2020 American Control Conference
Denver, CO, USA, July 1-3, 2020

Nonlinear Covariance Control via Differential Dynamic Programming

Zeji Yi, Zhefeng Cao, Evangelos Theodorou, and Yongxin Chen

***Abstract*— We consider covariance control problems for nonlinear stochastic systems. Our objective is to find an optimal control strategy to steer the state from an initial distribution to a terminal one with specified mean and covariance. This problem is considerably more complicated than previous studies on covariance control for linear systems. We leverage a widely used technique – differential dynamic programming – in nonlinear optimal control to achieve our goal. In particular, we adopt the stochastic differential dynamic programming framework to handle the stochastic dynamics. Additionally, to enforce the terminal statistical constraints, we apply a primal-dual type algorithm. Several examples are presented to demonstrate the effectiveness of our framework.**

I. Introduction

The theory of covariance control/assignment was first studied by Hotz and Skelton [1] in the 80's and further developed in [2]–[4]. The original goal was to find an optimal control strategy for a linear time-invariant system to achieve some specified stationary state covariance. Recently, the covariance control theory was extended to a finite horizon control setting [5]–[10] where the goal was to steer the state covariance of a continuous-time linear dynamic system from an initial value to a terminal one. This finite-horizon perspective was further extended to cases with discrete-time dynamics [11], [12], nonlinear dynamics [13], multiple systems [14] etc. The basic idea of covariance control is to relax the hard constraints in classical optimal control with soft probabilistic constraints; the former is usually unrealistically strong due to the stochastic disturbance. This relaxation makes it suitable for a range of applications in the presence of large uncertainties. Indeed, the covariance control theory has been applied to in aerospace [15], [16], robotics [17] and sensing [18] etc.

Most previous works on covariance control focused on linear systems. The goal of this paper is to generalize this framework to nonlinear stochastic dynamics. This will greatly expand the application domain as most real-world systems in robotics, autonomy, etc cannot be described within the linear dynamics regime. It turns out that the methods developed for linear system covariance control are not applicable to nonlinear problems. For one thing, the mean and the covariance in the linear system have independent dynamics and can be controlled separately and independently [5]; this is no longer the case in nonlinear problems.

In this work, we develop an efficient algorithm for nonlinear covariance control based on differential dynamic programming (DDP) [19], or more precisely, stochastic differential dynamic programming (SDDP) [20]. To ensure the terminal constraint on the state mean and covariance, we adopt a Lagrangian multiplier method [21]. A primal-dual method is used to update the primal and dual variables iteratively. More specifically, for the given Lagrangian multiplier, SDDP is executed first to obtain the nominal trajectory and control. The multiplier is then updated following gradient direction, which is computed through propagating the dynamics forward under optimal control.

The nonlinear covariance control problem was recently studied in [13] under different assumptions with a different method. In particular, the cost function used in [13] is quadratic. In addition, [13] assumes additive noise whose intensity is independent of state and control. In contrast, we consider more general cost functions and dynamics. A case of particular interest to us is multiplicative noise which is ubiquitous in robotics [22]. On another topic, even though differential dynamic programming has been developed for a long time, the literature with terminal constraints is relatively scarce [19], [23] and all of them are for deterministic dynamics. How to generalize the methods to stochastic dynamics was not clear. It turns out that for problems with terminal constraints, there is a significant difference between stochastic settings and deterministic settings, as can be seen later in Section III.

The paper is organized as follows. In Section II we provide the background on covariance control and differential dynamic programming. The main result and the algorithm are presented in Section III. We present two examples in Section IV to illustrate the framework. This follows by a short concluding remark in Section V.

II. Background

A. Covariance Control

Covariance steering/control [5] is about controlling the state of the stochastic system from the initial random vector $x(0)$ at $t = 0$ to a terminal one $x(T)$ at $t = T$ via a control input that minimizes a certainty cost; here $x(0)$ has mean μ_0 and covariance $\Sigma_0 = \mathbb{E}[(x(0) - \mu_0)(x(0) - \mu_0)^{\mathrm{T}}]$, similarly, $x(T)$ has mean μ_T and covariance $\Sigma_T = \mathbb{E}[(x(T) - \mu_T)(x(T) - \mu_T)^{\mathrm{T}}]$. The motivation stems from classical linear quadratic control problems [5], [9], in which, the objective is to steer the state of the system to the desired

Z. Yi is with the School of Aerospace Engineering, Georgia Institute of Technology, Atlanta, GA also the School of Aerospace Engineering, Tsinghua University, Bejing, China; yizj16@mails.tsinghua.edu.cn

Z. Cao is with the School of Aerospace Engineering, Georgia Institute of Technology, Atlanta, GA also the School of Aerospace Engineering, Zhejiang University, Hangzhou, China; zcao99@gatech.edu

The first two authors contributed equally to this work.

