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 as 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.

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

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 the performance measure as in [6]. The inverse of the variance is also used as a weighting source in two common robotic problems. Simulation results of a point mass robot and a two-link manipulator validate the applicability of the proposed approach.

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

\[\begin{align*}
\dot{x}^r(t) &= f^r(x^r(t), u^r(t)) + G^r dB^r(t) ,
\end{align*}\]

where \(x^r \in \mathbb{R}^n\) and \(u^r \in \mathbb{R}^m\) 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 \((\Omega_r, \mathcal{F}_r, P_r)\). \(G^r \in \mathbb{R}^{n \times n}\) is its infinitesimal variance and \(x^r_0\) is the initial state. Additionally, let \(x^g \in \mathbb{R}^n\) and \(x^o \in \mathbb{R}^n\) be a desired trajectory to follow and the state of an obstacle respectively

\[\begin{align*}
\dot{x}^g(t) &= f^g(x^g(t)) + G^g dB^g(t) ,
\dot{x}^o(t) &= f^o(x^o(t)) + G^o dB^o(t) ,
\end{align*}\]

where \(B^g\) and \(B^o\) are \(n\)-dimensional standard Brownian motion noises defined in complete probability spaces \((\Omega_g, \mathcal{F}_g, P_g)\) and \((\Omega_o, \mathcal{F}_o, P_o)\) respectively, \(G_g \in \mathbb{R}^{n \times n}\) and \(G_o \in \mathbb{R}^{n \times n}\) their infinitesimal variances and \(x^g_0\) and \(x^o_0\) their respective initial states.

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

\[J(u^*(\cdot)) = h_{T_c}(x^r(T_c), x^g(T_c), x^o(T_c)) + \int_{t=0}^{T_c} h(t, x^r(t), x^g(t), x^o(t), u^r(t)) dt ,\]

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 \(x^g\) and the necessary control efforts \(u^r\) while favoring configurations distant to obstacles \(x^o\).

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 \(\xi^*\) be the joint state such that

\[\xi^* = [(x^g - x^r) (x^o - x^r)]^T ,\]

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

\[\begin{align*}
\dot{\xi}^*(t) &= f(\xi^*(t), u^*(t)) + \sum_{s=1}^{S} G^s dB^s(t) ,
\end{align*}\]

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 \((\Omega_s, \mathcal{F}_s, P_s)\) for \(s = 1 \cdots S\), and \(G^s\) and \(f\) are defined such that (1), (2) and (4) hold.

The optimal control solution is given by the control law \(u^*(\cdot)\) 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. \(\mathbb{E}_P[J]\) and where the expectation is defined in the product probability space of all uncertainty sources, i.e. the probability space \((\Omega, \mathcal{F}, P)\) given by

\[\Omega = \Omega_1 \times \Omega_2 \times \cdots \times \Omega_S \quad \mathcal{F} = \mathcal{F}_1 \times \mathcal{F}_2 \times \cdots \times \mathcal{F}_S ,\]

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

\[\kappa_p^{(r)}(J) = \sum_{r=2}^{K} \gamma_r \kappa_r^{(r)}(J) ,\]

where \(\kappa_r^{(r)}\) denotes the \(r\)-th cumulant calculated in probability space \((\Omega, \mathcal{F}, P)\). 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

\[\Psi_p(\theta) = \log \mathbb{E}_P[\exp(\theta J)] .\]

Its power series expansion is given by

\[\Psi_p(\theta) = \sum_{r=1}^{\infty} \frac{\theta^r}{r!} \kappa_r^{(r)}(J) ,\]

where

\[\kappa_r^{(r)}(J) = \left. \frac{\partial^r \Psi_p(\theta)}{\partial \theta^r} \right|_{\theta=0} ,\]

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

\[\begin{align*}
\dot{\xi}^s(t) &= f(\xi^s(t), u^s(t)) + G^s dB^s(t) ,
\end{align*}\]

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

\[\min_{u^s(\cdot)} \xi_s = \min_{u^s(\cdot)} \frac{1}{S} \left( \sum_{s=1}^{S} \kappa_r^{(1)}(J) + \sum_{r=2}^{K} \gamma_r \kappa_r^{(r)}(J) \right) ,\]

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

\[\text{II. PROBLEM SETTING AND APPROACH} \]

