On the Convergence to Stationary Points of the Iterative Linear Exponential Quadratic Gaussian Algorithm

Vincent Roulet[1], Maryam Fazel[2], Siddhartha Srinivasa[3], Zaid Harchaoui[1],

[1] Department of Statistics, University of Washington, Seattle
[2] Department of Electrical and Computer Engineering, University of Washington, Seattle
[3] Paul G. Allen School of Computer Science & Engineering, University of Washington, Seattle

Abstract

We present a convergence analysis of the iterative linear exponential quadratic Gaussian algorithm from a first-order optimization viewpoint. The iterative linear exponential quadratic Gaussian algorithm is a classical method for risk-sensitive nonlinear control. We identify the objective that the algorithm actually minimizes and show how the addition of a proximal term guarantees convergence to a stationary point.

Introduction

We present a convergence analysis of the classical iterative linear quadratic exponential Gaussian controller (ILEQG) [Whittle, 1981] for finite-horizon risk-sensitive or safe nonlinear control. The ILEQG algorithm is particularly popular in robotics applications [Li and Todorov, 2007] and can be seen as a risk-sensitive or safe counterpart of the iterative linear quadratic Gaussian (ILQG) algorithm recently analyzed by Roulet et al. [2019]. We adopt here the viewpoint of the modern complexity analysis of first-order optimization algorithms.

We address the following questions: (i) what is the convergence rate to stationary point of ILEQG? (ii) how can we set the step-size to guarantee a decreasing objective along the iterations? The analysis we present here sheds light on these questions by highlighting the objective minimized by ILEQG which is a Gaussian approximation of a risk-sensitive cost around the linearized trajectory. We underscore the importance of the addition of a proximal regularization component for ILEQG to guarantee a worst-case convergence to a stationary point of the objective.

The main result of the paper is Theorem 2.5, where a sufficient decrease condition to choose the strength of the proximal regularization is given. The result also yields a complexity bound in terms of calls to a dynamic programming procedure implementable in a "differentiable programming" framework that is a computational framework equipped with an automatic differentiation software library. We illustrate the variant of the iterative regularized linear quadratic exponential Gaussian controller we recommend on simple risk-sensitive nonlinear control examples.

Related work. The linear exponential quadratic Gaussian algorithm is a fundamental algorithm for risk-sensitive or safe control [Whittle, 1981, Jacobson, 1973, Speyer et al., 1974]. The algorithm builds upon a risk-sensitive measure, a less conservative and more flexible framework than the H^∞ theory also used for robust control; see [Glover and Doyle, 1988, Hassibi et al., 1999, Helton and James, 1999] and references therein. An excellent review of the classical results in abstract dynamic programming and control theory, in particular for risk-sensitive control, can be found in [Bertsekas, 2018]. Risk-measures were analyzed as instances of the optimized certainty equivalent applied to specific utility functions; see [Ben-Tal and Teboulle, 1986, 2007] for a recent overview. Risk-averse model predictive control was also studied to account for ambiguity in the knowledge of the underlying probability distribution [Sopasakis et al., 2019].

Algorithms for nonlinear control problems are usually derived by analogy to the linear case, which is solved in linear time with respect to the horizon by dynamic programming [Bellman, 1971]. In particular, the iterative linear quadratic regulator (ILQR) and iterative linear quadratic Gaussian (ILQG) algorithms are usually informally motivated as iterative linearization algorithms [Li and Todorov, 2007]. A risk-sensitive variant with a straightforward optimization algorithm without theoreti-

cal guarantees was considered in [Farshidian and Buchli, 2015, Ponton et al., 2016].

On the first-order optimization front, optimization subproblems such as Newton or Gauss-Newton-steps were shown to be implementable by using dynamic programming in classical works [De O. Pantoja, 1988, Dunn and Bertsekas, 1989, Sideris and Bobrow, 2005]. Iterative linearized methods such as ILQR or ILQG were recently analyzed as Gauss-Newton-type algorithms and improved using proximal regularization and acceleration by extrapolation in [Roulet et al., 2019]. This work shares the same viewpoint and establishes worst-case complexity bounds for iterative linear quadratic exponential Gaussian controller (ILEQG) algorithms.

All proofs and notations are provided in the Appendix. The companion code is available at `https://github.com/vroulet/ilqc`.

1 Risk-sensitive control

Problem formulation. We consider discretized control problems stemming from continuous time settings with finite-horizon, see Appendix E for the discretization step. Those are off-line control problems used for example at each step of a model predictive control framework. We focus on the control of a trajectory of length τ composed of state variables $x_1, \dots, x_\tau \in \mathbb{R}^d$ and controlled by parameters $u_0, \dots, u_{\tau-1} \in \mathbb{R}^p$ through dynamics ψ_t perturbed by i.i.d. white noise $w_t \sim \mathcal{N}(0, \sigma^2 \mathrm{I}_q)$ such that

$$x_0 = \hat{x}_0, \qquad x_{t+1} = \psi_t(x_t, u_t, w_t), \tag{1}$$

for $t = 0, \dots, \tau - 1$, where $\hat{x}_0$ is a fixed starting point and the functions $\psi_t : \mathbb{R}^d \times \mathbb{R}^p \times \mathbb{R}^q \to \mathbb{R}^d$ are assumed to be continuously differentiable and bounded. Precise Assumptions for convergence are detailed in Sec. 2.

Optimality is measured through convex costs h_t, g_t, on the state and control variables x_t, u_t respectively, defining the objective

$$h(\bar{x}) + g(\bar{u}) = \sum_{t=1}^{\tau} h_t(x_t) + \sum_{t=0}^{\tau-1} g_t(u_t), \tag{2}$$

where $\bar{x} = (x_1; \dots; x_\tau) \in \mathbb{R}^{\tau d}$ is the trajectory, $\bar{u} = (u_0; \dots; u_{\tau-1}) \in \mathbb{R}^{\tau p}$ is the command, $h(\bar{x}) = \sum_{t=1}^{\tau} h_t(x_t)$ and $g(\bar{u}) = \sum_{t=0}^{\tau-1} g_t(u_t)$, and in the following we denote by $\bar{w} = (w_0; \dots; w_{\tau-1}) \in \mathbb{R}^{\tau q}$ the noise. For a given command $\bar{u}$, the dynamics in (1) define a probability distribution on the trajectories $\bar{x}$ that we denote $p(\bar{x}; \bar{u})$.

The standard objective consists in minimizing the expected cost $\min_{\bar{u} \in \mathbb{R}^{\tau p}} \mathbb{E}_{\bar{x} \sim p(\cdot; \bar{u})} [h(\bar{x})] + g(\bar{u})$, where $\bar{x}$ is a random variable following the model (1). We focus on risk-sensitive applications by minimizing

$$\min_{\bar{u} \in \mathbb{R}^{\tau p}} \frac{1}{\theta} \log \mathbb{E}_{\bar{x} \sim p(\cdot; \bar{u})} \left[\exp \theta h(\bar{x}) \right] + g(\bar{u}), \tag{3}$$

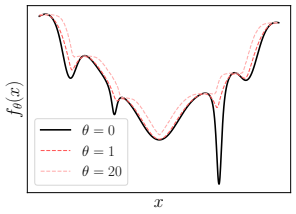

Figure 1: Effect of the risk-sensitive parameter θ for $f_\theta(x) = \frac{1}{\theta} \log \mathbb{E}_{w \sim \mathcal{N}(0,1)} \left[\exp \theta F(x+w) \right]$ with F illustrated by the black line.

for a given positive parameter $\theta > 0$. Since the dynamics are bounded, the risk-sensitive objective is well defined for any $\bar{u}$. The risk-sensitive objective (3) seeks to minimize not only the expected objective but also higher moments as can be seen by expanding it around $\theta = 0$,

$$\frac{1}{\theta} \log \mathbb{E}_{\bar{x} \sim p(\cdot; \bar{u})} \left[\exp \theta h(\bar{x}) \right] = \mathbb{E}_{\bar{x} \sim p(\cdot; \bar{u})} \left[h(\bar{x}) \right] + \frac{\theta}{2} \mathbb{V}\mathrm{ar}_{\bar{x} \sim p(\cdot; \bar{u})} \left[h(\bar{x}) \right] + \mathcal{O}(\theta^2), \tag{4}$$

which also shows that for $\theta \to 0$ we retrieve the expected cost. In Fig. 1 we illustrate the smoothness effect of the risk-sensitive objective, which, for larger values of θ, tends to select the most stable minimizers, i.e., the ones with the largest valley, see [Dvijotham et al., 2014] for a detailed discussion.

Linear Quadratic Exponential Gaussian control. The resolution of non-linear risk-sensitive control problems rest on the linear quadratic case whose properties are recalled below.

Proposition 1.1. *Consider quadratic objectives and linear dynamics defined by*

$$h_t(x_t) = \frac{1}{2} x_t^\top H_t x_t + \tilde{h}_t^\top x_t, \quad g_t(u_t) = \frac{1}{2} u_t^\top G_t u_t + \tilde{g}_t^\top u_t,$$
$$x_{t+1} = A_t x_t + B_t u_t + C_t w_t, \tag{5}$$

where $H_t \succeq 0$, $G_t \succ 0$, $w_t \sim \mathcal{N}(0, \sigma^2 \mathrm{I}_q)$ and denote by $H, \tilde{B}, \tilde{C}, \tilde{x}_0$ the matrices and vector such that for any trajectory $\bar{x}$, $H = \nabla^2 h(\bar{x})$, $\bar{x} = \tilde{B}\bar{u} + \tilde{C}\bar{w} + \tilde{x}_0$. We have that

(i) the risk sensitive control problem (3) is equivalent to

$$\min_{\bar{u} \in \mathbb{R}^{\tau p}} \sup_{\substack{\bar{w} \in \mathbb{R}^{\tau q} \\ \bar{x} \in \mathbb{R}^{\tau d}}} \sum_{t=1}^{\tau} \frac{1}{2} x_t^\top H_t x_t + \tilde{h}_t^\top x_t + \sum_{t=0}^{\tau-1} \frac{1}{2} u_t^\top G_t u_t + \tilde{g}_t^\top u_t - \sum_{t=0}^{\tau-1} \frac{1}{2\theta\sigma^2} \|w_t\|_2^2 \tag{6}$$
$$\text{subject to} \quad x_{t+1} = A_t x_t + B_t u_t + C_t w_t, \quad x_0 = \hat{x}_0,$$

(ii) if $(\theta\sigma^2)^{-1} < \lambda_{\max}(\tilde{C}^\top H \tilde{C})$ the risk-sensitive control problem is infeasible,

(iii) if $(\theta\sigma^2)^{-1} > \lambda_{\max}(\tilde{C}^\top H \tilde{C})$, the risk-sensitive control problem can be solved analytically by dynamic programming.

The resolution of the control problem by dynamic programming tracks if the problem is strongly concave in $\bar{w}$ when performing the computations, otherwise the problem is not feasible. Each cost-to-go function is indeed a quadratic whose positive-definiteness determines the feasibility of the problem. The detailed implementation is provided in Appendix B.

Iterative Linearized Quadratic Exponential Gaussian. A common method to tackle the non-linear control problem is the Iterative Linearized Quadratic Exponential Gaussian (ILEQG) algorithm, that (i) linearizes the dynamics and approximates quadratically the objectives around the current command and associated exact trajectory, (ii) solves the associated linear quadratic problem to get a descent direction, (iii) moves along the descent direction using a line-search. Formally, at a given command $\bar{u}^{(k)}$ with associated exact trajectory $\bar{x}^{(k)}$ given by $x_0^{(k)} = \hat{x}_0$, $x_{t+1}^{(k)} = \psi_t(x_t^{(k)}, u_t^{(k)}, 0)$, a descent direction is given by $\bar{v}^*$ solution, if it exists, of

$$\min_{\bar{v}\in\mathbb{R}^{\tau p}} \sup_{\substack{\bar{w}\in\mathbb{R}^{\tau p}\\ \bar{y}\in\mathbb{R}^{\tau d}}} \sum_{t=1}^{\tau}\left(\frac{1}{2}y_t^\top H_t y_t + \tilde{h}_t^\top y_t\right) \tag{7}$$
$$+\sum_{t=0}^{\tau-1}\left(\frac{1}{2}v_t^\top G_t v_t + \tilde{g}_t^\top v_t\right) - \sum_{t=0}^{\tau-1}\frac{1}{2\theta\sigma^2}\|w_t\|_2^2$$
$$\text{subject to}\quad y_{t+1} = A_t y_t + B_t v_t + C_t w_t$$
$$y_0 = 0,$$

with $H_t{=}\nabla^2 h_t(x_t^{(k)}), \tilde{h}_t{=}\nabla h_t(x_t^{(k)}), G_t{=}\nabla^2 g_t(u_t^{(k)}), \tilde{g}_t = \nabla g_t(u_t^{(k)}), A_t{=}\nabla_x \psi_t(x_t^{(k)}, u_t^{(k)}, 0)^\top, B_t{=}\nabla_u \psi_t(x_t^{(k)}, u_t^{(k)}, 0)^\top$ $C_t{=}\nabla_w \psi_t(x_t^{(k)}, u_t^{(k)}, 0)^\top$. The next command is given by

$$\bar{u}^{(k+1)} = \bar{u}^{(k)} + \gamma \bar{v}^*,$$

where γ is a step-size chosen by line-search. The complete pseudo-code is presented in Appendix C. The objective of this work is to understand the relevance of this method and to improve its implementation by answering the following questions:

1. Does ILEQG ensure the decrease of the risk-sensitive objective? If yes, what is its rate of convergence?
2. How can the step-size be chosen to ensure the monotonicity of the algorithm in a principled way?

2 Iterative linearized risk-sensitive control

2.1 Model minimization

We analyze the ILEQG method as a model-minimization scheme. To ease the exposition, we consider the case of additive noise, i.e., dynamics of the form,

$$x_0 = \hat{x}_0, \qquad x_{t+1} = \phi_t(x_t, u_t + w_t). \tag{8}$$

for bounded continuously differentiable dynamics $\phi_t : \mathbb{R}^d \times \mathbb{R}^p \to \mathbb{R}^d$. The algorithm and its interpretation can readily be extended to the general case (1). The analysis would require specific assumptions on the derivative of the dynamics w.r.t. the noise.

First, we consider the exact trajectory as a function $\tilde{x} : \mathbb{R}^{\tau p} \to \mathbb{R}^{\tau d}$ of the control variables, decomposed as $\tilde{x}(\bar{u}) = (\tilde{x}_1(\bar{u}); \dots; \tilde{x}_\tau(\bar{u}))$ where

$$\tilde{x}_1(\bar{u}) = \phi_0(\hat{x}_0, u_0), \quad \tilde{x}_{t+1}(\bar{u}) = \phi_t(\tilde{x}_t(\bar{u}), u_t), \tag{9}$$

such that the noisy trajectory is given by $\tilde{x}(\bar{u} + \bar{w})$. The risk sensitive objective (3) can then be written as

$$\min_{\bar{u}\in\mathbb{R}^{\tau p}} f_\theta(\bar{u}) = \eta_\theta(\bar{u}) + g(\bar{u}), \tag{10}$$
$$\text{with}\qquad \eta_\theta(\bar{u}) = \frac{1}{\theta}\log \mathbb{E}_{\bar{w}}\left[\exp \theta h\big(\tilde{x}(\bar{u}+\bar{w})\big)\right],$$

where, here and thereafter, $\bar{w} \sim \mathcal{N}(0, \sigma^2 \mathrm{I}_{\tau p})$ unless specified differently. Now, at a current command $\bar{u}$, for a given control deviation $\bar{v}$, the random trajectory $\tilde{x}(\bar{u}+\bar{v}+\bar{w})$ is approximated as a perturbed trajectory of $\tilde{x}(\bar{u})$, by

$$\tilde{x}(\bar{u} + \bar{v} + \bar{w}) \approx \tilde{x}(\bar{u}) + \nabla\tilde{x}(\bar{u})^\top(\bar{v} + \bar{w}). \tag{11}$$

The objective is then approximated as $f_\theta(\bar{u} + \bar{v}) \approx m_{f_\theta}(\bar{u} + \bar{v}; \bar{u})$, where

$$m_{f_\theta}(\bar{u}{+}\bar{v}; \bar{u}) \triangleq \frac{1}{\theta}\log \mathbb{E}_{\bar{w}} \exp \theta q_h\left(\bar{x}{+}\nabla\tilde{x}(\bar{u})^\top\bar{v}{+}\nabla\tilde{x}(\bar{u})^\top\bar{w}; \bar{x}\right)$$
$$+ q_g(\bar{u} + \bar{v}; \bar{u}), \tag{12}$$

$q_h(\bar{x}{+}\bar{y}; \bar{x}) \triangleq h(\bar{x}){+}\nabla h(\bar{x})^\top\bar{y}{+}\bar{y}^\top\nabla^2 h(\bar{x})\bar{y}/2$, $q_g(\bar{u}{+}\bar{v}; \bar{u})$ is defined similarly and $\bar{x} = \tilde{x}(\bar{u})$ is the exact trajectory. As the following proposition will clarify, the descent direction computed by ILEQG in (7) is given by minimizing directly the model m_{f_θ}. Yet, from an optimization viewpoint, a regularization term must be added to this minimization to ensure that the solutions stay in a region where the model is valid. Formally, we consider a regularized variant of ILEQG, we call RegILEQG, that starts at a point $\bar{u}_0$ and defines the next iterate as

$$\bar{u}^{(k+1)} = \bar{u}^{(k)} + \operatorname*{arg\,min}_{\bar{v}\in\mathbb{R}^{\tau p}} \left\{ m_{f_\theta}(\bar{u}^{(k)} + \bar{v}; \bar{u}^{(k)}) + \frac{1}{2\gamma_k}\|\bar{v}\|_2^2 \right\}, \tag{RegILEQG}$$

where γ_k is the step-size: the smaller γ_k is, the closer the solution is to the current iterate. The following proposition shows that the minimization step (RegILEQG) amounts to a linear quadratic exponential Gaussian risk-sensitive control problem.

Proposition 2.1. *The model minimization step* (RegILEQG) *is given as* $\bar{u}^{(k+1)} = \bar{u}^{(k)} + \bar{v}^*$ *where* $\bar{v}^*$ *is the solution, if it exists,*

of

$$\begin{aligned}\min_{\bar{v}\in\mathbb{R}^{\tau p}} \sup_{\substack{\bar{w}\in\mathbb{R}^{\tau p}\\ \bar{y}\in\mathbb{R}^{\tau d}}} \quad & \sum_{t=1}^{\tau}\left(\frac{1}{2}y_t^\top H_t y_t + \tilde{h}_t^\top y_t\right) \\ & + \sum_{t=0}^{\tau-1}\left(\frac{1}{2}v_t^\top (G_t+\gamma_k^{-1}\,\mathrm{I}_p)v_t + \tilde{g}_t^\top v_t\right) \quad (13)\\ & - \sum_{t=0}^{\tau-1}\frac{1}{2\theta\sigma^2}\|w_t\|_2^2 \\ \text{subject to} \quad & y_{t+1} = A_t y_t + B_t(v_t+w_t) \\ & y_0 = 0,\end{aligned}$$

where $x_t^{(k)} = \tilde{x}_t(\bar{u}^{(k)})$ $A_t = \nabla_x \phi_t(x_t^{(k)}, u_t^{(k)})^\top$, $B_t = \nabla_u \phi_t(x_t^{(k)}, u_t^{(k)})^\top$, $H_t = \nabla^2 h_t(x_t^{(k)})$, $\tilde{h}_t = \nabla h_t(x_t^{(k)})$, $G_t = \nabla^2 g_t(u_t^{(k)})$, $\tilde{g}_t = \nabla g_t(u_t^{(k)})$.