E. Theodorou and Y. Chen are with the School of Aerospace Engineering, Georgia Institute of Technology, Atlanta, GA; evangelos.theodorou@gatech.edu, yongchen@gatech.edu

Supported by the NSF under Grant ECCS-1901599.

one in a way that strikes a balance between keeping the deviations of the system's state tolerable on average while also using affordable control effort.

Covariance control provides enormous benefits in the integration of modeling, and control problems [1]. In a general control task, except for controller, identification and state estimation also use covariances as a measure of performance as well. Therefore, a theory which steer covariance allows fusing the entire class of problems (modeling and control) of concern in systems via the same measure of performance.

As for the linear system, the mean and the covariance have independent dynamics. Open-loop control is used to control the mean and the covariance is controlled by close-loop state feedback gain. However, in nonlinear system or constrained cases, the mean and covariance are usually coupled [13].

B. Differential Dynamic Programming

Differential dynamic programming (DDP) [19] is an iterative algorithm for nonlinear optimal control problems, which has high execution speed so that is widely adopted. Consider a system with discrete-time dynamics

$$x(i+1) = f(x, u, i), \quad x(0) = \bar{x}_0, \tag{1}$$

and cost function

$$J = \ell_f(x_N) + \sum_{i=0}^{N-1} \ell(x, u, i), \tag{2}$$

where N is the final time step, $x \in \mathbb{R}^n$ is the state of the system, $u \in \mathbb{R}^m$ is the control input, ℓ is the running cost, and ℓ_f is the final cost. Then, we define the value function at time i is the optimal cost-to-go starting at $x(i) = x$

$$V(x, i) \triangleq \min_u \left[\ell(x, u, i) + V(f(x, u, i), i+1)\right], \tag{3}$$

where $V(f(x, u, i), i+1)$ is value function at time $i+1$.

The algorithm begins with a nominal trajectory, which is sequence of states $(\bar{x}_0, \bar{x}_1, \cdots, \bar{x}_N)$ and corresponding controls $(\bar{u}_0, \bar{u}_1, \cdots, \bar{u}_{N-1})$, and then executes a backward pass and a forward pass at each iteration. In the backward pass, the algorithm expands value function to second-order around the nominal trajectory[1]. In the forward pass, the nominal trajectory will be updated using the optimal control law from the backward pass. The process will be repeated until convergence.

1) Backward pass: First let $Q(\delta x, \delta u, i)$ be the change in value function based on the i-th nominal $(\bar{x}_i, \bar{u}_i)$ pair

$$\begin{aligned} Q(\delta x, \delta u, i) &= \ell(x+\delta x, u+\delta u, i) \\ &\quad + V(f(x+\delta x, u+\delta u), i+1), \end{aligned} \tag{4}$$

then approximate the cost-to-go function as a quadratic function, i.e. expand Q to second-order around $(\bar{x}, \bar{u})$

$$\begin{aligned} Q(\delta x, \delta u, i) &= Q(i) + Q_x^{\mathrm{T}}(i)\delta x + Q_u^{\mathrm{T}}(i)\delta u \\ &\quad + \frac{1}{2}\delta x^{\mathrm{T}} Q_{xx}(i)\delta x + \frac{1}{2}\delta u^{\mathrm{T}} Q_{uu}(i)\delta u \\ &\quad + \frac{1}{2}\delta u^{\mathrm{T}} Q_{ux}(i)\delta x + \frac{1}{2}\delta x^{\mathrm{T}} Q_{xu}(i)\delta u, \end{aligned} \tag{5}$$

[1]A modification of DDP with only first-order approximation of the dynamics was developed in [22] under the name iLQR/iLQG.

where

$$\begin{aligned} Q_x(i) &= \ell_x(i) + f_x^{\mathrm{T}} V_x(i+1), \\ Q_u(i) &= \ell_u(i) + f_u^{\mathrm{T}} V_x(i+1), \\ Q_{xx}(i) &= \ell_{xx}(i) + f_x^{\mathrm{T}} V_{xx}(i+1) f_x + \kappa V_x(i+1) \cdot f_{xx}, \\ Q_{uu}(i) &= \ell_{uu}(i) + f_u^{\mathrm{T}} V_{xx}(i+1) f_u + \kappa V_x(i+1) \cdot f_{uu}, \\ Q_{ux}(i) &= \ell_{ux}(i) + f_u^{\mathrm{T}} V_{xx}(i+1) f_x + \kappa V_x(i+1) \cdot f_{ux}. \end{aligned}$$

When $\kappa = 1$, second order expansion of the dynamics is used [19]; When $\kappa = 0$, first order expansion of dynamics is used [22]. The parameter κ is introduced to unify the results derived from the first and second-order expansion of the dynamics.

