A Nonstochastic Information Theory for Feedback

Girish N. Nair

Abstract—This paper extends a recently proposed theory of nonstochastic maximin information to study feedback systems with erroneous channels. The concepts of conditional and directed maximin information are introduced, without assuming any statistical structure. It is proved that the zero-error feedback capacity of a stationary memoryless channel coincides with the largest rate of directed maximin information across it. This provides a nonstochastic counterpart to recent results in information theory. This characterization is then used to find a tight condition for exponential uniform stabilizability of a linear time-invariant plant over an erroneous channel, consistent with a result of Matveev and Savkin.

I. INTRODUCTION

In control systems, indeterminate quantities and signals are often modelled without imposing statistical structure. This may happen when performance must be guaranteed always, not just with a high chance, or when the dominant disturbances cannot be assumed to follow a probability law.

When such a control loop is closed over an error-prone communication channel, this nonstochastic view of the plant causes some challenges, since channels are usually modelled stochastically in communications. An obvious solution would be to treat the unknown variables in the control system as random variables (rv’s) with unknown joint distributions, and seek to guarantee the worst-case closed-loop performance almost surely (a.s.). However, such a probabilistic model of the plant may not always be justifiable on physical (or philosophical) grounds, and may impose more structure on the disturbances than is warranted.

In the paper [1], a mixed approach was pursued instead. The disturbances entering a linear, time-invariant (LTI) plant were modelled as bounded, nonstochastic unknowns, but the channel and initial plant state were modelled as a stochastic discrete memoryless channel (DMC) and random variable (rv) $X_0$ respectively. Assuming that each channel transition was statistically independent of past channel transitions and $X_0$, it was proved that to be able to ensure a.s. uniformly bounded states for any admissible disturbance sequence, it is necessary and sufficient that the sum $H$ of the unstable eigenvalues of the plant did not exceed $C_{0f}$, the zero-error feedback capacity [2] of the channel. In communications, the quantity $C_{0f}$ is defined as the largest block-coding bit-rate possible over the DMC such that the probability of a decoding error is exactly zero, with the encoder granted access to the past outputs of the channel. Importantly, this criterion remains tight even if there is no direct feedback from the channel output to the encoder; if the disturbances are bounded but the control inputs can be as large as pleased, then the plant effectively provides a perfect channel feedback path to the encoder. For the analogous state estimation problem, it was shown that the corresponding criterion is that $H \leq C_0$, the zero-error capacity $C_0$ without channel feedback. As $C_{0f}$ is (sometimes strictly) larger than $C_0$, this indicates that over an erroneous channel, stabilization is easier than state estimation, in terms of communication requirements. For the special case when the DMC is errorless, $C_{0f} = C_0$ and the so-called data rate theorem [3], [4], [5], [6], [7], [8], [9] is recovered.

For the purposes of this paper, the results from [1] are interesting for two reasons. Firstly, although its proof uses probability and a law of large numbers, the actual criterion itself reflects no statistical structure: $H$ is given by the fixed dynamical matrix of the plant, and neither $C_{0f}$ nor $C_0$ depend on the nonzero transition probabilities of the DMC. This suggests that imposing a probabilistic structure on the channel and initial state may be more than required for the objective at hand.

Secondly, although $H$ may be interpreted as the intrinsic rate at which the plant generates entropy, the analysis in [1] does not use concepts from information theory. Moreover, to the best of the author’s knowledge, no information-theoretic characterization of $C_{0f}$ has been available up to now, only an operational definition in terms of block codes. In contrast, the ordinary capacity $C$ of a DMC (defined as the largest block-coding bit-rate possible such that the probability of a decoding error can be made as small as pleased without (with) channel feedback, is known to coincide with the maximum rate of (directed) Shannon information $I (\cdot: \to \cdot)$ that the channel can support [10] ([11], [12], [13]). As $C_{0f}$ and $C_0$ are (typically strictly) smaller than $C$ and do not depend on the DMC transition probabilities, it seems plausible that any analogous identities for them should involve an information functional that is more conservative than $I$, and that is nonstochastic.

