

Linear Quadratic Control with Risk Constraints ^{*}

Anastasios Tsiamis [‡], Dionysios S. Kalogerias [‡], Alejandro Ribeiro[†], and George J. Pappas[†]

December 15, 2021

Abstract

We propose a new risk-constrained formulation of the classical Linear Quadratic (LQ) stochastic control problem for general partially-observed systems. Our framework is motivated by the fact that the risk-neutral LQ controllers, although optimal in expectation, might be ineffective under relatively infrequent, yet statistically significant extreme events. To effectively trade between average and extreme event performance, we introduce a new risk constraint, which explicitly restricts the total expected predictive variance of the state penalty by a user-prescribed level. We show that, under certain conditions on the process noise, the optimal risk-aware controller can be evaluated explicitly and in closed form. In fact, it is affine relative to the minimum mean square error (mmse) state estimate. The affine term pushes the state away from directions where the noise exhibits heavy tails, by exploiting the third-order moment (skewness) of the noise. The linear term regulates the state more strictly in riskier directions, where both the prediction error (conditional) covariance and the state penalty are simultaneously large; this is achieved by inflating the state penalty within a new filtered Riccati difference equation. We also prove that the new risk-aware controller is internally stable, regardless of parameter tuning, in the special cases of i) fully-observed systems, and ii) partially-observed systems with Gaussian noise. The properties of the proposed risk-aware LQ framework are lastly illustrated via indicative numerical examples.

1 Introduction

In the problem of Linear Quadratic (LQ) stochastic control, one is typically interested in optimizing average control performance for linear systems of the form

$$\begin{aligned} x_{t+1} &= Ax_t + Bu_t + w_{t+1} \\ y_t &= Cx_t + v_t, \end{aligned} \quad (1)$$

where $x_t \in \mathbb{R}^n$ is the state, $y_t \in \mathbb{R}^m$ is the measured output, u_t is input, and w_t, v_t are process and measurement noise disturbances. A standard approach is to minimize the expectation of the following quadratic cost comprising of stage-wise input and state penalties up to a horizon N

$$\begin{aligned} \min_u \quad & \mathbb{E} \left\{ x_N' Q x_N + \sum_{t=0}^{N-1} x_t' Q x_t + u_t' R u_t \right\}, \\ \text{s.t.} \quad & \text{Dynamics (1)} \end{aligned} \quad (2)$$

^{*}This work was supported by the AFOSR under grant FA9550-19-1-0265 (Assured Autonomy in Contested Environments) and by the ARL under grant DCIST CRA W911NF-17-2-0181.

[†]Department of Electrical and Systems Engineering, University of Pennsylvania, Philadelphia, PA 19104 (email: {atsiamis, aribeiro, pappas}@seas.upenn.edu).

[‡]Department of Electrical Engineering, Yale University, New Haven, CT 06511 (email: dionysis.kalogerias@yale.edu).

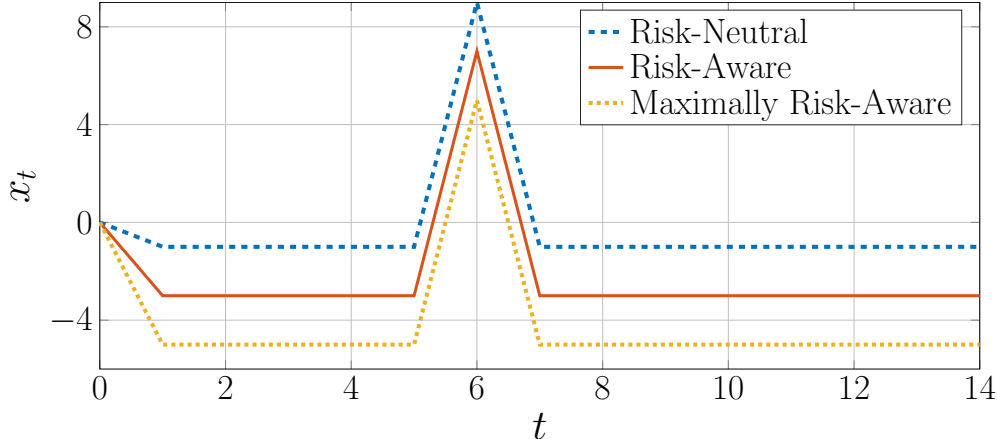


Figure 1: Comparison between risk-neutral and risk-aware control performance, when the system experiences rare but large shocks—here the shock occurs at time 6. By sacrificing average behavior, the risk-aware controllers push the state away from the direction of the shock.

where matrices Q , R are design choices.

While LQ control has been a standard approach to controlling stochastic systems, it only focuses on *average* performance, which might be an insufficient objective when dealing with critical applications. Examples of such applications appear naturally in many areas, including wireless industrial control [1], energy [2, 3], finance [4–6], robotics [7, 8], networking [9], and safety [10, 11], to name a few. Indeed, occurrence of less probable, non-typical or unexpected events might lead the underlying dynamical system to experience shocks with possibly catastrophic consequences, e.g., a drone diverging too much from a given trajectory in a hostile environment, or an autonomous vehicle crashing onto a wall or hitting a pedestrian. In such situations, design of effective *risk-aware* control policies is highly desirable, systematically compensating for those extreme events, at the cost of slightly sacrificing average performance under nominal conditions.

To highlight the usefulness of a risk-aware control policy, let us consider the following simple, motivating example. Let $x_{k+1} = x_k + u_k + w_{k+1}$ model an aerial robot, moving along a line. Assume that the process noise w_k is *i.i.d.* Bernoulli, taking the values $\beta > 2$ with probability $1/\beta$ and 0 with probability $1 - 1/\beta$. This noise represents shocks, e.g., wind gusts, that can occur with some small probability. We would then like to minimize the LQR cost $\mathbb{E} \sum_{t=0}^N \{x_t^2\}$, i.e., the total displacement of the robot over a horizon of N time steps. In this case, the LQR optimal controller is $u_k^{\text{LQR}} = -x_k - 1$, where $-1 \equiv -\mathbb{E}w_k$ cancels the mean of the process noise. We see that the LQR solution is *risk-neutral*, as it does not account for the fact that the shock β could be arbitrarily large. On the other hand, the risk-aware LQR formulation proposed in this work results in a *family* of optimal controllers of the form

$$u_t^*(\mu) = -x_t - 1 - \frac{\mu}{1 + 2\mu}(\beta - 2), \quad \mu \geq 0,$$

where μ controls the trade-off between average performance and risk. As μ increases, we move from the risk-neutral to the *maximally risk-aware* controller $u_t^*(\infty) = -x_t - \beta/2$, which treats the noise as adversarial—see Fig. 1.

In both classical and recent literature in linear-quadratic problems, risk awareness in estimation and control is typically achieved by replacing the respective random cost with its exponentiation [12–21]. Yet, the resulting stochastic control problem might not be well-defined for general classes of noise distributions, as it requires the moment generating function of the cost to be finite. Thus, heavy-tailed or skewed distributions,

which are precisely those exhibiting high risk, are naturally excluded. Also, even if the expectation of the exponential cost is finite, it does not lead to a general, closed-form and interpretable solution. A notable exception is that of Gaussian noise, also known as the Linear Exponential Quadratic Gaussian (LEQG) problem, which does enjoy a simple closed-form solution [13, 22]. Apparently though, the Gaussian assumption is unable to capture distributions with asymmetric (skewed) structure, as in the above example.

Our contributions are as follows:

–New Risk-Constrained Formulation. We introduce a new risk-constrained formulation for the problem of LQ control in the case of partially-observed systems. The standard LQ objective is minimized subject to a total expected predictive variance risk constraint with respect to the state penalties. By tuning the risk constraint, we can trade between average performance and statistical variability of the state penalties.

–General Noise Models. Contrary to the LEQG approach, our risk-constrained formulation is well-defined for general noise distributions, provided the associated fourth-order moments of the process noise are finite; thus, heavy-tailed or skewed noises are supported within our framework. For fully-observed systems, the optimal control law can be explicitly characterized under the same condition of finite fourth-order moments. In the case of general partially-observed systems, in order to characterize the optimal controller, we require the additional sufficient condition that all higher-order moments of the process noise exist. In any case, we do not require the existence of a moment generating function.

–Characterization of Optimal Risk-Aware Controls. Under the aforementioned regularity conditions on the process noise, the constrained LQ problem admits a closed-form solution with a natural interpretation. The optimal risk-aware feedback controller is affine with respect to the optimal observer. The affine component pushes the state away from directions where the state prediction error exhibits (skewed) heavy tails. Meanwhile, the state feedback gain satisfies a new risk-aware filtered Riccati recursion, in which the state penalty is inflated in riskier directions, where both the (conditional) covariance of the state prediction error and the state penalty are simultaneously larger. Interestingly, the separation principle holds, in the sense that the optimal observer is the minimum mean-square error estimator, which is designed independently of the control objective. To explicitly compute the parameters of the affine optimal control law, it is required to track several conditional moments, which might be a hard problem in general.

–Explicit Risk-Aware LQR and LQG controllers. In the special case of fully-observed systems (Linear Quadratic Regulator (LQR)) we can explicitly compute the optimal control law. The same is true for the case of partially-observed systems with Gaussian noise (Linear Quadratic Gaussian (LQG) control). Further, we show that our optimal risk-aware controllers are always stable, under standard controllability/observability conditions. Interestingly, *by appropriate re-parameterization*, our risk-aware LQR problem is equivalent to a generalized risk-neutral LQR problem with a tracking objective. Essentially, this implies that risk-neutral LQR formulations can provide inherent risk-averse behavior, as long as the involved parameters are selected in a principled way, as presented herein. A similar property holds for the risk-aware LQG problem.

1.1 Related Work

Risk-aware optimization. Risk-aware optimization has been studied in a wide variety of decision making contexts [11, 23–33]. The basic idea is to replace expectations by more general functionals, called risk measures [34], purposed to effectively quantify the statistical volatility of the involved random cost function, in addition to mean performance. Typical examples are mean-variance functionals [4, 34], mean-semideviations [27], and Conditional Value-at-Risk (CVaR) [35].

CVaR-optimal control. In the case of control systems, CVaR optimization techniques have also been considered for risk-aware constraint satisfaction [11]. Although CVaR captures variability and tail events well, CVaR optimization problems rarely enjoy closed-form expressions. Approximations are usually re-

quired to make computations tractable, e.g., process noise and controls are assumed to be finite-valued [11]. Recently, in [36] a tractable upper bound was derived for the CVaR-LQR problem based on the assumption that the noise is finite-valued.

Robust control. Another related concept is that of robust control, where the system model or the noise profile is unknown [37–39]. The objective is to optimally control the true system under worst case model uncertainty. On the contrary, in risk-aware control, extreme noise events are part of the system model; they are not the outcome of model mismatch. Even if the system is exactly modeled, we would still need to consider risk-aware control if the process noise is heavy-tailed or highly variable. From this point of view, robustness and risk are complementary concepts.

Mixed $\mathcal{H}_2/\mathcal{H}_\infty$ control. Regret-optimal control Interestingly, there is a connection between mixed $\mathcal{H}_2/\mathcal{H}_\infty$ control and risk-aware LEQG control [40, 41]. By increasing the exponential parameter in the LEQG control law, we trade average performance (\mathcal{H}_2) for closed-loop responses with smaller \mathcal{H}_∞ norm. Another way to trade between robustness and performance was introduced in [42], where the worst-case regret with respect to non-causal \mathcal{H}_2 policies is minimized.

Predictive variance. Recently, in our previous work [43], we introduced predictive variance as a new risk measure for LQR control and used it in a risk-constrained optimal control formulation. The results of [43] were extended to the infinite horizon case in [44]. The performance of the policy gradient algorithm in the case of risk-constrained Linear Quadratic Regulators was also studied in [45]. Predictive variance constraints have also been used as a measure of risk in portfolio optimization [46]; different from our paper, the noise is limited to Gaussian distributions and the variance is with respect to linear stage costs. Note that our previous work [43] contains only preliminary results for fully-observed systems. Here, we study the more general and challenging problem of LQ control in the case of partially observed systems. In fact, the optimal feedback law in the case of partially-observed systems can be quite different from the feedback law in the fully-observed case, even in the case of Gaussian noise—see Section 7 for more details.

Notation and Structure: The transpose operation is denoted by $(\cdot)'$. If x_k, \dots, x_t is a sequence of vectors, then $x_{k:t}$ denotes the batch vector of all x_i for $k \leq i \leq t$. We use the notation $\|\cdot\|_2$ to denote both the square norm of vectors and the spectral norm of matrices. The σ -algebra generated by a random vector x is denoted by $\sigma(x)$. By $\mathcal{L}_p(\mathcal{F})$, we denote the space of \mathcal{F} -measurable random variables (vectors) with finite p -order moments. The remaining paper is structured as follows. In Section 2, we introduce our risk-aware formulation of LQ control. In Sections 3, 4 we show that our the risk-aware LQ problem can be reformulated as a Quadratically Constrained Quadratic Problem and solved by exploiting Lagrangian duality. In Sections 5, 7, we provide explicit control laws for the problem of risk-aware LQR and risk-aware LQG control respectively. In Section 6, we characterize the optimal control laws in the case of general partially observed systems. We conclude with numerical simulations in Section 8 and with remarks in Section 9.

2 Risk-Constrained LQ Formulation

Consider system (1), where $x_t \in \mathbb{R}^n$ is the state, $u_t \in \mathbb{R}^p$ is the control signal, and $y_t \in \mathbb{R}^m$ is the measured output. Matrix $A \in \mathbb{R}^{n \times n}$ is the state transition matrix, $B \in \mathbb{R}^{n \times p}$ is the input matrix, and $C \in \mathbb{R}^{m \times n}$ is the output matrix. We assume that the initial value x_0 is deterministic and fixed. Signal $w_t \in \mathbb{R}^n$ is a random process noise, while $v_t \in \mathbb{R}^m$ is a random measurement noise. The process (w_t, v_t) is assumed to be *i.i.d* across time, but it can have any joint distribution (possibly non-Gaussian). For $t \geq 0$, let $\mathcal{F}_t = \sigma(y_{0:t}, u_{0:t})$ be the σ -algebra generated by all observables up to time t , and let \mathcal{F}_{-1} be the trivial σ -algebra. Based on this notation, u_t is \mathcal{F}_t -measurable, while (w_{t+1}, v_{t+1}) is independent of \mathcal{F}_t . We also make an additional

assumption on the process noise.

