Anytime Capacity
Week 8 Presentation
Summary of the Summary

Constantine Caramanis*
Massachusetts Institute of Technology

April 3, 2001

The work presented in this paper is all taken from the Doctoral Thesis of Anant Sahai [1]. In addition, there are some general ideas inspired by the Doctoral Thesis of Sekhar Tatikonda [2], and also by a series of talks given by Professor Mitter (see [3]).

1 Introduction

First we provide a motivation, giving a general description of the high–level ideas behind the material this exposition seeks to develop. Next, we give an outline of the content of this paper.

1.1 Motivation

The classical set–up for the problem of information transmission across a noisy channel, in the context of capacity achieving block codes, has an inherent causal structure, where the transmitter has access to the entire message to be sent at "time -1". However in many practical situations this may not at all be the case. Indeed, the causality relation may be much more complicated, due to some intricate feedback structure in a communications network, or, as will be the case in the situation examined in this exposition, the message may be given only implicitly, and may unfold in time. In this paper we examine the problem of "tracking" with finite distortion, an **unstable** Markov source, across a noisy channel.

In addition to the dynamic specification of the message, there is yet another difference in the problem statement, namely, the ultimate goal is not information transmission in and of itself, but rather the tracking of the unstable Markov source. In some sense, it is this interaction between control, and communication, that is the fundamental idea behind this paper. The causality of the set–up, and the dynamical nature of the message specification, as will be made clear, render the problem sensitive to errors made arbitrarily in the past. Indeed the total end to end distortion plays a crucial role. Because we are interested in this end–to–end performance, there is a conceptual difference from the usual set up. For instance, it is unclear whether or not we can speak of a channel source separation theorem, as is the case with block coding.

In short, the fundamental elements of this paper rest in the interaction between control and information theory, and therefore delay, and distortion sensitivity to delay, as well as the total end–to–end distortion. These issues are all at the core of the material presented, and examining the case of the unstable Markov source illuminates them all. In the words of the author of the Thesis, "For unstable processes, the issues of streaming, delays, and 'real–time' are absolutely essential to the problem of end–to–end communication" ([1]).

*cmcaram@mit.edu

1.2 Outline

The thread that leads us through the material developed in this paper is the problem of tracking an unstable (scalar) Markov source, across a noisy channel, where by tracking we mean achieving a finite expected distortion. This toy problem serves both as an indication that existing ideas are in some sense insufficient, but also to motivate the need and appropriateness of the new material introduced.

In this paper, we introduce a new notion of channel capacity, that turns out to be stronger than the classical Shannon capacity, yet weaker than Shannon's "zero–error" capacity. In fact, we introduce a parametric family of channel capacities, which in one direction, the limit of the parameter yields classical Shannon capacity, and in the other, the limit yields zero error capacity. This parameter has a quality of service interpretation, as discussed in section 5.

In section 2, we introduce the major problem motivating this paper, and in addition, we illustrate that current notions of capacity, and block coding, are inadequate, with respect to the natural performance measure induced by our problem. In section 3, we develop the new parametrized family of capacities, called "Anytime Capacity." Next we show that the notion of "anytime capacity" is meaningful, in the sense that it is indeed stronger than zero error capacity and weaker than Shannon capacity, in a host of interesting cases. That is, we show for a fairly wide class of channels, that anytime capacity is below Shannon capacity, yet strictly greater than zero error capacity (which, in most interesting cases, is zero). Having shown that anytime capacity is in some sense meaningful, we go on in section 4 to show that it is actually useful. We use anytime capacity, along with appropriate encoders and decoders, to solve the problem posed in section 2. Furthermore, we provide a converse statement, which shows that anytime capacity is in fact *exactly* the right thing to consider. Finally, in section 5, we discuss the quality of service interpretation of the anytime parameter, and in section 6 we conclude.

2 The Problem

In this section we introduce the problem of tracking an unstable Markov source. Before we formulate the general statement of the problem, we give some definitions.