In the recent paper [14], a theory of uncertainty and information was formulated for systems with no statistical structure. In place of Shannon information, a maximin information index $I_\ast [X;Y]$ was constructed to measure the information common to two nonstochastic uncertain variables $X,Y$; in essence, $I_\ast [X;Y]$ measures the most refined common value that can be agreed on by individually observing $X$ and $Y$. Though motivated by systems with worst-case-like criteria, $I_\ast$ does not generally coincide with the infimum of $I[X;Y]$ over joint distributions with given support, and depends solely on the shape of the support region,
as captured by its taxicab- or overlap-connectedness. It was shown that this index possesses a number of natural properties in common with Shannon information, and that the feedbackless zero-error capacity $C_0$ of a stationary memoryless uncertain channel coincides with the highest rate of maximin information across it. Using these results, it was then established that the state of a noiseless LTI system can be estimated via such a channel with exponential uniform accuracy iff an inequality linking the unstable plant eigenvalues, the desired exponential rate of convergence and $C_0$ was satisfied. Though other nonstochastic information functionals had been previously proposed in [15], [16], these results indicate that $I_*$ may be particularly relevant when dynamical systems or erroneous channels are involved.

This paper builds on [14] to present a theoretical analysis of feedback systems. Specifically, its objectives are to understand whether $I_*$ can be used to i) provide an intrinsic, information-theoretic characterization of the zero-error capacity with feedback, $C_{0f}$, and ii) obtain a condition for exponential uniform stabilizability for an LTI plant controlled via a stationary, erroneous channel in the feedback path from the controller to the actuator. As $C_{0f}$ is not generally equal to $C_0$, the same maximin-information-theoretic identity as in [14] cannot apply. Furthermore, since the controller only sees what was received over the erroneous channel, not what was sent, the problem of stabilization cannot be readily recast as one of state estimation. Thus, these problems are considerably different from those studied in [14].

In sec. II, necessary definitions and concepts from [14] are briefly recalled or generalized, and the new notion of conditional maximin information is defined. In sec. III, the problem of zero-error communication over a stationary memoryless uncertain channel is described, and the main result is that $C_{0f}$ for such a channel coincides with the largest possible rate of directed maximin information across it. This gives a nonstochastic analogue of recent characterizations of the ordinary feedback capacity of a stochastic channel in terms of the maximum directed Shannon information rate. In sec. IV, the main result is that the state of a noiseless LTI plant can be driven exponentially uniformly to the origin via an erroneous channel iff the same inequality as in [14] holds, but with $C_0$ replaced with $C_{0f}$. Due to space constraints, all proofs are omitted.

II. PRELIMINARIES

Before the main results of this paper are presented, the uncertain variable framework and nonstochastic information index $I_*$ of [14] are briefly described.

A. Uncertain Variables, Unrelatedness and Conditioning

The key idea in the framework of [14] is to keep the probability theory convention of regarding an unknown variable as a mapping $X$ from some underlying sample space $\Omega$ to a set $X$ of interest. Such a mapping $X$ is called an uncertain variable ($uv$). As in probability theory, the dependence on $\omega$ is usually suppressed for conciseness, so that a statement such as $X \in K$ means $X(\omega) \in K$. Unlike probability theory, no measure is imposed on $\Omega$.

Given another uv $Y$ taking values in $Y$, write

$$[X] := \{X(\omega) : \omega \in \Omega\},$$

$$[X|Y] := \{X(\omega) : Y(\omega) = y, \omega \in \Omega\},$$

$$[X,Y] := \{(X(\omega),Y(\omega)) : \omega \in \Omega\}.$$  (3)

Call $[X]$ the marginal range of $X$, $[X|Y]$ its conditional range (or range conditional on $Y = y$, and $[X,Y]$, the joint range of $X$ and $Y$. With some abuse of notation, denote the family of conditional ranges (2) as

$$[X|Y] := \{[X|y] : y \in [Y]\},$$

with empty sets omitted. In the absence of stochastic structure, the uncertainty associated with $X$ given $Y$ is determined by the set-family $[X|Y]$. Notice that $\cup_{B \in [X|Y]} B = [X]$, i.e. $[X|Y]$ is an $[X]$-cover. In addition,

$$[X,Y] = \bigcup_{y \in [Y]} [X|y] \times \{y\},$$

i.e. the joint range is fully determined by the conditional and marginal ranges in a manner that parallels the relationship between joint, conditional and marginal probability distributions.