Assumption 1 (Noise Regularity). *The process w_t has finite fourth-order moment, i.e., for every $t \in \mathbb{N}$, $\mathbb{E} \|w_t\|_2^4 < \infty$.*

The above mild regularity condition is required for our risk measure to be well-defined. It is satisfied by general noise distributions, including many heavy-tailed ones. Denote the mean of the noise by $\bar{w} \triangleq \mathbb{E}w_k$ and its variance by $W \triangleq \mathbb{E}(w_k - \bar{w})(w_k - \bar{w})'$.

As discussed in Section 1, the classical LQ problem is risk-neutral, since it optimizes performance only on average [47]. Still, even if average performance is good, the state can grow arbitrarily large under less probable, yet extreme events. In other words, the state can exhibit large variability. To deal with this issue, we propose a risk-constrained formulation of the LQ control problem, posed as

$$\begin{aligned} \min_u \quad & \mathbb{E} \left\{ x_N' Q x_N + \sum_{t=0}^{N-1} x_t' Q x_t + u_t' R u_t \right\} \\ \text{s.t.} \quad & \mathbb{E} \left\{ \sum_{t=1}^N [x_t' Q x_t - \mathbb{E}(x_t' Q x_t | \mathcal{F}_{t-1})]^2 \right\} \leq \epsilon, \\ & \text{Dynamics (1)} \\ & u_t \in \mathcal{L}_4(\mathcal{F}_t), t = 0, \dots, N-1 \end{aligned} \tag{3}$$

where $u = u_{0:N-1}$ are the inputs from time 0 up to time $N-1$, for some horizon $N \in \mathbb{N}$. For each t , the *causality constraint* on u_t restricts the inputs to the space of \mathcal{F}_t -measurable random vectors of appropriate dimension with bounded fourth-order moments, denoted as $\mathcal{L}_4(\mathcal{F}_t)$. Here, the risk measure adopted is the (*cumulative expected*) *predictive variance* of the state cost. The predictive variance incorporates information about the tail *and* skewness of the penalty $x_t' Q x_t$. This forces the controller to take higher-order noise statistics into account, mitigating the effect of rare though large noise values. Hence, our risk-aware LQ formulation not only forces the state x_t to be close to zero, but also explicitly restricts its variability. The initial state is fixed (for simplicity), so there is no associated risk term for $t = 0$. The fourth-order integrability constraint on the inputs along with Assumption 1 are sufficient to guarantee that the cumulative expected predictive variance is well-defined.

Remark 1 (Input integrability). The fourth-order integrability condition $u_t \in \mathcal{L}_4(\mathcal{F}_t)$ on the inputs is stricter compared with the risk-neutral formulation, where only square-integrability is needed. In the general case of partially-observed systems, this condition is needed to guarantee that the constraint in (3) is well-defined. However, in many cases of interest, this condition is not essential. For example, in the fully observed case (Section 5), we can pose problem (3) with the constraint $u_t \in \mathcal{L}_2(\mathcal{F}_t)$ and the optimal control is still guaranteed to be in $\mathcal{L}_4(\mathcal{F}_t)$; this is a byproduct of the noise regularity Assumption 1. The same holds for the case of partially-observed systems with Gaussian noise (Section 7). \diamond

Problem (3) offers a simple and interpretable way to control the trade-off between average performance and risk. By simply decreasing ϵ , we increase risk-awareness. Inspired by standard risk-aware formulations, in the above optimization problem our risk definition is tied to the specific state penalty $x_t' Q x_t$. However, all of our results are still valid if we employ the predictive variance of a different quadratic form, e.g., the norm of the state, $\|x_t\|_2^2$, in the constraint. In the following sections, we characterize the optimal controllers in the case of general partially-observed systems. We also provide explicit, finite-dimensional control laws for the case of i) fully-observed systems with general noise, which we term risk-aware LQR controllers; and ii) partially-observed systems with Gaussian noise, which we term risk-aware LQG controllers.

3 Quadratic Reformulation of Risk-Constrained LQ Control

The solution procedure of the risk-aware dynamic program (3) consists of the following steps. First, we ensure the well-definiteness of (3), also showing that (3) can be equivalently expressed as a *sequential variational Quadratically Constrained Quadratic Program (QCQP)*, or, more precisely, as a *Quadratically Constrained LQ (QC-LQ)* problem (Proposition 1). Then, we exploit Lagrangian duality (Theorem 1) to solve (3) *exactly and in closed form*. More specifically, we first derive an explicit expression for the optimal risk-aware controller (Theorems 4, 5, 7), given an arbitrary but fixed Lagrange multiplier. Then, we show how an optimal Lagrange multiplier may be efficiently discovered via trivial bisection (Theorem 2).

Since we are dealing with partially observed systems, we can only approximately estimate the current state x_t based on the information \mathcal{F}_t collected so far. Define the state estimate and the state prediction at time t respectively as

$$\begin{aligned}\hat{x}_{t|t} &= \mathbb{E}(x_t|\mathcal{F}_t) \\ \hat{x}_t &= \mathbb{E}(x_t|\mathcal{F}_{t-1}).\end{aligned}$$

Note that both values are mean-square optimal, i.e. they minimize the mean square estimation error (prediction error respectively) [48]. Under Assumption 1 on w_t , and since $u_t \in \mathcal{L}_4(\mathcal{F}_t)$ both expectations are well-defined. The state prediction and the state estimate are related via the expression

$$\hat{x}_t = A\hat{x}_{t-1|t-1} + Bu_{t-1} + \bar{w}.$$

The innovation (or prediction) error is defined as

$$\delta_t \triangleq x_t - \hat{x}_t. \quad (4)$$

Define also the refinement error between the prediction and the estimate:

$$e_t \triangleq \hat{x}_{t|t} - \hat{x}_t. \quad (5)$$

Both errors are martingale differences, satisfying the mean conditions $\mathbb{E}(e_t|\mathcal{F}_{t-1}) = 0$, $\mathbb{E}(\delta_t|\mathcal{F}_{t-1}) = 0$. Note that in the general case of non-Gaussian noise, the innovation error δ_t is not i.i.d. and not independent of the past in general.

In the following result, we show that the predictive variance constraint has an underlying quadratic structure.

Proposition 1 (Quadratic Reformulation). *Let Assumption 1 be in effect and define the (random) conditional moments*

$$\begin{aligned}W_{t-1} &\triangleq \mathbb{E}(\delta_t \delta_t' | \mathcal{F}_{t-1}) \\ m_{3,t-1} &\triangleq 2Q \mathbb{E} \{ \delta_t \delta_t' Q \delta_t | \mathcal{F}_{t-1} \} \\ m_{4,t-1} &\triangleq \mathbb{E} \{ (\delta_t' Q \delta_t - \text{tr}(QW_{t-1}))^2 | \mathcal{F}_{t-1} \}.\end{aligned}$$

Then, the risk-constrained LQ problem (3) is well-defined and equivalent to the sequential variational QCQP

$$\min_u J(u) \triangleq \mathbb{E} \left\{ x_N' Q x_N + \sum_{t=0}^{N-1} x_t' Q x_t + u_t' R u_t \right\} \quad (6)$$

$$\text{s.t.} \quad J_R(u) \triangleq \mathbb{E} \left\{ \sum_{t=1}^N 4\hat{x}_t' Q W_{t-1} Q \hat{x}_t + 2\hat{x}_t' m_{3,t-1} \right\} \leq \bar{\epsilon}$$

Dynamics (1)
 $u_t \in \mathcal{L}_4(\mathcal{F}_t), t = 0, \dots, N-1,$

where $\bar{\epsilon} \triangleq \epsilon - \sum_{t=1}^N \mathbb{E} m_{4,t-1}$.

Proof. Define the state penalty difference:

$$\Delta_t \triangleq x_t' Q x_t - \mathbb{E} (x_t' Q x_t | \mathcal{F}_{t-1}). \quad (7)$$

Since the inputs $u_t \in \mathcal{L}_4$ and the disturbances $w_t \in \mathcal{L}_4$ have finite fourth moments, it follows that $x_t \in \mathcal{L}_4$ since it is a linear combination of inputs and disturbances. As a result, $x_t' Q x_t$ is integrable, and Δ_t is well defined. Next, we find an expression for Δ_t^2 . By the definition of \hat{x}_t, δ_t

$$x_t = \hat{x}_t + \delta_t,$$

where the prediction $\hat{x}_t \in \mathcal{L}_4$ is well-defined since $x_t \in \mathcal{L}_4$. Similarly $\delta_t \in \mathcal{L}_4$. Based on the above decomposition, the quadratic form becomes

$$x_t' Q x_t = \hat{x}_t' Q \hat{x}_t + 2\hat{x}_t' Q \delta_t + \delta_t' Q \delta_t.$$

All three terms are integrable since $x_t \in \mathcal{L}_4$ and $\delta_t \in \mathcal{L}_4$. By orthogonality, the cross terms have zero expected value:

$$\mathbb{E}(\hat{x}_t' Q \delta_t | \mathcal{F}_{t-1}) = \hat{x}_t' Q (\mathbb{E}(\delta_t | \mathcal{F}_{t-1})) = 0.$$

This implies that

$$\mathbb{E}(x_t' Q x_t | \mathcal{F}_{t-1}) = \hat{x}_t' Q \hat{x}_t + \mathbb{E}(\delta_t' Q \delta_t | \mathcal{F}_{t-1})$$

As a result, we obtain the expression

$$\Delta_t = \delta_t' Q \delta_t - \text{tr}(Q W_{t-1}) + 2\hat{x}_t' Q \delta_t,$$

which leads to

$$\begin{aligned} \Delta_t^2 &= (\delta_t' Q \delta_t - \text{tr}(Q W_{t-1}))^2 + 4\hat{x}_t' Q \delta_t \delta_t' Q \hat{x}_t \\ &\quad + 4\hat{x}_t' Q \delta_t (\delta_t' Q \delta_t - \text{tr}(Q W_{t-1})). \end{aligned}$$

Finally, we show that Δ_t^2 are integrable. Integrability of all terms follows from the existence of the fourth moments of δ_t, u_t, \hat{x}_t and Hölder's inequality

$$\mathbb{E} \|\alpha\|_2^p \|\beta\|_2^q \leq (\mathbb{E} \|\alpha\|_2^4)^{p/4} (\mathbb{E} \|\beta\|_2^4)^{q/4},$$

for $p + q = 4, p, q \geq 0$. Hence, the total expected predictive-variance $\mathbb{E} \sum_{t=0}^{N-1} \Delta_t^2$ is well-defined. Moreover, we have

$$\mathbb{E} \{ \Delta_t^2 | \mathcal{F}_{t-1} \} = 4\hat{x}_t' Q W_{t-1} Q \hat{x}_t + 2\hat{x}_t' m_{3,t-1} + m_{4,t-1}.$$

To complete the proof, we take the expectation and move the $m_{4,t-1}$ terms to the right-hand side of the constraint. \square

The above reformulation enables us to apply duality theory, as discussed next. Note that the equivalent constraint is quadratic. The quadratic and linear penalties $W_{t-1}, m_{3,t-1}$ are random variables and depend on the observations up to time $t-1$. If the prediction error δ_t is independent of the past \mathcal{F}_{t-1} , e.g. in the special case of fully observed systems or Gaussian noise, then the penalties are deterministic and the above expressions can be simplified.

4 Lagrangian Duality

To tackle problem (3), we now consider the *variational Lagrangian* $\mathcal{L} : \mathcal{L}_2(\mathcal{F}_0) \times \cdots \times \mathcal{L}_2(\mathcal{F}_{N-1}) \times \mathbb{R}_+ \rightarrow \mathbb{R}$ of the sequential QCQP (6), defined as

$$\mathcal{L}(u, \mu) \triangleq J(u) + \mu J_R(u) - \mu \bar{\epsilon}, \quad (8)$$

where $\mu \in \mathbb{R}_+$ is a multiplier associated with the variational risk constraint of (6). Hereafter, problem (6) will be called the *primal problem*. Accordingly, the *dual function* $D : \mathbb{R}_+ \rightarrow [-\infty, \infty)$ is additionally defined as

$$D(\mu) \triangleq \inf_{u \in \mathcal{U}_0} \mathcal{L}(u, \mu), \quad (9)$$

where the *implicit feasible set* \mathcal{U}_0 obeys ($k \leq N - 1$)

$$\mathcal{U}_k \triangleq \left\{ u_{k:N-1} \in \prod_{t=k}^{N-1} \mathcal{L}_4(\mathcal{F}_t) \left| \begin{array}{l} x_{t+1} = Ax_t + Bu_t + w_{t+1} \\ y_t = Cx_t + v_t \end{array} \right. \right\},$$

and contains the constraints of (6) that have not been dualized in the construction of the Lagrangian in (8). Note that it is always the case that $D \leq J^*$ on \mathbb{R}_+ , where $J^* \in [0, \infty]$ denotes the optimal value of the primal problem (6). Then, the optimal value of the always concave *dual problem*

$$\sup_{\mu \geq 0} D(\mu) \equiv \sup_{\mu \geq 0} \inf_{u \in \mathcal{U}_0} \mathcal{L}(u, \mu), \quad (10)$$

$D^* \triangleq \sup_{\mu \geq 0} D(\mu) \in [-\infty, \infty]$, is the tightest under-estimate of J^* , when knowing only D .

Leveraging Lagrangian duality, we may now state the following result, which provides sufficient optimality conditions for the QCQP (6). The proof is omitted, as it follows as direct application of [49, Theorem 4.10].

Theorem 1 (Optimality Conditions). *Let Assumption 1 be in effect. Suppose that there exists a feasible policy-multiplier pair $(u^*, \mu^*) \in \mathcal{U}_0 \times \mathbb{R}_+$ such that*

1. $\mathcal{L}(u^*(\mu^*), \mu^*) = \min_{u \in \mathcal{U}_0} \mathcal{L}(u, \mu^*) = D(\mu^*)$;
2. $J_R(u^*) \leq \bar{\epsilon}$, i.e., the dualized variational risk constraint of (6) is satisfied by control policy u^* ;
3. $\mu^*(J_R(u^*) - \bar{\epsilon}) = 0$, i.e., complementary slackness holds.

Then, u^* is optimal for both the primal problem (6) and the initial problem (3), μ^* is optimal for the dual problem (10), and (6) exhibits zero duality gap, that is, $D^* \equiv P^* < \infty$.

Theorem 1 will be serving as the backbone of our analysis towards the solution to problem (6). It is sufficient to compute the relaxed optimal input $u^*(\mu)$ of the Lagrangian in (9), for any given multiplier $\mu \geq 0$. Then, we can also compute an optimal multiplier μ^* via bisection, thus providing a complete solution to the primal problem. The use of bisection is based on the following theorem (the proof can be found in the Appendix).