The optimal control law δu^* of minimizing the quadratic approximation with respect to δu is

$$\begin{aligned} \delta u^* &= \underset{\delta u}{\operatorname{argmin}}\ Q(\delta x, \delta u) \\ &= k + K\delta x, \end{aligned} \tag{6}$$

where

$$k = -Q_{uu}^{-1} Q_u \quad K = -Q_{uu}^{-1} Q_{ux}. \tag{7}$$

Substitute δu^* into the expansion of Q, we get a quadratic model of V

$$\begin{aligned} V_x(i) &= Q_x(i) - Q_{xu}(i) Q_{uu}^{-1}(i) Q_u(i), \\ V_{xx}(i) &= Q_{xx}(i) - Q_{xu}(i) Q_{uu}^{-1}(i) Q_{ux}(i). \end{aligned} \tag{8}$$

The backward pass begins by initializing the value function with the terminal cost also value function $V(N) = \ell_f(x_N)$ and its derivatives, then calculate the value at each time.

2) Forward pass: After backward pass, we update the nominal trajectory using the optimal feedback control law got from the previous backward pass with the given initial condition

$$x(0) = \bar{x}_0, \tag{9a}$$
$$u(i) = u(i) + k(i) + K(i)(x(i) - \bar{x}_i), \tag{9b}$$
$$x(i+1) = f(x, u, i). \tag{9c}$$

Finally, the backward pass and forward pass will be repeated until convergence when k tends to be zero vector.

III. Nonlinear covariance control

We consider a class of nonlinear stochastic optimal control problems [26] with cost

$$J = \mathbb{E}\left[\ell_f(x(T)) + \int_0^T \ell(x, u)\, \mathrm{d}t\right], \tag{10}$$

subject to the stochastic system described by the following stochastic differential equation [27]

$$\mathrm{d}x = f(x, u)\, \mathrm{d}t + F(x, u)\, \mathrm{d}\omega, \tag{11}$$

where $x \in \mathbb{R}^n$ is the state, $u \in \mathbb{R}^m$ is the control, $\omega \in \mathbb{R}^p$ is standard Brownian motion noise. The dynamic functions are $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n$ and $F : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^{n\times p}$. The term $\ell_f(x(T))$ is the terminal cost while $\ell(x(t), u)$ is the instantaneous cost rate. The system starts from the initial condition

$$\begin{aligned} &\mathbb{E}[x(0)] = \mu_0, \\ &\mathbb{E}[(x(0)-\mu_0)(x(0)-\mu_0)^{\mathrm{T}}] = \Sigma_0, \end{aligned} \tag{12}$$

where $\mu_0 \in \mathbb{R}^n$ is the initial mean and $\Sigma_0 \in \mathbb{R}^{n\times n}$ is the initial covariance. We impose statistical terminal constraint

$$\begin{aligned} &\mathbb{E}[x(T)] = \mu_T, \\ &\mathbb{E}[(x(T)-\mu_T)(x(T)-\mu_T)^{\mathrm{T}}] = \Sigma_T, \end{aligned} \tag{13}$$

where μ_T is the desired mean and Σ_T is the desired covariance. The overall value function $V(x,t)$ is defined as the optimal expected cost accumulated over the time horizon starting from the initial state x at t under optimal control.

We first rewrite the constraint to $\mathbb{E}[x(T)] = \mu_T$ and $\mathbb{E}[x(T)x(T)^{\mathrm{T}}] = \mu_T\mu_T^{\mathrm{T}} + \Sigma_T$, then use Lagrangian multiplier to derive an unconstrained problem of minimizing

$$\begin{aligned} L = \mathbb{E}\Big[&\int_0^T \ell(x,u)\,\mathrm{d}t + \lambda^{\mathrm{T}}(x(T)-\mu_T) \\ &+ \mathrm{tr}(\gamma^{\mathrm{T}}(x(T)x(T)^{\mathrm{T}} - \mu_T\mu_T^{\mathrm{T}} - \Sigma_T))\Big], \end{aligned} \tag{14}$$

where $\lambda \in \mathbb{R}^{n\times 1}$ and $\gamma \in \mathbb{R}^{n\times n}$ are the Lagrange multiplier. Now the term $\lambda^{\mathrm{T}}(x(T)-\mu_T) + \mathrm{tr}(\gamma^{\mathrm{T}}(xx^{\mathrm{T}} - \mu_T\mu_T^{\mathrm{T}} - \Sigma_T))$ is the terminal cost.

To solve the adjoint problem concerning u, λ, γ we may take the Lagrange multiplier into consideration in DDP and solve their optimal together to get a second order convergence rate. Another way is to use the duality, first solve u for the minima $g(\lambda,\gamma)$ of the Lagrangian L over primal values and then update the multiplier to find the optimal value for the dual problem with gradient ascend. After that solve the optimal u again, back and forth. Due to the existence of the Brownian motion term in dynamics, the introduction of the second order term with respect to λ, γ will recursively bring higher order terms of the value function. In particular, $Q_{\lambda\lambda}$ contains $V_{xx\lambda\lambda}$, because of the noise term in the expansion of $\delta x(t+\delta t)$. As a result, second order convergence rate is not achievable. Therefore we turn to the second option. Note that the value function must be convex at the ending point, thus γ should be positive definite in each iteration.