"
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 $\theta_s^{-1}\Psi_p(\theta_s)$ and we could similarly formulate problem (11) when $K = \infty$ as

$$
\min_{u^*(\cdot)} \mathcal{V} = \min_{u^*(\cdot)} \frac{1}{S} \sum_{s=1}^{S} \theta_s^{-1}\Psi_p(\theta_s)
$$

$$
= \min_{u^*(\cdot)} \frac{1}{S} \left( \sum_{s=1}^{S} \kappa_p(J) + \sum_{r=2}^{\infty} \frac{\theta_s^{-1}}{r!} \kappa_p(r)(J) \right),
$$

(12)

by fixing the the high order statistics weighting factors to the above-mentioned McLaurin coefficients divided by $\theta_s$. 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 $(\xi_0^{\ast},u_{0\ldots T-1}^{\ast})$, a discrete-time linear quadratic (LQ) approximation of state and control deviations, i.e. $\xi = (\delta\xi^* - \xi^*)$ and $u = (\delta u^* - u^*)$ 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

$$
\xi_{k+1} = A_k\xi_k + B_ku_k + e_k ,
$$

(13)

where $A_k \in \mathbb{R}^{n \times n}$, $B_k \in \mathbb{R}^{n \times m}$, $e_k \in \mathbb{R}^n$ is an independent identically distributed Gaussian random variable such that $\mathcal{N}(\xi_k^0,0,\Sigma_k)$. Assume that the system is at sample time $k$. Given a control law $u_{k\ldots T-1}$ in the form (14) and having observed $\xi_k$, the remaining cost is calculated by means of the backwards recursion

$$
\theta^{-1}\Psi_p(\xi_k, u_{k\ldots T-1}) = J(\xi_k, u_k)
$$

$$
+ \theta^{-1}\Psi_p(\theta, \xi_{k+1}, u_{k+1\ldots T-1}) ,
$$

(17)

where the quantity $\theta^{-1}\Psi_p(\theta, \xi_k, u_{k\ldots T-1})$ is denoted the cost-to-go.

The following lemma solves (17) in closed form yielding a quadratic form on $\xi_k$.

**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

$$
\theta^{-1}\Psi_p(\theta, \xi_k, u_{k\ldots T-1}) = \frac{1}{2}\xi_k^TW_k\xi_k + \xi_k^Tw_k + w_k ,
$$

(18)
where
\[ W_k = Q_k + A_k^T \tilde{W}_{k+1} A_k + L_k^T H_k L_k + G_k^T L_k + L_k^T G_k \tag{19} \]
\[ w_k = q_k + A_k^T \tilde{w}_{k+1} + L_k^T H_k l_k + L_k^T g_k + G_k^T l_k \tag{20} \]
\[ w_k = q_k + \tilde{w}_{k+1} + \frac{1}{2} l_k^T H_k l_k + l_k^T g_k \tag{21} \]
with
\[ H_k = R_k + B_k^T \tilde{W}_{k+1} B_k \quad G_k = B_k \tilde{W}_{k+1} A_k \]
\[ g_k = r_k + B_k^T \tilde{w}_{k+1} \]
and final conditions \( W_N = Q_N, w_N = q_N \) and \( \tilde{W}_{N+1} = W_k + \theta W_{k+1}(\Sigma_k^{-1} - \theta W_{k+1})^{-1} W_{k+1} \tag{22} \]
\[ \tilde{w}_{k+1} = w_{k+1} + \theta W_{k+1}(\Sigma_k^{-1} - \theta W_{k+1})^{-1} w_{k+1} \tag{23} \]
\[ \tilde{w}_{k+1} = w_{k+1} + \theta w_{k+1}(\Sigma_k^{-1} - \theta W_{k+1})^{-1} w_{k+1} \tag{24} \]
with \( F_k = [I_q - \theta W_{k+1} \Sigma_k] \) and \( F_T = 1 \).