Each model-minimization step can then be performed by dynamic programming. The overall algorithm is presented in Appendix C. If the costs depend only on the final state, i.e., $h(\bar{x}) = h_\tau(x_\tau)$, the steps can be computed more efficiently by making calls to automatic differentiation oracles, see Appendix C for more details.

2.2 Convergence analysis

We analyze the behavior of the regularized variant of ILEQG for quadratic convex costs h_t, g_t, a common setting in applications. This algorithm is then based on two different approximations:

(i) the random trajectories are approximated by Gaussians defined by the linearization of the dynamics,
(ii) the non-linear control of the trajectory is approximated a linear control defined by the linearization of the dynamics,

The first approximation makes the algorithm work on a surrogate of the true risk-sensitive objective. By identifying this surrogate, we get criteria for the choice of the step-size γ.

Approximate risk-sensitive cost. By approximating the noisy trajectory by a Gaussian variable using first-order information of the trajectory, we define the approximated risk-sensitive objective as follows

$$\begin{aligned}&\hat{f}_\theta(\bar{u}) = \hat{\eta}_\theta(\bar{u}) + g(\bar{u}),\\ \text{where} \quad &\hat{\eta}_\theta(\bar{u}) = \frac{1}{\theta}\log \mathbb{E}_{\bar{w}} \exp[\theta h(\tilde{x}(\bar{u}) + \nabla\tilde{x}(\bar{u})^\top \bar{w})].\end{aligned}$$

The approximated risk-sensitive objective is essentially the log-partition function of a Gaussian distribution as shown in the following proposition.

Proposition 2.2. *For $\bar{u}\in\mathbb{R}^{\tau p}$ with $\bar{x} = \tilde{x}(\bar{u})$, if*

$$\sigma^{-2}\,\mathrm{I}_{\tau p} \succ \theta \nabla\tilde{x}(\bar{u})\nabla^2 h(\bar{x})\nabla\tilde{x}(\bar{u})^\top, \qquad (14)$$

the approximated risk sensitive cost is defined and is the scaled log-partition function of

$$\hat{p}(\bar{w};\bar{u}) = \exp\left(\theta h(\tilde{x}(\bar{u}) + \nabla\tilde{x}(\bar{u})^\top\bar{w}) - \frac{1}{2\sigma^2}\|\bar{w}\|_2^2 - \theta\hat{\eta}_\theta(\bar{u})\right), \qquad (15)$$

which is the density of a Gaussian $\mathcal{N}(\bar{w}_*, \Sigma)$ with

$$\bar{w}_* = \theta\Sigma X\tilde{h}, \qquad \Sigma = (\sigma^{-2}\,\mathrm{I}_{\tau p} - \theta X H X^\top)^{-1}, \qquad (16)$$

where $X = \nabla\tilde{x}(\bar{u})$, $\tilde{h} = \nabla h(\bar{x})$, $H = \nabla^2 h(\bar{x})$ and $\bar{x} = \tilde{x}(\bar{u})$. Therefore, the approximated risk-sensitive loss can be computed analytically.

The approximation error induced by the linearization is illustrated in Sec. 3. Note that the risk-sensitive approximation shares similar properties as the original function in (4), since it can be extended around $\theta = 0$ to

$$\begin{aligned}\hat{\eta}_\theta(\bar{u}) = h(\tilde{x}(\bar{u})) + \mathbb{E}_{\bar{w}\sim\hat{p}(\cdot;\bar{u})}\,\bar{w}^\top\nabla\tilde{x}(\bar{u})\nabla^2 h(\tilde{x}(\bar{u}))\nabla\tilde{x}(\bar{u})^\top\bar{w}\\ + \frac{\theta}{2}\,\mathbb{V}\mathrm{ar}_{\bar{w}\sim\hat{p}(\cdot;\bar{u})}\,h(\tilde{x}(\bar{u}) + \nabla\tilde{x}(\bar{u})^\top\bar{w})) + \mathcal{O}(\theta^2).\end{aligned}$$

Namely, it accounts not only for the cost of the exact trajectory but also for the variance defined by the linearized trajectories. Provided that condition (14) holds, the gradient of the approximated risk-sensitive cost reads (see Appendix D)

$$\begin{aligned}\nabla\hat{\eta}_\theta(\bar{u}) = \mathbb{E}_{\bar{w}\sim\hat{p}(\cdot;\bar{u})}(\nabla\tilde{x}(\bar{u}) + \nabla^2\tilde{x}(\bar{u})[\cdot,\bar{w},\cdot])\\ \times\,\nabla h(\tilde{x}(\bar{u}) + \nabla\tilde{x}(\bar{u})^\top\bar{w}),\end{aligned}$$

where $\hat{p}(\cdot;\bar{u})$ is defined in (15). Denote the truncated gradient of the approximated risk-sensitive cost

$$\widehat{\nabla}\hat{\eta}_\theta(\bar{u}) = \mathbb{E}_{\bar{w}\sim\hat{p}(\cdot;\bar{u})}\,\nabla\tilde{x}(\bar{u})\nabla h(\tilde{x}(\bar{u}) + \nabla\tilde{x}(\bar{u})^\top w).$$

We link the model-minimization steps of the regularized variant of ILEQG to the truncated gradient in the following proposition.

Proposition 2.3. *Consider* (RegILEQG) *at iteration k, if condition* (14) *holds on $\bar{u}^{(k)}$, the step is defined and reads*

$$\begin{aligned}\bar{u}^{(k+1)} = \bar{u}^{(k)} - (G + \gamma_k^{-1}\,\mathrm{I}_{\tau p} + XHX^\top + \theta V)^{-1}\\ \times\,(\nabla g(\bar{u}^{(k)}) + \widehat{\nabla}\hat{\eta}_\theta(\bar{u}^{(k)})),\end{aligned}$$

where

$$\begin{aligned}V &= \mathbb{V}\mathrm{ar}_{\bar{w}\sim\hat{p}(\cdot;\bar{u}^{(k)})}\,\nabla\tilde{x}(\bar{u}^{(k)})\nabla h(\tilde{x}(\bar{u}^{(k)}) + \nabla\tilde{x}(\bar{u}^{(k)})^\top w)\\ &= XHX^\top(\sigma^{-2}\,\mathrm{I}_{\tau p} - \theta XHX^\top)^{-1}XHX^\top\end{aligned}$$

and $X{=}\nabla\tilde{x}(\bar{u}^{(k)})$, $H{=}\nabla^2 h(\bar{x})$, $G{=}\nabla^2 g(\bar{u}^{(k)})$, $\bar{x}{=}\tilde{x}(\bar{u}^{(k)})$.

Convergence to stationary points. We make the following assumptions for our analysis

Assumption 2.4.

1. *The dynamics ϕ_t are twice differentiable, bounded, Lipschitz, smooth such that the trajectory function $\tilde{x}$ is also twice differentiable, bounded, Lipschitz and*

smooth. Denote by $L_{\tilde{x}}$ and $\ell_{\tilde{x}}$ the Lipschitz continuity and smoothness constants respectively of $\tilde{x}$ and define $M_{\tilde{x}} = \max_{\bar{u}\in\tau p} \operatorname{dist}(\tilde{x}(\bar{u}), X^*)$, where $X^* = \arg\min_{\bar{x}\in\mathbb{R}^{\tau d}} h(\bar{x})$.

2. *The costs h and g are convex quadratics with smoothness constants L_h, L_g.*
3. *The risk-sensitivity parameter is chosen such that $\tilde{\sigma}^{-2} = \sigma^{-2} - \theta L_h \ell_{\tilde{x}}^2 > 0$, which ensures that condition (14) holds for any $\bar{u} \in \mathbb{R}^{\tau p}$.*

The following proposition shows stationary convergence for the regularized variant of ILEQG as an optimization method of the approximated risk-sensitive loss. The additional constant term is due to the truncation of the gradient of the approximated risk-sensitive cost.

Theorem 2.5. *Under Ass. 2.4, suppose that the step-sizes of (RegILEQG) are chosen such that*

$$\hat{f}_\theta(\bar{u}^{(k+1)}) \leq m_{f_\theta}(\bar{u}^{(k+1)}; \bar{u}^{(k)}) + \frac{1}{2\gamma_k}\|\bar{u}^{(k+1)} - \bar{u}^{(k)}\|_2^2, \tag{17}$$

with $\gamma_k \in [\gamma_{\min}, \gamma_{\max}]$. Then, the approximated objective $\hat{f}_\theta$ decreases and after K iterations we have

$$\min_{k=0,\dots,K-1} \|\nabla \hat{f}_\theta(\bar{u}^{(k)})\|_2 \leq L\sqrt{\frac{2(\hat{f}_\theta(\bar{u}^{(0)}) - \hat{f}_\theta(\bar{u}^{(K)}))}{K}} + \delta,$$

where $L = \max_{\gamma\in[\gamma_{\min},\gamma_{\max}]} \sqrt{\gamma}(L_g + \gamma^{-1} + (\tilde{\sigma}/\sigma)^2 \ell_{\tilde{x}}^2 L_h)$, $\delta = \theta\tilde{\sigma}^2 L_h^2 L_{\tilde{x}} \ell_{\tilde{x}} M_{\tilde{x}}^2 + \theta^2 \tilde{\sigma}^4 L_h^3 L_{\tilde{x}} \ell_{\tilde{x}}^3 M_{\tilde{x}}^2 + \tau p \tilde{\sigma}^2 L_h L_{\tilde{x}} \ell_{\tilde{x}}$.

Previous proposition gives a criterion (17) for line-searches. We show in Appendix D that there exists a step-size $\hat{\gamma}$ such that condition (17) is satisfied along the iterations. With this step-size, the number of steps to get an $\epsilon + \delta$ stationary point is at most

$$\frac{2\hat{\gamma}(L_g + \hat{\gamma}^{-1} + (\tilde{\sigma}/\sigma)^2 \ell_{\tilde{x}}^2 L_h)^2 (\hat{f}_\theta(\bar{u}^{(0)}) - \hat{f}_\theta^*)}{\epsilon^2}.$$

3 Numerical experiments

Control settings. We apply the risk-sensitive framework to two classical continuous time control settings: swinging-up a pendulum and moving a two-link arm robot, both detailed in Appendix E. Their discretization leads to dynamics of the form

$$\begin{aligned} x_{1,t+1} &= x_{1,t} + \delta x_{2,t} \\ x_{2,t+1} &= x_{2,t} + \delta f(x_{1,t}, x_{2,t}, u_t) \end{aligned} \tag{18}$$

for $t = 0, \dots \tau - 1$, where x_1, x_2 describe the position and the speed of the system respectively, f defines the dynamics derived by Newton's law, δ is the time step, u is a force that controls the system.

Noise modeling. The risk-sensitive cost is defined by an additional noisy force applied to the dynamics. Formally, the discretized dynamics (18) are modified as

$$\begin{aligned} x_{1,t+1} &= x_{1,t} + \delta x_{2,t} \\ x_{2,t+1} &= x_{2,t} + \delta f(x_{1,t}, x_{2,t}, u_t + w_t) \end{aligned} \tag{19}$$

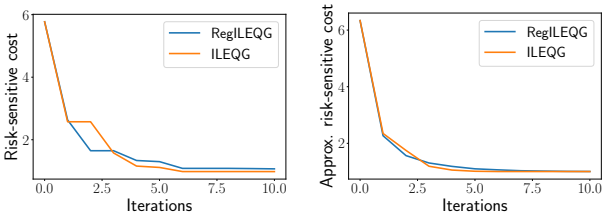

Figure 2: Convergence of iterative linearized methods, RegILEQG and ILEQG, on the pendulum problem.

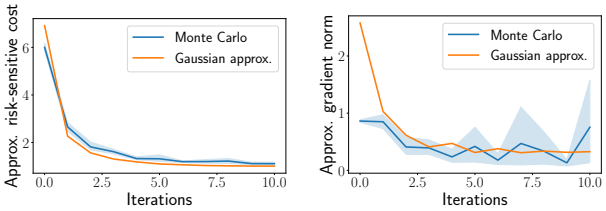

Figure 3: Risk-sensitive and gradient approximations.

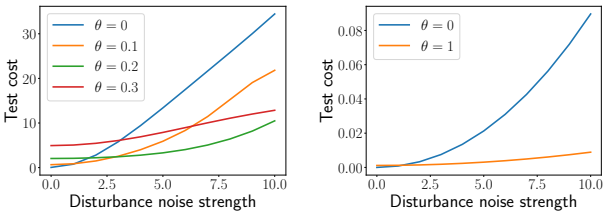

Figure 4: Robustness of controllers against disturbance noise. Left: pendulum. Right: robot arm.

for $t = 0, \dots, \tau - 1$, where $w_t \sim \mathcal{N}(0, \sigma^2 \mathrm{I}_p)$ and σ is chosen to avoid chaotic behavior, see Appendix E.

We test the optimized expected or risk-sensitive costs on a setting where the dynamics are perturbed at a given time t_w by a force of amplitude ρ. This models the robustness of the control against kicking the robot. Formally, we analyze the performance of the solutions of the expected cost (denoted $\theta = 0$) or the risk-sensitive cost (3) on dynamics of the form

$$\begin{aligned} x_{1,t+1} &= x_{1,t} + \delta x_{2,t} \\ x_{2,t+1} &= x_{2,t} + \delta f(x_{1,t}, x_{2,t}, u_t + \rho \mathbb{1}(t = t_w)) \end{aligned}$$

for $t = 0, \dots, \tau - 1$, where $\rho \sim \mathcal{N}(0, \sigma_{test} \mathrm{I}_p)$ with the same cost $h(\bar{x})$ computed as an average on $n = 100$ simulations. We call this cost the test cost.

3.1 Results

All detailed parameters are provided in Appendix E.

Convergence. In Fig. 2 we compare the convergence on the pendulum problem of RegILEQG and ILEQG. For both algorithms, we use a constant step-size sequence tuned after a burn-in phase of 5 iterations on a grid of step-sizes 2^i for $i \in [-5, 10]$.

The approximated risk-sensitive loss was used to tune the step-sizes. The best step-sizes found were 0.5 for ILEQG and 16 for RegILEQG. We plot the minimum values obtained until now, as the true function can be approximated. We observe that both ILEQG and RegILEQG minimize well the approximated risk-sensitive cost. Yet, the regularized variant provides smoother convergence. We leave as future work the implementation of line-search procedures as done for Levenberg-Marquardt methods.

Risk-sensitive cost approxiamtion. In Fig. 3, we show $\hat{f}_\theta(\bar{u}^{(k)})$, $\|\nabla \hat{f}_\theta(\bar{u}^{(k)})\|_2$ and $f_\theta(\bar{u}^{(k)})$, $\|\nabla f_\theta(\bar{u}^{(k)})\|_2$ approximated by Monte-Carlo for $N = 100$ samples along the iterations of the RegILEQG method for the pendulum (same experiment as in Fig. 2) for 10 runs of the Monte-Carlo approximation. We observe that the approximation $\tilde{f}_\theta(\bar{u}^{(k)})$ is fine compared to the approximation by Monte-Carlo. The sequence of compositions defining the trajectory leads to highly non-smooth functions (i.e. large smoothness constants), which contributes to the high variance of gradients computed by Monte-Carlo.

Robustness. In Fig. 4, we plot the test cost obtained by the expected or risk-sensitive optimizers on the movement perturbed by a dirac of increasing strength. We use our RegILEQG algorithm with constant-step-size tuned after a burn-in phase. The risk-sensitive approach provides smaller costs against perturbed trajectories. On the two-link-arm problem, we did not observe significant changes when varying the risk-sensitivity parameter. We leave the analysis of the choice of the parameter for future work.

4 Conclusion

We dissected the ILEQG algorithm to understand its correct implementation, this revealed: (i) the objective it minimizes, that is not the risk-sensitive cost but an approximation of it, (ii) the necessary introduction from an optimization viewpoint of a regularization inside the step, (iii) a sufficient decrease condition that ensures proven stationary convergence to a near-stationary point.

Acknowledgements

This work was funded by NIH R01 (#R01EB019335), NSF CPS (#1544797), NSF NRI (#1637748), NSF CCF (#1740551), NSF DMS (#1839371), DARPA Lagrange grant FA8650-18-2-7836, the program Learning in Machines and Brains of CIFAR, ONR, RCTA, Amazon, Google, Honda and faculty research awards

References

R. Bellman. *Introduction to the mathematical theory of control processes*, volume 2. Academic press, 1971.

A. Ben-Tal and M. Teboulle. Expected utility, penalty functions, and duality in stochastic nonlinear programming. *Management Science*, 32(11):1445–1466, 1986.

A. Ben-Tal and M. Teboulle. An old-new concept of convex risk measures: The optimized certainty equivalent. *Mathematical Finance*, 17 (3):449–476, 2007.

D. P. Bertsekas. *Abstract dynamic programming*. Athena Scientific, 2nd edition, 2018.

J. De O. Pantoja. Differential dynamic programming and Newton's method. *International Journal of Control*, 47(5):1539–1553, 1988.

J. C. Dunn and D. P. Bertsekas. Efficient dynamic programming implementations of Newton's method for unconstrained optimal control problems. *Journal of Optimization Theory and Applications*, 63(1): 23–38, 1989.

K. Dvijotham, M. Fazel, and E. Todorov. Universal convexification via risk-aversion. In *Proceedings of the Thirtieth Conference on Uncertainty in Artificial Intelligence*, pages 162–171, 2014.

F. Farshidian and J. Buchli. Risk sensitive, nonlinear optimal control: Iterative linear exponential-quadratic optimal control with Gaussian noise. *arXiv preprint arXiv:1512.07173*, 2015.

K. Glover and J. C. Doyle. State-space formulae for all stabilizing controllers that satisfy an h^∞-norm bound and relations to relations to risk sensitivity. *Systems & Control Letters*, 11(3):167–172, 1988.

B. Hassibi, A. H. Sayed, and T. Kailath. *Indefinite-Quadratic Estimation and Control: A Unified Approach to H2 and H-infinity Theories*, volume 16. SIAM, 1999.

J. W. Helton and M. R. James. *Extending H-infinity control to nonlinear systems: Control of nonlinear systems to achieve performance objectives*, volume 1. SIAM, 1999.