Theorem 2 (Optimal Multiplier). *Let Assumption 1 be in effect. Assume that for any $\mu \geq 0$ the minimum in (9) is attained by a control policy $u^*(\mu) \in \mathcal{U}_0$. Assume that the risk constraint functional $J_R(u^*(\cdot))$ is a continuous function of μ . Then, the following statements are true:*

1. The LQ cost $J(u^*(\mu))$ is increasing with $\mu \geq 0$, while the risk constraint functional $J_R(u^*(\mu))$ is decreasing.
2. Define the multiplier

$$\mu^* \triangleq \inf \{ \mu \geq 0 : J_R(u^*(\mu)) \leq \bar{\epsilon} \}. \quad (11)$$

If μ^* is finite, then the policy $u^*(\mu^*)$ is optimal for the primal problem (6), and this is the case as long as (6) satisfies Slater's condition:

$$J_R(u^\dagger) < \bar{\epsilon}, \text{ for some } u^\dagger \in \mathcal{U}_0.$$

The above result exploits the fact that, under the relaxed optimal policy $u^*(\cdot)$, both the LQ cost $J(u^*(\cdot))$ and the risk functional $J_R(u^*(\cdot))$ are monotone functions. Note that in order to apply Theorem 2, we need to verify three conditions i) existence of an optimal solution $u^*(\mu)$, ii) continuity of $J_R(u^*(\mu))$, and iii) satisfaction of Slater's condition. This is the subject of the following sections.

5 Optimal Risk-Aware LQR Control

Let us study first the simpler case of fully-observed systems, where $y_k = x_k$, i.e. there is no measurement noise $v_k = 0$ and the output matrix is the identity $C = I$. This problem is the risk-constrained version of the classical Linear Quadratic Regulator (LQR) problem. In this case, the conditional moments in Proposition 1 can be simplified significantly leading to an optimal control law which is easy to interpret, providing intuition for the solution of the general risk-aware LQ problem.

Let $\mu \geq 0$ be arbitrary but fixed. First, we may simplify the form of the Lagrangian \mathcal{L} and express it within a canonical dynamic programming framework. In this respect, we have the following straightforward, but key result.

Lemma 1 (Lagrangian Reformulation). *Assume that system (1) is fully-observed: $y_k = x_k$ for all $k \geq 0$. Let Assumption 1 be in effect. Consider the sequential variational QCQP problem (6). Consider the notation of Proposition 1. Define the inflated state penalty matrix*

$$Q_\mu \triangleq Q + 4\mu QWQ.$$

Then the innovation process $\delta_k = w_k - \bar{w}$ is i.i.d. and independent of \mathcal{F}_{k-1} with

$$\begin{aligned} W_{t-1} &= W, \quad m_{3,t-1} = m_3 \triangleq 2Q\mathbb{E} \{ \delta_t \delta_t' Q \delta_t \} \\ m_{4,t-1} &= m_4 \triangleq \mathbb{E} \{ (\delta_t' Q \delta_t - \text{tr}(QW))^2 \}. \end{aligned}$$

Moreover, for every $u_t \in \mathcal{L}_4(\mathcal{F}_t)$, $t \leq N-1$, the Lagrangian function \mathcal{L} can be expressed as

$$\mathcal{L}(u, \mu) = \mathbb{E} \left\{ \sum_{t=1}^N g_t(x_t, u_{t-1}, \mu) \right\} + g(\mu), \quad (12)$$

where

$$\begin{aligned} g_t(x_t, u_{t-1}, \mu) &\triangleq x_t' Q_\mu x_t + 2\mu m_3' x_t + u_{t-1}' R u_{t-1}, \quad t \leq N \\ g(\mu) &\triangleq \mu (-\bar{\epsilon} - 4N \text{tr}(WQ)^2) + x_0' Q x_0. \end{aligned}$$

Proof. The properties of δ_t follow immediately from (1), full observability, and the fact that w_k is i.i.d. As a result, all moments $W_t, m_{3,t}, m_{4,t}$ are deterministic and constant over time. For the Lagrangian reformulation, we used Proposition 1, the form of \mathcal{L} , and the identities

$$\begin{aligned}\mathbb{E}(\hat{x}'_k Q W Q \hat{x}_k) &= \mathbb{E}(x_k Q W Q x_k) - \mathbb{E}(\delta'_k Q W Q \delta_k) \\ \mathbb{E}(\delta'_k m_3) &= 0, \mathbb{E}(\hat{x}'_k m_3) = \mathbb{E}(x'_k m_3).\end{aligned}\quad \square$$

Remark 2 (Relation to LQR with tracking). The Lagrangian (12) has the structure of a generalized LQR problem with a tracking objective. Substituting for $m_3 = Q M_3$, where

$$M_3 \triangleq 2\mathbb{E} \{ \delta_t \delta'_t Q \delta_t \},$$

we can rewrite the stage cost as

$$\begin{aligned}g_t(x_t, u_t, \mu) &= (x_t + \mu M_3)' Q (x_t + \mu M_3) \\ &\quad + x'_t (4\mu Q W Q) x_t + u'_t R u_t - \mu^2 M_3' Q M_3,\end{aligned}$$

i.e., the state penalty is quadratic and consists of two distinct terms. The first one, i.e., $(x_t + \mu M_3)' Q (x_t + \mu M_3)$ is a tracking error term that forces the state to be close to the static target $-\mu M_3$. Informally, in the case of skewed noise, by tracking $-\mu M_3$ we pre-compensate for directions in which the distribution of the noise has heavy tails. This decreases the statistical variability of the predicted stage cost. The second term, $x'_t (4\mu Q W Q) x_t$, is a standard quadratic penalty term; notice that, contrary to the risk-neutral case, the covariance of the noise W now affects the penalty term. Informally, this term penalizes state directions which not only lead to high cost but are also more sensitive to noise, as captured by the product $Q W Q$. Hence, the risk-neutral LQR framework can exhibit inherent risk-averse properties, provided that its parameters are selected in a principled way. Of course, selecting those parameters *a priori* is not trivial. \diamond

The structure of the Lagrangian as suggested by Lemma 1 enables us to derive both a closed-form expression for its minimum and an explicit optimal control policy. To this end, define the *optimal cost-to-go* at stage $k \leq N - 1$ as

$$\mathcal{L}_k^*(x_k, \mu) \triangleq \inf_{u_{k:N-1} \in \mathcal{U}_k} \mathbb{E} \left\{ \sum_{t=k}^{N-1} g_{t+1}(x_{t+1}, u_t, \mu) \middle| \mathcal{F}_k \right\},$$

where we omit the constant components of the Lagrangian. Under this definition, it is true that

$$D(\mu) \equiv \inf_{u \in \mathcal{U}_0} \mathcal{L}(u, \mu) = \mathcal{L}_0^*(x_0, \mu) + g(\mu).$$

We may now derive the complete solution to (9), which provides optimal risk-aware control policies for every multiplier $\mu \geq 0$.

Theorem 3 (LQR Risk-Aware Controllers). *Assume that system (1) is fully-observed: $y_k = x_k$ for all $k \geq 0$. Let Assumption 1 be in effect, choose $\mu \geq 0$, and adopt the notation of Lemma 1. For $t \leq N - 1$, the optimal cost-to-go $\mathcal{L}_t^*(x_t, \mu)$ may be expressed as*

$$\mathcal{L}_t^*(x_t, \mu) = x'_t (V_t - Q_\mu) x_t + 2(\xi_t - \mu m_3)' x_t + c_t,$$

where the quantities V_t, ξ_t , and c_t are evaluated through the backward recursions

$$V_{t-1} = A' V_t A + Q_\mu - A' V_t B (B' V_t B + R)^{-1} B' V_t A, \quad (13)$$

$$K_{t-1} = -(B'V_tB + R)^{-1}B'V_tA, \quad (14)$$

$$\xi_{t-1} = (A + BK_{t-1})'(\xi_t + V_t\bar{w}) + \mu m_3, \quad (15)$$

$$l_{t-1} = -(B'V_tB + R)^{-1}B'(\xi_t + V_t\bar{w}), \quad (16)$$

$$c_{t-1} = c_t + \text{tr}(WV_t) + 2\xi_t'\bar{w} + \bar{w}'V_t\bar{w} - l_{t-1}'(B'V_tB + R)l_{t-1}, \quad (17)$$

with terminal values $V_N = Q_\mu$, $\xi_N = \mu m_3$, and $c_N = 0$. Additionally, an optimal control policy that achieves the dual value in (9) may be expressed as

$$u_t^*(\mu) = K_t x_t + l_t \in \mathcal{L}_4(\mathcal{F}_t), \quad \forall t \leq N-1, \quad (18)$$

and is unique up to sets of probability measure zero.

Proof. The proof is similar to that of Theorem 5 in Section 6 and is, thus, omitted. The only difference is that we need to verify that the input has bounded fourth moments $u_t^*(\mu) \in \mathcal{L}_4(\mathcal{F}_t)$ under Assumption 1. This can be inferred recursively by (18) and by the fact that K_t, l_t are deterministic constants for all $t \geq 0$ (at all time steps the input is a linear combination of random variables with bounded fourth moments). \square

As suggested by Remark 2, it turns out that the optimal controller (18) is affine with respect to the state. If we expand ξ_t , we can see that the affine term l_t consists of two components:

$$l_t = -(B'V_tB + R)^{-1}B'(S_t\mu m_3 + T_t\bar{w}),$$

for some appropriate matrices S_t, T_t :

$$\begin{aligned} S_t &= (A + BK_t)'S_{t+1} + I, & S_N &= I \\ T_t &= (A + BK_t)'(T_{t+1} + V_{t+1}), & T_N &= 0. \end{aligned}$$

One component forces the state to track the reference $-\mu m_3$, which points away from heavy-tailed regions of the noise distribution. The other component acts against the mean value of the noise—such a term also appears in risk-neutral LQR. Meanwhile, the state-feedback term accounts for the internal dynamics. Similar to the risk-neutral case, the controller's behavior is governed by a Riccati difference equation (13). However, we now have an inflated stage cost matrix $Q_\mu = Q + 4\mu QWQ$, instead of the original. As suggested by the product QWQ , the risk-aware control gain becomes more strict in directions that are simultaneously more costly and prone to noise, as captured by the covariance W . As a sanity check, we can verify that for $\mu = 0$, we recover the risk-neutral LQR optimal controller, i.e. $Q_0 = Q$ and l_t depends only on the mean value of the noise \bar{w} .

Since V_t in (13) satisfies a standard Riccati difference equation with an inflated matrix Q_μ , we immediately obtain from standard LQR theory that for any $\mu \geq 0$, the optimal controller (18) will be internally stable. Matrix $A + BK_t$ will converge and its spectral radius will eventually be bounded as $\rho(A + BK_t) < 1$, as the horizon N grows to infinity. The conditions for stability remain the same as in risk-neutral LQR.

Assumption 2 (Controllability). *The pair (A, B) is stabilizable, the pair $(A, Q^{1/2})$ is detectable, matrix $Q \succeq 0$ is positive semi-definite and matrix $R \succ 0$ is positive definite.*

Corollary 1 (Internal Stability). *Let Assumptions 1 and 2 be in effect, and adopt the notation of Lemma 1. For fixed $\mu \geq 0$, consider the control policy $u^*(\mu)$, as defined in (18). As $N \rightarrow \infty$, V_t converges exponentially fast to the unique stabilizing solution¹ of the algebraic Riccati equation*

$$V = A'VA + Q_\mu - A'VB(B'VB + R)^{-1}B'VA.$$

As a result, for every $t \geq 0$, it is true that, as $N \rightarrow \infty$,

$$\begin{aligned} K_t &\rightarrow K \triangleq -(B'VB + R)^{-1}B'VA, \\ \xi_t &\rightarrow \xi \triangleq (I - (A + BK)')^{-1} \{ (A + BK)'V\bar{w} + \mu m_3 \}, \\ l_t &\rightarrow l \triangleq -(B'VB + R)^{-1}B'(\xi + V\bar{w}), \end{aligned}$$

exponentially fast, and the closed-loop matrix $A + BK$ is stable (spectral radius $\rho(A + BK) < 1$).

Proof. Since $Q_\mu \succeq Q$ and $(A, Q^{1/2})$ is detectable, the pair $(A, Q_\mu^{1/2})$ is also detectable. Since (A, B) is stabilizable, $(A, Q_\mu^{1/2})$ is detectable, and $R \succ 0$, the exponential convergence of V_t and K_t to V and K respectively, and the stability of $A + BK$ follow from standard LQR theory [48, Chapter 4]. The proof of the convergence of the remaining terms follows similar steps. \square

5.1 Recovery of Primal-Optimal Solutions

Up to now we have discussed the properties of the optimal controller given a fixed $\mu \geq 0$. In what follows, we show how to compute an optimal multiplier μ^* based on Theorems 1, 2. For any fixed $\mu \geq 0$, we provide a closed-form expression for evaluating the risk functional $J_R(u^*(\mu))$. Moreover, we show that $J_R(u^*(\cdot))$ is a continuous function of μ . Hence, if Slater's condition is satisfied, then based on Theorem 2, we can find the optimal multiplier μ^* by trivially applying bisection on μ .

The evaluation of the risk constraint functional $J_R(u^*(\mu))$ may be performed in a recursive fashion, as the following result suggests.

Proposition 2 (Risk Functional Evaluation). *Assume that system (1) is fully-observed: $y_k = x_k$ for all $k \geq 0$. Let Assumption 1 be in effect, and adopt the notation of Lemma 1. For fixed $\mu \geq 0$, consider the control policy $u^*(\mu)$, as defined in (18). With terminal values $P_N = 4QWQ$, $\zeta_N = m_3$, $d_N = 0$, consider the backward recursions*

$$\begin{aligned} P_{t-1} &= (A + BK_{t-1})'P_t(A + BK_{t-1}) + 4QWQ, \\ \zeta_{t-1} &= (A + BK_{t-1})'\zeta_t + m_3 \\ &\quad + (A + BK_{t-1})'P_t(Bl_{t-1} + \bar{w}) \quad \text{and} \\ d_{t-1} &= d_t + \text{tr}([P_{t-1} - 4QWQ]W) \\ &\quad + 2\zeta_t'(\bar{w} + Bl_{t-1}) + (Bl_{t-1} + \bar{w})'P_t(Bl_{t-1} + \bar{w}). \end{aligned}$$

Then, the risk constraint in problem (6) may be evaluated by

$$\begin{aligned} J_R(u^*(\mu)) &= x_0'(P_0 - 4QWQ)x_0 + 2(\zeta_0' - m_3)x_0 \\ &\quad + d_0 - \text{tr}([P_0 - 4QWQ]W) \end{aligned}$$

Moreover, the risk-functional $J_R(u^*(\cdot))$ is continuous.