A. System Dynamics Linearization and Discretization

Given a nominal trajectory of states and controls we do expansion with respect to x and u up to the second order ($\kappa = 1$) for differential dynamic programming or the first order ($\kappa = 0$) for iLQG [22]. Then discretize the system with $\delta t = t_{i+1} - t_i$ corresponding to a small interval to transit it from continuous to discrete time. All the derivations from here on consider first order terms with respect to time and up to second order terms with respect to state. The expansion expressed as

$$\begin{aligned} \delta x(t+\delta t) = A_t\delta x(t) + B_t\delta u(t) + \Gamma_t\boldsymbol{\xi}(t) \\ + \kappa\mathbf{O}_d(\delta x, \delta u, \boldsymbol{\xi}, \delta t), \end{aligned} \tag{15}$$

where the random variable $\boldsymbol{\xi} \in \mathbb{R}^{p\times 1}$ is zero mean and Gaussian distributed with covariance $\sigma^2 I_{p\times p}$ while the matrices $A_t \in \mathbb{R}^{n\times n}$, $B_t \in \mathbb{R}^{n\times m}$ and $\Gamma_t \in \mathbb{R}^{n\times p}$ are defined as [20]

$$\begin{aligned} A_t &= I_{n\times n} + \nabla_x f(x,u)\delta t, \\ B_t &= \nabla_u f(x,u)\delta t, \\ \Gamma_t &= \begin{bmatrix} \Gamma^{(1)} & \Gamma^{(2)} & \dots & \Gamma^{(m)} \end{bmatrix}, \end{aligned} \tag{16}$$

with $\Gamma^{(i)} \in \mathbb{R}^{n\times 1}$ defined $\Gamma_t = \Delta_t(\delta x, \delta u) + F(x,u)$ where each column vector of Δ_t is defined as $\Delta_t^{(i)}(\delta x, \delta u) = \nabla_u F_c^{(i)}\delta u(t) + \nabla_x F_c^{(i)}\delta x(t)$. Notice that $F_c^{(i)}$ represent the ith columns of F. Also $\mathbf{O}(\delta x, \delta u, d\omega) \in \mathbb{R}^{n\times 1}$ contains all the second order terms in the deviations in states, controls and noise.

As in original DDP, the derivation of SDDP requires the second order expansion of the action value function around a nominal trajectory $\bar{x}$. Substitution of the discretized dynamics in the second order value function expansion results in

$$\begin{aligned} &V(\bar{x}+\delta x, t+\delta t) = \\ &V(\bar{x}, t+\delta t) + V_x(\bar{x}+\delta x, t+\delta t)^{\mathrm{T}} \\ &(A_t\delta x(t) + B_t\delta u(t) + \Gamma_t\boldsymbol{\xi}(t) + \kappa\mathbf{O}_d) \\ &+ (A_t\delta x(t) + B_t\delta u(t) + \Gamma_t\boldsymbol{\xi}(t) + \kappa\mathbf{O}_d)^{\mathrm{T}} \\ &V_{xx}(\bar{x}+\delta x, t+\delta t) \\ &(A_t\delta x(t) + B_t\delta u(t) + \Gamma_t\boldsymbol{\xi}(t) + \kappa\mathbf{O}_d). \end{aligned} \tag{17}$$

After calculating the expectations of the terms with uncertainty and reshape the matrix, we recollect all the first order and second order expansion of the value function V and get the following action value function

$$\begin{aligned} Q_x(i) &= \ell_x + A_t V_x(i+1) + \tilde{\mathcal{S}}, \\ Q_u(i) &= \ell_u + A_t V_x(i+1) + \tilde{\mathcal{U}}, \\ Q_{xx}(i) &= \ell_{xx} + A_t^{\mathrm{T}} V_{xx}(i+1) A_t + \kappa\mathcal{F} + \tilde{\mathcal{F}} + \tilde{\mathbf{M}}, \\ Q_{xu}(i) &= \ell_{xu} + A_t^{\mathrm{T}} V_{xu}(i+1) B_t + \kappa\mathcal{L} + \tilde{\mathcal{L}} + \tilde{\mathbf{N}}, \\ Q_{uu}(i) &= \ell_{uu} + B_t^{\mathrm{T}} V_{uu}(i+1) B_t + \kappa\mathcal{Z} + \tilde{\mathcal{Z}} + \tilde{\mathbf{G}}, \end{aligned} \tag{18}$$

where the matrices $\mathcal{F} \in \mathbb{R}^{n\times n}$, $\mathcal{L} \in \mathbb{R}^{m\times n}$, $\mathcal{Z} \in \mathbb{R}^{m\times m}$, $\tilde{\mathcal{U}} \in \mathbb{R}^{m\times 1}$, $\tilde{\mathcal{S}} \in \mathbb{R}^{n\times 1}$, $\tilde{\mathcal{F}} \in \mathbb{R}^{n\times n}$, $\tilde{\mathcal{L}} \in \mathbb{R}^{n\times m}$, $\tilde{\mathcal{Z}} \in \mathbb{R}^{m\times m}$, $\tilde{\mathbf{M}} \in \mathbb{R}^{n\times n}$, $\tilde{\mathbf{N}} \in \mathbb{R}^{n\times m}$ and $\tilde{\mathbf{G}} \in \mathbb{R}^{m\times m}$ are defined as in the SDDP paper [20]. To find the optimal control policy, we compute the local variations in control $\delta u(i)$ that minimize the Q-function

$$\begin{aligned} \delta u^* &= \operatorname*{argmin}_u Q(\bar{x}+\delta x, \bar{u}+\delta u) \\ &= -Q_{uu}^{-1}(Q_u + Q_{ux}\delta x). \end{aligned} \tag{19}$$

