

Adaptive Dual Covariance Steering with Active Parameter Estimation

Jacob W. Knaup

Panagiotis Tsiotras

Abstract—This work examines the optimal covariance steering problem for systems subject to unknown parameters that enter multiplicatively with the state and control, in addition to additive disturbances. In contrast to existing works, the unknown parameters are modeled as random variables and are estimated online. A dual control problem is formulated in which the effect of the planned control policy on the parameter estimates is modeled and optimized for. The parameter estimates are then used to modify the pre-computed control policy online in an adaptive control fashion. Finally, the proposed approach is demonstrated in a vehicle control example with closed-loop parameter identification.

I. INTRODUCTION

Optimal covariance steering deals with the problem of steering a stochastic system from a given initial mean and covariance to a terminal mean and covariance while minimizing a given performance criterion [1]–[4]. Covariance steering has been successfully applied to a variety of practical problems including path planning [5], [6]; differential games [7]; entry, descent, and landing for space systems [8], [9]; and orbital operations [10]–[12]. While originally the problem was formulated for Gaussian distributions—in which case the initial and terminal distributions were fully defined by the initial and terminal means and covariances—recent results have moved towards more general distributions with finite moments [13]–[15]. Moreover, in addition to additive disturbances, recent works have also considered systems subject to multiplicative disturbances [16]–[18] and unknown parameters [19].

While control policies are typically designed for idealized models, real physical systems, such as the vehicles and spacecraft considered in [8]–[12], [17], [19], depend on a variety of physical parameters that can be difficult to determine exactly. Typically, the parameter identification is performed offline and then the nominal identified parameter is taken as the ground truth for the purposes of control design. Although adaptive control extends this paradigm by performing the identification online, these methods, nonetheless, still only use the nominal estimate for the purposes of control design [20]–[22]. This, however, can lead to overconfidence in the control policy and poor performance if the parameter estimate is inaccurate [23]. In [19], we proposed an alternative paradigm in which the estimated parameter *distribution* is used for the purposes of control design. This

method accounts for lingering uncertainty in the true value of the parameter after estimation, and designs a control policy that is robust to this additional source of uncertainty.

An alternative method is to identify the parameters online in closed-loop. This has the advantage of enabling continuous, lifelong learning of the system’s parameters and is also applicable to systems for which generating suitable test signals for parameter identification is difficult (e.g., because the system must obey safety constraints while the parameters are being identified) [23]. This gives rise to the question of dual control, which addresses how to optimally control the system while performing identification to balance the performance of the system’s operation and the accuracy of the resulting parameter estimates [21], [23]. In general, there is a trade-off between these objectives as, typically, introducing increased excitation leads to better parameter estimates but at the expense of the performance and safety of the system [22], [24].

This work builds on our recent results in [19] dealing with covariance steering for systems subject to unknown parameters and additive disturbances. In contrast to [19], which designs a *fixed* optimal control policy for the given *distribution* of parameters, the present work proposes to estimate the parameters online and design an optimal control policy which *adapts* to the parameter *estimates*.

Specifically, we formulate an optimal dual control problem which models the effect of the control on the uncertainty of the parameter estimates, and designs an optimal feedback policy which adapts to the parameter estimates as they converge. The parameter estimation is performed using recursive least squares estimation and the feedback policy is design offline so that the policy may be efficiently implemented on embedded systems with extremely limited computational resources. This work represents the first introduction of online learning and parameter estimation into the covariance steering literature and, in particular, presents the first dual control formulation of the covariance steering problem.

II. PROBLEM FORMULATION

Consider an affine time-and-parameter-varying system given by

$$x_{k+1} = A_k(p)x_k + B_k(p)u_k + D_kw_k + r_k(p), \quad (1a)$$

where, $A_k(p) = \sum_{j=1}^{n_p} A_{j,k}p^j$, $B_k(p) = \sum_{j=1}^{n_p} B_{j,k}p^j$, $r_k(p) = \sum_{j=1}^{n_p} r_{j,k}p^j$, and where $p = [p^1, \dots, p^{n_p}]^\top \in \mathbb{R}^{n_p}$, $x_k \in \mathbb{R}^{n_x}$, and $u_k \in \mathbb{R}^{n_u}$. Let the parameter $p \sim \mathcal{P}$, such that $\mathbb{E}[p] = \bar{p}$ and $\mathbb{E}[(p - \bar{p})(p - \bar{p})^\top] = P \succeq 0_{n_p \times n_p}$, and let $w_k \sim \mathcal{W}$ such that $\mathbb{E}[w_k] = 0_{n_w}$, $\mathbb{E}[w_k w_k^\top] = I_{n_w}$, and $\mathbb{E}[w_{k_1} w_{k_2}^\top] = 0_{n_w \times n_w}$ for all $k_1 \neq k_2$. It is assumed that

J. Knaup is with the School of Interactive Computing, College of Computing and the Institute for Robotics and Intelligent Machines, Georgia Institute of Technology, Atlanta, GA 30332–0250, USA (e-mail: jacobk@gatech.edu)

P. Tsiotras is with the School of Aerospace Engineering and the Institute for Robotics and Intelligent Machines, Georgia Institute of Technology, Atlanta, GA 30332–0150, USA (e-mail: tsiotras@gatech.edu)

$A_{j,k}$, $B_{j,k}$, D_k , and $r_{j,k}$ are known for all $j = 1, \dots, n_p$ and $k = 0, \dots, N-1$. The system dynamics are therefore affine in the parameters, and thus can also be written as

$$x_{k+1} = \Gamma_k(x_k, u_k)p + D_k w_k, \quad (2)$$

where, $\Gamma_k(x, u) = [A_{1,k}x_k + B_{1,k}u_k + r_{1,k}, \dots, A_{n_p,k}x_k + B_{n_p,k}u_k + r_{n_p,k}]$.