Using this basic framework, a nonstochastic analogue of statistical independence can be defined:

Definition 2.1 (Unrelatedness [14]): A collection of uncertain variables $Y_1,\ldots,Y_m$ is said to be (unconditionally) unrelated if

$$[Y_1,\ldots,Y_m] = [Y_1] \times \cdots \times [Y_m].$$

They are said to be conditionally unrelated given (or unrelated conditional on) $X$ if

$$[Y_1,\ldots,Y_m|x] = [Y_1|x] \times \cdots \times [Y_m|x], \ \forall x \in [X].$$

Like independence, unrelatedness has an alternative characterization in terms of conditioning:

Lemma 2.1 ([14]): Given uncertain variables $X,Y,Z$,

a) $Y,Z$ are unrelated iff the conditional range

$$[Y|z] = [Y], \ \forall z \in [Z].$$

b) $Y,Z$ are unrelated conditional on $X$ iff

$$[Y|z,x] = [Y|x], \ \forall (z,x) \in [Z,X].$$

Definition 2.2 (Markov Uncertainty Chains): The uncertain variables ($uv$’s) $X$, $Y$ and $Z$ are said to form a Markov uncertainty chain $X \leftrightarrow Y \leftrightarrow Z$ if $X,Z$ are unrelated conditional on $Y$.

They are said to form a conditional Markov uncertainty chain given another uv $W$, denoted $(X \leftrightarrow Y \leftrightarrow Z)|W$ or $X \leftrightarrow Y,W \leftrightarrow Z$, if $X,Z$ are unrelated conditional on $(Y,W)$.

Note that by the symmetry of Def. 2.1, $Z \leftrightarrow Y \leftrightarrow X|(W)$ is also a (conditional) Markov uncertainty chain (given $W$).
B. Maximin Information

The framework above can be used to define a nonstochastic analogue \( I^* \) of Shannon’s mutual information functional. Throughout this subsection, \( X, Y, Z \) are arbitrary uncertain variables (uv’s) with marginal ranges \([\cdot]\) and (1) conditional range families \([\cdot|\cdot]\) (4). Set cardinality is denoted by \( |\cdot| \), with the value \( \equiv \) finite, and all logarithms are base 2.

The following notion is a minor generalization of Def. 2.3 in [14].

Definition 2.3 (Overlap Connectedness/Isolation): Let \( S \) be a cover of a set \( S^\circ \).

a) A pair of points \( x \) and \( x' \in S \) is called overlap connected (in \( S^\circ \)), denoted \( x \leftrightarrow x' \), if \( \exists \) a finite sequence \( (C_i)_{i=1}^n \) of sets in \( S^\circ \) such that \( x \in C_1, x' \in C_n \) and each set has nonempty intersection with its predecessor, i.e. \( C_i \cap C_{i-1} \neq \emptyset \), \( \forall i \in [2, \ldots, n] \).

b) A set \( A \subseteq S \) is called overlap connected if every pair of points in \( A \) is overlap connected.

c) A pair of sets \( A, B \subseteq S \) is called overlap isolated if no point in \( A \) is overlap connected with any point in \( B \).

d) An overlap-isolated partition (of \( S \) in \( S^\circ \)) is a cover of \( S \) where every pair of distinct member-sets is overlap isolated.

e) An overlap partition (of \( S \) in \( S^\circ \)) is an overlap-isolated partition each member-set of which is overlap connected.

