Uncertainty-dependent Optimal Control for Robot Control Considering High-order Cost Statistics

José Ramón Medina and Sandra Hirche

Abstract— **As the application of probabilistic models in robotic applications increases, the necessity of a systematic robot-control method that considers the effects of multiple uncertainty sources becomes more evident. Motivated by human sensorimotor findings, in this work we study the stochastic locally optimal feedback control problem with high-order cost statistics where dynamics have multiple additive noise sources and cost variability produced by each uncertainty source is evaluated marginally. We present risk-sensitive and cost-cumulant solutions for this problem for non-linear dynamics and non-quadratic costs. Locally optimal solutions are found by iteratively performing a linear quadratic approximation around a nominal trajectory, solving the local problem and updating the trajectory until convergence. Simulation results of a point mass robot and a two-link manipulator validate the applicability of the proposed approach and illustrate its peculiarities.**

I. Introduction

The use of probabilistic models and estimation methods in robotics is rapidly increasing. In the quest of producing autonomous cognitive robots, probabilistic methods are the most recurrent machine learning tool, with applications such us sensing the environment and the robot's state [1] or learning dynamical models of task execution behavior [2]. When aiming for robot control based on such models, the decision-making process has to cope with their intrinsic probabilistic uncertainty. Although the usual approach considers the expectation of uncertain estimations, the level of uncertainty itself might be a key parameter to enhance task performance. As an example, consider the task of grasping an object with uncertain pose. Its expected value is only a coarse estimation of its probabilistic representation. Its variance, though, may reveal more certainty in some degrees of freedom; control design targeting those more aggressively while allowing more variability in others may be key for task success. Similarly, in a navigation task though a cluttered environment where obstacles have different levels of uncertainty, the possibility of collision may vary significantly depending on the obstacles' variances. In order to exploit the potential of probabilistic estimations, a systematic and flexible control approach that considers potential variability appears as a necessary step.

Recent findings suggest that human sensorimotor behavior is the result of the minimization of a cost function with explicit consideration of noise in the dynamics [3]. The standard treatment in these settings is to minimize the expected value of the resulting stochastic cost. In contrast, human decisions under variability have been more accurately modeled by risk-sensitive decision-makers [4], which also considers cost variance. These results motivate us to consider high order cost statistics for control in robotic applications.

Several works explore the synthesis of uncertainty-dependent decisions in robotic settings from an optimality perspective. In order to do so, the performance measure applicable to deterministic settings is modified in order to account for the uncertainty's variance. For instance, enhanced collision avoidance is achieved including an additive cost term representing the probability of collision in [5]. Similarly, variance-dependent stiffness is synthesized adding prediction variance to the performance measure as in [6]. The inverse of the variance is also used as a weighting term for measuring distance to desired states in [7] or [8]. The heterogeneity of these solutions highlights the lack of a systematic approach to uncertainty-dependent control.

Local approximations to non-linear systems are an effective solution to nonlinear robot control problems. The iterative linear quadratic Gaussian (iLQG) method presented in [9] which is a simplified version of Stochastic Differential Dynamic Programming (SDDP) [10] is a recurrent example. The resulting locally optimal feedback policies have been applied in many robot navigation problems combined with belief roadmaps [11] or adding the state variance as a part of an augmented state [12]. All these approaches consider the expected value of the random cost as a performance measure, neglecting cost variance.

In this work we study the stochastic locally optimal feedback control problem considering high-order cost statistics. Dynamics are assumed to have multiple additive noise sources and cost variability produced by each uncertainty source is evaluated marginally. We present risk-sensitive and cost-cumulant solutions for this problem for non-linear dynamics and non-quadratic costs. Locally optimal solutions are found by iteratively performing a linear quadratic approximation around a nominal trajectory, solving the local problem and updating the trajectory until convergence. We further study the high-order statistics evaluation of static uncertainty sources in two common robotic problems. Simulation results of a point mass robot and a two-link manipulator validate the applicability of the proposed approach.

The rest of this paper is organized as follows. Section II formally defines the problem and approach. Section III and Section IV present the risk-sensitive and cost-cumulant solutions for the linear quadratic setting. The main iteration of the algorithm is described in Section V. Experimental simulations are illustrated in Section VI.

Authors are with the Institute for Information-Oriented Control, Department of Electrical Engineering and Information Technology, Technische Universität München, D-80290 Munich, Germany. {medina, hirche}@tum.de

II. Problem Setting and Approach

Consider a robot with dynamics given by the completely observed controlled diffusion process

$$d\boldsymbol{x}^r(t) = \boldsymbol{f}^r(\boldsymbol{x}^r(t), \boldsymbol{u}^*(t))dt + \mathcal{G}^r dB^r(t) \ , \qquad (1)$$

where $\boldsymbol{x}^r \in \mathbb{R}^n$, $\boldsymbol{u}^* \in \mathbb{R}^m$ represent the robot's state and control input respectively, B^r is a n-dimensional standard Brownian motion noise defined in the complete probability space $(\Omega_r, \mathcal{F}_r, \mathsf{P}_r)$, $\mathcal{G}^r \in \mathbb{R}^{n\times n}$ is its infinitesimal variance and $\boldsymbol{x}_0^r$ is the initial state. Additionally, let $\boldsymbol{x}^g \in \mathbb{R}^n$ and $\boldsymbol{x}^o \in \mathbb{R}^n$ be a desired trajectory to follow and the state of an obstacle respectively

$$\begin{aligned} d\boldsymbol{x}^g(t) &= \boldsymbol{f}^g(\boldsymbol{x}^g(t))dt + \mathcal{G}^g dB^g(t) \\ d\boldsymbol{x}^o(t) &= \boldsymbol{f}^o(\boldsymbol{x}^o(t))dt + \mathcal{G}^o dB^o(t) \ , \end{aligned} \qquad (2)$$

where B^g and B^o are n-dimensional standard Brownian motion noises defined in complete probability spaces $(\Omega_g, \mathcal{F}_g, \mathsf{P}_g)$ and $(\Omega_o, \mathcal{F}_o, \mathsf{P}_o)$ respectively, $\mathcal{G}_g \in \mathbb{R}^{n\times n}$ and $\mathcal{G}_o \in \mathbb{R}^{n\times n}$ their infinitesimal variances and $\boldsymbol{x}_0^g$ and $\boldsymbol{x}_0^o$ their respective initial states [1].

The measure evaluating the performance of a candidate control law $\boldsymbol{u}^*(\cdot)$ is given by cost function

$$\begin{aligned} J(\boldsymbol{u}^*(\cdot)) = \ & h_{T_c}(\boldsymbol{x}^r(T_c), x^g(T_c), x^o(T_c)) \\ & + \int_{t=0}^{T_c} h(t, \boldsymbol{x}^r(t), \boldsymbol{x}^g(t), \boldsymbol{x}^o(t), \boldsymbol{u}^*(t))dt \ , \end{aligned} \qquad (3)$$

where T_c is the time horizon, h is the cost rate and h_{T_c} the final cost. This performance index is usually designed penalizing both the distance to the desired goal $\boldsymbol{x}^g$ and the necessary control efforts $\boldsymbol{u}^*$ while favoring configurations distant to obstacles $\boldsymbol{x}^o$.

As (3) is usually defined in terms of errors between robot and obstacle and robot and goal and in order to keep a compact formulation, let $\boldsymbol{\xi}^*$ be the joint state such that

$$\boldsymbol{\xi}^* = [(\boldsymbol{x}^g - \boldsymbol{x}^r) \ \ (\boldsymbol{x}^o - \boldsymbol{x}^r)]^\mathsf{T}. \qquad (4)$$

The robot-obstacle-goal dynamics are then compactly formulated as

$$d\boldsymbol{\xi}^*(t) = \boldsymbol{f}(\boldsymbol{\xi}^*(t), \boldsymbol{u}^*(t))dt + \sum_{s=1}^{S} \mathcal{G}^s dB^s(t) \ , \qquad (5)$$

where S is the number of independent Brownian motions (three in our particular case), B^s is the corresponding n-dimensional standard Brownian motion noise defined in the s-th probability space $(\Omega_s, \mathcal{F}_s, \mathsf{P}_s)$ for $s = 1 \cdots S$, and $\mathcal{G}^s$ and $\boldsymbol{f}$ are defined such that (1), (2) and (4) hold.

