if the set

$$\mathcal{M}_{(\mu_z,M_z)}=\{(\mu_z,M_z)\in\mathcal{M}_2^{d_z}:\mathbb{M}\left((\mu_z,M_z),(\hat{\mu}_z,\hat{M}_z)\right)\leq\rho_z\}$$

is convex and compact for the given $(\hat{\mu}_z,\hat{M}_z)$. This is readily satisfied if the restriction of $\mathbb{M}$ to its first argument (μ_z,Σ_z) is a quasiconvex and coercive function.

H Proofs for Section B

H.1 Proofs for §B.2 (Upper Bound for Primal)

Proof of Proposition 1. In problem (A.11), u and x are given by $u=q+UDw+Uv$ and $x=Hu+Gw=Hq+(G+HUD)w+HUv$, respectively. By substituting these expressions into the objective function of problem (A.11), we obtain the following equivalent reformulation:

$$\begin{aligned}\min_{\substack{q\in\mathbb{R}^{pT}\\ U\in\mathcal{U}}}\max_{\mathbb{P}\in\overline{\mathcal{B}}}\mathbb{E}_{\mathbb{P}}\Big[&(U(Dw+v)+q)^\top\bar{R}(U(Dw+v)+q)+w^\top(G^\top QHUD+G^\top QG)w\Big]\\ &+\mathbb{E}_{\mathbb{P}}[2w^\top G^\top QHq].\end{aligned}$$

[5]This follows because $\mathbb{T}\big((\mu_z,M_z-\mu_z\mu_z^\top),(\hat{\mu}_z,\hat{\Sigma}_z)\big)$ is finite only if $\Sigma_z=M_z-\mu_z\mu_z^\top\succ 0$, due to the $-\log\det\Sigma_z$ term.

Because the objective function in the problem above is quadratic in w and v, we can express the expectation with respect to any $\mathbb{P} \in \overline{\mathcal{B}}$ in terms of the first two moments of w and v, (μ_w, M_w) and (μ_v, M_v) respectively, where:

$$\mu_w = \mathbb{E}_{\mathbb{P}}[w] = (\mu_{x_0}, \mu_{w_0}, \ldots, \mu_{w_{T-1}}), \quad M_w = \mathbb{E}_{\mathbb{P}}[ww^\top] = \mathrm{diag}(M_{x_0}, M_{w_0}, \ldots, M_{w_{T-1}})$$
$$\mu_v = \mathbb{E}_{\mathbb{P}}[v] = (\mu_{v_0}, \ldots, \mu_{v_{T-1}}), \quad M_v = \mathbb{E}_{\mathbb{P}}[vv^\top] = \mathrm{diag}(M_{v_0}, \ldots, M_{v_{T-1}}).$$

Thus, the problem becomes

$$\begin{array}{lll} \min\limits_{\substack{q\in\mathbb{R}^{pT}\\ U\in\mathcal{U}}} & \max\limits_{\substack{\mu_w, M_w,\\ \mu_v, M_v, \mathbb{P}}} & \mathrm{Tr}\left(\left((UD)^\top RUD + (G+HUD)^\top Q(G+HUD)\right)M_w + U^\top \bar{R}UM_v\right) \\ & & +2q^\top(\bar{R}UD + G^\top QH)\mu_w + 2q^\top \bar{R}U\mu_v + q^\top \bar{R}q \\ & \text{s.t.} & \mathbb{P}\in\overline{\mathcal{B}},\ \mu_w = \mathbb{E}_{\mathbb{P}}[w], M_w = \mathbb{E}_{\mathbb{P}}[ww^\top], \mu_v = \mathbb{E}_{\mathbb{P}}[v],\ M_v = \mathbb{E}_{\mathbb{P}}[vv^\top]. \end{array} \tag{A.45}$$

We first prove that the objective in problem (A.45) is at least as large as the objective in problem (A.12). It suffices to prove that any $(\mu_w, M_w), (\mu_v, M_v)$ feasible in the inner maximization problem in (A.45) are also feasible in the inner maximization problem in (A.12). This holds because any $\mathbb{P} \in \overline{\mathcal{B}}$ and $(\mu_w, M_w), (\mu_v, M_v)$ feasible in the inner maximization in (A.45) must satisfy $(\mu_w, M_w) \in \mathcal{M}_{(\mu_w, M_w)}$ and $(\mu_v, M_v) \in \mathcal{M}_{(\mu_v, M_v)}$ by the definition of $\overline{\mathcal{B}}$, and therefore $(\mu_w, M_w), (\mu_v, M_v)$ are feasible in the inner maximization problem in (A.12).

To conclude our proof, we show that the inner maximization problems in (A.45) and (A.12) have the same optimal value. Note that for any (μ_w, M_w) and (μ_v, M_v) feasible in the inner maximization problem in (A.12), the distribution obtained by taking independent couplings of Gaussian distributions with the same first two moments – i.e., $\mathbb{P} = \mathbb{P}_{x_0} \otimes (\otimes_{t=0}^{T-1}\mathbb{P}_{w_t}) \otimes (\otimes_{t=0}^{T}\mathbb{P}_{v_t})$ where $\mathbb{P}_z = \mathcal{N}(\mu_z, M_z)$ for every $z \in \mathcal{Z}$ and $\mathbb{P}_1 \otimes \mathbb{P}_2$ denotes the independent coupling of distributions $\mathbb{P}_1$ and $\mathbb{P}_2$ – would be feasible in the inner maximization problem in (A.45) and would result in the same objective value.

Therefore, the relaxation is exact and the optimal values of (A.11), (A.12), and (A.45) coincide. □

H.2 Proofs for §B.3 (Lower Bound for Dual)

Proof of Proposition 2. We can replace the feasible set $\mathcal{U}_\eta$ of the inner minimization with $\mathcal{U}_y$ because the space $\mathcal{U}_y$ of all causal output feedback policies coincides with the space $\mathcal{U}_\eta$ of all causal *purified* output feedback policies.

With this change, the inner minimization problem in (A.13) becomes an LQG problem where the uncertainties have a known Gaussian distribution $\mathbb{P} \in \mathcal{B}_{\mathcal{N}}$, for which classical results apply. By standard LQG theory, an *affine* output feedback policy $u = U'y + q'$ for some $U' \in \mathcal{U}$ and $q' \in \mathbb{R}^{pT}$ is optimal (see Appendix §F and Bertsekas, 2017). And because any such policy can be equivalently expressed as an affine *purified*-output feedback policy $u = U\eta + q$ for some $U \in \mathcal{U}$ and $q \in \mathbb{R}^{pT}$ (see Lemma E), the feasible set of the inner minimization problem in (A.13) can be taken as the set of all affine purified-output feedback policies without sacrificing optimality.

Thus, problem (A.13) has the same optimal value as problem:

$$\begin{array}{lll} \max\limits_{\mathbb{P}\in\mathcal{B}_{\mathcal{N}}} & \min\limits_{q,U,x,u} & \mathbb{E}_{\mathbb{P}}\left[u^\top Ru + x^\top Qx\right] \\ & \text{s.t.} & U\in\mathcal{U}, u = q + U\eta, x = Hu + Gw. \end{array}$$

Using a similar reasoning as in the proof of Proposition 1, we can now substitute the linear representations of u and x into the objective function and reformulate the above problem as

$$\begin{array}{lll} \max\limits_{\substack{\mu_w, M_w,\\ \mu_v, M_v, \mathbb{P}}} & \min\limits_{\substack{q\in\mathbb{R}^{pT}\\ U\in\mathcal{U}}} & \mathrm{Tr}\left(\left((UD)^\top RUD + (G+HUD)^\top Q(G+HUD)\right)M_w + U^\top \bar{R}UM_v\right) \\ & & +2q^\top(\bar{R}UD + G^\top QH)\mu_w + 2q^\top \bar{R}U\mu_v + q^\top \bar{R}q, \\ & \text{s.t.} & \mathbb{P}\in\mathcal{B}_{\mathcal{N}}, \mu_w = \mathbb{E}_{\mathbb{P}}[w],\ M_w = \mathbb{E}_{\mathbb{P}}[ww^\top],\ \mu_v = \mathbb{E}_{\mathbb{P}}[v],\ M_v = \mathbb{E}_{\mathbb{P}}[vv^\top]. \end{array} \tag{A.46}$$

Lastly, we use a similar reasoning as in the proof of Proposition 1 to prove that problem (A.46) has the same optimal objective as problem (A.14). Consider any $(\mu_w, M_w), (\mu_v, M_v), \mathbb{P}$ feasible

in the outer maximization problem in (A.46). Because $\mathbb{P} \in \mathcal{B}_{\mathcal{N}}$ is a *Gaussian* distribution and marginal distributions of Gaussians are also Gaussian, we can write the requirement $\mathbb{D}(\mathbb{P}_z^1, \hat{\mathbb{P}}_z) \leq \rho_z$ in the definition of $\mathcal{B}_{\mathcal{N}}$ equivalently as $\mathbb{D}(\mathcal{N}(\mu_z, M_z), \hat{\mathbb{P}}_z) \leq \rho_z$, for any $z \in \mathcal{Z}$. This implies that $(\mu_w, M_w), (\mu_v, M_v)$ is feasible in the outer maximization problem in (A.14), proving that the optimal value in (A.14) is at least as large as that in (A.46). However, for any $(\mu_w, M_w), (\mu_v, M_v)$ feasible in the outer maximization in (A.14), the Gaussian distribution obtained from independent couplings $\mathbb{P} = \mathbb{P}_{x_0} \otimes (\otimes_{t=0}^{T-1} \mathbb{P}_{w_t}) \otimes (\otimes_{t=0}^{T} \mathbb{P}_{v_t})$ where $\mathbb{P}_z = \mathcal{N}(\mu_z, M_z)$ for every $z \in \mathcal{Z}$, is feasible in (A.46), proving that the two problems have the same optimal value. □

H.3 Proofs for §B.5 (Optimality of Linear Policies and Zero-Mean Distributions)

Proof of Proposition 3. We prove the result separately, for each divergence of interest. Recall that $\hat{\mu}_z = 0$, which implies that $\hat{M}_z = \hat{\Sigma}_z$, so the nominal distribution of interest is $\hat{\mathbb{P}}_z = \mathcal{N}(0, \hat{\Sigma}_z)$. Because the proof is done separately for each $z \in \mathcal{Z}$, we simplify notation by dropping the subscript z from quantities of interest such as $\mathbb{P}_z, \hat{\mathbb{P}}_z, \hat{\mu}_z, M_z, \Sigma_z$.

Consider any two Gaussian distributions $\mathbb{P}, \mathbb{P}' \in \mathcal{B}$ such that $\mathbb{E}_{\mathbb{P}}[zz^\top] = \mathbb{E}_{\mathbb{P}'}[zz^\top] = M$ and $\mu' = \hat{\mu} = 0$, which implies that $\Sigma' = \mathbb{E}_{\mathbb{P}'}[(z-\mu')(z-\mu')^\top] = M$. For the subsequent arguments, it helps to note that

$$\|\mu - \mathbb{E}_{\hat{\mathbb{P}}}[z]\|^2 + \mathrm{Tr}(\Sigma) - \|\mu' - \mathbb{E}_{\hat{\mathbb{P}}}[z]\|^2 - \mathrm{Tr}(\Sigma') = \mathrm{Tr}(M) - \mathrm{Tr}(M) = 0. \tag{A.47}$$

Also, it is useful to recall that the maps $X \mapsto \hat{\Sigma}^{1/2} X \hat{\Sigma}^{1/2}$ and $X \mapsto X^{\frac{1}{2}}$ are operator monotone on the cone of positive semidefinite matrices $\mathbb{S}_+^d$, that is, are functions $f : \mathbb{S}_+^d \to \mathbb{S}_+^d$ that satisfy $f(X) \succeq f(Y)$ for any $X, Y \in \mathbb{S}_+^d$ with $X \succeq Y$. (For a proof of these facts, see Theorem 1.5.9 and Theorem 4.2.3 in Bhatia (2009).)

The 2-Wasserstein distance $\mathbb{W}$. Recall from §G.1.1 and Proposition 8 that in this case, we have:

$$\mathcal{M}_{(\mu,M)} = \Big\{(\mu, M) \in \mathcal{M}_2^d : \big(\mathbb{G}\big((\mu, M), (0, \hat{M})\big)\big)^2 \leq \rho^2\Big\}.$$

To prove that $(\mu, M) \in \mathcal{M}_{(\mu,M)}$ implies that $(0, M) \in \mathcal{M}_{(\mu,M)}$, it therefore suffices to show that $(\mathbb{G}((0, M), (0, \hat{M})))^2 \leq (\mathbb{G}((\mu, M), (0, \hat{M})))^2$. To that end, using the expression of $\mathbb{G}$ from Definition 7 and applying (A.47) implies that:

$$\begin{aligned}
&\big(\mathbb{G}\big((0, M), (0, \hat{M})\big)\big)^2 - \big(\mathbb{G}\big((\mu, M), (0, \hat{M})\big)\big)^2 \\
&= -2\left(\mathrm{Tr}\left(\hat{\Sigma}^{1/2} M \hat{\Sigma}^{1/2}\right)^{1/2} - \mathrm{Tr}\left(\hat{\Sigma}^{1/2}(M - \mu\mu^\top)\hat{\Sigma}^{1/2}\right)^{1/2}\right) \leq 0,
\end{aligned} \tag{A.48}$$

where the inequality follows because $M \succeq M - \mu\mu^\top$ and the mappings $X \mapsto \hat{\Sigma}^{1/2} X \hat{\Sigma}^{1/2}$ and $X \mapsto X^{1/2}$ are operator monotone and $\mathrm{Tr}(X) \geq \mathrm{Tr}(Y)$ if $X \succeq Y$. This completes the argument.

The KL Divergence $\mathbb{K}$. Recall from Section G.1.2 and Proposition 9 that in this case, we have:

$$\mathcal{M}_{(\mu,M)} = \Big\{(\mu, M) \in \mathcal{M}_2^d : \mathbb{T}\big((\mu, M), (0, \hat{M})\big) \leq \rho\Big\}.$$

With the same proof strategy as above, using the expression of $\mathbb{T}$ from Definition 8 and applying (A.47), we obtain:

$$\begin{aligned}
&\mathbb{T}\big((0, M), (0, \hat{M})\big) - \mathbb{T}\big((\mu, M), (0, \hat{M})\big) \\
&= -\frac{1}{2}\mu^\top \hat{\Sigma}^{-1}\mu + \frac{1}{2}\,\mathrm{Tr}\left((M - M + \mu\mu^\top)\hat{\Sigma}^{-1}\right) - \frac{1}{2}\Big(\log\det\left(M\hat{\Sigma}^{-1}\right) \\
&\qquad - \log\det\left((M - \mu\mu)\hat{\Sigma}^{-1}\right)\Big) \leq 0,
\end{aligned}$$

where the inequality follows because the first two terms cancel out and the term in the last bracket is positive because $\log\det(X) \geq \log\det(Y)$ if $X \succeq Y$ and $M \succeq M - \mu\mu^\top$.

The Moment-Based Ambiguity Set $\mathbb{M}$. Recall from Section G.1.3 that we have:

$$\mathcal{M}_{(\mu,M)} = \{(\mu, M) \in \mathcal{M}_2^d : \mathbb{M}\left((\mu, M), (\hat{\mu}, \hat{M})\right) \leq \rho\}.$$

Therefore, the condition stated in Proposition 3 exactly ensures that Assumption 3 is satisfied. □

Proof of Theorem B. Because problem (A.9) has the same optimal value as problem (A.13) by Corollary 2, we prove the results for the optimal solution $\mathbb{P}^\star \in \mathcal{B}_\mathcal{N}$ to the outer maximization in (A.13) and the corresponding optimal policy $u^\star$ to the inner minimization in (A.13). (All these optimal solutions exist in view of Theorem A.)

We first recall a few important facts from Proposition 2 and its proof. By Proposition 2, the optimal value in problem (A.13) is the same as the optimal value in problem (A.14). The proof of Proposition 2 also shows that for any $\mathbb{P} \in \mathcal{B}_\mathcal{N}$ feasible in (A.13), the means and second moments of w and v evaluated under $\mathbb{P}$ would be feasible in (A.14), i.e., $\mathbb{P} \in \mathcal{B}_\mathcal{N}$ implies that $(\mu_w, M_w) \in \mathcal{M}_{(\mu_w,M_w)}$ and $(\mu_v, M_v) \in \mathcal{M}_{(\mu_v,M_v)}$. Moreover, a policy of the form $u = U\eta + q$ for $U \in \mathcal{U}$ and $q \in \mathbb{R}^{pT}$ is optimal in the inner minimization problems in (A.14) and in (A.13). Lastly, recall that the objective in (A.14) evaluated for the moment pairs $(\mu_w, M_w), (\mu_v, M_v)$ and policy $u = U\eta + q$ is given by:

$$\begin{aligned}
&f\big((q, U); (\mu_w, M_w), (\mu_v, M_v)\big) \\
&= \mathrm{Tr}\left((UD)^\top RUD + (G + HUD)^\top Q(G + HUD)\right)M_w + U^\top \bar{R}UM_v\big) \\
&\qquad + 2q^\top(\bar{R}UD + G^\top QH)\mu_w + 2q^\top \bar{R}U\mu_v + q^\top \bar{R}q.
\end{aligned}$$

Let us fix any $U \in \mathcal{U}$ and consider the inner minimization problem in (A.14), which corresponds to finding $q \in \mathbb{R}^{pT}$ to minimize the function f above. Because this is an unconstrained, convex optimization problem, the solution is given by the first-order optimality condition,

$$q^\star(U, \mu_w, \mu_v) = -\bar{R}^{-1}\bar{K}(\mu_w, \mu_v), \tag{A.49}$$

where $\bar{K}(\mu_w, \mu_v) = (\bar{R}UD + G^\top QH)\mu_w + \bar{R}U\mu_v$. In particular, the optimal choice $q^\star$ only depends on U and on the means μ_w, μ_v, and moreover, $q^\star = 0$ if $\mu_w = 0$ and $\mu_v = 0$.

In this context, consider the distribution $\mathbb{P}^\star \in \mathcal{B}_\mathcal{N}$ that is optimal in (A.13) and let $\mu_z^\star = \mathbb{E}_{\mathbb{P}^\star}[z]$ and $M_z^\star = \mathbb{E}_{\mathbb{P}^\star}[zz^\top]$ denote the mean and second moment matrix of $z \in \mathcal{Z}$ under $\mathbb{P}^\star$, respectively.

We prove our desired result by contradiction. Suppose that $\mathbb{P}^\star$ is such that at least one of $\mu_w^\star$ and $\mu_v^\star$ is non-zero and consider a Gaussian distribution $\tilde{\mathbb{P}}$ under which the first and second moment pairs for the random vectors w and v are respectively given by $(0, M_w^\star)$ and $(0, M_v^\star)$. That is, $\tilde{\mathbb{P}}$ has zero mean and the same second moments as $\mathbb{P}^\star$.