D. Jacobson. Optimal stochastic linear systems with exponential performance criteria and their relation to deterministic differential games. *IEEE Transactions on Automatic control*, 18(2):124–131, 1973.

W. Li and E. Todorov. Iterative linear quadratic regulator design for nonlinear biological movement systems. In *1st International Conference on Informatics in Control, Automation and Robotics*, volume 1, pages 222–229, 2004.

W. Li and E. Todorov. Iterative linearization methods for approximately optimal control and estimation of non-linear stochastic system. *International Journal of Control*, 80(9):1439–1453, 2007.

Y. Nesterov. *Introductory lectures on convex optimization: A basic course*, volume 87. Springer Science & Business Media, 2013.

B. Ponton, S. Schaal, and L. Righetti. On the effects of measurement uncertainty in optimal control of contact interactions. In *The 12th International Workshop on the Algorithmic Foundations of Robotics WAFR*, 2016.

V. Roulet, S. Srinivasa, D. Drusvyatskiy, and Z. Harchaoui. Iterative linearized control: Stable algorithms and complexity guarantees. In *Proceedings of the 36th International Conference on Machine Learning*, 2019.

A. Sideris and J. E. Bobrow. An efficient sequential linear quadratic algorithm for solving nonlinear optimal control problems. In *Proceedings of the American Control Conference*, pages 2275–2280, 2005.

P. Sopasakis, D. Herceg, A. Bemporad, and P. Patrinos. Risk-averse model predictive control. *Automatica*, 100:281–288, 2019.

J. Speyer, J. Deyst, and D. Jacobson. Optimization of stochastic linear systems with additive measurement and process noise using exponential performance criteria. *IEEE Transactions on Automatic Control*, 19(4):358–366, 1974.

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

A Notations

A.1 Miscellaneous

We use semicolons to denote concatenation of vectors, namely for n d-dimensional vectors $a_1, \dots, a_n \in \mathbb{R}^d$, we have $(a_1; \dots; a_n) \in \mathbb{R}^{nd}$. The Kronecker product is denoted $\otimes$. For a sequence of matrices $X_1, \dots X_\tau \in \mathbb{R}^{d \times p}$ we denote

$$\mathbf{diag}(X_1, \dots, X_\tau) = \begin{pmatrix} X_1 & 0 & \dots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & X_\tau \end{pmatrix} \in \mathbb{R}^{d\tau \times p\tau}.$$

the corresponding block diagonal matrix. For a set $S \subset \mathbb{R}^d$ and $x \in \mathbb{R}^d$, denote $\operatorname{dist}(x, S)^2 = \min_{y \in \mathbb{R}^d} \|x - y\|_2^2$. Given a density function $p : \mathbb{R}^d \to \mathbb{R}^+$, such that $\int_{\mathbb{R}^d} p(w)dw = 1$ and a function $f : \mathbb{R}^d \to \mathbb{R}^p$ we denote

$$\mathbb{E}_{w \sim p} f(w) = \int_{\mathbb{R}^d} f(w)p(w)dw.$$

For a matrix $M \in \mathbb{R}^{d \times d}$, we denote $\|M\|_2 = \sup_{x \in \mathbb{R}^d_*} x^\top Mx / \|x\|_2^2$ the spectral norm induced by the Euclidean norm. We denote semi-definite positive matrices $S \in \mathbb{R}^{d \times d}$ as $S \succeq 0$ and denote $\lambda_{\max}(S) = \|S\|_2$ the maximal eigenvalue of S. For a matrix $A \in \mathbb{R}^{d \times n}$ we denote by $A^\dagger$ the pseudo-inverse of A.

A.2 Tensors

For a tensor $\mathcal{A} = (a_{i,j,k})_{i \in \{1,\dots,d\},\ j \in \{1,\dots,n\},\ k \in \{1,\dots,p\}} \in \mathbb{R}^{d \times n \times p}$, we denote $\mathcal{A}_{i,\cdot,\cdot} = (a_{i,j,k})_{j \in \{1,\dots,n\},\ k \in \{1,\dots,p\}} \in \mathbb{R}^{n \times p}$ the matrix obtained by fixing the first index at i. Similarly we define $\mathcal{A}_{\cdot,j,\cdot} \in \mathbb{R}^{d \times p}$ and $\mathcal{A}_{\cdot,\cdot,k} \in \mathbb{R}^{d \times n}$. A tensor $\mathcal{A}$ can be represented as the list of matrices $\mathcal{A} = (\mathcal{A}_{\cdot,\cdot,1}, \dots, \mathcal{A}_{\cdot,\cdot,k})$. Given matrices $P \in \mathbb{R}^{d \times d'}, Q \in \mathbb{R}^{n \times n'}, R \in \mathbb{R}^{p \times p'}$, we denote

$$\mathcal{A}[P, Q, R] = \left(\sum_{k=1}^{p} R_{k,1} P^\top \mathcal{A}_{\cdot,\cdot,k} Q, \dots, \sum_{k=1}^{p} R_{k,p'} P^\top \mathcal{A}_{\cdot,\cdot,k} Q \right) \in \mathbb{R}^{d' \times n' \times p'}$$

If P, Q or R are identity matrices, we use the symbol " $\cdot$ " in place of the identity matrix. For example, we denote $\mathcal{A}[P, Q, \mathrm{I}_p] = \mathcal{A}[P, Q, \cdot] = \left(P^\top \mathcal{A}_{\cdot,\cdot,1} Q, \dots, P^\top \mathcal{A}_{\cdot,\cdot,p} Q\right)$. If P, Q or R are vectors we consider the flatten object. In particular, for $x \in \mathbb{R}^d, y \in \mathbb{R}^n$, we denote

$$\mathcal{A}[x, y, \cdot] = \begin{pmatrix} x^\top \mathcal{A}_{\cdot,\cdot,1} y \\ \vdots \\ x^\top \mathcal{A}_{\cdot,\cdot,p} y \end{pmatrix} \in \mathbb{R}^p$$

rather than having $\mathcal{A}[x, y, \cdot] \in \mathbb{R}^{1 \times 1 \times p}$. Similarly, for $z \in \mathbb{R}^p$, we have

$$\mathcal{A}[\cdot, \cdot, z] = \sum_{k=1}^{p} z_k \mathcal{A}_{\cdot,\cdot,k} \in \mathbb{R}^{d \times n}.$$

For a tensor $\mathcal{A}$, we denote

$$\|\mathcal{A}\|_2 = \sup_{x \in \mathbb{R}^d_*, y \in \mathbb{R}^n_*, z \in \mathbb{R}^p_*} \frac{\mathcal{A}[x, y, z]}{\|x\|_2 \|y\|_2 \|z\|_2} \tag{20}$$

the norm induced by the Euclidean norm for the tensor $\mathcal{A}$.

A.3 Gradients

For a multivariate function $f : \mathbb{R}^d \mapsto \mathbb{R}^n$, composed of $f^{(j)}$ real functions with $j \in \{1, \dots, n\}$, we denote $\nabla f(x) = (\nabla f^{(1)}(x), \dots, \nabla f^{(n)}(x)) \in \mathbb{R}^{d \times n}$, that is the transpose of its Jacobian on x, $\nabla f(x) = (\frac{\partial f^{(j)}}{\partial x_i}(x))_{1 \le i \le d, 1 \le j \le n} \in \mathbb{R}^{d \times n}$. We represent its 2nd order information by a tensor $\nabla^2 f(x) = (\nabla^2 f^{(1)}(x), \dots, \nabla^2 f^{(n)}(x)) \in \mathbb{R}^{d \times d \times n}$

For a real function, $f : \mathbb{R}^d \times \mathbb{R}^p \mapsto \mathbb{R}$, whose value is denoted $f(x, y)$, we decompose its gradient $\nabla f(x, y) \in \mathbb{R}^{d+p}$ on $(x, y) \in \mathbb{R}^d \times \mathbb{R}^p$ as

$$\nabla f(x,y) = \begin{pmatrix} \nabla_x f(x,y) \\ \nabla_y f(x,y) \end{pmatrix} \qquad \text{with} \qquad \nabla_x f(x,y) \in \mathbb{R}^d, \quad \nabla_y f(x,y) \in \mathbb{R}^p.$$

Similarly for a multivariate function $f : \mathbb{R}^d \times \mathbb{R}^p \mapsto \mathbb{R}^n$ and (x, y), we denote $\nabla_x f(x, y) = (\nabla_x f^{(1)}(x, y), \ldots, \nabla_x f^{(n)}(x, y)) \in \mathbb{R}^{d\times n}$ and we define similarly $\nabla_y f(x, y) \in \mathbb{R}^{p\times n}$.

We drop the dependency to the time when it is clear from context, e.g., for a dynamic $\phi_t : \mathbb{R}^{d+p} \to \mathbb{R}^d$ we denote by $\nabla_u \phi_t(x_t, u_t) = \nabla_{u_t} \phi_t(x_t, u_t)$. Those definitions extend for noisy dynamics ψ_t, where we add the noise variable $w \in \mathbb{R}^q$.

All Lipschitz continuity constants are defined with respect to the norm induced by the Euclidean norm. In particular, for a multivariate twice differentiable function f, we say that it is smooth if its second order tensor has a bounded norm for the Euclidean induced norm of a tensor defined in (20).

B Linear quadratic risk sensitive control

B.1 Min-max formulation

Proposition 1.1. *Consider quadratic objectives and linear dynamics defined by*

$$\begin{aligned} h_t(x_t) = \frac{1}{2} x_t^\top H_t x_t + \tilde{h}_t^\top x_t, \quad g_t(u_t) = \frac{1}{2} u_t^\top G_t u_t + \tilde{g}_t^\top u_t, \\ x_{t+1} = A_t x_t + B_t u_t + C_t w_t, \end{aligned} \tag{5}$$

where $H_t \succeq 0$, $G_t \succ 0$, $w_t \sim \mathcal{N}(0, \sigma^2 \mathrm{I}_q)$ and denote by $H, \tilde{B}, \tilde{C}, \tilde{x}_0$ the matrices and vector such that for any trajectory $\bar{x}$, $H = \nabla^2 h(\bar{x})$, $\bar{x} = \tilde{B}\bar{u} + \tilde{C}\bar{w} + \tilde{x}_0$. We have that

(i) the risk sensitive control problem (3) is equivalent to

$$\begin{aligned} \min_{\bar{u}\in\mathbb{R}^{\tau p}} \sup_{\substack{\bar{w}\in\mathbb{R}^{\tau q} \\ \bar{x}\in\mathbb{R}^{\tau d}}} & \sum_{t=1}^{\tau} \frac{1}{2} x_t^\top H_t x_t + \tilde{h}_t^\top x_t + \sum_{t=0}^{\tau-1} \frac{1}{2} u_t^\top G_t u_t + \tilde{g}_t^\top u_t \\ & - \sum_{t=0}^{\tau-1} \frac{1}{2\theta\sigma^2} \|w_t\|_2^2 \\ \text{subject to} \quad & x_{t+1} = A_t x_t + B_t u_t + C_t w_t \\ & x_0 = \hat{x}_0, \end{aligned} \tag{6}$$

(ii) if $(\theta\sigma^2)^{-1} < \lambda_{\max}(\tilde{C}^\top H \tilde{C})$ the risk-sensitive control problem is infeasible,

(iii) if $(\theta\sigma^2)^{-1} > \lambda_{\max}(\tilde{C}^\top H \tilde{C})$, the risk-sensitive control problem can be solved analytically by dynamic programming.

Proof of (i). Since w_t are i.i.d, the states x_t given by the linear dynamics form a Markov sequence of random variables, i.e., denoting $\mathbb{P}$ the probability defined by the dynamics, for any $t \in \{0, \ldots, \tau - 1\}$, $\mathbb{P}(x_{t+1}|x_t, \ldots, x_0) = \mathbb{P}(x_{t+1}|x_t) \sim \mathcal{N}(A_t x_t + B_t u_t, \Sigma_t)$ where $\Sigma_t = \sigma^2 C_t C_t^\top$ is potentially degenerated and $x_0 = \hat{x}_0$. Precisely denote $d\mu(\bar{x})$ a measure such that $\mathbb{P}(\Pi_{\mathrm{Null}(\Sigma_t)}(x_{t+1} - A_t x_t - B_t u_t) = 0) = 1$ with $\Pi_{\mathrm{Null}(\Sigma_t)}$ the orthonormal projection on the null space of Σ_t, we have

$$\begin{aligned} \mathbb{E}_{\bar{x}\sim p(\cdot;\bar{u})} \left[\exp(\theta h(\bar{x}))\right] \propto \int \exp\Bigg(& -\sum_{t=0}^{\tau-1} \frac{1}{2} (x_{t+1} - A_t x_t - B_t u_t)^\top \Sigma_t^\dagger (x_{t+1} - A_t x_t - B_t u_t) \\ & + \theta \sum_{t=1}^{\tau} \frac{1}{2} x_t^\top H_t x_t + \tilde{h}_t^\top x_t \Bigg) d\mu(x_1, \ldots, x_\tau) \\ = \int \exp(-Q(\bar{x}, \bar{u})) d\mu(\bar{x}), & \end{aligned}$$

where $Q(\bar{x}, \bar{u})$ is a quadratic in $\bar{x}, \bar{u}$ and we ignored the normalization constants in the first line as we are interested in computing the minimum. The integral will then be finite if and only if $Q(\bar{x}, \bar{u})$ is bounded below in $\bar{x} \in \mathcal{X} = \{\bar{x} : \Pi_{\mathrm{Null}(\Sigma_t)}(x_{t+1} - A_t x_t - B_t u_t) = 0 \quad \text{for} \quad t \in \{0, \ldots, \tau - 1\}\}$. In that case, the quadratic reads $Q(\bar{x}, \bar{u}) = Q(\bar{x} - \bar{x}^*, \bar{u}) + \min_{\bar{x}\in\mathcal{X}} Q(\bar{x}, \bar{u})$ where $\bar{x}^* \in \arg\min_{\bar{x}\in\mathcal{X}} Q(\bar{x}, \bar{u})$. The expectation is then proportional to, the variance term being independent of $\bar{u}$,

$$\mathbb{E}_{\bar{x}\sim p(\cdot;\bar{u})} \left[\theta \exp(h(\bar{x}))\right] \propto \exp\left(-\min_{\bar{x}} Q(\bar{x}, \bar{u})\right).$$

By parameterizing the states as $x_{t+1} = A_t x_t + B_t u_t + C_t w_t$, using that C_t has the same image as Σ_t, the minimization can be rewritten

$$\begin{aligned}\min_{\bar{x}\in\mathcal{X}} Q(\bar{x},\bar{u}) = \quad & \min_{\bar{w}\in\mathbb{R}^{\tau d},\bar{x}\in\mathbb{R}^{\tau d}} \quad -\theta\sum_{t=1}^{\tau}\left(\frac{1}{2}x_t^\top H_t x_t + \tilde{h}_t^\top x_t\right) + \sum_{t=0}^{\tau-1}\frac{1}{2\theta\sigma^2}\|w_t\|_2^2 \\ & \text{subject to} \quad x_{t+1} = A_t x_t + B_t u_t + C_t w_t \\ & \qquad\qquad\quad x_0 = \hat{x}_0.\end{aligned}$$

The risk sensitive control problem (3) is then equivalent to

$$\begin{aligned}\min_{\bar{u}\in\mathbb{R}^{\tau p}} \sup_{\bar{w}\in\mathbb{R}^{\tau q},\bar{x}\in\mathbb{R}^{\tau d}} \quad & \sum_{t=1}^{\tau}\frac{1}{2}x_t^\top H_t x_t + \tilde{h}_t^\top x_t + \sum_{t=0}^{\tau-1}\frac{1}{2}u_t^\top G_t u_t + \tilde{g}_t^\top u_t - \sum_{t=0}^{\tau-1}\frac{1}{2\theta\sigma^2}\|w_t\|_2^2 \\ \text{subject to} \quad & x_{t+1} = A_t x_t + B_t u_t + C_t w_t \\ & x_0 = \hat{x}_0,\end{aligned}$$

which, if the sup is not attained, means that the problem is infeasible. □

Proof of (ii). The linear dynamics read $x_{t+1} - A_t x_t = B_t u_t + C_t w_t$ for $t = 0, \dots, \tau-1$. Denoting

$$L = \begin{pmatrix} I & 0 & \dots & 0 \\ -A_1 & I & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & -A_{\tau-1} & I \end{pmatrix} \quad \text{with} \quad L^{-1} = \begin{pmatrix} I & 0 & \dots & 0 \\ A_1 & I & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots \\ A_{\tau-1}\dots A_1 & A_{\tau-1}\dots A_2 & \dots & I \end{pmatrix},$$

we get

$$L\bar{x} = \bar{B}\bar{u} + \bar{C}\bar{w} + \tilde{x}_0 \quad \text{and so} \quad \bar{x} = L^{-1}(\bar{B}\bar{u} + \bar{C}\bar{w} + \tilde{x}_0),$$

where $\tilde{x}_0 = (A_0\hat{x}_0; 0; \dots; 0) \in \mathbb{R}^{\tau d}$, $\bar{x} = (x_1; \dots; x_\tau)$, $\bar{B} = \mathbf{diag}(B_0, \dots, B_{\tau-1})$, $\bar{C} = \mathbf{diag}(C_0, \dots, C_{\tau-1})$. Problem (6) reads then

$$\begin{aligned}\min_{\bar{u}\in\mathbb{R}^{\tau p}} \sup_{\bar{w}\in\mathbb{R}^{\tau q}} & \frac{1}{2}(\bar{B}\bar{u} + \bar{C}\bar{w} + \tilde{x}_0)^\top L^{-\top}\bar{H}L^{-1}(\bar{B}\bar{u} + \bar{C}\bar{w} + \tilde{x}_0) + \bar{h}^\top L^{-1}(\bar{B}\bar{u} + \bar{C}\bar{w} + \tilde{x}_0) \\ & + \frac{1}{2}\bar{u}^\top\bar{G}\bar{u} + \bar{g}^\top\bar{u} - \frac{1}{2\theta\sigma^2}\|\bar{w}\|_2^2,\end{aligned} \tag{21}$$

where $\bar{H} = \mathbf{diag}(H_1, \dots, H_\tau)$, $\bar{G} = \mathbf{diag}(G_0, \dots, G_{\tau-1})$, $\bar{h} = (h_1; \dots; h_\tau)$ and $\bar{g} = (g_0; \dots; g_{\tau-1})$. It is always a strongly convex problem in $\bar{u}$ by assumption on the G_t. If

$$(\theta\sigma^2)^{-1} < \lambda_{\max}(\bar{C}^\top L^{-\top}\bar{H}L^{-1}\bar{C})$$

i.e., $(\theta\sigma^2)^{-1}\,\mathrm{I}_{\tau q} \not\succeq \bar{C}^\top L^{-\top}\bar{H}L^{-1}\bar{C}$, then there exists $\bar{w}^*$ such that $\bar{w}^{*\top}(\bar{C}^\top L^{-\top}\bar{H}L^{-1}\bar{C} - (\theta\sigma^2)^{-1}\,\mathrm{I}_{\tau q})\bar{w}^* > 0$, by taking $\alpha\bar{w}^*$ with $\alpha \to +\infty$, the maximization problem in (21) is always infinite, independently of $\bar{u}$. The claim follows by identifying $H = \nabla^2 h(\bar{x}) = \bar{H}$, and $\tilde{C} = L^{-1}\bar{C}$. □

