

Covering Point Patterns

Amos Lapidoth
ETH Zurich
Zurich, Switzerland
lapidoth@isi.ee.ethz.ch

Andreas Malär
ETH Zurich
Zurich, Switzerland
amalaer@ee.ethz.ch

Ligong Wang
MIT
Cambridge, MA, USA
wlg@mit.edu

Abstract—A source generates a “point pattern” consisting of a finite number of points in an interval. Based on a binary description of the point pattern, a reconstructor must produce a “covering set” that is guaranteed to contain the pattern. We study the optimal trade-off (as the length of the interval tends to infinity) between the description length and the least average Lebesgue measure of the covering set. The trade-off is established for point patterns that are generated by a Poisson process. Such point patterns are shown to be the most difficult to describe. We also study a Wyner-Ziv version of this problem, where some of the points in the pattern are known to the reconstructor but not to the encoder. We show that this scenario is as good as when they are known to both encoder and reconstructor.

I. INTRODUCTION

Imagine a controller that receives a request that a computer be on at certain epochs. If the controller could describe these epochs to the computer with infinite precision, then the computer would turn itself on only at these epochs and be in sleep mode at all other times. If the controller cannot describe the epochs at all, then the computer must be on all the time. In this paper we study the trade-off between the bit rate with which the epochs can be described and the percentage of time the computer must be on.

More specifically, we consider a source that generates a “point pattern” consisting of a finite number of points in the interval $[0, T]$. Based on a binary description of the pattern, a reconstructor must produce a “covering-set”—a subset of $[0, T]$ containing all the points. There is a trade-off between the description length and the minimal Lebesgue measure of the covering-set. This trade-off is formulated as a continuous-time rate-distortion problem in Section III. In this paper we investigate this trade-off in the limit where T tends to infinity.

For point patterns that are generated by a Poisson process of intensity λ , we show that, for the reconstructor to produce covering-sets of average measure not exceeding DT , the required description rate in bits per second is $-\lambda \log D$. This result is closely related to results on the capacity of the ideal peak-limited Poisson channel [1]–[4]. In fact, the two problems can be considered dual in the sense of [5].¹

Rate-distortion problems for Poisson processes under different distortion measures were studied in [6]–[10]. It is interesting that our rate-distortion function, $-\lambda \log D$, is equal to the ones in [8] and in [10], where a queueing distortion

¹The results of [5] are not directly applicable here because our problem is of a continuous-time nature.

measure was considered. This is no coincidence, because the Poisson channel is closely related to the queueing channel introduced in [11].

We also show that the Poisson process is the most difficult to cover, in the sense that any point process that, with high probability, has no more than λT points in $[0, T]$ can be described with $-\lambda \log D$ bits per second. This is true even if an adversary selects the point pattern, provided that the encoder and the reconstructor are allowed to use random codes.

Finally, we consider a Wyner-Ziv setting [12] of the problem, where some points in the pattern are known to the reconstructor but the encoder does not know which ones. This can be viewed as a dual problem to the Poisson channel with noncausal side-information [13]. We show that in this setting one can achieve the same minimum rate as when the transmitter *does* know the reconstructor’s side-information.

The rest of this paper is arranged as follows: in Section II we introduce some notation; in Section III we present the result for Poisson processes; in Section IV we present the results for general point processes and arbitrary point patterns; and in Section V we present the results for the Wyner-Ziv setting.

II. NOTATION

We use lower-case letters like x to denote numbers, and upper-case letters like X to denote random variables. We use boldface lower-case letters like \mathbf{x} to denote vectors, functions from the reals, or point patterns, depending on the context. If \mathbf{x} is a vector, x_i denotes its i th element. If \mathbf{x} is a function, $x(t)$ denotes its value at $t \in \mathbb{R}$. If \mathbf{x} is a point pattern, $n_{\mathbf{x}}(\cdot)$ denotes its counting function, so $n_{\mathbf{x}}(t_2) - n_{\mathbf{x}}(t_1)$ is the number of points in \mathbf{x} that fall in the interval $(t_1, t_2]$. We use bold-face upper-case letters like \mathbf{X} to denote random vectors, random functions, or random point processes. The random counting function corresponding to a point process \mathbf{X} is denoted $N_{\mathbf{X}}(\cdot)$.