$\tilde{\mathbb{P}}$ is feasible in problem (A.13) by Assumption 3. We claim that the value of the objective in (A.14) corresponding to $\tilde{\mathbb{P}}$ is at least as large as the objective in (A.14) corresponding to $\mathbb{P}^\star$. Recalling the definition of f, we can evaluate the difference in these objectives for an arbitrary fixed $U \in \mathcal{U}$ and the optimal $q^\star$ from (A.49) as:

$$\begin{aligned}
&= f\left(\left(-\bar{R}^{-1}\bar{K}(\mu_w, \mu_v), U\right); (\mu_w, M_w), (\mu_v, M_v)\right) - f\left((0, U); (0, M_w), (0, M_v)\right) \\
&= 2\left(q^\star(U, \mu_w, \mu_v)\right)^\top \bar{K}(\mu_w, \mu_v) + \left(q^\star(U, \mu_w, \mu_v)\right)^\top \bar{R}q^\star(U, \mu_w, \mu_v)) \\
&= -\left(\bar{K}(\mu_w, \mu_v)\right)^\top \bar{R}^{-1}\bar{K}(\mu_w, \mu_v).
\end{aligned}$$

The last expression is non-positive because $\bar{R} \succ 0$. Because this holds for an arbitrary $U \in \mathcal{U}$, we conclude that $\tilde{\mathbb{P}}$ yields an objective in (A.14) at least as large as that achieved by $\mathbb{P}^\star$. Therefore, it is always optimal for nature to choose a distribution $\mathbb{P}^\star$ that is zero-mean.

For $\mathbb{P}^\star$ with zero mean, $q^\star = 0$ by (A.49), so the optimal policy can be taken as $u = U^\star\eta$ for some $U \in \mathcal{U}_\eta$. □

H.4 Proofs for §B.6 (Worst-Case Covariance)

H.4.1 Verifying Assumption 4.

We show that all ambiguity sets introduced in §A.2 also satisfy Assumption 4 (under a mild condition for the moment-based ambiguity set). This is summarized in the following result.

Proposition 10. *The ambiguity sets based on the divergences $\mathbb{W}$ and $\mathbb{K}$ introduced in §A.2 satisfy Assumption 4 if $\rho_z > 0$ for every $z \in \mathcal{Z}$. The moment-based ambiguity set also satisfies Assumption 4 if $\rho_z > 0$ and $\mathbb{M}((0,\Sigma_z),(0,\hat{\Sigma}_z))$ is differentiable in Σ_z on $\mathbb{S}^{d_z}_{++}$ and satisfies:*

$$\begin{aligned} \nabla_{\Sigma_z}\mathbb{M}((0,\Sigma_z),(0,\hat{\Sigma}_z))\Big|_{\Sigma_z=\Sigma_1} \succeq \nabla_{\Sigma_z}\mathbb{M}((0,\Sigma_z),(0,\hat{\Sigma}_z))\Big|_{\Sigma_z=\Sigma_2} \\ \Rightarrow\ \Sigma_1 \succeq \Sigma_2,\ \forall\,\Sigma_1,\Sigma_2 \in \mathbb{S}^{d_z}_{+},\ \forall\, z \in \mathcal{Z}. \end{aligned} \tag{A.50}$$

Proof. To simplify notation, we subsequently drop the subscript z from Σ_z, $\hat{\Sigma}_z$, ρ_z or d_z.

Wasserstein ambiguity sets. Set $g(\Sigma) = (\mathbb{G}((0,\Sigma),(0,\hat{\Sigma})))^2 - \rho^2$ where $\mathbb{G}$ is the Gelbrich distance. We claim that all requirements in Assumption 4 are satisfied if $\rho > 0$. We have:

$$g(\Sigma) = \mathrm{Tr}\Big(\Sigma + \hat{\Sigma} - 2\big(\hat{\Sigma}^{1/2}\Sigma\hat{\Sigma}^{1/2}\big)^{1/2}\Big) - \rho^2.$$

Proposition 8 and the discussion that verified Assumption 2-(ii) imply that the set $\mathcal{M}_\Sigma = \big\{\Sigma \in \mathbb{S}^d_+ : g(\Sigma) \le 0\big\}$ is convex and compact. Moreover, g is differentiable in Σ on $\mathbb{S}^d_{++}$ because $\mathbb{G}$ is. Requirements (i) and (ii) are readily satisfied because g and the Gelbrich distance $\mathbb{G}$ are minimized at $\Sigma = \hat{\Sigma}$ and we have $g(\hat{\Sigma}) = -\rho^2 < 0$ when $\rho > 0$. To check (iii), note that the gradient ∇g is:

$$\nabla g(\Sigma) = I - \hat{\Sigma}^{\frac{1}{2}}(\hat{\Sigma}^{\frac{1}{2}}\Sigma\hat{\Sigma}^{\frac{1}{2}})^{-\frac{1}{2}}\hat{\Sigma}^{\frac{1}{2}}.$$

Fix $\Sigma_1, \Sigma_2 \in \mathbb{S}^{d_z}_+$ and suppose that $\nabla g(\Sigma_1) \succeq \nabla g(\Sigma_2)$. This implies

$$\hat{\Sigma}^{\frac{1}{2}}(\hat{\Sigma}^{\frac{1}{2}}\Sigma_1\hat{\Sigma}^{\frac{1}{2}})^{-\frac{1}{2}}\hat{\Sigma}^{\frac{1}{2}} \preceq \hat{\Sigma}^{\frac{1}{2}}(\hat{\Sigma}^{\frac{1}{2}}\Sigma_2\hat{\Sigma}^{\frac{1}{2}})^{-\frac{1}{2}}\hat{\Sigma}^{\frac{1}{2}} \ \Rightarrow\ (\hat{\Sigma}^{\frac{1}{2}}\Sigma_1\hat{\Sigma}^{\frac{1}{2}})^{-\frac{1}{2}} \preceq (\hat{\Sigma}^{\frac{1}{2}}\Sigma_2\hat{\Sigma}^{\frac{1}{2}})^{-\frac{1}{2}} \ \Rightarrow\ \Sigma_1 \succeq \Sigma_2,$$

where the first implication follows because pre- and post-multiplying by $\hat{\Sigma}^{-\frac{1}{2}}$ preserves ordering, and the second implication follows because the operator $X \mapsto X^{-1/2}$ reverses ordering on $\mathbb{S}^d_+$ (i.e., $X_1 \succeq X_2$ implies $(X_1)^{-1/2} \preceq (X_2)^{-1/2}$) and pre- and post-multiplying by $\hat{\Sigma}^{1/2}$ preserves ordering.

Kullback-Leibler ambiguity sets. Set $g(\Sigma) = \mathbb{T}((0,\Sigma),(0,\hat{\Sigma})) - \rho$ where $\mathbb{T}$ is the KL-type divergence between moments introduced in Definition 8. We claim that all requirements in Assumption 4 are satisfied if $\rho > 0$. We have:

$$g(\Sigma) = \frac{1}{2}\Big(\mathrm{Tr}\big(\Sigma\hat{\Sigma}^{-1}\big) - \log\det\big(\Sigma\hat{\Sigma}^{-1}\big) - d\Big) - \rho.$$

Proposition 9 and the discussion that verified Assumption 2-(ii) imply that the set $\mathcal{M}_\Sigma = \big\{\Sigma \in \mathbb{S}^d_{++} : g(\Sigma) \le 0\big\}$ is convex and compact. Moreover, g is differentiable in Σ on $\mathbb{S}^d_{++}$ because the $\mathrm{Tr}(\cdot)$ and $\log\det(\cdot)$ operators are differentiable on $\mathbb{S}^d_{++}$. Requirements (i) and (ii) are readily satisfied because g and $\mathbb{T}$ are both minimized at $\Sigma = \hat{\Sigma}$ and we have $g(\hat{\Sigma}) = -\rho < 0$ when $\rho > 0$. To check (iii), note that the gradient ∇g can be written using matrix-calculus rules as:

$$\nabla g(\Sigma) = \frac{1}{2}(\hat{\Sigma}^{-1} - \Sigma^{-1}).$$

That $\nabla g(\Sigma_1) \succeq \nabla g(\Sigma_2)$ implies $\Sigma_1 \succeq \Sigma_2$ follows because $X \mapsto X^{-1}$ is order-reversing on $\mathbb{S}^d_{++}$.

Moment-Based ambiguity sets. Define $g(\Sigma) = \mathbb{M}((0,\Sigma),(0,\hat{\Sigma})) - \rho$. The conditions follow immediately by recognizing that $\mathbb{M}$ being differentiable and satisfying $\mathbb{M}(m,m) = 0$ for all $m \in \mathcal{M}^d_2$ (as required in its definition in §A.2) imply that $\nabla g(\hat{\Sigma}) = 0$ and $g(\hat{\Sigma}) < 0$ for $\rho > 0$. Condition (iii) is exactly equivalent to the stated condition (A.50) on $\mathbb{M}$. □

Proof of Theorem C. Problem (A.9) has the same optimal value as problem (A.14), so we consider the latter formulation. Recall that $\hat{\mu}_z = 0$, so Theorem B allows us to restrict attention to zero-mean, Gaussian distributions $\mathbb{P}_z$ for nature and linear control policies $u = U\eta$.

We prove a slightly stronger result: that for any linear control policy, nature's optimal choice satisfies $\Sigma_z^\star \succeq \hat{\Sigma}_z$, for every $z \in \mathcal{Z}$. Consider any $U \in \mathcal{U}$ and define auxiliary matrices

$$\Upsilon_w = (UD)^\top RUD + (G + HUD)^\top Q(G + HUD) \quad \text{and} \quad \Upsilon_v = U^\top RU + (HU)^\top RHU.$$

Because $R \succ 0$ and $Q \succeq 0$, we have $\Upsilon_w \succeq 0$ and $\Upsilon_v \succeq 0$. Substituting into the objective of (A.14) yields:

$$\mathrm{Tr}(\Upsilon_w \Sigma_w) + \mathrm{Tr}(\Upsilon_v \Sigma_v). \tag{A.51}$$

Function (A.51) is separable in Σ_w and Σ_v. In fact, because Σ_w and Σ_v are block-diagonal matrices (because the random noise terms $z \neq z' \in \mathcal{Z}$ satisfy the SMO condition), this objective further decomposes based on each component $z \in \mathcal{Z}$. For instance, with $\mathcal{Z}_1 = \{x_0, w_0, \ldots, w_{T-1}\}$, we have:

$$\mathrm{Tr}(\Upsilon_w \Sigma_w) = \sum_{z \in \mathcal{Z}_1} \Upsilon_w^{(z)} \Sigma_z,$$

where each $\Upsilon_w^{(z)}$ is the diagonal block of Υ_w corresponding to block Σ_z in $\Sigma_w^\star$. A similar decomposition applies to the objective over Σ_v, with $\Upsilon_v^{(z)}$ denoting the blocks in Υ_w. Because the constraints $(\Sigma_w, \Sigma_v) \in \mathcal{M}_{\Sigma_w} \times \mathcal{M}_{\Sigma_v}$ are also separable by $z \in \mathcal{Z}$, problem (A.14) is separable by $z \in \mathcal{Z}$:

$$\max_{(\Sigma_w, \Sigma_v) \in \mathcal{M}_{\Sigma_w} \times \mathcal{M}_{\Sigma_v}} \mathrm{Tr}(\Upsilon_w \Sigma_w + \Upsilon_v \Sigma_v) = \sum_{z \in \mathcal{Z}_1} \max_{\Sigma_z \in \mathcal{M}_{\Sigma_z}} \mathrm{Tr}(\Upsilon_w^{(z)} \Sigma_z) + \sum_{z \in \mathcal{Z} \setminus \mathcal{Z}_1} \max_{\Sigma_z \in \mathcal{M}_{\Sigma_z}} \mathrm{Tr}(\Upsilon_v^{(z)} \Sigma_z). \tag{A.52}$$

We prove the result for a given $z \in \mathcal{Z}$. Without loss of generality, consider the problem:

$$\text{(SDP)}_1 \qquad \begin{array}{ll} \max\limits_{\Sigma_z} & \mathrm{Tr}(\Upsilon_w^{(z)} \Sigma_z) \\ \text{s.t.} & g(\Sigma_z) \leq 0 \\ & \Sigma_z \succeq 0. \end{array}$$

Instead of solving this problem, we consider the following relaxation:

$$\text{(SDP)}_2 \qquad \begin{array}{ll} \max\limits_{\Sigma_z} & \mathrm{Tr}(\Upsilon_w^{(z)} \Sigma_z) \\ \text{s.t.} & g(\Sigma_z) \leq 0 \\ & \Sigma_z \succeq \epsilon I, \end{array}$$

where ϵ satisfies $0 < \epsilon < \lambda_{\min}(\hat{\Sigma}_z)$. Subsequently, we will prove that an optimal solution to (SDP)$_2$ exists satisfying $\Sigma_z^\star \succeq \hat{\Sigma}_z$, which (because $\hat{\Sigma}_z \succ \epsilon I$) will imply that the last constraint in (SDP)$_2$ will not be binding and $\Sigma_z^\star$ will also be optimal in (SDP)$_1$.

Consider problem (SDP)$_2$. Because $\Upsilon_w \succeq 0$, its principal sub-matrices are positive semidefinite, so $\Upsilon_w^{(z)} \succeq 0$. Let $\Sigma_z^\star$ be a maximizer in the optimization problem above. (The Weierstrass Theorem, which applies because the objective is linear and Assumption 2 guarantees that the feasible set $\mathcal{M}_{\Sigma_z}$ is compact, implies that a maximizer exists.) We prove that a maximizer exists so that $\Sigma_w^\star \succeq \hat{\Sigma}_w$.

(a) If $\Upsilon_w^{(z)} = 0$, we can replace $\Sigma_z^\star$ with $\hat{\Sigma}_z$, which satisfies $\Sigma_z^\star \succeq \hat{\Sigma}_z$ and is optimal.

(b) If $\Upsilon_w^{(z)} \neq 0$, with λ_z and Λ_z as the dual variables for the constraints above, the KKT conditions for this maximization problem are:

$$-\Upsilon_w^{(z)} + \lambda_z \nabla g(\Sigma_z^\star) - \Lambda_z = 0 \tag{Stationarity}$$

$$\lambda_z \, g(\Sigma_z^\star) = 0, \ \mathrm{Tr}(\Lambda_z^\top \Sigma_z^\star) = 0 \tag{Complementary Slackness}$$

$$\lambda_z \geq 0, \Lambda_z \in \mathbb{S}_+^n \tag{Dual Feasibility}$$

In writing these KKT conditions, the use of the gradient ∇g was valid because the feasible set of (SDP)$_2$ is contained in $\mathbb{S}_{++}^{d_z}$ and g is differentiable on the latter set, by Assumption 4. Moreover, the

KKT conditions are necessary for optimality here because we have a convex optimization problem and Slater's condition is satisfied because $g(\hat{\Sigma}_z) < 0$ due to Assumption 4-(i).

The Stationarity condition implies:

$$\lambda_z \, \nabla g(\Sigma_z^\star) = \Upsilon_w^{(z)} + \Lambda_z.$$

Because $\Upsilon_w^{(z)} \succeq 0$ and $\Lambda_z \succeq 0$, we have $\lambda_z \, \nabla g(\Sigma_z^\star) \succeq 0$. Moreover, if $\lambda_z = 0$ above, we must have $\Upsilon_w^{(z)} = \Lambda_z = 0$, which contradicts our standing assumption that $\Upsilon_w^{(z)} \neq 0$.

If $\lambda_z \neq 0$, because $\nabla g(\Sigma_z^\star) \succeq 0$ by the argument above, we have:

$$\nabla g(\Sigma_z^\star) \succeq 0 = \nabla g(\hat{\Sigma}_z), \tag{A.53}$$

where the last equality follows from Assumption 4-(ii), which together with the convexity and differentiability of g implies that $\nabla g(\hat{\Sigma}_z) = 0$. But then, (A.53) and Assumption 4-(iii) imply $\Sigma_z^\star \succeq \hat{\Sigma}_z$, which completes our proof. □

I Proofs for Section C

I.1 Concavity and β-Smoothness for the Objective

Recall the notation from Section C whereby $f(\Sigma_w, \Sigma_v)$ is the optimal value of the classical LQG problem for a Gaussian distribution with zero mean and covariances Σ_w, Σ_v.

Definition 9 (β-smoothness)**.** *For $\mathcal{M}_1, \mathcal{M}_2 \subseteq \mathbb{S}_+^d$, the function $f : \mathcal{M}_1 \times \mathcal{M}_2 \to \mathbb{R}$ is called β-smooth for some $\beta > 0$ if*

$$|\nabla f(\Sigma_1, \Sigma_2) - \nabla f(\Sigma_1', \Sigma_2')| \leq \beta \left(\|\Sigma_1 - \Sigma_1'\|_{\mathrm{F}}^2 + \|\Sigma_2 - \Sigma_2'\|_{\mathrm{F}}^2 \right)^{\frac{1}{2}} \quad \forall \Sigma_1, \Sigma_2 \in \mathcal{M}_1, \Sigma_2, \Sigma_2' \in \mathcal{M}_2.$$

Proof of Proposition 4. The function $f(\Sigma_w, \Sigma_v)$ is concave because the objective function of the inner minimization problem in (A.14) is linear (and hence concave) in Σ_w and Σ_v and because concavity is preserved under minimization. By Corollary 3, the value of $f(\Sigma_w, \Sigma_v)$ is given by (A.35), where Λ_t for $t \in [T-1]$, is a function of (Σ_w, Σ_v) defined recursively through the Kalman filter equations (A.33). Note that all inverse matrices in (A.33) are well-defined because any $\Sigma_v \in \mathcal{M}_{\Sigma_v}^+$ is strictly positive definite. Therefore, Λ_t constitutes a proper rational function, i.e., a ratio of two polynomials with the polynomial in the denominator being strictly positive, for every $t \in [T-1]$. Thus, $f(\Sigma_w, \Sigma_v)$ is infinitely often continuously differentiable on a neighborhood of $\mathcal{M}_{\Sigma_w} \times \mathcal{M}_{\Sigma_v}^+$.

Because $f(\Sigma_w, \Sigma_v)$ is concave and at least twice continuously differentiable, it is β-smooth on $\mathcal{M}_{\Sigma_w} \times \mathcal{M}_{\Sigma_v}^+$ if and only if the largest eigenvalue of the negative of the Hessian matrix, $|\lambda_{\max}\left(-\nabla^2 f(\Sigma_w, \Sigma_v)\right)|$, is bounded above by β on $\mathcal{M}_{\Sigma_w} \times \mathcal{M}_{\Sigma_v}^+$. The value $|\lambda_{\max}\left(-\nabla^2 (f(\Sigma_w, \Sigma_v))\right)|$ coincides with the spectral norm of $\nabla^2 f(\Sigma_w, \Sigma_v)$, so we can set:

$$\beta = \sup_{\Sigma_w \in \mathcal{M}_{\Sigma_w}, \Sigma_v \in \mathcal{M}_{\Sigma_v}^+} \|\nabla^2 f(\Sigma_w, \Sigma_v)\|_2, \tag{A.54}$$

where $\|\cdot\|_2$ denotes the spectral norm. The supremum in the above maximization problem is finite and attained thanks to Weierstrass' theorem, which applies because $f(\Sigma_w, \Sigma_v)$ is twice continuously differentiable and the spectral norm is continuous, while the sets $\mathcal{M}_{\Sigma_w}$ and $\mathcal{M}_{\Sigma_v}^+$ are compact by Assumption 2. This observation completes the proof. □

I.2 Bisection Algorithm for the Linearization Oracle

We now show that the direction-finding subproblem (A.18) can be solved efficiently via bisection for the Wasserstein and KL ambiguity sets.

