Control with side information

Gireeja Ranade[1] and Anant Sahai[2]
[1]Microsoft Research, Redmond, WA
[2]University of California, Berkeley EECS
giranade@microsoft.com, sahai@eecs.berkeley.edu

***Abstract*— This paper develops an information-theoretic understanding of side information in control systems. We build on the notion of "control capacity" — the rate at which a controller can dissipate uncertainty in a system over an actuation channel — and quantify the change in control capacity due to side information. The results build on the bit-level carry-free models that are widely used to understand information flows in wireless network information theory. These models motivate our definitions and theorems. We show that we are able to compute the value of side information for real i.i.d. actuation channels using a simple conditional expectation. Finally, we give an example of how side information affects the control capacity of an actuation channel with a uniform distribution.**

***Index Terms*— control, capacity, side-information, stabilization**

I. Introduction

New developments in the Internet of Things (IoT) and 5G networks keep pushing for shorter cycles times (on the order of milliseconds) in modern control systems. This means that we might be dealing with control for systems that might be changing at the same timescale at which the control is acting. System models might have unknown parameters due to sampling inaccuracies, linearization, or noisy and unreliable communication.

Clearly, the more information the controller has about the uncertainties in a system, the better off it is in terms of stabilizing it. But by how much? And how do we decide which information is valuable? To answer these questions, we need to think about side information in a principled way. Previously, the control community has studied how networked control systems behave with or without acknowledgements (a kind of side information) of dropped packets since this is relevant for choosing among practical protocols like TCP vs. UDP [1].

This paper takes an information-theoretic perspective to understand side information in control systems with parameter uncertainty. Classically, parameter uncertainty in control systems has been studied through the uncertainty threshold principle result [2]. This provides a threshold based on the means and variances of the system parameters beyond which a scalar system with Gaussian uncertainty cannot be means-square stabilized. The threshold increases as the parameters of the system become less uncertain.

The work in [3] gives an information-theoretic interpretation of this result, and generalizes it to beyond mean-square stability and Gaussian randomness. [3] provides an information-theoretically motivated definition of "control capacity," which we will build on in this paper to understand side information in control more generally. The control capacity of a system (or actuation channel) is the maximum amount of uncertainty that can be dissipated by a controller in a unit time step. Alternatively, we can think of it as the log (base 2) of the maximum intrinsic growth rate that the system can tolerate.

Information theory uses communication capacity as the standard metric to understand bottlenecks in information transmission systems, and side information has been long studied in information theory. We aim to develop a parallel theory for controls.

A. Previous work

The ideas in this paper are deeply connected to many results in both control theory and information theory. Our definitions and results are certainly inspired by all of these. We briefly outline related work in three areas — first, work at the intersection of information theory and control, second perspectives on information from within the control community, and last the characterizations of side information in the information theory community.

1) Control with communication constraints: There is extensive previous work regarding informational limitations in systems [4], [5]. The data-rate theorems [4], [6], [7] are a product of using information-theoretic techniques to understand the impact of unreliable communication channels (connecting the sensor to the controller) on the ability to control a system.

2) Value of information in control: Parameter uncertainty in systems has been studied previously in control, and there has been a long quest to understand a notion of "the value of information" in control. One perspective on this has been provided by the idea of preview control and how it improves performance. A series of works have examined the value of information in stochastic control problems with *additive* disturbances. For a standard LQG problem, Davis [8] defines the value of information as the control cost reduction due to using a clairvoyant controller that has knowledge of the complete past and future of the noise. He observes that future side information about noise realizations can effectively reduce a stochastic control problem to a set of deterministic ones: one for each realization of the noise sequence.

The area of non-anticipative control characterizes the price of not knowing the future realizations of the noise [9], [10].

Rockafellar and Wets [9] first considered this in discrete-time finite-horizon stochastic optimization problems, and then Dempster [11] and Flam [10] extended the result for infinite horizon problems. Finally, Back and Pliska [12] defined the shadow price of information for continuous time decision problems as well. Other related works and references for this also include [13], [14], [15].

An important related result is that of Martins et al. [16]. This paper studied the impact of a preview, or non-causal side information, on the frequency domain sensitivity function of the system using a robust control approach. This result looked at systems with an additive disturbance. In contrast, our paper considers multiplicative uncertainty in the system, and we take an information-theoretic approach.

3) Side information in information theory: There are two sources of uncertainty in point-to-point communication. The first is uncertainty about the system itself, i.e. uncertainty about the channel state, and the second is uncertainty about the message bits in the system. Information theory has looked at both channel-state side information in the channel coding setting and source side information in the source coding setting. Wolfowitz was among the first to investigate channel capacity when the channel state was time-varying [17]. Goldsmith and Varaiya [18] build on this to characterize how channel-state side information at the transmitter can improve channel capacity. Waterfilling (of power) in time is optimal. The time at which side information is received clearly also matters. For instance, in the absence of a probability model for the channel, waterfilling is only possible if all future channel realizations are known in advance[1]. Caire and Shamai [19] provided a unified perspective to understand both causal and imperfect CSI. Lapidoth and Shamai quantify the degradation in performance due to channel-state estimation errors by the receiver [20]. Medard in [21] examines the effect of imperfect channel knowledge on capacity for channels that are decorrelating in time and Goldsmith and Medard [22] further analyze causal side information for block memoryless channels. Their results recover Caire and Shamai [19] as a special case.