¹A stabilizing solution renders $A + BK$ stable.

Proof. Omitted; it is similar to that of Proposition 3 in Section 6. To prove continuity it is sufficient to invoke invertibility of R . We don't need Assumption 3 since the expressions above are deterministic. As a result, Assumption 1 suffices. \square

Now, we can obtain the optimal solution $u^*(\mu^*)$ to the original problem (6) for fully observed systems.

Theorem 4 (Primal-Optimal Solution). *Assume that system (1) is fully-observed: $y_k = x_k$ for all $k \geq 0$. Let Assumption 1 be in effect, and adopt the notation of Lemma 1. Define the minimum feasible $\bar{\epsilon}_{\text{inf}}$:*

$$\bar{\epsilon}_{\text{inf}} \triangleq \inf_{u \in \mathcal{U}_0} J_R(u).$$

Then, for any $\bar{\epsilon} > \bar{\epsilon}_{\text{inf}}$, problem (6) is feasible and the optimal solution is given by $u^(\mu^*)$ based on (11), (18).*

Proof. Omitted; it is similar to the proof of Theorem 6. \square

Note that solving the problem

$$\inf_{u \in \mathcal{U}_0} J_R(u) = \inf_{u \in \mathcal{U}_0} \mathbb{E} \sum_{k=1}^N \hat{x}_k' Q W Q \hat{x}_k + 2m'_3 \hat{x}_k$$

corresponds to finding a maximally risk-aware policy. Since the risk functional is quadratic, we can solve the problem following standard LQR theory. Note that the problem is singular since there is no input penalty in J_R and the matrix QWQ could be singular. Hence, there might be multiple optimal solutions. Nonetheless, we can still obtain an admissible optimal solution so that the infimum becomes minimum; such a solution will involve pseudo-inverses instead of inverses. More information about singular LQR control can be found in [50].

The above problem characterizes the minimum value of ϵ such that (3) is feasible. Note that as we increase ϵ , we relax the risk-aware requirements. Let $\epsilon_{LQR} \triangleq J_R(u^*(0))$ be the value of the risk functional evaluated at the risk-neutral LQR optimal controller. Then, trivially if ϵ reaches a large value, i.e. larger than ϵ_{LQR} , then the risk-neutral LQR controller will become feasible. After that point, if we keep increasing ϵ , the risk-neutral LQR controller will always be the optimal solution to (3) with $\mu^* = 0$. Hence, to obtain risk-aware behaviors, we need to select $\epsilon < \epsilon_{LQR}$.

6 Optimal Risk-Aware LQ Control

In this section we study problem (6) in its full generality, when we only have access to partial state measurements. Fix a Lagrange multiplier $\mu \geq 0$ and recall the definition of Lagrangian \mathcal{L} in (8). Before we derive the optimal control law, let us simplify the form of the Lagrangian \mathcal{L} . For brevity, denote the information up to time t (extended state) by $z_t = (y_{0:t}, u_{0:t-1})$, $z_0 = y_0$. Then, we get the following result.

Lemma 2 (Lagrangian Reformulation). *Let Assumption 1 be in effect. Consider the sequential variational QCQP problem (6) and define the inflated state penalty matrix*

$$Q_{\mu,t} \triangleq Q + 4\mu QW_tQ.$$

Then, for every $u_t \in \mathcal{L}_4(\mathcal{F}_t)$, $t \leq N - 1$, the Lagrangian function \mathcal{L} can be expressed as

$$\mathcal{L}(u, \mu) = \mathbb{E} \left\{ \sum_{t=0}^{N-1} g_t(z_t, u_t, \mu) \right\} + g(\mu), \quad (19)$$

where

$$g_t(z_t, u_t, \mu) = \hat{x}'_{t+1} Q_{\mu,t} \hat{x}_{t+1} + 2\mu \hat{x}'_{t+1} m_{3,t} + u'_t R u_t$$

$$g(\mu) = -\mu \bar{e} + \mathbb{E} \sum_{t=0}^{N-1} \text{tr}(QW_t) + x'_0 Q x_0.$$

Proof. It follows from Proposition 1 and

$$\begin{aligned} \mathbb{E} x'_t Q x_t &= \mathbb{E} \{ \mathbb{E}(x'_t Q x_t | \mathcal{F}_{t-1}) \} \\ &= \mathbb{E} \hat{x}'_t Q \hat{x}_t + \mathbb{E} \{ \mathbb{E}(\delta'_t Q \delta_t | \mathcal{F}_{t-1}) \} \\ &= \mathbb{E} \hat{x}'_t Q \hat{x}_t + \mathbb{E} \text{tr}(QW_{t-1}). \end{aligned} \quad \square$$

Following the same convention as in the fully-observed case, we define the *optimal cost-to-go* at stage $k \leq N-1$ as

$$\mathcal{L}_k^*(z_k, \mu) \triangleq \inf_{u_{k:N-1} \in \mathcal{U}_k} \mathbb{E} \left\{ \sum_{t=k}^{N-1} g_t(z_t, u_t, \mu) \middle| \mathcal{F}_k \right\}.$$

where we omit the constant components of the Lagrangian.

We may now derive a closed-form solution to (10), which provides optimal risk-aware control policies for every fixed multiplier $\mu \geq 0$. The above formulation in Lemma 2 only requires the noise signals to satisfy Assumption 1. However, to guarantee that our closed-form optimal controller below is well-defined and has finite fourth moments, we need the following sufficient stricter assumption, which strengthens Assumption 1.

Assumption 3 (Strict noise regularity). *The process noise w_t has finite moments $\mathbb{E} \|w_t\|_2^p < \infty$, for any order $p \geq 1$ and any time $t \in \mathbb{N}$.*

Theorem 5 (LQ Risk-Aware Controllers). *Let Assumption 3 be in effect. Fix a multiplier $\mu \geq 0$ and adopt the notation of Lemma 2. Recall the definition of the refinement error e_t in (5). Then, the optimal cost-to-go at time t is given by:*

$$\mathcal{L}_t^*(z_t, \mu) = \hat{x}'_{t|t} P_t \hat{x}_{t|t} + 2\zeta'_t \hat{x}_{t|t} + c_t, \quad (20)$$

where P_t, ζ_t, c_t are \mathcal{F}_t -measurable, given by the recursions:

$$V_t = \mathbb{E}(P_{t+1} | \mathcal{F}_t) + Q_{\mu,t} \quad (21)$$

$$\xi_t = \mathbb{E}(P_{t+1} e_{t+1} + \zeta_{t+1} | \mathcal{F}_t) + \mu m_{3,t} \quad (22)$$

$$d_t = \mathbb{E} \{ c_{t+1} + e'_{t+1} P_{t+1} e_{t+1} + 2e'_{t+1} \zeta_{t+1} | \mathcal{F}_t \} \quad (23)$$

$$K_t = -(B' V_t B + R)^{-1} B' V_t A \quad (24)$$

$$l_t = -(B' V_t B + R)^{-1} B' (\xi_t + V_t \bar{w}) \quad (25)$$

$$P_t = (A + BK_t)' V_t (A + BK_t) + K'_t R K_t \quad (26)$$

$$\zeta_t = (A + BK_t)' (\xi_t + V_t \bar{w}) \quad (27)$$

$$c_t = d_t + \bar{w}' V_t \bar{w} - l'_t (B' V_t B + R) l_t + 2\xi'_t \bar{w} \quad (28)$$

with initial values $P_N = 0, \zeta_N = 0, c_N = 0$. Additionally, an optimal control policy that achieves the dual value in (9) may be expressed as:

$$u_t^*(\mu) = K_t \hat{x}_{t|t} + l_t \in \mathcal{L}_4(\mathcal{F}_t). \quad (29)$$

The proof can be found in the Appendix. Note that for the well-posedness of the solution, we only need i) the conditional moments in (21)-(23) to be well-defined, and ii) fourth-moment integrability of the optimal inputs in (29). Assumption 3 is only a sufficient condition so that the above conditions are satisfied. It might not be a necessary condition. For example, in the fully-observed case (Section 5), the milder Assumption 1 suffices. Here, the technical difficulty stems from the fact that in the general case, K_t is stochastic and potentially unbounded. We leave further discussion for future work.

Interestingly, in the partially observed case the control law is still affine-like. However, the linear and affine terms are no longer constants. They evolve based on a new filtered version of the Riccati difference equation, see (21), (26). They are random variables that depend exclusively on the stochastic dynamics (noises) of the system. The intuition behind the linear gain and the affine term is similar to the fully-observed case, however, there is a major difference. Instead of accounting only for the process noise w_k , we account for the whole prediction error $x_k - \hat{x}_k$, which captures also the estimation uncertainty. The control policy and the estimation process are intertwined, in the sense that the latter affects the gain and the affine part. Hence, the control policy cannot be designed independently of the estimation process/noise statistics. In other words, the certainty equivalence property (see [22]) does not hold. However, separation holds weakly in the sense that the optimal state estimator $\hat{x}_{t|t}$ in (29) is the minimum mean-square error (mmse) estimator and can be designed separately from the optimal controller.

6.1 Recovery of Primal-Optimal Solutions

In this subsection, we provide a closed-form expression to evaluate the risk functional $J_R(u^*(\mu))$. Moreover, we show that $J_R(u^*(\cdot))$ is a continuous function of μ . Similar to the fully-observed case, if Slater's condition is satisfied, then we can find the optimal multiplier μ^* by trivially applying bisection.

Proposition 3 (Risk Functional Evaluation). *Let Assumption 3 be in effect, and adopt the notation of Lemma 2. Recall the definition of the refinement error e_t in (5). For fixed $\mu \geq 0$, consider the control policy $u^*(\mu)$, as defined in (29). With terminal values $H_{N-1} = 4QW_{N-1}Q$, $f_{N-1} = m_{3,N-1}$, $g_N = 0$, consider the backward recursions*

$$\begin{aligned}\Theta_t &= (A + BK_t)'H_t(A + BK_t) \\ H_{t-1} &= \mathbb{E}(\Theta_t|\mathcal{F}_{t-1}) + 4QW_{t-1}Q \\ \eta_t &= (A + BK_t)'(f_t + H_t(Bl_t + \bar{w})) \\ f_{t-1} &= \mathbb{E}(\eta_t + \Theta_t e_t|\mathcal{F}_{t-1}) + m_{3,t-1} \\ \gamma_t &= g_t + (Bl_t + \bar{w})'H_t(Bl_t + \bar{w}) + 2(Bl_t + \bar{w})'f_t \\ g_{t-1} &= \mathbb{E}(\gamma_t + e_t'\Theta_t e_t + 2e_t'\eta_t|\mathcal{F}_{t-1})\end{aligned}$$

Then, the risk constraint in problem (6) may be evaluated by

$$J_R(u^*(\mu)) = x_0'\mathbb{E}(\Theta_0)x_0 + 2\mathbb{E}(\eta_0')x_0 + \mathbb{E}(\gamma_0).$$

Moreover, the risk functional $J_R(u^*(\mu))$ is a continuous function of μ .

Finally, for completeness we state the following theorem which completely characterizes the solution to the primal problem (6).

Theorem 6 (Primal-Optimal Solution). *Let Assumption 3 be in effect, and adopt the notation of Lemma 2. Define the minimum feasible \bar{c}_{inf} :*

$$\bar{c}_{\text{inf}} \triangleq \inf_{u \in \mathcal{U}_0} J_R(u).$$

Then, for any $\bar{\epsilon} > \bar{\epsilon}_{\text{inf}}$, problem (6) is feasible and the optimal solution is given by $u^*(\mu^*)$ based on (11), (29).

Proof. It follows from Proposition 3 that $J_R(u^*(\cdot))$ is continuous. Hence, if Slater's condition is satisfied, optimality follows from Theorems 1, 2. We only need to show that Slater's condition is satisfied under $\bar{\epsilon} > \bar{\epsilon}_{\text{inf}}$. This follows from the definition of infimum; there exists $u^\dagger \in \mathcal{U}_0$ such that $J_R(u^\dagger) \leq \bar{\epsilon}_{\text{inf}} + (\bar{\epsilon} - \bar{\epsilon}_{\text{inf}})/2 < \bar{\epsilon}$. \square

Although the recursions of Theorem 5 and Proposition 3 provide a closed-form solution, they require knowledge of several conditional moments. In reality, these conditional moments might be hard to track. Another side effect of the stochasticity of K_t is that the result of Corollary 1 will no longer apply for general partially-observed systems; the gains will not converge pointwise in general. As a consequence, any stability analysis of (29) might require further assumptions, beyond the scope of this paper. Nonetheless, in the case of Gaussian noise we can solve both challenges; we can compute these moments exactly and we can prove stability under certain controllability/observability conditions.

7 Optimal Risk-Aware LQG Control

In the special case of Gaussian measurement and process noise, the innovation error δ_t is actually independent of the past \mathcal{F}_{t-1} . Therefore, the moments defined in Proposition 1 are deterministic, and the recursive formulas for the control policy and the risk-evaluation can be simplified dramatically. In this section, we focus on exactly this case and provide explicit formulas for optimal risk-aware LQG controllers.