I.2.1 Wasserstein Ambiguity.

We establish that (A.18) can be reduced to the solution of $2T+1$ univariate algebraic equations. To this end, let $\Gamma_z = \nabla_{\Sigma_z} f(\Sigma_w^{(k)}, \Sigma_v^{(k)})$ denote the gradient at the current iterate with respect to Σ_z. Note that a routine calculation (see proof of Theorem C) shows that $\Gamma_z \succeq 0$. If $\Gamma_z = 0$, then $\Sigma_z^\star = \hat{\Sigma}_z$ is trivially optimal in (1). From now on we thus assume that $\Gamma_z \neq 0$.

Proposition 11. *(Nguyen et al., 2023, Proposition A.4) If $\Gamma_z \in \mathbb{S}_+^d$, $\Gamma_z \neq 0$, $\hat{\Sigma}_z \in \mathbb{S}_{++}^{d_z}$, and $\rho_z > 0$, then the unique optimal solution to the problem*

$$\begin{array}{ll} \max & \langle \Gamma_z, \Sigma_z - \Sigma_z^{(k)} \rangle \\ \text{s.t.} & \Sigma_z \in \mathbb{S}_+^d \\ & \mathbb{G}(\Sigma_z, \hat{\Sigma}_z) \leq \rho_z \\ & \Sigma_z \succeq \lambda_{\min}(\hat{\Sigma}_z) I \end{array} \tag{A.55}$$

is $\Sigma_z^\star = (\gamma^\star)^2 (\gamma^\star I - \Gamma_z)^{-1} \hat{\Sigma}_z (\gamma^\star I - \Gamma_z)^{-1}$, where $\gamma^\star$ is the unique solution of

$$\rho_z^2 - \langle \hat{\Sigma}_z, (I - \gamma^\star (\gamma^\star I - \Gamma_z)^{-1})^2 \rangle = 0 \tag{A.56}$$

satisfying

$$\underline{\gamma} = \lambda_1 (1 + (p_1^\top \hat{\Sigma}_z p_1)^{\frac{1}{2}} / \rho_z) \leq \gamma^\star \leq \lambda_1 (1 + \mathrm{Tr}(\hat{\Sigma}_z)^{\frac{1}{2}} / \rho_z) = \overline{\gamma}, \tag{A.57}$$

where $\lambda_1 = \lambda_{\max}(\Gamma_z)$ and p_1 is an eigenvector for λ_1.

In practice, we need to solve the algebraic equation (A.56) numerically. The numerical error in approximating $\gamma^\star$ should be contained to ensure that $\Sigma_z^\star$ approximates the exact maximizer of problem (A.55). The next proposition shows that for any tolerance $\delta \in (0,1)$, a δ-approximate solution of (A.55) can be computed with an efficient bisection algorithm.

Proposition 12. *(Nguyen et al., 2023, Theorem 6.4) For any fixed $\rho_z > 0, \hat{\Sigma}_z \in \mathbb{S}_{++}^d$ and $\Gamma_z \in \mathbb{S}_+^d, \Gamma_z \neq 0$, define $\mathcal{G}_{\Sigma_z} = \{\Sigma_z \in \mathbb{S}_+^d : \mathbb{G}(\Sigma_z, \hat{\Sigma}_z) \leq \rho_z, \hat{\Sigma}_z \succeq \lambda_{\min}(\hat{\Sigma}_z)\}$ as the feasible set of problem (A.55), and let $\hat{\Sigma}_z \in \mathcal{G}_{\Sigma_z}$ be any reference covariance matrix. Then, Algorithm A.2 with $\delta \in (0,1)$, $\varphi(\gamma) = \gamma(\rho_z^2 + \langle \gamma(\gamma I - \Gamma_z)^{-1} - I, \hat{\Sigma}_z \rangle) - \langle \Gamma_z, \Sigma_z^{(k)} \rangle$, $L(\gamma) = (\gamma)^2 (\gamma I - \Gamma_z)^{-1} \hat{\Sigma}_z (\gamma I - \Gamma_z)^{-1}$ for any $\gamma > \lambda_{\max}(\Gamma_z)$, $\overline{\gamma}$ and $\underline{\gamma}$ as defined in (A.57) returns in finite time a matrix $\Sigma_z^\delta \in \mathbb{S}_+^d$ that satisfies: $\Sigma_z^\delta \in \mathcal{G}_{\Sigma_z}$ (feasible) and $\langle \Gamma_z, \Sigma_z^\delta - \Sigma_z^{(k)} \rangle \geq \delta \max_{\Sigma_z \in \mathcal{G}_{\Sigma_z}} \langle \Gamma_z, \Sigma_z - \Sigma_z^{(k)} \rangle$ (δ-Suboptimal).*

Algorithm A.2 Bisection algorithm to compute $L_{\Sigma_z}^\delta$

Input: nominal covariance matrix $\hat{\Sigma}_z \in \mathbb{S}_{++}^{d_z}$, radius $\rho_z \in \mathbb{R}_{++}$,
reference covariance matrix $\Sigma_z \in \mathcal{M}_z$,
gradient matrix $\Gamma_z \in \mathbb{S}_+^d$, $\Gamma_z \neq 0$, precision $\delta \in (0,1)$,
dual objective function $\phi : \mathbb{R} \to \mathbb{R}$ and estimation function $L(\gamma) : \mathbb{R} \to \mathbb{S}_+^d$
upper bound $\overline{\gamma}$ and lower bound $\underline{\gamma}$

repeat
  set $\gamma \leftarrow (\overline{\gamma} + \underline{\gamma})/2$
  if $\frac{\mathrm{d}\phi}{\mathrm{d}\gamma}(\gamma) < 0$ **then** set $\underline{\gamma} \leftarrow \gamma$ **else** set $\overline{\gamma} \leftarrow \gamma$ **endif**
until $\frac{\mathrm{d}\phi}{\mathrm{d}\gamma}(\gamma) > 0$ and $\langle L(\gamma) - \Sigma_z^{(k)}, \Gamma_z \rangle \geq \delta \phi(\gamma)$

Output: $L(\gamma)$

In summary, for any $z \in \mathcal{Z}$, Algorithm A.2 computes a δ-approximate solutions to the direction-finding subproblem (A.18) with $\Gamma_z = \nabla_{\Sigma_z} f(\Sigma_w^{(k)}, \Sigma_v^{(k)})$.

I.2.2 Kullback-Leibler Divergence

We first establish that (A.18) can be reduced to the solution of a univariate algebraic equation. Let $\Gamma_z = \nabla_{\Sigma_z} f(\Sigma_w^{(k)}, \Sigma_v^{(k)})$ denote the gradient at the current iterate.

Theorem A. *If $\hat{\Sigma}_z \in \mathbb{S}^{d_z}_{++}$, $\Gamma_z \in \mathbb{S}^d_+$, $\Gamma_z \neq 0$, $\Sigma_z^{(k)} \in \mathbb{S}^d_+$, and $\rho_z \in \mathbb{R}_{++}$, then the problem*

$$\begin{array}{ll} \max\limits_{\Sigma_z \in \mathbb{S}^d_+} & \langle \Gamma_z, \Sigma_z - \Sigma_z^{(k)} \rangle \\ \text{s.t.} & \mathbb{T}(\Sigma_z, \hat{\Sigma}_z) \leq \rho_z \end{array} \tag{A.58}$$

is uniquely solved by $L^\star_{\Sigma_z} = \gamma^\star(\gamma^\star \hat{\Sigma}_z^{-1} - \Gamma_z)^{-1}$, where $\gamma^\star$ is the unique solution of the following nonlinear equation

$$2\rho_z = \log\det(I - \hat{\Sigma}_z\Gamma_z/\gamma^\star) + \mathrm{Tr}((\gamma^\star I - \hat{\Sigma}_z\Gamma_z)^{-1}\hat{\Sigma}_z\Gamma_z), \tag{A.59}$$

satisfying

$$\lambda_1 < \gamma^\star \leq \lambda_1 \left(1 + \frac{d}{\rho_z}\right), \tag{A.60}$$

where λ_1 is the largest eigenvalue of $\hat{\Sigma}^{\frac{1}{2}}\Gamma_z\hat{\Sigma}^{\frac{1}{2}}$.

The proof relies on Lemma A.1 in Taşkesen et al. (2021), restated below for completeness.

Lemma G. *(Taşkesen et al., 2021, Lemma A.1) Fix $\hat{\Sigma} \in \mathbb{S}^{d_z}_{++}$, then for any $F \in \mathbb{S}^d$ and $\rho > 0$, the optimization problem*

$$\begin{array}{ll} \sup\limits_{\Sigma \in \mathbb{S}^d_{++}} & \mathrm{Tr}(F\Sigma) \\ \text{s.t.} & \mathbb{T}(\Sigma, \hat{\Sigma}) \leq \rho \end{array} \tag{A.61}$$

admits the strong dual formulation

$$\begin{array}{ll} \inf\limits_{\gamma \in \mathbb{R}} & 2\gamma\rho - \gamma\log\det(I - \hat{\Sigma}F/\gamma) \\ \text{s.t.} & \gamma \geq 0,\ \gamma\hat{\Sigma}^{-1} \succ F. \end{array} \tag{A.62}$$

Proof of Theorem A. By Lemma G, for any $\hat{\Sigma}_z \in \mathbb{S}^{d_z}_{++}$, problem (A.58) admits the strong dual

$$\begin{array}{ll} \inf\limits_{\gamma \in \mathbb{R}} & 2\gamma\rho_z - \gamma\log\det(I - \hat{\Sigma}_z\Gamma_z/\gamma) - \langle \Gamma_z, \Sigma_z^{(k)} \rangle \\ \text{s.t.} & \gamma \geq 0,\ \gamma\hat{\Sigma}_z^{-1} \succ \Gamma_z. \end{array} \tag{A.63}$$

For ease of exposition, we let $h(\gamma) = 2\gamma\rho_z - \gamma\log\det(I - \hat{\Sigma}_z\Gamma_z/\gamma)$, which is the objective function of problem (A.63) without the constant term $\langle \Gamma_z, \Sigma_z^{(k)} \rangle$. The gradient of $h(\gamma)$ satisfies

$$\nabla h(\gamma) = 2\rho_z - \log\det(I - \hat{\Sigma}_z\Gamma_z/\gamma) - \mathrm{Tr}((\gamma - \hat{\Sigma}_z\Gamma_z)^{-1}\hat{\Sigma}_z\Gamma_z).$$

By the above expression of $\nabla h(\gamma)$ and the strict convexity of $h(\gamma)$ for $\gamma \geq 0$ and $\gamma\hat{\Sigma}_z^{-1} \succ \Gamma_z$, the value $\gamma^\star$ that solves (A.59) is the unique minimizer of (A.63).

Now, we show that $\Sigma_z^\star$ obtained as $\gamma^\star(\gamma^\star\hat{\Sigma}_z^{-1} - \Gamma_z)^{-1}$ is feasible in (A.58), that is, it satisfies

$$\mathrm{Tr}(\Sigma_z^\star\hat{\Sigma}_z^{-1}) - \log\det(\Sigma_z^\star\hat{\Sigma}_z^{-1}) - d_z \leq 2\rho_z. \tag{A.64}$$

Through the Woodbury matrix identity we have

$$\Sigma_z^\star\hat{\Sigma}_z^{-1} = \gamma^\star(\gamma^\star\hat{\Sigma}_z^{-1} - \Gamma_z)^{-1}\hat{\Sigma}_z^{-1} = I + \hat{\Sigma}_z(\gamma^\star - \Gamma_z\hat{\Sigma}_z)^{-1}\Gamma_z.$$

By replacing the $\Sigma_z^\star\hat{\Sigma}_z^{-1}$ term inside the trace operator in (A.64), we have

$$\mathrm{Tr}(I + \hat{\Sigma}_z(\gamma^\star - \Gamma_z\hat{\Sigma}_z)^{-1}\Gamma_z) - \log\det(\gamma^\star(\gamma^\star\hat{\Sigma}_z^{-1} - \Gamma_z)^{-1}\hat{\Sigma}_z^{-1}) - d_z = 2\rho_z,$$

where the equality follows because $\gamma^\star$ solves (A.59).

Next, we show that $\Sigma_z^\star = \gamma^\star(\gamma^\star\hat{\Sigma}_z^{-1} - \Gamma_z)^{-1}$ is optimal in (A.58). To this end, note that the objective value of $\Sigma_z^\star$ in (A.58) coincides with the optimal value of its strong dual (A.63) because

$$\begin{aligned} \mathrm{Tr}(\Sigma_z^\star\Gamma_z) &= \mathrm{Tr}((I - \hat{\Sigma}_z\Gamma_z/\gamma^\star)^{-1}\hat{\Sigma}_z\Gamma_z) \\ &= \gamma^\star\rho_z - \gamma^\star\log\det(I - \hat{\Sigma}_z\Gamma_z/\gamma^\star) \end{aligned}$$

$$= \inf_{\gamma \geq 0,\, \gamma \hat{\Sigma}_z^{-1} \succ \Gamma_z} \gamma \rho_z - \gamma \log\det(I - \hat{\Sigma}_z \Gamma_z / \gamma),$$

where the second equality follows because $\gamma^\star$ solves (A.59).

It remains to show the upper and lower bounds on $\gamma^\star$. The lower bound on $\gamma^\star$ follows because $\gamma^\star$ is feasible in (A.63) and thus satisfies $\gamma^\star \hat{\Sigma}_z^{-1} \succ \Gamma_z$. For any $\gamma^\star$ such that $\gamma^\star \hat{\Sigma}_z^{-1} \succ \Gamma_z$, the algebraic equation (A.56) yields

$$\begin{aligned}
0 &= \rho_z - \log\det(I - \hat{\Sigma}_z \Gamma_z / \gamma^\star) - \mathrm{Tr}((\gamma^\star I - \hat{\Sigma}_z \Gamma_z)^{-1} \hat{\Sigma}_z \Gamma_z) \\
&> \rho_z - \mathrm{Tr}((\gamma^\star I - \hat{\Sigma}_z \Gamma_z)^{-1} \hat{\Sigma}_z \Gamma_z) \\
&= \rho_z - \mathrm{Tr}((\gamma^\star I - \hat{\Sigma}_z^{\frac{1}{2}} \Gamma_z \hat{\Sigma}_z^{\frac{1}{2}})^{-1} \hat{\Sigma}_z^{\frac{1}{2}} \Gamma_z \hat{\Sigma}_z^{\frac{1}{2}}) \\
&= \rho_z - \sum_{i=1}^{d_z} \frac{\lambda_i}{\gamma^\star - \lambda_i},
\end{aligned} \tag{A.65}$$

where the inequality holds because the eigenvalues of $I - \hat{\Sigma}_z \Gamma_z / \gamma^\star$ are in $(0, 1)$ under the condition $\gamma^\star \hat{\Sigma}_z^{-1} \succ \Gamma_z$, and this implies that $\log\det(I - \hat{\Sigma}_z \Gamma_z / \gamma^\star)$ is negative. In the last expression, $\lambda_1, \ldots, \lambda_{d_z}$ denote the eigenvalues of $\hat{\Sigma}_z^{\frac{1}{2}} \Gamma_z \hat{\Sigma}_z^{\frac{1}{2}}$ indexed in descending order. We thus have

$$\rho_z < \sum_{i=1}^{d_z} \frac{\lambda_i}{\gamma^\star - \lambda_i} \leq \frac{d_z \lambda_1}{\gamma^\star - \lambda_1},$$

where the second inequality holds immediately because $\gamma^\star > \lambda_1$. The above inequality implies the second inequality in (A.60), and therefore the correctness of the upper bound. This observation concludes the proof. □

J Proofs for §E

Proof of Lemma A. Assertions (i) and (ii) are easy to verify. Details are omitted for brevity. Assertion (iii) is a direct consequence of assertion (ii). Indeed, if $N \in \mathcal{T}^{k \times k}$ is the unique Toeplitz matrix whose blocks satisfy (A.23), then $O = MN$ is the identity matrix in $\mathcal{T}^{k \times k}$, that is, we have $O_0 = I$ and $O_t = 0$ for every $t \in \mathbb{N}_+$. □

Proof of Lemma B. To facilitate following the proof, we restate the identities to prove:

$$u_t = \sum_{s=0}^{t} [(UD)_{t-s} w_s + (U)_{t-s} v_s] \quad \text{and} \quad x_t = \sum_{s=0}^{t} [(G + HUD)_{t-s} w_s + (HU)_{t-s} v_s]$$

$$\Sigma_{u_t} = \mathbb{E}_{\mathbb{P}}\left[u_t u_t^\top\right] = \sum_{s=0}^{t} \left((UD)_{t-s} \Sigma_{w_s} (UD)_{t-s}^\top + (U)_{t-s} \Sigma_{v_s} (U)_{t-s}^\top\right)$$

$$\Sigma_{x_t} = \mathbb{E}_{\mathbb{P}}\left[x_t x_t^\top\right] = \sum_{s=0}^{t} \left((G + HUD)_{t-s} \Sigma_{w_s} (G + HUD)_{t-s}^\top + (HU)_{t-s} \Sigma_{v_s} (HU)_{t-s}^\top\right)$$

$$\frac{1}{T} \sum_{t=0}^{T-1} \mathbb{E}_{\mathbb{P}}\left[x_t^\top Q_0 x_t + u_t^\top R_0 u_t\right] = \frac{1}{T} \sum_{s=0}^{T-1} \mathrm{Tr}\left(\Sigma_{w_s} \left(\sum_{t=0}^{T-1-s} M_t\right) + \Sigma_{v_s} \left(\sum_{t=0}^{T-1-s} N_t\right)\right).$$

Recall from Lemma D that $u = U\eta = U(Dw + v)$. As $U \in \mathcal{T}^{m \times p}$, $C \in \mathcal{T}^{p \times n}$ and $G \in \mathcal{T}^{n \times n}$, Lemma A (ii) further implies that $UD = UCG \in \mathcal{T}^{m \times n}$. Hence, we find

$$u_t = \sum_{s=0}^{t} \left((UD)_{t-s} w_s + (U)_{t-s} v_s\right).$$

As $x = Hu + Gw = HU(Dw + v) + Gw$ and as $H \in \mathcal{T}^{n\times m}$, $UD \in \mathcal{T}^{m\times n}$ and $G \in \mathcal{T}^{n\times n}$, the expression for x_t readily follows from a similar argument.