We denote by $\text{Ber}(p)$ the Bernoulli distribution of parameter p , which assigns probability p to the outcome 1 and probability $(1 - p)$ to the outcome 0.

III. COVERING A POISSON PROCESS

Consider a Poisson process \mathbf{X} of intensity λ on the interval $[0, T]$. Its counting function $N_{\mathbf{X}}(\cdot)$ satisfies

$$\Pr[N_{\mathbf{X}}(t + \tau) - N_{\mathbf{X}}(t) = k] = \frac{e^{-\lambda\tau} (\lambda\tau)^k}{k!}$$

for all $\tau \in [0, T]$, $t \in [0, T - \tau]$ and $k \in \{0, 1, \dots\}$.

The encoder maps the realization of the Poisson process to a message in $\{1, \dots, 2^{TR}\}$, where R is the description rate in bits per second. The reconstructor then maps this message to a $\{0, 1\}$ -valued, Lebesgue-measurable, signal $\hat{x}(t)$, $t \in [0, T]$. We wish to minimize the length of the region where $\hat{x}(t) = 1$ while guaranteeing that all points in the original Poisson process lie in this region. See Figure 1 for an illustration.

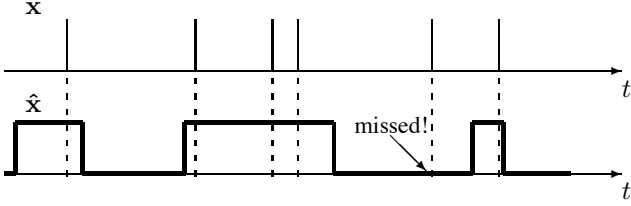


Fig. 1. Illustration of the problem.

More formally, we formulate this problem as a continuous-time rate-distortion problem, where the distortion between the point pattern \mathbf{x} and the reproduction signal $\hat{\mathbf{x}}$ is

$$d(\mathbf{x}, \hat{\mathbf{x}}) \triangleq \begin{cases} \frac{\mu(\hat{x}^{-1}(1))}{T}, & \text{if all points in } \mathbf{x} \text{ are in } \hat{x}^{-1}(1) \\ \infty, & \text{otherwise.} \end{cases} \quad (1)$$

Here $\mu(\cdot)$ denotes the Lebesgue measure.

We say that (R, D) is an achievable rate-distortion pair for \mathbf{X} if, for every $\epsilon > 0$, there exists some $T_0 > 0$ such that, for every $T > T_0$, there exist an encoder $f_T(\cdot)$ and a reconstructor $\phi_T(\cdot)$ of rate $R + \epsilon$ bits per second that, when applied to \mathbf{X} on $[0, T]$, result in

$$\mathbb{E}[d(\mathbf{X}, \phi_T(f_T(\mathbf{X})))] \leq D + \epsilon.$$

Denote by $R(D, \lambda)$ the minimal rate R such that (R, D) is achievable for the Poisson process of intensity λ . Define

$$R_{\text{Pois}}(D, \lambda) \triangleq \begin{cases} -\lambda \log D \text{ bits per second,} & D \in (0, 1) \\ 0, & D \geq 1. \end{cases} \quad (2)$$

Theorem 1: For all $D, \lambda > 0$,

$$R(D, \lambda) = R_{\text{Pois}}(D, \lambda). \quad (3)$$

To prove Theorem 1, we propose a scheme to reduce the original problem to one for a discrete memoryless source. This is reminiscent of Wyner's scheme for reducing the peak-limited Poisson channel to a discrete memoryless channel [3]. We shall show the optimality of this scheme in Lemma 1, and we shall then prove Theorem 1 by computing the best rate that is achievable using this scheme.

Scheme 1: We divide the time interval $[0, T]$ into T/Δ slots² of duration Δ . The encoder first maps the original point pattern \mathbf{x} to a $\{0, 1\}$ -valued vector \mathbf{x}^Δ of $\frac{T}{\Delta}$ components in the following way: if \mathbf{x} has at least one point in the slot

²If T is not divisible by Δ , we replace it with $T' = \lceil \frac{T}{\Delta} \rceil \Delta$. When Δ tends to zero, the difference between RT and RT' tends to zero. Consequently, we shall ignore this edge effect and assume that T is divisible by Δ .