Proof. The solution to the Bellman equation (17) enlists only some complication in the term \( \theta^{-1} \psi_p(\theta, \xi_{k+1}, u_{k+1}, \ldots, \xi_{T-1}) \). Assuming quadratic form (18) holds yields
\[ \theta^{-1} \psi_p(\theta, \xi_{k+1}, u_{k+1}, \ldots, \xi_{T-1}) = \theta^{-1} \log \mathbb{E}_p[\exp(\theta(\xi_{k+1}^T W_{k+1} \xi_{k+1} + \xi_{k+1}^T w_{k+1} + w_{k+1}))] . \]
Let now \( M_k = A_k \xi_k + B_k u_k \). Considering the system dynamics (16), the expression for the expectation is given by the Gaussian integral
\[ \frac{1}{\sqrt{(2\pi)^n |\Sigma_k|}} \exp \left\{ \frac{1}{2} M_k^T W_{k+1} M_k + M_k^T w_{k+1} + w_{k+1} \right\} \]
\[ = \int \exp \left\{ -\frac{1}{2} \tilde{e}_k^T (\Sigma_k^{-1} - \theta W_{k+1}) \tilde{e}_k + \tilde{e}_k^T (20W_{k+1}M_k + \theta w_{k+1}) \right\} d\tilde{e}_k . \]
If \( (\Sigma_k^{-1} - \theta W_{k+1}) > 0 \), this expression has analytical solution [16] yielding
\[ \sqrt{(\Sigma_k^{-1} - \theta W_{k+1})^{-1}} \]
\[ \cdot \exp \left\{ \frac{1}{2} M_k^T \tilde{W}_{k+1} M_k + \theta M_k^T \tilde{w}_{k+1} + \theta \tilde{w}_{k+1} \right\} \]
where \( \tilde{W}_{k+1}, \tilde{w}_{k+1}, \) and \( \tilde{w}_{k+1} \) are defined as in (22), (23) and (24). Substituting into (17) and constraining \( u_k \) 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 straightforwardly as
\[ \tilde{W}(\xi_k, u_{k}, \ldots, \xi_{T-1}) = \frac{1}{2} \xi_k^T \tilde{W}_{k+1} \xi_k + \xi_k^T w_k + w_k \tag{25} \]
and where \( W_k^e, w_k^e \) and \( w_k^e \) 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 \( u_{0:T-1} \) corresponds to the cost-to-go at \( k = 0 \), i.e. \( \Psi_0(u_0, u_{0:T-1}) \).

The risk-sensitive solution is computed applying Bellman’s optimal equation, i.e. minimizing (25) w.r.t \( L_k \) and \( l_k \) at each step of the backwards recursion yielding
\[ L_k = -(R_k + B_k^T \tilde{W}_{k+1} B_k)^{-1}(B_k^T \tilde{W}_{k+1} A_k) \tag{26} \]
\[ l_k = -(R_k + B_k^T \tilde{W}_{k+1} B_k)^{-1}(B_k^T \tilde{w}_{k+1} + r_k) \]
where
\[ \tilde{W}_k = \frac{1}{S} \sum_{s=1}^{S} \tilde{W}_k^s \quad \tilde{w}_k = \frac{1}{S} \sum_{s=1}^{S} \tilde{w}_k^s \quad \bar{w}_k = \frac{1}{S} \sum_{s=1}^{S} \bar{w}_k^s , \]
and where \( \tilde{W}_k^e, \tilde{w}_k^e \) and \( \bar{w}_k^e \) 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
\[ \psi_p(\theta, \xi_k, u_{k}, \ldots, \xi_{T-1}) = \theta \left( \frac{1}{2} \xi_k^T W_k \xi_k + \xi_k^T w_k + w_k \right) . \]
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 \( \theta \) at point \( \theta = 0 \), i.e. \( f(\theta)^{[r]} = \frac{\partial^r f(\theta)}{\partial \theta^r} |_{\theta=0} \). An \( r = 0 \) index corresponds to the function itself at \( \theta = 0 \). Additionally, \( \mathcal{C}_k^j \) denotes the binomial coefficient \( \binom{j}{k} \).