Let the initial conditions be given by $x_0 \sim \mathcal{X}_0$, where $\mathbb{E}[x_0] = \mu_I$ and $\mathbb{E}[(x_0 - \mu_I)(x_0 - \mu_I)^\top] = \Sigma_I$, for $\mu_I \in \mathbb{R}^{n_x}$ and $\Sigma_I \succeq \mathbf{0}_{n_x \times n_x}$. We wish to steer (1a) to a given final mean $\mu_F \in \mathbb{R}^{n_x}$ at time N and regulate the terminal covariance with respect to $\Sigma_F \succ \mathbf{0}_{n_x \times n_x}$, such that $\mathbb{E}[x_N] = \mu_F$ and $\mathbb{E}[(x_N - \mathbb{E}[x_N])(x_N - \mathbb{E}[x_N])^\top] \preceq \Sigma_F$, while minimizing the cost function $J(\mu_I, \Sigma_I; u_0, \dots, u_{N-1}) = \mathbb{E}\left[\sum_{k=0}^{N-1} \ell_k(x_k, u_k)\right]$, where, $\ell_k(x_k, u_k) = x_k^\top Q_k x_k + u_k^\top R_k u_k$, and $Q_k \succeq \mathbf{0}_{n_x \times n_x}$, $R_k \succ \mathbf{0}_{n_u \times n_u}$.

Thus, the problem may be summarized by designing an N -step control policy $\pi^N = \{\pi_0, \dots, \pi_{N-1}\}$ that solves the problem

$$\min_{\pi^N} J_N(\pi^N) = \mathbb{E}\left[\sum_{k=0}^{N-1} \ell_k(x_k, u_k)\right], \quad (3a)$$

subject to

$$x_0 \sim \mathcal{X}_0, \quad w_k \sim \mathcal{W}, \quad (3b)$$

$$x_{k+1} = A_k(p_{\text{gt}})x_k + B_k(p_{\text{gt}})u_k + D_k w_k + r_k(p_{\text{gt}}), \quad (3c)$$

$$u_k = \pi_k(x_k), \quad (3d)$$

$$\mathbb{E}[x_N] = \mu_F, \quad \mathbb{E}[(x_N - \mu_F)(x_N - \mu_F)^\top] \preceq \Sigma_F, \quad (3e)$$

for $k = 0, 1, \dots, N-1$, and where the dynamics (3c) evolve according to the particular, unknown realization of $p_{\text{gt}} \sim \mathcal{P}$.

III. COVARIANCE STEERING CONTROLLER DESIGN

A. Existing Methods

While Problem (3) represents the ideal problem we would like to solve using the true parameter value, it is not tractable as the realization of $p_{\text{gt}} \sim \mathcal{P}$ is not known ahead of time, nor can it be observed directly. One alternative is to solve the *certainty equivalence* problem given by

$$\min_{\pi^N} J_N(\pi^N) = \mathbb{E}\left[\sum_{k=0}^{N-1} \ell_k(x_k, u_k)\right], \quad (4a)$$

subject to

$$x_0 \sim \mathcal{X}_0, \quad w_k \sim \mathcal{W}, \quad (4b)$$

$$x_{k+1} = A_k(\bar{p})x_k + B_k(\bar{p})u_k + D_k w_k + r_k(\bar{p}), \quad (4c)$$

$$u_k = \pi_k(x_k), \quad (4d)$$

$$\mathbb{E}[x_N] = \mu_F, \quad \mathbb{E}[(x_N - \mu_F)(x_N - \mu_F)^\top] \preceq \Sigma_F, \quad (4e)$$

for $k = 0, 1, \dots, N-1$, and where the expected value of p , given by $\bar{p}$, is used in place of the true value, p_{gt} . This is the case, for example, when system identification is performed offline and then the nominal estimated parameter values are taken as the true values for the purpose of control design. However, this approach does not consider the uncertainty in the parameters and will fail to meet the terminal constraints in practice if $p_{\text{gt}} \neq \bar{p}$.

Another alternative, similar to the approach used in [19], is to solve the stochastic problem given by

$$\min_{\pi^N} J_N(\pi^N) = \mathbb{E}\left[\sum_{k=0}^{N-1} \ell_k(x_k, u_k)\right], \quad (5a)$$

subject to

$$x_0 \sim \mathcal{X}_0, \quad w_k \sim \mathcal{W}, \quad p \sim \mathcal{P}, \quad (5b)$$

$$x_{k+1} = A_k(p)x_k + B_k(p)u_k + D_k w_k + r_k(p), \quad (5c)$$

$$u_k = \pi_k(x_k), \quad (5d)$$

$$\mathbb{E}[x_N] = \mu_F, \quad \mathbb{E}[(x_N - \mu_F)(x_N - \mu_F)^\top] \preceq \Sigma_F, \quad (5e)$$

for $k = 0, 1, \dots, N-1$, and where the known distribution of p , given by $\mathcal{P}$, is used to propagate the uncertainty in the state dynamics. In contrast to Problem (4), the expectations in Problem (5) are also taken over $p \sim \mathcal{P}$ in order to account for the additional source of uncertainty. This approach is robust to the uncertainty and will meet the terminal constraints in practice. However, this approach is overly conservative, as it must design a single state-feedback control policy which statistically performs well for all possible parameter realizations.