We propagate backward in time iteratively to get the second-order local approximation of the value function. The controller is used to generate a locally optimal trajectory by propagating the dynamics forward in time. Substituting δu in second order expansion of action value function with the optimal control policy, we get the update law of value function at each time.

$$\begin{aligned} V(i) &= V(i) - \frac{1}{2}Q_u(i)Q_{uu}^{-1}(i)Q_u(i), \\ V_x(i) &= Q_x(i) - Q_u(i)Q_{uu}^{-1}(i)Q_{ux}(i), \\ V_{xx}(i) &= Q_{xx}(i) - Q_{x(i)u}Q_{uu}^{-1}(i)Q_{ux}(i). \end{aligned} \tag{20}$$

B. Multiplier Update

We continue with the control value we get from with fixed λ and γ and extract the control policy for each time step. Notice that what we get is not only a open loop control value u_0 but also a closed loop policy $\delta u = k + K\delta x$. When we execute the SDDP method to convergence we will get $Q_u = 0$ at each time step. Therefore $\delta u = Q_{uu}^{-1}Q_{ux}\delta x$. When the algorithm converges, $\delta x = 0$ but the control policy is still valuable and provides the closed loop part which represented as $K = Q_{uu}^{-1}Q_{ux}$ here. The overall control policy can be expressed as

$$u(i) = K(i)(x(i) - \bar{x}_i) + \bar{u}_i, \tag{21}$$

where $\bar{x}$ is the trajectory which we propagate according to dynamics without noise as in SDDP. $\bar{u}$ is what we solved for in SDDP. When convergence is reached for fixed λ and γ, we compute $\mathbb{E}[x(T)]$ and $\mathbb{E}[x(T)x(T)^{\mathrm{T}}]$, and then update the multipliers with subgradient ascend method [24]

$$\begin{aligned} V_\lambda(T) &= \mathbb{E}[x(T)] - \mu_T, \\ V_\gamma(T) &= \mathbb{E}[x(T)x(T)^{\mathrm{T}}] - \mu_T\mu_T^{\mathrm{T}} - \Sigma_T. \end{aligned} \tag{22}$$

There are two ways to compute the above terms. In the first one, we sample the noise, propagate the stochastic dynamics and calculate the expectation in a statistic way. The feedback law is employed with each step, where x in the control law is the actual state we propagate with real noise. The second method is to propagate the mean and covariance half analytically [28]. The first way has better feasibility but will oscillate slightly even at the fixing point. Consider the samples as i.i.d. and the covariance of $\mathbb{E}[x]$ and $\mathbb{E}[xx^{\mathrm{T}}]$ are of $1/\sqrt{n}$ order of magnitude. The required number of samples is proportional to the square of resolution. For higher resolution, increasing the number of samples is necessary, which will also decrease the randomness in gradient descent, therefore decrease the number of steps to convergence. However, there is always a trade-off between computation speed and precision, which we recommend a combination of several hundred samples and a ten percent resolution. As for the second method, the problem is that for some highly nonlinear cases it may be hard to propagate analytically and the approximation of each state to be Gaussian may have an unneglectable offset for some certain cases especially after many steps. The proposed algorithm is summarized in Algorithm 1. The algorithm stops when k in (7) is sufficiently small, the same as standard DDP.

IV. NUMERICAL EXAMPLES

In this section, we demonstrate the algorithm's performance with the specified terminal constraint on two different systems. We focus on the final state mean and covariance to meet the terminal constraint.

Algorithm 1 Covariance Control with SDDP

Given: Initial state mean μ_0 and covariance Σ_0, total time step N, control sequence $\bar{u}$, initial Lagrange multiplier λ_0 and γ_0, and target mean μ_T and covariance Σ_T. Set $\kappa = 1$ for the case of 2nd order dynamics expansion and $\kappa = 0$ for 1st order expansion of dynamics.

Goal: Optimal control sequence u^*, corresponding state trajectory x^* and local state feedback control strategy $K_i, i = 1, \cdots, N-1$.