$((i-1)\Delta, i\Delta]$, then we set the i th component of \mathbf{x}^Δ to 1. Otherwise, we set it to zero. The encoder then maps \mathbf{x}^Δ to a message in $\{1, \dots, 2^{TR}\}$.

Based on the encoder's message, the reconstructor produces a $\{0, 1\}$ -valued length- $\frac{T}{\Delta}$ vector $\hat{\mathbf{x}}^\Delta$ that meets the distortion criterion

$$\mathbb{E}[d^\Delta(\mathbf{X}^\Delta, \hat{\mathbf{X}}^\Delta)] \leq D + \epsilon,$$

where the distortion measure $d^\Delta(\cdot, \cdot)$ between vectors is defined in terms of the single-letter distortion function

$$\begin{aligned} d^\Delta(0, 0) &= 0 \\ d^\Delta(0, 1) &= d^\Delta(1, 1) = 1 \\ d^\Delta(1, 0) &= \infty. \end{aligned}$$

It then maps $\hat{\mathbf{x}}^\Delta$ to the piecewise-constant continuous-time signal \hat{x}

$$\hat{x}(t) = \hat{x}_{\lceil \frac{t}{\Delta} \rceil}^\Delta, \quad t \in [0, T].$$

Scheme 1 reduces the task of designing a code for \mathbf{X} subject to the distortion $d(\cdot, \cdot)$ to the task of designing a code for the vector \mathbf{X}^Δ subject to the distortion $d^\Delta(\cdot, \cdot)$ because

$$d(\mathbf{x}, \hat{\mathbf{x}}) = d^\Delta(\mathbf{x}^\Delta, \hat{\mathbf{x}}^\Delta). \quad (4)$$

When \mathbf{X} is a Poisson process of intensity λ , the components of \mathbf{X}^Δ are independent and identically distributed (IID), and each is $\text{Ber}(1 - e^{-\lambda\Delta})$. Let $R_\Delta(D, \lambda)$ denote the rate-distortion function for \mathbf{X}^Δ and $d^\Delta(\cdot, \cdot)$. If we combine Scheme 1 with an optimal code for \mathbf{X}^Δ subject to $\mathbb{E}[d^\Delta(\mathbf{X}^\Delta, \hat{\mathbf{X}}^\Delta)] \leq D + \epsilon$, we can achieve any rate that is larger than

$$\frac{R_\Delta(D, \lambda) \text{ bits}}{\Delta \text{ seconds}}.$$

The next lemma, which is reminiscent of [4, Theorem 2.1], shows that when we let Δ tend to zero, there is no loss in optimality in using Scheme 1.

Lemma 1: For all $D, \lambda > 0$,

$$R(D, \lambda) = \lim_{\Delta \downarrow 0} \frac{R_\Delta(D, \lambda)}{\Delta}. \quad (5)$$

Proof: See Appendix. ■

Proof of Theorem 1: We derive $R(D, \lambda)$ by computing the right-hand side (RHS) of (5). To compute $R_\Delta(D, \lambda)$ we apply Shannon's formula [14]³ for the rate-distortion function of a discrete memoryless source

$$R_\Delta(D, \lambda) = \min_{P_{\hat{Z}|Z}: \mathbb{E}[d_\Delta(Z, \hat{Z})] \leq D} I(Z; \hat{Z}). \quad (6)$$

When $D \in (0, 1)$, the conditional distribution $P_{\hat{Z}|Z}$ that achieves the minimum on the RHS of (6) is

$$\begin{aligned} P_{\hat{Z}|Z}^*(1|0) &= De^{\lambda\Delta} - e^{\lambda\Delta} + 1, \\ P_{\hat{Z}|Z}^*(1|1) &= 1. \end{aligned}$$

³Although our distortion function is unbounded, the all-one reconstruction sequence yields a bounded distortion, so Shannon's formula applies.

Computing the mutual information $I(Z; \hat{Z})$ corresponding to this conditional law yields that for all $\lambda > 0$ and $D \in (0, 1)$,

$$R_\Delta(D, \lambda) = H_b(D) - e^{-\lambda\Delta} H_b(De^{\lambda\Delta} - e^{\lambda\Delta} + 1), \quad (7)$$

where $H_b(\cdot)$ denotes the binary entropy function.