The optimal control solution is given by the control law $\boldsymbol{u}^*(\cdot)$ that minimizes (3) constrained to dynamics (5) and (4). Due to the stochastic nature of (5), the cost to be optimized is a random variable. Hence, prior to finding an optimal solution, an interpretation of the random cost in terms of a deterministic performance measure is necessary. A valid approach consists of evaluating a statistical measure of (3) usually limited to the expected value, i.e. $\mathsf{E}_\mathsf{P}[J]$ and where the expectation is defined in the product probability space of all uncertainty sources, i.e. the probability space $(\Omega, \mathcal{F}, \mathsf{P})$ given by

$$\Omega = \Omega_1 \times \Omega_2 \times \cdots \times \Omega_S \quad \mathcal{F} = \mathcal{F}_1 \times \mathcal{F}_2 \times \cdots \times \mathcal{F}_S, \qquad (6)$$

where $\times$ denotes the Cartesian product and P is the joint measure defined on the measurable space $(\Omega, \mathcal{F})$. However, this formulation has two drawbacks: (i) all statistics further than the expected value are neglected even when higher-order measures enable richer uncertainty-dependent decisions and (ii) the evaluation of the statistics in probability space (6) considers all random variables *jointly*. This limits the way cost variability influence decisions: it might be beneficial that cost fluctuations induced *marginally* by the obstacle's randomness influence the robot's decisions in a different manner as the goal's stochasticity.

Concerning the first issue, we consider an arbitrary linear combination of K high-order cumulants

$$\kappa_\mathsf{P}^{(1)}(J) + \sum_{r=2}^{K} \gamma_r \kappa_\mathsf{P}^{(r)}(J) \ , \qquad (7)$$

where $\kappa_\mathsf{P}^{(r)}$ denotes the r-th cumulant calculated in probability space $(\Omega, \mathcal{F}, \mathsf{P})$. In the following we will informally refer to higher order statistical terms present in the summation of (7) as *cost variability*. Note that high order statistical terms are also denoted *risk measures* in modern portfolio theory [15]. Cumulants are derived by means of the cumulant generating function

$$\Psi_\mathsf{P}(\theta) = \log \mathsf{E}_\mathsf{P}[\exp\{\theta J\}] \ .$$

Its power series expansion is given by

$$\Psi_\mathsf{P}(\theta) = \sum_{r=1}^{\infty} \frac{\theta^r}{r!} \kappa_\mathsf{P}^{(r)}(J) \ , \qquad (8)$$

where

$$\kappa_\mathsf{P}^{(r)}(J) = \left.\frac{\partial^r \Psi_\mathsf{P}(\theta)}{\partial \theta^r}\right|_{\theta=0}, \qquad (9)$$

providing a compact way to calculate the desired cumulants.

Regarding the second issue, in order to evaluate the cost variability produced marginally by the s-th stochastic process, we define its corresponding *marginal dynamics* as

$$d\boldsymbol{\xi}^*(t) = \boldsymbol{f}(\boldsymbol{\xi}^*(t), \boldsymbol{u}^*(t))dt + \mathcal{G}^s dB^s(t) \ . \qquad (10)$$

We then aim for the analysis of (7) constrained to each of the marginal dynamics. The optimization criterion considered through this work is given by

$$\min_{\boldsymbol{u}(\cdot)} \Xi = \min_{\boldsymbol{u}(\cdot)} \frac{1}{S} \Big(\sum_{s=1}^{S} \kappa_{\mathsf{P}_s}^{(1)}(J) + \sum_{r=2}^{K} \gamma_{r,s} \kappa_{\mathsf{P}_s}^{(r)}(J) \Big) \qquad (11)$$

where the cumulants defined in the s-th probability space are constrained to the s-th marginal dynamics (10). Although

[1] The fact that all elements (robot, goal and obstacle) follow stochastic dynamics might appear as an extreme case. However, the application of probabilistic models for approximating plant dynamics [6], dynamic obstacle motions [13] or desired goal dynamics [14] gives cause for this setting.

we do not solve the original problem with dynamics (5) this formulation provides more flexible decision-makers, enabled only by means of a marginal treatment of uncertainty sources. Instrumental in this problem is the choice of the weighting factors: cost variability terms with a positive weighting factor $\gamma_{r,s}$ will increase the overall cost while negative weighting factors produce the opposite effect. When all $\gamma_{r,s} = 0$ the expected value is recovered, which neglects any risk measure. In line with portfolio theory nomenclature, we will denote these three cases where $\gamma_{r,s} > 0$, $\gamma_{r,s} = 0$ and $\gamma_{r,s} < 0$ as *risk-averse*, *risk-seeking* and *risk-neutral* respectively, as a description of the optimizer's attitude towards a certain risk measure.

Interestingly, risk-sensitive control [16] is a particular case of cost cumulant minimization [17], where weighting factors are given by the McLaurin coefficients of the power series (8). In fact, the risk sensitive functional defined in the s-th probability space is given by $\theta_s^{-1}\Psi_{\mathsf{P}_s}(\theta_s)$ and we could similarly formulate problem (11) when $K = \infty$ as

$$\begin{aligned}\min_{\boldsymbol{u}^*(\cdot)} \overline{\Psi} &= \min_{\boldsymbol{u}^*(\cdot)} \frac{1}{S}\sum_{s=1}^{S} \theta_s^{-1}\Psi_{\mathsf{P}_s}(\theta_s) \\ &= \min_{\boldsymbol{u}^*(\cdot)} \frac{1}{S}\Big(\sum_{s=1}^{S} \kappa_{\mathsf{P}_s}^{(1)}(J) + \sum_{r=2}^{\infty} \frac{\theta_s^{r-1}}{r!}\kappa_{\mathsf{P}_s}^{(r)}(J)\Big) \ , \quad (12)\end{aligned}$$

by fixing the the high order statistics weighting factors to the above-mentioned McLaurin coefficients divided by θ_s. The cost functional to minimize reduces therefore to an average of standard risk-sensitive functionals. Due to its relevance we will explore both the risk-sensitive (12) and the K-cost-cumulant control (11) problems.

A. Approach

The solution of (12) or (11) with non-linear dynamics and arbitrary costs is in general not attainable. As an alternative, we aim for a local optimum by means of a numerical solution. By linearizing the dynamics and quadratically approximating the cost around a discretized nominal trajectory $\left(\boldsymbol{\xi}^*_{0\cdots T}, \boldsymbol{u}^*_{0\cdots T-1}\right)$, a discrete-time linear quadratic (LQ) approximation of state and control *deviations*, i.e. $\boldsymbol{\xi} = (\delta\boldsymbol{\xi}^* - \boldsymbol{\xi}^*)$ and $\boldsymbol{u} = (\delta\boldsymbol{u}^* - \boldsymbol{u}^*)$ is obtained. Its solution is a gradient towards the local optimum, found by iteratively updating the nominal trajectory and repeating the whole process until convergence.

The local deviations problem is defined as follows. Time is discretized in T steps with sample time $\Delta = T_c/(T)$. The linearized marginal dynamics at time step k are given by

$$\boldsymbol{\xi}_{k+1} = A_k\boldsymbol{\xi}_k + B_k\boldsymbol{u}_k + \boldsymbol{\epsilon}_k^s \quad , \tag{13}$$

where $A_k \in \mathbb{R}^{n\times n}$, $B_k \in \mathbb{R}^{n\times m}$, $\boldsymbol{\epsilon}_k^s \in \mathbb{R}^n$ is an independent identically distributed Gaussian random variable such that $\mathcal{N}(\boldsymbol{\epsilon}_k^s|\mathbf{0}, \Sigma_k^s)$ and $\boldsymbol{x}_0 = \mathbf{0}$. These quantities are calculated as

$$A_k = I_n + \Delta\frac{\partial f_k}{\partial \boldsymbol{\xi}_k^*} \quad B_k = \Delta\frac{\partial f_k}{\partial \boldsymbol{u}_k^*} \quad \Sigma_k^s = \Delta \mathcal{G}_k^{s\,2} \ .$$