Assumption 4 (Gaussian Noise). *The process noise and measurement noise w_k, v_k are jointly i.i.d. Gaussian with mean \bar{w} , 0 respectively and covariance*

$$\mathbb{E} \begin{bmatrix} w_k \\ v_k \end{bmatrix} \begin{bmatrix} w_k \\ v_k \end{bmatrix}' = \begin{bmatrix} W & 0 \\ 0 & S \end{bmatrix}.$$

Theorem 7 (LQG Risk-Aware Controllers). *Let Assumption 4 be in effect. Fix a multiplier $\mu \geq 0$ and consider the notation of Theorem 5. The innovation sequence $\delta_t, t \geq 0$ is Gaussian and independent with covariance given by the forward recursion (Kalman Filter)*

$$W_{t+1} = AW_tA' + W - AW_tC'(CW_tC' + S)^{-1}CW_tA'$$

for $W_0 = 0$. The mean square estimate $\hat{x}_{t|t}$ is given by

$$\hat{x}_{t|t} = \hat{x}_t + W_tC'(CW_tC' + S)^{-1}(y_t - \hat{x}_t).$$

Consider the backward recursion

$$V_{t-1} = (A + BK_t)'V_t(A + BK_t) + K_t'RK_t + Q_{\mu,t-1} \quad (30)$$

$$\xi_{t-1} = (A + BK_t)'(\xi_t + V_t\bar{w}) \quad (31)$$

$$K_t = -(B'V_tB + R)^{-1}B'V_tA \quad (32)$$

$$l_t = -(B'V_tB + R)^{-1}B'\xi_t \quad (33)$$

with initial values $V_{N-1} = Q_{\mu,N-1}$, $\xi_{N-1} = 0$. An optimal control policy that achieves the dual value in (9) may be expressed as

$$u_t^*(\mu) = K_t\hat{x}_{t|t} + l_t. \quad (34)$$

Proof. The properties of δ_t and the recursions for $W_t, \hat{x}_{t|t}$ follow from standard Kalman Filter theory [48]. We also have $m_{3,t} = 0$ since the variables δ_t are Gaussian and, thus, symmetric. Since both $W_t, m_{3,t}$ are deterministic, matrices V_t, P_t, ξ_t, ζ_t in the statement of Theorem 5 are also deterministic. Hence, we have $\mathbb{E}(P_t e_t | \mathcal{F}_t) = 0$ and we can remove the conditional expectations from V_t, ξ_t as

$$V_t = P_{t+1} + Q_{\mu,t}, \quad \xi_t = \zeta_{t+1},$$

and the result now follows from Theorem 5. \square

Unlike the fully-observed case, the inflated matrix takes into account not only the instantaneous process error W but the whole prediction error. In other words, we account also for the uncertainty in the prediction due to partial observability.

The aforementioned property also differentiates our controller from classical risk-neutral Linear Quadratic Gaussian (LQG) control. The state penalties $Q_{\mu,t}$ are inflated, time-varying, and they depend on the filtering process W_t itself, whereas, in risk-neutral LQG, the control design is completely independent of the noise statistics. Hence, we obtain a novel family of risk-aware LQG policies, which include the classical LQG ($\mu = 0$) as a special case.

In the case of Gaussian noise the third moment $m_{3,t} = 0$ is always zero due to symmetry. Contrary to the non-Gaussian noise case, the affine term only accounts for the mean value of the noise. As a result, risk-aware behavior is mainly imposed through the gain K_t .

Finally, we prove a stability result for the closed-loop system under certain observability conditions.

Assumption 5 (Observability). *The pair (A, C) is detectable, the pair $(A, W^{1/2})$ is stabilizable, and the covariance of the measurement noise is strictly positive definite $S \succ 0$.*

To prove stability let us assume that we start estimating/controlling the system at some arbitrary time $t_0 < N$ instead of 0, with x_{t_0} deterministic and known. Based on this, all recursions in the statement of Theorem 7 are extended to hold for any $t = t_0, \dots, N$, with $W_{t_0} = 0, \hat{x}_{t_0|t_0} = x_{t_0}$. We will prove that stability is achieved as we let the initial state t_0 and the horizon N go to $-\infty$ and $+\infty$ respectively.

To simplify the proof, we also assume that the state penalty is strictly positive definite $Q \succ 0$. The proof can be extended to the case $Q \succeq 0$ at the cost of more complicated arguments, but we omit it in this paper-see proof in the Appendix for discussion.

Theorem 8 (Stability). *Consider the forward and backward recursions of Theorem 7 extended to the interval $t \in \{t_0, \dots, N - 1\}$, with initial conditions $W_{t_0} = 0, \hat{x}_{t_0|t_0} = 0$ and terminal conditions as is. Let Assumptions 2, 4, 5 be in effect. Let V, W_∞ be the stabilizing solutions to the Discrete Algebraic Riccati Equations (DARE):*

$$\begin{aligned} W_\infty &= AW_\infty A' + W - AW_\infty C' (CW_\infty C + S)^{-1} CW_\infty A' \\ V &= A'VA + Q_{\mu,\infty} - A'VB(B'VB + R)^{-1}B'VA, \end{aligned}$$

with $Q_{\mu,\infty} = Q + \mu Q W_\infty Q$. Let $K \triangleq -(B'VB + R)^{-1}B'VA$ be the respective control gain, with $\bar{A} \triangleq A + BK$ the closed-loop matrix. Then, the closed-loop matrix \bar{A} is stable and

$$\|V_t - V\|_2 \leq C_1 \|\bar{A}^{N-t-1}\| + C_2 \|W_t - W_\infty\|_2, \quad (35)$$

where C_1, C_2 are some positive constants that depend on the system parameters and V . As a result, if we let $N \rightarrow \infty, t_0 \rightarrow -\infty$:

$$V_t \rightarrow V, \quad K_t \rightarrow K, \quad W_t \rightarrow W$$

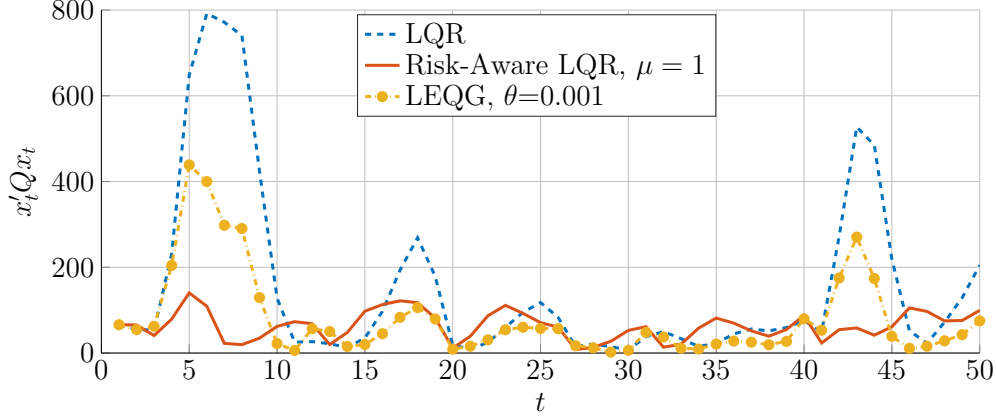


Figure 2: Evolution of the state penalties $x'_k Q x_k$, over the first 50 steps. Notice that our risk-aware LQR controller indeed limits the variability of $x'_k Q x_k$. In fact, it sacrifices performance under small wind forces, but protects the system against large wind gusts, for example at time 5 – 10.

$$\begin{aligned}\xi_t &\rightarrow \xi \triangleq (I - (A + BK)')^{-1}(A + BK)'V\bar{w} \\ l_t &\rightarrow -(B'VB + R)^{-1}B'\xi,\end{aligned}$$

exponentially fast.

The intuition behind sending t_0 and N to $-\infty, \infty$ is the following. This corresponds to a doubly infinite horizon problem, where the estimation process has started infinitely long ago in the past and the control process is running continuously for an infinite amount of time. At first sight, the result seems to be equivalent to proving stability of the classical risk-neutral LQG controller. However, by a more careful examination, equation (35) is different from classical LQG. The reason is that the second term $\mathcal{C}_2 \|W_t - W_\infty\|_2$ shows up in the error only in the case of risk-aware LQG. While estimation and control are designed independently in risk-neutral LQG, in the case of risk-aware LQG, the estimation procedure affects the convergence of the controller to its steady-state.

Similar to LEQG control for partially observed systems, our risk-aware controller regulates the state more strictly. However, this is achieved via a different mechanism, that is, via the inflation of the state penalty term. As a result, in our formulation, stability is guaranteed for any choice of μ in (34). This is unlike LEQG control, which might be unstable if we do not tune the exponential parameter θ carefully [22]. Notice also that the optimal estimator for our risk-aware LQG controller in (34) is the minimum mean-square estimator. This is different from LEQG control, where the estimator is a biased version of the minimum mean-square estimator.

8 Simulations And Discussion

Consider a flying robot that moves on a horizontal plane, i.e., the Euclidean space \mathbb{R}^2 . We assume that its linearized dynamics can be abstracted by a double integrator as

$$x_{k+1} = \begin{bmatrix} 1 & T_s & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & T_s \\ 0 & 0 & 0 & 1 \end{bmatrix} x_k + \begin{bmatrix} \frac{T_s^2}{2} & 0 \\ T_s & 0 \\ 0 & \frac{T_s^2}{2} \\ 0 & T_s \end{bmatrix} (\eta_k + d_k),$$

where $T_s = 0.5$ is the sampling time, $x_{k,1}$, $x_{k,3}$ are the position coordinates, $x_{k,2}$, $x_{k,4}$ the respective velocities and η_k is the acceleration input. Let d_k be a wind disturbance force that acts on the robot, which is modeled as follows: We assume that $d_{k,1}$ constitutes the dominant wind direction with non-zero mean and large variability, while the orthogonal direction $d_{k,2}$ is a weak wind direction with zero mean and small variability. We model $d_{k,1}$ as a mixture of two Gaussians $\mathcal{N}(30, 30)$, $\mathcal{N}(80, 60)$ with weights 0.8 and 0.2, respectively. This bimodal distribution models the presence of infrequent but large wind gusts. The weak direction $d_{k,2}$ is modeled as zero-mean Gaussian $\mathcal{N}(0, 5)$. If we cancel the mean of d_k by applying $\eta_k = u_k - \mathbb{E}d_k$, then the system can be re-written in terms of (1), where $w_k = B(d_k - \mathbb{E}d_k)$ is now a zero-mean disturbance $\bar{w} = 0$, and u_k is the exogenous input.

Consider now the LQR problem with parameters

$$Q = \text{diag}(1, 0.1, 2, 0.1) \quad \text{and} \quad R = I,$$

and a horizon of length $N = 5000$. We primarily compare our risk-aware LQR formulation with the classical, risk-neutral LQR via simulations. To tune our controller, we vary μ in (18) directly instead of varying ϵ . We also (heuristically) compare our controller with the exponential (LEQG) method, even though the noise is not Gaussian, by plugging in the second order statistics W . Let the tuning parameter of LEQG be θ . Note that the exponential problem is well defined only if $\theta < 0.001276$ (roughly), where the “neurotic breakdown” occurs [13]. For the purpose of comparison, we simulate all schemes under the same noise sequence $w_{0:N}$.

In Fig. 2, we see the evolution of the state penalty terms $x_k' Q x_k$, for the first 50 time steps, under the different control schemes. By slightly sacrificing performance under small wind forces, our risk-aware LQR controller forces the state to have less variability and protects the robot against large gusts. On the other hand, the state penalty can grow very large under the risk-neutral and LEQG schemes. This behavior is illustrated more clearly in Fig. 3, where we present the time-empirical cumulative distribution of the state penalties for all N time steps. The time-empirical “probability” of suffering large state penalties is drastically smaller compared to LQR or LEQG.

To better illustrate how the proposed risk-aware controller works, we also discuss the evolution of the position $x_{k,1}$ and the input $u_{k,1}$, as shown in Fig. 4, for the first 50 steps. First, we observe that the controller pushes the state $x_{k,1}$ towards negative values, away from the direction of the large gusts. Second, notice that we penalize $x_{k,3}$ more in Q . In fact, the risk-neutral LQR results in the steady state gains $K_{\text{LQR},11} = -0.697$, $K_{\text{LQR},12} = -1.201$, $K_{\text{LQR},23} = -0.925$, $K_{\text{LQR},24} = -1.376$, i.e., it is stricter with direction $x_{k,3}$. However, $x_{k,1}$ exhibits more variability due to the strong wind direction. In contrast, our risk-aware scheme adapts to the noise in a principled way. Due to the inflation term $\mu Q W Q$, our scheme returns the steady-state gains $K_{11} = -2.1008$, $K_{12} = -2.2132$, $K_{23} = -1.1161$, $K_{24} = -1.5131$, which means that the risky direction $x_{k,1}$ is controlled more strictly. Naturally, being more cautious with the state leads to higher control effort, as shown in Fig. 4. Lastly, although the LEQG controller is also more state-cautious,

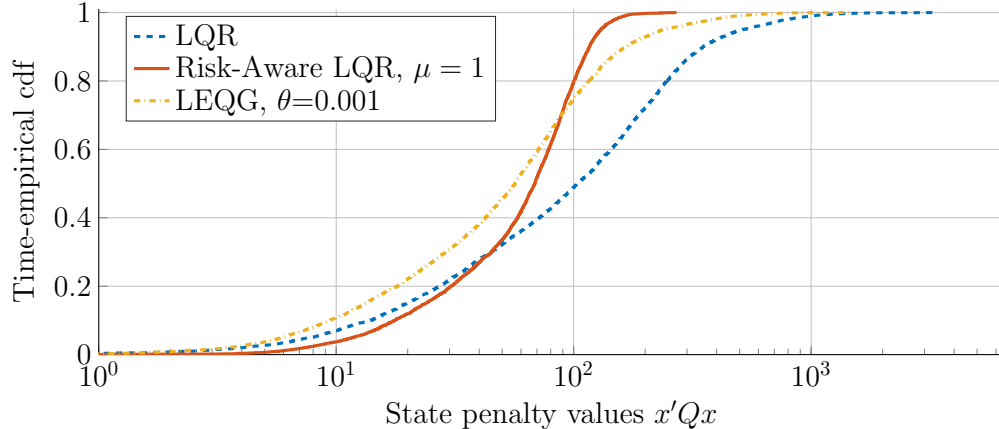


Figure 3: The time-empirical cdf for the state penalties $x'_k Q x_k$, $k \leq N$, for the LQR (risk-neutral), our method, and LEQG. Our method sacrifices some average performance but exhibits much smaller variability for the state penalties. It protects the system against rare but large wind gusts.

it is agnostic to the heavy tails of the wind distribution. Hence, it still suffers from large perturbation due to the wind gusts.

8.1 Risk-aware LQG control

In this section we evaluate the risk-aware LQG controller developed in Section 7. We use the penalty matrices

$$Q = \text{diag}(1, 0.5, 2, 0.5) \quad \text{and} \quad R = I.$$

However, the process noise is now mean-zero Gaussian, with $d_{k,1} \sim \mathcal{N}(0, 30)$ and $d_{k,2} \sim \mathcal{N}(0, 5)$. For the measurement model, we assume

$$C = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}, \quad \mathbb{E}v_k v'_k = \begin{bmatrix} 5 & 2 \\ 2 & 2 \end{bmatrix},$$

which implies that we have access to position measurements. We compare our risk-aware LQG controller with the risk-neutral LQG and the LEQG schemes. For the LEQG scheme, we used the non-delayed version [22][Th 10.5].