\( \diamond \)

Symmetry and transitivity guarantee that a unique overlap partition always exists:

Lemma 2.2 (Unique Overlap Partition): For any cover \( S^\circ \) of a set \( S \), there is a unique overlap partition of \( S \) in \( S^\circ \) (Def. 2.3) , denoted \( S^\circ \). Every set \( C \in S^\circ \) satisfies the identities

\[
C = \{ x \in [X] : x \leftrightarrow C \} = \bigcup_{B \in S^\circ ; B \leftrightarrow C} B. \tag{6}
\]

Furthermore, \( S^\circ \) is the unique overlap-isolated partition (of \( S \) in \( S^\circ \)) of maximum cardinality.

Proof: similar to that of Lem. 2.2 in [14].

Using this lemma, a nonstochastic notion of information can be defined:

Definition 2.4 (Maximin Information [14]): The maximin information shared in common by two uncertain variables \( X, Y \) is

\[
I_m[X;Y] := \log |[X|Y]|_S,
\]

where \([X|Y]_S\) is the overlap partition of \([X]\) in the set-family \([X|Y]\) (Lem. 2.2).

\( \diamond \)

Remark: Maximin information can be thought of as the largest number of distinct values that can always be agreed on from individual observations of \( X \) and \( Y \), and can be shown to be symmetric. It is defined in two other equivalent ways in [14], and the name arises from the maximin argument used in one of them. However, for present purposes it is sufficient to consider only (7).

In the following, a new, nonstochastic analogue of conditional Shannon information is developed and shown to have certain natural properties. To begin, for any \( z \in [Z] \), define the conditional-range subfamily

\[
[X|Y,z] := \{ [x|y,z] : y \in [Y|z] \}. \tag{8}
\]

Noting that this family covers \([X|z]\), Lem. 2.2 guarantees the existence of a unique overlap partition \([X|Y,z]_X\) of \([X|z]\) in \([X|Y,z]\). This yields the following definition:

Definition 2.5: The conditional maximin information of \( X, Y \) given \( Z \) is

\[
I_m[X;Y|Z] := \min_{z \in [Z]} \log |([X|Y,z]|_X|. \tag{9}
\]

\( \diamond \)

Remark: Conditional maximin information can be interpreted as the largest (log)-number of distinct functions \( f : [Z] \to [X] \) that can always be distinguished from observing \( Y, Z \) in the sense that for any possible value of \( (y, z) \), at most one valid function \( f \) takes value \( f(z) \in [X|y,z] \). This insight is the key to proving Thm. 4.1. In contrast, the unconditional maximin information \( I_m[X;Y,Z] \) represents the largest (log)-number of distinct points \( g \in [X] \) that can always be distinguished from observing \( Y, Z \), i.e. for any possible value of \( (y, z) \), at most one valid point \( g \) lies in \([X|y,z]\).

Conditional maximin information possesses three important properties in common with its Shannon analogue:

Lemma 2.3 (Properties of Conditional \( I_m \) ): For any uncertain variables \( W, X, Y, Z \), the conditional maximin information (9) satisfies the following properties:

Symmetry:

\[
I_m[X;Y|W] = I_m[Y;X|W].
\]

Monotonicity:

\[
I_m[X;Y|W] \leq I_m[X;Y,Z|W].
\]

Data Processing:

If \( X \leftrightarrow Y \leftrightarrow Z | W \) is a Markov uncertainty-chainconditional on \( W \) (Def. 2.2), then

\[
I_m[X;Z|W] \leq I_m[X;Y|W].
\]

Proof: Omitted.

Note that by the symmetry of Markov uncertainty chains and maximin information, \( I_m[X;Z|W] \leq I_m[Y;Z|W] \).

C. Stationary Memoryless Uncertain Channels

Let \( \mathbb{X} \) be the space of all \( X \)-valued, discrete-time functions \( x : \mathbb{Z}_{\geq 0} \to \mathbb{X} \). An uncertain (discrete-time) signal \( X \) is a mapping from the sample space \( \Omega \) to some function space \( \mathcal{X} \subseteq \mathbb{X} \) of interest. Confining this mapping to any time \( t \in \mathbb{Z}_{\geq 0} \) yields an uncertain variable (uv), denoted \( X(t) \). The uv sequence \( (X(t))_{t=0}^\infty \) is denoted \( X(a : b) \) as with uv’s, the dependence on \( \omega \in \Omega \) will not usually be indicated. Also note that \([X]\) is a subset of the function space \( \mathcal{X} \).

