Optimization-based estimation of expected values with applications to optimal stochastic control

Raphael Chinchilla and João P. Hespanha

Abstract—This paper constructs bounds on the expected value of a scalar function of a random vector. The bounds are obtained using an optimization method, which can be computed efficiently using state-of-the-art solvers, and do not require integration or sampling the random vector. This optimization based approach is especially useful in stochastic optimization, where the criteria to be minimized takes the form of an expected value. In particular, we minimize the bounds to solve problems of finite time control with stochastic perturbations and also uncertainty in the system’s parameters. We illustrate this application with two numerical examples.

I. INTRODUCTION

This paper addresses the development of efficient methods to estimate the expected value of a function of a random variable and optimize it with respect to a set of (deterministic) parameters that affect the function. The expected value for continuous random variables is given by an integral, and its exact computation is computationally very intensive in high dimensions. Hence, an array of techniques has been developed in order to tackle this problem.

Numerical integration methods such as the trapezoid method or the adaptive Gauss-Kronrod quadrature method [1] provide a numerical solution, but do not scale well with the dimension. Another approach is to use first order approximation, where the expected value of the cost function is substituted by the cost function of the expected value. It is a simple method and in the case where the cost function is a convex function, it provides a lower bound on the expected value via Jensen’s inequality. However, this method provides no formal guarantees for more general optimization criteria.

Another set of methods is based on sampling. Markov Chain Monte Carlo [2], [3] and Monte Carlo integration can be used to estimate the expected value by computing the empirical mean of samples. In the context of stochastic optimization, stochastic search algorithms, like Recursive Least Square and Stochastic Gradient Descent [4], provide algorithms analogous to their deterministic counterpart. However, such algorithms have slow rates of convergence. The Scenario Approach method [5], [6], substitutes minimizing the expected value by the sum of the cost function in different "scenarios", obtained through sampling the random variables. It requires optimizations with large numbers of samples, which can make it slower. For both of these methods, the sampling process is a computational bottleneck.

While sampling from unconditioned distributions is generally straightforward, sampling from conditional distributions (i.e. where there is knowledge that a certain informative event has occurred) is more complicated. Few conditional distributions have efficient samplers [7]–[11]. In the other cases, one can use generic samplers such as the Gibbs Sampler [4] and Metropolis Hastings algorithm [12], but they are computationally demanding.

As an alternative, this paper constructs bounds on an expected value that are obtained without explicit integration or sampling. The starting point for this paper is a result that provides a lower and an upper bound on the expected value of a scalar function \( V(\cdot) \) of a random vector \( \Psi \) in terms of a minimization and maximization, respectively. The optimization penalizes a criterion that consists of a combination of the function \( V(\Psi) \), the logarithm of the probability density function of \( \Psi \) and its differential entropy. This result permits estimating the value of the expected value through an optimization (and thus optimality conditions that involve differentiation) rather than through an integration. The results in Section II actually provide a family of bounds parameterized by a scalar parameter \( \epsilon \) that sometimes can be optimized to improve tightness of the bound.

Our optimization-based approach to compute upper/lower bounds to expected values is especially useful in stochastic optimization problems, where the criteria to optimize appears in the form of an expected value. In such problem, one can replace the minimization of the expected value by the minimization of our lower or upper bounds, which results in an optimization over a larger space or a min-max problem, respectively. Using TensCalc [13], a high-performance numerical solver that combines symbolic computations with interior point methods, such problem can often be solved very accurately with modest computation.

Our bounds on the expected value are used in Section III to solve the problem of optimal Bayesian estimation [14], [15]. The goal is to find an estimator that minimizes the expected value of a loss function given a set of measurements. For this particular problem, we show that the approach outlined above actually leads to the maximum a posterior estimation for a class of loss functions and probability density functions.

In Section IV the bounds are used to solve two problems in stochastic optimal control. In both problems the goal is to optimize a finite horizon criteria that depends on the trajectory of a dynamical system with stochastic uncertainty in the form of additive perturbations and also parametric uncertainty in the dynamics.

The first problem corresponds to a state-feedback scenario,
for which the initial state is known, whereas the second problem corresponds to an output feedback scenario, for which we have past noisy output measurements based on which the initial state can be inferred.

A criterion similar to the one developed for the output feedback can be found in [16]–[18], where it is used to construct output feedback model predictive controllers. In those references, this criterion was justified on the basis of that it enabled formal stability proofs for the control scheme in a purely deterministic setting and, because of that, did not include terms related to a priori distributions. The results in the present paper provide a formal justification for those criteria in a stochastic control setting.

**Notation:** Unless otherwise specified, regular (non calligraphic) lower case denotes column vectors. The symbol $v'$ indicates the transpose of $v$. If $v$ evolves in discrete time, we write $v_t$ to indicate the value of $v$ at the time instant $t$ and $v_{t_0:t_1} := \{v_{t_0}, v_{t_0+1}, \ldots, v_{t_1}\}$ to indicate the values of $v$ between $t_0$ and $t_1$.