A more optimal policy would *adapt* to the parameter realization to approximate the solution to (3). To this end, introduce the control policy

$$u_k = \rho_k(x_k, p), \quad (6)$$

where $\rho_k(\cdot, \cdot) : \mathbb{R}^{n_x} \times \mathbb{R}^{n_p} \rightarrow \mathbb{R}^{n_u}$ for all $k = 0, 1, \dots, N-1$. Such a control policy is referred to as *adaptive*, which we define below.

Definition 1 (Adaptive Control [23], [25], [26]): An adaptive control system is one that automatically changes the parameters of the control policy in order to improve performance by adjusting its behavior to the properties of the system to be controlled.

However, policy (6) is inadmissible because the control policy depends on the unobservable parameter realization p . Instead, we propose a control policy which adapts to the *estimated* parameter.

B. Proposed Approach

We propose to estimate the unknown parameters using the method of weighted least squares given by

$$\begin{aligned} \min_{p \in \mathbb{R}^{n_p}} & \gamma^k (p - \bar{p})^\top P^{-1} (p - \bar{p}) \\ & + \sum_{t=0}^{k-1} \gamma^{k-t-1} (x_{t+1} - \Gamma_k(x_t, u_t)p)^\top (x_{t+1} - \Gamma_k(x_t, u_t)p), \end{aligned} \quad (7)$$

for $k = 0, 1, \dots, N-1$, and where P is the covariance of $\mathcal{P}$ and $\gamma \in (0, 1]$. If $\gamma < 1$, then γ acts as an exponentially decaying weight and reduces the influence of the prior and older measurements over time, ensuring the estimated parameter fits the most recent data. The case of $\gamma = 1$ is also permitted, and in this case, all observations and the prior are weighted equally. The inclusion of the prior, which is constructed using the first and second moments of the parameter distribution, ensures that Problem (7) has a

unique solution regardless of the available data $x_0, \dots, x_k$, and $u_0, \dots, u_{k-1}$. The optimal parameter estimate which solves (7) at time k is given by $\hat{p}_k \in \mathbb{R}^{n_p}$ and depends on the observations available up to and including time k as well as on the first and second moments of the parameter distribution.

We then employ the adaptive control policy (6), and use the estimated parameter $\hat{p}_k$ in place of the true parameter p . Thus, the control is given by

$$u_k = \rho_k(x_k, \hat{p}_k), \quad (8)$$

which clearly possesses the adaptive property given in Definition 1. Additionally, we wish to design the policy (8) in such a way that it also possesses the *dual* control properties given in Definition 2 below,

Definition 2 (Adaptive Dual Control [21]–[24]): A dual control system is one that operates under model uncertainty and incorporates this uncertainty into the control strategy such that the control signal has the dual properties:

- (i) The system tracks the desired reference value and obeys the constraints in the presence of the model uncertainty.
- (ii) The system is excited in order to improve the estimation (by reducing the higher order moments of the parameter error) so that the quality of the adaptive control policy using the parameter estimate can be improved considerably in future time intervals.

We are now ready to introduce the adaptive dual covariance steering problem, which is given as

$$\min_{\rho^N} J_N(\rho^N) = \mathbb{E} \left[\sum_{k=0}^{N-1} \ell_k(x_k, u_k) \right], \quad (9a)$$

subject to

$$x_0 \sim \mathcal{X}_0, \quad w_k \sim \mathcal{W}, \quad p \sim \mathcal{P}, \quad (9b)$$

$$x_{k+1} = A_k(p)x_k + B_k(p)u_k + D_k w_k + r_k(p), \quad (9c)$$

$$u_k = \rho_k(x_k, \hat{p}_k), \quad (9d)$$

$$\hat{p}_k = \operatorname{argmin}_p \gamma^k (p - \bar{p})^\top P^{-1} (p - \bar{p}) \quad (9e)$$

$$+ \sum_{t=0}^{k-1} \gamma^{k-t-1} (x_{t+1} - \Gamma_k(x_t, u_t)p)^\top (x_{t+1} - \Gamma_k(x_t, u_t)p), \quad (9f)$$

$$\mathbb{E}[x_N] = \mu_F, \quad \mathbb{E}[(x_N - \mu_F)(x_N - \mu_F)^\top] \preceq \Sigma_F, \quad (9f)$$

for $k = 0, 1, \dots, N-1$, and where $\rho^N = \{\rho_0(\cdot, \cdot), \dots, \rho_{N-1}(\cdot, \cdot)\}$, $p \in \mathbb{R}^{n_p}$, and $\gamma \in (0, 1]$.

Most adaptive control policies are based on the separation of parameter estimation and controller design and only identify a model passively as a byproduct. In such cases, the control law is designed using the estimated parameters as an exact representation for the system in a certainty equivalence fashion, without accounting for the uncertainty of estimation [20], [21]. Problem (9), on the other hand, ensures the constraints are satisfied statistically for any parameter realization $p \sim \mathcal{P}$, regardless of the estimate. In contrast to Problem (5), though, Problem (9) also accounts for the online estimation of the parameter and adapts the control policy to the parameter estimate $\hat{p}_k$. Thus, the solution is an adaptive *dual* control policy, as per Definition 2.