A nonstochastic parallel of the standard notion of a stationary memoryless channel in communications can be defined as follows:

Definition 2.6: A stationary memoryless uncertain channel (SMUC) consists of an input function space \( \mathcal{X} \subseteq \mathbb{X} \), a
set-valued transition function $T : X \to 2^Y$, and a family $\mathcal{G}$ of
valid uncertain input-output signal pairs $(X, Y)$, with ranges
$[X] \subseteq \mathcal{X}$ and $[Y] \subseteq 2^Y$, such that
$$
[Y(t)|x(0:t),y(0:t-1)] = [Y(t)|x(0:t-1)], \quad \forall (x,y) \in [X,Y], t \in \mathbb{Z}_{\geq 0},
$$
The set-valued reverse transition function $R : Y \to 2^X$ of
the channel is
$$R(y) := \{ x \in X : T(x) \ni y \}, \quad \forall y \in Y. \quad (11)
$$

**Remarks:** This is a generalization of Def. 7 in [14],
coinciding with it if there is no feedback from the channel
output to the channel input, i.e. if
$$[X(t)|x(0:t-1),y(0:t-1)] = [X(t)|x(0:t-1)], \quad \forall (x,y) \in [X,Y], t \in \mathbb{Z}_{\geq 0}.
$$

The mapping $T$ in this definition plays a similar role to
a transition probability matrix or kernel of a stochastic
memoryless channel. In particular, any stationary stochastic
memoryless channel can be modelled as a SMUC by remov-
ing zero-probability events and setting
$$T(x) = \{ y \in Y : p_{Y|X}(y|x) > 0 \}. $$

The input function class $\mathcal{X}$ is included to handle possible
constraints such as limited time-averaged transmission power
or input run-lengths. However, in the rest of this paper it will
be assumed that $\mathcal{X} = X^\omega$. 

### III. ZERO-ERROR COMMUNICATION WITH
PERFECT FEEDBACK

The connection between maximin information and the
problem of errorless communication over an erroneous
channel with feedback is now discussed.

As mentioned in the introduction, the zero-error capacity
$C_0$ [2], [17] of a stochastic channel is defined operationally
as the highest average bit-rate at which messages can be
encoded, transmitted and decoded with exactly zero prob-
ability of error, with the encoder not having any access
to the channel outputs. Clearly, $C_0$ is a more conservative
measure than Shannon’s ordinary capacity $C$ [10], which
allows an arbitrarily small probability of decoder error. The
well-known channel coding theorem implies that $C$ for a
stationary, stochastic memoryless channel coincides with the
largest average rate of Shannon information $I$ across it. This
supplies an intrinsic characterization of $C$ that coincides with
the operational definition. In [14], it was shown that $I$, plays
an analogous role for the zero-error capacity. That is, $C_0$ for
a stationary, memoryless uncertain channel coincides with the
highest rate of maximin information $I_*$ across it.

In this section, the focus is on channels where the encoder
has access to all past channel outputs. Shannon showed that
despite this feedback, for a stationary stochastic memoryless
channel the highest message bit-rate yielding an arbitrarily
small decoder error probability does not increase, i.e. the
feedback capacity $C_1$ remains the same as $C$. However, he also showed that such counterintuitive behaviour was
not generally exhibited if the goal was strengthened to
communication with exactly zero probability of error. That
is, if the encoder can see all past channel outputs, then the
highest message bit-rate $C_{0f}$ yielding no decoding errors can
be strictly larger than $C_0$ (and definitely no smaller). The
quantity $C_{0f}$ is called the zero-error feedback capacity of
the channel.