The control input is constrained to the form

$$\boldsymbol{u}_k = \boldsymbol{l}_k + L_k\boldsymbol{x}_k \ , \tag{14}$$

where $\boldsymbol{l}_k$ is an affine input and $L_k \in \mathbb{R}^{m\times n}$ is the feedback matrix. The cost functional (3) results in the quadratic approximation

$$\begin{aligned}J(\boldsymbol{\xi}_0, \boldsymbol{u}_{0\cdots T-1}) = \frac{1}{2}\boldsymbol{\xi}_T^\mathsf{T} Q_T\boldsymbol{\xi}_T + \boldsymbol{\xi}_T^\mathsf{T}\boldsymbol{q}_T + \sum_{k=0}^{T-1}\frac{1}{2}\boldsymbol{\xi}_k^\mathsf{T} Q_k\boldsymbol{\xi}_k + \\ + \boldsymbol{\xi}_k^\mathsf{T}\boldsymbol{q}_k + q_k + \frac{1}{2}\boldsymbol{u}_k^\mathsf{T} R_k\boldsymbol{u}_k \ + \boldsymbol{u}_k^\mathsf{T}\boldsymbol{r}_k, \quad (15)\end{aligned}$$

where $Q_k \in \mathbb{R}^{n\times n}$, $R_k \in \mathbb{R}^{m\times m}$, $\boldsymbol{q}_k \in \mathbb{R}^n$ and $\boldsymbol{r}_k \in \mathbb{R}^m$ and $q_k \in \mathbb{R}$, with $Q_k \geq 0$ and $R_k > 0$, are computed as

$$\begin{aligned} q_k = \Delta h_k \qquad \boldsymbol{q}_k = \Delta\frac{\partial h_k}{\partial \boldsymbol{\xi}_k^*} \qquad Q_k = \Delta\frac{\partial^2 h_k}{\partial(\boldsymbol{\xi}_k^*)^2} \\ \boldsymbol{r}_k = \Delta\frac{\partial h_k}{\partial \boldsymbol{u}_k^*} \quad R_k = \Delta\frac{\partial^2 h_k}{\partial(\boldsymbol{u}_k^*)^2} \ . \end{aligned}$$

The local deviations problem consists of minimizing (11) or (12) with J as in (15) and constrained to (13) and (14). In Section III, we first explore the risk-sensitive solution solving problem (12). Results on cost cumulant control follow in Section IV solving the more general problem (11). The main iteration which drives the nominal trajectory towards the local optimum is explained in Section V.

III. Risk-sensitive Solution

In this section we study the solution to the average of marginal risk-sensitive LQ problems $\overline{\Psi}$. A discrete-time LQ optimal control problem is solved by means of dynamic programming applying Bellman's optimality principle.

For simplicity, consider first the risk-sensitive problem in a single probability space. The cost function to minimize is then $\theta^{-1}\Psi_\mathsf{P}(\theta)$ constrained to dynamics

$$\boldsymbol{\xi}_{k+1} = A_k\boldsymbol{\xi}_k + B_k\boldsymbol{u}_k + \boldsymbol{\epsilon}_k \quad , \tag{16}$$

where $\boldsymbol{\epsilon}_k \in \mathbb{R}^q$ is an independent Gaussian random variable with $\mathcal{N}(\boldsymbol{\epsilon}_k|\mathbf{0}, \Sigma_k)$. Assume the system is at sample time k. Given a control law $\boldsymbol{u}_{k\cdots T-1}$ in the form (14) and having observed $\boldsymbol{\xi}_k$, the remaining cost is calculated by means of the backwards recursion

$$\begin{aligned}\theta^{-1}\Psi_\mathsf{P}(\theta, \boldsymbol{\xi}_k, \boldsymbol{u}_{k\cdots T-1}) = J(\boldsymbol{\xi}_k, \boldsymbol{u}_k) \\ + \theta^{-1}\Psi_\mathsf{P}(\theta, \boldsymbol{\xi}_{k+1}, \boldsymbol{u}_{k+1\cdots T-1}) \ , \quad (17)\end{aligned}$$

where the quantity $\theta^{-1}\Psi_\mathsf{P}(\theta, \boldsymbol{\xi}_k, \boldsymbol{u}_{k\cdots T-1})$ is denoted the *cost-to-go*.

The following lemma solves (17) in closed form yielding a quadratic form on $\boldsymbol{\xi}_k$.

Lemma 1. *If* $({\Sigma_k}^{-1} - \theta W_{k+1}) > 0$ *for* $k = 0\cdots T-1$*, the analytic solution to* (17) *is given recursively by the quadratic form*

$$\theta^{-1}\Psi_\mathsf{P}(\theta, \boldsymbol{\xi}_k, \boldsymbol{u}_{k\cdots T-1}) = \tfrac{1}{2}\boldsymbol{\xi}_k^\mathsf{T} W_k\boldsymbol{\xi}_k + \boldsymbol{\xi}_k^\mathsf{T}\boldsymbol{w}_k + w_k \ , \tag{18}$$

where

$$W_k = Q_k + A_k^\mathsf{T}\widetilde{W}_{k+1}A_k + L_k^\mathsf{T}H_kL_k + G_k^\mathsf{T}L_k + L_k^\mathsf{T}G \tag{19}$$

$$\boldsymbol{w}_k = \boldsymbol{q}_k + A_k^\mathsf{T}\widetilde{\boldsymbol{w}}_{k+1} + L_k^\mathsf{T}H_k\boldsymbol{l}_k + L_k^\mathsf{T}\boldsymbol{g}_k + G_k^\mathsf{T}\boldsymbol{l}_k \tag{20}$$

$$w_k = q_k + \widetilde{w}_{k+1} + \frac{1}{2}\boldsymbol{l}_k^\mathsf{T}H\boldsymbol{l}_k + \boldsymbol{l}_k^\mathsf{T}\boldsymbol{g}_k \tag{21}$$

with

$$H_k = R_k + B_k^\mathsf{T}\widetilde{W}_{k+1}B_k \qquad G_k = B_k\widetilde{W}_{k+1}A_k$$
$$\boldsymbol{g}_k = \boldsymbol{r}_k + B_k^\mathsf{T}\widetilde{\boldsymbol{w}}_{k+1}$$

and final conditions $W_N = Q_N, \boldsymbol{w}_N = \boldsymbol{q}_N, w_N = q_N$ *and*

$$\widetilde{W}_{k+1} = W_{k+1} + \theta W_{k+1}(\Sigma_k{}^{-1} - \theta W_{k+1})^{-1}W_{k+1} \tag{22}$$

$$\widetilde{\boldsymbol{w}}_{k+1} = \boldsymbol{w}_{k+1} + \theta W_{k+1}(\Sigma_k{}^{-1} - \theta W_{k+1})^{-1}\boldsymbol{w}_{k+1} \tag{23}$$

$$\widetilde{w}_{k+1} = w_{k+1} + \theta \boldsymbol{w}_{k+1}{}^\mathsf{T}(\Sigma_k{}^{-1} - \theta W_{k+1})^{-1}\boldsymbol{w}_{k+1} - \frac{1}{2}\theta^{-1}\log F_k \ , \tag{24}$$

with $F_k = |I_q - \theta W_{k+1}\Sigma_k|$ *and* $F_T = 1$.

Proof. The solution to the Bellman equation (17) entails only some complication in the term $\theta^{-1}\Psi_\mathsf{P}(\theta, \boldsymbol{\xi}_{k+1}, \boldsymbol{u}_{k+1\cdots T-1})$. Assuming quadratic form (18) holds yields

$$\theta^{-1}\Psi_\mathsf{P}(\theta, \boldsymbol{\xi}_{k+1}, \boldsymbol{u}_{k+1\cdots T-1}) =$$
$$\theta^{-1}\log \mathsf{E}_\mathsf{P}[\exp\{\theta(\boldsymbol{\xi}_{k+1}^\mathsf{T}W_{k+1}\boldsymbol{\xi}_{k+1} + \boldsymbol{\xi}^\mathsf{T}\boldsymbol{w}_{k+1} + w_{k+1})\}] \ .$$

Let now $M_k = A_k\boldsymbol{\xi}_k + B_k\boldsymbol{u}_k$. Considering the system dynamics (16), the expression for the expectation is given by the Gaussian integral

$$\frac{1}{\sqrt{(2\pi)^n|\Sigma_k|}}\exp\Big\{\theta\Big(\frac{1}{2}M_k^\mathsf{T}W_{k+1}M_k + M_k^\mathsf{T}\boldsymbol{w}_{k+1} + w_{k+1}\Big)\Big\} \cdot \int \exp\Big\{-\frac{1}{2}\epsilon_k^\mathsf{T}(\Sigma_k^{-1} - \theta W_{k+1})\epsilon_k + \epsilon_k^\mathsf{T}\Big(2\theta W_{k+1}M_k + \theta \boldsymbol{w}_{k+1}\Big)\Big\}d\boldsymbol{\epsilon}_k \ .$$

If $(\Sigma_k^{-1} - \theta W_{k+1}) > 0$, this expression has analytical solution [16] yielding

$$\sqrt{\frac{|\left(\Sigma_k^{-1} - \theta_s W_{k+1}\right)^{-1}|}{|\Sigma_k|}} \cdot \exp\{\theta_s\frac{1}{2}M_k^\mathsf{T}\widetilde{W}_{k+1}M_k + \theta_s M_k^\mathsf{T}\widetilde{\boldsymbol{w}}_{k+1} + \theta_s\widetilde{w}_{k+1}\}$$

where $\widetilde{W}_{k+1}$, $\widetilde{\boldsymbol{w}}_{k+1}$, and $\widetilde{w}_{k+1}$ are defined as in (22), (23) and (24). Substituting into (17) and constraining $\boldsymbol{u}_k$ to (14) yields expressions (19), (20) and (21). ∎