Random variables - scalar or vectors - are denoted by either calligraphic capital letters or capital greek letters. Let $\Psi$ be a random vector taking values in $\mathbb{R}^n$, its probability density function (p.d.f.) is $p_{\Psi} : \mathbb{R}^n \rightarrow \mathbb{R}_{>0}$. Let $\mathcal{Y}$ be another random vector. The conditional p.d.f. of $\Psi$ with respect to $\mathcal{Y}$ is denoted by $p_{\psi|\mathcal{Y}}(\cdot)$. Let $V(\cdot)$ be a scalar valued function. The expected value of $V(\Psi)$ is denoted by $E_{\Psi}[V(\Psi)]$ and the conditional probability of $\Psi$ given $\mathcal{Y}$ is denoted by $E_{\Psi|\mathcal{Y}}[V(\Psi)]$.

**II. Bounds on the expected value**

The following result provides upper and lower bounds on the expected value of a random variable:

**Theorem 1 (Bounds on the expected value):** Let $V : \mathbb{R}^N \rightarrow \mathbb{R}$ be a scalar valued function and $\Psi$ a random vector taking values in $\mathbb{R}^N$ with p.d.f. $p_{\Psi}(\cdot)$. For every scalar $\epsilon$:

$$
\inf_{\Psi} J(\psi, \epsilon) \leq E_{\Psi}[V(\Psi)] \leq \sup_{\Psi} J(\psi, \epsilon) \tag{1}
$$

with

$$
J(\psi, \epsilon) := V(\psi) + \epsilon \log(p_{\Psi}(\psi)) + \epsilon \mathcal{H}_\Psi
$$

and where

$$
\mathcal{H}_\Psi = -\int \log p_{\Psi}(\psi)p_{\Psi}(\psi)d\psi = E_{\Psi}\{-\log p_{\Psi}(\Psi)\}
$$

is the differential entropy of $\Psi$.

**Proof.** Starting with the upper bound, let us define

$$
V_{up}(\epsilon) := \sup_{\Psi} V(\psi) + \epsilon \log(p_{\Psi}(\psi)).
$$

Then, by definition of supremum,

$$
V(\psi) + \epsilon \log(p_{\Psi}(\psi)) \leq V_{up}(\epsilon) \quad \forall \psi \text{ and } \forall \epsilon
$$

$$
\Rightarrow V(\psi) \leq V_{up}(\epsilon) - \epsilon \log(p_{\Psi}(\psi)) \quad \forall \psi \text{ and } \forall \epsilon
$$

$$
\Rightarrow V(\psi)p_{\Psi}(\psi) \leq \left(V_{up}(\epsilon) - \epsilon \log(p_{\Psi}(\psi))\right)p_{\Psi}(\psi) \quad \forall \psi \text{ and } \forall \epsilon
$$

and integrating in both sides yields

$$
E_{\Psi}[V(\Psi)] \leq V_{up}(\epsilon) + \epsilon \mathcal{H}_\Psi \quad \forall \epsilon. \tag{2}
$$

The lower bound can be obtained analogously.

**Remark 1:** [Conditional expected value] Let the function $V(\cdot)$ and the random vector $\Psi$ from Theorem 1 and another random vector $\mathcal{Y}$. Given a particular realization $y$ from $\mathcal{Y}$, an almost identical deduction for Theorem 1 can be done to conditional expected values. In this case, $J(\psi, \epsilon) := V(\psi) + \epsilon \log(p_{\Psi|\mathcal{Y}}(\psi | y)) + \mathcal{H}_{\Psi|\mathcal{Y}}(y)$, where

$$
\mathcal{H}_{\Psi|\mathcal{Y}}(y) := \int -p_{\Psi|\mathcal{Y}}(\psi | y) \log p_{\Psi|\mathcal{Y}}(\psi | y)d\psi
$$

is a pseudo conditional differential entropy.

Since Theorem 1 holds for any $\epsilon \in \mathbb{R}$, the tightest upper bound can be found by minimizing the upper bound with respect to $\epsilon$:

$$
E_{\Psi}[V(\Psi)] \leq \inf_{\epsilon} \sup_{\Psi} V(\psi) + \epsilon \log(p_{\Psi}(\psi)) + \epsilon \mathcal{H}_\Psi \tag{3}
$$

which can be computed using the following algorithm.

**Determining the tightest upper bound:** To introduce the algorithm, we define the function $U(\epsilon) = \sup_{\psi} V(\psi) + \epsilon \log(p_{\Psi}(\psi)) + \epsilon \mathcal{H}_\Psi$. This function is convex as it is the supremum over $\psi$ of a family of linear (thus convex) functions on $\epsilon$ [19]. Let us define $\epsilon^* = \arg \inf \epsilon U(\epsilon)$.

**Assumption 1:** Consider the following assumptions on the criterion $V(\cdot)$ and the random variable $\psi$:

- i) There exists an $\epsilon \in \mathbb{R}$ such that $\sup_{\psi} J(\psi, \epsilon) < \infty$.
- ii) The function $V(\cdot)$ is positive semi-definite.
- iii) $\sup_{\psi} \log(p_{\Psi}(\psi)) < \infty$.
- iv) The function $V(\cdot)$ and the p.d.f. $p_{\Psi}(\cdot)$ are continuously differentiable, the supremum in the definition of $U(\epsilon)$ takes place at a unique maximum, uniquely defining a continuously differentiable function $\psi^*(\epsilon) := \arg \sup_{\psi} V(\psi) + \epsilon \log(p_{\Psi}(\psi)) + \epsilon \mathcal{H}_\Psi$.