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
\[ \kappa_p^{(d)}(J_k, \ldots, \xi_{T-1}) = d \left( \frac{1}{2} \xi_k^T W_k^{[d-1]} \xi_k + \xi_k^T w_k^{[d-1]} + w_k^{[d-1]} \right) \]
where for \( d = 1 \)
\[ W_k^{[0]} = Q_k + A_k^T W_{k+1}^0 A_k + L_k^T H_k^0 L_k + G_k^0 T L_k + L_k^T G_k^0 \]
\[ w_k^{[0]} = q_k + A_k^T w_{k+1}^0 + L_k^T H_k^0 l_k + L_k^T g_k^0 + G_k^0 T l_k \]
\[ w_k^{[0]} = q_k + \tilde{w}_{k+1} + \frac{1}{2} l_k^T H_k^0 l_k + l_k^T g_k^0 \]
with 
\[ H_k^{[0]} = R_k + B_k^T \tilde{w}_{k+1} B_k \quad G_k^{[0]} = B_k^T \tilde{w}_{k+1} A_k \]
\[ g_k^{[0]} = r_k + B_k^T \tilde{w}_{k+1} \]
and final conditions \( W_N^{[0]} = Q_T \), \( \tilde{w}_N^{[0]} = q_T \), \( w_N^{[0]} = q_T \) and for \( d > 1 \)
\[ W_k^{[r]} = A_k^T \tilde{w}_{k+1} B_k + L_k^T H_k^{[r]} L_k + G_k^{[r]} T L_k \]
\[ w_k^{[r]} = A_k^T \tilde{w}_{k+1} B_k + L_k^T H_k^{[r]} L_k + L_k^T G_k^{[r]} T I_k \]
\[ w_k^{[r]} = \bar{w}_{k+1} + \frac{1}{2} L_k^T H_k^{[r]} I_k + L_k^T g_k^{[r]} \]
with
\[ H_k^{[r]} = B_k^T \tilde{w}_{k+1} B_k \quad G_k^{[r]} = B_k^T \tilde{w}_{k+1} A_k \]
\[ g_k^{[r]} = B_k^T \tilde{w}_{k+1} \]
and final conditions \( W_N^{[r]} = 0_{n \times n} \), \( \tilde{w}_N^{[r]} = 0 \), \( w_N^{[r]} = 0 \) and
\[ \tilde{W}_{k+1}^{[r]} = W_{k+1}^{[r]} + r \sum_{j=0}^{r-1} C_j^{r-1} W_k^{[j]} I_k^{[r-j]} \]
(27)
\[ \tilde{w}_{k+1}^{[r]} = w_{k+1}^{[r]} + r \sum_{j=0}^{r-1} C_j^{r-1} w_k^{[j]} I_k^{[r-j]} \]
(28)
\[ \tilde{w}_{k+1}^{[r]} = w_{k+1}^{[r]} + r \sum_{j=0}^{r-1} C_j^{r-1} w_k^{[j]} I_k^{[r-j]} \]
(29)
and where \( W_k^{[r]} \) and \( \tilde{w}_{k+1}^{[r]} \) 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 \( \gamma_{s,s} = 1 \).

The solution to the cost cumulant control problem (11) follows from minimizing (30) w.r.t \( L_k \) and \( I_k \) at each step of the backwards recursion yielding
\[ L_k = - (R_k + B_k^T \tilde{w}_{k+1} B_k)^{-1} (B_k^T \tilde{w}_{k+1} A_k) \]
\[ I_k = - (R_k + B_k^T \tilde{w}_{k+1} B_k)^{-1} (B_k^T \tilde{w}_{k+1} + r_k) \]
(31)

The evaluation of the average of marginal cost-cumulants from (11) at time step \( k \) follows immediately as
\[ \Xi(\xi_k) = \frac{1}{2} \xi_k^T \tilde{W}_k \xi_k + \xi_k^T \tilde{w}_k + \hat{w}_k \]
(30)
where
\[ \tilde{W}_k = \frac{1}{S} \sum_{s=1}^{S} \sum_{r=1}^{K} r_{\gamma,s} W_k^{[r-1]} \]
\[ \hat{w}_k = \frac{1}{S} \sum_{s=1}^{S} \sum_{r=1}^{K} r_{\gamma,s} \tilde{w}_k^{[r-1]} \]
and where \( W_k^{[r]} \) and \( \tilde{w}_{k+1}^{[r]} \) and \( \tilde{w}_{k+1}^{[r]} \) 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 \( \gamma_{s,s} = 1 \).

The solution to the cost cumulant control problem (11) follows from minimizing (30) w.r.t \( L_k \) and \( I_k \) at each step of the backwards recursion yielding
\[ L_k = - (R_k + B_k^T \tilde{w}_{k+1} B_k)^{-1} (B_k^T \tilde{w}_{k+1} A_k) \]
\[ I_k = - (R_k + B_k^T \tilde{w}_{k+1} B_k)^{-1} (B_k^T \tilde{w}_{k+1} + r_k) \]
(31)
where
\[ \tilde{W}_k = \frac{1}{S} \sum_{s=1}^{S} \sum_{r=1}^{K} r_{\gamma,s} W_k^{[r-1]} \]
\[ \tilde{w}_k = \frac{1}{S} \sum_{s=1}^{S} \sum_{r=1}^{K} r_{\gamma,s} \tilde{w}_k^{[r-1]} \]
and where \( \tilde{w}_{k+1}^{[r]} \) and \( \tilde{w}_{k+1}^{[r]} \) 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 \( \xi^*_0 \), the algorithm iterates around the nominal control trajectory \( \pi^*_0 \cdots \pi^*_{T-1} \) by calculating the optimal control deviations that improve the expected performance. At the \( r \)-th iteration, the optimal solution is denoted \( \pi^*_0 \cdots \pi^*_{T-1} = \pi^*_0 \cdots \pi^*_{T-1} + \pi^*_{T-1} \cdots \pi^*_0 \gamma_{s,s} \pi^*_0 \cdots \pi^*_{T-1} \).