Get initial trajectory $\bar{x}$ by integrating the dynamics forward with $\bar{x}_0$ and $\bar{u}$;
while *not convergence,* **do**
  Differentiate value function at final time, get $V_x(N)$, $V_{xx}(N)$;
  for $i = N-1, \cdots, 1$ **do**
    Compute the value of Q_x, Q_u, Q_{xx}, Q_{xu}, Q_{ux}, Q_{uu} at time i according to (18);
    Compute the value of V_x, V_{xx} at time i according to (20);
  end
  for $i = 1, \cdots, N-1$ **do**
    Update control sequence $\bar{u} = \bar{u} + \delta u^*$ with control policy $\delta u^* = k + K\delta x$ at time i;
    Update state trajectory $\bar{x}$ at time i without noise;
  end
  if *SDDP not convergence,* **then**
    continue;
  end
  Sample trajectories from initial state mean μ_0 and covariance Σ_0 using forward dynamics with noise;
  Get statistic final state's mean $\mu(N)$ and covariance $\Sigma(N)$;
  Compute the gradients of Lagrange multiplier λ and γ according to (22) and update Lagrange multiplier $\lambda = \lambda + \eta_1 V_\lambda(T)$ and $\gamma = \gamma + \eta_2 V_\gamma(T)$, where $\eta_1 \in [0,1]$ and $\eta_2 \in [0,1]$ are search parameters ;
end

A. One-Dimensional Dynamics

First, we consider one-dimensional stochastic nonlinear system of the form

$$\mathrm{d}x = \cos(x)\,\mathrm{d}t + u\,\mathrm{d}t + x^2\,\mathrm{d}\omega. \tag{23}$$

Our goal is to manipulate the system with a state dependent noise to reach a final state $x(T)$ with zero mean and covariance 0.03. The initial state has mean 0 and covariance 0.25. The running cost is $\ell = \mathbb{E}\left[\int_0^T ru^2 d\tau\right]$. This is augmented by terminal cost $\ell_f = \mathbb{E}\left[\lambda(x(T) - \mu_T) + \gamma(x(T)^2 - \mu_T^2 - \Sigma_T)\right]$ where $r = 10^{-4}$ induced from Lagrangian multiplier. The time step is set to be 10 msec and time horizon is $T = 1$ sec. We sample the noise and get several trajectories to calculate gradients $V_\lambda(T) = \mathbb{E}[x(T)] - \mu_T$, and $V_\gamma(T) = \mathbb{E}[x(T)^2 - \mu_T^2 - \Sigma_T]$ statistically, while the number of samples is related to the distance to the optimal point. It is worth noting that γ should be positive during the updating process to guarantee the convexity of the terminal cost. In the beginning, only a rough direction is needed for those λ and γ who are distant from the optimal point, i.e. V_λ and V_γ are rather large, thus the number of samples can be small, in our case we choose 80. While for those λ and γ which are rather close to the optimum, i.e. V_λ and V_γ are rather small, the gradient ascent direction and length should be relatively precise. So that the next step will get towards the optimum instead of drifting around because of the uncertainty in the calculation of mean and covariance.

From the experiment, the resolution of the final state is about 0.01 for both mean and covariance with 800 samples around the neighborhood. Also the covariance becomes much smaller under the feedback control while starting with $\Sigma_0 = 0.25$ and ending with $\Sigma_T = 0.01$. As a comparison, for the uncontrolled case, the mean and covariance nearly remain the same. Fig. 1 display the nominal state trajectory under feedback control with some sampled trajectories. Fig. 2 illustrates the nominal trajectory's shift with covariance which corresponds to our controlled system well.

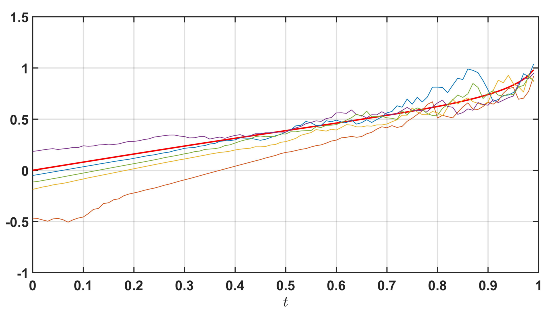

Fig. 1: The nominal trajectory is marked with the red line. The colored lines are several sampled trajectories.

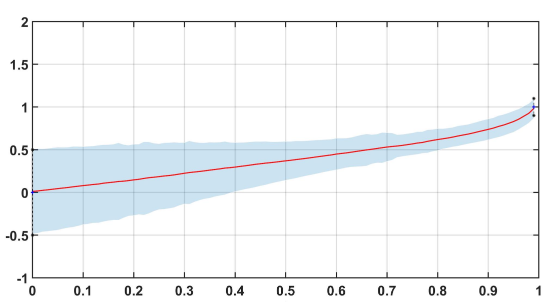

Fig. 2: The whole process's mean is overlapped with the nominal trajectory as in the red line. The shadow part illustrates state covariance.