Consider the following algorithm:

**Algorithm 1 Bisection algorithm**

**Require:** a tolerance $\mathcal{T} > 0$, $\epsilon_{up} > \epsilon^*$

1. $C \leftarrow \sup_{\psi} \log(p_{\Psi}(\psi)) + \mathcal{H}_\Psi$
2. $\epsilon_{up}(0) \leftarrow \epsilon_{up}$
3. $\epsilon_{low}(0) \leftarrow 0$
4. $g \leftarrow \log(p_{\Psi}(\psi^*(\epsilon_{up}(0)))) + \mathcal{H}_\Psi$
5. $k \leftarrow 0$
6. **while** $\left(\epsilon_{up}(k) - \epsilon_{low}(k)\right) > \mathcal{T}$ **or** $g \neq 0$ **do**
7. $k \leftarrow k + 1$
8. $\bar{\epsilon} \leftarrow \epsilon_{up}(k-1) - \epsilon_{low}(k-1)/2$
9. $g \leftarrow \log(p_{\Psi}(\psi^*(\epsilon_{up}(k)))) + \mathcal{H}_\Psi$
10. **if** $U(\epsilon) = +\infty$ **or** $g < 0$ **then**
11. $\epsilon_{low}(k) \leftarrow \bar{\epsilon}$
12. $\epsilon_{up}(k) \leftarrow \epsilon_{up}(k-1)$
13. **else if** $g > 0$ **then**
14. $\epsilon_{low}(k) \leftarrow \epsilon_{low}(k-1)$
15. $\epsilon_{up}(k) \leftarrow \bar{\epsilon}$
16. **end if**
17. **end while**
18. $K \leftarrow k$

**return** $\epsilon_{up}(K)$ and $K$
Lemma 1 (Convergence of Algorithm 1): Algorithm 1 converges to \( \epsilon^* \) and the error obtained by using \( \epsilon_{up}^{(k)} \) in lieu of \( \epsilon^* \) satisfies \( U(\epsilon_{up}^{(k)}) - U(\epsilon^*) \leq (0.5)^k \epsilon_{up} C < T \).

Two technical Lemmas are needed to prove this result.

Lemma 2 (Linear error on the upper bound): Take \( \epsilon_{up} > \epsilon^* \). Then \( U(\epsilon_{up}) - U(\epsilon^*) \leq (\epsilon_{up} - \epsilon^*) C \).

Proof of Lemma 2. Using the definition of \( U(\epsilon) \) we obtain
\[
U(\epsilon_{up}) = \sup_{\psi} J(\psi, \epsilon^* + \epsilon_{up} - \epsilon^*)
= \sup_{\psi} V(\psi) + (\epsilon^* + \epsilon_{up} - \epsilon^*)(\log(p(\psi)) + \mathcal{H}_\psi)
\leq \sup_{\psi} [V(\psi) + \epsilon^* \log(p(\psi)) + \epsilon^* \mathcal{H}_\psi]
+ (\epsilon_{up} - \epsilon^*) \sup_{\psi} [\log(p(\psi)) + \mathcal{H}_\psi],
\]
from which the results follows with \( C := \sup_{\psi} [\log(p(\psi)) + \mathcal{H}_\psi] \), which is finite in view of item (iii) in Assumption 1 and positive as \( \mathcal{H}_\psi = -E(\log(p(\psi))) \Rightarrow \sup_{\psi} [\log(p(\psi)) + \mathcal{H}_\psi] \geq 0 \).

Lemma 3 (Derivative of \( U(\epsilon) \)): The derivative of \( U(\epsilon) \) is given by:
\[
\frac{dU(\epsilon)}{d\epsilon} = \log(p(\psi^*(\epsilon))) + \mathcal{H}_\psi.
\]

Proof. Taking the derivative, one obtains
\[
\frac{dU(\epsilon)}{d\epsilon} = \frac{dV(\psi^*(\epsilon)) + \epsilon \log(p(\psi^*(\epsilon))) + \epsilon \mathcal{H}_\psi}{d\epsilon}
= \frac{d\psi^*(\epsilon)}{d\epsilon} \left| \frac{dV(\psi) + \epsilon \log(p(\psi))}{d\psi}\right|_{\psi = \psi^*(\epsilon)}
+ \log(p(\psi^*(\epsilon))) + \mathcal{H}_\psi.
\]
As \( \psi^*(\epsilon) \) is determined in an open set, the gradient of \( J(\psi, \epsilon) \) is zero at that point, i.e.
\[
\frac{dJ(\psi, \epsilon)}{d\psi}\bigg|_{\psi = \psi^*(\epsilon)} = \frac{dV(\psi) + \epsilon \log(p(\psi))}{d\psi}\bigg|_{\psi = \psi^*(\epsilon)} = 0.
\]
Therefore \( dU(\epsilon)/d\epsilon = \log(p(\psi^*(\epsilon))) + \mathcal{H}_\psi \).