The cost-to-go for problem (12), i.e. the remaining average cost of all marginal problems at time step k, follows now straightforwardly as

$$\overline{\Psi}(\boldsymbol{\xi}_k, \boldsymbol{u}_{k\cdots T-1}) = \frac{1}{2}\boldsymbol{\xi}_k^\mathsf{T}\overline{W}_k\boldsymbol{\xi}_k + \boldsymbol{\xi}_k^\mathsf{T}\overline{\boldsymbol{w}}_k + \overline{w}_k \ , \tag{25}$$

where

$$\overline{W}_k = \frac{1}{S}\sum_{s=1}^{S}W_k^s \quad \overline{\boldsymbol{w}}_k = \frac{1}{S}\sum_{s=1}^{S} = \boldsymbol{w}_k^s \quad \overline{w}_k = \frac{1}{S}\sum_{s=1}^{S}w_k^s$$

and where W_k^s, $\boldsymbol{w}_k^s$ and w_k^s are the quadratic coefficients resulting from applying lemma 1 in the s-th probability space constrained to the s-th marginal dynamics, i.e. with $\Sigma_k = \Sigma_k^s$ and $\theta = \theta_s$. Note that the overall cost given control policy $\boldsymbol{u}_{0\cdots T-1}$ corresponds to the cost-to-go at $k = 0$, i.e. $\overline{\Psi}(\boldsymbol{\xi}_0, \boldsymbol{u}_{0\cdots T-1})$.

The risk-sensitive solution is computed applying Bellman's optimal equation, i.e. minimizing (25) w.r.t L_k and $\boldsymbol{l}_k$ at each step of the backwards recursion yielding

$$L_k = -\left(R_k + B_k^\mathsf{T}\overline{\widetilde{W}}_{k+1}B_k\right)^{-1}(B_k^\mathsf{T}\overline{\widetilde{W}}_{k+1}A_k)$$
$$\boldsymbol{l}_k = -\left(R_k + B_k^\mathsf{T}\overline{\widetilde{W}}_{k+1}B_k\right)^{-1}(B_k^\mathsf{T}\overline{\widetilde{\boldsymbol{w}}}_{k+1} + \boldsymbol{r}_k) \ , \tag{26}$$

where

$$\overline{\widetilde{W}}_k = \frac{1}{S}\sum_{s=1}^{S}\widetilde{W}_k^s \quad \overline{\widetilde{\boldsymbol{w}}}_k = \frac{1}{S}\sum_{s=1}^{S} = \widetilde{\boldsymbol{w}}_k^s \quad \overline{\widetilde{w}}_k = \frac{1}{S}\sum_{s=1}^{S}\widetilde{w}_k^s \ ,$$

and where $\widetilde{W}_k^s$, $\widetilde{\boldsymbol{w}}_k^s$ and $\widetilde{w}_k^s$ follow from computing (25).

IV. Cost-cumulant solution

In this section we approach the cost-cumulant control problem (11) studied in [15] for the continuous-time case and in [18] for the discrete-time case. We extend it here to our problem setting with a more general family of LQ systems and explore the average of marginal cost-cumulant problems. In contrast to the risk-sensitive solution, the cost-cumulant control problem allows more flexible decision-makers as the number of considered cumulants and their respective weighting factors are design parameters.

Following a similar treatment to the previous section, we first consider a single probability space and dynamics (16). From lemma 1, the cumulant generating function of the cost-to-go at time step k is given by

$$\Psi_\mathsf{P}(\theta, \boldsymbol{\xi}_k, \boldsymbol{u}_{k\cdots T-1}) = \theta\Big(\frac{1}{2}\boldsymbol{\xi}_k^\mathsf{T}W_k\boldsymbol{\xi}_k + \boldsymbol{\xi}_k^\mathsf{T}\boldsymbol{w}_k + w_k\Big) \ .$$

Recall that the r-th cumulant is calculated by means of expression (9). In the following, in line with previous work, the notation $f(\theta)^{[r]}$ denotes the r-th derivative of $f(\theta)$ w.r.t θ at point $\theta = 0$, i.e. $f(\theta)^{[r]} = \frac{\partial^r f(\theta)}{\partial\theta^r}\Big|_{\theta=0}$. An $r = 0$ index corresponds to the function itself at $\theta = 0$. Additionally, $\mathcal{C}_r^j$ denotes the binomial coefficient $\binom{j}{r}$.

Cost cumulants are calculated recursively in a similar manner to [18], yielding the following lemma.

Lemma 2. *The d-th cost cumulant of the random cost* (15) *at sample time k is given by*

$$\kappa_\mathsf{P}^{(d)}(J_{k\cdots T-1}) = d\Big(\tfrac{1}{2}\boldsymbol{\xi}_k^\mathsf{T}W_k^{[d-1]}\boldsymbol{\xi}_k + \boldsymbol{\xi}_k^\mathsf{T}\boldsymbol{w}_k^{[d-1]} + w_k^{[d-1]}\Big)$$

where for $d = 1$

$$W_k^{[0]} = Q_k + A_k^\mathsf{T}\widetilde{W}_{k+1}^{[0]}A_k + L_k^\mathsf{T}H_k^{[0]}L_k + G_k^{[0]\mathsf{T}}L_k + L_k^\mathsf{T}G^{[0]}{}_k$$
$$\boldsymbol{w}_k^{[0]} = \boldsymbol{q}_k + A_k^\mathsf{T}\widetilde{\boldsymbol{w}}_{k+1}^{[0]} + L_k^\mathsf{T}H_k^{[0]}\boldsymbol{l}_k + L_k^\mathsf{T}\boldsymbol{g}_k^{[0]} + G_k^{[0]\mathsf{T}}\boldsymbol{l}_k$$
$$w_k^{[0]} = q_k + \widetilde{w}_{k+1}^{[0]} + \frac{1}{2}\boldsymbol{l}_k^\mathsf{T}H_k^{[0]}\boldsymbol{l}_k + \boldsymbol{l}_k^\mathsf{T}\boldsymbol{g}_k^{[0]}$$

with

$$H_k^{[0]} = R_k + B_k^\mathsf{T} \widetilde{W}_{k+1}^{[0]} B_k \qquad G_k^{[0]} = B_k^\mathsf{T} \widetilde{W}_{k+1}^{[0]} A_k$$
$$\boldsymbol{g}_k^{[0]} = \boldsymbol{r}_k + B_k^\mathsf{T} \widetilde{\boldsymbol{w}}_{k+1}^{[0]}$$

and final conditions $W_N^{[0]} = Q_T, \boldsymbol{w}_N^{[0]} = \boldsymbol{q}_T, w_N^{[0]} = q_T$ *and for* $d > 1$

$$\begin{aligned} W_k^{[r]} =& A_k^\mathsf{T} \widetilde{W}_{k+1}^{[r]} A_k + L_k^\mathsf{T} H_k^{[r]} L_k + G_k^{[r]\,\mathsf{T}} L_k + L_k^\mathsf{T} G_k^{[r]} \\ \boldsymbol{w}_k^{[r]} =& A_k^\mathsf{T} \widetilde{\boldsymbol{w}}_{k+1}^{[r]} + L_k^\mathsf{T} H_k^{[r]} \boldsymbol{l}_k + L_k^\mathsf{T} \boldsymbol{g}_k^{[r]} + G_k^{[r]\,\mathsf{T}} \boldsymbol{l}_k \\ w_k^{[r]} =& \widetilde{w}_{k+1}^{[r]} + \frac{1}{2} \boldsymbol{l}_k^\mathsf{T} H_k^{[r]} \boldsymbol{l}_k + \boldsymbol{l}_k^\mathsf{T} \boldsymbol{g}_k^{[r]} \end{aligned}$$

with

$$H_k^{[r]} = B_k^\mathsf{T} \widetilde{W}_{k+1}^{[r]} B_k \qquad G_k^{[r]} = B_k^\mathsf{T} \widetilde{W}_{k+1}^{[r]} A_k$$
$$\boldsymbol{g}_k^{[r]} = B_k^\mathsf{T} \widetilde{\boldsymbol{w}}_{k+1}^{[r]}$$