C. Tractable Formulation

The solution to Problem (9) depends on optimizing over the space of N arbitrary (infinite-dimensional) control policies which each depend on the estimated parameters at that time step. The estimated parameters, in-turn, depend on the solution of N quadratic programs. We address these computational issues by first introducing an analytical, recursive solution to the quadratic program (7). Then, we introduce a finite-dimensional approximation of $\rho_k(\cdot, \cdot)$.

The solution to Problem (7) requires storing the entire history of the observations, which grows with k , making the approach impractical for large time horizons. To resolve this issue, we utilize recursive least squares, given by the following lemma.

Lemma 1 (Recursive Least Squares [27]): The solution to Problem (7) is given by

$$\hat{p}_{k+1} = \hat{p}_k + P_{k+1} \Gamma_k(x_k, u_k)^\top \hat{e}_{k+1}, \quad (10a)$$

$$P_{k+1} = \frac{1}{\gamma} P_k - \frac{1}{\gamma} P_k \Gamma_k(x_k, u_k)^\top (\gamma I + \Gamma_k(x_k, u_k) P_k \Gamma_k(x_k, u_k)^\top)^{-1} \Gamma_k(x_k, u_k) P_k, \quad (10b)$$

where $\hat{e}_{k+1} = x_{k+1} - \Gamma_k(x_k, u_k) \hat{p}_k$, for $k = 0, 1, \dots, N-1$, where $\hat{p}_0 = \bar{p}$ and $P_0 = P$.

Proof: The result follows immediately from Theorem 1 of [27]. ■

Lemma 1 allows for the parameter to be updated using only the two most recent data points. Thus, the sizes of the matrices and vectors in (10) are fixed and do not depend on k . For more details on recursive least squares, see, for example, [28], [29].

Having rendered the parameter estimation tractable, we turn to the control policy parameterization. Since optimizing over arbitrary feedback policies amounts to an intractable infinite dimensional optimization problem, in practice, prior works, such as [1], [19], [30], introduce to Problems (3), (4), (5) the affine state-feedback policy parameterization given by

$$\pi_k(x_k) = v_k + L_k x_k, \quad (11)$$

where $v_k \in \mathbb{R}^{n_u}$ and $L_k \in \mathbb{R}^{n_u \times n_x}$ for $k = 0, 1, \dots, N-1$. The affine state feedback parameterization given by (11) is reasonable due to the following lemma.

Lemma 2 ([16]): The optimal solution for the exact covariance steering (equality constrained) version of Problem (3) is given by

$$\pi^*(x_k) = \rho_k^*(x_k, p_{\text{gt}}) = \tilde{v}_k^*(p_{\text{gt}}) + \tilde{L}_k^*(p_{\text{gt}}) x_k, \quad (12)$$

where $\tilde{v}_k^*(\cdot) : \mathbb{R}^{n_p} \rightarrow \mathbb{R}^{n_u}$ and $\tilde{L}_k^*(\cdot) : \mathbb{R}^{n_p} \rightarrow \mathbb{R}^{n_u \times n_x}$ for $k = 0, 1, \dots, N-1$, and where $\tilde{v}_k^*(\cdot)$ and $\tilde{L}_k^*(\cdot)$ are given in Theorem 3 of [16]. Furthermore, the result also holds for Problem (4), using $\bar{p}$ in place of p_{gt} .

Proof: The result follows from Theorem 3 of [16], with the following considerations. While [16] presents the optimal policy as $u_k = K_k(x_k - \mu_k) + z_k$, one may observe that the mean sequence is deterministic and thus it may be incorporated into the feed-forward gain, to arrive at

$\tilde{L}_k^*(\cdot) = K_k$ and $\tilde{v}_k^*(\cdot) = z_k - K_k \mu_k$. Additionally, observe the expressions for the optimal values of z_k and K_k given in Theorem 3 of [16] depend on the initial moments as well as the system matrices, which in-turn, depend on the system parameter realization, hence the mappings $\tilde{v}_k^*(p_{\text{gt}})$ and $\tilde{L}_k^*(p_{\text{gt}})$. ■

Thus, as Lemma 2 illustrates, the optimal control policy depends on the system parameter realization which is unknown a-priori at the time the control policy is designed in the cases of Problems (5) and (9). Therefore, the parameterization proposed in (6) and (8) is reasonable. As optimizing over functions is intractable, we replace (8) with a first-order approximation of the optimal policy presented in Lemma 2, given by

$$\rho_k(x_k, \hat{p}_k) = v_k^0 + L_k^0 x_k + \sum_{j=1}^{n_p} v_k^j \hat{p}_k^j + L_k^j \hat{p}_k^j x_k, \quad (13)$$

where $v_k^j \in \mathbb{R}^{n_u}$ and $L_k^j \in \mathbb{R}^{n_u \times n_x}$ for $k = 0, 1, \dots, N-1$ and $j = 0, 1, \dots, n_p$.