The operational definition of $C_{0f}$ follows. In a feedback
code, a block-length $t+1 \in \mathbb{Z}_{\geq 0}$ is selected, and each distinct
‘message’ is represented by a sequence $f$ of $t+1$ coding
functions $f_k : Y \to X$, $k = 0, \ldots, t$, so that if $y(0:k-1)$ is
the sequence of past channel outputs at any time $k \in [0:t]$, then the input is
$$x(k) = f_k(y(0:k-1)). \quad (12)$$

With some abuse of notation, write
$$x(0:t) \equiv f(y(0:t-1)). \quad (13)$$

The code may thus be represented by a finite set $F$ of such
function sequences $f = f_k_{0:k}$. Define $\mathcal{F}$ to be the set of
all unambiguous, length-$(t+1)$ codes $F$, i.e. such that any
output sequence $y(0:t) \in Y^{t+1}$ is yielded by at most one
coding function sequence in $F$. In other words, there is at
least one $f \in F$ s.t.
$$f(y(0:t-1)) \in \prod_{k=0}^{t} R(y(k)) \quad (14)$$
$$\iff f_k(y(0:k-1)) \in R(y(k)), \forall k \in [0:t]. \quad (15)$$

Then
$$C_{0f} := \sup_{f \in F} |F| = \limsup_{t \to \infty} \frac{|F|}{t+1}, \quad (16)$$
where the limit follows by superadditivity and Fekete’s
lemma.

The main result of this section can now stated.

**Theorem 3.1 (C_{0f} via Conditional Maximin Information):**
For any stationary memoryless uncertain channel (Def. 2.6)
with transition function $T : X \to 2^Y$, input signal space
$\mathcal{X} = X^\omega$ and class $\mathcal{G}$ of valid uncertain input-output signal
pairs, the zero-error feedback capacity $C_{0f}$ (16) satisfies
$$C_{0f} = \sup_{f \in \mathcal{F}} \frac{\sum_{k=0}^{t} \mathbb{I}_* [X(k);Y(k)|Y(0:k-1)]}{t+1}$$
$$= \limsup_{t \to \infty} \frac{\sum_{k=0}^{t} \mathbb{I}_* [X(k);Y(k)|Y(0:k-1)]}{t+1} \quad (17)$$
where $\mathcal{F}$, the projection of $\mathcal{G}$ onto the $[0:t]$ time-interval,
consists of all signal segments $(X(0:t),Y(0:t))$ that cor-
respond to $(X,Y) \in \mathcal{G}$, and $\mathbb{I}_* [\cdot,\cdot|\cdot]$ is conditional maximin
information (Def. 2.5).

**Proof:** Omitted.

**Remarks:** The characterization above is different from the
corresponding, unconditional formula
$$C_0 = \sup_{f \in \mathcal{F}} \frac{\mathbb{I}_* [X(0:t);Y(0:t)]}{t+1}$$
for the feedbackless zero-error capacity [14]. Further observe
that for any \((X(0 : t), Y(0 : t)) \in G_t\),
\[
\mathbb{E}[Y(k)|X(k), y(0 : k - 1)] = \mathbb{E}[Y(k)|X(0 : k - 1), y(0 : k - 1)], \forall y(0 : k - 1) \in \mathbb{E}[Y(0 : k - 1)]
\]
Consequently,
\[
I_s[Y(k); X(k)|Y(0 : k - 1)] = I_s[Y(k); X(0 : k - 1)|Y(0 : k - 1)]
\]
and (17) may be rewritten as
\[
C_{0f} = \lim_{t \to \infty} \sum_{k=0}^{t-1} I_s[X(0 : k); Y(k)|Y(0 : k - 1)]
\]
The sum in (18) is a nonstochastic analogue of directed Shannon information [18], and is introduced here as the directed maximin information from \(X(0 : t)\) to \(Y(0 : t)\). It is known that the largest directed (or undirected) Shannon information rate across a stationary memoryless channel coincides with its feedback capacity \(C_f\) (which in turn coincides with its capacity \(C\)). Theorem 3.1 states that a similar characterization is possible for zero-error communication: the largest directed maximin information rate across a stationary memoryless uncertain channel coincides with its zero-error feedback capacity \(C_{0f}\). However, unlike the stochastic framework, \(C_{0f}\) may be strictly larger than the feedbackless zero-error capacity \(C_0\), even for a memoryless channel.