Definition 1 *Given a distortion measure ρ, we say that a random process $\hat{X}_t$ ρ–tracks another random process X_t, if the expected distortion is always bounded, i.e.*

$$\sup_{t>0} E[\rho(X_t, \hat{X}_t)] < \infty.$$

For the purposes of this paper, we restrict our attention to squared error distortion:

$$\rho(X_t, \hat{X}_t) = (X_t - \hat{X}_t)^2.$$

Furthermore, in all that follows, by "noisy channel" we mean a discrete time channel with 0 delay, and unit sampling time, that is a random map from the input alphabet $\mathcal{A}$, to the output alphabet $\mathcal{B}$, neither of which need be discrete. Then we have,

The Problem: *Let X_t be a Markov process with parameter $A > 1$, driven by noise $\{W_t\}$:*

$$X_{t+1} = AX_t + W_t.$$

Is it possible to transmit information across a noisy channel, so that the receiver can construct a process $\hat{X}_t$ that ρ-tracks the process X_t?

Note that the dual problem to this is the rate distortion formulation, namely, can we track an unstable Markov source, as in the above statement, across a noiseless channel with finite rate.

2.1 Usual Methods are Insufficient

First, it is clear that the most naive attempt at tracking, namely, to encode and send the actual state of the source X_t, cannot work when we restrict ourselves to any finite rate. Consider the simplified case of a random walk, i.e. $A = 1$, and W_t is a zero mean binary random variable. By the instability of the source (since $A \geq 1$) the process X_t will assume arbitrarily large values with probability 1. A finite rate code can only encode finitely many values. Therefore this naive attempt is clearly inadequate. Intuition tells us immediately that there must be a better way, for over a noiseless channel, all we need to do is transmit the single bit representing the outcome of the random noise, and we can track the random walk perfectly. This is true for $A \geq 1$ as well. However the assumption of perfect transmission is crucial to this solution. In fact, it is insurmountable. This follows because in the proposed tracking method, the receiver's estimate $\hat{X}_t$ *never* loses sensitivity to bit errors made in the past. In fact, if $A > 1$, the squared distortion can grow exponentially as A^{2n} for bit errors that occurred n time units in the past. Even for $A = 1$ we are in trouble. For consider any block code approach, where we can encode the outcome of the random noise W_t with arbitrarily low probability of error. The point is that while the probability of error may be arbitrarily low, it is nonetheless nonzero. Therefore there is a nonzero probability or error for any given bit. By the memoryless nature of the errors (namely, the errors in decoding block k are independent of the errors made in decoding block $k+l$) the difference of the two processes X_t and $\hat{X}_t$, will behave like a random walk, and thus will attain arbitrarily large values with probability one. Therefore we have (loosely) shown the following

Proposition 1 *Traditional block coding arguments are inadequate to track unstable Markov sources, as in the problem statement above.*

If block coding does not meet our needs, after asking "what will," a second, yet equally important question is to ask what happens of our source channel separation theorems. We come to this in section 4.

2.2 Bit Sensitivity

There is an important distinction that has been implicitly drawn above between block codes, and the tracking of the Markov process. Block codes have the property that the dependence on errors made in the past eventually completely disappears. In fact, this happens as soon as we begin decoding the next block. In other words, if we are not transmitting a Markov process, but rather just codewords, $X_1, X_2, X_3, \ldots$, then if we use traditional block codes to transmit the codewords across the channel, then $\rho(X_t, \hat{X}_t)$ will be independent of any (and all) errors made in decoding X_k, for all $k < t$. Therefore, if the source is stationary,

$$\sup_{t>0} E[\rho(X_t, \hat{X}_t)] = E[\rho(X_1, \hat{X}_1)] < \infty,$$

if our channel has a sufficient capacity.

On the other hand, as we have seen, in the case of an unstable Markov source, not only is this sensitivity never lost, but to the contrary, it may grow exponentially. Something qualitatively different is needed. The first question then is, what channel classification is necessary in order to render this problem tractable? Is a nonzero classical Shannon capacity sufficient? Is a nonzero zero–error capacity necessary? To this question, we now turn.