B. Simple Inverted Pendulum

The second model we study is an inverted pendulum, with constraints on both final state mean and covariance. The stochastic nonlinear dynamics is

$$\mathrm{d}x_1 = x_2\,\mathrm{d}t, \tag{24a}$$

$$\mathrm{d}x_2 = 4sin(x_1)\,\mathrm{d}t + u\,\mathrm{d}t + \alpha u\,\mathrm{d}\omega, \tag{24b}$$

where the state variables are $x_1 = \theta, x_2 = \dot{\theta}$, the measurement parameter of noise $\alpha = 0.04$, the time step is 10 msec and time horizon is $T = 4$ sec. The goal is to let the suspended pendulum swing up from initial condition (corresponding to a -180 deg angle) to an inverted position (corresponding to a 0 deg angle) and stay static at the final time T. The first step is to find the optimal trajectory and the corresponding feedback control strategy with the following cost function when the Lagrange multipliers are fixed

$$J = \mathbb{E}\{\lambda^{\mathrm{T}}(x(T) - \mu_T) + x(T)^{\mathrm{T}}\gamma x(T) - \mu_T^{\mathrm{T}}\gamma\mu_T - \Sigma_T + 0.01\int_0^T u(t)^2\,\mathrm{d}t\}. \tag{25}$$

As in the previous example, we sample the noise and get some trajectories to calculate gradients. Now it is necessary to guarantee that the multiplier matrix γ should be positive definite. In this case, we choose 800 for the number of samples. Fig. 3 demonstrates the variety of the mean and covariance with several sampled trajectories. Fig. 4 displays the control sequences of the sampled trajectories in Fig. 3. Fig. 5 illustrates all nominal trajectory, sampled trajectories, mean and covariance variety in a phase graph.

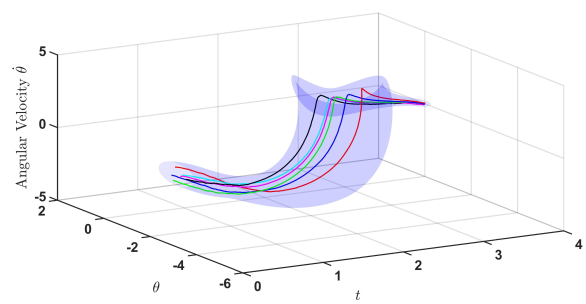

Fig. 3: Colored lines are the sampled trajectories and the blue shadowed part indicates their covariance (3-σ region).

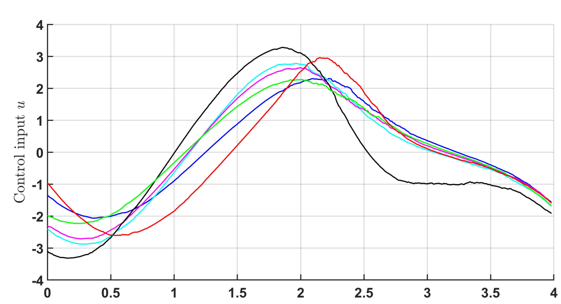

Fig. 4: Control trajectories of the sampled ones are displayed correspondingly.

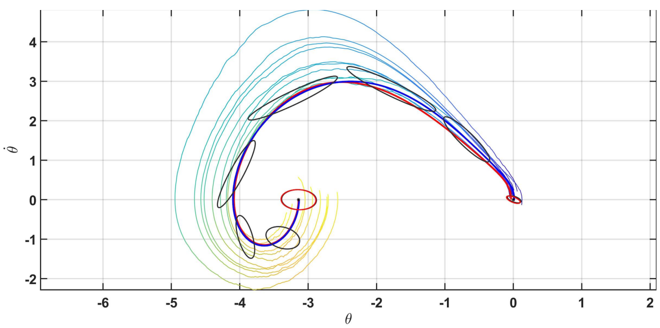

Fig. 5: The Red line indicates the nominal trajectories. The Blue line indicates the mean of samples. Several sampled trajectories are presented with colored lines whose variation in color symbolizes the passage of time. The one-sigma ellipses of some selected time points' are presented as well.

V. Conclusion

In this paper, we developed an algorithm for covariance control of nonlinear stochastic systems. Our method is based on stochastic differential dynamic programming [20]. In order to achieve targeted state mean and covariance, SDDP is combined with the Lagrangian multiplier method to handle the terminal constraints. We remark that DDP with terminal constraints for deterministic systems does not apply in a stochastic setting. We tested our algorithm on severally examples and observed satisfying performance. The next step is to apply this algorithm for high-dimensional robotics path/trajectory planning problems. There are also several potential directions to improve the performance of our algorithm including working on belief space [28] and employing more advanced algorithms for constrained optimizations.

References

[1] A. Hotz and R. E. Skelton, "Covariance control theory," *International Journal of Control*, vol. 46, no. 1, pp. 13–32, 1987.

[2] T. Iwasaki and R. Skelton, "Quadratic optimization for fixed order linear controllers via covariance control," in *1992 American Control Conference*. IEEE, 1992, pp. 2866–2870.

[3] J.-H. Xu and R. Skelton, "An improved covariance assignment theory for discrete systems," *IEEE transactions on Automatic Control*, vol. 37, no. 10, pp. 1588–1591, 1992.