Proof of Lemma 1. The proof consists of the following steps:

a. \( \epsilon^* = 0 \) is the minimum possible value for \( \epsilon^* \). Proof: By contradiction, suppose \( \epsilon^* < 0 \), then \( \epsilon^* \log(p(\psi^*(\epsilon))) \) is unbounded above, thus \( U(\epsilon^*) = +\infty \) contradicting item (i) in Assumption 1.

b. A corollary of Lemma 2 is that if \( \epsilon < +\infty \) and \( U(\epsilon) = +\infty \) then \( \epsilon < \epsilon^* \).

c. The gradient of \( U(\epsilon) \) is \( \log(p(\psi^*(\epsilon))) + \mathcal{H}_\psi \). Thus, as \( U(\epsilon) \) is convex, \( \log(p(\psi^*(\epsilon))) + \mathcal{H}_\psi > 0 \Rightarrow \epsilon > \epsilon^* \) and \( \log(p(\psi^*(\epsilon))) + \mathcal{H}_\psi < 0 \Rightarrow \epsilon < \epsilon^* \).

d. The initial range contains \( \epsilon^* \) and we are able to classify \( \epsilon \) with respect to \( \epsilon^* \) using the derivative. Thus the bisection algorithm converges.

e. The last step is to prove the decay on the error. Using that
\[
\epsilon_{up}^{(k)} - \epsilon_{low}^{(k)} = 0.5(\epsilon_{up}^{(k-1)} - \epsilon_{low}^{(k-1)}),
\]
by recursion and applying it to Lemma 2 we obtain
\[
U(\epsilon_{up}^{(k)}) - U(\epsilon^*) \leq (\epsilon_{up}^{(k)} - \epsilon^*) C
\leq (\epsilon_{up}^{(k)} - \epsilon_{low}^{(k)}) C
= 0.5^k \epsilon_{up} C,
\]
which finishes the proof of Lemma 1.

Nota Bene: In the case where item (iv) in Assumption 1 does not hold, one can use a Golden-section search [20] which converges at a slower rate, but is applicable to discontinuous functions \( V(\cdot) \), which are useful, for example, in optimization criteria that involved the probability of an event, for which \( V(\cdot) \) would be an indicator function.

III. APPLICATION OF THE BOUNDS TO OPTIMAL ESTIMATION

A common Bayesian formulation [14, 15] for estimating an unknown parameter \( \Theta \) based on a set of measurements \( Y \) consists on finding the estimate \( \phi \) that minimizes the conditional expectation of a semi-positive loss function \( L(\Theta, \phi) \) given the measurements:
\[
\min_{\phi} E_{\Theta | Y}[L(\Theta, \phi)].
\]
Motivated by Theorem 1, one could replace the conditional expectation in the right-hand side of (4) by the upper bound in Theorem 1 and solve instead:
\[
\phi^* := \arg \min_{\phi} \sup_{\theta} \left\{ \min_{\phi} L(\Theta, \phi) + \epsilon \log(p(\Theta | Y)) \right\},
\]
where we omitted the term related to the pseudo differential entropy since it does not depend on the optimization variables \( \theta \) and \( \phi \). It turns out that \( \phi^* \) has a strong relationship with the maximum a posteriori estimation, as we state in the next Lemma.

Lemma 4 (Relationship with maximum a posteriori): Suppose the \( \min \) and the \( \sup \) in the definition of (5) commute. For any semi-positive loss function such that \( L(\Theta, \phi) = 0 \Leftrightarrow \theta = \phi \), \( \phi^* \) corresponds to the maximum a posteriori estimate of \( \theta \).

Proof. If the \( \min \) and \( \sup \) commute, then
\[
\min_{\phi} \sup_{\theta} L(\Theta, \phi) + \epsilon \log(p(\Theta | Y)) = \sup_{\theta} \min_{\phi} L(\Theta, \phi) + \epsilon \log(p(\Theta | Y)).
\]
As \( L(\Theta, \phi) \) is minimized by setting \( \phi = \theta \), the minimizer will necessarily picks \( \phi^*(\theta) = \theta \), and we conclude that
\[
\min_{\phi} L(\Theta, \phi) + \epsilon \log(p(\Theta | Y)) = \sup_{\theta} \epsilon \log(p(\Theta | Y)),
\]
which is the maximum a posteriori estimate of \( \theta \).
IV. APPLICATION OF THE BOUNDS TO OPTIMAL STOCHASTIC CONTROL

This section addresses the use of the upper bound in Theorem 1 to solve stochastic optimal control problems. The key idea is to optimize the upper bound of an expected value, rather than the expected value itself. We will consider that the Assumptions 1 hold and that we take $\epsilon$ sufficiently large such that the upper bound is finite.