Proof of (iii). If

$$(\theta\sigma^2)^{-1} > \lambda_{\max}(\bar{C}^\top L^{-\top}\bar{H}L^{-1}\bar{C}), \tag{22}$$

i.e., $(\theta\sigma^2)^{-1}\,\mathrm{I}_{\tau q} \succ \bar{C}^\top L^{-\top}\bar{H}L^{-1}\bar{C}$, the maximization problem in (21) is a strongly concave problem in $\bar{w}$ such that the max on $\bar{w}$ is finite. For the dynamic programming resolution, define cost-to-go functions starting from y at time t as

$$\begin{aligned}c_t(y) = \min_{u_t,\dots,u_{\tau-1}} \sup_{\substack{w_t,\dots,w_{\tau-1} \\ x_t,\dots,x_\tau}} & \sum_{s=t}^{\tau}\frac{1}{2}x_s^\top H_s x_s + \tilde{h}_s^\top x_s + \sum_{s=t}^{\tau-1}\frac{1}{2}u_s^\top G_s u_s + \tilde{g}_s^\top u_s - \sum_{s=t}^{\tau-1}\frac{1}{2\theta\sigma^2}\|w_s\|_2^2 \\ \text{subject to} \quad & x_{s+1} = A_s x_s + B_s u_s + C_s w_s \quad \text{for } s = t, \dots, \tau-1 \\ & x_t = y,\end{aligned}$$

with the convention $H_0 = 0$, $\tilde{h}_0 = 0$. Cost-to-go functions satisfy the Bellman equation

$$c_t(y) = \frac{1}{2}y^\top H_t y + \tilde{h}_t^\top y + \min_{u_t\in\mathbb{R}^p}\sup_{w_t\in\mathbb{R}^q}\left\{\frac{1}{2}u_t^\top G_t u_t + \tilde{g}_t^\top u_t - \frac{1}{2\theta\sigma^2}\|w_t\|_2^2 + c_{t+1}(A_t y + B_t u_t + C_t w_t)\right\}, \tag{23}$$

with optimal control

$$u_t^*(y) = \underset{u_t \in \mathbb{R}^p}{\arg\min} \left\{ \frac{1}{2} u_t^\top G_t u_t + \tilde{g}_t^\top u_t + \sup_{w_t \in \mathbb{R}^q} \left\{ -\frac{1}{2\theta\sigma^2} \|w_t\|_2^2 + c_{t+1}(A_t y + B_t u_t + C_t w_t) \right\} \right\},$$

and optimal noise, if the sup is finite,

$$w_t^*(u_t, y) = \underset{w_t \in \mathbb{R}^d}{\arg\max} \left\{ -\frac{1}{2\theta\sigma^2} \|w_t\|_2^2 + c_{t+1}(A_t y + B_t u_t + C_t w_t) \right\}.$$

The final cost initializing the recursion is defined as $c_\tau(y) = \frac{1}{2} y^\top H_\tau y + \tilde{h}_\tau^\top y$. For quadratic costs and linear dynamics, the cost-to-go functions are quadratic and can be computed analytically through the recursive equation (23). If the quadratic defining the supremum problem is not negative semi-definite the problem is infeasible.

If condition (22) holds, the overall maximization is feasible, all supremums are reached. The solution of (6) is given by computing $c_0(\hat{x}_0)$, which amounts to solve iteratively the Bellman equations starting from $x_0 = \hat{x}_0$, i.e., getting the optimal control at the given state and moving along the dynamics to compute the next cost-to-go:

$$u_t^* = u_t^*(x_t), \quad w_t^* = w_t^*(u_t^*, x_t), \quad x_{t+1} = A_t x_t + B_t u_t^* + C_t w_t^*.$$

□

B.2 Dynamic programming resolution

Detailed computations of the dynamic programming approach are given in the following proposition that supports Algo. 1. Though finer sufficient conditions to get a solution can be derived in the case $(\theta\sigma^2)^{-1} = \lambda_{\max}(C_t^\top P_{t+1} C_t)$, simply reducing the risk sensitivity parameter is enough to get the condition in line 5. For simplicity, in Algo. 1, if condition (24) is not satisfied, we consider the problem to be infeasible.

Proposition B.1. *Consider Algo. 1 applied for the linear quadratic risk sensitive control problem (6) with $H_t \succeq 0$ and $G_t \succ 0$. If condition*

$$(\theta\sigma^2)^{-1} > \lambda_{\max}(C_t^\top P_{t+1} C_t) \tag{24}$$

in line 5 is satisfied for all $t = \tau - 1, \dots, 0$, then the cost-to-go functions are quadratics of the form

$$c_t(y) = \frac{1}{2} y^\top P_t y + p_t^\top y + c \quad \textit{with} \quad P_t \succeq 0, \tag{25}$$

where c is a constant and P_t, p_t are defined recursively in line 6.

If for any $t = \tau - 1, \dots, 0$,

$$(\theta\sigma^2)^{-1} < \lambda_{\max}(C_t^\top P_{t+1} C_t),$$

the linear quadratic risk sensitive control problem (6) is infeasible.

Proof. The cost-to-go function at time τ reads $c_\tau(y) = \frac{1}{2} y^\top H_\tau y + \tilde{h}_\tau^\top y$. It has then the form (25) with $p_\tau = \tilde{h}_\tau$ and $P_\tau = H_\tau \succeq 0$. Assume now that at time $t+1$, the cost-to-go function has the form of (25), i.e., $c_{t+1}(y) = \frac{1}{2} y^\top P_{t+1} y + p_{t+1}^\top y$ with $P_{t+1} \succeq 0$. Then, the Bellman equation reads, ignoring the constant terms,

$$\begin{aligned}
c_t(y) = \frac{1}{2} y^\top H_t y + \tilde{h}_t^\top y + \min_{u_t \in \mathbb{R}^p} \sup_{w_t \in \mathbb{R}^q} \Big\{ & \frac{1}{2} u_t^\top G_t u_t + \tilde{g}_t^\top u_t - \frac{1}{2\theta\sigma^2} \|w_t\|_2^2 \\
& + p_{t+1}^\top (A_t y + B_t u_t + C_t w_t) \\
& + \frac{1}{2} (A_t y + B_t u_t + C_t w_t)^\top P_{t+1} (A_t y + B_t u_t + C_t w_t) \Big\} \\
= \frac{1}{2} y^\top H_t y + \tilde{h}_t^\top y + \min_{u_t \in \mathbb{R}^p} \Big\{ & \frac{1}{2} u_t^\top G_t u_t + \tilde{g}_t^\top u_t \\
& + \frac{1}{2} (A_t y + B_t u_t)^\top P_{t+1} (A_t y + B_t u_t) + p_{t+1}^\top (A_t y + B_t u_t) \\
& + \sup_{w_t \in \mathbb{R}^q} \Big[\frac{1}{2} w_t^\top C_t^\top [P_{t+1}(A_t y + B_t u_t) + p_{t+1}] \\
& \qquad - \frac{1}{2} w_t^\top ((\theta\sigma^2)^{-1} \mathrm{I}_q - C_t^\top P_{t+1} C_t) w_t \Big] \Big\}.
\end{aligned}$$

If $(\theta\sigma^2)^{-1} < \lambda_{\max}(C_t^\top P_{t+1} C_t)$, the supremum in w_t is infinite. If $(\theta\sigma^2)^{-1} > \lambda_{\max}(C_t^\top P_{t+1} C_t)$, the supremum is finite and reads

$$w_t^* = ((\theta\sigma^2)^{-1}\,\mathrm{I}_q - C_t^\top P_{t+1} C_t)^{-1} C_t^\top [P_{t+1}(A_t y + B_t u_t) + p_{t+1}]. \tag{26}$$

So we get, ignoring the constant terms,

$$\begin{aligned} c_t(y) = \frac{1}{2} y^\top H_t y + \tilde{h}_t^\top y + \min_{u_t \in \mathbb{R}^p} \Big\{ & \frac{1}{2} u_t^\top G_t u_t + \tilde{g}_t^\top u_t \\ & + \frac{1}{2}(A_t y + B_t u_t)^\top \tilde{P}_{t+1}(A_t y + B_t u_t) + \tilde{p}_{t+1}^\top (A_t y + B_t u_t) \Big\}, \end{aligned} \tag{27}$$

where

$$\begin{aligned} \tilde{P}_{t+1} &= P_{t+1} + P_{t+1} C_t ((\theta\sigma^2)^{-1}\,\mathrm{I}_q - C_t^\top P_{t+1} C_t)^{-1} C_t^\top P_{t+1} \succeq 0 \\ \tilde{p}_{t+1} &= p_{t+1} + P_{t+1} C_t ((\theta\sigma^2)^{-1}\,\mathrm{I}_q - C_t^\top P_{t+1} C_t)^{-1} C_t^\top p_{t+1}. \end{aligned}$$

We then get, ignoring the constant terms,

$$c_t(y) = \frac{1}{2} y^\top (H_t + A_t^\top \tilde{P}_{t+1} A_t) y + (\tilde{h}_t + A_t^\top \rho_t)^\top y - \frac{1}{2} y^\top A_t^\top \tilde{P}_{t+1} B_t (G_t + B_t^\top \tilde{P}_{t+1} B_t)^{-1} B_t^\top \tilde{P}_{t+1} A_t y.$$

where $\rho_t = \tilde{p}_{t+1} - \tilde{P}_{t+1} B_t (G_t + B_t^\top \tilde{P}_{t+1} B_t)^{-1} [B_t^\top \tilde{p}_{t+1} + \tilde{g}_t]$. The cost function is then a quadratic defined by

$$P_t = H_t + A_t^\top \tilde{P}_{t+1} A_t - A_t^\top \tilde{P}_{t+1} B_t (G_t + B_t^\top \tilde{P}_{t+1} B_t)^{-1} B_t^\top \tilde{P}_{t+1} A_t.$$

Denoting $\tilde{P}_{t+1}^{1/2}$ a square root matrix of $\tilde{P}_{t+1}$ such that $\tilde{P}_{t+1}^{1/2} \succeq 0$ and $\tilde{P}_{t+1}^{1/2} \tilde{P}_{t+1}^{1/2} = \tilde{P}_{t+1}$, we get

$$\begin{aligned} P_t &= H_t + A_t^\top \tilde{P}_{t+1}^{1/2} \big(\mathrm{I}_d - \tilde{P}_{t+1}^{1/2} B_t (G_t + B_t^\top \tilde{P}_{t+1} B_t)^{-1} B_t^\top \tilde{P}_{t+1}^{1/2}\big) \tilde{P}_{t+1}^{1/2} A_t \\ &= H_t + A_t^\top \tilde{P}_{t+1}^{1/2} \big(\mathrm{I}_d + \tilde{P}_{t+1}^{1/2} B_t G_t^{-1} B_t^\top \tilde{P}_{t+1}^{1/2}\big)^{-1} \tilde{P}_{t+1}^{1/2} A_t \succeq 0, \end{aligned}$$

where we use Sherman-Morisson-Woodbury formula for the last equality. This proves that $c_t(y)$ satisfies (25) at time t with P_t defined above and

$$p_t = \tilde{h}_t + A_t^\top \Big(\tilde{p}_{t+1} - \tilde{P}_{t+1} B_t (G_t + B_t^\top \tilde{P}_{t+1} B_t)^{-1} [B_t^\top \tilde{p}_{t+1} + \tilde{g}_t] \Big).$$

The optimal control is given from (27) as

$$u_t^*(y) = -(G_t + B_t^\top \tilde{P}_{t+1} B_t)^{-1} [B_t^\top \tilde{P}_{t+1} A_t y + \tilde{g}_t + B_t^\top \tilde{p}_{t+1}]$$

and the optimal noise is given by (26), i.e.,

$$w_t^*(y, u_t) = ((\theta\sigma^2)^{-1}\,\mathrm{I}_q - C_t^\top P_{t+1} C_t)^{-1} C_t^\top [P_{t+1}(A_t y + B_t u_t) + p_{t+1}].$$

□

Remark B.2. *Consider the case $\tilde{h}_t = 0$, $\tilde{g}_t = 0$ such that $\tilde{p}_{t+1} = 0$ and $p_{t+1} = 0$. Then Algorithm 1 is a modified version of the classical Linear Quadratic Regulator (LQR) algorithm where the value function at time $t+1$ is $\tilde{c}_{t+1}(y) = y^\top \tilde{P}_{t+1} y/2$ instead of $c_{t+1}(y) = y^\top P_{t+1} y/2$ for the LQR derivations.*

In particular, denoting $P_{t+1}^{1/2}$ a square root matrix of P_{t+1} and using Sherman-Morisson-Woodbury formula, we have that

$$\begin{aligned} \tilde{P}_{t+1} &= P_{t+1}^{1/2} \big(\mathrm{I}_d - P_{t+1}^{1/2} C_t (C_t^\top P_{t+1} C_t - (\theta\sigma^2)^{-1}\,\mathrm{I}_d)^{-1} C_t^\top P_{t+1}^{1/2}\big) P_{t+1}^{1/2} \\ &= P_{t+1}^{1/2} (\mathrm{I}_d - \theta\sigma^2 P_{t+1}^{1/2} C_t C_t^\top P_{t+1}^{1/2})^{-1} P_{t+1}^{1/2} \end{aligned}$$

such that for $\theta = 0$ we get $\tilde{P}_{t+1} = P_{t+1}$, so we retrieve the minimization of a Linear Quadratic Gaussian control problem by dynamic programming.

C Iterative linearized algorithms

C.1 Model minimization

We present the implementation of RegILEQG for general noisy dynamics of the form

$$x_{t+1} = \psi_t(x_t, u_t, w_t). \tag{28}$$

We define the trajectory as a function $\tilde{x} : \mathbb{R}^{\tau p \times \tau q} \to \mathbb{R}^{\tau d}$ of the control and noise variables decomposed as $\tilde{x}(\bar{u}, \bar{w}) = (\tilde{x}_1(\bar{u}, \bar{w}); \dots; \tilde{x}_\tau(\bar{u}, \bar{w}))$ where

$$\tilde{x}_1(\bar{u}, \bar{w}) = \psi_0(\hat{x}_0, u_0, w_0), \quad \tilde{x}_{t+1}(\bar{x}, \bar{w}) = \psi_t(\tilde{x}_t(\bar{u}, \bar{w}), u_t, w_t) \tag{29}$$

The risk sensitive objective (3) can be written

$$\min_{\bar{u} \in \mathbb{R}^{\tau p}} f_\theta(\bar{u}) = \eta_\theta(\bar{u}) + g(\bar{u}) \quad \text{where} \quad \eta_\theta(\bar{u}) = \frac{1}{\theta} \log \mathbb{E}_{\bar{w}} \Big[\exp \theta h\big(\tilde{x}(\bar{u}, \bar{w})\big) \Big]. \tag{30}$$

The model we consider for the trajectory reads

$$\tilde{x}(\bar{u} + \bar{v}, \bar{w}) \approx \tilde{x}(\bar{u}, 0) + \nabla \tilde{x}(\bar{u}, 0)^\top (\bar{v}, \bar{w}) = \tilde{x}(\bar{u}, 0) + \nabla_{\bar{u}} \tilde{x}(\bar{u}, 0)^\top \bar{v} + \nabla_{\bar{w}} \tilde{x}(\bar{u}, 0)^\top \bar{w}, \tag{31}$$

where $\tilde{x}(\bar{u}, 0)$ is the exact trajectory, $\nabla_{\bar{u}} \tilde{x}$ and $\nabla_{\bar{w}} \tilde{x}$ denote the gradient w.r.t. the command and the noise, respectively, see Appendix A for gradient notations.

We approximate the objective as $f_\theta(\bar{u} + \bar{v}) \approx m_{f_\theta}(\bar{u} + \bar{v}; \bar{u})$, where

$$m_{f_\theta}(\bar{u} + \bar{v}; \bar{u}) \triangleq \frac{1}{\theta} \log \mathbb{E}_{\bar{w}} \Big[\exp \theta q_h\big(\bar{x} + \nabla_{\bar{u}} \tilde{x}(\bar{u}, 0)^\top \bar{v} + \nabla_{\bar{w}} \tilde{x}(\bar{u}, 0)^\top \bar{w}; \bar{x}\big) \Big] + q_g(\bar{u} + \bar{v}; \bar{u}) \tag{32}$$

where $q_h(\bar{x} + \bar{y}; \bar{x}) \triangleq h(\bar{x}) + \nabla h(\bar{x})^\top \bar{y} + \bar{y}^\top \nabla^2 h(\bar{x}) \bar{y}/2$, $q_g(\bar{u} + \bar{v}; \bar{u})$ is defined similarly and $\bar{x} = \tilde{x}(\bar{u}, 0)$ is the exact trajectory.

This model is then minimized with an additional proximal term. Formally, the algorithm starts at a point $\bar{u}_0$ and defines the next iterate as

$$\bar{u}^{(k+1)} = \bar{u}^{(k)} + \operatorname*{arg\,min}_{\bar{v} \in \mathbb{R}^{\tau p}} \left\{ m_{f_\theta}(\bar{u}^{(k)} + \bar{v}; \bar{u}^{(k)}) + \frac{1}{2\gamma_k} \|\bar{v}\|_2^2 \right\} \tag{33}$$

where γ_k is the step-size: the smaller γ_k is, the closer the solution is to the current iterate.

The following proposition shows that the minimization step (33) amounts to a linear quadratic risk-sensitive control problem. Prop. 2.1 is then a sub-case of the following proposition.

Proposition C.1. *The model minimization step* (33) *is given as* $\bar{u}^{(k+1)} = \bar{u}^{(k)} + \bar{v}^*$ *where* $\bar{v}^*$ *is the solution of*

$$\begin{aligned} \min_{\bar{v} \in \mathbb{R}^{\tau p}} \sup_{\bar{w} \in \mathbb{R}^{\tau q}\, \bar{y} \in \mathbb{R}^{\tau d}} \quad & \sum_{t=1}^{\tau} \left(\frac{1}{2} y_t^\top H_t y_t + \tilde{h}_t^\top y_t \right) + \sum_{t=0}^{\tau-1} \left(\frac{1}{2} v_t^\top (G_t + \gamma_k^{-1} \mathrm{I}_p) v_t + \tilde{g}_t^\top v_t \right) - \sum_{t=0}^{\tau-1} \frac{1}{2\theta\sigma^2} \|w_t\|_2^2 \\ \text{subject to} \quad & y_{t+1} = A_t y_t + B_t v_t + C_t w_t \\ & y_0 = 0, \end{aligned} \tag{34}$$

where $x_t^{(k)} = \tilde{x}_t(\bar{u}^{(k)}, 0)$, $A_t = \nabla_x \psi_t(x_t^{(k)}, u_t^{(k)}, 0)^\top$, $B_t = \nabla_u \psi_t(x_t^{(k)}, u_t^{(k)}, 0)^\top$, $C_t = \nabla_w \psi_t(x_t^{(k)}, u_t^{(k)}, 0)^\top$, $H_t = \nabla^2 h_t(x_t^{(k)})$, $\tilde{h}_t = \nabla h_t(x_t^{(k)})$, $G_t = \nabla^2 g_t(u_t^{(k)})$, $\tilde{g}_t = \nabla g_t(u_t^{(k)})$.