Thus, we have an alternative to Problem (9), given by

$$\min_{\mathbf{v}^N, \mathbf{L}^N} J_N(\mathbf{v}^N, \mathbf{L}^N) = \mathbb{E} \left[\sum_{k=0}^{N-1} \ell_k(x_k, u_k) \right], \quad (14a)$$

subject to

$$x_0 \sim \mathcal{X}_0, \quad w_k \sim \mathcal{W}, \quad p \sim \mathcal{P}, \quad \hat{p}_0 = \bar{p}, \quad P_0 = P, \quad (14b)$$

$$x_{k+1} = A_k(p)x_k + B_k(p)u_k + D_k w_k + r_k(p), \quad (14c)$$

$$u_k = v_k^0 + L_k^0 x_k + \sum_{j=1}^{n_p} v_k^j \hat{p}_k^j + L_k^j \hat{p}_k^j x_k, \quad (14d)$$

$$\hat{p}_{k+1} = \hat{p}_k + P_{k+1} \Gamma_k(x_k, u_k)^\top \hat{e}_{k+1}, \quad (14e)$$

$$P_{k+1} = \frac{1}{\gamma} P_k - \frac{1}{\gamma} P_k \Gamma_k(x_k, u_k)^\top$$

$$(\gamma I + \Gamma_k(x_k, u_k) P_k \Gamma_k(x_k, u_k)^\top)^{-1} \Gamma_k(x_k, u_k) P_k, \quad (14f)$$

$$\mathbb{E}[x_N] = \mu_F, \quad \mathbb{E}[(x_N - \mu_F)(x_N - \mu_F)^\top] \preceq \Sigma_F, \quad (14g)$$

for $k = 0, 1, \dots, N-1$, and where $\mathbf{v}^N = \{v_k^j\}_{k=0, j=0}^{N-1, n_p}$ and $\mathbf{L}^N = \{L_k^j\}_{k=0, j=0}^{N-1, n_p}$. Problem (14) replaces the batch least squares estimation defined in (9e) with the recursive least squares estimation (14e)-(14f), and Problem (14) parameterizes the arbitrary feedback policies ρ^N in terms of $\mathbf{v}^N$ and $\mathbf{L}^N$.

D. Sample Average Approximation

As discussed in [19], analytically propagating the state uncertainty through (14c) is challenging due to the dependence between x_k and p for all $k > 0$. In [19], we derived analytical expressions for the propagation of the moments for a specific class of problems subject to constant parametric uncertainty, similar to Problem (5). In our current work, the moment dynamics are further complicated by the online estimation of the parameters and the feedback in the control policy on the parameter estimates. As a result, in this work, we take an alternative approach and utilize a Monte Carlo sampling-based approximation to propagate the uncertainty.

The approximation of Problem (14) is given by

$$\min_{\mathbf{v}^N, \mathbf{L}^N, V} J_N^M(\mathbf{v}^N, \mathbf{L}^N) = \sum_{i=1}^M \sum_{k=0}^{N-1} \ell_k(x_k^i, u_k^i), \quad (15a)$$

subject to

$$x_0^i \sim \mathcal{X}_0, \quad w_k^i \sim \mathcal{W}, \quad p^i \sim \mathcal{P}, \quad \hat{p}_0^i = \bar{p}, \quad P_0^i = P, \quad (15b)$$

$$x_{k+1}^i = A_k(p^i)x_k^i + B_k(p^i)u_k^i + D_k w_k^i + r_k(p^i), \quad (15c)$$

$$u_k^i = v_k^0 + L_k^0 x_k^i + \sum_{j=1}^{n_p} v_k^j \hat{p}_k^{i,j} + L_k^j \hat{p}_k^{i,j} x_k^i, \quad (15d)$$

$$\hat{p}_{k+1}^i = \hat{p}_k^i + P_{k+1}^i \Gamma_k(x_k^i, u_k^i)^\top \hat{e}_{k+1}^i, \quad (15e)$$

$$P_{k+1}^i = P_k^i - P_k^i \Gamma_k(x_k^i, u_k^i)^\top (I + \Gamma_k(x_k^i, u_k^i) P_k^i \Gamma_k(x_k^i, u_k^i)^\top)^{-1} \Gamma_k(x_k^i, u_k^i) P_k^i, \quad (15f)$$

$$\left| \sum_{i=1}^M x_N^i / M - \mu_F \right| \leq \Delta_\mu, \quad (15g)$$

$$|\text{vec}(\sum_{i=1}^M (x_N^i - \mu_F)(x_N^i - \mu_F)^\top / M + VV^\top - \Sigma_F)| \leq \Delta_\Sigma, \quad (15h)$$

for $i = 1, \dots, M$, $k = 0, 1, \dots, N-1$, and where $(\star)$ represents repeated terms, $\hat{e}_{k+1}^i = x_{k+1}^i - \Gamma_k(x_k^i, u_k^i) \hat{p}_k^i$, $\hat{p}_k^i = [\hat{p}_k^{i,1}, \dots, \hat{p}_k^{i,n_p}]^\top$, $\hat{p}_k^{i,j} \in \mathbb{R}$ for $j = 1, \dots, n_p$, $\text{vec}(\cdot)$ is the vector (flattening) operator, $V \in \mathbb{R}^{n_x \times n_x}$, and $\Delta_\mu \in \mathbb{R}^{n_x}$ and $\Delta_\Sigma \in \mathbb{R}^{n_x \times n_x}$ are small quantities. For details on V , Δ_μ , and Δ_Σ , refer to Remark 1. Problem (15) approximates Problem (14) using M Monte Carlo trials. The use of Monte Carlo sampling in this context is referred to as the sample average approximation (SAA), a well-established technique in the field of stochastic optimization [31]. For results on the convergence and feasibility of SAA, see, for example, [31], [32]. Although the SAA introduces additional computational burden to the solver, Problem (15) only needs to be solved once offline, so it does not hinder the online application of the parameter estimation and adaptive policy.