The expressions for the covariance matrices $\Sigma_{u_t}, \Sigma_{x_t}$ follow readily by recognizing that the random vectors in the set $\mathcal{Z}$ (i.e., $x_0, \{w_s\}_{s=0}^{\infty}, \{v_s\}_{s=0}^{\infty}$) have zero mean and are uncorrelated.

The average expected cost of (x, u) over the first T periods can then be expressed as

$$\begin{aligned}
\frac{1}{T}\sum_{t=0}^{T-1}\mathbb{E}_{\mathbb{P}}\left[x_t^\top Q_0 x_t + u_t^\top R_0 u_t\right] &= \frac{1}{T}\sum_{t=0}^{T-1}\mathrm{Tr}\left(\Sigma_{x_t}Q_0 + \Sigma_{u_t}R_0\right) \\
&= \frac{1}{T}\sum_{t=0}^{T-1}\sum_{s=0}^{t}\mathrm{Tr}\left(\Sigma_{w_s}M_{t-s} + \Sigma_{v_s}N_{t-s}\right) \\
&= \frac{1}{T}\sum_{s=0}^{T-1}\sum_{t=s}^{T-1}\mathrm{Tr}\left(\Sigma_{w_s}M_{t-s} + \Sigma_{v_s}N_{t-s}\right) \\
&= \frac{1}{T}\sum_{s=0}^{T-1}\mathrm{Tr}\left(\Sigma_{w_s}\left(\sum_{t=0}^{T-1-s}M_t\right) + \Sigma_{v_s}\left(\sum_{t=0}^{T-1-s}N_t\right)\right).
\end{aligned}$$

Here, the first equality holds because u_t, x_t have zero mean; the second equality uses (A.24b)-(A.24c) and the expressions for M_t and N_t; the third equality is obtained by interchanging the order of summing over t and s; and the fourth inequality follows from an index shift $t \leftarrow t - s$. □

Proof of Proposition 5. Select any $\mathbb{P} \in \overline{\mathcal{B}}^{\infty}$ and $U \in \mathcal{U}_{\infty}$. We claim that the expression of the long-run-average costs in Lemma B achieves a finite limit as $T \to \infty$. To that end, recalling the definition of the matrices M_t and N_t from Lemma B, we claim that the infinite matrix sums

$$M_{\infty} = \sum_{t=0}^{\infty} M_t \quad \text{and} \quad N_{\infty} = \sum_{t=0}^{\infty} N_t$$

exist and are finite. Specifically, an infinite sum of matrices such as $\sum_{t=0}^{\infty} M_t$ exists and is finite if and only if each corresponding sum of matrix entries, $\sum_{t=0}^{\infty}(M_t)_{ij}$, converges for all i, j. Equivalently, this condition holds if and only if the scalar series $\sum_{t=0}^{\infty} a^\top M_t b$ converges for all vectors $a, b \in \mathbb{R}^n$.

Note first that, for any $a \in \mathbb{R}^n$, the limit $\lim_{T\to\infty}\sum_{t=0}^{T} a^\top M_t a$ exists because $a^\top M_t a \geq 0$ for every $t \in \mathbb{N}$, so $\sum_{t=0}^{T} a^\top M_t a$ is non-decreasing in T. The limit is also finite because

$$\begin{aligned}
\sum_{t=0}^{T} a^\top M_t a &\leq \sum_{t=0}^{T}\left(\lambda_{\max}(R_0)\, a^\top (UD)_t^\top (UD)_t a + \lambda_{\max}(Q_0)\, a^\top (G + HUD)_t^\top (G + HUD)_t a\right) \\
&\leq \sum_{t=0}^{T}\left(\lambda_{\max}(R_0)\, \|(UD)_t\|_{\mathrm{F}}^2 \|a\|_2^2 + \lambda_{\max}(Q_0)\, \|(G + HUD)_t\|_{\mathrm{F}}^2 \|a\|_2^2\right) \\
&\leq \lambda_{\max}(R_0)\, \|UD\|_{\mathcal{T}}^2 \|a\|_2^2 + \lambda_{\max}(Q_0)\, \|G + HUD\|_{\mathcal{T}}^2 \|a\|_2^2 < \infty \quad \forall T \in \mathbb{N}.
\end{aligned}$$

Above, the first two inequalities follow from the definition of M_t and the Cauchy-Schwarz inequality, respectively. The third inequality exploits our definition of the Toeplitz norm $\|\cdot\|_{\mathcal{T}}$ and holds because $U \in \mathcal{U}_{\infty}$. Hence, $\sum_{t=0}^{\infty} a^\top M_t a$ exists and is finite. In addition, we have

$$\lim_{T\to\infty}\sum_{t=0}^{T} a^\top M_t b = \lim_{T\to\infty}\frac{1}{4}\sum_{t=0}^{T}(a+b)^\top M_t (a+b) - \lim_{T\to\infty}\frac{1}{4}\sum_{t=0}^{T}(a-b)^\top M_t (a-b)$$

for all $a, b \in \mathbb{R}^n$. Both limits on the right hand side exist and are finite thanks to the above arguments. Hence, the limit on the left hand side exists and is finite, too. Because the choice of $a, b \in \mathbb{R}^m$ was arbitrary, every element of the matrix $\sum_{t=0}^{T} M_t$ has a finite limit as $T \to \infty$, proving that M_{∞} exists and is finite. Similarly, one can show that N_{∞} exists and is finite.

Select any $\Sigma_w^\star \in \arg\max_{\Sigma_w \in \mathcal{M}_{\Sigma_w}} \mathrm{Tr}(\Sigma_w M_\infty)$ and $\Sigma_v^\star \in \arg\max_{\Sigma_v \in \mathcal{M}_{\Sigma_v}} \mathrm{Tr}(\Sigma_v N_\infty)$, which exist because $\mathcal{M}_{\Sigma_w}$ and $\mathcal{M}_{\Sigma_v}$ are compact. Next, define $\mathbb{P}^\star = \mathbb{P}^\star_{x_0} \otimes (\otimes_{t=0}^\infty (\mathbb{P}^\star_{w_t} \otimes \mathbb{P}^\star_{v_t}))$, where $\mathbb{P}^\star_{x_0} = \mathbb{P}^\star_{w_t} = \mathcal{N}(0, \Sigma_w^\star)$ and $\mathbb{P}^\star_{v_t} = \mathcal{N}(0, \Sigma_v^\star)$ for all $t \in \mathbb{N}$. By construction, $\mathbb{P}^\star$ is a time-invariant Gaussian distribution. By (A.19), $\mathcal{M}_{\Sigma_{x_0}} = \mathcal{M}_{\Sigma_{w_t}} = \mathcal{M}_{\Sigma_w}$ and $\mathcal{M}_{\Sigma_{v_t}} = \mathcal{M}_{\Sigma_v}$ for all $t \in \mathbb{N}$, so one readily verifies that $\mathbb{P}^\star$ belongs to $\mathcal{B}_{\mathcal{N}}^\infty$. Then, because $\mathcal{B}_{\mathcal{N}}^\infty \subseteq \overline{\mathcal{B}}^\infty$, we have that $\mathbb{P}^\star$ is feasible in the inner maximization problem in (A.25). In the remainder of the proof we will show that $\mathbb{P}^\star$ is also optimal. To this end, note first that for any $\mathbb{P} \in \overline{\mathcal{B}}^\infty$ and $T \in \mathbb{N}$ we have

$$\frac{1}{T}\sum_{t=0}^{T-1} \mathbb{E}_{\mathbb{P}}\left[x_t^\top Q_0 x_t + u_t^\top R_0 u_t\right] \leq \frac{1}{T}\sum_{s=0}^{T-1} \mathrm{Tr}\left(\Sigma_{w_s} M_\infty + \Sigma_{v_s} N_\infty\right) \leq \mathrm{Tr}\left(\Sigma_w^\star M_\infty + \Sigma_v^\star N_\infty\right).$$

Above, the first inequality follows from (A.24d) and the construction of M_∞ and N_∞, while the second inequality follows from the choice of $\Sigma_w^\star$ and $\Sigma_v^\star$. Because the inequality above holds for all $\mathbb{P} \in \overline{\mathcal{B}}^\infty$ and $T \in \mathbb{N}$, we may conclude that $\max_{\mathbb{P} \in \overline{\mathcal{B}}^\infty} J_{\mathbb{P}}(x, u) \leq \mathrm{Tr}\left(\Sigma_w^\star M_\infty + \Sigma_v^\star N_\infty\right)$. To show that this upper bound is attained by $\mathbb{P}^\star$, choose any $\varepsilon > 0$. Then, there exists $T_0 \in \mathbb{N}$ such that

$$\left\| M_\infty - \sum_{t=0}^{T} M_t \right\|_2 \leq \frac{\varepsilon}{2(1 + \mathrm{Tr}(\Sigma_w^\star))} \quad \text{and} \quad \left\| N_\infty - \sum_{t=0}^{T} N_t \right\|_2 \leq \frac{\varepsilon}{2(1 + \mathrm{Tr}(\Sigma_v^\star))} \quad \forall T \geq T_0. \tag{A.66}$$

Thus, we have

$$\begin{aligned}
\frac{1}{T}\sum_{t=0}^{T-1} \mathbb{E}_{\mathbb{P}^\star}\left[x_t^\top Q_0 x_t + u_t^\top R_0 u_t\right] &\geq \frac{1}{T}\sum_{s=0}^{T-1-T_0} \mathrm{Tr}\left(\Sigma_w^\star \left(\sum_{t=0}^{T-1-s} M_t\right) + \Sigma_v^\star \left(\sum_{t=0}^{T-1-s} N_t\right)\right) \\
&\geq \frac{T - T_0}{T}\left(\mathrm{Tr}\left(\Sigma_w^\star M_\infty + \Sigma_v^\star N_\infty\right) - \mathrm{Tr}(\Sigma_w^\star)\frac{\varepsilon}{2(1 + \mathrm{Tr}(\Sigma_w^\star))} - \mathrm{Tr}(\Sigma_v^\star)\frac{\varepsilon}{2(1 + \mathrm{Tr}(\Sigma_v^\star))}\right) \\
&\geq \frac{T - T_0}{T}\left(\mathrm{Tr}\left(\Sigma_w^\star M_\infty + \Sigma_v^\star N_\infty\right) - \varepsilon\right) \quad \forall T > T_0.
\end{aligned}$$

The first inequality in the above expression follows from (A.24d) applied to $\mathbb{P} = \mathbb{P}^\star$ and from the observation that all terms in the sum over s are nonnegative. The second inequality exploits (A.66). Taking the limit superior over T on both sides then shows that

$$J_{\mathbb{P}^\star}(x, u) \geq \limsup_{T \to \infty} \frac{T - T_0}{T}\left(\mathrm{Tr}\left(\Sigma_w^\star M_\infty + \Sigma_v^\star N_\infty\right) - \varepsilon\right) = \mathrm{Tr}\left(\Sigma_w^\star M_\infty + \Sigma_v^\star N_\infty\right) - \varepsilon.$$

As the choice of $\varepsilon > 0$ was arbitrary, we may conclude that $J_{\mathbb{P}^\star}(x, u) = \mathrm{Tr}(\Sigma_w^\star M_\infty + \Sigma_v^\star N_\infty)$. This shows that $\mathbb{P}^\star$ solves indeed the inner maximization problem in (A.25). □

Proof of Proposition 6. Fix any $\mathbb{P} \in \underline{\mathcal{B}}_{\mathcal{N}}^\infty$. By Proposition II.1 in Hadjiyiannis et al. (2011), the inner minimization problem in (A.28) over *purified* output feedback controls $u \in \mathcal{U}_\eta$ is equivalent to a classical infinite-horizon LQG problem with output feedback controls $u \in \mathcal{U}_y$, of the form

$$\begin{aligned}
\min_{u,x,y} \quad & J_{\mathbb{P}}(u, x) \\
\text{s.t.} \quad & u \in \mathcal{U}_y,\ x = Hu + Gw,\ y = Cx + v.
\end{aligned} \tag{A.67}$$

Under Assumption 5, the results in Appendix§F.4 are directly applicable to the latter problem. Specifically, there exist $K \in \mathbb{R}^{m \times n}$ and $L \in \mathbb{R}^{n \times p}$ such that both $A_0 + B_0 K$ and $A_0 - LC_0$ are Schur stable, and problem (A.67) is solved by $u_t = K\hat{x}_t$ for every $t \in \mathbb{N}$, where $\hat{x}_t$ is a state estimator that follows the recursion in (A.38) (from Appendix§F.4) and satisfies

$$\hat{x}_{t+1} = (I - LC_0)(A_0 + B_0 K)\hat{x}_t + Ly_{t+1} \quad \forall t \in \mathbb{N}.$$

This recursion implies that $u = U'y$, where $U' \in \mathcal{T}^{m \times p}$ is a block lower triangular Toeplitz matrix with blocks

$$U'_t = K(I - LC_0)^t (A_0 + B_0 K)^t L \quad \forall t \in \mathbb{N}.$$

Consequently, problem (A.67) is solved by a *stationary* linear output feedback policy.

An immediate generalization of Lemma E to infinite-horizon control problems ensures that the linear output feedback policy $u = U'y$ can equivalently be represented as a linear purified output feedback policy $u = U\eta$, where $U = (I - U'CH)^{-1}U'$ and I stands for the identity matrix in $\mathcal{T}^{m\times m}$. Observe now that $C \in \mathcal{T}^{p\times n}$ and $H \in \mathcal{T}^{n\times m}$. As $U' \in \mathcal{T}^{m\times p}$, assertions (i) and (ii) of Lemma A imply that $I - U'CH \in \mathcal{T}^{m\times m}$. In addition, as H has zero blocks on the diagonal, one can readily verify that all blocks on the main diagonal of $I - U'CH$ are m-dimensional identity matrices. This ensures via Lemma A (iii) that $U = (I - U'CH)^{-1}U' \in \mathcal{T}^{m\times p}$. Consequently, $u = U\eta$ constitutes a *stationary* linear purified output feedback policy. By construction, this policy solves the inner minimization problem in (A.28), which is equivalent to (A.67)

We verify that $U \in \mathcal{U}_\infty$. Recall from Lemma B that with $u = U\eta$ for $U \in \mathcal{T}^{m\times p}$, the covariance of u_t is:

$$\Sigma_{u_t} = \mathbb{E}_{\mathbb{P}}\left[u_t u_t^\top\right] = \sum_{s=0}^{t}\left((UD)_{t-s}\Sigma_{w_s}(UD)_{t-s}^\top + (U)_{t-s}\Sigma_{v_s}(U)_{t-s}^\top\right)$$

where $(UD)_s$ is shorthand for the block matrix on the s-the subdiagonal of the Toeplitz matrix UD. Moreover, by Lemma F in Appendix§F.4, the state estimator $\hat{x}_t$ must admit a stationary covariance matrix in the limit $t \to \infty$, which implies that the feedback control $u_t = K\hat{x}_t$ also admits a stationary covariance matrix, and thus the limit

$$\lim_{t\to\infty}\Sigma_{u_t} = \sum_{t=0}^{\infty}(UD)_t\Sigma_w(UD)_t^\top + \sum_{t=0}^{\infty}U_t\Sigma_v U_t^\top$$

exists and is finite. Because $\Sigma_w, \Sigma_v \succ 0$ by our standing assumption that $\mathbb{P} \in \underline{\mathcal{B}}_{\mathcal{N}}^{\infty}$, this implies that

$$\sum_{t=0}^{\infty}\mathrm{Tr}\left((UD)_t(UD)_t^\top\right) = \|UD\|_{\mathcal{T}}^2 < \infty \quad \text{and} \quad \sum_{t=0}^{\infty}\mathrm{Tr}\left(U_tU_t^\top\right) = \|U\|_{\mathcal{T}}^2 < \infty.$$

Next, recall that $x = Hu + Gw = (HUD + G)w + HUv$. From Lemma F, we know that x_t admits a stationary covariance matrix, and thus the limit

$$\lim_{t\to\infty}\Sigma_{x_t} = \sum_{t=0}^{\infty}(HUD+G)_t\Sigma_w(HUD+G)_t^\top + \sum_{t=0}^{\infty}(HU)_t\Sigma_v(HU)_t^\top$$

exists and is finite. Because $\Sigma_w, \Sigma_v \succ 0$ (because $\mathbb{P} \in \underline{\mathcal{B}}_{\mathcal{N}}^{\infty}$), this implies that $\|HUD + G\|_{\mathcal{T}}^2 < \infty$ and $\|HU\|_{\mathcal{T}}^2 < \infty$. In summary, we have shown that $U \in \mathcal{U}_\infty$, and thus the claim follows. □

Proof of Lemma C. Recall that $R_0 \succ 0$ and $Q_0 \succ 0$ by Assumption 5-(iii). This ensures that $\mathrm{Tr}(R_0)$ and $\mathrm{Tr}(Q_0)$ are both strictly positive. By Lemma F, which applies because $\mathbb{P} \in \underline{\mathcal{B}}_{\mathcal{N}}^{\infty}$, the limits $\lim_{t\to\infty}\Sigma_{x_t} = \Sigma_{x_\infty}$ and $\lim_{t\to\infty}\Sigma_{u_t} = \Sigma_{u_\infty}$ exist and are finite. We will show that

$$J_{\mathbb{P}}(x,u) = \mathrm{Tr}(\Sigma_{x_\infty}Q_0) + \mathrm{Tr}(\Sigma_{u_\infty}R_0).$$

As Σ_{u_t} converges to Σ_{u_∞} and Σ_{x_t} converges to Σ_{x_∞}, for any $\varepsilon > 0$ there exists $t_0 \in \mathbb{N}$ such that

$$\|\Sigma_{x_T} - \Sigma_{x_\infty}\|_2 \le \frac{\varepsilon}{3\,\mathrm{Tr}(Q_0)} \quad \text{and} \quad \|\Sigma_{u_T} - \Sigma_{u_\infty}\|_2 \le \frac{\varepsilon}{3\,\mathrm{Tr}(R_0)} \quad \forall t \ge t_0. \tag{A.68}$$

In addition, one easily verifies that there exists $T_0 > t_0$ such that

$$\left|\frac{1}{T}\sum_{t=1}^{t_0}\left[\mathrm{Tr}\big((\Sigma_{x_t} - \Sigma_{x_\infty})Q_0\big) + \mathrm{Tr}\big((\Sigma_{u_t} - \Sigma_{u_\infty})R_0\big)\right]\right| \le \frac{\varepsilon}{3} \quad \forall T \ge T_0. \tag{A.69}$$