The impact of side information has also been studied in multi-terminal settings. For example, Kotagiri and Laneman [23] consider a helper with non-causal knowledge of the state in a multiple access channel. Finally, there is the surprising result by Maddah-Ali and Tse [24]. They showed that in multi-terminal settings stale channel state information at the encoder can be useful even for a memoryless channel. It can enable a retroactive alignment of the signals. Such stale information is completely useless in a point-to-point setting.

Then there are the classic Wyner-Ziv and Slepian-Wolf results for source coding with source side information that are found in standard textbooks [25]. Pradhan et al. [26] showed that even with side information, the duality between source and channel coding continues to hold: this is particularly interesting given the well-known parallel between source coding and portfolio theory. Source coding with fixed-delay side information can be thought of as the dual problem to channel coding with feedback [27].

Another related body of work looks at the uncertainty in the distortion function as opposed to uncertainty of the source[2]. Rate-distortion theory serves as a primary connection between information theory and control. The distortion function quantifies the importance of the various bits in the message, and in a sense confers a meaning upon them. Uncertainty regarding the distortion function is a way to model uncertainty of meaning. Martinian et al. [29] quantified the value of side information regarding the distortion function used to evaluate the decoder [30], [29].

Portfolio theory also gives us an understanding of side information. In the context of horse races, if each race outcome is distributed according a random variable X, then the mutual information between X and Y, $I(X;Y)$, measures the gain in the doubling rate that the side information Y provides the gambler [31]. Directed mutual information captures exactly the causal information that is shared between two random processes. This connection has been made explicit for portfolio theory in [32], [33] by showing that the directed mutual information $I(X^n \to Y^n)$ is the gain in the doubling rate for a gambler due to causal side information Y^n. Of course, directed mutual information is central to control and information theory as the appropriate measure of the capacity of a channel with feedback [34].

B. Organization of the paper

The paper first recaps the notion of an actuation channel and the associated control capacity in Sec. II. Then Sec. III describe the bit-level carry free models that provide the intuition for the results. Sec. IV defines control capacity with side information for real-valued systems and shows that it can be computed as a conditional expectation. It also gives an example calculation of the change in the control capacity of an actuation channel with a uniform distribution. We conclude with some directions for future work.

II. Actuation channels and control capacity

To develop a parallel to the information theoretic notion of a communication channel, we defined an actuation channel for a system in [3], and use this to characterize the informational bottlenecks in a controls system (see Fig. 1). Consider a simple scalar control system, $\mathcal{S}_a$ with perfect state observation, and no additive driving disturbance as in [2].

$$\begin{aligned} X[n+1] &= a(X[n] + B[n]U[n]), \\ Y[n] &= X[n]. \end{aligned} \tag{1}$$

In this system, the actuation channel is captured in the block $(X[n] + B[n]U[n])$ in the dynamics. Here, the $B[n]$'s represent the uncertainty in how the applied control will

[1]With a probability model on the channel, the encoder can simulate the future channel realizations, and hence it has an idea of whether the current realization is good or bad.

[2]A dual notion on the channel coding side is the channel cost function, i.e. the power that it takes per input symbol [28]. What if the encoder does not know this perfectly?

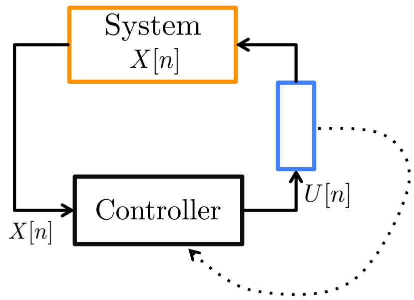

Fig. 1. The actuation channel between the controller and the system was defined in [3]. This paper considers the case where the controller might have side information about the actuation channel uncertainty.

actually influence the state, and are i.i.d. random variables with mean μ_B and variance σ_B^2. $X[0] \sim \mathcal{N}(0,1)$. The aim is to choose $U[n]$, a function of $Y[n]$, so as to stabilize the system. Historically, the stability of this system has been studied in the mean-squared case.

The classic uncertainty threshold principle gives the achievable strategy and converse for this system for mean-squared error and shows that linear strategies are optimal in this case [2]. We know that system (1) is mean-square stabilizable using linear strategies if $a^2 < \left(1 + \frac{\mu_B^2}{\sigma_B^2}\right)$.