We simulated the system for a horizon of length $N = 3000$. The evolution of the state penalties for the first 50 time steps is shown in Fig. 5. As expected from (34), the controller regulates the state more strictly compared to the risk-neutral LQG controller, by inflating the Q matrix. Note that contrary to the fully-observed example, the noise is zero-mean Gaussian here, hence, there is no affine term in the optimal controller. We observed that the LEQG controller has similar behavior for small values of the exponential parameter θ .

A more detailed comparison is shown in Fig. 6, where the time-empirical cumulative distributions of the state penalties and the input penalties over 3000 time steps are shown. As we require our controller to be more risk-aware (we increase μ), the state penalties become smaller since the risky directions of the state are regulated more strictly. Naturally, regulating the state more strictly requires more control effort, hence the input penalties become larger. As we approach the maximally-risk aware controller ($\mu = 100$), we achieve the smallest state penalties but the largest input penalties.

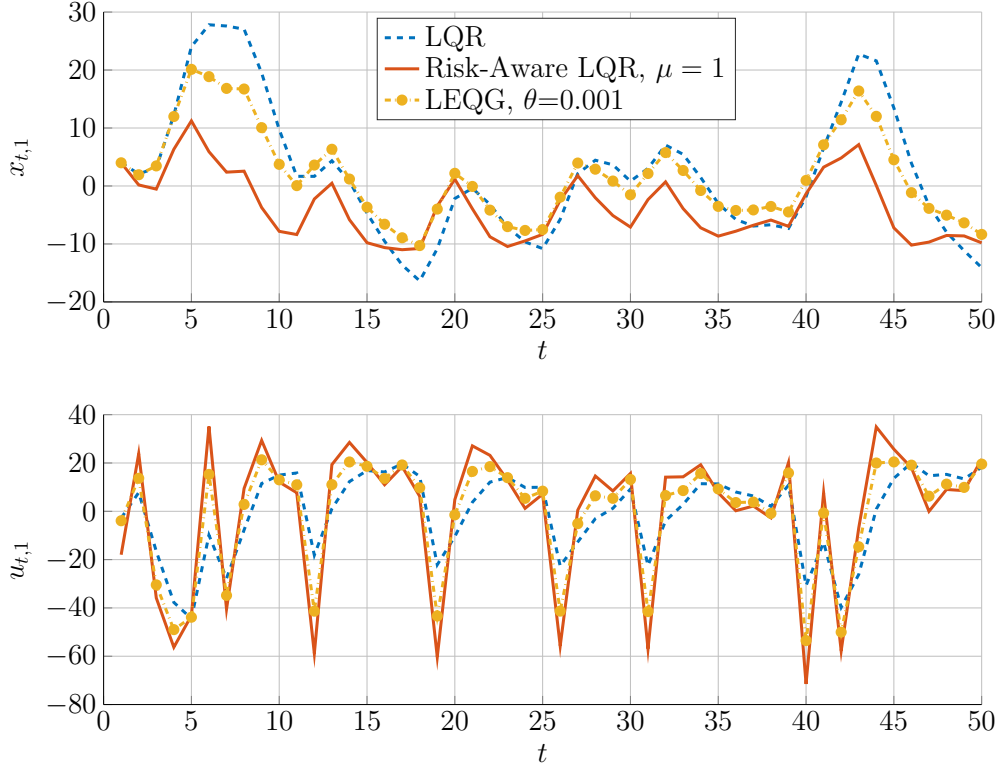


Figure 4: Evolution of the state $x_{k,1}$, and the input $u_{k,1}$ over the first 50 steps. The controller pushes the state away from the direction of the large gusts, which helps the robot to avoid extreme perturbations. Meanwhile, by inflating the state penalty with the μQWQ , we force the state-feedback component to be more cautious with the state. Naturally, being more cautious with the state requires extra control effort.

For small exponential parameters θ (below 0.006) a similar behavior is observed in the case of the LEQG controller. As we increase θ , the state is regulated more strictly at the expense of increased control effort. We achieve the smallest state penalties for roughly $\theta = 0.006$. After this value, the tradeoff between control effort and state regulation becomes worse; for example, here both the state penalties and the input penalties increase as we increase θ past 0.006. In fact, as θ approaches the “neurotic breakdown” point, e.g. for $\theta = 0.01$, both penalties become excessively large. This might be expected since the LEQG maximally risk-aware controller is very conservative, treating the noise as being adversarial rather than being stochastic, which is a different regime. On the contrary, our risk-aware LQG controller is well-behaved regardless the value of μ . Hence it is more easy to tune and offers a wider variety of tradeoff curves between control effort and state regulation. For example, if we compare the risk-aware LQG controller for $\mu = 0.5$ and the LEQG controller for $\theta = 0.006$, then the risk-aware LQG controller achieves similar state penalties with less control effort.

9 Conclusion

We studied a novel risk-aware formulation of the classical Linear Quadratic control problem, where we minimize average performance, subject to predictive variance constraints. This gives rise to risk-aware con-

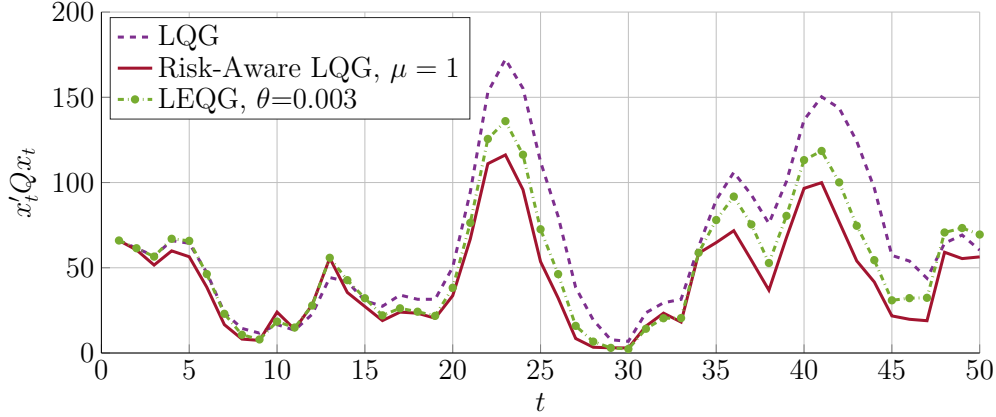


Figure 5: Evolution of the state penalties $x'_k Q x_k$, over the first 50 steps. Our risk-aware LQG controller regulates the state more strictly by using more control effort. A similar property holds for the LEQG controller. The risk-neutral LQG controller suffers from larger state perturbations.

trollers which trade between average performance and protection against uncommon but strong random disturbances. Our formulation is well-defined for general noise distributions, without requiring the existence of the respective moment generating functions. We characterized the optimal control laws for general partially-observed systems, which are affine with respect to the minimum mean-square state estimate. We provided explicit risk-aware control formulas for the special cases of i) fully-observed systems and ii) Gaussian noise. The optimal controllers are easy to tune and are internally stable under standard controllability/observability conditions.

Moving forward, there are numerous interesting research directions. First, our formulation places more emphasis on regulating the state at the cost of increased control effort. To mitigate this, we could potentially include input power constraints [51] in the quadratic formulation (6). Another open problem is explicitly computing the optimal control (29) in the case of partially-observed systems with non-Gaussian noise. Providing explicit closed-form expressions in this case is a hard problem, since it requires tracking of conditional moments. However, it might be possible to provide computational methods, which solve the problem approximately. Lastly, our predictive variance constraint is based on one-step-ahead prediction. In some cases, this might make our controller more myopic. Increasing the prediction horizon, however, might not always preserve the quadratic form of the constraint. In future work, we would also like to address this issue.

References

- [1] A. Ahlén, J. Akerberg, M. Eriksson, A. L. F. J. Isaksson, T. Iwaki, K. H. Johansson, S. Knorn, T. Lindh, and H. Sandberg, “Towards Wireless Control in Industrial Process Automation: A Case Study at a Paper Mill,” *Control Systems, IEEE*, vol. 39, no. 5, pp. 36–57, 2019.
- [2] S. Bruno, S. Ahmed, A. Shapiro, and A. Street, “Risk-Neutral and Risk-Averse Approaches to Multi-stage Renewable Investment Planning under Uncertainty,” *European Journal of Operational Research*, vol. 250, no. 3, pp. 979–989, May 2016.

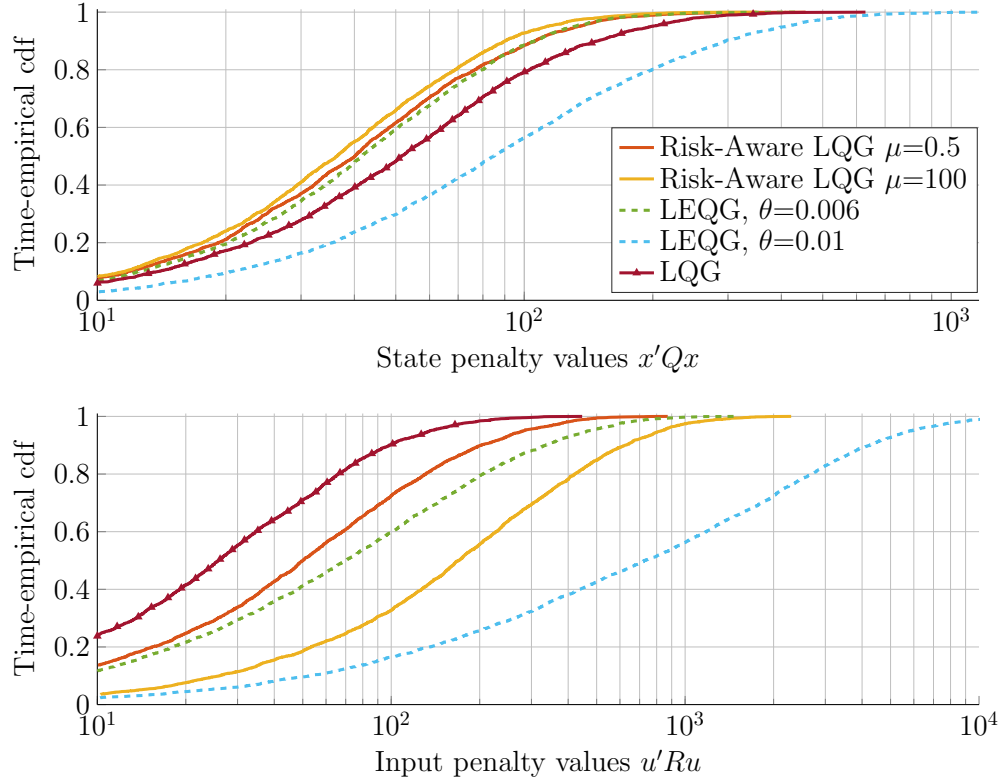


Figure 6: The time-empirical cdfs for the state and input penalties $x'_k Q x_k$, $u'_k R u_k$ under our method, the LEQG controller, and the LQG (risk-neutral) controller. For better visibility we show the probabilities (y-axis) for penalties above the value of 10 (x-axis). Our risk-aware formulation is more flexible, resulting in a wider variety of tradeoff curves between control effort and state regulation. It is more intuitive and easy to tune; as we increase μ we give more emphasis into regulating the risky directions of the state. The LEQG controller is less intuitive to tune. If θ increases too much there is a sharp decline in performance.

- [3] S. Moazeni, W. B. Powell, and A. H. Hajimiragha, “Mean-Conditional Value-at-Risk Optimal Energy Storage Operation in the Presence of Transaction Costs,” *IEEE Transactions on Power Systems*, vol. 30, no. 3, pp. 1222–1232, May 2015.
- [4] H. Markowitz, “Portfolio Selection,” *The Journal of Finance*, vol. 7, no. 1, pp. 77–91, Mar. 1952.
- [5] H. Föllmer and A. Schied, “Convex Measures of Risk and Trading Constraints,” *Finance and Stochastics*, vol. 6, no. 4, pp. 429–447, Oct. 2002.
- [6] D. Shang, V. Kuzmenko, and S. Uryasev, “Cash Flow Matching with Risks Controlled by Buffered Probability of Exceedance and Conditional Value-at-Risk,” *Annals of Operations Research*, vol. 260, no. 1-2, pp. 501–514, Jan. 2018.
- [7] S.-K. Kim, R. Thakker, and A.-A. Agha-Mohammadi, “Bi-Directional Value Learning for Risk-Aware Planning Under Uncertainty,” *IEEE Robotics and Automation Letters*, vol. 4, no. 3, pp. 2493–2500, 2019.

- [8] A. A. Pereira, J. Binney, G. A. Hollinger, and G. S. Sukhatme, “Risk-Aware Path Planning for Autonomous Underwater Vehicles using Predictive Ocean Models,” *Journal of Field Robotics*, vol. 30, no. 5, pp. 741–762, Sep. 2013.
- [9] W.-J. Ma, C. Oh, Y. Liu, D. Dentcheva, and M. M. Zavlanos, “Risk-Averse Access Point Selection in Wireless Communication Networks,” *IEEE Transactions on Control of Network Systems*, vol. 5870, no. c, pp. 1–1, 2018.
- [10] S. Samuelson and I. Yang, “Safety-Aware Optimal Control of Stochastic Systems Using Conditional Value-at-Risk,” in *2018 Annual American Control Conference (ACC)*, 2018, pp. 6285–6290.
- [11] M. P. Chapman, J. Lacotte, A. Tamar, D. Lee, K. M. Smith, V. Cheng, J. F. Fisac, S. Jha, M. Pavone, and C. J. Tomlin, “A Risk-Sensitive Finite-Time Reachability Approach for Safety of Stochastic Dynamic Systems,” in *2019 American Control Conference (ACC)*, 2019, pp. 2958–2963.
- [12] D. Jacobson, “Optimal Stochastic Linear Systems with Exponential Performance Criteria and their Relation to Deterministic Differential Games,” *IEEE Transactions on Automatic Control*, vol. 18, no. 2, pp. 124–131, 1973.
- [13] P. Whittle, “Risk-Sensitive Linear/Quadratic/Gaussian Control,” *Adv. Appl. Prob.*, vol. 13, pp. 764–777, 1981.
- [14] T. Başar, “Risk-Averse Designs: From Exponential Cost to Stochastic Games,” in *System Theory*. Springer, 2000, pp. 131–143.
- [15] K. D. Pham, *Linear-Quadratic Controls in Risk-Averse Decision Making: Performance-Measure Statistics and Control Decision Optimization*. Springer Science & Business Media, 2012.
- [16] V. Roulet, M. Fazel, S. Srinivasa, and Z. Harchaoui, “On the Convergence of the Iterative Linear Exponential Quadratic Gaussian Algorithm to Stationary Points,” 2019.
- [17] J. Speyer, C.-H. Fan, and R. Banavar, “Optimal Stochastic Estimation with Exponential Cost Criteria,” in *Proceedings of the 31st IEEE Conference on Decision and Control*. Institute of Electrical and Electronics Engineers (IEEE), Aug. 1992, pp. 2293–2298.
- [18] S. Dey and J. B. Moore, “Finite-Dimensional Risk-Sensitive Filters and Smoothers for Discrete-Time Nonlinear Systems,” *IEEE Transactions on Automatic Control*, vol. 44, no. 6, pp. 1234–1239, 1999.
- [19] J. B. Moore, R. J. Elliott, and S. Dey, “Risk-Sensitive Generalizations of Minimum Variance Estimation and Control,” *Journal of Mathematical Systems, Estimation, and Control*, vol. 7, no. 1, pp. 123–126, 1997.
- [20] S. Dey and J. B. Moore, “Risk-Sensitive Filtering and Smoothing via Reference Probability Methods,” *IEEE Transactions on Automatic Control*, vol. 42, no. 11, pp. 1587–1591, 1997.
- [21] N. Bäuerle and U. Rieder, “More Risk-Sensitive Markov Decision Processes,” *Mathematics of Operations Research*, vol. 39, no. 1, pp. 105–120, 2014.
- [22] J. L. Speyer and W. H. Chung, *Stochastic Processes, Estimation, and Control*. Siam, 2008, vol. 17.