Then, we have

$$\left|\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}[x_t^\top Q_0 x_t + u_t^\top R_0 u_t] - \mathrm{Tr}(\Sigma_{x_\infty}Q_0) - \mathrm{Tr}(\Sigma_{u_\infty}R_0)\right|$$

$$
\begin{aligned}
&= \left| \frac{1}{T} \sum_{t=1}^{T} \Big[\mathrm{Tr}\big((\Sigma_{x_t} - \Sigma_{x_\infty}) Q_0\big) + \mathrm{Tr}\big((\Sigma_{u_t} - \Sigma_{u_\infty}) R_0\big) \Big] \right| \\
&\leq \left| \frac{1}{T} \sum_{t=1}^{t_0} \Big[\mathrm{Tr}\big((\Sigma_{x_t} - \Sigma_{x_\infty}) Q_0\big) + \mathrm{Tr}\big((\Sigma_{u_t} - \Sigma_{u_\infty}) R_0\big) \Big] \right| \\
&\quad + \left| \frac{1}{T} \sum_{t=t_0+1}^{T} \mathrm{Tr}((\Sigma_{x_t} - \Sigma_{x_\infty}) Q_0) \right| + \left| \frac{1}{T} \sum_{t=t_0+1}^{T} \mathrm{Tr}((\Sigma_{u_t} - \Sigma_{u_\infty}) R_0) \right| \\
&\leq \frac{\varepsilon}{3} + \frac{\varepsilon(T - t_0)}{3T} + \frac{\varepsilon(T - t_0)}{3T} \leq \varepsilon
\end{aligned}
$$

for all $T \geq T_0$. Here, the first inequality exploits the triangle inequality, while the second inequality follows from (A.68) and (A.69) and from the matrix Hölder inequality $\mathrm{Tr}(\Sigma Q) \leq \|\Sigma\|_2 \, \mathrm{Tr}(Q)$, which holds for any symmetric matrices Σ and Q with $Q \succeq 0$. As ε was chosen arbitrarily, the above estimate proves that the limit in (A.30) exists and is indeed equal to $\mathrm{Tr}(\Sigma_{x_\infty} Q_0) + \mathrm{Tr}(\Sigma_{u_\infty} R_0)$. □

Proof of Proposition 7. Consider any $U \in \mathcal{U}_\infty$ and any $\mathbb{P} \in \underline{\mathcal{B}}_{\mathcal{N}}^{\infty}$ with covariance matrices $\Sigma_w \in \mathcal{M}_{\Sigma_w}^{+}$ and $\Sigma_v \in \mathcal{M}_{\Sigma_v}^{+}$. For any $T \in \mathbb{N}_+$, define $J_T(U; \Sigma_w, \Sigma_v)$ as shorthand for the average expected cost under the control policy $u = U(Dw + v)$ and associated states $x = Hu + Gw$ over the first T periods, with an expression given by (A.24d). Hence, $J_T(U; \Sigma_w, \Sigma_v)$ is convex quadratic in U and linear in (Σ_w, Σ_v) for every $T \in \mathbb{N}_+$. By the definition of the long-run average expected cost $J(U; \Sigma_w, \Sigma_v)$, we have

$$
J(U; \Sigma_w, \Sigma_v) = \limsup_{T \to \infty} J_T(U; \Sigma_w, \Sigma_v) = \lim_{T' \to \infty} \sup_{T \geq T'} J_T(U; \Sigma_w, \Sigma_v).
$$

Because convexity is preserved under pointwise suprema and limits, this shows that $J(U; \Sigma_w, \Sigma_v)$ is convex in $U \in \mathcal{U}_\infty$ for any $(\Sigma_w, \Sigma_v) \in \mathcal{M}_{\Sigma_w}^{+} \times \mathcal{M}_{\Sigma_v}^{+}$. Moreover, Lemma C implies that

$$
J(U; \Sigma_w, \Sigma_v) = \lim_{T \to \infty} J_T(U; \Sigma_w, \Sigma_v). \tag{A.70}
$$

Hence, the claim follows. □

K Extension to Entropy-Regularized Optimal Transport

In this section, we discuss a different example of ambiguity set that is compatible with our framework. Consider taking the divergence $\mathbb{D}$ as the entropy-regularized optimal transport discrepancy with regularization parameter $\epsilon \geq 0$, denoted with $\mathbb{W}_\epsilon$ and defined as follows.

Definition 10 (Entropy-regularized Optimal Transport Discrepancy)**.** *The entropy-regularized optimal transport discrepancy between two distributions $\mathbb{P}_z$ and $\hat{\mathbb{P}}_z$ on $\mathbb{R}^{d_z}$ with regularization parameter $\epsilon \geq 0$ is*

$$
\mathbb{W}_\epsilon(\mathbb{P}_z, \hat{\mathbb{P}}_z) = \min_{\pi \in \Pi(\mathbb{P}_z, \hat{\mathbb{P}}_z)} \int_{\mathbb{R}^{d_z} \times \mathbb{R}^{d_z}} \|z - \hat{z}\|^2 \, \mathrm{d}\pi(z, \hat{z}) + \epsilon \mathbb{H}(\pi),
$$

where $\Pi(\mathbb{P}_z, \hat{\mathbb{P}}_z)$ is the set of couplings of $\mathbb{P}_z$ and $\hat{\mathbb{P}}_z$ as in Definition 1, and the entropy of a coupling π is defined as:

$$
\mathbb{H}(\pi) = \int_{\mathbb{R}^{d_z} \times \mathbb{R}^{d_z}} \frac{\mathrm{d}\pi}{\mathrm{d}\mathbb{L}}(z, \hat{z}) \log\Big(\frac{\mathrm{d}\pi}{\mathrm{d}\mathbb{L}}(z, \hat{z}) \Big) \, \mathrm{d}\mathbb{L}(z, \hat{z}),
$$

where $\frac{\mathrm{d}\pi}{\mathrm{d}\mathbb{L}}$ denotes the Radon-Nikodym derivative of π with respect to the Lebesgue measure $\mathbb{L}$ on $\mathbb{R}^{d_z} \times \mathbb{R}^{d_z}$, defined for any π that is absolutely continuous with respect to $\mathbb{L}$. If π is not absolutely continuous with respect to $\mathbb{L}$, then we set $\mathbb{H}(\pi) = +\infty$.

Ambiguity sets based on discrepancies similar to $\mathbb{W}_\epsilon$ have appeared before in the DRO literature (see, e.g., Wang et al., 2021; Azizian et al., 2023). Although these sets are compatible with our framework, a few notable differences arise. First, the divergence $\mathbb{W}_\epsilon$ no longer satisfies the identity of indiscernibles principle, i.e., $\mathbb{W}(\mathbb{P},\mathbb{P})$ may not be identically zero for all $\mathbb{P} \in \mathcal{P}(\mathbb{R}^d)$. Moreover, the ambiguity set could be empty when the ambiguity radius ρ_z is small (this arises due to the regularization term $\epsilon\mathbb{H}(\pi)$, for a sufficiently large ϵ). As a result, small changes are required in our assumptions to ensure the problem is non-trivial. We point these out subsequently, as we verify each assumption.

K.1 Verifying Assumption 2

Mirroring our earlier developments, for two non-degenerate distributions with finite second moments, we define a distance function between two mean and covariance matrix pairs as follows.

Definition 11. *The entropy-regularized Bures-Wasserstein distance between* $(\mu_z, \Sigma_z) \in \mathbb{R}^{d_z} \times \mathbb{S}^{d_z}_{++}$ *and* $(\hat{\mu}_z, \hat{\Sigma}_z) \in \mathbb{R}^{d_z} \times \mathbb{S}^{d_z}_{++}$ *with regularization parameter* $\epsilon \in \mathbb{R}_+$ *is given by*

$$\mathbb{G}_\epsilon((\mu_z,\Sigma_z),(\hat{\mu}_z,\hat{\Sigma}_z)) = \sqrt{\|\mu_z-\hat{\mu}_z\|^2 + \operatorname{Tr}(\Sigma_z) + \operatorname{Tr}(\Sigma_2) - 2\operatorname{Tr}(X_\epsilon(\Sigma_z)) - \frac{\epsilon}{2}\log\left((2\pi e)^{2d}\left(\frac{\epsilon}{2}\right)^d |X_\epsilon(\Sigma_z)|\right)},$$

where

$$X_\epsilon(\Sigma_z) = \left(\hat{\Sigma}_z^{1/2}\Sigma_z\hat{\Sigma}_z^{1/2} + \left(\frac{\epsilon}{4}\right)^2 I\right)^{1/2} - \frac{\epsilon}{4}I. \tag{A.71}$$

The following proposition shows that the entropy-regularized OT between any non-degenerate distribution and a non-degenerate Gaussian distribution is bounded below by the entropy-regularized OT between their respective mean vectors and covariance matrices. Moreover, if both distributions are Gaussian, then this bound is tight.

Proposition 13 (del Barrio and Loubes (2020))**.** *For any distribution* $\mathbb{P}_z$ *with mean and covariance matrix* $(\mu_z, \Sigma_z) \in \mathbb{R}^{d_z} \times \mathbb{S}^{d_z}_{++}$ *and any Gaussian distribution* $\hat{\mathbb{P}}_z$ *on* $\mathbb{R}^{d_z}$ *with mean and covariance matrix* $(\hat{\mu}_z, \hat{\Sigma}_z) \in \mathbb{R}^{d_z} \times \mathbb{S}^{d_z}_{++}$*, we have*

i) $\mathbb{W}_\epsilon(\mathbb{P}_z, \hat{\mathbb{P}}_z) \geq \mathbb{G}_\epsilon((\mu_z,\Sigma_z),(\hat{\mu}_z,\hat{\Sigma}_z))$

ii) $\mathbb{W}_\epsilon(\mathbb{P}_z, \hat{\mathbb{P}}_z) = \mathbb{G}_\epsilon((\mu_z,\Sigma_z),(\hat{\mu}_z,\hat{\Sigma}_z))$ *if* $\mathbb{P}_z$ *is Gaussian.*

The results in (i) and (ii) of Proposition 13 follow from Theorem 2.3 and Theorem 2.2 in del Barrio and Loubes (2020), respectively. Notably, for the nominal Gaussian distribution $\hat{\mathbb{P}}_z = \mathcal{N}(\hat{\mu}_z, \hat{M}_z)$, the minimum value of the square of $\mathbb{G}_\epsilon(\mathbb{P}_z, \mathcal{N}(\hat{\mu}_z, \hat{\Sigma}_z))$ over $\mathbb{P}_z \in \mathcal{P}(\mathbb{R}^{d_z})$ is attained by $\mathcal{N}(\hat{\mu}_z, \hat{\Sigma}_z + \epsilon/2I)$ (del Barrio and Loubes, 2020, Theorem 2.4). Importantly, this value is nonzero (and is not attained by the nominal distribution $\hat{\mathbb{P}}_z$), so to ensure that the ambiguity set is nonempty, we require

$$\rho_z \geq \underline{\rho}_z := \mathbb{G}_\epsilon\big((\hat{\mu}_z, \hat{\Sigma}_z + \epsilon/2I), (\hat{\mu}_z, \hat{\Sigma}_z)\big), \text{ for all } z \in \mathcal{Z}. \tag{A.72}$$

Under this premise, we have the following result.

Proposition 14. *If* $\rho_z \geq \underline{\rho}_z$*, the divergence* $\mathbb{W}_\epsilon$ *satisfies Assumption 2.*

Proof of Proposition 14. If $\rho_z \geq \underline{\rho}_z$, Proposition 13 implies that Assumption 2-(i) holds. To see that Assumption 2-(ii) also holds, note that:

$$\begin{aligned}\mathcal{M}^{\mathbb{G}_\epsilon}_{(\mu_z,M_z)} &= \{(\mu_z, M_z) \in \mathcal{M}_2^{d_z} : M_z \succ \mu_z\mu_z^\top,\ \mathbb{W}_\epsilon(\mathcal{N}(\mu_z, M_z), \hat{\mathbb{P}}_z) \leq \rho_z\} \\ &= \{(\mu_z, M_z) \in \mathcal{M}_2^{d_z} : M_z \succ \mu_z\mu_z^\top,\ \big(\mathbb{G}_\epsilon((\mu_z, M_z - \mu_z\mu_z^\top), (\hat{\mu}_z, \hat{\Sigma}_z)\big)^2 \leq \rho_z^2\},\end{aligned}$$

where the second equality follows from Proposition 13-(ii). Note that $\mathcal{M}^{\mathbb{G}_\epsilon}_{(\mu_z, M_z)}$ contains only non-degenerate distributions because otherwise, $\mathbb{G}_\epsilon(\mathcal{N}(\mu_z, \Sigma_z), \hat{\mathbb{P}}_z)$ would not be finite.

We next establish that the set $\mathcal{M}^{\mathbb{G}_\epsilon}_{(\mu_z, M_z)}$ is convex and compact. As a first step, we prove that the map $\Sigma_z \mapsto X_\epsilon(\Sigma_z)$ is matrix concave and operator monotone on the cone of positive definite matrices, $\mathbb{S}^{d_z}_{++}$. By Theorem 1.5.9 and Theorem 4.2.3 in Bhatia (2009), the map $X \mapsto X^{\frac{1}{2}}$ is operator monotone and matrix concave on $\mathbb{S}^{d_z}_{+}$. The map $\Sigma_z \mapsto \hat{\Sigma}_z^{1/2}\Sigma_z\hat{\Sigma}_z^{1/2} + \left(\frac{\epsilon}{4}\right)^2 I$ is linear and operator monotone for all $\hat{\Sigma}_z \in \mathbb{S}^{d_z}_{++}$. Because matrix convexity/concavity is preserved under linear maps[6] and the composition of operator monotone functions is operator monotone, it follows that $X_\epsilon(\Sigma_z)$ is matrix concave and operator monotone.

Next, we prove that the function $\mathbb{G}_\epsilon((\mu_z, M_z - \mu_z\mu_z^\top), (\hat{\mu}_z, \hat{\Sigma}_z))^2$ is jointly convex in $(\mu_z, M_z) \in \mathcal{M}_2^{d_z}$ for any fixed $(\hat{\mu}_z, \hat{\Sigma}_z) \in \mathcal{M}_2^{d_z}$. As a function of (μ_z, M_z) (ignoring constant terms), this can be rewritten as $\|\mu_z - \hat{\mu}_z\|^2 + \mathrm{Tr}(M_z) - \|\mu_z\|^2 - 2\,\mathrm{Tr}(X_\epsilon(M_z - \mu_z\mu_z^\top)) - \epsilon/2\log(|X_\epsilon(M_z - \mu_z\mu_z^\top)|)$. Expanding the first term and canceling the third term, it can be seen that the first three terms yield a jointly convex function of (μ_z, M_z). So it suffices to prove that $f(X_\epsilon(M_z - \mu_z\mu_z^\top))$ is jointly concave in (μ_z, M_z) on $\mathcal{M}_2^{d_z}$, where $f(X) = 2\,\mathrm{Tr}(X) + \epsilon/2\log(|X|)$.

To analyze this, we first prove that $f(X_\epsilon(\Sigma_z))$ is operator monotone and concave in Σ_z on $\mathbb{S}^{d_z}_{++}$. Note that the function $f(X)$ is operator monotone on $\mathbb{S}^{d_z}_{+}$ because $\mathrm{Tr}(\cdot)$ and $\log\det(\cdot)$ are both operator monotone. Because $X_\epsilon(\Sigma_z)$ is also operator monotone (as argued above) and the composition of operator monotone functions remains operator monotone, we conclude that $f(X_\epsilon(\Sigma_z))$ is operator monotone on $\mathbb{S}^{d_z}_{++}$. To prove concavity, consider any $\Sigma_z, \Sigma_z' \in \mathbb{S}^{d_z}_{+}$ and any $\lambda \in [0, 1]$. We have:

$$\begin{aligned} &f\left(X_\epsilon(\lambda\Sigma_z + (1-\lambda)\Sigma_z')\right) \\ &\qquad \geq f\left(\lambda X_\epsilon(\Sigma_z) + (1-\lambda)X_\epsilon(\Sigma_z')\right) \geq \lambda f\left(X_\epsilon(\Sigma_z)\right) + (1-\lambda) f\left(X_\epsilon(\Sigma_z')\right), \end{aligned}$$

where the first inequality follows because $f(X)$ is operator monotone on $\mathbb{S}^{d_z}_{+}$ and $X_\epsilon(\Sigma_z)$ is matrix concave on $\mathbb{S}^{d_z}_{+}$, and the second inequality follows because $f(X)$ is concave on $\mathbb{S}^{d_z}_{+}$. This shows that $f(X_\epsilon(\Sigma_z))$ is concave in Σ_z.

To prove that $f(X_\epsilon(M_z - \mu_z\mu_z^\top))$ is jointly concave in (μ_z, M_z) on $\mathcal{M}_2^{d_z}$, consider any $(\mu_z, M_z), (\mu_z', M_z') \in \mathcal{M}_2^{d_z}$ and $\lambda \in [0, 1]$. With $\mu_\lambda = \lambda\mu_z + (1-\lambda)\mu_z'$, we can verify the identity:

$$\mu_\lambda\mu_\lambda^\top = \lambda\mu_z\mu_z^\top + (1-\lambda)\mu_z'(\mu_z')^\top - \lambda(1-\lambda)(\mu_z - \mu_z')(\mu_z - \mu_z')^\top. \tag{A.73}$$

This implies that:

$$\begin{aligned} f\left(X_\epsilon(\lambda M_z + (1-\lambda)M_z' - \mu_\lambda\mu_\lambda^\top)\right) &\geq f\left(X_\epsilon(\lambda M_z + (1-\lambda)M_z' - \lambda\mu_z\mu_z^\top - (1-\lambda)\mu_z'(\mu_z')^\top)\right) \\ &\geq \lambda f\left(X_\epsilon(M_z - \mu_z\mu_z^\top)\right) + (1-\lambda) f\left(X_\epsilon(M_z' - \mu_z'(\mu_z')^\top)\right), \end{aligned}$$

where the first inequality follows from (A.73) and because $f(X_\epsilon(\Sigma_z))$ is operator monotone in Σ_z. The second inequality follows because $f(X_\epsilon(\Sigma_z))$ is concave in Σ_z.

These arguments imply that $\mathbb{G}_\epsilon((\mu_z, M_z - \mu_z\mu_z^\top), (\hat{\mu}_z, \hat{\Sigma}_z))^2$ is jointly convex in (μ_z, M_Z) of $\mathcal{M}_2^{d_z}$.

To conclude the proof, we argue that $\mathcal{M}^{\mathbb{G}_\epsilon}_{\Sigma_z}$ is convex and compact. $\mathcal{M}^{\mathbb{G}_\epsilon}_{\Sigma_z}$ is convex because it corresponds to the $\rho_z^2 > 0$ sublevel set of the convex function $\mathbb{G}_\epsilon^2((\mu_z, M_z - \mu_z\mu_z^\top), (\hat{\mu}_z, \hat{\Sigma}_z))$. Moreover, $\mathbb{G}_\epsilon^2((\mu_z, M_z - \mu_z\mu_z^\top), (\hat{\mu}_z, \hat{\Sigma}_z))$ is a continuous and coercive function in (μ_z, M_z), and thus its sublevel sets are closed and bounded, and thus are compact (see, e.g., Lemma 1.24 and Proposition 11.12 in Bauschke and Combettes, 2017, respectively). □

[6] For an elementary proof, assume $f : \mathbb{S}^{d_z}_{+} \to \mathbb{S}^{d_z}_{+}$ is matrix convex and consider an affine map $X(Y)$. Then, $f\left(X(\lambda Y_1 + (1-\lambda)Y_2)\right) = f\left(\lambda X(Y_1) + (1-\lambda)X(Y_2)\right) \preceq \lambda f(X(Y_1)) + (1-\lambda)f(X(Y_2))$, where the first equality follows because $X(Y)$ is affine and the inequality follows because f is matrix convex.