3 Any Time Capacity

In this section we define a parametrized family of capacities, thus introducing a stronger notion of reliable transmission. We establish the relation with classical Shannon capacity, and then we show that this new notion is meaningful, with respect to a fairly wide class of channels. The fundamental idea behind anytime capacity, and the associated encoders and decoders, is that we have the ability to actually *correct* past errors. Evidently (yet loosely), from the informal discussion above, it is something along these lines that is necessary for tracking an unstable source, if we are to allow our channel to produce any transmission errors. We have the following key definition. Let $f(d) > 0$ be any decreasing function of delay.

Definition 2 *The f–anytime capacity $C_{anytime}(f) = C_a(f)$ of a channel is the supremum of rates at which the channel can transmit data, such that the probability of error is arbitrarily small, and moreover decays at least as fast as the function f does, as a function of the delay d.*

We specialize this definition to the case of exponential functions $f(d) = 2^{-\alpha d}$.

Definition 3 *The α–anytime capacity $C_a(\alpha)$ of a channel, is defined using the above definition, and*

$$C_a(\alpha) = C_a(2^{-\alpha d}).$$

As mentioned above, this introduces a stronger sense of reliable communication, for rather than merely requiring an arbitrarily low probability of error, we insist that it also go to zero at some rate. The next result shows that if this error decay rate is sufficiently fast, then all bits are eventually decoded correctly, with probability one.

Proposition 2 *If $f(d)$ is summable, i.e. $\sum_{d=1}^{\infty} f(d) < \infty$, then there exists an encoder/decoder pair such that for any i, the decoder's estimate for the i^{th} bit converges to the correct value.*

PROOF. This can essentially be viewed as an application of the Borel-Cantelli lemma. ☐

In section 3.1 below, we show that the binary erasure, and AWGN channels have nonzero α–anytime capacity, and we also cite the general result from [1] which lower bounds the anytime capacity away from zero, for a general memoryless channel. First, however, it is clear that we have, for any $\alpha > 0$,

$$C_{\text{zero–error}} \leq C_{\text{anytime}}(\alpha) \leq C_{\text{Shannon}},$$

and in general all inequalities are strict. The first inequality becomes an equality when $\alpha \to \infty$, and the second when $\alpha \to 0$.

3.1 A Lower Bound on Anytime Capacity

In this section, we use random encoders. Notice in the following definition that unlike in the usual definition of block codes, the dependence on past bits never disappears. As above, we define $N_i := \lfloor (\theta + i)R \rfloor$.

Definition 4 *A random rate R offset θ encoder $\mathcal{E}$ generated according to a distribution $P(a)$, generates the channel inputs $a_i = \mathcal{E}_i(s_1^{N_i})$ such that each a_i is drawn according to the distribution $P(a)$ on the channel input alphabet $\mathcal{A}$. $\mathcal{E}_i(x_1^{N_i})$ and $\mathcal{E}_j(y_1^{N_j})$ are independent whenever $i \neq j$, or $x_1^{N_i} \neq y_1^{N_j}$.*

We illustrate this definition in figure 1, for a rate 1/3 encoder. The input bits navigate our way through the binary tree, and the encoder outputs the realizations of *iid* random variables which can be thought of sitting on each leg of each branch of the tree. Since all the random variables sitting on this tree are *iid*, it is clear that the code generated has the property of independence asserted at the end of the definition above.

3.1.1 Binary Erasure Channel

Proposition 3 *For the binary erasure channel with probability of erasure e, we have*

$$C_{anytime}(\alpha) \geq 1 - \log_2(1 + e) - \alpha,$$

and moreover, this lower bound is achievable.

In particular then, $C_a(\alpha) > 0$ whenever $\alpha < 1 - \log_2(1 + e)$.