- [23] L. A. Prashanth and M. Fu, “Risk-Sensitive Reinforcement Learning: A Constrained Optimization Viewpoint,” *arXiv preprint, arXiv:1810.09126*, Oct. 2018.
- [24] A. R. Cardoso and H. Xu, “Risk-Averse Stochastic Convex Bandit,” in *International Conference on Artificial Intelligence and Statistics*, vol. 89, Apr. 2019, pp. 39–47.
- [25] W. Huang and W. B. Haskell, “Risk-Aware Q-learning for Markov Decision Processes,” in *2017 IEEE 56th Annual Conference on Decision and Control, CDC 2017*, vol. 2018-Janua. IEEE, Dec. 2018, pp. 4928–4933.
- [26] D. R. Jiang and W. B. Powell, “Risk-Averse Approximate Dynamic Programming with Quantile-Based Risk Measures,” *Mathematics of Operations Research*, vol. 43, no. 2, pp. 554–579, Nov. 2018.
- [27] D. S. Kalogieras and W. B. Powell, “Recursive Optimization of Convex Risk Measures: Mean-Semideviation Models,” *arXiv preprint, arXiv:1804.00636*, Apr. 2018.
- [28] A. Tamar, Y. Chow, M. Ghavamzadeh, and S. Mannor, “Sequential Decision Making with Coherent Risk,” *IEEE Transactions on Automatic Control*, vol. 62, no. 7, pp. 3323–3338, Jul. 2017.
- [29] C. A. Vitt, D. Dentcheva, and H. Xiong, “Risk-Averse Classification,” *Annals of Operations Research*, Aug. 2019.
- [30] L. Zhou and P. Tokekar, “An Approximation Algorithm for Risk-averse Submodular Optimization,” *arXiv preprint, arXiv:1807.09358*, Jul. 2018.
- [31] A. Ruszczyński, “Risk-Averse Dynamic Programming for Markov Decision Processes,” *Mathematical Programming*, vol. 125, no. 2, pp. 235–261, Oct. 2010.
- [32] P. Sopasakis, D. Herceg, A. Bemporad, and P. Patrinos, “Risk-Averse Model Predictive Control,” *Automatica*, vol. 100, pp. 281 – 288, 2019.
- [33] D. S. Kalogieras, “Noisy Linear Convergence of Stochastic Gradient Descent for CV@R Statistical Learning under Polyak-Lojasiewicz conditions,” *arXiv preprint arXiv:2012.07785*, 2020.
- [34] A. Shapiro, D. Dentcheva, and A. Ruszczyński, *Lectures on Stochastic Programming: Modeling and Theory*, 2nd ed. Society for Industrial and Applied Mathematics, 2014.
- [35] R. T. Rockafellar and S. Uryasev, “Optimization of Conditional Value-at-Risk,” *Journal of Risk*, vol. 2, pp. 21–41, 1997.
- [36] M. P. Chapman and L. Lessard, “Toward a Scalable Upper Bound for a CVaR-LQ Problem,” *IEEE Control Systems Letters*, vol. 6, pp. 920–925, 2022.
- [37] K. Zhou, J. Doyle, and K. Glover, *Robust and Optimal Control*. Prentice Hall, 1996.
- [38] I. Tzortzis, C. D. Charalambous, T. Charalambous, C. K. Kourtellaris, and C. N. Hadjicostis, “Robust Linear Quadratic Regulator for Uncertain Systems,” in *2016 IEEE 55th Conference on Decision and Control (CDC)*, 2016, pp. 1515–1520.
- [39] S. Dean, H. Mania, N. Matni, B. Recht, and S. Tu, “On the Sample Complexity of the Linear Quadratic Regulator,” *Foundations of Computational Mathematics*, vol. 20, no. 4, pp. 633–679, 2020.

- [40] K. Glover and J. C. Doyle, “State-Space Formulae for All Stabilizing Controllers that Satisfy an H_∞ -norm Bound and Relations to Risk Sensitivity,” *Systems & control letters*, vol. 11, no. 3, pp. 167–172, 1988.
- [41] K. Zhang, X. Zhang, B. Hu, and T. Başar, “Derivative-Free Policy Optimization for Risk-Sensitive and Robust Control Design: Implicit Regularization and Sample Complexity,” *arXiv preprint arXiv:2101.01041*, 2021.
- [42] G. Goel and B. Hassibi, “Regret-optimal measurement-feedback control,” in *Learning for Dynamics and Control*. PMLR, 2021, pp. 1270–1280.
- [43] A. Tsiamis, D. S. Kalogerias, L. F. Chamon, A. Ribeiro, and G. J. Pappas, “Risk-Constrained Linear-Quadratic Regulators,” in *59th IEEE Conference on Decision and Control (CDC)*, 2020, pp. 3040–3047.
- [44] F. Zhao, K. You, and T. Basar, “Infinite-horizon Risk-constrained Linear Quadratic Regulator with Average Cost,” *arXiv preprint arXiv:2103.15363*, 2021.
- [45] F. Zhao, K. You, and T. Başar, “Global Convergence of Policy Gradient Primal-dual Methods for Risk-constrained LQRs,” *arXiv preprint arXiv:2104.04901*, 2021.
- [46] M. Abeille, A. Lazaric, X. Brokmann *et al.*, “LQG for Portfolio Optimization,” *Available at SSRN: <https://ssrn.com/abstract=2863925>*, 2016.
- [47] D. P. Bertsekas, *Dynamic Programming and Optimal Control*, 4th ed. Athena Scientific, 2017, vol. 1.
- [48] B. Anderson and J. Moore, *Optimal Filtering*. Dover Publications, 2005.
- [49] A. Ruszczyński, *Nonlinear Optimization*. Princeton university press, 2011.
- [50] F. Lewis, “A Generalized Inverse Solution to the Discrete-Time Singular Riccati Equation,” *IEEE Transactions on Automatic Control*, vol. 26, no. 2, pp. 395–398, 1981.
- [51] A. Gattami, “Generalized Linear Quadratic Control,” *IEEE Transactions on Automatic Control*, vol. 55, no. 1, pp. 131–136, 2009.
- [52] D. P. Bertsekas, *Dynamic Programming and Optimal Control*, 4th ed. Athena Scientific, 2012, vol. 2: Approximate Dynamic Programming.

Proof of Theorem 2

To prove part 1), let $\mu_2 > \mu_1 \geq 0$. From the definition of the Lagrangian and optimality of the controller $u^*(\mu)$, we obtain the inequalities

$$\begin{aligned} J(u^*(\mu_1)) + \mu_1 J_R(u^*(\mu_1)) &\leq J(u^*(\mu_2)) + \mu_1 J_R(u^*(\mu_2)) \\ J(u^*(\mu_1)) + \mu_2 J_R(u^*(\mu_1)) &\geq J(u^*(\mu_2)) + \mu_2 J_R(u^*(\mu_2)). \end{aligned}$$

By subtracting, we get

$$(\mu_2 - \mu_1) \{J_R(u^*(\mu_1)) - J_R(u^*(\mu_2))\} \geq 0,$$

which shows that $J_R(u^*(\mu_1)) \geq J_R(u^*(\mu_2))$. The proof of $J(u^*(\mu_1)) \leq J(u^*(\mu_2))$ is similar.

To prove part 2), we first show that, whenever $\mu^* < \infty$, $\mu^*(J_R(u^*(\mu^*)) - \bar{\epsilon}) = 0$, i.e., complementary slackness holds. We have two cases: either $\mu^* = 0$, where complementary slackness is satisfied trivially; or $\mu^* > 0$, $J_R(u^*(\mu^*)) \leq \bar{\epsilon}$. Therefore, it will be sufficient to show that in the latter case we can only have $J_R(u^*(\mu^*)) = \bar{\epsilon}$. Since $\mu^* > 0$, it is true that $J_R(u^*(0)) > \bar{\epsilon}$. Now, assume that $J_R(u^*(\mu^*)) < \bar{\epsilon}$. Then by the assumption of continuity of $J_R(u^*(\mu))$, there exists a $0 < \bar{\mu} < \mu^*$ such that $J_R(u^*(\bar{\mu})) = \bar{\epsilon}$, contradicting the definition of μ^* . Hence, we can only have $J_R(u^*(\mu^*)) = \bar{\epsilon}$, which shows that complementary slackness is satisfied.

Now, complementary slackness, along with the trivial fact that $J_R(u^*(\mu^*)) \leq \bar{\epsilon}$ imply that the policy-multiplier pair $(u^*(\mu^*), \mu^*) \in \mathcal{U}_0 \times \mathbb{R}_+$ satisfies the sufficient conditions for optimality provided by Theorem 1. Enough said.

To prove the last claim of part 2), suppose that (6) satisfies Slater's condition. For every $\mu \geq 0$, we have

$$\begin{aligned} D(\mu) &\leq J(u^\dagger) + \mu(J_R(u^\dagger) - \bar{\epsilon}) \\ \implies D(\mu) - \mu(J_R(u^\dagger) - \bar{\epsilon}) &\leq J(u^\dagger) < \infty. \end{aligned}$$

Next, suppose that, for every $\mu \geq 0$, $J_R(u^*(\mu)) - \bar{\epsilon} \geq 0$. Because $J(u^*(\cdot))$ is increasing on \mathbb{R}_+ , it must be true that

$$\begin{aligned} J(u^\dagger) &\geq \sup_{\mu \geq 0} D(\mu) - \mu(J_R(u^\dagger) - \bar{\epsilon}) \\ &= \sup_{\mu \geq 0} J(u^*(\mu)) + \mu(J_R(u^*(\mu)) - \bar{\epsilon}) - \mu(J_R(u^\dagger) - \bar{\epsilon}) \\ &= \infty, \end{aligned}$$

which contradicts the fact that $J(u^\dagger) < \infty$. Therefore, there must exist $\mu^\dagger \geq 0$, such that $J_R(u^*(\mu^\dagger)) - \bar{\epsilon} < 0$. But $J_R(u^*(\cdot))$ is decreasing on \mathbb{R}_+ and, consequently, it must be the case that $\mu^* \in [0, \mu^\dagger)$. The proof is now complete. \square

Proof of Theorem 5

The quadratic and linear penalties $Q_{\mu,t}$, $m_{3,t}$, the errors δ_t , ϵ_t , and all parameters V_t , ξ_t , d_t , K_t , l_t , P_t , ζ_t , c_t are a function of the stochastic dynamics of the system. To see why this holds, define the stochastic part of the system as:

$$x_k^s \triangleq x_k - \sum_{i=0}^k BA^i u_{k-i} = Ax_{k-1}^s + w_k, \quad y_k^s \triangleq Cx_k^s + v_k,$$

with $x_0^s = x_0$. We can define the deterministic part of the system as:

$$x_t^d \triangleq \sum_{i=0}^t BA^i u_{t-i} = x_t - x_t^s.$$

Since the deterministic part $x_t^d \in \mathcal{F}_{t-1}$ is measurable with respect to the current information:

$$\mathbb{E}(x_t^d | \mathcal{F}_{t-1}) = x_t^d.$$

As a result,

$$\delta_t = x_t^s - \mathbb{E}(x_t^s | \mathcal{F}_{t-1}), \quad e_t = \mathbb{E}(x_t^s | \mathcal{F}_t) - \mathbb{E}(x_t^s | \mathcal{F}_{t-1}).$$

Let now $u^1, u^2 \in \mathcal{U}_0$ be two arbitrary policies. Denote by $x_t^{s,1}, x_t^{s,2}$ the respective stochastic component of the state. Then for any sample of the probability space (everywhere):

$$x_t^{s,1} = x_t^{s,2}.$$

In other words, the stochastic states are point-wise independent of the inputs. The same holds for the quadratic and linear penalties $Q_{\mu,t}, m_{3,t}$, the errors δ_t, ϵ_t , and all parameters $V_t, \xi_t, d_t, K_t, l_t, P_t, \zeta_t, c_t$. Then, this recursively implies that pointwise everywhere:

$$\frac{\partial V_t}{\partial u_t} = 0, \quad \frac{\partial \xi_t}{\partial u_t} = 0, \quad \frac{\partial d_t}{\partial u_t} = 0.$$

Meanwhile, since all moments of w_t exist and $R \succ 0$ is strictly positive definite, all moments of $V_t, \xi_t, d_t, K_t, l_t, P_t, \zeta_t, c_t$ also exist (follows from Hölder's inequality). In particular, all moments of K_t, l_t exist.