K.2 Verifying Assumption 3

Proposition 15. *Fix $\hat{\mu}_z = 0$. If $\rho_z \geq \underline{\rho}_z$, the divergence $\mathbb{W}_\epsilon$ satisfies Assumption 3.*

Proof. We simplify notation by omitting the subscript z. Recall that $\hat{\mu} = 0$, which implies that $\hat{M} = \hat{\Sigma}$, and the nominal distribution is $\hat{\mathbb{P}} = \mathcal{N}(0, \hat{\Sigma})$. As in the proof of Proposition 3, consider any two Gaussian distributions $\mathbb{P}, \mathbb{P}' \in \mathcal{B}$ such that $\mathbb{E}_{\mathbb{P}}[zz^\top] = \mathbb{E}_{\mathbb{P}'}[zz^\top] = M$ and $\mu' = \hat{\mu} = 0$, which implies that $\Sigma' = \mathbb{E}_{\mathbb{P}'}[(z-\mu')(z-\mu')^\top] = M$. For the subsequent arguments, it helps to note that

$$\|\mu - \mathbb{E}_{\hat{\mathbb{P}}}[z]\|^2 + \mathrm{Tr}(\Sigma) - \|\mu' - \mathbb{E}_{\hat{\mathbb{P}}}[z]\|^2 - \mathrm{Tr}(\Sigma') = \mathrm{Tr}(M) - \mathrm{Tr}(M) = 0. \tag{A.74}$$

Also, we recall from the proof of Proposition 14 that the maps $\Sigma_z \mapsto X_\epsilon(\Sigma_z)$ and $X \mapsto X^{\frac{1}{2}}$ are matrix-concave and operator monotone on the cone of positive semidefinite matrices $\mathbb{S}^{d_z}_{++}$, that is, are functions $f : \mathbb{S}^{d_z}_{+} \to \mathbb{S}^{d_z}_{+}$ that satisfy $f(X) \succeq f(Y)$ for any $X, Y \in \mathbb{S}^{d_z}_{+}$ with $X \succeq Y$.

Recall from Proposition 13 that the set of moments $\mathcal{M}_{(\mu_z, M_z)}$ can be written in this case as:

$$\mathcal{M}_{(\mu_z, M_z)} = \{(\mu_z, M_z) \in \mathcal{M}_2^{d_z} : \Big(\mathbb{G}_\epsilon\big((\mu_z, M_z), (\hat{\mu}, \hat{M})\big)\Big)^2 \leq \rho_z^2\}. \tag{A.75}$$

To prove that $(\mu, M) \in \mathcal{M}_{(\mu, M)}$ implies that $(0, M) \in \mathcal{M}_{(\mu, M)}$, it suffices to show that $\mathbb{G}_\epsilon^2((0, M), (0, \hat{M})) \leq \mathbb{G}_\epsilon^2((\mu_z, M), (0, \hat{M}))$. By Proposition 13 and (A.74), we have:

$$\begin{aligned}
&\mathbb{G}_\epsilon^2\big((0, M), (0, \hat{M})\big) - \mathbb{G}_\epsilon^2\big((\mu, M), (0, \hat{M})\big) \\
&\quad = -2(\mathrm{Tr}\left(X_\epsilon(M)\right) - \mathrm{Tr}(X_\epsilon(M - \mu\mu^\top))) - \frac{\epsilon}{2}\left(\log(|X_\epsilon(M)|) - \log\left(|X_\epsilon(M - \mu\mu^\top)|\right)\right) \leq 0,
\end{aligned}$$

where the inequality follows because $M \succeq M - \mu\mu^\top$, the maps X_ϵ and $X^{1/2}$ are operator monotone on $\mathbb{S}^{d_z}_{++}$, and $\mathrm{Tr}(X) \geq \mathrm{Tr}(Y)$ and $\log\det(X) \geq \log\det(Y)$ if $X \succeq Y$. □

K.3 Theorem C for the $\mathbb{W}_\epsilon$ Divergence

For the $\mathbb{W}_\epsilon$ divergence, Assumption 4 as stated in the main text is not satisfied. Instead, we require the following slightly modified assumption.

Assumption 7. *Fix $\mu_z = \hat{\mu}_z = 0$ for all $z \in \mathcal{Z}$. There exists $g : \mathbb{S}^{d_z}_{+} \to \mathbb{R}$ such that the sets defined in Assumption 2 can be represented as $\mathcal{M}_{\Sigma_z} = \{\Sigma_z \in \mathbb{S}^{d_z}_{+} : g(\Sigma_z) \leq 0\}$ for any $z \in \mathcal{Z}$, where g is convex on $\mathbb{S}^{d_z}_{+}$, differentiable on $\mathbb{S}^{d_z}_{++}$, and satisfies: (i) $g(\bar{\Sigma}_z) < 0$, (ii) $\bar{\Sigma}_z \in \mathrm{argmin}_{\Sigma_z \succeq 0}\, g(\Sigma_z)$, and (iii) $\nabla g(\Sigma_1) \succeq \nabla g(\Sigma_2)$ implies $\Sigma_1 \succeq \Sigma_2$, for any $\Sigma_1, \Sigma_2 \in \mathbb{S}^{d_z}_{+}$, where $\bar{\Sigma}_z = \hat{\Sigma}_z + \epsilon/2I$.*

The key difference compared to Assumption 4 is that conditions (i) and (ii) are required for the matrix $\bar{\Sigma}_z$ rather than $\hat{\Sigma}_z$. We claim that Theorem C holds for the $\mathbb{W}_\epsilon$-based ambiguity sets if Assumptions 1-3 and Assumption 7 hold: one can follow the same steps in the proof of Theorem C to argue that $\Sigma_z^\star \succeq \bar{\Sigma}_z = \hat{\Sigma}_z + \epsilon/2I$, which implies that $\Sigma_z^\star \succeq \hat{\Sigma}_z$, as desired.

What remains is to show that Assumption 7 is satisfied. This is always the case if the ambiguity set is not a singleton, as summarized in the next result.

Proposition 16. *Fix $\hat{\mu}_z = 0$. If $\rho_z > \underline{\rho}_z$, the divergence $\mathbb{W}_\epsilon$ divergence satisfies Assumption 7.*

Proof. We show that $g(\Sigma_z) = (\mathbb{G}_\epsilon((0, \Sigma_z), (0, \hat{\Sigma}_z)))^2 - \rho_z^2$ satisfies Assumption 7 for $\rho_z > \underline{\rho}_z$.

Proposition 14 (and its proof) show that g is a convex function. That g is differentiable in Σ_z on $\mathbb{S}^{d_z}_{++}$ and achieves its minimum value at $\bar{\Sigma}_z = \hat{\Sigma}_z + \epsilon/2I$ follows from the same properties that hold for $\mathbb{G}_\epsilon$ (see Lemma 1.24 and Proposition 11.12 in Bauschke and Combettes, 2017.) As $\underline{\rho}_z$ was chosen so that $g(\bar{\Sigma}_z) = 0$ for $\rho_z = \underline{\rho}_z$, the continuity of g implies that $g(\bar{\Sigma}_z) < 0$ for $\rho_z > \underline{\rho}_z$.

These arguments show that requirements (i)-(ii) of Assumption 7 are satisfied. For (iii), write the gradient of g as:

$$\nabla g(\Sigma_z) = I - \hat{\Sigma}_z^{\frac{1}{2}} \left(\frac{\epsilon}{4} I + \left(\frac{\epsilon^2}{16} I + \hat{\Sigma}_z^{\frac{1}{2}} \Sigma_z \hat{\Sigma}_z^{\frac{1}{2}} \right)^{\frac{1}{2}} \right)^{-1} \hat{\Sigma}_z^{\frac{1}{2}}.$$

Then, consider $\Sigma_1, \Sigma_2 \in \mathbb{S}_+^{d_z}$ and note that $\nabla g(\Sigma_1) \succeq \nabla g(\Sigma_2)$ implies that $\Sigma_1 \succeq \Sigma_2$ because the mapping $X \mapsto X^{1/2}$ preserves ordering and the mapping $X \mapsto X^{-1}$ reverses ordering on $\mathbb{S}_{++}^{d_z}$. $\square$

L Extension to Fisher Divergence

Our final example involves another information-theoretic divergence inspired by the Fisher information matrix. To formalize it, let $\tilde{\mathcal{P}}(\mathbb{R}^d)$ denote the set of probability distributions $\mathbb{P}$ on $\mathbb{R}^d$ that admit densities $p(z)$ that are continuously differentiable, everywhere positive and satisfy the condition:

$$\exists \epsilon > 0,\ \bar{p} > 0 \ \ : \ \ \|z\|^{d_z+\epsilon} p(z) \leq \bar{p},\ \forall\, z \in \mathbb{R}^{d_z}. \tag{A.76}$$

All distributions with sub-exponential/sub-Gaussian tails or tails with a sufficiently fast polynomial decay satisfy the requirement; this includes many common distributions such as Gaussian, Laplace, or (with minor parameter restrictions) Student-t or Pareto. The assumption rules out heavy-tailed distributions whose density decays slower than $\|z\|^{-d_z}$.

For the rest of the section, we will be interested in a divergence $\mathbb{D}$ corresponding to the *relative Fisher divergence* (or *score-matching distance*), which we formalize next.

Definition 12 (Fisher divergence)**.** *Consider $\mathbb{P}_z, \hat{\mathbb{P}}_z \in \mathcal{P}(\mathbb{R}^{d_z})$. If $\mathbb{P}_z, \hat{\mathbb{P}}_z \in \tilde{\mathcal{P}}(\mathbb{R}^{d_z})$, we define the* Fisher divergence *(also called the* relative Fisher information *or* score-matching distance*) between $\mathbb{P}_z$ and $\hat{\mathbb{P}}_z$ as*

$$\mathbb{F}(\mathbb{P}_z, \hat{\mathbb{P}}_z) = \frac{1}{2} \int_{\mathbb{R}^{d_z}} \|\nabla \log p(z) - \nabla \log \hat{p}(z)\|_2^2 \, p(z) \, \mathrm{d}z,$$

where p and $\hat{p}$ denote the densities of $\mathbb{P}_z$ and $\hat{\mathbb{P}}_z$, respectively, and ∇ denotes the gradient with respect to z. If $\mathbb{P}_z \notin \tilde{\mathcal{P}}(\mathbb{R}^{d_z})$ or $\hat{\mathbb{P}}_z \notin \tilde{\mathcal{P}}(\mathbb{R}^{d_z})$, we set $\mathbb{F}(\mathbb{P}_z, \hat{\mathbb{P}}_z) = \infty$ if $\mathbb{P}_z \neq \hat{\mathbb{P}}_z$ and $\mathbb{F}(\mathbb{P}_z, \hat{\mathbb{P}}_z) = 0$ if $\mathbb{P}_z = \hat{\mathbb{P}}_z$.

Originally considered in information theory and applied probability (Johnson, 2004), this divergence has recently been used to derive novel concentration inequalities (Rioul, 2011) and in many applications in machine learning and computer vision (Hyvärinen, 2005; Song and Ermon, 2019). However, to the best of our knowledge, it has never been considered in distributionally robust optimization or control models before.

If can be readily seen that $\mathbb{F}(\mathbb{P}_z, \hat{\mathbb{P}}_z)$ is non-negative and equals 0 iff $\mathbb{P}_z = \hat{\mathbb{P}}_z$. Our definition extends the Fisher divergence to all distributions from $\mathcal{P}(\mathbb{R}^{d_z})$, although the distributions of interest are from $\tilde{\mathcal{P}}(\mathbb{R}^{d_z})$. (Note that this also requires the nominal distribution $\hat{\mathbb{P}}_z$ to belong to $\tilde{\mathcal{P}}(\mathbb{R}^{d_z})$, which means that setting $\mathbb{P}_z$ as an empirical distribution from a finite set of samples is not possible.) The quantity $s(z) = \nabla \log p(z)$ is commonly referred to as the *score* of the distribution $\mathbb{P}$, which explains the alternative naming for the divergence. We are interested in a Gaussian nominal distribution $\hat{\mathbb{P}}_z$, and we subsequently use $\hat{s}(z) = \nabla \log \hat{p}(z) = -\hat{\Sigma}_z^{-1}(z - \hat{\mu}_z)$ to denote its score.

L.1 Verifying Assumption 2

Proposition 17. *If $\rho_z \geq 0$, the Fisher divergence $\mathbb{F}$ satisfies Assumption 2.*

Proof. To simplify notation, we subsequently drop the subscript z. We verify parts (i) and (ii) separately.

Part (i). Define the ambiguity set $\mathcal{C}(\mu, M) = \{\mathbb{P} \in \tilde{\mathcal{P}}(\mathbb{R}^d) : \mathbb{E}_{\mathbb{P}}[z] = \mu,\ \mathbb{E}_{\mathbb{P}}[zz^\top] = M\}$. It suffices to prove that:

$$\mathbb{F}(\mathbb{P}, \hat{\mathbb{P}}) \geq \mathbb{F}\big(\mathcal{N}(\mu, M), \hat{\mathbb{P}}\big)\ \forall\, \mathbb{P} \in \mathcal{C}(\mu, M),$$

and equality holds if only if $\mathbb{P}_z = \mathcal{N}(\mu_z, M_z)$. To prove this inequality, consider any distribution $\mathbb{P} \in \mathcal{C}(\mu, M)$ and expand the expression for $\mathbb{F}(\mathbb{P}, \hat{\mathbb{P}})$:

$$\mathbb{F}(\mathbb{P}, \hat{\mathbb{P}}) = \frac{1}{2}\,\mathbb{E}_{\mathbb{P}}\left[\|\nabla \log p(z)\|_2^2\right] + \frac{1}{2}\,\mathbb{E}_{\mathbb{P}}\left[\|\hat{s}(z)\|_2^2\right] - \mathbb{E}_{\mathbb{P}}\left[(\nabla \log p(z))^\top \hat{s}(z)\right].$$

We evaluate each term separately. The first term is exactly equal to $\frac{1}{2}\Phi(\mathbb{P})$, where $\Phi(\mathbb{P}) = \mathbb{E}_{\mathbb{P}}\left[\|\nabla \log p(z)\|_2^2\right]$ is the trace of the Fisher information matrix of $\mathbb{P}$.

Because $\hat{s}(z)$ is linear in z, the second term $\mathbb{E}_{\mathbb{P}}[\|\hat{s}(z)\|_2^2]$ only depends on the first two moments of the distribution $\mathbb{P}$, so is a constant on $\mathcal{C}(\mu, M)$.

For the last term, we use integration by parts to prove that $\mathbb{E}_{\mathbb{P}}[(\nabla \log p(z))^\top \hat{s}(z)]$ is also constant on $\mathcal{C}(\mu, \Sigma)$. Recall that for any scalar function f and vector field g,

$$(\nabla f)^\top g = \operatorname{div}(fg) - f\ \operatorname{div} g,$$

where $\operatorname{div}(g) = \nabla \cdot g = \sum_i \frac{\partial g_i(z)}{\partial z_i}$. Take $f(z) = p(z)$ and $g(z) = \hat{s}(z)$. Then, integrating over all $\mathbb{R}^d$,

$$\mathbb{E}_{\mathbb{P}}[(\nabla \log p(z))^\top \hat{s}(z)] = \underbrace{\int_{\mathbb{R}^d} \operatorname{div}(p\hat{s})\mathrm{d}z}_{A} - \underbrace{\int_{\mathbb{R}^d} p(z)\ \operatorname{div} \hat{s}(z)\mathrm{d}z}_{B}. \tag{A.77}$$

We claim that $A = 0$ and B is a constant for all distributions in $\mathcal{C}$. We first argue the latter: note that $\hat{s}(z) = -\hat{\Sigma}^{-1}(z - \hat{\mu})$, so $\operatorname{div} \hat{s}(z) = -\operatorname{Tr}(\hat{\Sigma}^{-1})$. Integrating this constant over the density $p(z)$ thus proves that B is a constant. To prove that $A = 0$, note that A can be expressed via the Gauss-Ostrogradski Theorem as:

$$A = \lim_{R\to\infty} A_R,\ \text{where } A_R = \int_{B_R} \operatorname{div} F(z)\mathrm{d}z = \int_{S_R} (F(z))^\top n(z)\mathrm{d}S(z).$$

Above, $F(z) = p(z)\hat{s}(z)$ and for $R > 0$, we define $B_R = \{z \in \mathbb{R}^d : \|z\| \leq R\}$ and $S_R = \partial B_R = \{z \in \mathbb{R}^d : \|z\| = R\}$ as the sphere with radius R, with outward unit normal $n(z) = z/\|z\|$ and unit volume/area given by $\mathrm{d}S(z)$. On S_R, we have

$$|(\hat{s}(z)^\top n(z)| = \left|\left(-\hat{\Sigma}^{-1}(z-\hat{\mu})\right)^\top n(z)\right| \leq \|\hat{\Sigma}^{-1}\|\big(\|z\|+\|\hat{\mu}\|\big) = \|\hat{\Sigma}^{-1}\|R+\|\hat{\Sigma}^{-1}\|\|\hat{\mu}\|,\ \forall\, z \in S_R.$$

Therefore,

$$|A_R| \leq \left(\|\hat{\Sigma}^{-1}\|R + \|\hat{\Sigma}^{-1}\|\|\hat{\mu}\|\right) \underbrace{\int_{S_R} p(z)\mathrm{d}S(z)}_{=I_R}.$$

We claim that requirement (A.76) implies that $RI_R \to 0$ as $R \to \infty$, which in turn implies that $I_R \to 0$ and completes the argument that $A = 0$. We have:

$$R \cdot I_R = R\int_{S_R} p(z)\, dS(z) \leq R\int_{S_R} \frac{\bar{p}}{\|z\|^{d+\epsilon}}\mathrm{d}S(z) = \omega_d R^{-\epsilon}\bar{p}, \tag{A.78}$$

where the inequality follows from (A.76) and the last step follows by recalling that the sphere S_R in $\mathbb{R}^d$ has surface $\omega_d R^{d-1}$, for a constant ω_d that depends on d. Then, we readily conclude that the right-hand-side in (A.78) converges to 0 as $R \to \infty$, proving that $A = 0$.

The arguments above show that $\mathbb{F}(\mathbb{P}, \hat{\mathbb{P}}) = \frac{1}{2}\,\Phi(\mathbb{P})$ plus a term that is constant over $\mathcal{C}(\mu, \Sigma)$. By the Stam Inequality (Stam, 1959) and its multidimensional generalizations (see inequality (16) and the related discussion in Rioul (2011)), it is well known that the Fisher information $\Phi(\mathbb{P})$ for any random variable with covariance $\Sigma_{\mathbb{P}}$ satisfies the inequality:

$$\Phi(\mathbb{P}) \geq \operatorname{Tr}(\Sigma_{\mathbb{P}}^{-1}),$$

with equality only if the random variable is Gaussian. This implies that among all distributions $\mathbb{P} \in \mathcal{C}$, the one that minimizes $\mathbb{F}(\mathbb{P}, \hat{\mathbb{P}})$ is a Gaussian distribution.