A. Trajectory Planning with State Feedback

Consider the dynamical system

$$X_{t+1} = f(X_t, u_t, \Theta) + D_t,$$

where $X_t$ takes values in $\mathbb{R}^{N_x}$ and is the state of the system, $\Theta$ is a random vector representing unknown parameters, $u_t \in \mathbb{R}^{N_u}$ is the control signal and $D_t$ is a zero mean random vector taking values in $\mathbb{R}^{N_d}$ called disturbance. The random parameter $\Theta$ and the random process $D_t$ are all independent and have p.d.f. $p_\Theta(\cdot)$ and $p_D(\cdot)$ and differential entropies $H_\Theta$ and $H_D$, respectively.

For simplicity, we assume full knowledge of the initial state $X_0$, but the results could be easily generalized to the case where we only know its distribution. Given a time horizon $T$ and a cost function $V(\cdot)$ that depends on the unknown parameter $\theta$, the sequence of states $X_{0:T} := \{X_0, X_1, \ldots, X_T\}$ and control inputs $u_{0:T-1} := \{u_0, u_1, \ldots, u_{T-1}\}$, our goal is to solve the finite horizon optimal control problem

$$u_{0:T-1}^* := \arg \min_{u_{0:T-1}} \mathbb{E}_{X_t, \Theta}[V(X_{0:T}, \Theta, u_{0:T-1})]$$

s.t $u_{0:T-1} \in \mathcal{U}$,  

(7)

where $\mathcal{U}$ denotes an admissible set of values for $u_{0:T-1}$.

To introduce the next result, we define the cost function

$$G(x_{0:T}, \theta, u_{0:T-1}, \epsilon) := V(x_{0:T}, \theta, u_{0:T-1}) + \epsilon \log \left( \mathbb{E} \left[ \prod_{t=0}^{T-1} p_{D_t}(x_{t+1} - f(x_t, u_t, \theta)) \right] \right).$$

**Theorem 2 (Control for systems with state feedback):** Let $u_{0:T-1}^\epsilon$ be the solution to the optimization problem

$$u_{0:T-1}^\epsilon := \arg \min_{u_{0:T-1}} \max_{X_{0:T}, \theta} G(x_{0:T}, \theta, u_{0:T-1}, \epsilon)$$

s.t $u_{0:T-1} \in \mathcal{U}$.

The following bounds hold:

$$\min_{u_{0:T-1}} \inf_{x_{0:T}, \theta} G(x_{0:T}, \theta, u_{0:T-1}, \epsilon) + \epsilon H_\Theta + \epsilon \sum_{t=0}^{T-1} H_{D_t} = \mathbb{E}_{X_t, \Theta} \left[ V(X_{0:T}, \Theta, u_{0:T-1}^\epsilon) \right]$$

(9a)

$$\leq \mathbb{E}_{X_t, \Theta} \left[ V(X_{0:T}, \Theta, u_{0:T-1}^*) \right]$$

(9b)

$$\leq \sup_{X_{0:T}, \theta} G(x_{0:T}, \theta, u_{0:T-1}^*, \epsilon) + \epsilon H_\Theta + \epsilon \sum_{t=0}^{T-1} H_{D_t} = \mathbb{E}_{X_t, \Theta} \left[ V(X_{0:T}, \Theta, u_{0:T-1}^*) \right]$$

(9c)

Theorem 2 provides the formal justification to use the solution $u_{0:T-1}^\epsilon$ from (8) in lieu of the actual optimum $u_{0:T-1}^*$. from (7) by guaranteeing a performance. On one hand, the expected value using $u_{0:T-1}^\epsilon$ in (9c) will not exceed the optimal expected value in (9b) by more than the upper bound in (9d). On the other hand, the control $u_{0:T-1}^*$ can never do better than the lower bound in (9a).

**Proof of Theorem 2.** In order to determine the bounds, we apply Theorem 1 such that $X_{1:T}$ and $\Theta$ play the role of $\Psi$ in (1). The proof is based on showing that the p.d.f. of $X_{1:T}$ and $\Theta$ as well as their differential entropy can be determined from the disturbance and the prior.

Using the Markov Chain property of stochastic dynamic systems we obtain

$$p_{X_{1:T}, \Theta}(x_{0:T}, \theta) = p_\theta(\theta) \prod_{t=0}^{T-1} p_{X_{t+1} | x_t, \theta}(x_{t+1} | x_t, \theta).$$

From the perspective of $p_{X_{1:T}, \Theta}$, $x_{1:T}$ and $\theta$ are constant. Using the property of change of variable of probability density functions we obtain

$$p_{X_{t+1} | X_t, \theta}(x_{t+1} | x_t, \theta) = \frac{\det \left( \frac{dx_{t+1} - f(x_t, u_t, \theta)}{dx_{t+1}} \right)}{p_{D_t}(x_{t+1} - f(x_t, u_t, \theta))}.$$

For differential entropy, we use the chain rule to obtain

$$H_{X_{1:T}, \Theta} = H_\Theta + \sum_{t=0}^{T-1} H_{X_{t+1} | x_t, \theta}.$$

Therefore, the p.d.f. of $X_{1:T}$ and $\Theta$ can be determined from the p.d.f. of $D_t$ and of $\Theta$.