Remark 1: As most nonlinear programming (NLP) solvers do not support semidefinite constraints, we add the slack variable V to convert the semidefinite-constrained NLP (14) to an equality-constrained NLP, which is then relaxed to the inequality-constrained NLP given by (15). Note that $VV^\top \succeq \mathbf{0}_{n_x \times n_x}$ for all $V \in \mathbb{R}^{n_x \times n_x}$, thus the relaxation using the slack variable V is lossless. As recommended in [32], we relax the equality constraints to inequalities allowing for a small margin of error, given by Δ_μ and Δ_Σ , dependent on the sample size.

We are now ready to introduce the following theorem concerning the proposed adaptive dual covariance steering method, which ensures the control policy given by (15) will optimally balance the trade-off between the system's performance and the quality of the closed-loop parameter.

Theorem 1: Problem (15) preserves the dual control effect given in Definition 2. That is, the role of the control actions in affecting the parameter estimate is accounted for in the problem formulation, and the control policy adapts to the estimated parameters so that the following dual control properties are present.

- (i) The control policy minimizes the performance criteria and satisfies the terminal constraints for $p \sim \mathcal{P}$.
- (ii) The control policy affects the parameter uncertainty which is exploited to adapt the control policy using the improved parameter estimate at later time steps.

Proof: Satisfaction of property (i) may be seen from the dynamics (15c), in which, the future empirical state distributions $\{x_k^i\}_{k=0, i=1}^{N, M}$ depend on the parameter realizations $\{p^i\}_{i=1}^M$ sampled from the distribution $\mathcal{P}$. The cost and constraint equations (15a), (15g), (15h), then enforce the objective function and terminal constraints on the empirical state distributions $\{x_k^i\}_{k=0, i=1}^{N, M}$. Note that the parameter estimate only enters into the problem as a source of feedback through the control policy. This is in contrast to methods that enforce the constraints using the estimated parameter and which may not provide the desired performance in practice.

Satisfaction of property (ii) may be seen from the effect of the control on the parameter estimate in (15e)-(15f). Specifically, note that, as per (15b), $p_0^i = \bar{p}$ and $P_0^i = P$, for all $i = 1, \dots, M$, but for $k > 0$, p_k^i and P_k^i depend on the previous controls u_t^i , where $0 \leq t < k$. Therefore, the control policy may improve the parameter estimate and reduce the corresponding uncertainty. The improvement in the parameter estimate at later time steps is then exploited through the control policy (15d), which allows for improved adaptation as $\hat{p}_k^i$ approaches the true realization of p^i . ■

IV. NUMERICAL EXAMPLE

We compare the proposed approach (15) against the non-adaptive method given in [19] in an example of controlling a vehicle with uncertainty in the steering column dynamics. The equations of motion for the nonlinear kinematic bicycle model, shown in Fig. 1, are given by the following

$$\begin{aligned} \dot{e}_\psi &= \dot{\psi} - \dot{\psi}_{\text{ref}} + \dot{y}_y/\nu_x, \quad \dot{e}_y = \nu_y \cos e_\psi + \nu_x \sin e_\psi, \\ \dot{s} &= (\nu_x \cos e_\psi - \nu_y \sin e_\psi)/(1 - e_y \sigma), \end{aligned}$$

where $\nu_y = \nu_x \delta \frac{\ell_r}{\ell_{fr}}$, $\dot{\psi} = \tan \delta \frac{\nu_x}{\ell_{fr}}$, and where e_ψ is the heading error with respect to the reference path, ψ is the vehicle's absolute heading, ψ_{ref} is the heading of the reference path, ν_y is the vehicle's lateral velocity, ν_x is the vehicle's longitudinal velocity, e_y is the vehicle's lateral error with respect to the reference path, s is the vehicle's longitudinal position with respect to the reference path, σ is the curvature of the reference path, δ is the front steering angle, $\ell_{fr} = \ell_f + \ell_r$, and ℓ_f and ℓ_r are the locations of the vehicle's center of mass. Given a constant velocity ν_x and curvature σ , and assuming δ and e_ψ remain small, a linear approximation for the lateral motion is given by

$$\dot{e}_\psi = \frac{\nu_x}{\ell_{fr}} \delta - \nu_x \sigma + \dot{\delta} \frac{\ell_r}{\ell_{fr}}, \quad \dot{e}_y = \frac{\ell_r}{\ell_{fr}} \nu_x \delta + \nu_x e_\psi. \quad (16)$$

We define the state $x = [\delta, e_\psi, e_y]^\top$ and the control $u = \dot{\delta}/p_\delta$, and use Euler integration with a time-step of $\Delta t = 0.2$ sec to obtain the affine system $x_{k+1} = Ax_k + B(p_\delta)u_k +$

$Dw_k + r$, where,

$$A = \begin{bmatrix} 1 & 0 & 0 \\ \frac{\nu_x}{\ell_f + \ell_r} \Delta t & 1 & 0 \\ \frac{\ell_r}{\ell_f + \ell_r} \nu_x \Delta t & \nu_x \Delta t & 1 \end{bmatrix}, \quad B(p_\delta) = \begin{bmatrix} p_\delta \Delta t \\ p_\delta \frac{\ell_r}{\ell_f + \ell_r} \Delta t \\ 0 \end{bmatrix},$$

$$D = \text{diag}(\theta_\delta \Delta t, \theta_\psi \Delta t, \theta_y \Delta t), \quad r = [0, -\sigma \nu_x \Delta t, 0]^\top,$$

where we have added the additive noise term to account for errors owing to the linear approximation.