Proof. To ease notations denote $\bar{u}^{(k)} = \bar{u}$. Recall that the trajectory defined by $\bar{u}, \bar{w}$ reads

$$\tilde{x}_1(\bar{u}, \bar{w}) = \psi_0(\hat{x}_0, F_0^\top \bar{u}, E_0^\top \bar{w}), \qquad \tilde{x}_{t+1}(\bar{u}, \bar{w}) = \psi_t(\tilde{x}_t(\bar{u}, \bar{w}), F_t^\top \bar{u}, E_t^\top \bar{w})$$

where $F_t = e_{t+1} \otimes \mathrm{I}_p \in \mathbb{R}^{\tau p \times p}$ satisfies $F_t^\top \bar{u} = u_t$, $E_t = e_{t+1} \otimes \mathrm{I}_q \in \mathbb{R}^{\tau q \times q}$ satisfies $E_t^\top \bar{w} = w_t$ and $e_t \in \mathbb{R}^\tau$ is the t^{th} canonical vector in $\mathbb{R}^\tau$. The gradient is then given by

$$\begin{aligned} \nabla \tilde{x}_1(\bar{u}, \bar{w}) &= \begin{pmatrix} F_0 \nabla_u \psi_0(\hat{x}_0, u_0, w_0) \\ E_0 \nabla_w \psi_0(\hat{x}_0, u_0, w_0) \end{pmatrix} \\ \nabla \tilde{x}_{t+1}(\bar{u}, \bar{w}) &= \nabla \tilde{x}_t(\bar{u}, \bar{w}) \nabla_x \psi_t(\tilde{x}_t(\bar{u}, \bar{w}), u_t, w_t) + \begin{pmatrix} F_t \nabla_u \psi_t(\tilde{x}_t(\bar{u}, \bar{w}), u_t, w_t) \\ E_t \nabla_w \psi_t(\tilde{x}_t(\bar{u}, \bar{w}), u_t, w_t) \end{pmatrix} \end{aligned}$$

For a given $\bar{v} = (v_0; \dots; v_{\tau-1})$, the product $\bar{y} = (y_1; \dots; y_\tau) = \nabla \tilde{x}(\bar{u}, 0)^\top (\bar{v}, \bar{w})$ reads

$$\begin{aligned} y_1 &= \nabla_u \psi_0(x_0, u_0, 0)^\top v_0 + \nabla_w \psi_0(x_0, u_0, 0)^\top w_0 \\ y_{t+1} &= \nabla_x \psi_t(x_t, u_t, 0)^\top y_t + \nabla_u \psi_t(x_t, u_t, 0)^\top v_t + \nabla_w \psi_t(x_t, u_t, 0)^\top w_t, \end{aligned}$$

where $x_t = \tilde{x}_t(\bar{u}, 0)$, $x_0 = \hat{x}_0$ and we used that $y_t = \nabla \tilde{x}_t(\bar{u}, 0)^\top (\bar{v}, \bar{w})$.

The approximate state objective inside the exponential in (32) reads then

$$\begin{aligned} q_h\big(\bar{x} + \nabla_{\bar{u}} \tilde{x}(\bar{u}, 0)^\top \bar{v} + \nabla_{\bar{w}} \tilde{x}(\bar{u}, 0)^\top \bar{w}; \bar{x}\big) = & \sum_{t=1}^{\tau} q_{h_t}(x_t + y_t; x_t) \\ \text{s.t.} \quad & y_{t+1} = A_t y_t + B_t v_t + C_t w_t \\ & y_0 = 0, \end{aligned}$$

where $A_t = \nabla_x \psi_t(x_t, u_t, 0)^\top, B_t = \nabla_u \psi_t(x_t, u_t, 0)^\top, C_t = \nabla_w \psi_t(x_t, u_t, 0)^\top$. We retrieve the model of a linear quadratic control problem perturbed by noise $\bar{w}$. The risk sensitive objective can then be decomposed as in Proposition 1.1, leading to the claimed formulation. □

C.2 ILEQG and RegILEQG implementations

C.2.1 Implementations by dynamic programming

We present in Algo. 2 the Regularized variant of ILEQG that calls Algo. 1 at each step to solve the linear quadratic problem by dynamic programming. We present it for constant step-size. Line-searches are left for future work. We also present in Algo. 3 the classical ILEQG method equipped with a line-search on the Monte-Carlo approximation of the objective. Implementations proposed in [Farshidian and Buchli, 2015] or [Ponton et al., 2016] do not mention the choice of the line-search.

C.2.2 Implementation by automatic differentiation

We consider here problems whose objective rely only in the last state, i.e.

$$h(\bar{x}) = h_\tau(x_\tau), \tag{35}$$

and assume h_τ strictly convex. In that case we can use automatic differentiation oracles as defined in [Roulet et al., 2019] and recalled below.

Definition C.2 (Automatic-differentiation oracle)**.** *Let $\tilde{x}_\tau : \mathbb{R}^{\tau\pi} \to \mathbb{R}^d$ be a chain of compositions defined by*

$$x_0 = \hat{x}_0, \qquad x_{t+1} = \psi(x_t, \omega_t) \qquad \text{for } t \in \{0, \dots, \tau - 1\}$$

for differentiable functions $\psi_t : \mathbb{R}^d \times \mathbb{R}^\pi$, $\hat{x}_0 \in \mathbb{R}^d$ An automatic-differentiation oracle *is any procedure that computes $\nabla \tilde{x}_\tau(\bar{\omega}) z$ for any $\bar{\omega} = (\omega_0, \dots, \omega_{\tau-1}) \in \mathbb{R}^{\tau\pi}$, $z \in \mathbb{R}^d$.*

We can then use the dual optimization problem of (33) as shown in the following proposition. For final-state cost (35), the automatic differentiation implementation is computationally less expensive than a dynamic programming approach whose naive implementation requires the inversion of multiple matrices. The detailed implementation by automatic-differentiation oracle is provided in Algo. 4.

Proposition C.3. *Consider the model minimization subproblem (33) for strictly convex last state cost (35) and notations defined in Prop. C.1. If $\nabla^2 h_\tau(x_\tau^{(k)})^{-1} \succ \theta\sigma^2 \nabla_{\bar{w}} \tilde{x}_\tau(\bar{u}^{(k)}, 0)^\top \nabla_{\bar{w}} \tilde{x}_\tau(\bar{u}^{(k)}, 0)$, then*

(i) the dual of subproblem (34) reads

$$\min_{z \in \mathbb{R}^d} \tilde{q}^*_{h_\tau}(z) + \tilde{q}^*_g(-\nabla_{\bar{u}} \tilde{x}_\tau(\bar{u}^{(k)}, 0) z) - \frac{\theta\sigma^2}{2} \|\nabla_{\bar{w}} \tilde{x}_\tau(\bar{u}^{(k)}, 0) z\|_2^2, \tag{36}$$

where $\tilde{q}_{h_\tau}(y) = \frac{1}{2} y_\tau^\top H_\tau y_\tau + \tilde{h}_\tau^\top y_\tau$, $\tilde{q}_g(\bar{v}) = \frac{1}{2} \bar{v}^\top (\bar{G} + \gamma_k^{-1} \mathrm{I}_{\tau p}) \bar{v} + \tilde{g}^\top \bar{v}$, $\bar{G} = \mathbf{diag}(G_0, \dots, G_{\tau-1})$, $\tilde{g} = (\tilde{g}_0, \dots, \tilde{g}_{\tau-1})$ and for a function f, we denote by f^* its convex conjugate,

(ii) the model minimization step is then given as $\bar{u}^{(k+1)} = \bar{u}^{(k)} + \nabla \tilde{q}^*_g(-\nabla_{\bar{u}} \tilde{x}(\bar{u}^{(k)}, 0) z^*)$, where z^* is solution of (36),

(iii) the model minimization step makes $10d+1$ calls to an automatic differentiation oracle defined in Def. C.2 by using a conjugate gradient method to solve (36).

Proof. To ease notations denote $\bar{u}^{(k)} = \bar{u}$. Denoting $\tilde{A} = \nabla_{\bar{u}} \tilde{x}_\tau(\bar{u}, 0)^\top$, $\tilde{B} = \nabla_{\bar{w}} \tilde{x}_\tau(\bar{u}, 0)^\top$, $\tilde{q}_{h_\tau}(y) = \frac{1}{2} y_\tau^\top H_\tau y_\tau + \tilde{h}_\tau^\top y_\tau$, $\tilde{q}_g(\bar{v}) = \frac{1}{2} \bar{v}^\top (\bar{G} + \gamma_k^{-1} \mathrm{I}_{\tau p}) \bar{v} + \tilde{g}^\top \bar{v}$, $\bar{G} = \mathbf{diag}(G_0, \dots, G_{\tau-1})$, $\tilde{g} = (\tilde{g}_0, \dots, \tilde{g}_{\tau-1})$, the model minimization subproblem (34) for last state cost (35) reads

$$\begin{aligned}
&\min_{\bar{v} \in \mathbb{R}^{\tau p}} \sup_{\bar{w} \in \mathbb{R}^{\tau q}} \quad \tilde{q}_g(\bar{v}) + \tilde{q}_{h_\tau}(\tilde{A}\bar{v} + \tilde{B}\bar{w}) - \frac{1}{2\theta\sigma^2} \|\bar{w}\|_2^2 \\
&= \min_{\bar{v} \in \mathbb{R}^{\tau p}} \tilde{q}_g(\bar{v}) + \sup_{\bar{w} \in \mathbb{R}^{\tau q}} \sup_{z \in \mathbb{R}^d} z^\top (\tilde{A}\bar{v} + \tilde{B}\bar{w}) - \tilde{q}^*_{h_\tau}(z) - \frac{1}{2\theta\sigma^2} \|\bar{w}\|_2^2 \\
&= \min_{\bar{v} \in \mathbb{R}^{\tau p}} \sup_{z \in \mathbb{R}^d} \tilde{q}_g(\bar{v}) + z^\top \tilde{A}\bar{v} - \tilde{q}^*_{h_\tau}(z) + \frac{\theta\sigma^2}{2} \|\tilde{B}^\top z\|_2^2.
\end{aligned} \tag{37}$$

Recall that for a function $f(x) = x^\top q + x^\top Qx/2$ with $Q \succ 0$, we have $f^*(z) = \sup_x \{z^\top x - f(x)\} = (z-q)^\top Q^{-1}(z-q)/2$. If $H_\tau^{-1} \not\succeq \theta\sigma^2 \tilde{B}\tilde{B}^\top$ the supremum in z is infinite. If $H_\tau^{-1} \succ \theta\sigma^2 \tilde{B}\tilde{B}^\top$, the supremum in z is finite. The problem is then a strongly convex-concave problem such that min and max can be inverted leading to the dual problem

$$\max_{z \in \mathbb{R}^d} -\tilde{q}^*_{h_\tau}(z) - \tilde{q}^*_g(-\tilde{A}^\top z) + \frac{\theta\sigma^2}{2}\|\tilde{B}^\top z\|_2^2.$$

The primal solution is obtained from a dual solution z^* by the mapping $\bar{v}^* = \nabla \tilde{q}^*_g(-\tilde{A}^\top z^*)$ obtained from (37).

The dual problem (36) is a quadratic problem, which can then be solved in d iterations by a conjugate gradients method. The gradients of $z \to \tilde{q}^*_g(-\nabla_{\bar{u}}\tilde{x}(\bar{u}^{(k)}, 0)z)$ and $z \to \frac{\theta\sigma^2}{2}\|\nabla_{\bar{w}}\tilde{x}_\tau(\bar{u}^{(k)}, 0)z\|_2^2$ can be computed by an automatic differentiation procedure defined in C.2. Each gradient computation requires the equivalent of two calls to an automatic differentiation oracle as detailed in [Roulet et al., 2019]. The mapping to the primal solution costs an additional call. Finally, checking if the problem is feasible requires to compute the Hessian of $z \to \tilde{q}^*_{h_\tau}(z) - \frac{\theta\sigma^2}{2}\|\tilde{B}^\top z\|_2^2$ which costs $4d$ additional calls (each call computes the second order derivative with respect to a given coordinate in $\mathbb{R}^d$ and computing the second order derivative amounts to back-propagate through the computation of the gradient of $z \to \frac{\theta\sigma^2}{2}\|\tilde{B}^\top z\|_2^2$ which itself cost 2 calls to an automatic differentiation procedure). □

We detail the complete implementation by automatic differentiation in Algo. 4. We assume that we have access to a conjugate gradients method `conjgrad` for quadratic problems of the form

$$\min_{z \in \mathbb{R}^d} f(z) = \frac{1}{2} z^\top A z + b^\top z,$$

with $A \succ 0$, that given an oracle on the gradient of f outputs the solution of the quadratic problem. Formally, it reads $\texttt{conjgrad}(\nabla f) = \arg\min_{z \in \mathbb{R}d} f(z)$. This can be implemented following Nesterov [2013].

Algorithm 1 Dynamic programming for Linear Quadratic Exponential Gaussian (LEQG) (6)

Inputs: Initial state $\hat{x}_0$, risk-sensitivity parameter θ, variance σ^2, convex quadratic costs $H_t \succeq 0, \tilde{h}_t$, strictly convex quadratic costs $G_t \succ 0, \tilde{g}_t$, linear dynamics A_t, B_t, C_t

Backward pass:

Initialize $P_\tau = H_\tau$, $p_\tau = \tilde{h}_\tau$, `feasible` = True

for $t = \tau - 1, \ldots, 0$ **do**

  if $(\theta\sigma^2)^{-1} > \lambda_{\max}(C_t^\top P_{t+1} C_t)$ **then**

    Compute

$$\tilde{P}_{t+1} = P_{t+1} + P_{t+1}C_t((\theta\sigma^2)^{-1}\,\mathrm{I}_q - C_t^\top P_{t+1}C_t)^{-1}C_t^\top P_{t+1} \tag{38}$$

$$\tilde{p}_{t+1} = p_{t+1} + P_{t+1}C_t((\theta\sigma^2)^{-1}\,\mathrm{I}_q - C_t^\top P_{t+1}C_t)^{-1}C_t^\top p_{t+1} \tag{39}$$

$$P_t = H_t + A_t^\top \tilde{P}_{t+1}A_t - A_t^\top \tilde{P}_{t+1}B_t(G_t + B_t^\top \tilde{P}_{t+1}B_t)^{-1}B_t^\top \tilde{P}_{t+1}A_t \tag{40}$$

$$p_t = \tilde{h}_t + A_t^\top\left[\tilde{p}_{t+1} - \tilde{P}_{t+1}B_t(G_t + B_t^\top \tilde{P}_{t+1}B_t)^{-1}[B_t^\top \tilde{p}_{t+1} + \tilde{g}_t]\right] \tag{41}$$

    Store

$$K_t = -(G_t + B_t^\top \tilde{P}_{t+1}B_t)^{-1}B_t^\top \tilde{P}_{t+1}A_t \qquad L_t^x = ((\theta\sigma^2)^{-1}\,\mathrm{I}_q - C_t^\top P_{t+1}C_t)^{-1}C_t^\top P_{t+1}A_t$$

$$k_t = -(G_t + B_t^\top \tilde{P}_{t+1}B_t)^{-1}(\tilde{g}_t + B_t^\top \tilde{p}_{t+1}) \qquad L_t^u = ((\theta\sigma^2)^{-1}\,\mathrm{I}_q - C_t^\top P_{t+1}C_t)^{-1}C_t^\top P_{t+1}B_t$$

$$l_t = ((\theta\sigma^2)^{-1}\,\mathrm{I}_q - C_t^\top P_{t+1}C_t)^{-1}C_t^\top p_{t+1}$$

  else

    State `feasible` = False

    break

  end if

end for

Rollout phase:

if `feasible` **then**

  Initialize $x_0 = \hat{x}_0$

  for $t = 0, \ldots, \tau - 1$ **do**

    Compute

$$u_t^* = K_t x_t + k_t \qquad w_t^* = L_t^x x_t + L_t^u u_t^* + l_t \tag{42}$$

$$x_{t+1} = A_t x_t + B_t u_t^* + C_t w_t^* \tag{43}$$

  end for

else

  $u_t^* =$ None for all t

end if

Output: $\bar{u}^* = (u_0^*; \ldots; u_{\tau-1}^*)$

Algorithm 2 Regularized Iterative Linear Exponential Quadratic Gaussian (RegILEQG)

Inputs: Initial state $\hat{x}_0$, risk sensitive parameter θ, variance σ^2, fixed step-size γ, initial command $\bar{u}^{(0)}$, number of iterations K, convex costs h_t, g_t, dynamics ψ_t

for $k = 0, \dots, K$ **do**

Forward pass

Compute along the exact trajectory $\bar{x}^{(k)} = \tilde{x}(\bar{u}^{(k)}, 0)$ defined by $\bar{u}^{(k)}$,

$$H_t = \nabla^2 h_t(x_t^{(k)}) \quad \tilde{h}_t = \nabla h_t(x_t^{(k)}) \quad G_t = \nabla^2 g_t(u_t^{(k)}) \quad \tilde{g}_t = \nabla g_t(u_t^{(k)})$$
$$A_t = \nabla_x \psi_t(x_t^{(k)}, u_t^{(k)}, 0)^\top \quad B_t = \nabla_u \psi_t(x_t^{(k)}, u_t^{(k)}, 0)^\top \quad C_t = \nabla_w \psi_t(x_t^{(k)}, u_t^{(k)}, 0)^\top$$

Backward pass

Apply Algo. 1 to

$$\begin{aligned} \min_{\bar{v} \in \mathbb{R}^{\tau p}} \sup_{\bar{w} \in \mathbb{R}^{\tau d}} \quad & \sum_{t=1}^{\tau} \left(\frac{1}{2} y_t^\top H_t y_t + \tilde{h}_t^\top y_t \right) + \sum_{t=0}^{\tau-1} \left(\frac{1}{2} v_t^\top (G_t + \gamma^{-1} \mathrm{I}_p) v_t + \tilde{g}_t^\top v_t \right) - \sum_{t=0}^{\tau-1} \frac{1}{2\theta\sigma^2} \|w_t\|_2^2 \\ \text{subject to} \quad & y_{t+1} = A_t y_t + B_t v_t + C_t w_t \\ & y_0 = 0. \end{aligned}$$

if Algo. 1 cannot output a solution $\bar{v}^*$ **then**

State `feasible` = False

break

else

Update $\bar{u}^{(k+1)} = \bar{u}^{(k)} + \bar{v}^*$, with $\bar{v}^*$ found by Algo. 1

end if

end for

Output: $\bar{u}^{(K)}$ if `feasible` or last iterate $\bar{u}^{(k)}$ if not `feasible`