When $D \geq 1$, it is optimal to choose $\hat{Z} = 1$ (deterministically), yielding

$$R_\Delta(D, \lambda) = 0, \quad D \geq 1. \quad (8)$$

Combining (5), (7), and (8) and computing the limit as Δ tends to zero yields (3). ■

IV. COVERING GENERAL POINT PROCESSES AND ARBITRARY POINT PATTERNS

We next consider a general point process \mathbf{Y} . We assume that there exists some λ such that

$$\lim_{t \rightarrow \infty} \Pr \left[\frac{N_{\mathbf{Y}}(t)}{t} > \lambda + \delta \right] = 0 \quad \text{for all } \delta > 0. \quad (9)$$

For example, \mathbf{Y} could be an ergodic process whose expected number of points per second is less than or equal to λ .

Since the Poisson process is memoryless, one naturally expects it to be the most difficult to describe. This is indeed the case, as the next theorem shows.

Theorem 2: The pair $(R_{\text{Pois}}(D, \lambda), D)$ is achievable on any point process satisfying (9).

Before proving Theorem 2, we state a stronger result. Suppose that a point pattern \mathbf{z} is generated by an adversary with the only constraint that it be in the interval $[0, T]$ and that it contain no more than λT points. The corresponding counting function $n_{\mathbf{z}}(\cdot)$ must hence satisfy

$$n_{\mathbf{z}}(T) \leq \lambda T. \quad (10)$$

The encoder and the reconstructor are allowed to use random codes. That is, they fix a distribution on all (deterministic) codes of a given rate on $[0, T]$, and they use this distribution to generate a code, which is not revealed to the adversary. They then apply it to the point pattern \mathbf{z} chosen by the adversary. We say that (R, D) is achievable with random coding against an adversary subject to (10) if, for every $\epsilon > 0$, there exists some T_0 such that, for every $T > T_0$, there exists a random code on $[0, T]$ of rate $R + \epsilon$ such that the expected distortion between *any* \mathbf{z} satisfying (10) and its reconstruction is smaller than $D + \epsilon$.

Theorem 3: The pair $(R_{\text{Pois}}(D, \lambda), D)$ is achievable with random coding against any adversary respecting (10).

Proof: When $D \geq 1$, the encoder does not need to describe the pattern: the reconstructor simply produces the all-one function, yielding distortion 1 for any \mathbf{z} . Hence the pair $(0, D)$ is achievable with random coding.

Next consider $D \in (0, 1)$. We use Scheme 1 of Section III to reduce the original problem to one of random coding for an arbitrary discrete-time sequence \mathbf{z}^Δ . Here the vector \mathbf{z}^Δ is $\{0, 1\}$ -valued, has $\frac{T}{\Delta}$ components, and satisfies

$$\sum_{i=1}^{T/\Delta} z_i^\Delta \leq \lambda T. \quad (11)$$

We shall construct a random code of rate $\frac{R}{\Delta}$ which, when applied to any \mathbf{z}^Δ satisfying (11), yields

$$\mathbb{E} \left[d^\Delta(\mathbf{z}^\Delta, \hat{\mathbf{Z}}^\Delta) \right] < D + \epsilon,$$

where the random vector $\hat{\mathbf{Z}}^\Delta$ is the result of applying the random encoder and decoder to \mathbf{z}^Δ . Combined with Scheme 1 this random code will yield a random code for the continuous-time point pattern \mathbf{z} that achieves the rate-distortion pair (R, D) .

Our discrete-time random code consists of 2^{TR} $\{0, 1\}$ -valued, length- $\frac{T}{\Delta}$ random sequences $\hat{\mathbf{Z}}_m^\Delta$, $m \in \{1, \dots, 2^{TR}\}$. The first sequence $\hat{\mathbf{Z}}_1^\Delta$ is chosen deterministically to be the all-one sequence. The other $2^{TR} - 1$ sequences are drawn independently, with each sequence drawn IID $\text{Ber}(D)$.

To describe a source sequence \mathbf{z}^Δ , the encoder looks for a codeword $\hat{\mathbf{z}}_m^\Delta$, $m \in \{2, \dots, 2^{TR}\}$ such that

$$\hat{z}_{m,i}^\Delta = 1 \text{ whenever } z_i^\Delta = 1. \quad (12)$$

If it finds one or more such codewords, it sends the index of the first one; otherwise it sends the index 1. The reconstructor produces the sequence $\hat{\mathbf{z}}_m^\Delta$, where m is the index it received from the encoder.