IV. CONTROLLING LTI PLANTS VIA ERRONEOUS CHANNELS

In this section, \(I_s\) is used to study the problem of controlling a noiseless, linear time-invariant (LTI) plant via an erroneous stationary memoryless uncertain channel (SMUC) (Def. 2.6). No ‘channel feedback’ from the receiver back to the transmitter is assumed, other than through the plant. First, some related work is discussed.

In the case where the channel is an errorless digital bit-pipe, the ‘data rate theorem’ states that the state can be bounded or taken to zero iff \(H\), the sum of the logarithms of the unstable eigenvalues of the plant dynamical matrix, is less than the channel bit-rate. This tight condition holds in both nonstochastic and probabilistic settings, and under different notions of convergence or boundedness, e.g. uniform, in \(r\)th moment or almost surely (a.s.) [3, 4, 5, 6, 7, 8, 9].

However, with channel errors the controller does not necessarily know what encoder sent, and the stabilizability condition becomes highly dependent on the setting and objective. For instance, given a stochastic discrete memoryless channel (DMC) and a noiseless LTI system with random initial state, a.s. convergence of the state or estimation error to zero is possible if and (almost) only if the ordinary channel capacity \(C \geq H\); this was proved for digital packet-drop channels with acknowledgements in [19], and for general DMC’s with or without channel feedback in [20]. The same result also holds for asymptotic stabilizability via an additive white Gaussian noise channel [21], with no channel feedback.

Suppose next that noise perturbs the plant additionally. Assuming channel feedback, bounded stochastic noise, and scalar states, the achievability of bounded \(r\)th state moments is determined not by \(C\), but by the anytime capacity of the channel [22]. However, if the plant noise is treated instead as a bounded, deterministic disturbance, with the channel and initial plant state still modelled stochastically, then a.s. uniformly bounded states are achievable iff \(H < C_{0f}\), the zero-error feedback capacity \(C_{0f}\) of the channel [1]. The quantity \(C_{0f}\) is defined operationally as the largest block-coding bit-rate possible over the DMC such that the probability of a decoding error is exactly zero, with the encoder granted access to the past outputs of the channel. This criterion holds even without direct channel feedback, since the control loop can be used to provide a feedback path through the plant. As \(C_{0f}\) is (sometimes strictly) smaller than \(C\), this is a more restrictive condition than for disturbance-free systems, even if the disturbance bound is arbitrarily small. In rough terms, the reason is that nonstochastic disturbances do not enjoy a law of large numbers that averages them out in the long run. As a result it becomes crucial for no decoding errors to occur in the channel, not just for their average probability to be arbitrarily small. This important result was proved using probability theory, a law of large numbers and volume-partitioning arguments, but no information theory.

In this section, neither the initial state nor the erroneous channel are modelled stochastically. As a consequence, probability and the law of large numbers cannot be employed in the analysis. Instead, directed maximin information is used. For simplicity, it is assumed that there is no plant noise.

In what follows, \(\|\cdot\|\) denotes either the maximum norm on a finite-dimensional real vector space or the matrix norm it induces, and \(B_l(x)\) denotes the corresponding \(l\)-ball \(\{y : \|y - x\| \leq l\}\) centered at \(x\).