By using dynamic programming and assuming (temporarily) that involved measurability issues are resolved [52], we have, for every $k \leq N - 1$, the recursive optimality condition (i.e., the Bellman equation)

$$\mathcal{L}_k^*(z_k, \mu) = \inf_{u_k} g_{k+1}(z_k, u_k, \mu) + \mathbb{E} \{ \mathcal{L}_{k+1}^*(z_{k+1}, \mu) | \mathcal{F}_k \}.$$

The base case is obvious since $\mathcal{L}_N^*(z_N, \mu) = 0$. Assume it is true for $k = t + 1$, we will show that the same holds for $k = t$. Writing $\hat{x}_{t+1|t+1} = \hat{x}_{t+1} + e_{t+1}$, we have:

$$\begin{aligned} & \mathbb{E} \{ \mathcal{L}_{t+1}^*(z_{t+1}, \mu) | \mathcal{F}_t \} \\ &= \mathbb{E} \left\{ \hat{x}'_{t+1|t+1} P_{t+1} \hat{x}'_{t+1|t+1} + 2\zeta'_{t+1} \hat{x}_{t+1|t+1} + c_{t+1} | \mathcal{F}_t \right\} \\ &= \hat{x}'_{t+1} \mathbb{E}(P_{t+1} | \mathcal{F}_t) \hat{x}_{t+1} + 2\mathbb{E}(e'_{t+1} P_{t+1} + \zeta'_{t+1} | \mathcal{F}_t) \hat{x}_{t+1} \\ &+ \mathbb{E}(c_{t+1} | \mathcal{F}_t) + \mathbb{E}(e'_{t+1} P_{t+1} e_{t+1} | \mathcal{F}_t) + 2\mathbb{E}(\zeta'_{t+1} e_{t+1} | \mathcal{F}_t). \end{aligned}$$

As a result, we obtain the following quadratic form

$$\begin{aligned} & g_t(z_t, u_t, \mu) + \mathbb{E} \{ \mathcal{L}_{t+1}^*(z_{t+1}, \mu) | \mathcal{F}_t \} \\ &= \hat{x}'_{t+1} V_t \hat{x}_{t+1} + u_t' R u_t + 2\xi_t' \hat{x}_{t+1} + d_t. \end{aligned} \tag{36}$$

Based on the above we can also verify that all measurability issues are now resolved in a recursive way, retrospectively. Recall that $\hat{x}_{t+1} = A\hat{x}_{t|t} + Bu_t + \bar{w}$. Since all V_t, ξ_t, d_t are independent of u_t pointwise, the unique stationary point of the above convex quadratic form is given by:

$$u_t^* = K_t \hat{x}_{t|t} + l_t.$$

Plugging the optimal input (29) into (36) gives the optimal cost-to-go (20). In detail:

$$\begin{aligned} & \{A\hat{x}_{t|t} + Bu_t^* + \bar{w}\}' V_t \{A\hat{x}_{t|t} + Bu_t^* + \bar{w}\} \\ &+ (u_t^*)' R u_t^* + 2\xi_t' \{A\hat{x}_{t|t} + Bu_t^* + \bar{w}\} + d_t \\ &= \hat{x}'_{t|t} P_t \hat{x}_{t|t} + 2 \{Bl_t + \bar{w}\}' V_t (A + BK_t) \hat{x}_{t|t} \\ &+ \{Bl_t + \bar{w}\}' V_t \{Bl_t + \bar{w}\} + 2l_t' R K_t \hat{x}_{t|t} + l_t' R l_t \\ &+ 2\xi_t' (A + BK_t) \hat{x}_{t|t} + 2\xi_t' (Bl_t + \bar{w}) + d_t \\ &\stackrel{i)}{=} \hat{x}'_{t|t} P_t \hat{x}_{t|t} + 2\bar{w}' V_t (A + BK_t) \hat{x}_{t|t} \end{aligned}$$

$$\begin{aligned}
& + 2\xi'_t(A + BK_t)\hat{x}_{t|t} + \{Bl_t + \bar{w}\}'V_t\{Bl_t + \bar{w}\} \\
& + l'_tRl_t + 2\xi'_t(Bl_t + \bar{w}) + d_t \\
& \stackrel{ii)}{=} \hat{x}'_{t|t}P_t\hat{x}_{t|t} + 2\bar{w}'V_t(A + BK_t)\hat{x}_{t|t} + 2\xi'_t(A + BK_t)\hat{x}_{t|t} \\
& + \bar{w}'V_t\bar{w} - l'_t(B'V_tB + R)l_t + 2\xi'_t\bar{w} + d_t
\end{aligned}$$

where the cancellations in *i*) follow from the identity:

$$(B'V_tB + R)K_t = -B'V_tA,$$

and *ii*) follows from:

$$2\xi'_tBl_t + 2\bar{w}'V_tBl_t = -2l'_t(B'V_tB + R)l_t.$$

Recursively, we can verify that the optimal control has finite fourth moments. In fact, all higher-order moments of u_t^* exist. This follows from the fact that all moments of K_t, l_t, w_t exist. \square

Proof of Proposition 3

We will only sketch the proof of continuity. The proof of the recursive expressions is omitted since it is similar to the proof of Theorem 5. Let $\mu_s \geq 0, s = 1, 2, \dots$ be a sequence such that $\mu_s \rightarrow \mu$ and let $\|\mu\|_\infty \triangleq \sup_{s \geq 1} \mu_s$. We will use Dominated Convergence Theorem (DCT) for conditional expectation to prove that $J_R(u^*(\mu_s))$ converges to $J_R(u^*(\mu))$. To emphasize the dependence on μ_s , we will use the notation $V_t(\mu_s), \dots, c_t(\mu_s)$ and $\Theta_t(\mu_s), \dots, g_t(\mu_s)$ for the quantities appearing in Theorem 5 and in the statement.

We will show that the terms $V_{N-2}(\mu_s), H_{N-2}(\mu_s)$ all converge almost surely to their respective limits. The proof for the remaining terms is similar. The idea is to show that K_{N-1} is dominated by a function which is independent of the index s . Note that $\|V_{N-1}(\mu_s)\|_2 \leq \|\mu\|_\infty \|W_{N-1}\|_2 \|Q\|_2^2 + \|Q\|_2$. Since R is invertible, we obtain:

$$\|K_{N-1}(\mu_s)\|_2 \leq O(\|W_{N-1}\|_2 + 1).$$

We interpret the notation $\alpha = O(\beta)$ as follows: there is a deterministic constant $\mathcal{C} = \mathcal{C}(\|\mu\|_\infty)$ such that $\alpha \leq \mathcal{C}\beta$ almost surely. As a result, we also obtain:

$$\|P_{N-1}(\mu_s)\|_2, \|\Theta_{N-1}(\mu_s)\|_2 \leq O(\|W_{N-1}\|_2^3 + 1).$$

Note that since all moments of δ_N exist (follows from Assumption 3), the term $\|W_{N-1}\|_2^3$ has finite expectation. Meanwhile, almost surely $P_{N-1}(\mu_s) \rightarrow P_{N-1}(\mu), \Theta_{N-1}(\mu_s) \rightarrow \Theta_{N-1}(\mu)$. Hence, by the conditional DCT, we also have:

$$\begin{aligned}
\mathbb{E}(P_{N-1}(\mu_s)|\mathcal{F}_{N-2}) & \rightarrow \mathbb{E}(P_{N-1}(\mu)|\mathcal{F}_{N-2}), \text{ a.s.} \\
\mathbb{E}(\Theta_{N-1}(\mu_s)|\mathcal{F}_{N-2}) & \rightarrow \mathbb{E}(\Theta_{N-1}(\mu)|\mathcal{F}_{N-2}), \text{ a.s.}
\end{aligned}$$

This proves that $V_{N-2}(\mu_s), H_{N-2}(\mu_s)$ converge to $V_{N-2}(\mu), H_{N-2}(\mu)$ almost surely. Meanwhile, they remain dominated by:

$$\begin{aligned}
& \|V_{N-2}(\mu_s)\|_2, \|H_{N-2}(\mu_s)\|_2 \leq \\
& O(\mathbb{E}(\|W_{N-1}\|_2^3|\mathcal{F}_{N-2}) + \|W_{N-2}\|_2 + 1).
\end{aligned}$$

Proceeding similarly, by successive use of DCT and the existence of all moments of δ_t , we can show that all matrices $V_t(\mu_s)$, $P_t(\mu_s)$, $H_t(\mu_s)$, $\Theta_t(\mu_s)$ and their (conditional) means converge to the respective limits almost surely. We omit the details to avoid repetition. The same will hold for terms $\eta_t(\mu_s)$ and $\gamma_0(\mu_s)$ and their (conditional) means. \square

Proof of Theorem 8

For convenience, we drop the index μ from $Q_{\mu,t}$. Note that from standard Kalman filter theory [48][Ch. 4.4], the covariance W_t converges exponentially fast to W_∞ as t_0 goes to $-\infty$. Moreover, since $W_{t_0} = 0$, the sequence W_t is increasing with respect to the positive semi-definite cone [48][Ch. 4.4]. Since $Q_\infty \succeq Q$, we also obtain that $(A, Q_\infty^{1/2})$ is detectable. Hence, the stabilizing solution V is well-defined and $A + BK$ is stable.

We will only prove (35). The remaining results follow using similar arguments. First we state a lemma that generalizes Problem 4.5, Section 4 in [48].

Lemma 3. *Let V_t, \bar{V}_t satisfy the Riccati Difference Equations*

$$\begin{aligned} V_{t-1} &= (A + BK_t)'V_t(A + BK_t) + Q_{t-1} + K_t'RK_t \\ \bar{V}_{t-1} &= (A + BL_t)'\bar{V}_t(A + BL_t) + \bar{Q}_{t-1} + L_t'RL_t, \end{aligned}$$

with gains

$$\begin{aligned} K_t &= -(B'V_tB + R)^{-1}B'V_tA \\ L_t &= -(B'\bar{V}_tB + R)^{-1}B'\bar{V}_tA, \end{aligned}$$

Then, their difference satisfies the identity:

$$V_{t-1} - \bar{V}_{t-1} = (A + BL_t)'(V_t - \bar{V}_t)(A + BK_t) + Q_{t-1} - \bar{Q}_{t-1}$$

Proof. Omitted; similar to the proof of Problem 4.5 in [48]. \square

For $N - 1 \geq k_1 > k_2$, define the products:

$$\Psi_{k_1:k_2}(N) \triangleq (A + BK_{k_1}) \times \cdots \times (A + BK_{k_2+1}), \quad (37)$$

with $\Psi_{k_1:k_1}(N) = I$. Applying Lemma 3 to V_t and V , we get:

$$V_t - V = \bar{A}'(V_{t+1} - V)(A + BK_{t+1}) + Q_t - Q_\infty.$$

Repeating multiple times, we obtain a fundamental identity:

$$\begin{aligned} V_t - V &= (\bar{A}')^{N-t-1}(Q_{N-1} - V)\Psi_{N-1:t}(N) \\ &\quad + \sum_{k=0}^{N-t-2} (\bar{A}')^k(Q_{t+k} - Q_\infty)\Psi_{t+k:t}(N). \end{aligned} \quad (38)$$

This almost gives us the result. What remains is to show that $\Psi_{t+k:t}(N)$ are uniformly bounded over all k, N, t . This follows from the following lemma.

Lemma 4. Let Q be positive definite. For any $k_1 \geq k_2 \geq t_0$, $N \geq t_0$ such that $N - 1 \geq k_1$:

$$\|\Psi_{k_1:k_2}(N)\|_2 \leq \sqrt{\sigma_{\min}^{-1}(Q)}\|V\|_2.$$

Proof. It is sufficient to show that:

$$V \succeq \Psi_{k_1:k_2}(N)'Q\Psi_{k_1:k_2}(N).$$

Then, we can get the result from the bound

$$\|V\|_2 \geq \sigma_{\min}(Q)\|\Psi_{k_1:k_2}(N)\|_2^2.$$

The proof will proceed in two steps. First, we show that for any k , N such that $k \geq t_0$, $N - 1 \geq k$ the stabilizing solution V overbounds V_k . Second, we use this property to upper-bound the products $\Psi_{k_1:k_2}(N)$.

Step 1. We show that $V_k \preceq V$ via induction. Recall that the sequence W_t is increasing, which implies that $Q_t \preceq Q_\infty$, for any $t \geq t_0$. For the base case, we have $V_{N-1} = Q_{N-1} \preceq Q_\infty \preceq V$. Assume that $V_t \preceq V$. Define the operator:

$$g(F) \triangleq A'FA + Q_\infty - A'FB(B'FB + R)^{-1}B'FA.$$

Then, since $Q_{t-1} \preceq Q_\infty$, we have:

$$V_{t-1} \preceq g(V_t) \stackrel{i)}{\preceq} g(V) = V.$$

The second inequality i) follows from the fact (e.g. see pages 79-80 Ch. 4.4 of [48]) that $g(\cdot)$ preserves positive semi-definite order, i.e. if $F_1 \succeq F_2$, then $g(F_1) \succeq g(F_2)$. This completes the proof of the first step.

Step 2. We will show that:

$$V_{k_2} \succeq \Psi_{k_1:k_2}(N)'Q\Psi_{k_1:k_2}(N).$$

Then, using the result of step 1 proves the desired inequality. From the definition of the Riccati Difference Equation:

$$\begin{aligned} V_{k_2} &\succeq (A + BK_{k_2+1})'V_{k_2+1}(A + BK_{k_2+1}) \\ &\succeq \cdots \succeq \Psi_{k_1:k_2}(N)'V_{k_1}\Psi_{k_1:k_2}(N)'. \end{aligned}$$

To complete the proof, note that $V_{k_1} \succeq Q_{k_1} \succeq Q$ for any $k_1 \leq N - 1$. □

Now choose:

$$\begin{aligned} \mathcal{C}_1 &= \|V\|_2^{3/2} \sqrt{\sigma_{\min}^{-1}(Q)} \\ \tilde{\mathcal{C}}_2 &= \sqrt{\sigma_{\min}^{-1}(Q)}\|V\|_2 \sum_{k=0}^{\infty} \|\bar{A}^k\|_2 \end{aligned}$$

Based on the above lemma, we obtain:

$$\|V_t - V\|_2 \leq \mathcal{C}_1 \|\bar{A}^{N-t-1}\|_2 + \tilde{\mathcal{C}}_2 \sup_{k \geq t} \|Q_k - Q_\infty\|_2.$$

Since the sequence W_t is increasing, we can replace the supremum by

$$\sup_{k \geq t} \|Q_k - Q_\infty\|_2 = \|Q_t - Q_\infty\|_2.$$

Finally, since:

$$Q_t - Q_\infty = \mu Q(W_t - W_\infty)Q,$$

it suffices to select $\mathcal{C}_2 = \mu \tilde{\mathcal{C}}_2 \|Q\|^2$. □

Note that if Q is singular, then the result of Lemma 4 no longer applies. Instead we could bound the products $V^{1/2}\Psi(k_1 : k_2)(N)$. A way to do this is to use $N - k_1$ large enough in the proof of step 2, so that V_{k_1} has the same range space as V . To avoid technicalities, we defer the proof for future work.