The threshold here shows that to tolerate a larger system growth a we need to somehow reduce the randomness in $B[n]$. This can be done using side information. The uncertainty threshold principle gives us one perspective on this matter, and here would like to provide an information-theoretic perspective.

In this paper, in addition to the second moment stability, we would like to consider the notion of logarithmic stability of a system, i.e. the "weakest" notion of stability in a moment sense.

Definition 2.1: The system (1) is said to be logarithmically stabilizable (η-th moment stabilizable) if there exists a causal control strategy $U_0^n(\cdot) \in \sigma(Y_0^n)$ such that $\exists M \in \mathbb{R}$, $M < \infty$, s.t. $\mathbb{E}[\log |X[n]|] < M$ ($\mathbb{E}[|X[n]|^\eta] < M$) for all n.

We recount the definitions of control capacity from [3] for convenience here. Consider the following system $\mathcal{S}$ with no explicit growth, i.e. $a = 1$, and i.i.d. $B[n]$ as above

$$\begin{aligned} X[n+1] &= X[n] + B[n]U[n], \\ Y[n] &= X[n]. \end{aligned} \tag{2}$$

Definition 2.2: The control capacity in the Shannon-sense of the system $\mathcal{S}$ in eq. (2) is defined as

$$C_{sh}(\mathcal{S}) = \lim_{n\to\infty} \min_{\substack{U_0^{n-1}(\cdot) \\ \text{s.t. } U[i](\cdot)\in\sigma(Y_0^n)}} -\frac{1}{n}\mathbb{E}\left[\log\left|\frac{X[n]}{X[0]}\right|\right]. \tag{3}$$

Definition 2.3: The η-th moment control capacity of the system $\mathcal{S}$ in eq. (2) is defined as

$$C_\eta(\mathcal{S}) = \lim_{n\to\infty} \min_{\substack{U_0^{n-1}(\cdot) \\ \text{s.t. } U[i](\cdot)\in\sigma(Y_0^n)}} -\frac{1}{n}\frac{1}{\eta}\log\mathbb{E}\left[\left|\frac{X[n]}{X[0]}\right|^\eta\right]. \tag{4}$$

We show in [3] that the Shannon control capacity of the system in eq. (2) actually characterizes the maximum growth through a that the original system can tolerate. With some technical conditions, the theorem effectively states that the system $\mathcal{S}_a$ is stabilizable if and only if the control capacity of the actuation channel $\mathcal{S}$, $C_{sh}(\mathcal{S}) \geq \log|a|$. Similar results are also shown for the η-th moment capacity.

What if we now had some side information $T[n]$ at time n about the random variable $B[n]$? How does affect the control capacity of the system? The current paper answers this question.

III. Carry-free systems

"Carry-free" models introduced in [35], [36] provide an information-theoretic intuition for the uncertainty threshold principle. The bit-level representation of how the information is flowing in the system proves to be useful here. They help model the uncertainty introduced by the actuations channels in a simplified way. The name "carry-free" is derived from the fact that the addition operation in this model does not involve carryovers. Multiplication between two bit-strings is convolution: if $y_m y_{m-1}\cdots y_0 y_{-1}\cdots = (c_m c_{m-1}\cdots)\cdot(x_m x_{m-1}\cdots)$, then $y_i = \sum_j c_j x_{i-j}$.

Let the state as it evolves in time be represented by the bits of the state in binary representation, i.e., $x[n] = x_{d_n}[n]x_{d_n-1}[n]\cdots x_1[n]x_0[n].x_{-1}[n]x_{-2}[n]\cdots$. n is the time index. The subscript denotes the bit-level, i.e. d_n is the highest non-zero bit-level occupied by the state at time n.

In polynomial notation, we denote the state as $x[n](z) = x_{d_n}[n]z^{d_n} + x_{d_n-1}[n]z^{d_n-1} + \cdots + x_0[n] + x_{-1}z^{-1}\cdots$. We denote the carry-free state as $x[n](z)$. We will suppress the dummy polynomial variable z wherever obvious. Thus the degree of $x[n]$ is d_n.

Now, consider the carry-free system, $\mathcal{S}$, with a random gain for the control input, with unit system gain (Fig. 2).

$$\begin{aligned} x[n+1](z) &= x[n](z) + b[n](z)\cdot u[n](z) \\ y[n](z) &= x[n](z). \end{aligned} \tag{5}$$

and

$$\begin{aligned} b[n](z) = {} & 1\cdot z^{g_{det}} + 0\cdot z^{g_{det}-1} + 0\cdot z^{g_{det}-2} + \cdots \\ & + b_{g_{ran}}[n]\cdot z^{g_{ran}} + b_{g_{ran}-1}[n]\cdot z^{g_{ran}-1} + \cdots \end{aligned}$$

Basically, b is a 1, followed by 0's, followed by random numbers. g_{det} is the highest deterministic level for b and g_{ran} is the highest random level. So, there are $g_{det} - g_{ran}$ deterministic bits in b.