We next analyze the expected distortion of this random code for a fixed \mathbf{z}^Δ satisfying (11). Define

$$\mu \triangleq T^{-1} \sum_{i=1}^{T/\Delta} z_i^\Delta,$$

and note that by (11) $\mu \leq \lambda$. Let \mathcal{E} be the event that the encoder cannot find $\hat{\mathbf{z}}_m^\Delta$, $m \in \{2, \dots, 2^{TR}\}$ satisfying (12). If \mathcal{E} occurs, the encoder produces the index 1, and the resulting distortion is 1. The probability that a randomly drawn codeword $\hat{\mathbf{Z}}_m^\Delta$ satisfies (12) is

$$D^{\mu T} \geq D^{\lambda T} = 2^{(\lambda \log D)T}.$$

Because the codewords $\hat{\mathbf{Z}}_m^\Delta$, $m \in \{2, \dots, 2^{TR}\}$ are chosen independently, if $R > -\lambda \log D$, then $\Pr[\mathcal{E}] \rightarrow 0$ as $T \rightarrow \infty$. Hence, for large enough T , the contribution to the expected distortion from the event \mathcal{E} can be ignored.

We next analyze the expected distortion conditional on \mathcal{E}^c . The reproduction $\hat{\mathbf{Z}}^\Delta$ has the following distribution: at positions where \mathbf{z}^Δ is 1, $\hat{\mathbf{Z}}^\Delta$ must also be 1; at other positions the components of $\hat{\mathbf{Z}}^\Delta$ are IID $\text{Ber}(D)$. (These components were not “looked at” in the process of generating the index.) Thus, the expected value of $\sum_{i=1}^{T/\Delta} \hat{z}_i^\Delta$ is $\mu T + D(\frac{T}{\Delta} - \mu T)$, and

$$\mathbb{E} \left[d^\Delta(\mathbf{z}^\Delta, \hat{\mathbf{Z}}^\Delta) \mid \mathcal{E}^c \right] = D + (1 - D)\mu\Delta,$$

which tends to D as Δ tend to zero. We have thus shown that, for small enough Δ , we can achieve the pair $(R/\Delta, D)$ on \mathbf{z}^Δ using random coding whenever $R > -\lambda \log D$. Consequently, if $R > -\lambda \log D$ then we can also use random coding to achieve (R, D) on the continuous-time point pattern \mathbf{z} . ■

We next use Theorem 3 to prove Theorem 2.

Proof of Theorem 2: It follows from Theorem 3 that, on any point process satisfying (9), the pair $(R_{\text{Pois}}(D, \lambda + \delta), D)$

is achievable with *random coding*. Further, since there is no adversary, the existence of a good random code guarantees the existence of a good deterministic code. Hence $(R_{\text{Pois}}(D, \lambda + \delta), D)$ is also achievable on this process with deterministic coding. Theorem 2 now follows when we let δ tend to zero, because $R_{\text{Pois}}(D, \cdot)$ is a continuous function. ■

V. SOME POINTS ARE KNOWN TO THE RECONSTRUCTOR

In this section we consider a Wyner-Ziv setting for our problem. We first consider the case where \mathbf{X} is a Poisson process of intensity λ . (Later we consider an arbitrary point pattern.) Assume that each point in \mathbf{X} is known to the reconstructor independently with probability p . Also assume that the encoder does not know which points are known to the reconstructor. The encoder maps \mathbf{X} to a message in $\{1, \dots, 2^{TR}\}$, and the reconstructor produces a Lebesgue-measurable, $\{0, 1\}$ -valued signal $\hat{\mathbf{X}}$ on $[0, T]$ based on this message and the positions of the points that it knows. The achievability of a rate-distortion pair is defined in the same way as in Section III. Denote the smallest rate R for which (R, D) is achievable by $R_{\text{WZ}}(D, \lambda, p)$.