and final conditions $W_N^{[r]} = 0^{n\times n}, \boldsymbol{w}_N^{[r]} = \mathbf{0}, w_N^{[r]} = 0$ *and*

$$\widetilde{W}_{k+1}^{[r]} = W_{k+1}^{[r]} + r \sum_{j=0}^{r-1} \mathcal{C}_{r-1}^j W_{k+1}^{[j]} \Sigma_k \widetilde{W}_{k+1}^{[r-1-j]} \tag{27}$$

$$\widetilde{\boldsymbol{w}}_{k+1}^{[r]} = \boldsymbol{w}_{k+1}^{[r]} + r \sum_{j=0}^{r-1} \mathcal{C}_{r-1}^j W_{k+1}^{[j]} \Sigma_k \widetilde{\boldsymbol{w}}_{k+1}^{[r-1-j]} \tag{28}$$

$$\widetilde{w}_{k+1}^{[r]} = w_{k+1}^{[r]} + r \sum_{j=0}^{r-1} \mathcal{C}_{r-1}^j \boldsymbol{w}_{k+1}^{[j]} \Sigma_k \widetilde{\boldsymbol{w}}_{k+1}^{[r-1-j]} + \sum_{t=k}^{T-1} \mathrm{Tr}(H_t^{[r-1]}) \tag{29}$$

with $H_t^{[r]} = (r+1) W_{t+1}^{[r]} \Sigma_t + r \sum_{j=1}^{r-1} \mathcal{C}_{r-1}^j W_{t+1}^{[j]} \Sigma_t H_t^{[r-1-j]}$.

Proof. The results for $W_k^{[r]}$, $\boldsymbol{w}_k^{[r]}$ and $w_k^{[r]}$ are straightforward from their definitions in (19), (20) and (21). For the proof of expression $\widetilde{W}_{k+1}^{[r]}$ and $H_t^{[r]}$ see [18]. The solution to $\widetilde{\boldsymbol{w}}_{k+1}^{[r]}$ and $\widetilde{w}_{k+1}^{[r]}$ is derived with the same procedure. The definition of $\widetilde{\boldsymbol{w}}_{k+1}$ from (23) is reformulated as

$$\begin{aligned} \widetilde{\boldsymbol{w}}_{k+1} &= \boldsymbol{w}_{k+1} + \theta W_{k+1} (\Sigma_k^{-1} - \theta W_{k+1})^{-1} \boldsymbol{w}_{k+1} \\ &= \boldsymbol{w}_{k+1} - (\Sigma_k^{-1} - \theta W_{k+1} - \Sigma_k^{-1}) \\ &\qquad (\Sigma_k^{-1} - \theta W_{k+1})^{-1} \boldsymbol{w}_{k+1} \\ &= (I_q - \theta W_{k+1} \Sigma_k)^{-1} \boldsymbol{w}_{k+1} \ , \end{aligned}$$

from which we can write

$$\widetilde{\boldsymbol{w}}_{k+1} = \boldsymbol{w}_{k+1} + \theta W_{k+1} \Sigma_k \widetilde{\boldsymbol{w}}_{k+1} \ .$$

Applying Leibniz's Theorem twice we get

$$\begin{aligned} \widetilde{\boldsymbol{w}}_{k+1}^{[r]} &= \boldsymbol{w}_{k+1}^{[r]} + \theta (W_{k+1} \Sigma_k \widetilde{\boldsymbol{w}}_{k+1})^{[r]} + r (W_{k+1} \Sigma_k \widetilde{\boldsymbol{w}}_{k+1})^{[r-1]} \\ &= \boldsymbol{w}_{k+1}^{[r]} + \theta (W_{k+1} \Sigma_k \widetilde{\boldsymbol{w}}_{k+1})^{[r]} + r \sum_{j=0}^{r-1} \mathcal{C}_{r-1}^j W_{k+1}^{[j]} \Sigma_k \widetilde{\boldsymbol{w}}_{k+1}^{[r-1-j]}, \end{aligned}$$

which after evaluating $\theta = 0$ yields (28).

The procedure for $\widetilde{w}_{k+1}^{[r]}$ is similar. ■

The evaluation of the average of marginal cost-cumulants from (11) at time step k follows immediately as

$$\Xi(\boldsymbol{\xi}_k) = \tfrac{1}{2} \boldsymbol{\xi}_k^\mathsf{T} \widehat{W}_k \boldsymbol{\xi}_k + \boldsymbol{\xi}_k^\mathsf{T} \widehat{\boldsymbol{w}}_k + \widehat{w}_k \ , \tag{30}$$

where

$$\widehat{W}_k = \frac{1}{S} \sum_{s=1}^{S} \sum_{r=1}^{K} r \gamma_{r,s} W_k^{s\,[r-1]} \quad \widehat{\boldsymbol{w}}_k = \frac{1}{S} \sum_{s=1}^{S} \sum_{r=1}^{K} r \gamma_{r,s} \boldsymbol{w}^{s\,[r-1]}_{k+1}$$
$$\widehat{w}_k = \frac{1}{S} \sum_{s=1}^{S} \sum_{r=1}^{K} r \gamma_{r,s} w_k^{s\,[r-1]}$$

and where $W_k^{s\,[r]}$ and $\boldsymbol{w}_k^{s\,[r]}$ and $w_k^{s\,[r]}$ are the quadratic coefficients of the r-th cumulant resulting from applying lemma 2 in the s-th probability space constrained to the s-th marginal dynamics and $\gamma_{1,s} = 1$.

The solution to the cost cumulant control problem (11) follows from minimizing (30) w.r.t L_k and $\boldsymbol{l}_k$ at each step of the backwards recursion yielding

$$\begin{aligned} L_k =& -(R_k + B_k^\mathsf{T} \widehat{\widetilde{W}}_{k+1} B_k)^{-1} (B_k^\mathsf{T} \widehat{\widetilde{W}}_{k+1} A_k) \\ \boldsymbol{l}_k =& -(R_k + B_k^\mathsf{T} \widehat{\widetilde{W}}_{k+1} B_k)^{-1} (B_k^\mathsf{T} \widehat{\widetilde{\boldsymbol{w}}}_{k+1} + \boldsymbol{r}_k) \ , \end{aligned} \tag{31}$$

where

$$\widehat{\widetilde{W}}_k = \frac{1}{S} \sum_{s=1}^{S} \sum_{r=1}^{K} \gamma_{r,s} \widetilde{W}_k^{s[r-1]} \quad \widehat{\widetilde{\boldsymbol{w}}}_k = \frac{1}{S} \sum_{s=1}^{S} \sum_{r=1}^{K} \gamma_{r,s} \widetilde{\boldsymbol{w}}_k^{s[r-1]}$$

and where $\widetilde{W}_{k+1}^{s[r-1]}$ and $\widetilde{\boldsymbol{w}}_{k+1}^{s[r-1]}$ result from the computation of the r-th cumulant in the s-th probability space in (30).

The effect of weightings $\gamma_{r,s}$ on the resulting cost (30) is evident: risk-aversion is achieved by selecting $\gamma_{r,s} > 0$ and thereby increasing the resulting quadratic coefficients. Selecting $\gamma_{r,s} < 0$ has the opposite effect yielding a risk-seeking evaluation.

V. Main Iteration

A numerical approximation that computes locally optimal solutions in nonlinear and non-quadratic problems requires a procedure that iteratively approximates and updates a nominal trajectory. The main iteration of our approach follows the iLQG algorithm presented in [9]. In this section we summarize it pointing out the subtle changes that arise due to the different problem setting. The resulting algorithm is either an iterative Linear Exponential Quadratic Regulator (iLEQR) for the risk-sensitive case or an iterative K-Cost Cumulant Regulator (iKCCR) for the cost cumulant optimization.

In order to find a local optimum and given an initial state $\boldsymbol{\xi}_0^*$, the algorithm iterates around the nominal control trajectory $\overline{\boldsymbol{u}}_{0\cdots T-1}^*$ by calculating the optimal control deviations that improve the expected performance. At the i-th iteration, the optimal solution is denoted $\boldsymbol{u}_{0\cdots T-1}^{*(i)} = \overline{\boldsymbol{u}}_{0\cdots T-1}^{*(i)} + L_{0\cdots T-1}^{(i)} \boldsymbol{\xi}_{0\cdots T-1}$ and is updated to obtain the next $\boldsymbol{u}_{0\cdots T-1}^{*(i+1)}$ following the next steps:

1) The corresponding state trajectory $\boldsymbol{\xi}_{0\cdots T}^*$ is computed simulating the discretized dynamics, for instance by Euler integration .i.e. $\boldsymbol{\xi}_{k+1}^* = \boldsymbol{\xi}_k^* + \Delta f(\boldsymbol{\xi}_k^*, \overline{\boldsymbol{u}}_k^{*(i)})$.