Our aim is to understand the stability of this system, which is captured by the growth of the random variabled_n. If we think of a pictorial representation of what is going on, we want to make sure that the stack of bits in Fig. 2 keeps shrinking to zero height as fast as possible.

The picture clearly illustrates the ideas that we want to capture with control capacity. The number of bit levels that can be reduced with probability 1 is the zero-error control

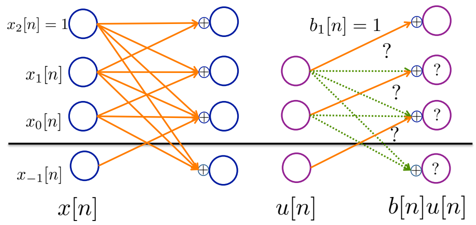

Fig. 2. This figure represents a carry-free model for the basic system model with system gain 1 (Eq. (5)).

capacity of the system [3]. We repeat the definition here for convenience.

Definition 3.1: The zero-error control capacity of the system $\mathcal{S}$ from eq. (5) is defined as the largest constant $C_{ze}(\mathcal{S})$ such that there exists a control strategy $u[0](z), \ldots, u[n](z)$ such that

$$P\left(\frac{1}{n}(d_0 - d_n) \geq C_{ze}(\mathcal{S})\right) = 1. \tag{6}$$

The time index n does not matter.
C_{ze} is essentially the largest decay exponent that is possible for the system state.

A. Carry-free model with side information

Now, we explore the system with side information. In Figure 3, we consider a simple bit-level carry-free model that is the counterpart of system (1). Say the control gain $B[n]$ has one deterministic bit, so that $g_{det} = 1$, but all lower bits are random Bernoulli-$\frac{1}{2}$ bits. Then the controller can only reliably cancel $1 - 0 = 1$ bits of the state each time. The difference between the level of the deterministic bits and the level of the random bits is what determines the number of reliably controllable bits. If the value of the bit at level 0, i.e. b_0, were also known, then we could tolerate a growth through a of two bits at a time. We can think of this as the value of the side information b_0 for this problem.

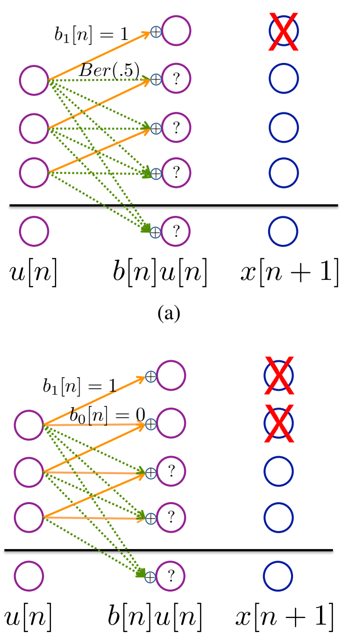

Fig. 3. This system has the highest deterministic link at level $g_{det} = 1$ and the highest unknown link at $g_{ran} = 0$. Bits $b_{-1}[n], b_{-2}[n], \cdots$ are all random Bernoulli-($\frac{1}{2}$). As a result the controller can only influence the top bits of the control going in, and can only cancel one bit of the state. If one extra bit b_0 were known to the controller, it could cancel a second bit of the state.

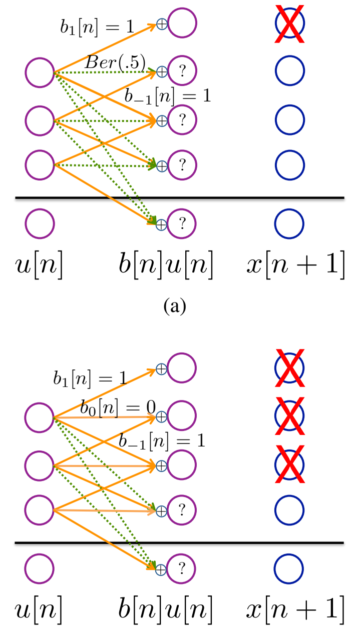

Fig. 4. Consider the following gain for the controller in (a): $b_1[n] = 1, b_{-1}[n] = 1$ are deterministically known, but all other links are Bernoulli-($\frac{1}{2}$). Only a gain of $\log a = 1$ can be tolerated in this case. Now, say side information regarding the value of $b_0[n]$ is received as in (b). This suddenly buys the controller not just one, but two bits of growth.

B. A carry-free counterexample

In the portfolio theory literature, it is known that the maximum increase in doubling rate due to side information Z for a set of stocks distributed as T is upper bounded by $I(T; Z)$. It is tempting to conjecture that "a bit buys a bit" for information in control systems as well given the previous carry-free example. However, we see that the following counterexample rejects this conjecture. Consider the carry-free model in Fig. 4. Here $u[n]$ is the control action, and $b[n]u[n] = z[n]$ is the control effect. In Fig. 4(a) the uncertainty in $b_0[n]$ does not allow the controller to utilize the knowledge that $b_{-1}[n] = 1$ and arbitrarily set the bits of $b[n]u[n]$. However, one bit of information $b_0[n]$ in Fig. 4(b), lets the controller buy two bits of gain in the tolerable growth rate as explained in the caption.