We compared the proposed approach (15) with the method presented in [19], which solves Problem (5) using policy parameterization (11), as a baseline. We used CasADi [33] and IPOPT [34] to formulate and solve the NLP given by Problem (15). The two methods are evaluated with additive Gaussian disturbances and with parameters drawn from the distributions shown in Fig. 1. Table I presents a comparison of the proposed approach (Adaptive Dual) vs. the baseline method [19] (Static Robust).

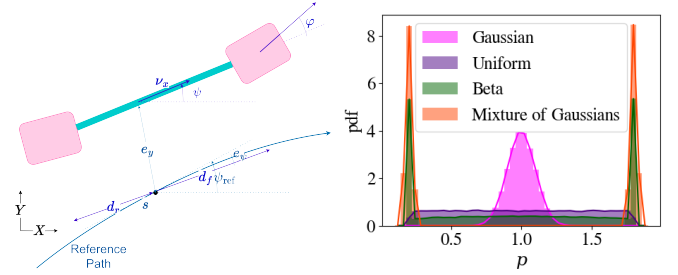


Fig. 1: Kinematic bicycle model in curvilinear coordinates (left) and parameter distributions (right).

TABLE I: Average cost over 1,000 Monte Carlo trials, normalized using the proposed method for the given problem data, for various parameter distributions and terminal constraints, given by $\Sigma_F = \text{diag}(0.01, \Sigma_F^\theta, \Sigma_F^\theta)$.

Control Formulation	Static Robust	Adaptive Dual
Gaussian, $\Sigma_F^\theta = 10^{-4}$	1.03	1.0
Uniform, $\Sigma_F^\theta = 10^{-3}$	1.41	1.0
Uniform, $\Sigma_F^\theta = 10^{-4}$	∞	1.0
Beta, $\Sigma_F^\theta = 10^{-3}$	1.62	1.0
Beta, $\Sigma_F^\theta = 10^{-4}$	∞	1.0
Gaussian Mixture, $\Sigma_F^\theta = 10^{-4}$	1.09	1.0

It may be seen in Table I that the proposed adaptive dual covariance steering method significantly outperforms the baseline static robust covariance steering method for the uniform and beta distributions. Moreover, the proposed approach performs comparably to the baseline method when the parameters are drawn from a Gaussian distribution or mixture of Gaussians, which, as may be seen in Fig. 1, have a lower entropy than the uniform and beta distributions, causing the advantages of an adaptive control policy to be less apparent. Fig. 2 shows sampled realized trajectories when a control policy generated by solving Problem (15) using the proposed method (blue) is compared with the baseline static control policy (red). As seen in Fig. 2, the proposed adaptive dual covariance steering is able to successfully achieve smaller terminal covariance targets than the baseline method by adapting the control policy to the parameter estimate online.

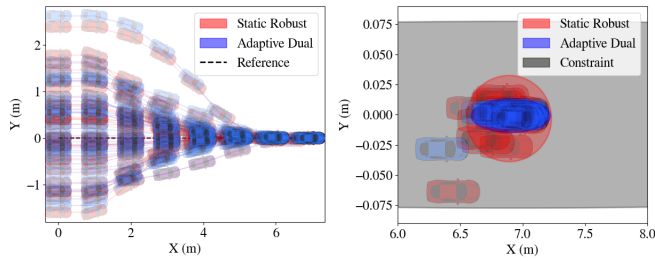


Fig. 2: Sampled vehicle trajectories (left) and terminal constraint visualization (right).

V. CONCLUSION

In this paper, we formulated the first dual control variation of the covariance steering problem for systems subject to unknown parameters. The adaptive dual covariance steering problem includes online parameter identification and a feedback policy parameterization which enables online adaptation to the estimated parameters while still ensuring constraint satisfaction in practice. We showed that the proposed method preserves the dual control property of encoding the control policy's effect on reducing model uncertainty through the parameter estimation, and demonstrated the effectiveness in a vehicle control example. Future work may explore reducing the suboptimality of the affine feedback parameterization by employing a deep neural network as the parameterization.