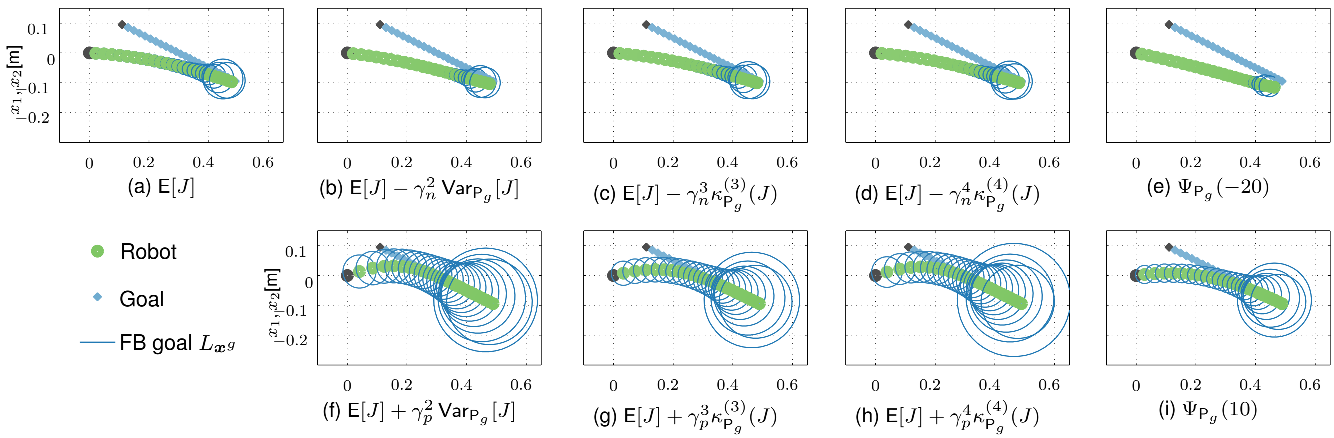


Fig. 1. Optimal trajectories and feedback gains for a 2D point mass damper robot $\boldsymbol{x}^r$ tracking goal $\boldsymbol{x}^g$ with noisy passive mass-damper system dynamics and where $\gamma_n = 2$ and $\gamma_p = 7$ for a horizon of $T_c = 0.5s$. Initial states of robot and goal are $\boldsymbol{x}_0^r = [0\ 0]$, $\dot{\boldsymbol{x}}_0^r = [0\ 0]$, $\boldsymbol{x}_0^g = [0.1\ 0.1]$ and $\dot{\boldsymbol{x}}_0^g = [1\ -0.5]$. Results are shown every 0.02s. Ellipses represent the corresponding feedback gain matrices in terms of their eigenvectors.

2) An LQ approximation of state and control deviations, i.e. $\boldsymbol{\xi}_k = \delta\boldsymbol{\xi}_k^* - \boldsymbol{\xi}_k^*$ and $\boldsymbol{u}_k = \delta\boldsymbol{u}_k^* - \boldsymbol{u}_k^*$ around $(\boldsymbol{\xi}_{0\cdots T}^*, \overline{\boldsymbol{u}}_{0\cdots T-1}^*)$ is computed as explained in Section II-A.
3) The optimal deviations law $\boldsymbol{u}_k = \boldsymbol{l}_k + L_k\boldsymbol{\xi}_k$ is computed by means of either the risk-sensitive (26) or cost-cumulant solution (31).
4) As solution $\boldsymbol{u}_k$ holds only in the close vicinity of the current nominal trajectory, a line search algorithm aims for an adapted step in the feedforward component of $\boldsymbol{u}_k$ that yields a policy improvement as $\boldsymbol{u}_k^{*(i+1)}(\alpha) = \overline{\boldsymbol{u}}_k^{*(i)} + \alpha\boldsymbol{l}_k + L_k^{(i+1)}\boldsymbol{\xi}_k$, where α is the line search parameter. If it converges, finish, otherwise go back to 1).

Note that the evaluation of the expected performance in step 4) is not straightforward. As analytical expressions of (11) or (12) for arbitrary problems are rarely available, an LQ approximation of the cost around the new trajectory is obtained as in step 2) and its expected performance is computed by means of lemma 1 or 2 respectively.

VI. Evaluation

In order to illustrate the peculiarities of the proposed approach in simple robotic scenarios, we apply the iterative algorithm from Section V on simulations of a point-mass robot and a two-link manipulator. In both settings we consider a cost function

$$J(\boldsymbol{u}(\cdot)) = h_g(\boldsymbol{x}^r(T_c), \boldsymbol{x}^g(T_c)) + \int_{t=0}^{T_c} h_g(\boldsymbol{x}^r(t), \boldsymbol{x}^g(t)) + h_o(\boldsymbol{x}^r(t), \boldsymbol{x}^o(t)) + h_{so}(\boldsymbol{x}^r(t), \boldsymbol{x}^{so}) + h_u(\boldsymbol{u}(t))dt \ , \tag{32}$$

where $h_g = \frac{1}{2}\boldsymbol{\xi}^{*\mathsf{T}} Q_g \boldsymbol{\xi}^*$ and $h_u = \frac{1}{2}\boldsymbol{u}^{*\mathsf{T}} R \boldsymbol{u}^*$ penalize distance to goal and control efforts, $h_o = w \exp\{-0.5(\boldsymbol{x}^r - \boldsymbol{x}^o)^{\mathsf{T}} Q_o (\boldsymbol{x}^r - \boldsymbol{x}^o)\}$ favor configurations distant to dynamic obstacles and h_{so} has similar definition considering a static obstacle $\boldsymbol{x}^{so}$.

A. 2D Point-Mass robot

Consider a two-dimensional point robot with position $\boldsymbol{x}^r \in \mathrm{R}^2$, no orientation and second-order dynamics given by the mass-damper system

$$M\ddot{\boldsymbol{x}}^{\boldsymbol{r}} + D\dot{\boldsymbol{x}}^{\boldsymbol{r}} = \boldsymbol{u}^* \ ,$$

where $M, D \in \mathbb{R}^2$ are the mass and the damping. We first consider a task consisting only of following a goal $\boldsymbol{x}^g \in \mathrm{R}^2$ whose stochastic dynamics are constrained to a passive mass-damper system with noise, i.e. $h_o = h_{so} = 0$ in (32). This is an equivalent scenario to the problem of tracking a learned Dynamical Movement Primitive (DMP) [19] with a probabilistic estimator for its nonlinear forcing term. The error state of the system is given by $\boldsymbol{\xi}^* = [(\boldsymbol{x}^g - \boldsymbol{x}^r)^{\mathsf{T}}\ (\dot{\boldsymbol{x}}^g - \dot{\boldsymbol{x}}^r)^{\mathsf{T}}]^{\mathsf{T}}$. The optimal policy takes the form $\boldsymbol{u}^* = \overline{\boldsymbol{u}}^* + [L_{\boldsymbol{x}^g}\ L_{\dot{\boldsymbol{x}}^g}]\boldsymbol{\xi}$. In the following results, parameters were fixed to $M = I_2$ kg, $D = I_2$ Ns/m, $R = 10^{-2}I_2$, $Q_g = diag\{\ 1\ \ 1\ \ 0.1\ \ 0.1\ \}$, $\Delta = 10^{-2}$ and $T_c = 0.5s$. Goal dynamics are assumed to have infinitesimal variance $\mathcal{G}^g = I_4$ and identical mass and damping to the robot's.