This carry-free model represents a real system where $B[n]$ is drawn from a mixture of disjoint uniform distributions, as in Fig. 5(a). The first most significant bit and the third most significant bit are known, but the second most significant

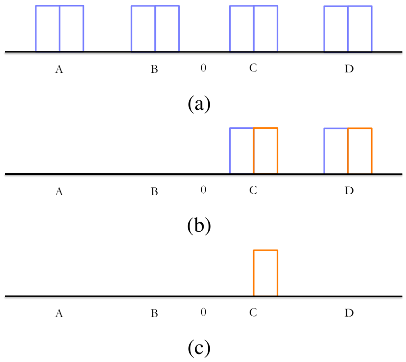

Fig. 5. (a) The distribution of $B[n]$ for the carry-free model. (b) The first and the third bit together tell the controller that $B[n]$ comes from one of the orange parts of the distribution. Since there are two orange sections that are far away, the effective variance of $B[n]$ is not reduced. (c) Once the second bit is also known, the fact that the controller already knew the third bit becomes useful.

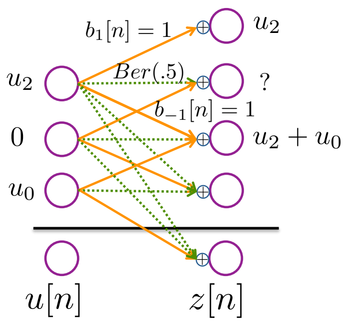

Fig. 6. With two known bits on the gain, the decoder can decode two additional bits about the message $u[n]$ from the received signal $z[n]$. However, these bits are decoded at specific positions. It is not possible to get information out of the received signal position z_2.

bit is not known. The first bit tells us whether $B[n]$ comes from group A and B or group C and D. The third bit only discriminates to the level that is shown in Fig. 5(b), i.e. the controller only knows that $B[n]$ belongs to one of the two orange boxes. So the variance of the distribution isn't actually lowered by much due to the side information. The side information containing the second bit finally lowers the variance, as in Fig. 5(c). Thus, the gain this one-bit of information provides is worth more than a bit.

A communication aside

If the problem was that of pure communication, we could still decode two bits of information about $u[n]$ from $b[n]u[n] = z[n]$. See Fig. 6. Let u_2 and u_0 be the information carrying bits, and set $u_1 = 0$ to zero. Then $z_3 = u_2$ and $z_1 = u_2 + u_0$. With two equations and two unknowns, both $u_2[n]$ and $u_0[n]$ can be recovered at the decoder. In the control problem, this is not possible because of the contamination that is introduced by b_0 at z_2. While communication systems can choose which bits contain relevant information, control systems do not have that flexibility. A bit at a predetermined position must be cancelled or moved by the control action.

In the case of portfolio theory, it is possible to hedge across uncertainty in the system and get "partial-credit" for uncertain quantities. This is not possible in communication and control systems since it is not possible to hedge a control signal in the same way one can hedge a bet.

IV. Control capacity with side information

Now, we define control capacity with side information. The definitions and theorems naturally follow from the definitions of control capacity.

Consider the following system $\mathcal{S}(p_{B,T})$,

$$\begin{aligned} X[n+1] &= X[n] + B[n]U[n] \\ Y[n] &= X[n], \end{aligned} \tag{7}$$

and let $T[i]$ is the side information received by the controller at time i, and $(B[i], T[i]), 0 \leq i \leq n$ are drawn i.i.d. from $p_{B,T}(\cdot,\cdot)$ at each time. $p_B(\cdot), p_{B|T}(\cdot)$ both do not have any atoms. The control signal $U[n]$ can causally depend on $Y[i], 0 \leq i \leq n$ as well as on the side information $T[i], 0 \leq i \leq n$.

Let $\mathcal{F}_n = \sigma(Y_0^n, T_0^n)$ be the sigma-algebra generated by the observations and the side information. $X[0]$ is a random variable with density $p_{X[0]}(\cdot)$ and also does not have an atom at 0.

Now, we can naturally extend the definition in [3] to define control capacity with side information.

Definition 4.1: The Shannon control capacity with causal side information of the system $\mathcal{S}(p_{B,T})$ is

$$C_{sh}(\mathcal{S}|T) = \lim_{n\to\infty} - \min_{U_0^{n-1}(\cdot) \text{ s.t. } U[i](\cdot)\in\mathcal{F}_i} \frac{1}{n}\mathbb{E}\left[\log\left|\frac{X[n]}{X[0]}\right|\right]. \tag{8}$$

The control capacity with side information is the maximum uncertainty (in bits) that can be dissipated from the state using both the observation and the side-information. Using the theorems developed in [3] we can immediately characterize the logarithmic stabilizability of a system.