Obviously, $R_{\text{WZ}}(D, \lambda, p)$ is lower-bounded by the smallest achievable rate when the transmitter *does* know which points are known to the reconstructor. The latter rate is given by $R_{\text{Pois}}(D, (1-p)\lambda)$, where $R_{\text{Pois}}(\cdot, \cdot)$ is given by (2). Indeed, when the encoder knows which points are known to the reconstructor, it is optimal for it to describe only the remaining points, which themselves form a Poisson process of intensity $(1-p)\lambda$. The reconstructor then selects a set based on this description to cover the points unknown to it and adds to this set the points it knows. Thus,

$$R_{\text{WZ}}(D, \lambda, p) \geq R_{\text{Pois}}(D, (1-p)\lambda). \quad (13)$$

The next theorem shows that (13) holds with equality.

Theorem 4: Knowing the points at the reconstructor only is as good as knowing them also at the encoder:

$$R_{\text{WZ}}(D, \lambda, p) = R_{\text{Pois}}(D, (1-p)\lambda). \quad (14)$$

To prove Theorem 4, it remains to show that the pair $(R_{\text{Pois}}(D, (1-p)\lambda), D)$ is achievable. We shall show this as a consequence of a stronger result concerning arbitrarily varying sources.

Consider an arbitrary point pattern \mathbf{z} on $[0, T]$ chosen by an adversary. The adversary is allowed to put at most λT points in \mathbf{z} . Also, it must reveal all but at most νT points to the reconstructor, without telling the encoder which points it has revealed. The encoder and the reconstructor are allowed to use random codes, where the encoder is a random mapping from \mathbf{z} to a message in $\{1, \dots, 2^{TR}\}$, and where the reconstructor is a random mapping from this message, together with the point pattern that it knows, to a $\{0, 1\}$ -valued, Lebesgue-measurable signal $\hat{\mathbf{z}}$. The distortion $d(\mathbf{z}, \hat{\mathbf{z}})$ is defined as in (1).

Theorem 5: Against an adversary who puts at most λT points on $[0, T]$ and reveals all but at most νT points to the reconstructor, the rate-distortion pair $(R_{\text{Pois}}(D, \nu), D)$ is achievable with random coding.

Proof: The case $D \geq 1$ is trivial, so we shall only consider the case where $D \in (0, 1)$. The encoder and the reconstructor first use Scheme 1 as in Section III to reduce the point pattern \mathbf{z} to a $\{0, 1\}$ -valued vector \mathbf{z}^Δ of length $\frac{T}{\Delta}$. Define

$$\mu \triangleq T^{-1} \sum_{i=1}^{T/\Delta} z_i^\Delta,$$

and note that, by assumption, $\mu \leq \lambda$. If $\mu \leq \nu$, then we can ignore the reconstructor's side-information and use the random code of Theorem 3. Henceforth we assume $\mu > \nu$.

Denote by \mathbf{s} the point pattern known to the reconstructor and by \mathbf{s}^Δ the vector obtained from \mathbf{s} through the discretization in time of Scheme 1. Since there are at most νT points that are unknown to the reconstructor,

$$\sum_{i=1}^{T/\Delta} s_i^\Delta \geq (\mu - \nu)T. \quad (15)$$

The encoder conveys the value of μT to the receiver using bits. Since μT is an integer between 0 and λT , the number of bits per second needed to describe it tends to zero as T tends to infinity.

Next, the encoder and the reconstructor randomly generate $2^{T(R+\tilde{R})}$ independent codewords

$$\hat{\mathbf{z}}_{m,l}^\Delta, \quad m \in \{1, \dots, 2^{TR}\}, \quad l \in \{1, \dots, 2^{T\tilde{R}}\},$$

where each codeword is generated IID $\text{Ber}(D)$.

To describe \mathbf{z}^Δ , the encoder looks for a codeword $\hat{\mathbf{z}}_{m,l}^\Delta$ such that

$$\hat{z}_{m,l,i}^\Delta = 1 \text{ whenever } z_i^\Delta = 1. \quad (16)$$

If it finds one or more such codewords, it sends the index m of the first one; otherwise it tells the reconstructor to produce the all-one sequence.

When the reconstructor receives the index m , it looks for an index $\tilde{l} \in \{1, \dots, 2^{T\tilde{R}}\}$ such that

$$\hat{z}_{m,\tilde{l},i}^\Delta = 1 \text{ whenever } s_i^\Delta = 1. \quad (17)$$

If there is only one such codeword, it produces it as the reconstruction; if there are more than one such codewords, it produces the all-one sequence.