Consider an undisturbed linear time-invariant (LTI) plant
\[
X(t + 1) = AX(t) + BU(t) \in \mathbb{R}^n, \quad (20)
\]
\[
Y(t) = GX(t) \in \mathbb{R}^p, \quad \forall t \in \mathbb{Z}_{\geq 0}, \quad (21)
\]
where the initial state \(X(0)\) is an uncertain variable (uv). Assume that \((A, B)\) is controllable and \((G, A)\) is observable. The output \(Y(t)\) of this system is causally encoded via an operator \(\gamma\) as
\[
S(t) = \gamma(t, Y(0 : t)) \in \mathcal{S}, \quad \forall t \in \mathbb{Z}_{\geq 0}, \quad (22)
\]
and each symbol \(S(t)\) is then transmitted over a stationary memoryless uncertain channel, with set-valued transition function \(S \mapsto \mathcal{Q}\) and input function space \(\mathcal{Q}\) (Def. 2.6), yielding a received symbol \(Q(t) \in \mathcal{Q}\). The received symbols are used to produce a control input \(U(t)\) by means of another operator \(\eta\) as
\[
U(t) = \eta(t, Q(0 : t)) \in \mathbb{R}^m, \quad \forall t \in \mathbb{Z}_{\geq 0}. \quad (23)
\]
The pair $(\gamma, \eta)$ is called a coder-controller. Such a pair is said to yield $\rho$-exponential uniform stability on the maxnorm ball $B_l(0)$ if for any $X(0)$ with range $\subseteq B_l(0),$
\[ \limsup_{t \to \infty} \sup_{\omega \in \Omega} \|X(t)\| = \limsup_{t \to \infty} \|\rho^{-t} \|X(t)\|\| = 0, \] (24)
where $l, \rho > 0$ are specified parameters. The main result of this section follows:

**Theorem 4.1:** Consider a linear time-invariant plant (20)–(21), with plant matrix $A \in \mathbb{R}^{n \times n}$, uncertain initial state $X(0)$ and outputs that are coded and controlled (22)–(23) via a stationary memoryless uncertain channel (Def. 2.6) with zero-error feedback capacity $C_{0f}$ (16). Assume that at any time $t \in \mathbb{Z}_{\geq 0}$, the channel output $Q(t)$ is conditionally unrelated with $(X(0), S(0:t - 1))$ given the channel input $S(t)$ (Def. 2.1).

If there exists a coder-controller (22)–(23) yielding $\rho$-uniform exponential stability (24) on an $l$-ball $B_l(0) \subset \mathbb{R}^n$ with $l > 0$, then
\[ C_{0f} \geq \sum_{i=1}^{n} \log \left| \frac{\lambda_i}{\rho} \right| = : H_\rho, \] (25)
where $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of $A$.

Conversely, if (25) holds as a strict inequality, then a coder-controller can be constructed to yield $\rho$-exponential uniform convergence on any $l$-ball.

**Proof:** Omitted.

Like the result of Matveev and Savkin [1], Thm. 4.1 involves the zero-error feedback capacity of the channel. However, the formulation here is completely nonstochastic, and the proof of necessity relies on maxim information techniques rather than a law of large numbers. In addition, Thm. 4.1 is concerned with performance (as measured by a specific convergence rate) in the absence of plant noise, and does not assume that the plant is open-loop unstable. It can be shown that if bounded plant disturbances are present, then (25) with $\rho = 1$ is a tight condition to be able to achieve uniformly bounded states.

Note that the RHS of (25) depends only on the plant eigenvalues that exceed $\rho$ in magnitude. This reflects the fact that the other eigenvalues correspond to plant modes that decay faster than $\rho^{-t}$, and so do not need to be controlled. The criterion derived in [14] for the analogous state estimation problem takes the same form, but with the feedbackless zero-error capacity $C_0$ replacing $C_{0f}$. As $C_{0f}$ can be strictly larger than $C_0$, this implies that over erroneous channels, feedback control is easier than state estimation, in terms of communication requirements.

**V. CONCLUSION**

In this paper, feedback systems were analyzed non-stochastically, using uncertain variables and maxim information. The notion of *conditional* maxim information was introduced and used to show that the zero-error feedback capacity $C_{0f}$ of a stationary memoryless channel coincides with the largest possible rate of *directed* maxim information across it. Using this characterization, a tight criterion was found to be able to exponentially uniformly stabilize the state of an undisturbed linear time-invariant plant with specified exponential speed over an erroneous channel.

Future work will consider the application of maxim information to decentralized control/coordination problems and channels with memory.

**REFERENCES**