Calculating the differential entropy, we obtain

$$H_{X_{t+1} | x_t, \theta} = -\int p_{X_{t+1} | x_t, \theta}(x_{t+1} | x_t, \theta) \log (p_{X_{t+1} | x_t, \theta}(x_{t+1} | x_t, \theta)) dx_{t+1} dx_t d\theta.$$

Therefore, the differential entropy does not depend on the control signal $u_{0:T-1}$. Combining the results, the bounds on Theorem 2 follow from the bounds on Theorem 1.

**Example 1:** Let us illustrate Theorem 2 with the following example, which was selected so that we could compute the solution analytically and compare it with the control obtained by Theorem 2. Consider a linear time invariant system given by

$$X_{t+1} = AX_t + Bu_t + D_t,$$

(10)
where $\mathcal{D}_t \sim \mathcal{N}(0, S)$ and for which there is no unknown parameter $\Theta$. We use as cost function the Linear Quadratic Regulator (LQR) $V(X_{0:T}, u_{0:T-1}) = \sum_{t=0}^{T-1} (x_t'(Q_x + u_t'R_u) + x_t'^T F x_t)$.

The function $\mathcal{G}(x_{0:T}, u_{0:T-1}, \epsilon)$ from Theorem 2 is

$$\mathcal{G}(x_{0:T}, u_{0:T-1}, \epsilon) = \sum_{t=0}^{T-1} (x_t'Q x_t + u_t'R u_t) + x_T'^T F x_T$$

$$- \frac{\epsilon}{2} \sum_{t=0}^{T-1} (x_{t+1} - Ax_t - Bu_t)'S^{-1}(x_{t+1} - Ax_t - Bu_t).$$

For this problem, the analytical expression of the expected value is known and given by

$$E_{X_{0:T}}[V(X_{0:T}, u_{0:T-1})] = \sum_{t=0}^{T-1} (x_t'Q x_t + u_t'R u_t) + x_T'^T F x_T$$

$$+ \sum_{t=0}^{T-1} \log |S_t| + \log |F S_T|,$$

where $\bar{x}_t$ is the state of the nominal system (i.e. with no perturbations) and $S_t$ is the covariance matrix of $X_t$.

We select a horizon $T = 10$,

$$A = \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix}, B = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, Q = \begin{bmatrix} 0.5 & -0.25 \\ -0.25 & 0.5 \end{bmatrix}.$$ 

$R = 0.5, F = Q, S = 2I$, and initial state $x_0 = [1, 1]'$. We can compute the value for the parameter $\epsilon$ that minimizes the upper bound, as in (3), which turns out to be $\epsilon^* \approx 3.4710^4$.

Figure 1 shows the controls $u_{0:T-1}^*$ given by Theorem 2 as well as the true minimum obtained by solving (11), which turn out to be essentially the same.

**B. Trajectory Planning with Output Feedback**

Considered the system in (6) complemented by an observation equation:

$$X_{t+1} = f(X_t, u_t, \Theta) + D_t$$

$$Y_t = h(X_t, \theta) + N_t,$$

where $Y_t$ is the observation vector and $N_t$ is a zero mean random vector called noise, both taking values in $\mathbb{R}^{N_t}$. The random processes $N_t$ is independent of $D_t$ and $\Theta$ and has p.d.f. $p_{N_t}(\cdot)$ and differential entropy $\mathcal{H}_{N_t}$.

We assume given a past window of size $K$ with a set of controls $u_{-K:T-1} := \{u_{-K}, u_{-K+1}, \ldots, u_{-1}\}$ applied before time $t = 0$ and outputs $y_{-K:0} := \{y_{-K}, y_{-K+1}, \ldots, y_0\}$ a realization of $\mathcal{Y}_{-K:0}$ observed before time 0. Given a time horizon $T$ and a function $V(\cdot)$ that depends on the unknown parameter $\Theta$ and on the sequence of states $X_{0:T} := \{X_0, X_1, \ldots, X_T\}$ and control inputs $u_{0:T-1} := \{u_0, u_1, \ldots, u_{T-1}\}$ from time 0 to time $T - 1$, our goal is to solve the finite horizon optimal control problem

$$u_{0:T-1}^* := \arg \min_{u_{0:T-1} \in \mathcal{U}} E_{X_{-K:T}, \Theta|Y_{-K:0}}[V(X_{0:T}, \Theta, u_{0:T-1})]$$

s.t. $u_{0:T-1} \in \mathcal{U},$

where $\mathcal{U}$ denotes an admissible set of values for $u_{0:T-1}$. Differently from the problem in (7), the expected value is now also taken with respect to past values of $X_t$ and is conditioned to the observations.

To introduce the next theorem, we define the cost function

$$G(x_{-K:T}, \theta, u_{0:T-1}, \epsilon) := V(x_{0:T}, \theta, u_{0:T-1}) + \epsilon \log p_{\hat{\theta}(\theta)} p_{x_{-K}(x_{-K})} + \epsilon \log \left( \prod_{t=-K}^{T-1} p_{D_t}(x_{t+1} - f(x_t, u_t, \theta)) \right),$$

where $p_{x_{-K}}(\cdot)$ is the p.d.f. of the prior on the initial state.