Theorem 4.1: Consider the system $\mathcal{S}_a(p_{B,T})$ below:

$$\begin{aligned} X[n+1] &= a \cdot X[n] + B[n]U[n] \\ Y[n] &= X[n]. \end{aligned} \tag{9}$$

The distributions of B, T and $X[0]$ are identical to those of system $\mathcal{S}(p_{B,T})$. Then, system $\mathcal{S}_a(p_{B,T})$ is logarithmically stabilizable with side information $T[i]$ received by the controller at time i if and only if

$$C_{sh}(\mathcal{S}|T) \geq \log|a|. \tag{10}$$

This theorem follows from the results in [3], [37] and the proof style is identical to that in [37].

The next theorem shows that the value of the side information is actually computable and can be thought of as a conditional expectation.

Theorem 4.2: The Shannon control capacity of the system $\mathcal{S}(p_{B,T})$ with side information $T[i]$ at time i, is given by:

$$C_{sh}(\mathcal{S}|T) = \mathbb{E}_T\left[\max_k \mathbb{E}\left[-\log\left|1 + B\cdot k\right|\middle|T\right]\right]. \tag{11}$$

Please see [37] for the full details of the proof.

Remark 4.1: For scalar systems like the one considered above, non-causal side information about the actuation channel would actually not provide any capacity gains over that provided by the same information causally. Receiving $T[2]$ at timestep-1 over timestep-2 would not change the system performance.

We can similarly define the η-th moment control capacity with side information for the system $\mathcal{S}$.

Definition 4.2: The η-th moment control capacity with causal side information of the system $\mathcal{S}(p_{B[0],T[0]}, p_{B[1],T[1]}, \cdots)$ is given by

$$C_\eta(\mathcal{S}|T) = \lim_{n\to\infty} \; - \min_{U_0^{n-1}(\cdot) \text{ s.t. } U[i](\cdot)\in\mathcal{F}_i} \frac{1}{n}\frac{1}{\eta} \log \mathbb{E}\left[\left|\frac{|X[n]|^\eta}{|X[0]|^\eta}\right|\right]. \tag{12}$$

A similar theorem holds for computing the value of this control capacity.

Theorem 4.3: The η-th moment control capacity of the system $\mathcal{S}(p_{B,T})$ is given by

$$C_\eta(\mathcal{S}|T) = -\frac{1}{\eta}\log\ \mathbb{E}_T\left[\min_k \mathbb{E}_{B|T}\left[\left|1 + B\cdot k\right|^\eta \Big| T\right]\right]. \tag{13}$$

Example

As an example, we plot the change in Shannon and 2nd-moment control capacities with zero to four bits of side information for a set of actuation channels in Fig. 7. We know that the critical parameter to calculate control capacity is the squared ratio of the mean to the standard deviation of the distribution [37]. For the remainder of this discussion we will refer to this ratio as the SNR of the actuation channel. We plot the control capacities for actuation channels with base "SNR" $\frac{1}{100}, 1$ and 100, and thus mean to standard deviation ratios of $\frac{1}{10}, 1$ and 10. All distributions here have mean 10, however the plots are identical regardless of the mean value and only depends on the SNR.

We consider a uniform distribution on the unreliability in the actuation channel. $B \sim$ Uniform$[b_1, b_2]$. We consider the case where the controller is provided one bit of side information that divides the interval into two halves, and tells the controller whether the realization of B is in $[b_1, \frac{b_1+b_2}{2}]$ or in $(\frac{b_1+b_2}{2}, b_2]$. Two bits of side information resolves the interval more finely, and so on.

The green curves represent the second-moment (solid) and Shannon (dashed) control capacity for the uniform distribution with SNR = 100. For both these curves, as the number of bits of side information increases the slope of both these curves approaches 1 but are a shade below 1 close to 0. (The line extends with the same slope as more side information is provided — this is not shown in the figure).

On the other hand, consider the pink lines that represent the second moment (solid) and Shannon (dashed) control capacities when the SNR $= \frac{1}{100}$. The slope of the dashed line (Shannon) between 0 and 1 is actually slightly greater than 1! In this case, half the time, the first bit of side information reveals perfectly the sign of the distribution and can increase the control capacity by more than one bit. The carry-free model predicted this behavior — the value of a bit can be more than a bit! Of course, as the side information increases the capacity steadily increases and eventually and one bit of side information only increases the control capacity by one bit — we can see that the slope of the curve tends to 1.

The slope of the second-moment control capacity (for SNR $= 1/100$) between 0 and 1 is still less than 1, but we see here that the value of the first bit of side information (that reveals the sign) is still more valuable than the second bit of side information. This curve also converges to slope 1 as the controller gets more side information.