1) The corresponding state trajectory \( \xi^*_0 \cdots \pi^*_{T-1} \) is computed simultaneously discretizing the simulated dynamics, for instance by Euler integration. i.e. \( \xi^*_{k+1} = \xi^*_k + \Delta \xi (\xi^*_k, \pi^*_{k+1} \cdots \pi^*_{T-1}) \).
2) An LQ approximation of state and control deviations, i.e. \( \xi_k = \delta \xi_k^r - \xi_k^r \) and \( u_k = \delta u_k^r - u_k^r \) around \( (\xi_{k-1}, T, P_{0\rightarrow T-1}) \) is computed as explained in Section II-A.

3) The optimal deviations law \( u_k = \omega_k + L_k \xi_k \) is computed by means of either the risk-sensitive (26) or cost-cumulant solution (31).

4) As solution \( u_k \) 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 \( u_k \) that yields a policy improvement as \( u_k^{(i+1)}(\alpha) = u_k^{(i)} + \alpha \omega_k + L_k^{(i+1)} \xi_k \), where \( \alpha \) 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

\[
J(u(\cdot)) = h_g(x^r(T_c), x^g(T_c)) + \int_{t=0}^{T_c} h_g(x^r(t), x^g(t)) + h_o(x^r(t), x^o(t)) + h_{so}(u(t))dt,
\]

(32)

where \( h_g = \frac{1}{2} \xi^T Q_g \xi \) and \( h_o = \frac{1}{2} u^T R u \) penalize distance to goal and control efforts, \( h_{so} \) is the distance to a static obstacle \( x^o \).

A. 2D Point-Mass robot

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

\[
M \ddot{x}^r + D \dot{x}^r = u^r,
\]

where \( M, D \in \mathbb{R}^2 \) are the mass and the damping. We first consider a task consisting only of following a goal \( x^g \in \mathbb{R}^2 \) 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 \( \xi^r = [(\dot{x}^r - x^g)^T (\ddot{x}^r - \ddot{x}^g)^T]^T \). The optimal policy takes the form \( u^r = u^r + [L_{x^r} \ L_{x^o}] \xi \). In the following results, parameters were fixed to \( M = I_2 \) kg, \( D = I_2 \) Ns/m, \( R = 10^{-2} I_2 \), \( Q_g = diag\{1 \ 0.1 \ 0.1 \} \), \( \Delta = 10^{-2} \) and \( T_c = 0.5s \). Goal dynamics are assumed to have infinitesimal variance \( \gamma_g = \gamma_o = 0 \), and identical mass and damping to the robot’s.

Optimal trajectories and positional gains \( L_{x^r} \), \( L_{x^o} \) 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 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 PhD (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
B. Two-link manipulator

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

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

where $\theta \in \mathbb{R}^2$ are the joint angles, $\mathcal{M}(\theta)$ is the inertia matrix, $\mathcal{C}(\dot{\theta}, \dot{\theta})$ is the vector of centripetal and Coriolis forces, $\mathcal{B}$ is the joint friction matrix and $\tau \in \mathbb{R}^2$ are the joint torques. Following [9] we set the mass of each
link to \( m_1 = 1.4 \text{kg} \) and \( m_2 = 1.1 \text{kg} \), the length of each link to \( l_1 = 0.3 \text{m} \) and \( l_2 = 0.23 \), the moments of inertia to \( I_1 = 0.025 \text{kg}\cdot\text{m}^2 \) and \( I_2 = 0.045 \text{kg}\cdot\text{m}^2 \) and we assume the center of mass of each link is placed at the link’s center.

The joint friction matrix is set to \( B = \begin{bmatrix} 0.05 & 0.025 \\ 0.025 & 0.05 \end{bmatrix} \). We consider a similar setting as in Fig. 3 but adding a static normally distributed obstacle in the scene by enabling the term \( h_{st} \) in (32) with \( Q_o = I_2 \) and \( w = 0.1 \) and evaluating its expectation a priori. 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.

\[ \text{VII. CONCLUSION AND FUTURE WORK} \]

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