Part (ii). Consider the set $\mathcal{M}^{\mathbb{F}}_{\mu,M} = \{(\mu, M) \in \mathcal{M}^d_2 : \mathbb{F}(\mathcal{N}(\mu, M), \hat{\mathbb{P}}) \leq \rho\}$. The Fisher divergence between two Gaussian distributions with mean-covariance pairs (μ, Σ) and $(\hat{\mu}, \hat{\Sigma})$, respectively, has the closed form expression (Chafaï, 2021, Eq. (2.2))

$$\left\|\hat{\Sigma}^{-1}(\mu - \hat{\mu})\right\|_2^2 + \operatorname{Tr}\left(\hat{\Sigma}^{-2}\Sigma - 2\hat{\Sigma}^{-1} + \Sigma^{-1}\right), \tag{A.79}$$

We can therefore rewrite:

$$\begin{aligned}\mathbb{F}\left(\mathcal{N}(\mu, M), \hat{\mathbb{P}}\right) &= \left\|\hat{\Sigma}^{-1}(\mu - \hat{\mu})\right\|_2^2 + \operatorname{Tr}\left(\hat{\Sigma}^{-2}(M - \mu\mu^\top) - 2\hat{\Sigma}^{-1} + (M - \mu\mu^\top)^{-1}\right) \\ &= \operatorname{Tr}\Big(\hat{\Sigma}^{-2}\big((\mu - \hat{\mu})(\mu - \hat{\mu})^\top + M - \mu\mu^\top\big)\Big) + \operatorname{Tr}\left(-2\hat{\Sigma}^{-1} + (M - \mu\mu^\top)^{-1}\right).\end{aligned}$$

Recognizing that quadratic terms in μ cancel out, the first trace term above is linear in (μ, M). Thus, the expression is jointly convex in (μ, M) if we can argue that the mapping $(\mu, M) \mapsto \operatorname{Tr}((M - \mu\mu^\top)^{-1})$ is jointly convex in (μ, M) on $\mathcal{M}^d_2$. We claim that for any $S \in \mathbb{S}^d_{++}$, we have the identity:

$$\operatorname{Tr}(S^{-1}) = \sup_{W \in \mathbb{S}^d_+} 2\operatorname{Tr}(W) - \operatorname{Tr}(SW^2). \tag{A.80}$$

To prove this, fix $S \succ 0$ and set $\phi(W) = 2\operatorname{Tr}(W) - \operatorname{Tr}(SW^2)$ for $W \succeq 0$. Because ϕ is concave in W, its maximizers must satisfy the first-order condition $0 = 2I - SW - WS$. The choice $W^\star = S^{-1}$ satisfies this equation and is feasible, and evaluating the objective for $W^\star$ proves (A.80). Applying (A.80) in our case, with $S = M - \mu\mu^\top$, yields:

$$\operatorname{Tr}\left((M - \mu\mu^\top)^{-1}\right) = \sup_{W \in \mathbb{S}^d_+} 2\operatorname{Tr}(W) - \operatorname{Tr}\left((M - \mu\mu^\top)W^2\right).$$

It can be readily seen that the expression maximized above is jointly convex in (μ, M) for any W. Because the supremum of convex functions preserves convexity, we readily obtain that $(\mu, M) \mapsto \operatorname{Tr}((M - \mu\mu^\top)^{-1})$ is jointly convex in (μ, M) on the set $\{(\mu, M) : M - \mu\mu^\top \succ 0\}$, which proves that the set $\mathcal{M}^{\mathbb{F}}_{(\mu,\Sigma)}$ is convex for any $\rho \geq 0$.

That $\mathcal{M}^{\mathbb{F}}_{(\mu,\Sigma)}$ is compact follows because the term $\left\|\hat{\Sigma}^{-1}(\mu - \hat{\mu})\right\|_2^2$ is a coercive function in $\mu \in \mathbb{R}^n$ and $\operatorname{Tr}(\hat{\Sigma}^{-2}M)$ is coercive in $M \in \mathbb{S}^n_+$. □

L.2 Verifying Assumption 3

Proposition 18. *Fix $\hat{\mu}_z = 0$. For any $\rho_z \geq 0$, the Fisher Divergence $\mathbb{F}$ satisfies Assumption 3.*

Proof. Again, we omit subscripts z for notational simplicity. The set of feasible moments $\mathcal{M}^{\mathbb{F}}_{(\mu,M)}$ exactly corresponds to the 0 sublevel set of the function $h : \mathcal{M}^d_2 \times \mathcal{M}^d_2 \to \mathbb{R}$ defined as:

$$\begin{aligned}h\big((\mu, M), (\hat{\mu}, \hat{M})\big) &= \mathbb{F}\big(\mathcal{N}(\mu, M), \mathcal{N}(\hat{\mu}, \hat{M})\big) - \rho \\ &= \|\hat{M}^{-1}(\mu - \hat{\mu})\|^2 + \operatorname{Tr}\Big(\hat{M}^{-2}\big(M - \mu\mu^\top\big)\Big) - 2\operatorname{Tr}(\hat{M}^{-1}) + \operatorname{Tr}\left((M - \mu\mu^\top)^{-1}\right) - \rho.\end{aligned}$$

To complete the proof, note that:

$$h\big((\mu, M), (0, \hat{M})\big) - h\big((0, M), (0, \hat{M})\big) = \operatorname{Tr}\left((M - \mu\mu^\top)^{-1}\right) - \operatorname{Tr}\left(M^{-1}\right) \geq 0,$$

where the inequality follows because the operator $X \mapsto X^{-1}$ reverses order on $\mathbb{S}^d_{++}$ and $M - \mu\mu^\top \preceq M$ for any $M \in \mathbb{S}^d_+$. □

L.3 Verifying Assumption 4

Proposition 19. *Fix $\hat{\mu}_z = 0$. For any $\rho_z > 0$, the Fisher Divergence $\mathbb{F}$ satisfies Assumption 4.*

Proof. We omit the subscript z. Let $g(\Sigma) = \mathbb{F}\big(\mathcal{N}(0,\Sigma), \mathcal{N}(0,\hat{\Sigma})\big) - \rho$. By (A.79), the expression for g is:

$$g(\Sigma) = \mathrm{Tr}\big(\hat{\Sigma}^{-2}\Sigma - 2\hat{\Sigma}^{-1} + \Sigma^{-1}\big) - \rho.$$

That g is differentiable on $\mathbb{S}^d_{++}$ is immediate. That g is minimized at $\Sigma = \hat{\Sigma}$ follows because $\mathbb{F}$ is a divergence, so $\mathbb{F}\big(\mathcal{N}(0,\Sigma), \mathcal{N}(0,\hat{\Sigma})\big)$ is minimized for $\Sigma = \hat{\Sigma}$. That g is convex follows readily from Proposition 17 and its proof. These arguments imply that properties (i)-(ii) are readily satisfied. To verify (iii), note that the gradient is given by $\nabla g(\Sigma) = \hat{\Sigma}^{-2} - \Sigma^{-2}$. Therefore, for $\Sigma_1, \Sigma_2 \in \mathbb{S}^d_{++}$,

$$\nabla g(\Sigma_1) \succeq \nabla g(\Sigma_2) \;\Leftrightarrow\; \Sigma_2^{-2} \succeq \Sigma_1^{-2} \;\Leftrightarrow\; \Sigma_1 \succeq \Sigma_2,$$

where the equivalences follow because $X \mapsto X^{-2}$ is order reversing on $\mathbb{S}^d_{++}$. □

M Extension to Elliptical Nominal Distribution

In this section, we focus on an ambiguity set where $\mathbb{D}$ is the 2-Wasserstein distance $\mathbb{W}$. We will show that our results from Section B continue to hold for any *elliptical* nominal distribution $\hat{\mathbb{P}}$ with finite second moments. To this end, we recall the following definition of elliptical distributions.

Definition 13 (Elliptical distribution)**.** *The distribution $\mathbb{P}_z$ for $z \in \mathbb{R}^{d_z}$ is called elliptical if the characteristic function $\Phi_{\mathbb{P}_z}(t) = \mathbb{E}_{\mathbb{P}_z}[\exp(it^\top z)]$ of $\mathbb{P}_z$ has a representation of the form $\Phi_{\mathbb{P}_z}(t) = \exp(it^\top \mu)\psi(t^\top S t)$ for some $\mu_z \in \mathbb{R}^{d_z}$ (location parameter), some $S_z \in \mathbb{S}^{d_z}_+$ (dispersion matrix), and some function $\psi : \mathbb{R}_+ \to \mathbb{R}$ (characteristic generator).*

This class of distributions includes many non-Gaussian distributions, such as the Laplace, logistic, or hyperbolic distributions, among others (Fang, 2018). Gaussian distributions are a special case of elliptical distributions, with the characteristic generator $\psi(r) = \exp(-r/2)$.

In the remainder of this section, we focus on elliptical distributions with finite second moments. Such distributions can be re-parameterized to ensure that the dispersion matrix S_z exactly equals the covariance matrix Σ_z, which will be convenient for our subsequent discussion. The following remark formalizes this idea.

Remark 3 (Technical remark on re-parametrization)**.** *Denote by $\mathcal{E}$ an arbitrary elliptical distribution of z with finite second moments. Recall that the characteristic function $\Phi_{\mathcal{E}}$ always exists and is finite even if some moments of $\mathcal{E}$ do not exist. Denote the right derivative of $\psi(u)$ at $u = 0$ as $\psi'(0)$. The mean and covariance of $\mathcal{E}$ exist if and only if $\psi'(0)$ exists and is finite, and they can be expressed as μ_z and $-2\psi'(0)S_z$ (Cambanis et al., 1981, Theorem 4), respectively. Denote by $\mathcal{E}^\psi(\mu_z, S_z)$ the elliptical distribution with mean μ_z, dispersion matrix S_z, and characteristic generator ψ. It can be checked that for any admissible ψ with $|\psi'(0)| < \infty$, we have $\mathcal{E}^\psi(\mu_z, S_z) = \mathcal{E}^{\tilde{\psi}}(\mu_z, \tilde{S}_z)$, where $\tilde{\psi}(u) = \psi(-u/(2\psi'(0)))$ and $\tilde{S}_z = -2\psi'(0)S_z$. But then, a choice that ensures $\tilde{\psi}'(0) = -1/2$ will also imply that $\tilde{S}_z = \Sigma_z$, so we can re-parameterize any elliptical distribution so that its dispersion matrix equals its covariance matrix. (The re-parametrization has no effect on the actual distribution.)*

Remark 3 illustrates that any elliptical distribution can be defined by a characteristic generator ψ, a mean μ_z, and a covariance matrix Σ_z – or equivalently, a second-moment matrix M_z – where, without loss of generality, we have $S_z = \Sigma_z$. In the following, we use $\mathcal{E}^\psi(\mu_z, M_z)$ to denote an elliptical distribution with characteristic generator ψ, mean μ_z, and second-moment matrix M_z (and where no confusion can arise, we also use $\mathcal{E}^\psi(\mu_z, \Sigma_z)$ to denote an elliptical distribution with characteristic generator ψ, mean μ_z, and covariance matrix Σ_z). Henceforth, we use the terms *dispersion matrix* and *covariance matrix* interchangeably.

M.1 Assumptions

We now formally present the counterparts of the tractability assumptions from the main text for the setup considered in this appendix section.

Assumption 8. *$\hat{\mathbb{P}}$ is an elliptical distribution with finite second moments.*

Requiring $\hat{\mathbb{P}}$ to be elliptical renders the model studied in this appendix section computationally tractable and is consistent with the assumptions of the so-called linear quadratic-elliptical (LQE) model, which assumes that the exogenous uncertainties follow a known, elliptical distribution and are uncorrelated. Note that if the joint distribution of the random variables is elliptical, mutual uncorrelatedness does not imply independence as in the case of Gaussians, but linear conditioning preserves ellipticity. Similarly to the LQG model, the minimum mean-squared-error state estimators $\hat{x}_t = \mathbb{E}_{\mathbb{P}}[x_t|y_0, \ldots, y_t]$ for the LQE model can be obtained through Kalman filtering and dynamic programming techniques (Basu and Das, 1994; Chu, 1972). In fact, the Kalman filter equations and the expressions for the optimal control inputs discussed in Appendix §F remain unchanged for the LQE case, with the sole difference being that Σ_t and $\Sigma_{t+1|t}$ do not necessarily represent the covariance matrices of the estimated state in the LQE case.

For the rest of our discussion, we also recall that marginal distributions of elliptical distributions are elliptical with the same characteristic generator (Hult and Lindskog, 2002, Corollary 3.1).

Assumption 9. *$\mathbb{D}$ is the 2-Wasserstein distance $\mathbb{W}$.*

Assumption 9 guarantees that a variant of Assumption 2 holds, where Gaussian distributions are replaced by elliptical distributions that share the same characteristic generator as the elliptical nominal distribution $\hat{\mathbb{P}}_z$. This is formalized in Proposition 20.

Proposition 20. *The 2-Wasserstein distance $\mathbb{W}$ satisfies the following properties:*

(i) For every $(\mu_z, M_z) \in \mathcal{M}_2^{d_z}$, an elliptical distribution that shares the same characteristic generator ψ as $\hat{\mathbb{P}}_z$ minimizes the distance $\mathbb{W}$ from $\hat{\mathbb{P}}_z$ among all distributions with mean μ_z and second moment matrix M_z, that is,

$$\mathbb{W}(\mathcal{E}^{\psi}(\mu_z, M_z), \hat{\mathbb{P}}_z) = \left\{ \begin{array}{ll} \inf\limits_{\mathbb{P}_z \in \mathcal{P}(\mathbb{R}^{d_z})} & \mathbb{W}(\mathbb{P}_z, \hat{\mathbb{P}}_z) \\ \text{s.t.} & \mathbb{E}_{\mathbb{P}_z}[z] = \mu_z, \quad \mathbb{E}_{\mathbb{P}_z}[zz^\top] = M_z. \end{array} \right.$$

(ii) The set $\mathcal{M}^{\mathcal{E}^{\psi}}_{(\mu_z, M_z)} = \{(\mu_z, M_z) \in \mathcal{M}_2^{d_z} : \mathbb{W}(\mathcal{E}^{\psi}(\mu_z, M_z), \hat{\mathbb{P}}_z) \leq \rho_z\}$ is convex and compact.

Proposition 20 follows from Theorem 2.1 in Gelbrich (1990), which shows that, for any two distributions $\mathbb{P}_z^1$ and $\mathbb{P}_z^2$ on $\mathbb{R}^{d_z}$ with first and second moments given by $(\mu_1, M_1) \in \mathcal{M}_2^{d_z}$ and $(\mu_2, M_2) \in \mathcal{M}_2^{d_z}$, respectively,

$$\mathbb{W}(\mathbb{P}_z^1, \mathbb{P}_z^2) \geq \mathbb{G}\left((\mu_1, M_1 - \mu_1\mu_1^\top), (\mu_2, M_2 - \mu_2\mu_2^\top)\right),$$

with equality holding if $\mathbb{P}_z^1$ and $\mathbb{P}_z^2$ are elliptical with the same characteristic generator. Here, $\mathbb{G}((\mu_1, \Sigma_1), (\mu_2, \Sigma_2))$ is the Gelbrich distance between two mean-covariance pairs, defined in (A.6). This implies that Proposition 20-(i) is satisfied. Proposition 20-(ii) is also satisfied because the set $\mathcal{M}^{\mathcal{E}^{\psi}}_{(\mu_z, M_z)}$ exactly corresponds to the set of pairs of first and second moments with Gelbrich distance at most ρ_z from the mean-covariance pair for the nominal distribution, which is known to be convex and compact (Nguyen, 2019, Proposition 3.17).

Our treatment considers only 2-Wasserstein distance $\mathbb{W}$ because we are not aware of other divergences $\mathbb{D}$ that satisfy the requirements in Proposition 20 for elliptical distributions $\hat{\mathbb{P}}$.

Subsequently, we refer to the DRLQ problem formulation with elliptical nominal and 2-Wasserstein distance as the “elliptical DRLQ problem" for conciseness.

M.2 Results for the Elliptical DRLQ Problem

It is worth emphasizing that the elliptical DRLQ problem exactly mirrors the Gaussian case with 2-Wasserstein distance discussed in Section A.2. The reason is that the 2-Wasserstein distance between two elliptical distributions $\mathbb{P}$ and $\hat{\mathbb{P}}$ with the same generator ψ exactly equals the 2-Wasserstein distance between two Gaussian distributions with the same first two moments, and both distances are given by the Gelbrich distance between the respective mean-covariance pairs:

$$\begin{aligned}\mathbb{W}\big(\mathcal{E}^{\psi}(\mu_z, M_z), \mathcal{E}^{\psi}(\hat{\mu}_z, \hat{M}_z)\big) &= \mathbb{W}\big(\mathcal{N}(\mu_z, M_z), \mathcal{N}(\hat{\mu}_z, \hat{M}_z)\big) \\ &= \mathbb{G}\big((\mu_z, M_z - \mu_z\mu_z^\top), (\hat{\mu}_z, \hat{M}_z - \hat{\mu}_z\hat{\mu}_z^\top)\big). \end{aligned} \tag{A.81}$$

This directly parallels the Gaussian case in Section A.2 (under the assumption that the nominal distributions have the same means and second moments). For instance, the set of valid pairs of first and second moments $\mathcal{M}^{\mathcal{E}^{\psi}}_{(\mu_z, M_z)}$ defined in Proposition 20 exactly equals the set $\mathcal{M}_{(\mu_z, M_z)}$ defined in Section A.2. Moreover, because most of our constructions and results rely on the set $\mathcal{M}_{(\mu_z, M_z)}$, we can mirror all the main results in Section B. For conciseness, we refrain from formally restating and reproving all results and instead just discuss how they apply to the elliptical DRLQ case.

We consider the reformulation (A.8) of the DRLQ problem in terms of purified observations, and its dual (A.9). These reformulations are applicable here because they do not rely on Assumption 8 and Assumption 9 (or their counterparts in the main text).

Upper Bound for Primal.

Similar to the construction of the primal upper bound in Section B, we can obtain an upper bound for $p^\star$ by enlarging the ambiguity set $\mathcal{B}$ and restricting the policies u_t to affine dependencies. We define the following outer approximation for the ambiguity set:

$$\overline{\mathcal{B}} = \Big\{\mathbb{P} \in \mathcal{P}(\mathbb{R}^{n+T(n+p)}) : \mathbb{P}_z \in \overline{\mathcal{B}}_z,\ \mathbb{E}_{\mathbb{P}}[z'z^\top] = 0\ \forall z, z' \in \mathcal{Z},\ z' \neq z\Big\},$$

where, for all $z \in \mathcal{Z}$,

$$\begin{aligned}\overline{\mathcal{B}}_z = \Big\{&\mathbb{P}_z \in \mathcal{P}(\mathbb{R}^{d_z}) : \exists (\mu_z, M_z) \in \mathcal{M}_2^{d_z} \text{ with} \\ &\mathbb{E}_{\mathbb{P}_z}[z] = \mu_z, \mathbb{E}_{\mathbb{P}_z}[zz^\top] = M_z, \mathbb{W}\left(\mathcal{E}^{\psi}(\mu_z, M_z), \hat{\mathbb{P}}_z\right) \leq \rho_z\Big\},\end{aligned}$$

where ψ is the characteristic generator of $\hat{\mathbb{P}}_z$ (the same for all $z \in \mathcal{Z}$ because the characteristic generator of the marginals is the same as that of $\hat{\mathbb{P}}$). Affine policies take the same form as before, i.e., $u = q + U\eta = q + U(Dw + v)$, where $q = (q_0, \ldots, q_{T-1})$, and U is a block lower triangular matrix as defined in (A.10). The optimal value of Problem (A.11) now constitutes an upper bound for the primal here. Proposition 1 also holds, and the proof follows the same arguments.