Optimal trajectories and positional gains $L_{\boldsymbol{x}^g}$ for several cost cumulant and risk-sensitive controllers are depicted in Fig. 1. The solution corresponding to the expected cost depicted in Fig. 1(a) serves as the risk-neutral reference. Fig. 1(e) and Fig. 1(i) show risk-sensitive solutions in their seeking and averse variants respectively. The risk-seeking policy not only adopts significantly lower feedback gains but also adapts its feedforward trajectory to a less accurate positional tracking. This policy is desirable when goal uncertainty suggests more flexibility, e.g. in PbD (Programming by demonstration) settings [14]. The risk-averse solution has the opposite effect, tracking goal dynamics more aggressively as well as increasing feedback gains. This behavior is in accordance with navigation scenarios where uncertainty may hinder performance [5]. Cost-cumulant solutions are shown in Fig. 1(f), Fig. 1(g), Fig. 1(h) for the second, third and fourth cumulant in their risk-averse variants. All three cases are similar: the feedforward

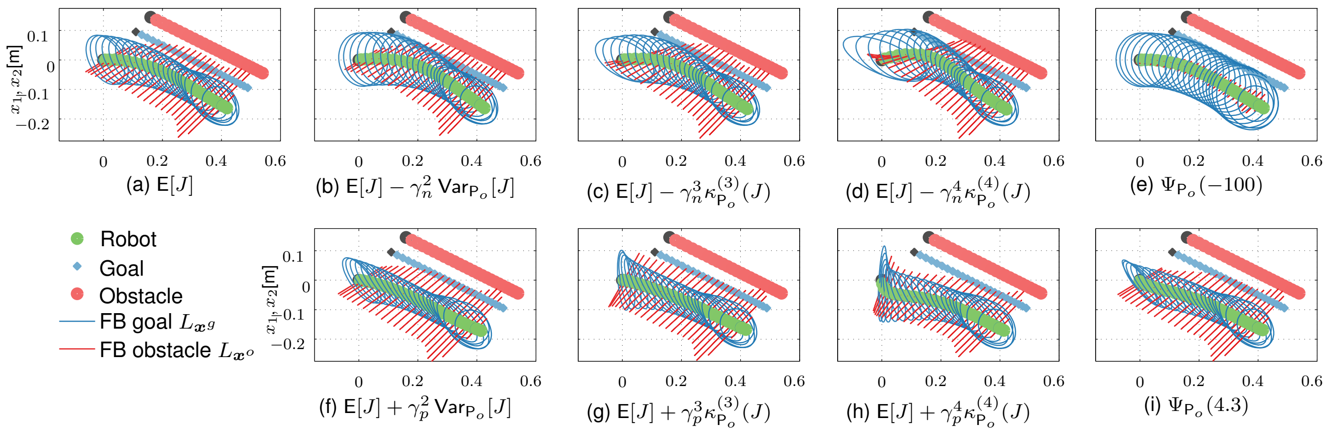


Fig. 2. Optimal trajectories and feedback gains for a 2D point mass damper robot $\boldsymbol{x}^r$ tracking goal $\boldsymbol{x}^g$ with passive mass-damper system dynamics and avoiding obstacle $\boldsymbol{x}^o$ with noisy passive mass-damper system dynamics and where $\gamma_n = 1.9$ and $\gamma_p = 3$ for a horizon of $T_c = 0.5s$. Initial states of robot, goal and obstacle are $\boldsymbol{x}_0^r = [0\ 0]$, $\dot{\boldsymbol{x}}_0^r = [0\ 0]$, $\boldsymbol{x}_0^g = [0.1\ 0.1]$, $\dot{\boldsymbol{x}}_0^g = [1\ \ -0.5]$, $\boldsymbol{x}_0^o = [0.15\ 0.15]$ and $\dot{\boldsymbol{x}}_0^o = [1\ \ -0.5]$. Results are shown every 0.02s.

trajectory is significantly adapted aiming for more accurate tracking together with higher feedback gains. Their risk-seeking counterparts depicted in Fig. 1(b), Fig. 1(c), Fig. 1(d) show little difference w.r.t the risk-neutral policy Fig. 1(a).

We now include a dynamic obstacle with position $\boldsymbol{x}^o$ with passive mass-damper dynamics with noise. This is a similar setting to the one considered in [13]. The error state of the system is now given by

$$\boldsymbol{\xi}^* = [(\boldsymbol{x}^g - \boldsymbol{x}^r)^\mathsf{T}\ (\dot{\boldsymbol{x}}^g - \dot{\boldsymbol{x}}^r)^\mathsf{T}\ (\boldsymbol{x}^o - \boldsymbol{x}^r)^\mathsf{T}\ (\dot{\boldsymbol{x}}^o - \dot{\boldsymbol{x}}^r)^\mathsf{T}]^\mathsf{T}\ .$$

The h_o term in (3) cost is now parameterized as $Q_o = 200I_2$ and $w = 0.1$. The optimal policy is now $\boldsymbol{u}^* = \overline{\boldsymbol{u}}^* + \begin{bmatrix} L_{\boldsymbol{x}^g} & L_{\dot{\boldsymbol{x}}^g} \\ L_{\boldsymbol{x}^o} & L_{\dot{\boldsymbol{x}}^o} \end{bmatrix} \boldsymbol{\xi}$. Obstacle dynamics are assumed to have infinitesimal variance $\mathcal{G}^o = I_4$ and identical mass and damping to the robot's.

The optimal trajectory, positional goal gain $L_{\boldsymbol{x}^g}$ and obstacle gain $L_{\boldsymbol{x}^o}$ for several cost cumulant and risk-sensitive controllers are illustrated in Fig. 2 considering only the marginal variability produced by the obstacle's uncertainty. The standard risk-neutral policy is shown in Fig. 2(a). Risk-sensitive solutions are depicted in Fig. 2(e) and Fig. 2(i) for the risk-seeking and the risk-averse variants. The risk-seeking solution decreases obstacle gains $L_{\boldsymbol{x}^o}$ dramatically and as a consequence positional gains increase. In contrast, the risk-averse solution adapts feedforward terms, choosing a trajectory further away to the obstacle. Cost-cumulant solutions are illustrated in Fig. 2(f), Fig. 2(g), Fig. 2(h) for risk-averse variants of the second, third and fourth cumulant respectively. Especially for the fourth cumulant, solutions show adapted trajectories aiming for configurations distant to the obstacle. The risk-seeking cost-cumulant solutions are shown in Fig. 2(b), Fig. 2(c), Fig. 2(d) for the second, third and fourth cumulant respectively. In contrast to the risk-sensitive solution from Fig. 2(e), feedforward terms are significantly adapted obtaining a trajectory closer to the obstacle, especially for the fourth cumulant.

We now explore solutions considering both the goal's and the obstacle's marginal variabilities. We consider goal and obstacle variability it in a risk-seeking and risk-averse way respectively. Results for a risk-sensitive and a mean-variance controller are depicted in Fig. 3. The mean-variance solution is depicted in Fig. 3(b), which clearly shows a different trajectory to the risk-neutral one from Fig. 3(a). It is also interesting to compare this solution with Fig. 2(f), as both policies consider the same risk-averse marginal obstacle variability $\gamma_p^2 \,\mathsf{Var}_{\mathsf{P}_o}[J]$. The inclusion of the risk-seeking marginal goal variability $-\gamma_n^2 \,\mathsf{Var}_{\mathsf{P}_o}[J]$ demands less accurate goal tracking enabling a more pronounced adaptation to the obstacle's stochasticity. Similar although less remarkable are the results for the risk-sensitive optimization shown in Fig. 3(c).

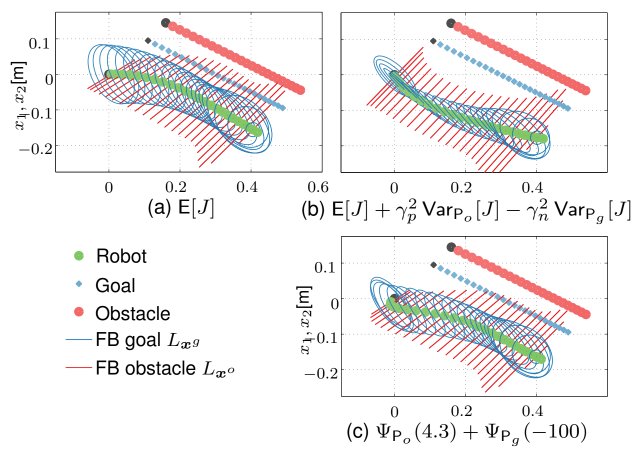


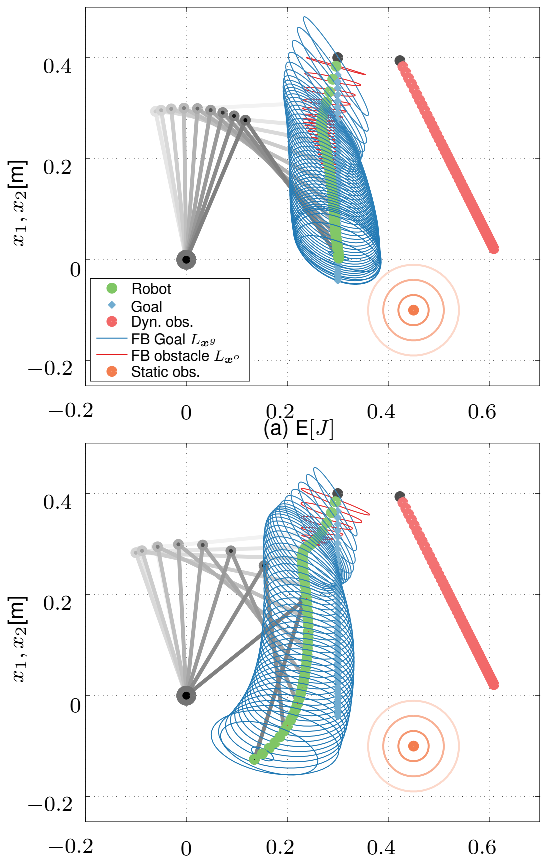
Fig. 3. Optimal trajectories and feedback gains of the same problem from Fig.2 considering both the goal's and the obstacle's marginal variability, where $\gamma_n = 3.5$ and $\gamma_p = 3$. Results are shown every 0.02s.