[4] K. M. Grigoriadis and R. E. Skelton, "Minimum-energy covariance controllers," *Automatica*, vol. 33, no. 4, pp. 569–578, 1997.

[5] Y. Chen, T. T. Georgiou, and M. Pavon, "Optimal steering of a linear stochastic system to a final probability distribution, part I," *IEEE Transactions on Automatic Control*, vol. 61, no. 5, pp. 1158–1169, 2015.

[6] Y. Chen, T. T. Georgiou, and M. Pavon, "Optimal steering of a linear stochastic system to a final probability distribution, part II," *IEEE Transactions on Automatic Control*, vol. 61, no. 5, pp. 1170–1180, 2015.

[7] Y. Chen, T. Georgiou, and M. Pavon, "Optimal steering of inertial particles diffusing anisotropically with losses," in *2015 American Control Conference (ACC)*. IEEE, 2015, pp. 1252–1257.

[8] A. Halder and E. D. Wendel, "Finite horizon linear quadratic gaussian density regulator with wasserstein terminal cost," in *2016 American Control Conference (ACC)*. IEEE, 2016, pp. 7249–7254.

[9] E. Bakolas, "Optimal covariance control for stochastic linear systems subject to integral quadratic state constraints," in *2016 American Control Conference (ACC)*. IEEE, 2016, pp. 7231–7236.

[10] Y. Chen, T. T. Georgiou, and M. Pavon, "Optimal steering of a linear stochastic system to a final probability distributionpart III," *IEEE Transactions on Automatic Control*, vol. 63, no. 9, pp. 3112–3118, 2018.

[11] E. Bakolas, "Optimal covariance control for discrete-time stochastic linear systems subject to constraints," in *2016 IEEE 55th Conference on Decision and Control (CDC)*. IEEE, 2016, pp. 1153–1158.

[12] E. Bakolas, "Finite-horizon covariance control for discrete-time stochastic linear systems subject to input constraints," *Automatica*, vol. 91, pp. 61–68, 2018.

[13] J. Ridderhof, K. Okamoto, and P. Tsiotras, "Nonlinear uncertainty control with iterative covariance steering," *arXiv preprint arXiv:1903.10919*, 2019.

[14] Y. Chen, T. T. Georgiou, and M. Pavon, "Covariance steering in zero-sum linear-quadratic two-player differential games," *arXiv preprint arXiv:1909.05468*, 2019.

[15] G. Zhu, K. M. Grigoriadis, and R. E. Skelton, "Covariance control design for hubble space telescope," *Journal of Guidance, Control, and Dynamics*, vol. 18, no. 2, pp. 230–236, 1995.

[16] J. Ridderhof and P. Tsiotras, "Uncertainty quantication and control during mars powered descent and landing using covariance steering," in *2018 AIAA Guidance, Navigation, and Control Conference*, 2018, p. 0611.

[17] K. Okamoto and P. Tsiotras, "Optimal stochastic vehicle path planning using covariance steering," *IEEE Robotics and Automation Letters*, vol. 4, no. 3, pp. 2276–2281, 2019.

[18] M. Kalandros, "Covariance control for multisensor systems," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 38, no. 4, pp. 1138–1157, 2002.

[19] D. H. Jacobson and D. Q. Mayne, "Differential dynamic programming," 1970.

[20] E. Theodorou, Y. Tassa, and E. Todorov, "Stochastic differential dynamic programming," in *Proceedings of the 2010 American Control Conference*. IEEE, 2010, pp. 1125–1132.

[21] D. P. Bertsekas, *Constrained optimization and Lagrange multiplier methods*. Academic press, 2014.

[22] E. Todorov and W. Li, "A generalized iterative lqg method for locally-optimal feedback control of constrained nonlinear stochastic systems," in *Proceedings of the 2005, American Control Conference, 2005.* IEEE, 2005, pp. 300–306.

[23] W. Sun, E. A. Theodorou, and P. Tsiotras, "Continuous-time differential dynamic programming with terminal constraints," in *2014 IEEE Symposium on Adaptive Dynamic Programming and Reinforcement Learning (ADPRL)*. IEEE, 2014, pp. 1–6.

[24] M. X. Goemans and D. P. Williamson, "The primal-dual method for approximation algorithms and its application to network design problems," *Approximation algorithms for NP-hard problems*, pp. 144–191, 1997.

[25] R. D. Monteiro and I. Adler, "Interior path following primal-dual algorithms. part i: Linear programming," *Mathematical programming*, vol. 44, no. 1-3, pp. 27–41, 1989.

[26] K. J. Åström, *Introduction to stochastic control theory*. Courier Corporation, 2012.

[27] R. Van Handel, "Stochastic calculus, filtering, and stochastic control," *Course notes., URL http://www. princeton. edu/rvan/acm217/ACM217. pdf*, 2007.

[28] Y. Pan and E. Theodorou, "Probabilistic differential dynamic programming," in *Advances in Neural Information Processing Systems*, 2014, pp. 1907–1915.