Algorithm 3 Iterative Linear Exponential Quadratic Gaussian (ILEQG) (7)

Inputs: Initial state $\hat{x}_0$, risk sensitive parameter θ, variance σ^2, initial command $\bar{u}^{(0)}$, number of iterations K, convex costs h_t, g_t, dynamics ψ_t, line-search precision ϵ,
for $k = 0, \dots, K$ **do**
    Forward pass
    Compute along the exact trajectory $\bar{x}^{(k)} = \tilde{x}(\bar{u}^{(k)}, 0)$ defined by $\bar{u}^{(k)}$,

$$H_t = \nabla^2 h_t(x_t^{(k)}) \quad \tilde{h}_t = \nabla h_t(x_t^{(k)}) \quad G_t = \nabla^2 g_t(u_t^{(k)}) \quad \tilde{g}_t = \nabla g_t(u_t^{(k)})$$
$$A_t = \nabla_x \psi_t(x_t^{(k)}, u_t^{(k)}, 0)^\top \quad B_t = \nabla_u \psi_t(x_t^{(k)}, u_t^{(k)}, 0)^\top \quad C_t = \nabla_w \psi_t(x_t^{(k)}, u_t^{(k)}, 0)^\top$$

    Backward pass
    Apply Algo. 1 to

$$\begin{aligned} \min_{\bar{v} \in \mathbb{R}^{\tau p}} \sup_{\bar{w} \in \mathbb{R}^{\tau d}} \quad & \sum_{t=1}^{\tau} \left(\frac{1}{2} y_t^\top H_t y_t + \tilde{h}_t^\top y_t \right) + \sum_{t=0}^{\tau-1} \left(\frac{1}{2} v_t^\top G_t v_t + \tilde{g}_t^\top v_t \right) - \sum_{t=0}^{\tau-1} \frac{1}{2\theta\sigma^2} \|w_t\|_2^2 \\ \text{subject to} \quad & y_{t+1} = A_t y_t + B_t v_t + C_t w_t \\ & y_0 = 0. \end{aligned}$$

    if Algo. 1 cannot output a solution $\bar{v}^*$ **then**
        State `feasible` = False
        break
    else
        Find $\alpha > 0$ such that $\bar{u}^{(k+1)} = \bar{u}^{(k)} + \alpha \bar{v}^*$, with $\bar{v}^*$ found by Algo. 1, satisfies

$$\tilde{f}_\theta(\bar{u}^{(k+1)}) \leq \tilde{f}_\theta(\bar{u}^k) + \epsilon$$

        where $\tilde{f}_\theta(\bar{u})$ is the Monte-Carlo approximation of the risk-sensitive loss.
    end if
end for
Output: $\bar{u}^{(K)}$ if `feasible` or last iterate $\bar{u}^{(k)}$ if not `feasible`

Algorithm 4 RegILEQG by automatic differentiation for final-state cost (35)

Inputs: Initial state $\hat{x}_0$, risk sensitive parameter θ, variance σ^2, step-size γ, initial command $\bar{u}^{(0)}$, number of iterations K, convex costs g_t, final strictly convex cost h_τ, dynamics ψ_t.
for $k = 0, \dots, K$ **do**
  Forward pass
  Compute $\bar{x}^{(k)} = \tilde{x}(\bar{u}^{(k)}, 0)$ along the trajectory
  Store $\nabla\psi_t(x_t^{(k)}, u_t^{(k)}, 0))$ to compute any $\nabla_{\bar{u}}\tilde{x}(\bar{u}^{(k)}, 0)z$ or $\nabla_{\bar{w}}\tilde{x}(\bar{u}^{(k)}, 0)z$ by automatic-differentiation
  Dual problem definition
  Compute $H_\tau = \nabla^2 h_\tau(\bar{x}_\tau^{(k)}), h_\tau = \nabla h(\bar{x}_\tau^{(k)})$, $G_t = \nabla^2 g_t(u_t^{(k)})$, $\tilde{g}_t = \nabla g_t(u_t^{(k)})$
  Define $\tilde{q}^*_{h_\tau} : z \to \frac{1}{2}(z - \tilde{h}_\tau)^\top H_\tau^{-1}(z - \tilde{h}_\tau)$
  Define $\tilde{q}^*_g : \bar{\zeta} \to \frac{1}{2}(\bar{\zeta} - \tilde{g})^\top(\bar{G} + \gamma_k^{-1}\,\mathrm{I}_{\tau p})(\bar{\zeta} - \tilde{g})$ where $\bar{G} = \mathbf{diag}(G_0, \dots, G_{\tau-1})$, $\tilde{g} = (g_0; \dots; g_{\tau-1})$.
  Define $\nabla\tilde{q}^*_g : \bar{\zeta} \to (\bar{G} + \gamma_k^{-1}\,\mathrm{I}_{\tau p})(\bar{\zeta} - \tilde{g})$
  Define

$$f : z \to \tilde{q}^*_{h_\tau}(z) + \tilde{q}^*_g(-\nabla_{\bar{u}}\tilde{x}_\tau(\bar{u}^{(k)}, 0)z) - \frac{\theta}{2}\|\nabla_{\bar{w}}\tilde{x}_\tau(\bar{u}^{(k)}, 0)z\|_2^2$$

  where $\nabla_{\bar{u}}\tilde{x}_\tau(\bar{u}^{(k)}, 0)z$ and $\nabla_{\bar{w}}\tilde{x}_\tau(\bar{u}^{(k)}, 0)z$ are computed by automatic differentiation.
  Resolution
  Define $r : z \to q^*_{h_\tau}(z) - \frac{\theta}{2}\|\nabla_{\bar{w}}\tilde{x}_\tau(\bar{u}^{(k)}, 0)z\|_2^2$
  Compute $\nabla^2 r(z)$ for e.g. $z = 0$
  if $\nabla^2 r(z) \not\succ 0$ **then**
    State `feasible` = False and **break**
  else
    Compute $z^* = \texttt{conjgrad}(\nabla f) = \arg\min_{z\in\mathbb{R}^d} f(z)$ where ∇f is provided by automatic differentiation.
    Map to primal solution $\bar{u}^{(k+1)} = \bar{u}^{(k)} + \nabla\tilde{q}^*_g(-\nabla_{\bar{u}}\tilde{x}(\bar{u}^{(k)}, 0)z^*)$.
  end if
end for
Output: $\bar{u}^{(K)}$ or last iterate $\bar{u}^{(k)}$ if not `feasible`

D Convergence analysis proofs

D.1 Risk-sensitive gradient

We recall the derivation of a risk-sensitive objective below. The proof follows from standard derivations.

Proposition D.1. *Given a differentiable function $f : \mathbb{R}^{\tau p+\tau q} \to \mathbb{R}$, define*

$$F : \bar{u} \to \frac{1}{\theta} \log \mathbb{E}_{\bar{w}\sim\mathcal{N}(0,\sigma^2 \mathrm{I}_{\tau q})} \exp(\theta f(\bar{u}, \bar{w})).$$

Then for $\bar{u} \in \mathbb{R}^{\tau p}$ such that $F(\bar{u}) < +\infty$,

$$\nabla F(\bar{u}) = \frac{\mathbb{E}_{\bar{w}\sim\mathcal{N}(0,\sigma^2 \mathrm{I}_{\tau q})} \exp(\theta f(\bar{u}, \bar{w}))\nabla_{\bar{u}} f(\bar{u}, \bar{w})}{\mathbb{E}_{\bar{w}\sim\mathcal{N}(0,\sigma^2 \mathrm{I}_{\tau q})} \exp(\theta f(\bar{u}, \bar{w}))} = \mathbb{E}_{\bar{w}\sim p(\cdot;\bar{u})} \nabla_{\bar{u}} f(\bar{u}, \bar{w}),$$

where

$$p(\bar{w}; \bar{u}) = \exp\left(\theta f(\bar{u}, \bar{w}) - \frac{1}{2\sigma^2}\|\bar{w}\|_2^2 - \theta F(\bar{u})\right).$$

D.2 Approximated risk-sensitive objective

We study the approximated risk-sensitive objective, its truncated gradient and the link with ILEQG in the following propositions. Note that those results also hold for non-quadratic costs by considering

$$\tilde{\eta}_\theta(\bar{u}) = \frac{1}{\theta} \log \mathbb{E}_{\bar{w}} \exp[\theta q_h(\tilde{x}(\bar{u}) + \nabla\tilde{x}(\bar{u})^\top \bar{w}; \tilde{x}(\bar{u}))].$$

in place of $\hat{\eta}_\theta$ and

$$\widetilde{\nabla}\tilde{\eta}_\theta(\bar{u}) = \mathbb{E}_{\bar{w}\sim\tilde{p}(\cdot;\bar{u})} \nabla\tilde{x}(\bar{u})\nabla q_h(\tilde{x}(\bar{u}) + \nabla\tilde{x}(\bar{u})^\top \bar{w}; \tilde{x}(\bar{u}))$$

in place of $\widehat{\nabla}\hat{\eta}_\theta(\bar{u})$ where

$$\tilde{p}(\bar{w}; \bar{u}) = \exp\left(\theta q_h(\tilde{x}(\bar{u}) + \nabla\tilde{x}(\bar{u})^\top \bar{w}; \tilde{x}(\bar{u})) - \frac{1}{2\sigma^2}\|\bar{w}\|_2^2 - \theta\tilde{\eta}_\theta(\bar{u})\right)$$

Precisely, the approximated risk-sensitive loss $\tilde{\eta}_\theta(\bar{u})$ is defined if condition (14) holds, the probability distribution and the expression are the same. Prop. 2.3 is valid by replacing $\widehat{\nabla}\hat{\eta}_\theta(\bar{u})$ by $\widetilde{\nabla}\tilde{\eta}_\theta(\bar{u})$.

Proposition 2.2. *For $\bar{u} \in \mathbb{R}^{\tau p}$ with $\bar{x} = \tilde{x}(\bar{u})$, if*

$$\sigma^{-2} \mathrm{I}_{\tau p} \succ \theta\nabla\tilde{x}(\bar{u})\nabla^2 h(\bar{x})\nabla\tilde{x}(\bar{u})^\top, \tag{14}$$

the approximated risk sensitive cost is defined and is the scaled log-partition function of

$$\hat{p}(\bar{w}; \bar{u}) = \exp\left(\theta h(\tilde{x}(\bar{u}) + \nabla\tilde{x}(\bar{u})^\top \bar{w}) - \frac{1}{2\sigma^2}\|\bar{w}\|_2^2 - \theta\hat{\eta}_\theta(\bar{u})\right), \tag{15}$$

which is the density of a Gaussian $\mathcal{N}(\bar{w}_*, \Sigma)$ with

$$\bar{w}_* = \theta\Sigma X\tilde{h}, \qquad \Sigma = (\sigma^{-2} \mathrm{I}_{\tau p} - \theta X H X^\top)^{-1}, \tag{16}$$

where $X = \nabla\tilde{x}(\bar{u})$, $\tilde{h} = \nabla h(\bar{x}), H = \nabla^2 h(\bar{x})$ and $\bar{x} = \tilde{x}(\bar{u})$. Therefore, the approximated risk-sensitive loss can be computed analytically.

Proof. For $\bar{u} \in \mathbb{R}^{\tau p}$, since h is quadratic and $\bar{w} \to \theta h(\tilde{x}(\bar{u}) + \nabla\tilde{x}(\bar{u})^\top \bar{w}) - \|\bar{w}\|_2^2/2\sigma^2$ is strongly concave, the function $p(\cdot; \bar{u})$ is the density of a Gaussian where $\theta\hat{\eta}(\bar{u})$ is its log-partition function. It can be factorized as follows using $h(\bar{x} + \bar{y}) = h(\bar{x}) + \nabla h(\bar{x})^\top \bar{y} + \frac{1}{2}\bar{y}^\top \nabla^2 h(\bar{x})\bar{y}$ and denoting $X = \nabla\tilde{x}(\bar{u}), \tilde{h} = \nabla h(\bar{x}), H = \nabla^2 h(\bar{x}), \bar{x} = \tilde{x}(\bar{u})$,

$$\begin{aligned}\theta h(\bar{x} + \nabla\tilde{x}(\bar{u})^\top \bar{w}) - \frac{1}{2\sigma^2}\|\bar{w}\|_2^2 =& \theta h(\bar{x}) + \theta(X\tilde{h})^\top \bar{w} + \frac{\theta}{2}\bar{w}^\top X H X^\top \bar{w} - \frac{1}{2\sigma^2}\|\bar{w}\|_2^2 \\ =& \theta h(\bar{x}) - \frac{1}{2}(\bar{w} - \bar{w}_*)^\top \Sigma^{-1}(\bar{w} - \bar{w}_*) + \frac{1}{2}\bar{w}_*^\top \Sigma^{-1}\bar{w}_*\end{aligned} \tag{44}$$

where $\Sigma^{-1} = (\sigma^{-2} \mathrm{I}_{\tau p} - \theta X H X^\top) \succ 0$ and

$$\bar{w}_* = \operatorname*{arg\,max}_{\bar{w}\in\mathbb{R}^{\tau p}} \left\{\theta(X\tilde{h})^\top \bar{w} - \frac{1}{2}\bar{w}^\top(\sigma^{-2} \mathrm{I}_{\tau p} - \theta X H X^\top)\bar{w}\right\} = \theta(\sigma^{-2} \mathrm{I}_{\tau p} - \theta X H X^\top)^{-1} X\tilde{h}.$$

The first claim follows from the factorization in (44). The approximated risk-sensitive loss can then be computed analytically and reads

$$\begin{aligned}\hat{\eta}(\bar{u}) &= \frac{1}{\theta}\log\int(2\pi\sigma^2)^{-\tau p/2}\exp\left[\theta h(\tilde{x}(\bar{u})+\nabla\tilde{x}(\bar{u})^\top\bar{w})-\frac{1}{2\sigma^2}\|\bar{w}\|_2^2\right]d\bar{w}\\ &= \frac{1}{\theta}\log\left(\sqrt{\det(\sigma^{-2}\Sigma)}\exp\left[\theta h(\bar{x})+\frac{1}{2}\bar{w}_*^\top\Sigma^{-1}\bar{w}_*\right]\right)\\ &= -\frac{1}{2\theta}\log\det(\mathrm{I}_{\tau p}-\theta\sigma^2XHX^\top)+h(\bar{x})+\frac{\theta\sigma^2}{2}\tilde{h}^\top X^\top(\mathrm{I}_{\tau p}-\theta\sigma^2XHX^\top)^{-1}X\tilde{h}.\end{aligned}$$

□

As a corollary we get an expression for the truncated gradient.

Corollary D.2. *Given $\bar{u}\in\mathbb{R}^{\tau p}$ such that condition (14) holds, the truncated gradient of the approximated risk sensitive loss reads*

$$\widehat{\nabla}\hat{\eta}_\theta(\bar{u}) = \nabla\tilde{x}(\bar{u})\nabla h(\tilde{x}(\bar{u})+\nabla\tilde{x}(\bar{u})^\top w^*)$$

where $\bar{w}_*$ is given in (16).

Proof. The truncated gradient is the mean of an affine function of w under the distribution $\hat{p}(\cdot;\bar{u})$ defined in (15), it reads then

$$\widehat{\nabla}\hat{\eta}_\theta(\bar{u}) = \mathbb{E}_{\bar{w}\sim\hat{p}(\cdot;\bar{u})}[A\bar{w}+b] = A\bar{w}_*+b$$

with A, b defined by $h, \tilde{x}, \bar{u}$. □

We can then link the truncated gradient to the RegILEQ step.

Proposition 2.3. *Consider (RegILEQG) at iteration k, if condition (14) holds on $\bar{u}^{(k)}$, the step is defined and reads*

$$\begin{aligned}\bar{u}^{(k+1)} = \bar{u}^{(k)} &- (G+\gamma_k^{-1}\mathrm{I}_{\tau p}+XHX^\top+\theta V)^{-1}\\ &\times(\nabla g(\bar{u}^{(k)})+\widehat{\nabla}\hat{\eta}_\theta(\bar{u}^{(k)})),\end{aligned}$$

where

$$\begin{aligned}V &= \mathbb{V}\mathrm{ar}_{\bar{w}\sim\hat{p}(\cdot;\bar{u}^{(k)})}\nabla\tilde{x}(\bar{u}^{(k)})\nabla h(\tilde{x}(\bar{u}^{(k)})+\nabla\tilde{x}(\bar{u}^{(k)})^\top w)\\ &= XHX^\top(\sigma^{-2}\mathrm{I}_{\tau p}-\theta XHX^\top)^{-1}XHX^\top\end{aligned}$$

and $X{=}\nabla\tilde{x}(\bar{u}^{(k)})$, $H{=}\nabla^2h(\bar{x})$, $G{=}\nabla^2g(\bar{u}^{(k)})$, $\bar{x}{=}\tilde{x}(\bar{u}^{(k)})$.