Finally, we come to the control capacities of the distributions with SNR $= 1$ with the yellow curves. (Note here that the values for both the Shannon and second-moment control capacities with one bit of side information (i.e. the points corresponding to x-coordinate 1) are slightly higher than the points for SNR $= 1/100$ even though it is not apparent in this figure). These curves shows a very intriguing phenomenon — the first bit of side information is actually worth less than a bit, and the first bit of side-information is worth less than the second bit that is received. This is certainly something we plan to investigate further.

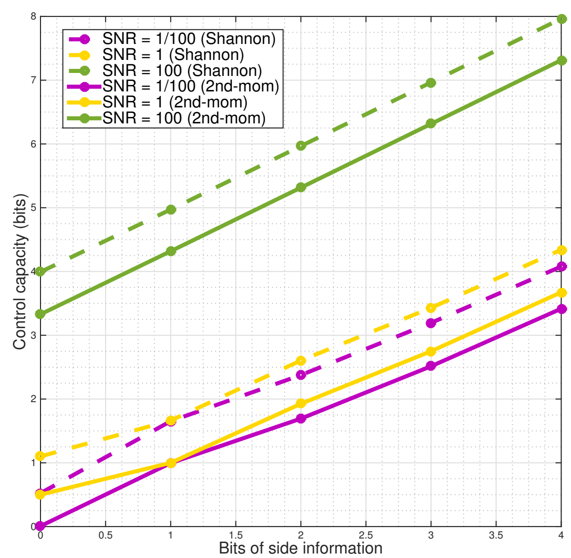

Fig. 7. This plot shows the increase in control capacity with side information. The actuation channels considered have a uniform distribution with different mean/standard deviation ratios (SNR) as in the legend.

V. Conclusions and Future Work

This paper presented a method to calculate the value of side-information in control systems. The fact that we can do this in logarithmic and second-moment (and any other moment) senses develops a full information-theoretic parallel to the uncertainty threshold principle in control and formalizes it.

Calculating the control capacities of different actuation channels with side information revealed some very interesting properties and these certainly must be investigated further. An important open question is: when can a bit be worth more than a bit? When can it only be worth less?

Further, we notice that all the "lines" in Fig. 7 tend to have slope 1 with more side information — is it possible to bound the value of a bit in this limit?

To move beyond the second moment, we believe that side-information should have diminishing returns as the moment $\eta \to \infty$, since we are approaching the worst-case "robust control" limit. This is exactly what preliminary plots indicate but this also warrants a detailed study and analysis.

VI. Acknowledgements

The authors would like to thank the NSF for a GRF and CNS-0932410, CNS-1321155, and ECCS-1343398.

References

[1] L. Schenato, B. Sinopoli, M. Franceschetti, K. Poolla, and S. S. Sastry, "Foundations of control and estimation over lossy networks," *Proceedings of the IEEE*, vol. 95, no. 1, pp. 163–187, 2007.

[2] M. Athans, R. Ku, and S. Gershwin, "The uncertainty threshold principle: Some fundamental limitations of optimal decision making under dynamic uncertainty," *Automatic Control, IEEE Transactions on.*, vol. 22, no. 3, pp. 491–495, 1977.

[3] G. Ranade and A. Sahai, "Control capacity." International Symposium on Information Theory (ISIT), 2015.

[4] S. Tatikonda and S. Mitter, "Control under communication constraints," *Automatic Control, IEEE Transactions on.*, vol. 49, no. 7, pp. 1056–1068, 2004.

[5] A. Sahai and S. Mitter, "The necessity and sufficiency of anytime capacity for stabilization of a linear system over a noisy communication link part I: scalar systems," *Information Theory, IEEE Transactions on.*, vol. 52, no. 8, pp. 3369–3395, 2006.

[6] P. Minero, M. Franceschetti, S. Dey, and G. N. Nair, "Data-rate theorem for stabilization over time-varying feedback channels," *Automatic Control, IEEE Transactions on.*, vol. 54, no. 2, pp. 243–255, 2009.

[7] G. N. Nair, F. Fagnani, S. Zampieri, and R. J. Evans, "Feedback control under data rate constraints: An overview," *Proceedings of the IEEE*, vol. 95, no. 1, pp. 108–137, 2007.

[8] M. Davis, "Anticipative LQG control," *IMA Journal of Mathematical Control and Information.*, vol. 6, no. 3, pp. 259–265, 1989.

[9] R. T. Rockafellar and R.-B. Wets, "Nonanticipativity and L 1-martingales in stochastic optimization problems," in *Stochastic Systems: Modeling, Identification and Optimization, II.* Springer, 1976, pp. 170–187.

[10] S. Flåm, "Nonanticipativity in stochastic programming," *Journal of optimization theory and applications.*, vol. 46, no. 1, pp. 23–30, 1985.

[11] M. A. H. Dempster, *The expected value of perfect information in the optimal evolution of stochastic systems*. Springer, 1981.

[12] K. Back and S. R. Pliska, "The shadow price of information in continuous time decision problems," *Stochastics: An International Journal of Probability and Stochastic Processes.*, vol. 22, no. 2, pp. 151–186, 1987.