B. Two-link manipulator

We consider now a torque-controlled arm with two joints moving in the horizontal plane with inverse dynamics

$$\mathcal{M}(\theta)\ddot{\theta} + \mathcal{C}(\theta, \dot{\theta}) + \mathcal{B}\dot{\theta} = \boldsymbol{\tau}\ ,$$

where $\theta \in \mathrm{R}^2$ are the joint angles, $\mathcal{M}(\theta)$ is the inertia matrix, $\mathcal{C}(\theta, \dot{\theta})$ is the vector of centripetal and Coriolis forces, $\mathcal{B}$ is the joint friction matrix and $\boldsymbol{\tau} \in \mathrm{R}^2$ are the joint torques. Following [9] we set the mass of each



(a) $\mathsf{E}[J]$

(b) $\mathsf{E}[J] + \gamma_p^2 \,\mathsf{Var}_{\mathsf{P}_o}[J] - \gamma_n^2 \,\mathsf{Var}_{\mathsf{P}_g}[J]$

Fig. 4. Optimal trajectories and feedback gains for a two-link manipulator $\boldsymbol{x}^r$ tracking goal $\boldsymbol{x}^g$ with passive mass-damper system dynamics and avoiding obstacle $\boldsymbol{x}^o$ with noisy passive mass-damper system dynamics including static obstacles and where $\gamma_n = 6$ and $\gamma_p = 3$. The static obstacle considered is normally distributed, centered at $(-0.1, 0.45)$ with covariance matrix $0.3 I_2$. Initial states of robot, goal and obstacle are $\boldsymbol{x}_0^r = [0.3\ 0.4]$, $\dot{\boldsymbol{x}}_0^r = [0\ 0]$, $\boldsymbol{x}_0^g = [0.3\ 0.4]$, $\dot{\boldsymbol{x}}_0^g = [0\ -0.7]$, $\boldsymbol{x}_0^o = [0.42\ 0.4]$ and $\dot{\boldsymbol{x}}_0^o = [0.3\ -0.6]$. Results are shown every 0.02s.

link to $m_1 = 1.4$kg and $m_2 = 1.1$kg, the length of each link to $l_1 = 0.3$m and $l_2 = 0.33$, the moments of inertia to $I_1 = 0.025\ \mathrm{kg \cdot m^2}$ and $I_2 = 0.045\ \mathrm{kg \cdot m^2}$ and we assume the center of mass of each link is placed at the link's center. The joint friction matrix is set to $\mathcal{B} = \begin{bmatrix} 0.05 & 0.025 \\ 0.025 & 0.05 \end{bmatrix}$. We consider a similar setting as in Fig. 3 but adding a static normally distributed obstacle in the scene by enabling the term h_{so} in (32) with $Q_o = I_2$ and $w = 0.1$ and evaluating its expectation a priori [2]. The rest of the parameters are identical to the previous subsection.

Optimal trajectories and feedback gains are shown in Fig. 4 for the expected cost policy in comparison with the mean variance solution. The risk-averse evaluation of obstacle variability drives the optimal trajectory away from the expected obstacle trajectory acting in a conservative manner. This effect is also boosted by the risk-seeking evaluation of goal variability, which provides less tracking accuracy thereby enabling a more pronounced avoidance of the static obstacle.

[2]The expectation of an exponential already considers obstacle variability, see for instance [20]

VII. Conclusion and Future Work

In this work we present a systematic approach to the design of uncertainty-dependent decision-makers. Our approach extends the family of solutions of the iLQG framework by considering high-order cost statistics either by means of a risk-sensitive optimization or an arbitrary number of cost cumulants. The exploration of multiplicative noise settings and model predictive control and belief-space implementations are the matter of our ongoing work.

ACKNOWLEDGMENTS

This research is partly supported by the ERC Starting Grant "Control based on Human Models (con-humo)" under grant agreement 337654.

References

[1] S. Thrun, W. Burgard, and D. Fox, *Probabilistic Robotics*. MIT Press, 2005.

[2] A. Billard, S. Calinon, R. Dillmann, and S. Schaal, "Robot Programming by Demonstration," in *Handbook of Robotics*, B. Siciliano and O. Khatib, Eds. Springer, 2008, pp. 1371–1394.

[3] E. Todorov, "Optimality principles in sensorimotor control." *Nature Neuroscience*, vol. 7, pp. 907–15, 2004.

[4] D. Braun, A. Nagengast, and D. Wolpert, "Risk-sensitivity in sensorimotor control," *Frontiers in Human Neuroscience*, vol. 5, no. 1, 2011.

[5] J. Müller and G. Sukhatme, "Risk-aware trajectory generation with application to safe quadrotor landing," in *Proc. IEEE/RSJ IROS*, Chicago, IL, USA, September 2014.

[6] S. K. D. Mitrovic and S. Vijayakumar, "Learning impedance control of antagonistic systems based on stochastic optimization principles," *IJRR*, no. 5, pp. 556–573, 2010.

[7] S. Calinon, F. Guenter, and A. Billard, "On learning, representing, and generalizing a task in a humanoid robot," *IEEE Trans. Syst., Man, Cybern. B*, vol. 37, no. 2, pp. 286 –298, 2007.

[8] J. Medina, D. Lee, and S. Hirche, "Risk Sensitive Optimal Feedback Control for Haptic Assistance," in *Proc. IEEE ICRA*, 2012.

[9] W. Li and E. Todorov, "Iterative linearization methods for approximately optimal control and estimation of non-linear stochastic system," *International Journal of Control*, vol. 80, no. 9, pp. 1439–1453, 2007.

[10] D. Jacobsen and D. Mayne, *Differential Dynamic Programming*. Elsevier, 1970.

[11] V. Huynh and N. Roy, "icLQG: Combining local and global optimization for control in information space," in *Proc. IEEE IROS*, 2009, pp. 2851–2858.

[12] J. Berg, S. Patil, and R. Alterovitz, "Motion planning under uncertainty using iterative local optimization in belief space," *The International Journal of Robotics Research*, vol. 31, no. 11, pp. 1263–1278, 2012.

[13] G. Aoude, B. Luders, J. Joseph, N. Roy, and J. How, "Probabilistically safe motion planning to avoid dynamic obstacles with uncertain motion patterns," *Autonomous Robots*, vol. 35, no. 1, pp. 51–76, 2013.

[14] S. Calinon, F. D'Halluin, E. Sauser, D. Caldwell, and A. Billard, "Learning and reproduction of gestures by imitation," *IEEE Robot. Autom. Mag.*, vol. 17, no. 2, pp. 44 –54, 2010.

[15] K. Pham, *Linear-Quadratic Controls in Risk-Averse Decision Making: Performance-Measure Statistics and Control Decision Optimization*. Springer, 2012.

[16] D. Jacobson, "Optimal stochastic linear systems with exponential performance criteria and their relation to deterministic differential games," *IEEE Trans. Autom. Control*, vol. 18, no. 2, pp. 124 – 131, 1973.

[17] C. Won, "Cost distribution shaping: the relation between bode integral, entropy, risk-sensitivity, and cost cumulant control," in *Proc. ACC*, vol. 3, 2004, pp. 2160–2165 vol.3.

[18] F. Qian, J. Gao, and D. Li, "Complete statistical characterization of discrete-time lqg and cumulant control," *IEEE Trans. Autom. Control*, vol. 57, no. 8, pp. 2110–2115, 2012.

[19] A. J. Ijspeert, J. Nakanishi, H. Hoffmann, P. Pastor, and S. Schaal, "Dynamical movement primitives: learning attractor models for motor behaviors," *Neural computation*, vol. 25, no. 2, pp. 328–373, 2013.

[20] M. Deisenroth, D. Fox, and C. Rasmussen, "Gaussian Processes for Data-Efficient Learning in Robotics and Control," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PP, no. 99, pp. 1–1, 2013.