**Theorem 3 (Control for systems with output feedback):**

Let $u_{0:T-1}^*$ be the solution to the optimization problem

$$u_{0:T-1}^* := \arg \min_{u_{0:T-1} \in \mathcal{X}_{-K:T}, \theta} G(x_{-K:T}, \theta, u_{0:T-1}, \epsilon)$$

s.t. $u_{0:T-1} \in \mathcal{U}.$

The following bounds hold:

$$\begin{align}
\min_{u_{0:T-1} \in \mathcal{X}_{-K:T}, \theta} G(x_{0:T}, \theta, u_{0:T-1}, \epsilon) + \kappa \epsilon = &\ \inf_{u_{0:T-1} \in \mathcal{X}_{-K:T}, \theta} G(x_{0:T}, \theta, u_{0:T-1}, \epsilon) + \kappa \epsilon \tag{15a} \\
\leq &\ E_{X_{-K:T}, \Theta|Y_{-K:0}}[V(X_{-K:T}, \theta, u_{0:T-1}^*)] \tag{15b} \\
\leq &\ \sup_{x_{-K:T}, \theta} G(x_{0:T}, \theta, u_{0:T-1}^*) + \kappa \epsilon, \tag{15c}
\end{align}$$

with $\kappa := - \log \left( p_{Y_{-K:0}}(y_{-K:0}) \right) \mathcal{H}_{X_{-K:T}, \Theta|Y_{-K:0}}(y_{-K:0})$, where $p_{Y_{-K:0}}(y_{-K:0})$ is the marginal p.d.f. of $Y_{-K:0}$, which we assume different from 0, and $\mathcal{H}_{X_{-K:T}, \Theta|Y_{-K:0}}(y_{-K:0})$ is the pseudo conditional differential entropy as defined in Remark 1.

As it was the case in Theorem 2, Theorem 3 provides the formal justification to use the solution $u_{0:T-1}^*$ from (14) in lieu of the optimum $u_{0:T-1}^*$ from (13).

**Remark 2:** When the min and sup in the definition of $u_{0:T-1}^*$ commute, this optimization provides both a control $u_{0:T-1}^*$ and a corresponding state trajectory $x_{-K:T}$. In this case, one can interpret $x_{-K:T}$ as a state estimate of the past values and $x_{-K:T}^\ast$ as a state estimate, or prediction, of the future values given $u_{0:T-1}^*$. In fact, in the absence of a control objective, this would be a maximum a posteriori estimate, as we saw in Section III.
**Proof of Theorem 3.** Similar to the proof of Theorem 2, the first part of the proof is based on showing that the p.d.f. of $X_{-K:T}$ and $\Theta$ can be determined from the p.d.f. of the disturbance, noise, and priors. Using Bayes Theorem

$$P(X_{-K:T}, \Theta | y_{-K:0}) = \frac{P(X_{-K:T}, \Theta | y_{-K:0})P(X_{-K:T} | \Theta)p(\Theta)}{P(y_{-K:0})}.$$  

Proving the equivalence in the p.d.f. such that

$$p_{X_{-K:T}}(x_{-K:T} | \theta) = p_{X_{-K}}(x_{-K}) \prod_{t=-K}^{T-1} \mathcal{D}_t(x_{t+1} - f(x_t, u_t, \theta))$$

is analogous to what was done in the proof in Theorem 2. For the likelihood, as the observations are conditionally independent, we obtain that

$$p_{Y_{-K:0}}(y_{-K:0} | x_{-K:0}, \theta) = \prod_{t=-K}^{0} p_{Y_t}(y_t | x_t, \theta),$$

from which a change of variable gives the result

$$p_{Y_t}(y_t | x_t, \theta) = p_{X_t}(y_t - h(x_t, \theta)).$$

Therefore, the p.d.f. of $X_{-K:T}$ can be determined from the p.d.f. of $X_{-K}, \Theta, D_t$, and $N_t$.

The second part of the proof is to show that $\mathcal{H}_{X_{-K:T}, \Theta | Y_{-K:0}}(y_{-K:0})$ does not depend on $u_{0:T-1}$. The key element is to separate $\mathcal{H}_{X_{-K:T}, \Theta | Y_{-K:0}}(y_{-K:0})$ into past and future values and show that the future values depend only on the differential entropy of the disturbance $\mathcal{H}_{D_t}$. From its definition, we obtain that

$$\mathcal{H}_{X_{-K:T}, \Theta | Y_{-K:0}}(y_{-K:0}) = - \int p_{X_{-K:T}, \Theta | Y_{-K:0}}(x_{-K:T}, \theta | y_{-K:0}) \log(p_{X_{-K:T}, \Theta | Y_{-K:0}}(x_{-K:T}, \theta | y_{-K:0})) dx_{-K:T} d\theta$$

$$= - \int p_{X_{-K:T}, \Theta | Y_{-K:0}}(x_{-K:T}, \theta | y_{-K:0}) \log(p_{X_{1:T}, \Theta | X_0}(x_{1:T} | \theta | x_0)) dx_{-K:T} d\theta$$

$$= - \int p_{X_{1:T}, \Theta | X_0}(x_{1:T} | \theta | x_0) \log(p_{X_{1:T}, \Theta | X_0}(x_{1:T} | \theta | x_0)) dx_{1:T} d\theta$$