[13] G. Tadmor and L. Mirkin, "H-infinity control and estimation with preview part II: fixed-size ARE solutions in discrete time," *Automatic Control, IEEE Transactions on.*, vol. 50, no. 1, pp. 29–40, 2005.

[14] R. H. Middleton, J. Chen, and J. S. Freudenberg, "Tracking sensitivity and achievable H-infinity performance in preview control," *Automatica.*, vol. 40, no. 8, pp. 1297–1306, 2004.

[15] J. Chen, Z. Ren, S. Hara, and L. Qin, "Optimal tracking performance: preview control and exponential signals," *Automatic Control, IEEE Transactions on.*, vol. 46, no. 10, pp. 1647–1653, 2001.

[16] N. Martins, M. Dahleh, and N. Elia, "Feedback stabilization of uncertain systems in the presence of a direct link," *Automatic Control, IEEE Transactions on.*, vol. 51, no. 3, pp. 438–447, 2006.

[17] J. Wolfowitz, J. Wolfowitz, P. Statistiker, J. Wolfowitz, and J. Wolfowitz, *Coding theorems of information theory*. Springer, 1961, no. 31.

[18] A. J. Goldsmith and P. P. Varaiya, "Capacity of fading channels with channel side information," *Information Theory, IEEE Transactions on.*, vol. 43, no. 6, pp. 1986–1992, 1997.

[19] G. Caire and S. Shamai, "On the capacity of some channels with channel state information," *Information Theory, IEEE Transactions on.*, vol. 45, no. 6, pp. 2007–2019, 1999.

[20] A. Lapidoth and S. Shamai, "Fading channels: how perfect need perfect side information be?" *Information Theory, IEEE Transactions on.*, vol. 48, no. 5, pp. 1118–1134, 2002.

[21] M. Medard, "The effect upon channel capacity in wireless communications of perfect and imperfect knowledge of the channel," *Information Theory, IEEE Transactions on.*, vol. 46, no. 3, pp. 933–946, 2000.

[22] A. J. Goldsmith and M. Medard, "Capacity of time-varying channels with causal channel side information," *Information Theory, IEEE Transactions on.*, vol. 53, no. 3, pp. 881–899, 2007.

[23] S. Kotagiri and J. Laneman, "Multiaccess channels with state known to some encoders and independent messages," *EURASIP Journal on Wireless Communications and Networking.*, vol. 2008, pp. 1–14, 2008.

[24] M. A. Maddah-Ali and D. Tse, "Completely stale transmitter channel state information is still very useful," *Information Theory, IEEE Transactions on.*, vol. 58, no. 7, pp. 4418–4431, 2012.

[25] T. Cover and J. Thomas, *Elements of information theory*, 2nd ed. Wiley, 2006.

[26] S. S. Pradhan, J. Chou, and K. Ramchandran, "Duality between source coding and channel coding and its extension to the side information case," *Information Theory, IEEE Transactions on.*, vol. 49, no. 5, pp. 1181–1203, 2003.

[27] E. Martinian and G. W. Wornell, "Source coding with fixed lag side information," *arXiv preprint cs/0410008*, 2004.

[28] M. Gastpar, "To code or not to code," Ph.D. dissertation, 2002.

[29] E. Martinian, G. W. Wornell, and R. Zamir, "Source coding with distortion side information," *Information Theory, IEEE Transactions on.*, vol. 54, no. 10, pp. 4638–4665, 2008.

[30] ——, "Source coding with distortion side information at the encoder," in *Data Compression Conference (DCC).* IEEE, 2004, pp. 172–181.

[31] J. Kelly Jr, "A new interpretation of information rate," *Information Theory, IEEE Transactions on.*, vol. 2, no. 3, pp. 185–189, 1956.

[32] H. H. Permuter, Y.-H. Kim, and T. Weissman, "On directed information and gambling," in *Information Theory (ISIT), IEEE International Symposium on.* IEEE, 2008, pp. 1403–1407.

[33] ——, "Interpretations of directed information in portfolio theory, data compression, and hypothesis testing," *Information Theory, IEEE Transactions on.*, vol. 57, no. 6, pp. 3248–3259, 2011.

[34] S. Tatikonda, "Control under communication constraints," Ph.D. dissertation, 2000.

[35] U. Niesen and M. Maddah-Ali, "Interference alignment: From degrees-of-freedom to constant-gap capacity approximations," in *Information Theory (ISIT), IEEE International Symposium on.*, 2012, pp. 2077–2081.

[36] S. Park, G. Ranade, and A. Sahai, "Carry-free models and beyond," in *Information Theory, IEEE International Symposium on.*, 2012, pp. 1927–1931.

[37] G. Ranade, "Active systems with uncertain parameters: an information-theoretic perspective," Ph.D. dissertation, University of California, Berkeley, 2014.