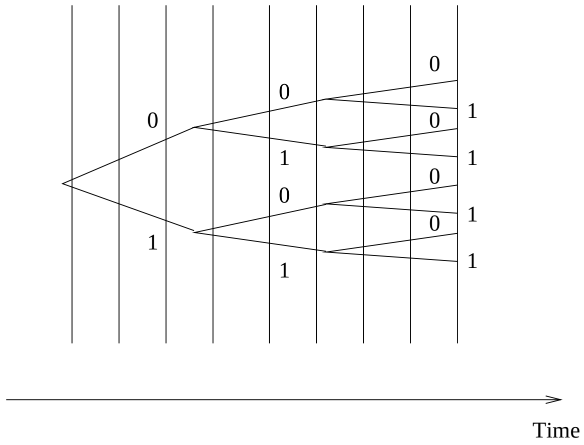

Figure 1: Random Encoder

3.1.2 Additive White Gaussian Channel

The encoder used for this channel will be much the same as the one used in the binary erasure channel. The same tree is formed, yet rather than drawing *iid* binary random variables, we draw *iid* zero mean Gaussian random variables, with variance P (thus satisfying the channel's power constraint). In this case, we use the density to obtain a maximum likelihood decoder. We have an analogous result: a lower bound on the anytime capacity.

Proposition 4 *The power constrained AWGN channel has*

$$C_{anytime}(\alpha) \geq \frac{1}{2}\log_2(1+\frac{P}{2}) - \alpha,$$

and this bound is achievable.

3.1.3 The General Case

We state, but do not prove, the general statement (but, as for the proof of all the results here, see [1]).

Theorem 1 *For a memoryless channel with block random coding exponent $E_r(R)$, we have*

$$C_{anytime}(E_r(R)\log_2 e) \geq R\log_2 e,$$

where $E_r(R)$ is Gallager's standard block coding exponent (see [4]) and $e = 2.71828\ldots$.

Note that all of our statements have been made and then proved for random encoders, that is to say, encoders/decoders that share common randomness. We can think of our results on performance of random encoders to be the expected performance of deterministic encoders with respect to the measure generated by the randomness used above. Recall that the usual achievability proofs, e.g. in [5], work in much the same way, and then argue that there must exist a deterministic encoder that achieves the given bounds. In our case, however, we are considering an expectation over an *uncountable* number of infinite encoders, so showing the existence of a deterministic encoder with desired performance is no longer as simple. See [1] for further details.

4 The Solution to the Problem

In this section we use the notion of anytime capacity introduced in section 3 above, to solve the problem of tracking the unstable source. We show first, that if a noisy channel has a sufficiently large α–anytime capacity for a sufficiently large α, then we can track a scalar unstable Markov process across the channel. Conversely, we show that this is exactly what we need, by showing that if we can track an unstable Markov source across a noisy channel, then the channel has appropriate α–anytime capacity.

4.1 The Direct Part

In section 2 of this paper, we saw that block codes are inadequate to address the problem of tracking an unstable Markov source. The primary reason for this is that the effect of errors is at the very least cumulative, rather than dying off with time. In section 3 above, we developed encoders that create codes whose dependence in time does not die off. Therefore while rendering the decoding sensitive to past errors, nevertheless this dependence allows for the possibility of *correcting* past errors. In this section we show that if a channel has sufficient α–anytime capacity, for a sufficiently large α, then it can track an unstable source. We expect that the amount of anytime capacity we require is related to the exponential growth rate of the unstable source. Indeed this turns out to be the case. This is made precise by introducing distortion measures $\Delta^+(j,t)$, and $\Delta^-(j,t)$, that measure the greatest possible distortion caused in $\hat{X}_t$, by making errors on bits *after* j, and *before* j, respectively. Then we have,

Theorem 2 *Consider a source code with finite expected distortion D_R, that emits a rate R stream of bits. Suppose our noisy channel has $C_a(f) > R$ for a function $f(d)$ that decays quickly enough so that*

$$\lim_{\delta' \to \infty} \sum_{n=0}^{\infty} f(\delta' + \frac{n}{R}) \Delta^+(\frac{n}{R}) = 0.$$

Then, we can trasmit the source code across the noisy channel, with finite expected distortion that can be made to approach D_R arbitrarily (as long as we are willing to accept sufficient additional end–to–end delay δ').

Now the following theorem, solving our problem from section 2, follows as a corollary to the above.