$$= \sum_{t=0}^{T-1} \mathcal{H}_{D_t} + \mathcal{H}_{X_{-K:T}, \Theta | Y_{-K:0}}(y_{-K:0}),$$

where the last line follows from the expression of $\mathcal{H}_{X_{-K:T}, \Theta | Y_{-K:0}}(y_{-K:0})$ deduced in Theorem 2. From that proof, $\sum_{t=0}^{T-1} \mathcal{H}_{D_t}$ does not depend on $u_{0:T-1}$ and, by causality, neither does $\mathcal{H}_{X_{-K:T}, \Theta | Y_{-K:0}}(y_{-K:0})$. Therefore, $\mathcal{H}_{X_{-K:T}, \Theta | Y_{-K:0}}(y_{-K:0})$ does not depend on $u_{0:T-1}$. The bounds on Theorem 3 follow from those on Theorem 1.

**Example 2:** We illustrate Theorem 3 with the following example, which, to the best of our knowledge, cannot be solved in closed form. Consider a linear system

$$X_{t+1} = A X_t + B u_t + D_t$$

$$Y_t = C X_t + N_t,$$

with $D_t \sim \mathcal{N}(0, S)$ and $N_t \sim \mathcal{N}(0, W)$ and prior on the initial state $X_0 \sim \mathcal{N}(\bar{x}, P)$. The system is time-invariant, but the matrices $A$ and $B$ are unknown stochastic parameters with distribution vec$([A, B]) \sim \mathcal{N}(\text{vec}([\bar{A}, \bar{B}]), O)$ where vec is the operator that stacks the columns of a matrix to create a vector. As before, we chose an LQR cost function, thus $G(x_{-K:T}, \theta, u_{0:T-1}, \epsilon)$ is given by

$$G(x_{-K:T}, \theta, u_{0:T-1}, \epsilon) = \sum_{t=0}^{T-1} (x_t' Q x_t + u_t' R u_t) + x_T' F x_T$$

$$- \epsilon (\text{vec}[A, B] - \text{vec}[\bar{A}, \bar{B}])' O^{-1} (\text{vec}[A, B] - \text{vec}[\bar{A}, \bar{B}])$$

$$- \frac{\epsilon}{2} (x_{-K} - \bar{x}_{-K})' P^{-1} (x_{-K} - \bar{x}_{-K})$$

$$- \frac{\epsilon}{2} \sum_{t=-K}^{T-1} (x_{t+1} - A x_t - B u_t)' S^{-1} (x_{t+1} - A x_t - B u_t)$$

$$- \frac{\epsilon}{2} \sum_{t=-K}^{0} (y_t - C x_t)' W^{-1} (y_t - C x_t).$$

We selected a future horizon $T = 10$ and a past horizon $K = 15$. $A$ and $B$ are given by

$$A \sim \begin{bmatrix} 1 + \mathcal{N}(0, 1) & 2 + \mathcal{N}(0, 1) \\ 0 & 1 + \mathcal{N}(0, 1) \end{bmatrix}$$

and initial state $x_{-K} = [1, 1]'$.

We solved the problem on MATLAB® on a computer using Intel® i7 with 8 cores of 2.2GHz each and 8Gb of RAM. TensCalc compiles a code for MATLAB® that can then be used to solve problems for different sets of parameters. Compiling took 4.6 seconds and solving the problem took 0.0062 seconds. Figure 3 shows the values of the controls $u_{0:T-1}$.

![Fig. 3](image-url) Simulation results for the control $u_{0:T-1}$ from Example 2. The control drives $x_{-K:T}$ to 0 as it can be seen in Figure 2a.
Figure 2a shows the trajectory $\mathbf{x}^{\circ}_{\hat{K}:T}$ obtained using Theorem 3 and Figure 2b shows the estimated trajectory using a maximum a posteriori estimator given the control inputs $\mathbf{u}_{0:T-1}$. We verified the maximum a posteriori computed through an optimization by comparing it with a minimum variance estimate that we computed using Markov Chain Monte Carlo; they are indistinguishable. As we can see, $\mathbf{x}^{\circ}_{\hat{K}:T}$ matches very closely the trajectory estimated using the maximum a posteriori. This reinforces the interpretation that $\mathbf{x}^{\circ}_{\hat{K}:T}$ can be seen as an estimation of the states trajectory (see Remark 2).

V. CONCLUSION AND FUTURE WORK

We presented a method to determine lower and upper bounds on the expected value of a scalar function of a random vector. The bounds are computed through an optimization requiring only the probability density function of the underlying random vector and its differential entropy. We showed how the bounds can be applied to the problem of Bayesian estimation and how they relate to maximum a posteriori estimation. Moreover, we showed how the bounds can be used to compute, as well as guarantee the performance, of a control in the context of trajectory planning of finite time dynamical stochastic systems with either state or output feedback. We were able to efficiently compute the control using solvers generated from TensCalc.

Directions for future research include further investigations of the relation between $\mathbf{x}^{\circ}_{\hat{K}:T}$ and the maximum a posteriori estimation as well as the possibility of simultaneously using the lower and upper bounds to compute a control.

REFERENCES