Lower Bound for Dual.

We restrict nature's feasible set to the family $\mathcal{B}_{\mathcal{E}^{\psi}}$ of all elliptical distributions from the ambiguity set $\mathcal{B}$ that have the same characteristic generator ψ as $\hat{\mathbb{P}}$. Paralleling the construction in Section B, we formulate a lower bound problem:

$$\underline{d}^\star = \left\{\begin{array}{lll} \max\limits_{\mathbb{P} \in \mathcal{B}_{\mathcal{E}^{\psi}}} & \min\limits_{x,u} & \mathbb{E}_{\mathbb{P}}\left[u^\top R u + x^\top Q x\right] \\ & \text{s.t.} & u \in \mathcal{U}_\eta, x = Hu + Gw. \end{array}\right. \tag{A.82}$$

We can now prove that a variant of Proposition 2 holds.

Proposition 21. *Under the standing assumptions in this section, problem* (A.82) *has the same optimal value as problem* (A.14).

Proof of Proposition 21. The proof follows from arguments similar to those in the proof of Proposition 2. Noting that the inner minimization problem in (A.82) is solved by an affine policy according

to standard LQE theory, we can reformulate problem (A.82) as

$$\begin{aligned}\max_{\substack{\mu_w, M_w,\\ \mu_v, M_v, \mathbb{P}}} \quad \min_{\substack{q\in\mathbb{R}^{p^T}\\ U\in\mathcal{U}}} \quad & \mathrm{Tr}\Big(\big((UD)^\top RUD + (G+HUD)^\top Q(G+HUD)\big)M_w + U^\top \bar{R}UM_v\Big)\\ & +2q^\top(\bar{R}UD + G^\top QH)\mu_w + 2q^\top \bar{R}U\mu_v + q^\top \bar{R}q,\\ \text{s.t.} \quad & \mathbb{P}\in\mathcal{B}_{\mathcal{E}^\psi},\ \mu_w = \mathbb{E}_{\mathbb{P}}[w],\ M_w = \mathbb{E}_{\mathbb{P}}[ww^\top],\ \mu_v = \mathbb{E}_{\mathbb{P}}[v],\ M_v = \mathbb{E}_{\mathbb{P}}[vv^\top].\end{aligned} \tag{A.83}$$

It remains to prove that problem (A.83) has the same optimal value as problem (A.14). Consider any $(\mu_w, M_w), (\mu_v, M_v), \mathbb{P}$ feasible in the outer maximization problem in (A.83). Because $\mathbb{P}\in B_{\mathcal{E}^\psi}$ is an elliptical distribution with characteristic generator ψ and marginal distributions of elliptical distributions are also elliptical with the same characteristic generator, we can write the requirement $\mathbb{W}(\mathbb{P}_z, \hat{\mathbb{P}}_z)\le\rho_z$ in the definition of $\mathcal{B}_{\mathcal{E}^\psi}$ equivalently as $\mathbb{W}(\mathcal{E}^\psi(\mu_z, M_z), \hat{\mathbb{P}}_z)\le\rho_z$, for any $z\in\mathcal{Z}$. By (A.81), this implies that $(\mu_w, M_w), (\mu_v, M_v)$ is feasible in the outer maximization problem in (A.14), proving that the optimal value in (A.14) is at least as large as that in (A.83). However, for any $(\mu_w, M_w), (\mu_v, M_v)$ feasible in the outer maximization in (A.14), the elliptical distribution with characteristic generator ψ, mean (μ_w, μ_v) and covariance matrix $\mathrm{diag}(M_w - \mu_w\mu_w^\top, M_v - \mu_v\mu_v^\top)$ is feasible in (A.83), proving that the two problems have the same optimal value. □

Optimality of Affine Policies and Elliptical Distributions.

A counterpart of Theorem A – i.e., the strong duality of the primal (A.8) and its dual (A.9) – holds under the assumptions in this section, and the proof relies on mirroring arguments. This implies that an affine policy $u = q + U\eta$ is optimal in the elliptical DRLQ problem and that an elliptical distribution with the same characteristic generator as $\hat{\mathbb{P}}$ is optimal for the dual of the elliptical DRLQ problem.

Optimality of Linear Policies and Zero-Mean Distributions.

Mirroring our discussion in Section B, we can readily argue that if $\hat{\mu}_z = 0$ for every $z\in\mathcal{Z}$, Assumption 3 readily holds for the elliptical DRLQ problem. This follows because the distances between two *elliptical* distributions $\mathbb{P}_z^1, \mathbb{P}_z^2\in\mathcal{P}(\mathbb{R}^{d_z})$ (with the same characteristic generator as $\hat{\mathbb{P}}$) is exactly given by the Gelbrich distance between their pairs of first two moments, by (A.81). So an argument that mirrors the one for the 2-Wasserstein distance can be used to show that Assumption 3 holds.

This allows us to extend Theorem B to this case, with an identical line of arguments. We can therefore conclude that the dual of the elliptical DRLQ problem admits an optimal solution $\mathbb{P}^\star$ that is elliptical and has the same mean as the nominal mean (i.e., zero), and under this distribution, there is an optimal *linear* control policy, $u^\star = U^\star\eta$.

Efficient Numerical Solution.

The results in Section C can also be readily extended to the elliptical DRLQ problem because our algorithms rely on solving problem (A.13) and the objective and feasible set of that problem are identical in the elliptical DRLQ case and in the Gaussian case due to (A.81).

N Conclusions, Limitations, and Future Directions

This work formulated a distributionally robust version of the classical LQG problem by replacing the fixed disturbance model with a divergence ball around a nominal distribution. Under zero-mean Gaussian nominal noise, an orthogonality requirement on the second moments of the distributions (equivalent to uncorrelatedness under zero means), and suitable structural requirements on the divergence, we proved that affine output-feedback policies and a Gaussian distribution form a Nash equilibrium in our mini-max game. We also showed that it is optimal for the adversary to set the mean of the distribution to zero, in which case the decision maker's policies become linear and the adversary

optimally "inflates" the nominal covariance matrix. The results generalize and rationalize many results from the (robust) LQG literature, provide an intuitive rule of thumb to address distributional misspecification in practice, and enable a very efficient Frank-Wolfe algorithm whose iterations are standard LQG subproblems. All results extend to an infinite-horizon, average-cost setting – yielding stationary linear policies and a time-invariant Gaussian worst-case model – and to entropy-regularized optimal transport, Fisher divergence, or an elliptical nominal distribution with 2-Wasserstein distance.

Future work could be aimed at extending this framework or addressing some of its limitations. For instance, one could leverage our results to compute control policies with performance guarantees when the disturbance noise distribution is known but otherwise *general*. More specifically, consider a distribution $\mathbb{Q}$ such that $\mathbb{Q} \in \mathcal{B}$ for some ambiguity set $\mathcal{B}$ that is compatible with our assumptions. Then, the optimal solution in the DRLQ problem $\inf_{u \in \mathcal{U}_y} \sup_{\mathbb{P} \in \mathcal{B}} \mathbb{E}_{\mathbb{P}}[J(u)]$ will be feasible in the (intractable) linear quadratic control problem $\inf_{u \in \mathcal{U}_y} \mathbb{E}_{\mathbb{Q}}[J(u)]$, and its optimality gap will be upper bounded by the difference between the optimal value of the DRLQ model and the optimal value in the distributionally robust "optimistic" problem $\inf_{\mathbb{P} \in \mathcal{B}} \inf_{u \in \mathcal{U}_y} \mathbb{E}_{\mathbb{P}}[J(u)]$. For this procedure to be effective, one must be able to solve the latter problem (which may be non-convex) and also test whether a given distribution $\mathbb{Q}$ belongs to $\mathcal{B}$, and possibly adjust $\mathcal{B}$ to guarantee this, without significantly expanding the ambiguity set too much. These tasks are likely challenging for general distributions $\mathbb{Q}$, so future work could be devoted to identifying tractable cases and designing algorithms for membership testing, ambiguity set calibration, and solving the optimistic problem.

One could also consider distributionally robust formulations for other stochastic control problems that extend the LQR or LQG frameworks, such as the linear-exponential-quadratic Gaussian model due to Whittle (1981) or the one recently considered in de Zegher et al. (2019), or the model with affine dynamics and extended quadratic costs from Barratt and Boyd (2022). Alternatively, one could consider formulations that involve state or control constraints and derive policies with provable performance guarantees, or consider control problems while learning the ambiguity set, as in Iancu et al. (2021) and related literature.

References

M. Abeille, S. Ciliberti, A. Rej, and J.-P. Bouchaud. LQG for portfolio optimization. *arXiv:1611.00997*, 2016.

A. Agrawal, R. Verschueren, S. Diamond, and S. Boyd. A rewriting system for convex optimization problems. *Journal of Control and Decision*, 5(1):42–60, 2018.

F. Auger, M. Hilairet, J. M. Guerrero, E. Monmasson, T. Orlowska-Kowalska, and S. Katsura. Industrial applications of the Kalman filter: A review. *IEEE Transactions on Industrial Electronics*, 60(12):5458–5471, 2013.

W. Azizian, F. Iutzeler, and J. Malick. Regularization for Wasserstein distributionally robust optimization. *ESAIM: Control, Optimisation and Calculus of Variations*, 29:1–33, 2023.

S. Barratt and S. Boyd. Stochastic control with affine dynamics and extended quadratic costs. *IEEE Transactions on Automatic Control*, 67(1):320–335, 2022.

A. Basu and J. Das. A Bayesian approach to Kalman filter for elliptically contoured distribution and its application in time series models. *Calcutta Statistical Association Bulletin*, 44(1-2):11–28, 1994.

H. H. Bauschke and P. L. Combettes. *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*. CMS Books in Mathematics. Springer, 2017.

A. Ben-Tal, S. Boyd, and A. Nemirovski. Control of uncertainty-affected discrete time linear systems via convex programming. *Available at Optimization Online*, 2005.

A. Ben-Tal, S. Boyd, and A. Nemirovski. Extending scope of robust optimization: Comprehensive robust counterparts of uncertain problems. *Mathematical Programming*, 107(1):63–89, 2006.

A. Bensoussan, M. Çakanyıldırım, and S. P. Sethi. Partially observed inventory systems: The case of zero-balance walk. *SIAM Journal on Control and Optimization*, 46(1):176–209, 2007.

D. Bertsekas. *Dynamic Programming and Optimal Control*, volume I. Athena Scientific, 2017.

R. Bhatia. *Positive Definite Matrices*. Princeton University Press, 2009.

J. Blanchet, K. Murthy, and V. A. Nguyen. Statistical analysis of Wasserstein distributionally robust estimators. In *Emerging Optimization Methods and Modeling Techniques with Applications*, chapter 8, pages 227–254. INFORMS, 2021.

S. Cambanis, S. Huang, and G. Simons. On the theory of elliptically contoured distributions. *Journal of Multivariate Analysis*, 11(3):368–385, 1981.

D. Chafaï. Fisher information inequalities and convexity of entropy power. In *Entropy and the Quantum*, pages 1–29. American Mathematical Society, 2021.

S. Y. Chen. Kalman filter for robot vision: A survey. *IEEE Transactions on Industrial Electronics*, 59(11):4409–4420, 2012.

K.-c. Chu. Estimation and decision for linear systems with elliptical random processes. In *IEEE Conference on Decision and Control*, pages 647–651, 1972.

T. M. Cover and J. A. Thomas. *Elements of Information Theory*. Wiley-Interscience, 2006.

J. F. de Zegher, D. A. Iancu, and E. L. Plambeck. Sustaining rainforests and smallholders by eliminating payment delay in a commodity supply chain. Technical report, Stanford University, 2019.

E. del Barrio and J.-M. Loubes. The statistical effect of entropic regularization in optimal transportation. *arXiv:2006.05199*, 2020.

S. Diamond and S. Boyd. CVXPY: A Python-embedded modeling language for convex optimization. *Journal of Machine Learning Research*, 17(83):1–5, 2016.

L. El Ghaoui, M. Oks, and F. Oustry. Worst-case value-at-risk and robust portfolio optimization: A conic programming approach. *Operations Research*, 51(4):543–556, 2003.

K. Fan. Minimax theorems. *Proceedings of the National Academy of Sciences*, 39(1):42–47, 1953.

K. W. Fang. *Symmetric Multivariate and Related Distributions*. CRC Press, 2018.

M. Frank and P. Wolfe. An algorithm for quadratic programming. *Naval Research Logistics*, 3(1-2): 95–110, 1956.

M. Gelbrich. On a formula for the L^2 Wasserstein metric between measures on Euclidean and Hilbert spaces. *Mathematische Nachrichten*, 147(1):185–203, 1990.

M. J. Hadjiyiannis, P. J. Goulart, and D. Kuhn. An efficient method to estimate the suboptimality of affine controllers. *IEEE Transactions on Automatic Control*, 56(12):2841–2853, 2011.

B. Han. Distributionally robust Kalman filtering with volatility uncertainty. *arXiv:2302.05993*, 2023.

L. P. Hansen and T. J. Sargent. Robust estimation and control under commitment. *Journal of Economic Theory*, 124(2):258–301, 2005.

Z. Hu and L. J. Hong. Kullback-Leibler divergence constrained distributionally robust optimization. *Available at Optimization Online*, 2013.

H. Hult and F. Lindskog. Multivariate extremes, aggregation and dependence in elliptical distributions. *Advances in Applied Probability*, 34(3):587–608, 2002.

A. Hyvärinen. Estimation of non-normalized statistical models by score matching. *Journal of Machine Learning Research*, 6:695–709, 2005.

D. A. Iancu, N. Trichakis, and D. Y. Yoon. Monitoring with limited information. *Management Science*, 67(7):4233–4251, 2021.

M. Jaggi. Revisiting Frank-Wolfe: Projection-free sparse convex optimization. In *International Conference on Machine Learning*, pages 427–435, 2013.

O. Johnson. *Information Theory and the Central Limit Theorem*. Imperial College Press, 2004.

A. Juditsky and A. Nemirovski. On well-structured convex–concave saddle point problems and variational inequalities with monotone operators. *Optimization Methods and Software*, 37(5): 1567–1602, 2022.

K. Kim and I. Yang. Distributional robustness in minimax linear quadratic control with Wasserstein distance. *SIAM Journal on Control and Optimization*, 61(2):458–483, 2023.

D. Kuhn, P. Mohajerin Esfahani, V. A. Nguyen, and S. Shafieezadeh-Abadeh. Wasserstein distributionally robust optimization: Theory and applications in machine learning. In *Operations Research & Management Science in the Age of Analytics*, pages 130–166. INFORMS, 2019.

P. R. Kumar and P. Varaiya. *Stochastic Systems: Estimation, Identification, and Adaptive Control*. SIAM, 2015.

P. Lancaster and L. Rodman. *Algebraic Riccati Equations*. Clarendon Press, 1995.

E. S. Levitin and B. T. Polyak. Constrained minimization methods. *USSR Computational Mathematics and Mathematical Physics*, 6(5):1–50, 1966.

MOSEK ApS. *The MOSEK Optimization Toolbox. Version 9.2.*, 2019.

V. A. Nguyen. *Adversarial Analytics*. PhD thesis, EPFL, 2019.

V. A. Nguyen, S. Shafieezadeh-Abadeh, D. Kuhn, and P. Mohajerin Esfahani. Bridging Bayesian and minimax mean square error estimation via Wasserstein distributionally robust optimization. *Mathematics of Operations Research*, 48(1):1–37, 2023.

A. Paszke, S. Gross, S. Chintala, G. Chanan, E. Yang, Z. DeVito, Z. Lin, A. Desmaison, L. Antiga, and A. Lerer. Automatic differentiation in PyTorch. In *Neural Information Processing Systems 2017 Workshop on Autodiff*, 2017.

A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala. Pytorch: An imperative style, high-performance deep learning library. In *Advances in Neural Information Processing Systems Workshop*, page 8026–8037, 2019.

S. D. Patek, M. D. Breton, Y. Chen, C. Solomon, and B. Kovatchev. Linear quadratic Gaussian-based closed-loop control of type 1 diabetes. *Journal of Diabetes Science and Technology*, 1(6):834–841, 2007.

L. A. Rademacher. Approximating the centroid is hard. In *Annual Symposium on Computational Geometry*, pages 302–305, 2007.

O. Rioul. Information theoretic proofs of entropy power inequalities. *IEEE Transactions on Information Theory*, 57(1):33–55, 2011.

P. Schiele, E. S. Luxenberg, and S. P. Boyd. Disciplined saddle programming. *Transactions on Machine Learning Research*, 1:1–25, 2024.

M. Sion. On general minimax theorems. *Pacific Journal of Mathematics*, 8(1):171–176, 1958.

J. Skaf and S. P. Boyd. Design of affine controllers via convex optimization. *IEEE Transactions on Automatic Control*, 55(11):2476–2487, 2010.

Y. Song and S. Ermon. Generative modeling by estimating gradients of the data distribution. In *Advances in Neural Information Processing Systems*, pages 11895–11907, 2019.

A. J. Stam. Some inequalities satisfied by the quantities of information of fisher and shannon. *Information and Control*, 2(2):101–112, 1959.

B. Taşkesen, M.-C. Yue, J. Blanchet, D. Kuhn, and V. A. Nguyen. Sequential domain adaptation by synthesizing distributionally robust experts. In *International Conference on Machine Learning*, pages 10162–10172, 2021.

B. Taşkesen, D. A. Iancu, Ç. Koçyiğit, and D. Kuhn. Distributionally robust linear quadratic control. In *Advances in Neural Information Processing Systems*, pages 18613–18632, 2023.

E. Todorov and M. I. Jordan. Optimal feedback control as a theory of motor coordination. *Nature Neuroscience*, 5(11):1226–1235, 2002.

J. Townsend, N. Koep, and S. Weichwald. Pymanopt: A Python toolbox for optimization on manifolds using automatic differentiation. *Journal of Machine Learning Research*, 17(137):1–5, 2016.

J. Wang, R. Gao, and Y. Xie. Sinkhorn distributionally robust optimization. *arXiv:2109.11926*, 2021.

P. Whittle. Risk-sensitive linear/quadratic/Gaussian control. *Advances in Applied Probability*, 13(4): 764–777, 1981.