To analyze the expected distortion for \mathbf{z}^Δ over this random code, first consider the event that the encoder cannot find a codeword satisfying (16). Note that the probability that a randomly generated codeword satisfies (16) is $D^{\mu T}$, so the probability of this event tends to zero as T tends to infinity provided that

$$R + \tilde{R} > -\mu \log D. \quad (18)$$

Next consider the event that the reconstructor finds more than one \tilde{l} satisfying (17). The probability that a randomly generated codeword satisfies (17) is $D^{\sum_{i=1}^{T/\Delta} s_i^\Delta}$. Consequently, by (15) the probability of this event tends to zero as T tends to infinity provided

$$\tilde{R} < -(\mu - \nu) \log D. \quad (19)$$

Finally, if the encoder finds a codeword satisfying (16) and the reconstructor finds only one codeword satisfying (17), then the two codewords must be the same. Following the same calculations as in the proof of Theorem 3, the expected distortion in this case tends to D as Δ tends to zero.

Combining (18) and (19), we can make the expected distortion arbitrarily close to D as $T \rightarrow \infty$ if

$$R > -\nu \log D.$$

Proof of Theorem 4: The claim follows from (13), Theorem 5, and the Law of Large Numbers. ■

APPENDIX

In this appendix we prove Lemma 1. Given any rate-distortion code with 2^{TR} codewords $\hat{\mathbf{x}}_m$, $m \in \{1, \dots, 2^{TR}\}$ that achieves expected distortion D , we shall construct a new code that can be constructed through Scheme 1, that contains $(2^{TR} + 1)$ codewords, and that achieves an expected distortion that is arbitrarily close to D .

Denote the codewords of our new code by $\hat{\mathbf{w}}_m$, where $m \in \{1, \dots, 2^{TR} + 1\}$. We choose the last codeword to be the constant 1. We next describe our choices of the other codewords. For every $\epsilon > 0$ and every $\hat{\mathbf{x}}_m$, we can approximate the set $\{t: \hat{x}_m(t) = 1\}$ by a set \mathcal{A}_m that is equal to a finite, say N_m , union of open intervals. More specifically,

$$\mu(\hat{x}_m^{-1}(1) \Delta \mathcal{A}_m) \leq 2^{-TR}\epsilon, \quad (20)$$

where Δ denotes the symmetric difference between two sets (see, e.g., [15, Chapter 3, Proposition 15]). Define

$$\mathcal{B} \triangleq \bigcup_{m=1}^{2^{TR}} (\hat{x}_m^{-1}(1) \setminus \mathcal{A}_m),$$

and note that by (20)

$$\mu(\mathcal{B}) \leq \epsilon. \quad (21)$$

For each \mathcal{A}_m , $m \in \{1, \dots, 2^{TR}\}$, define

$$\mathcal{T}_m \triangleq \{t \in [0, T]: ((\lceil t/\Delta \rceil - 1)\Delta, \lceil t/\Delta \rceil \Delta) \cap \mathcal{A}_m \neq \emptyset\}.$$

We now construct $\hat{\mathbf{w}}_m$, $m \in \{1, \dots, 2^{TR}\}$ as

$$\hat{\mathbf{w}}_m = \mathbf{1}_{\mathcal{T}_m},$$

where $\mathbf{1}_{\mathcal{S}}$ denotes the indicator function of the set \mathcal{S} . Note that $\mathcal{A}_m \subseteq \mathcal{T}_m = \hat{w}_m^{-1}(1)$. See Figure 2 for an illustration of this construction. Let

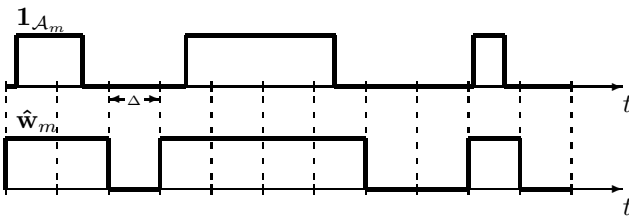


Fig. 2. Constructing $\hat{\mathbf{w}}_m$ from \mathcal{A}_m .