Proof. To ease notations denote $\bar{u}^{(k)}=\bar{u}$, $\bar{u}^{(k+1)}=\bar{u}^+$ and $\gamma_k=\gamma$ such that the ILEQG step reads $\bar{u}^+=\bar{u}+\bar{v}^*$ where $\bar{v}^*$ is the solution of the min-max problem

$$\min_{\bar{v}\in\mathbb{R}^{\tau p}}\max_{\bar{w}\in\mathbb{R}^{\tau p}} q_h(\bar{x}+\nabla\tilde{x}(\bar{u})^\top(\bar{v}+\bar{w});\bar{x})+q_g(\bar{u}+\bar{v};\bar{u})+\frac{1}{2\gamma}\|\bar{v}\|_2^2-\frac{1}{2\theta\sigma^2}\|\bar{w}\|_2^2$$

where $\bar{x}=\tilde{x}(\bar{u})$, $q_h(\bar{x}+\bar{y};\bar{x})=h(\bar{x}+\bar{y})=h(\bar{x})+\nabla h(\bar{x})^\top\bar{y}+\frac{1}{2}\bar{y}^\top\nabla^2h(\bar{x})\bar{y}$, same for q_g. Denote $\tilde{g}=\nabla g(\bar{u}), G=\nabla^2g(\bar{u}), \tilde{h}=\nabla h(\bar{x}), H=\nabla^2h(\bar{x})$ and $X=\nabla\tilde{x}(\bar{u})$. The problem is then equivalent to

$$\begin{aligned}&\min_{\bar{v}\in\mathbb{R}^{\tau p}}(\tilde{g}+X\tilde{h})^\top\bar{v}+\frac{1}{2}\bar{v}^\top(G+\gamma^{-1}\mathrm{I}_{\tau p}+XHX^\top)\bar{v}+\max_{\bar{w}\in\mathbb{R}^{\tau p}}(X\tilde{h}+XHX^\top\bar{v})^\top\bar{w}-\frac{1}{2}\bar{w}^\top((\theta\sigma^2)^{-1}\mathrm{I}_{\tau p}-XHX^\top)\bar{w}\\ &=\min_{\bar{v}\in\mathbb{R}^{\tau p}}(\tilde{g}+X\tilde{h})^\top\bar{v}+\frac{1}{2}\bar{v}^\top(G+\gamma^{-1}\mathrm{I}_{\tau p}+XHX^\top)\bar{v}+\frac{1}{2}(X\tilde{h}+XHX^\top\bar{v})^\top((\theta\sigma^2)^{-1}\mathrm{I}_{\tau p}-XHX^\top)^{-1}(X\tilde{h}+XHX^\top\bar{v})\end{aligned}$$

where we used $(\sigma^{-2}\mathrm{I}_{\tau p}-\theta XHX^\top)\succ 0$. Denote

$$\bar{w}_* = ((\theta\sigma^2)^{-1}\mathrm{I}_{\tau p}-XHX^\top)^{-1}X\tilde{h}$$

which is equal to $\bar{w}_*$ defined in Prop. 2.2. The solution of the problem reads then

$$\bar{v}^* = -(G+\gamma^{-1}\mathrm{I}_{\tau p}+R)^{-1}(\tilde{g}+X\tilde{h}+XHX^\top\bar{w}^*)$$

where

$$R = XHX^\top+XHX^\top((\theta\sigma^2)^{-1}\mathrm{I}_{\tau p}-XHX^\top)^{-1}XHX^\top$$

The truncated gradient from Prop. D.2 reads

$$\begin{aligned}\widehat{\nabla}\hat{\eta}_\theta(\bar{u}) &= \nabla\tilde{x}(\bar{u})\nabla h(\tilde{x}(\bar{u}) + \nabla\tilde{x}(\bar{u})^\top \bar{w}_*) \\ &= X(\tilde{h} + HX^\top \bar{w}_*)\end{aligned}$$

which concludes the proof. □

D.3 Convergence analysis

Recall the assumptions made for the convergence analysis.

Assumption 2.4.

1. *The dynamics ϕ_t are twice differentiable, bounded, Lipschitz, smooth such that the trajectory function $\tilde{x}$ is also twice differentiable, bounded, Lipschitz and smooth. Denote by $L_{\tilde{x}}$ and $\ell_{\tilde{x}}$ the Lipschitz continuity and smoothness constants respectively of $\tilde{x}$ and define $M_{\tilde{x}} = \max_{\bar{u}\in\tau p} \operatorname{dist}(\tilde{x}(\bar{u}), X^*)$, where $X^* = \arg\min_{\bar{x}\in\mathbb{R}^{\tau d}} h(\bar{x})$.*
2. *The costs h and g are convex quadratics with smoothness constants L_h, L_g.*
3. *The risk-sensitivity parameter is chosen such that $\tilde{\sigma}^{-2} = \sigma^{-2} - \theta L_h \ell_{\tilde{x}}^2 > 0$, which ensures that condition (14) holds for any $\bar{u} \in \mathbb{R}^{\tau p}$.*

On $\mathcal{X} = \tilde{x}(\mathbb{R}^{\tau p})$, h is Lipschitz continuous, denote $\ell_h(\mathcal{X})$ the Lipschitz parameter. Using that $h(\bar{x}) = \frac{1}{2}(\bar{x} - \bar{x}^*)^\top H(x - x^*) + \min_{\bar{x}} h(\bar{x})$ with $H = \nabla^2 h(\bar{x})$ and $\bar{x}^* \in \arg\min_{\bar{x}} h(\bar{x})$, we get $\|\nabla h(\bar{x})\|_2 \leq L_h \|\bar{x} - \bar{x}^*\|_2$ and so

$$\ell_h(\mathcal{X}) \leq L_h M_{\tilde{x}} \tag{45}$$

We detail the approximation made by the truncated gradient in the following proposition.

Proposition D.3. *Under Ass. 2.4, we have for any $\bar{u} \in \mathbb{R}^{\tau p}$,*

$$\|\nabla\hat{\eta}_\theta(\bar{u}) - \widehat{\nabla}\hat{\eta}_\theta(\bar{u})\|_2 \leq \theta\tilde{\sigma}^2 L_h^2 L_{\tilde{x}} \ell_{\tilde{x}} M_{\tilde{x}}^2 + \theta^2\tilde{\sigma}^4 L_h^3 L_{\tilde{x}} \ell_{\tilde{x}}^3 M_{\tilde{x}}^2 + \tau p \tilde{\sigma}^2 L_h L_{\tilde{x}} \ell_{\tilde{x}}.$$

Proof. We have with $\hat{p}(\cdot;\bar{u})$ defined in (15), and denoting $\tilde{h} = \nabla h(\bar{x})$, $H = \nabla^2 h(\bar{x})$ and $X = \nabla\tilde{x}(\bar{u})$ for $\bar{x} = \tilde{x}(\bar{u})$,

$$\begin{aligned}\nabla\hat{\eta}_\theta(\bar{u}) - \widehat{\nabla}\hat{\eta}_\theta(\bar{u}) &= \mathbb{E}_{\bar{w}\sim\hat{p}(\cdot;\bar{u})} \nabla^2\tilde{x}(\bar{u})[\cdot, \bar{w}, \nabla h(\tilde{x}(\bar{u}) + \nabla\tilde{x}(\bar{u})^\top \bar{w})] \\ &= \mathbb{E}_{\bar{w}\sim\hat{p}(\cdot;\bar{u})} \left[\nabla^2\tilde{x}(\bar{u})[\cdot, \bar{w}, \tilde{h}] + \nabla^2\tilde{x}(\bar{u})[\cdot, \bar{w}, HX^\top \bar{w}]\right] \qquad (46) \\ &= \nabla^2\tilde{x}[\cdot, \bar{w}_*, \tilde{h}] + \begin{pmatrix} \mathbf{Tr}(\mathcal{X}_{1,\cdot,\cdot} HX^\top \mathbb{E}_{\bar{w}\sim\hat{p}(\cdot;\bar{u})}[\bar{w}\bar{w}^\top]) \\ \vdots \\ \mathbf{Tr}(\mathcal{X}_{\tau p,\cdot,\cdot} HX^\top \mathbb{E}_{\bar{w}\sim\hat{p}(\cdot;\bar{u})}[\bar{w}\bar{w}^\top]), \end{pmatrix} \qquad (47)\end{aligned}$$

where $\mathcal{X} = \nabla^2\tilde{x}(\bar{u})$ and we used the notations defined in Appendix A. We have then

$$\mathbb{E}_{\bar{w}\sim\hat{p}(\cdot;\bar{u})}[\bar{w}\bar{w}^\top] = \mathbb{V}\mathrm{ar}_{\bar{w}\sim\hat{p}(\cdot;\bar{u})}(\bar{w}) + \mathbb{E}_{\bar{w}\sim\hat{p}(\cdot;\bar{u})}(\bar{w})\, \mathbb{E}_{\bar{w}\sim\hat{p}(\cdot;\bar{u})}(\bar{w})^\top = \Sigma + \bar{w}_*\bar{w}_*^\top$$

where $\bar{w}_*$ and Σ are defined in (16). So we get

$$\nabla\hat{\eta}_\theta(\bar{u}) - \widehat{\nabla}\hat{\eta}_\theta(\bar{u}) = \nabla^2\tilde{x}[\cdot; \bar{w}_*, \tilde{h}] + \nabla^2\tilde{x}(\bar{u})[\cdot, \bar{w}_*, HX^\top \bar{w}_*] + \sum_{i=1}^{\tau p} \nabla^2\tilde{x}(\bar{u})[\cdot, u_i, HX^\top u_i]$$

where $\Sigma = \sum_{i=1}^{\tau p} u_i u_i^\top$ with $\|u_i\|_2^2 \leq \lambda_{\max}(\Sigma)$. Therefore

$$\|\nabla\hat{\eta}_\theta(\bar{u}) - \widehat{\nabla}\hat{\eta}_\theta(\bar{u})\|_2 \leq L_{\tilde{x}} \|\bar{w}_*\|_2 \ell_h(\mathcal{X}) + L_{\tilde{x}} \|\bar{w}_*\|_2^2 L_h \ell_{\tilde{x}} + \tau p L_{\tilde{x}} \|\Sigma\|_2 L_h \ell_{\tilde{x}}$$

where $\ell_h(\mathcal{X})$ is the Lipschitz parameter of h on $\mathcal{X} = \tilde{x}(\mathbb{R}^{\tau p})$ that can be bounded by (45) and we used the tensor norm defined in (20). The bound follows, using the definitions of $\bar{w}_*$ and Σ, i.e.,

$$\begin{aligned}\|\bar{w}_*\|_2 &\leq \theta(\sigma^{-2} - \theta L_h \ell_{\tilde{x}}^2)^{-1} \ell_{\tilde{x}} \ell_h(\mathcal{X}), \\ \|\Sigma\|_2 &\leq (\sigma^{-2} - \theta L_h \ell_{\tilde{x}}^2)^{-1}.\end{aligned}$$

□

The convergence under appropriate sufficient decrease condition is presented in the following proposition.

Theorem 2.5. *Under Ass. 2.4, suppose that the step-sizes of* (RegILEQG) *are chosen such that*

$$\hat{f}_\theta(\bar{u}^{(k+1)}) \leq m_{f_\theta}(\bar{u}^{(k+1)}; \bar{u}^{(k)}) + \frac{1}{2\gamma_k}\|\bar{u}^{(k+1)} - \bar{u}^{(k)}\|_2^2, \tag{17}$$

with $\gamma_k \in [\gamma_{\min}, \gamma_{\max}]$. *Then, the approximated objective* $\hat{f}_\theta$ *decreases and after* K *iterations we have*

$$\min_{k=0,\dots,K-1} \|\nabla \hat{f}_\theta(\bar{u}^{(k)})\|_2 \leq L\sqrt{\frac{2(\hat{f}_\theta(\bar{u}^{(0)}) - \hat{f}_\theta(\bar{u}^{(K)}))}{K}} + \delta,$$

where $L = \max_{\gamma\in[\gamma_{\min},\gamma_{\max}]} \sqrt{\gamma}(L_g + \gamma^{-1} + (\tilde{\sigma}/\sigma)^2\ell_{\tilde{x}}^2 L_h)$, $\delta = \theta\tilde{\sigma}^2 L_h^2 L_{\tilde{x}}\ell_{\tilde{x}} M_{\tilde{x}}^2 + \theta^2\tilde{\sigma}^4 L_h^3 L_{\tilde{x}}\ell_{\tilde{x}}^3 M_{\tilde{x}}^2 + \tau p\tilde{\sigma}^2 L_h L_{\tilde{x}}\ell_{\tilde{x}}$.

Proof. Under Ass. 2.4, the model $m_{f_\theta}(\bar{v}; \bar{u}^{(k)})$ defined in (12) is convex as shown for example in the proof of Prop. 2.3. By using that $\bar{v} \to m_{f_\theta}(\bar{v}; \bar{u}^{(k)}) + \frac{1}{2\gamma_k}\|\bar{v} - \bar{u}^{(k)}\|_2^2$ is γ_k^{-1} strongly convex with minimum achieved on $\bar{u}_{k+1}$ we get

$$\begin{aligned}\hat{f}_\theta(\bar{u}^{(k)}) = m_{f_\theta}(\bar{u}^{(k)}; \bar{u}^{(k)}) &\geq m_{f_\theta}(\bar{u}^{(k+1)}; \bar{u}^{(k)}) + \frac{1}{\gamma_k}\|\bar{u}^{(k+1)} - \bar{u}^{(k)}\|_2^2 \\ &\overset{(17)}{\geq} \hat{f}_\theta(\bar{u}^{(k+1)}) + \frac{1}{2\gamma_k}\|\bar{u}^{(k+1)} - \bar{u}^{(k)}\|_2^2. \end{aligned} \tag{48}$$

Rearranging the terms and summing the inequalities we get

$$\frac{1}{K}\sum_{k=0}^{K-1} \frac{1}{2\gamma_k}\|\bar{u}^{(k+1)} - \bar{u}^{(k)}\|_2^2 \leq \frac{\hat{f}_\theta(\bar{u}^{(0)}) - \hat{f}_\theta(\bar{u}^{(K)})}{K}.$$

Now using Proposition 2.3, we have that

$$\|\nabla g(\bar{u}^{(k)}) + \widehat{\nabla}\hat{\eta}_\theta(\bar{u}^{(k)})\|_2 \leq (L_g + \gamma^{-1} + \|R\|_2)\|\bar{u}^{(k+1)} - \bar{u}^{(k)}\|_2,$$

where

$$\begin{aligned}\|R\|_2 &= \|XH^{1/2}(\mathrm{I} - H^{1/2}X^\top(XHX^\top - (\theta\sigma^2)^{-1}\mathrm{I})^{-1}XH^{1/2})H^{1/2}X^\top\|_2 \\ &= \|XH^{1/2}(\mathrm{I} - \theta\sigma^2 H^{1/2}XX^\top H^{1/2})^{-1}H^{1/2}X^\top\|_2 \\ &\leq \frac{\ell_{\tilde{x}}^2 L_h}{1 - \theta\sigma^2\ell_{\tilde{x}}^2 L_h},\end{aligned}$$

using that for a semi-definite positive matrix A s.t $0 \preceq A \prec \mathrm{I}$, $\|I - A\|_2 \geq 1 - \lambda_{\max}(A)$ and $\|H^{1/2}\|_2^2 = \|H\|_2$. Therefore we get

$$\min_{k=0,\dots,K-1} \|\nabla g(\bar{u}^{(k)}) + \widehat{\nabla}\hat{\eta}_\theta(\bar{u}^{(k)})\|_2^2 \leq \frac{2L^2(\hat{f}_\theta(\bar{u}^{(0)}) - \hat{f}_\theta(\bar{u}^{(K)}))}{K}$$

where $L = \max_{\gamma\in[\gamma_{\min},\gamma_{\max}]} \sqrt{\gamma}(L_g + \gamma^{-1} + (\tilde{\sigma}/\sigma)^2\ell_{\tilde{x}}^2 L_h)$. Finally, using Prop. D.3, we get

$$\min_{k=0,\dots,K-1} \|\nabla \hat{f}_\theta(\bar{u}^{(k)})\|_2 \leq L\sqrt{\frac{2(\hat{f}_\theta(\bar{u}^{(0)}) - \hat{f}_\theta(\bar{u}^{(K)}))}{K}} + \theta\tilde{\sigma}^2 L_h^2 L_{\tilde{x}}\ell_{\tilde{x}} M_{\tilde{x}}^2 + \theta^2\tilde{\sigma}^4 L_h^3 L_{\tilde{x}}\ell_{\tilde{x}}^3 M_{\tilde{x}}^2 + \tau p\tilde{\sigma}^2 L_h L_{\tilde{x}}\ell_{\tilde{x}}.$$

□

The following proposition ensures that on any compact set there exists a step-size such that this criterion is satisfied.

Proposition D.4. *Under Ass. 2.4, for any compact set* C *there exists* $M_C > 0$ *such that for any* $\bar{u} \in C, \bar{v} \in C$, *the model* m_{f_θ} *approximates the approximated risk-sensitive loss as*

$$|\hat{f}_\theta(\bar{u} + \bar{v}) - m_{f_\theta}(\bar{u} + \bar{v}; \bar{u})| \leq \frac{M_C\|\bar{v}\|^2}{2}.$$

Proof. Denote $R_C = \max_{\bar{u}\in C} \|\bar{u}\|_2$. Denote $X = \nabla\tilde{x}(\bar{u})$, $H = \nabla^2 h(\bar{x})$. Following proof of Prop. 2.2, we have

$$\begin{aligned}m_{f_\theta}(\bar{u} + \bar{v}; \bar{u}) =& h(\tilde{x}(\bar{u}) + \nabla\tilde{x}(\bar{u})^\top\bar{v}) - \frac{1}{2\theta}\log\det(\mathrm{I} - \theta\sigma^2 XHX^\top) \\ &+ \frac{\theta\sigma^2}{2}\nabla h(\tilde{x}(\bar{u}) + \nabla\tilde{x}(\bar{u})^\top\bar{v})^\top X^\top(\mathrm{I}_{\tau p} - \theta\sigma^2 XHX^\top)^{-1}X\nabla h(\tilde{x}(\bar{u}) + \nabla\tilde{x}(\bar{u})^\top\bar{v}) \\ &+ g(\bar{u} + \bar{v})\end{aligned}$$

In the following denote $\mathring{h} = \nabla h(\tilde{x}(\bar{u}) + \nabla\tilde{x}(\bar{u})^\top\bar{v})$. On the other side, denote $\bar{y} = \tilde{x}(\bar{u}+\bar{v})$, $Y = \nabla\tilde{x}(\bar{u}+\bar{v})$ and $\hat{h} = \nabla h(\tilde{x}(\bar{u}+\bar{v})) = \nabla h(\bar{y})$, such that

$$\hat{f}_\theta(\bar{u}+\bar{v}) = h(\bar{y}) - \frac{1}{2\theta}\log\det(\mathrm{I}-\theta\sigma^2 YHY^\top) + \frac{\theta\sigma^2}{2}\hat{h}^\top Y^\top(\mathrm{I}-\theta\sigma^2 YHY^\top)^{-1}Y\hat{h} + g(\bar{u}+\bar{v})$$

First we have using $\bar{x}_* \in \arg\min_{\bar{x}\in\mathbb{R}^{\tau d}} h(\bar{x})$,

$$\begin{aligned}|h(\tilde{x}(\bar{u}+\bar{v})) - h(\tilde{x}(\bar{u}) + \nabla\tilde{x}(\bar{u})^\top\bar{v})| &= |\frac{1}{2}(\tilde{x}(\bar{u}+\bar{v}) + \tilde{x}(\bar{u}) + \nabla\tilde{x}(\bar{u})^\top\bar{v} - 2\bar{x}^*)^\top H(\tilde{x}(\bar{u}+\bar{v}) - \tilde{x}(\bar{u}) - \nabla\tilde{x}(\bar{u})^\top\bar{v})| \\ &\leq \frac{1}{4}(2M_{\tilde{x}} + \ell_{\tilde{x}}R_C)L_h L_{\tilde{x}}\|\bar{v}\|_2^2.\end{aligned}$$

Then denote

$$f(X) = -\frac{1}{2\theta}\log\det(\mathrm{I}-\theta\sigma^2 XHX^\top)$$

such that

$$\|\nabla f(X)\|_2 = \sigma^2\|(\mathrm{I}-\theta\sigma^2 XHX^\top)^{-1}XH\|_2 \leq \frac{\sigma^2 L_h\ell_{\tilde{x}}}{1-\theta\sigma^2 L_h\ell_{\tilde{x}}^2}.$$

Therefore

$$\begin{aligned}|f(X) - f(Y)| &\leq \ell_f\|\nabla\tilde{x}(\bar{u}+\bar{v}) - \nabla\tilde{x}(\bar{u})\|_2 \\ &\leq \frac{L_h\ell_{\tilde{x}}L_{\tilde{x}}}{1-\theta\sigma^2 L_h\ell_{\tilde{x}}^2}\|\bar{v}\|_2\end{aligned}$$

where ℓ_f is the Lipschitz continuity of f for X s.t. $\|X\|_2 \leq \ell_{\tilde{x}}$.