Theorem 3 *A scalar discrete time unstable Markov process as given above, driven by bounded noise, can be tracked with finite squared–error–distortion across any noisy channel such that there is some $\varepsilon > 0$ for which $C_a(2\log_2 A + \varepsilon) > \log_2 A$, for the channel.*

Besides providing a direct solution to our problem, theorem 2 above also provides the answer to another question we posed earlier: The question of source channel separation. Indeed theorem 2 says that if our aim is to encode a source to some distortion D_R, and then transmit it across a noisy channel, we can optimally encode our source, and then cascade it with an anytime encoder, as described above.

4.1.1 The Converse Direction

In this section we state the theorem that says that in fact, we *need* to have a sufficient anytime capacity with sufficiently large α, in order to track an unstable Markov source. The proof of the necessity follows a different direction than the usual converse arguments follow. We show by construction that if we can track an unstable Markov source across a noisy channel, then that noisy channel must have a certain anytime capacity. We show this by actually constructing the anytime encoders and decoders.

In this section we give the main idea behind the proof, and leave the details to [1]. The fundamental idea is to take the input sequence S_1^∞, and construct a simulation of an unstable source. Then from the finite–distortion tracking of this source at the receiver, we try to recover the original bit sequence with probability of error decaying at the appropriate rate. Before we discuss the ideas behind the proof, we give the main statment:

Theorem 4 *Suppose we can track any scalar unstable Markov process X_t with parameter $A > 1$, driven by a bounded noise signal W_t, with $||W_t|| \leq \frac{\Omega}{2}$, across a given channel. Specifically, suppose that for this channel we have*

$$P(|\hat{X}_t - X_t| \geq \Delta) \leq f(\Delta).$$

Then, $\forall \varepsilon > 0$, this channel has

$$C_{anytime}(\tilde{f}) > \log_2 A - \varepsilon,$$

where $\tilde{f} := f(\delta' 2^{d \log_2 A})$, for some $\delta' > 0$.

Note that in particular, this theorem tells us that if we can η–track any process of the class above, then we can take $\tilde{f}$ to be exponential, and therefore $C_a(\eta \log_2 A) > \log_2 A - \varepsilon$.

The fundamental idea lies in finding a clever representation of the real numbers $\mathbb{R}$ in a way that decoding the first bits will somehow be more reliable than the subsequent bits. We achieve this by simulating an unstable Markov source using the incoming bits. It is plausible (and indeed true) that since $A > 1$, the "noise" we add at the beginning (i.e. the earlier bits) somehow have a larger impact on X_t, as their effect is amplified exponentially by A. This is indeed what happens, and it allows us to create an anytime encoder/decoder pair as specified in the theorem.

5 Quality of Service

In this section we illustrate at a high level the interpretation of the α parameter of anytime capacity as a quality of service measure.

The rate of a particular channel can be viewed as a resource, which is additively distributed in some manner, among the "customers" using the channel, as it may be divided among the various demands on the channel by straightforward time sharing. However, if we have feedback at our disposal, there is more we can do than simply time sharing. In other words, there are other ways of differentiating service, besides allocating a larger or smaller fraction of the available rate R to a given user. Using a simple vector unstable Markov process, we can see that dividing the rate up among the "users" or "customers" may not be enough, if subsequently the bits of the source code are all treated alike in the transmission layer. However, by considering α–anytime capacity requirements, it is possible to provide the bits with differentiated service, for instance different priorities in transmission, thus effectively achieving different α–capacities for different customers.

6 Conclusions

In this paper we have developed and discussed a new notion of reliability in noisy channel communication, called anytime capacity. One of the key motivations behind this development has been the interconnection of control and communication. When there is a value of information in regards to controlling a system, then the paradigm of reliable communication must, as [1] tries to show, be qualitatively different. This is intimately related to issues of delay, and causality, and how both control, and affect the system at hand.

References

[1] Sahai, A. "Anytime Information Theory." Ph.D. Dissertation, MIT 2001.

[2] Tatikonda, S. "Control Under Communication Constraints." Ph.D. Dissertation, MIT 2000.

[3] Mitter, S. "Control with Limited Information: the role of Systems Theory and Information Theory."

[4] Gallager, R. "Information Theory and Reliable Communication," NY: John Wiley and Sons, 1971.

[5] Cover, T.; Thomas, J, "Elements of Information Theory," NY: John Wiley, 1991.