$$N \triangleq \max_{m \in \{1, \dots, 2^{TR}\}} N_m.$$

It can be seen that

$$\mu(\hat{w}_m^{-1}(1)) - \mu(\mathcal{A}_m) \leq 2N\Delta, \quad m \in \{1, \dots, 2^{TR}\}. \quad (22)$$

Our encoder works as follows: if \mathbf{x} contains no point in \mathcal{B} , it maps \mathbf{x} to the same message as the given encoder; otherwise it maps \mathbf{x} to the index $(2^{TR} + 1)$ of the all-one codeword. To analyze the distortion, first consider the case where \mathbf{x} contains no point in \mathcal{B} . In this case, all points in \mathbf{x} must be covered by the selected codeword $\hat{\mathbf{w}}_m$. By (20) and (22), the difference $d(\mathbf{x}, \hat{\mathbf{w}}_m) - d(\mathbf{x}, \hat{\mathbf{x}}_m)$, if positive, can be made arbitrarily small by choosing small ϵ and Δ . Next consider the case where \mathbf{x} does contain points in \mathcal{B} . By (21), the probability that this happens can be made arbitrarily small by choosing ϵ small, therefore its contribution to the expected distortion can also be made arbitrarily small. We conclude that our code $\{\hat{\mathbf{w}}_m\}$ can achieve a distortion that is arbitrarily close to the distortion achieved by the original code $\{\hat{\mathbf{x}}_m\}$. This concludes the proof of Lemma 1.

REFERENCES

- [1] Y. Kabanov, "The capacity of a channel of the Poisson type," *Theory of Probability and Its Appl.*, vol. 23, pp. 143–147, 1978.
- [2] M. H. A. Davis, "Capacity and cutoff rate for Poisson-type channels," *IEEE Trans. Inform. Theory*, vol. 26, pp. 710–715, Nov. 1980.
- [3] A. D. Wyner, "Capacity and error exponent for the direct detection photon channel — part I," *IEEE Trans. Inform. Theory*, vol. 34, no. 6, pp. 1449–1461, Nov. 1988.
- [4] —, "Capacity and error exponent for the direct detection photon channel — part II," *IEEE Trans. Inform. Theory*, vol. 34, pp. 1462–1471, Nov. 1988.
- [5] T. M. Cover and M. Chiang, "Duality between channel capacity and rate distortion with two-sided state information," *IEEE Trans. Inform. Theory*, vol. 48, no. 6, pp. 1629–1638, June 2002.
- [6] I. Rubin, "Information rates and data-compression schemes for Poisson processes," *IEEE Trans. Inform. Theory*, vol. 20, no. 2, pp. 200–210, Mar. 1974.
- [7] R. G. Gallager, "Basic limits on protocol information in data communication networks," *IEEE Trans. Inform. Theory*, vol. 22, no. 4, pp. 385–398, July 1976.
- [8] S. Verdú, "The exponential distribution in information theory," *Problems of Information Transmission*, vol. 31, no. 1, pp. 86–95, 1996.
- [9] A. S. Bedekar, "On the information about message arrival times required for in-order decoding," in *Proc. IEEE Int. Symposium on Inf. Theory*, Washington DC, June 24–29, 2001.
- [10] T. P. Coleman, N. Kiyavash, and V. G. Subramanian, "The rate-distortion function of a Poisson process with a queueing distortion measure," in *Proceedings Data Compression Conference 2008*, Cliff Lodge, Snowbird, Utah, USA, Mar. 2008.
- [11] V. Anantharam and S. Verdú, "Bits through queues," *IEEE Trans. Inform. Theory*, vol. 42, no. 1, pp. 4–18, Jan. 1996.
- [12] A. D. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Trans. Inform. Theory*, vol. 22, no. 1, pp. 1–10, Jan. 1976.
- [13] S. Bross, A. Lapidoth, and L. Wang, "The Poisson channel with side information," in *Proceedings Forty-Seventh Allerton Conf. Comm., Contr. and Comp.*, Allerton House, Monticello, Illinois, September 30–October 2, 2009.
- [14] C. E. Shannon, "A mathematical theory of communication," *Bell System Techn. J.*, vol. 27, pp. 379–423 and 623–656, July and Oct. 1948.
- [15] H. L. Royden, *Real Analysis*, 3rd ed. Macmillan Publishing Company, 1988.