Now for the last term we have

$$\mathbf{Tr}(F(Y)\hat{h}\hat{h}^\top) - \mathbf{Tr}(F(X)\mathring{h}\mathring{h}^\top) = \mathbf{Tr}((F(Y)-F(X))\hat{h}\hat{h}^\top) + \mathbf{Tr}(F(X)(\hat{h}\hat{h}^\top - \mathring{h}\mathring{h}^\top))$$

where $F(X) = X^\top(\mathrm{I}-\theta\sigma^2 XHX^\top)^{-1}X$. Define for $M \in \mathbb{R}^{\tau d\times\tau d}$ with $M \succeq 0$,

$$f_M(X) = \frac{1}{2}\mathbf{Tr}(MX^\top(\mathrm{I}-\theta\sigma^2 XHX^\top)^{-1}X).$$

We have

$$\begin{aligned}\|\nabla f_M(X)\|_2 =& \|(\mathrm{I}-\theta\sigma^2 XHX^\top)^{-1}XM + \theta\sigma^2(\mathrm{I}-\theta\sigma^2 XHX^\top)^{-1}XMX^\top(\mathrm{I}-\theta\sigma^2 XHX^\top)^{-1}XH\|_2 \\ &\leq \frac{\|M\|_2\ell_{\tilde{x}}}{1-\theta\sigma^2 L_h\ell_{\tilde{x}}^2} + \frac{\theta\sigma^2\|M\|_2\ell_{\tilde{x}}^3 L_h}{(1-\theta\sigma^2 L_h\ell_{\tilde{x}}^2)^2}.\end{aligned}$$

Therefore

$$\begin{aligned}|\,\mathbf{Tr}((F(Y)-F(X))\hat{h}\hat{h}^\top)| &\leq \ell_{f_{\hat{h}\hat{h}^\top}}\|Y-X\|_2 \\ &\leq \ell_{h,\tilde{x}}^2\left(\frac{\ell_{\tilde{x}}}{1-\theta\sigma^2 L_h\ell_{\tilde{x}}^2} + \frac{\theta\sigma^2\ell_{\tilde{x}}^3 L_h}{(1-\theta\sigma^2 L_h\ell_{\tilde{x}}^2)^2}\right)L_{\tilde{x}}\|\bar{v}\|_2,\end{aligned}$$

where $\ell_{f_{\hat{h}\hat{h}^\top}}$ is the Lipschitz continuity of $f_{\hat{h}\hat{h}^\top}$ for X s.t. $\|X\|_2 \leq \ell_{\tilde{x}}$. Finally

$$\begin{aligned}|\,\mathbf{Tr}(F(X)(\hat{h}\hat{h}^\top - \mathring{h}\mathring{h}^\top))| &= |\,\mathbf{Tr}(\hat{h}+\mathring{h})^\top F(X)(\hat{h}-\mathring{h})| \\ &\leq (2\ell_{h,\tilde{x}} + L_h\ell_{\tilde{x}}R_C)\frac{\ell_{\tilde{x}}^2}{1-\theta\sigma^2 L_h\ell_{\tilde{x}}^2}L_h L_{\tilde{x}}\frac{\|\bar{v}\|_2^2}{2}.\end{aligned}$$

Combining all terms we get

$$\begin{aligned}|\hat{f}_\theta(\bar{u}+\bar{v}) - m_{f_\theta}(\bar{u}+\bar{v})| \leq& \frac{1}{2}(2M_{\tilde{x}} + \ell_{\tilde{x}}R_C)L_h L_{\tilde{x}}\frac{\|\bar{v}\|_2^2}{2} \\ &+ \frac{2L_h\ell_{\tilde{x}}L_{\tilde{x}}}{(1-\theta\sigma^2 L_h\ell_{\tilde{x}}^2)R_C}\frac{\|\bar{v}\|_2^2}{2} \\ &+ \theta\sigma^2\ell_{h,\tilde{x}}^2\left(\frac{\ell_{\tilde{x}}}{1-\theta\sigma^2 L_h\ell_{\tilde{x}}^2} + \frac{\theta\sigma^2\ell_{\tilde{x}}^3 L_h}{(1-\theta\sigma^2 L_h\ell_{\tilde{x}}^2)^2}\right)L_{\tilde{x}}\frac{\|\bar{v}\|_2^2}{2} \\ &+ \frac{\theta\sigma^2}{2}(2\ell_{h,\tilde{x}} + L_h\ell_{\tilde{x}}R_C)\frac{\ell_{\tilde{x}}^2}{1-\theta\sigma^2 L_h\ell_{\tilde{x}}^2}L_h L_{\tilde{x}}\frac{\|\bar{v}\|_2^2}{2}\end{aligned}$$

This concludes the proof with

$$M_C = \frac{1}{2}(2M_{\tilde{x}} + \ell_{\tilde{x}} R_C) L_h L_{\tilde{x}} + \frac{2\sigma^2 L_h \ell_{\tilde{x}} L_{\tilde{x}}}{(1-\theta\sigma^2 L_h \ell_{\tilde{x}}^2) R_C}$$
$$+ \theta\sigma^2 \ell_{h,\tilde{x}}^2 \left(\frac{\ell_{\tilde{x}}}{1-\theta\sigma^2 L_h \ell_{\tilde{x}}^2} + \frac{\theta\sigma^2 \ell_{\tilde{x}}^3 L_h}{(1-\theta\sigma^2 L_h \ell_{\tilde{x}}^2)^2} \right) L_{\tilde{x}} + \frac{\theta\sigma^2}{2}(2\ell_{h,\tilde{x}} + L_h \ell_{\tilde{x}} R_C) \frac{\ell_{\tilde{x}}^2}{1-\theta\sigma^2 L_h \ell_{\tilde{x}}^2} L_h L_{\tilde{x}}.$$

□

Finally the iterates can be forced to stay in a compact set such that the overall convergence is ensured as shown in the following proposition.

Proposition D.5. *Let $S_0 = \{\bar{u} : \hat{f}_\theta(\bar{u}) \leq \hat{f}_\theta(\bar{u}^{(0)})\}$ be the initial sub-level set of $\hat{f}_\theta$ and assume S_0 is compact. Consider the iterations of RegILEQG in (RegILEQG). Assume that*

$$\gamma_k = \hat{\gamma} = \min\{\ell_0^{-1}, M_C^{-1}\},$$

where M_C is defined in Prop. D.4, and denoting $\mathcal{B}_{2,1}$ the Euclidean ball of radius 1 centered at 0,

$$\ell_0 = \max_{\bar{u} \in S_0} \|\nabla g(\bar{u}) + \widehat{\nabla}\hat{\eta}_\theta(\bar{u})\|_2, \qquad C = S_0 + \mathcal{B}_{2,1}.$$

Then the sufficient decrease condition (17) is satisfied for all k.

Proof. Given $\bar{u}^{(k)} \in S_0$, we have from Proposition 2.3, using $\gamma_k \leq \ell_0^{-1}$

$$\|\bar{u}^{(k+1)} - \bar{u}^{(k)}\|_2 \leq \gamma_k \|\nabla g(\bar{u}^{(k)}) + \widehat{\nabla}\hat{\eta}_\theta(\bar{u}^{(k)})\|_2 \leq 1.$$

Therefore $\bar{u}^{(k+1)} \in S_0 + \mathcal{B}_{2,1} = C$ and $\bar{u}^{(k)} \in C$. They satisfy then, using $\gamma_k \leq M_C^{-1}$,

$$\hat{f}_\theta(\bar{u}^{(k+1)}) \leq m_{f_\theta}(\bar{u}^{(k+1)}; \bar{u}^{(k)}) + \frac{M_C}{2}\|\bar{u}^{(k+1)} - \bar{u}^{(k)}\|_2^2 \leq m_{f_\theta}(\bar{u}^{(k+1)}; \bar{u}^{(k)}) + \frac{1}{2\gamma_k}\|\bar{u}^{(k+1)} - \bar{u}^{(k)}\|_2^2$$

Therefore $\bar{u}^{(k+1)} \in S$. The claim follows by recursion starting from $\bar{u}^{(k)} = \bar{u}^{(0)} \in S_0$. □

E Detailed experimental setting

E.1 Discretization of the continuous time settings

The physical systems we consider below are described by continuous dynamics of the form

$$\ddot{z}(t) = f(z(t), \dot{z}(t), u(t))$$

where $z(t), \dot{z}(t), \ddot{z}(t)$ denote respectively the position, the speed and the acceleration of the system and $u(t)$ is a force applied on the system. The state $x(t) = (x_1(t), x_2(t))$ of the system is defined by the position $x_1(t) = z(t)$ and the speed $x_2(t) = \dot{z}(t)$ and the continuous cost is defined as

$$J(x,u) = \int_0^T h(x(t))dt + \int_0^T g(u(t))dt \quad \text{or} \quad J(x,u) = h(x(T)) + \int_0^T g(u(t))dt,$$

where T is the time of the movement and h, g are given convex costs. The discretization of the dynamics with a time step δ starting from a given state $\hat{x}_0 = (z_0, 0)$ reads then

$$\begin{aligned} x_{1,t+1} &= x_{1,t} + \delta x_{2,t} \\ x_{2,t+1} &= x_{2,t} + \delta f(x_{1,t}, x_{2,t}, u_t) \end{aligned} \qquad \text{for } t = 0, \dots \tau - 1$$

where $\tau = \lceil T/\delta \rceil$ and the discretized cost reads

$$J(\bar{x}, \bar{u}) = \sum_{t=1}^{\tau} h(x_t) + \sum_{t=0}^{\tau-1} g(u_t) \quad \text{or} \quad J(\bar{x}, \bar{u}) = h(x_\tau) + \sum_{t=0}^{T-1} g(u_t).$$

E.2 Continuous control settings

The control settings are illustrated in Fig. 5.

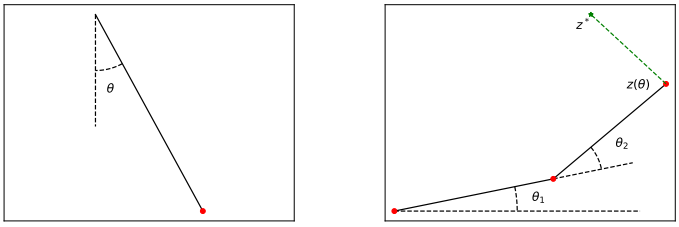

Figure 5: Control settings considered. From left to right: pendulum, two-link arm robot.

Pendulum. We consider a simple pendulum illustrated in Fig. 5, where $m = 1$ denotes the mass of the bob, $l = 1$ denotes the length of the rod, θ describes the angle subtended by the vertical axis and the rod, and $\mu = 0.01$ is the friction coefficient. Its dynamical evolution reads

$$\ddot{\theta}(t) = -\frac{g}{l}\sin\theta(t) - \frac{\mu}{ml^2}\dot{\theta}(t) + \frac{1}{ml^2}u(t)$$

The goal is to make the pendulum swing up (i.e. make an angle of π radians) and stop at a given time T. Formally, the continuous cost reads

$$J(x,u) = (\pi - \theta(T))^2 + \lambda_1\dot{\theta}(T)^2 + \lambda_2\int_0^T u^2(t)dt, \tag{49}$$

where $x(t) = (\theta(t), \dot{\theta}(t))$, $\lambda_1 > 0$ and $\lambda_2 > 0$.

Two-link arm. We consider the arm model with 2 joints (shoulder and elbow), moving in the horizontal plane presented by [Li and Todorov, 2004] and illustrated in Figure 5. The dynamics read

$$M(\theta(t))\ddot{\theta}(t) + C(\theta(t), \dot{\theta}(t)) + B\dot{\theta}(t) = u(t), \tag{50}$$

where $\theta = (\theta_1, \theta_2)$ is the joint angle vector, $M(\theta) \in \mathbb{R}^{2\times 2}$ is a positive definite symmetric inertia matrix, $C(\theta, \dot{\theta}) \in \mathbb{R}^2$ is a vector centripetal and Coriolis forces, $B \in \mathbb{R}^{2\times 2}$ is the joint friction matrix, and $u \in \mathbb{R}^2$ is the joint torque controlling the arm. See below for the complete definitions.

The goal is to make the arm reach a feasible target z^* and stop at that point. Denoting $\theta^*(z^*)$ a joint angle pairs that reach the target, the objective reads then

$$J(x,u) = \|\theta(T) - \theta^*(z^*)\|_2^2 + \lambda_1\|\dot{\theta}(T)\|_2^2 + \lambda_2\int_0^T \|u(t)\|_2^2 dt, \tag{51}$$

where $x(t) = (\theta(t), \dot{\theta}(t))$, $\lambda_1 > 0, \lambda_2 > 0$.

Detailed two-link arm model. We detail the the forward dynamics drawn from (50). We drop the dependence on t for readability. The dynamics read

$$\ddot{\theta} = M(\theta)^{-1}(u - C(\theta, \dot{\theta}) - B\dot{\theta}).$$

The expressions of the different variables and parameters are given by

$$M(\theta) = \begin{pmatrix} a_1 + 2a_2\cos\theta_2 & a_3 + a_2\cos\theta_2 \\ a_3 + a_2\cos\theta_2 & a_3 \end{pmatrix} \qquad C(\theta,\dot{\theta}) = \begin{pmatrix} -\dot{\theta}_2(2\dot{\theta}_1 + \dot{\theta}_2) \\ \dot{\theta}_1^2 \end{pmatrix} a_2\sin\theta_2$$

$$B = \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} \qquad \begin{aligned} a_1 &= k_1 + k_2 + m_2 l_1^2 \\ a_2 &= m_2 l_1 d_2 \\ a_3 &= k_2, \end{aligned}$$

where $b_{11} = b_{22} = 0.05$, $b_{12} = b_{21} = 0.025$, l_i and k_i are respectively the length (30cm, 33cm) and the moment of inertia (0.025kgm^2 , 0.045kgm^2) of link i , m_2 and d_2 are respectively the mass (1kg) and the distance (16cm) from the joint center to the

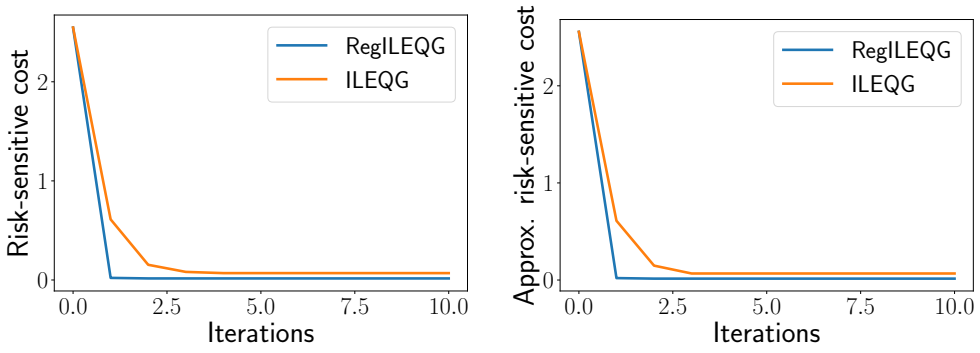

Figure 6: Convergence of iterative linearized methods, with regularization, RegILEQG, without regularization, ILEQG, on the two-link arm problem.

center of the mass for the second link. The inverse of the inertia matrix reads[1]

$$M(\theta)^{-1} = \frac{1}{(a_1 + 2a_2\cos(\theta_2))a_3 - (a_3 + a_2\cos\theta_2)^2}\begin{pmatrix} a_3 & -(a_3 + a_2\cos\theta_2) \\ -(a_3 + a_2\cos\theta_2) & a_1 + 2a_2\cos\theta_2 \end{pmatrix}.$$

E.3 Noise modeling details

For the two-link arm we use $\sigma_0 = 1/\|M(\theta)^{-1}\|$ to normalize the noise in the risk-sensitive and the test costs. Otherwise the modeled noise led experimentally to a chaotic behavior. Precisely we use for the risk-sensitive cost,

$$\begin{aligned} x_{1,t+1} &= x_{1,t} + \delta x_{2,t} \\ x_{2,t+1} &= x_{2,t} + \delta f(x_{1,t}, x_{2,t}, u_t + w_t) \end{aligned} \qquad \text{for } t = 0, \dots, \tau - 1.$$

with $w_t \sim \mathcal{N}(0, \sigma_0^2\,\mathrm{I})$ and for the test cost,

$$\begin{aligned} x_{1,t+1} &= x_{1,t} + \delta x_{2,t} \\ x_{2,t+1} &= x_{2,t} + \delta f(x_{1,t}, x_{2,t}, u_t + \rho\mathbb{1}(t = t_w)) \end{aligned} \qquad \text{for } t = 0, \dots, \tau - 1.$$

where $\rho \sim \mathcal{N}(0, \sigma_{test}/\sigma_0\,\mathrm{I}_p)$ and the plots are shown for increasing σ_{test}. For the pendulum problem we did not normalize the noise. We leave the analysis of the choice of σ for future work.

E.4 Optimization details

Convergence results. For Fig. 2, we took $\lambda_1 = 0.1$, $\lambda_2 = 0.01$, $T = 5$, in (49) for an horizon $\tau = 100$ and $\theta = 4$. We present in Fig. 6 the convergence obtained for the two-link arm problem, where we used the same parameters for $\lambda_1, \lambda_2, T, \tau, \theta$. The best step-sizes found after the burn-in phase were 8 for RegILEQG and 0.5 for ILEQG. Again the advantage of the regularized approach is that it can select bigger step-sizes while staying stable.

Robustness results. For both settings we used RegILEQG with a burn-in phase of 10 iterations and a grid of step-sizes 2^i for $i \in \{-5, 5\}$. We run the algorithm for 50 iterations and take the best solution according to the approximate risk-sensitive function.

For the pendulum problem we used $\lambda_1 = 10$, $\lambda_2 = 10^{-3}$, $T = 5$, for an horizon $\tau = 100$. For the two-link arm problem we used $\lambda_1 = 10^{-2}$ and $\lambda_2 = 10^{-3}$, $T = 5$, and the same horizon.

[1]Note that the dynamics have continuous derivatives if the norm of the denominator is bounded below by a positive constant 0. We have

$$(a_1 + 2a_2\cos(\theta_2))a_3 - (a_3 + a_2\cos\theta_2)^2 = \alpha - \beta\cos^2\theta_2$$

with

$$\alpha = a_3(a_1 - a_3) = k_1k_2 + m_2l_1^2k_2 \qquad \beta = a_2^2 = m_2^2l_1^2d_2^2,$$

which gives $\alpha = 9.1125 \times 10^{-2}$ and $\beta = 2.304 \times 10^{-3}$. Therefore it is bounded below by a positive constant, the function is continuously differentiable.