REFERENCES

- [1] M. Goldshtein and P. Tsiotras, "Finite-horizon covariance control of linear time-varying systems," in *IEEE 56th Annual Conference on Decision and Control (CDC)*, Melbourne, Australia, Dec. 12–15, 2017, pp. 3606–3611.
- [2] 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.
- [3] —, "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.
- [4] —, "Optimal steering of a linear stochastic system to a final probability distribution—part III," *IEEE Transactions on Automatic Control*, vol. 63, no. 9, pp. 3112–3118, 2018.
- [5] 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.
- [6] D. Zheng, J. Ridderhof, P. Tsiotras, and A. Agha-mohammadi, "Belief space planning: A covariance steering approach," in *International Conference on Robotics and Automation*, Philadelphia, PA, May 23–27, 2022, pp. 11 051–11 057.
- [7] Y. Chen, T. T. Georgiou, and M. Pavon, "Covariance steering in zero-sum linear-quadratic two-player differential games," in *IEEE 58th Conference on Decision and Control (CDC)*, Nice, France, Dec. 11–13, 2019, pp. 8204–8209.
- [8] J. Ridderhof and P. Tsiotras, "Uncertainty quantification and control during mars powered descent and landing using covariance steering," in *AIAA Guidance, Navigation, and Control Conference*, Kissimmee, FL, Jan. 8–12, 2018, p. 0611.
- [9] R. Goyal, M. Majji, and R. E. Skelton, "Optimal actuator/sensor precision for covariance steering with soft convex constraints on state and control," in *IEEE American Control Conference*, New Orleans, LA, May 25–28, 2021, pp. 5015–5022.
- [10] J. Ridderhof and P. Tsiotras, "Chance-constrained covariance steering in a Gaussian random field via successive convex programming," *Journal of Guidance, Control, and Dynamics*, vol. 45, no. 4, pp. 599–610, 2022.
- [11] B. Benedikter, A. Zavoli, Z. Wang, S. Pizzurro, and E. Cavallini, "Convex approach to covariance control with application to stochastic low-thrust trajectory optimization," *Journal of Guidance, Control, and Dynamics*, vol. 45, no. 11, pp. 2061–2075, 2022.
- [12] —, "Covariance control for stochastic low-thrust trajectory optimization," in *AIAA SCITECH 2022 Forum*, San Diego, CA, Jan. 3–7, 2022, p. 2474.
- [13] F. Liu and P. Tsiotras, "Optimal covariance steering for continuous-time linear stochastic systems with additive generic noise," *arXiv preprint arXiv:2206.11201*, 2022.
- [14] J. Ridderhof, K. Okamoto, and P. Tsiotras, "Nonlinear uncertainty control with iterative covariance steering," in *IEEE 58th Conference on Decision and Control (CDC)*, Nice, France, Dec. 11–13, 2019, pp. 3484–3490.
- [15] Y. Chen, "Covariance steering for nonlinear control-affine systems," *arXiv preprint arXiv:2108.09530*, 2021.
- [16] F. Liu and P. Tsiotras, "Optimal covariance steering for continuous-time linear stochastic systems with multiplicative noise," *arXiv preprint arXiv:2206.11735*, 2022.
- [17] J. W. Knaup and P. Tsiotras, "Computationally efficient covariance steering for systems subject to parametric disturbances and chance constraints," in *IEEE 62nd Conference on Decision and Control (CDC)*, Singapore, Dec. 13–15, 2023, pp. 1796–1801.
- [18] I. M. Balci and E. Bakolas, "Covariance steering of discrete-time linear systems with mixed multiplicative and additive noise," in *IEEE American Control Conference (ACC)*, San Diego, CA, May 31–June 2, 2023, pp. 2586–2591.
- [19] J. W. Knaup and P. Tsiotras, "Covariance steering for systems subject to unknown parameters," in *IEEE 62nd Conference on Decision and Control (CDC)*, Singapore, Dec. 13–15, 2023, pp. 1790–1795.
- [20] R. Soloperto, J. Köhler, M. A. Müller, and F. Allgöwer, "Dual adaptive mpc for output tracking of linear systems," in *IEEE 58th Conference on Decision and Control (CDC)*, Nice, France, Dec. 11–13, 2019, pp. 1377–1382.
- [21] N. M. Filatov and H. Unbehauen, "Survey of adaptive dual control methods," *IEE Proceedings-Control Theory and Applications*, vol. 147, no. 1, pp. 118–128, 2000.
- [22] A. Mesbah, "Stochastic model predictive control with active uncertainty learning: A survey on dual control," *Annual Reviews in Control*, vol. 45, pp. 107–117, 2018.
- [23] H. Unbehauen, "Adaptive dual control systems: a survey," in *Proceedings of the IEEE Adaptive Systems for Signal Processing, Communications, and Control Symposium*, Lake Louise, AB, Canada, Oct. 4, 2000, pp. 171–180.
- [24] Y. Bar-Shalom and E. Tse, "Dual effect, certainty equivalence, and separation in stochastic control," *IEEE Transactions on Automatic Control*, vol. 19, no. 5, pp. 494–500, 1974.
- [25] D. Seborg, T. F. Edgar, and S. Shah, "Adaptive control strategies for process control: a survey," *AIChE Journal*, vol. 32, no. 6, pp. 881–913, 1986.
- [26] R. Isermann, "Parameter adaptive control algorithms—a tutorial," *Automatica*, vol. 18, no. 5, pp. 513–528, 1982.
- [27] S. A. U. Islam and D. S. Bernstein, "Recursive least squares for real-time implementation [lecture notes]," *IEEE Control Systems Magazine*, vol. 39, no. 3, pp. 82–85, 2019.
- [28] Y. Liu and F. Ding, "Convergence properties of the least squares estimation algorithm for multivariable systems," *Applied Mathematical Modelling*, vol. 37, no. 1–2, pp. 476–483, 2013.
- [29] R. Isermann and M. Münchhof, *Identification of Dynamic Systems: An Introduction with Applications*. Springer, 2011, vol. 85.
- [30] G. Rapakoulias and P. Tsiotras, "Discrete-time optimal covariance steering via semidefinite programming," in *IEEE 62nd Conference on Decision and Control (CDC)*, Singapore, Dec. 13–15, 2023, pp. 1802–1807.
- [31] A. Shapiro, D. Dentcheva, and A. Ruszczyński, *Lectures on Stochastic Programming: Modeling and Theory*. SIAM, 2021.
- [32] T. Lew, R. Bonalli, and M. Pavone, "Sample average approximation for stochastic programming with equality constraints," *arXiv preprint arXiv:2206.09963*, 2022.
- [33] J. A. E. Andersson, J. Gillis, G. Horn, J. B. Rawlings, and M. Diehl, "CasADi – A software framework for nonlinear optimization and optimal control," *Mathematical Programming Computation*, vol. 11, no. 1, pp. 1–36, 2019.
- [34] A. Wächter and L. T. Biegler, "On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming," *Mathematical programming*, vol. 106, pp. 25–57, 2006.