Proceedings of the International Workshop on Statistical-Mechanical Informatics
September 13–16, 2009, Kyoto, Japan

Capacity of a Single Spiking Neuron

Shiro Ikeda[1], Jonathan H. Manton[2]

[1] The Institute of Statistical Mathematics, Tokyo, 106-8569, Japan
[2] The University of Melbourne, Victoria 3010, Australia

E-mail: shiro@ism.ac.jp

Abstract. It is widely believed the neurons transmit information in the form of spikes. Since the spike patterns are known to be noisy, the neuron information channel is noisy. We have investigated the channel capacity of this "Spiking neuron channel" for both of the "temporal coding" and the "rate coding,"which are two main coding considered in the neuroscience[1, 2]. As the result, we've proved that the distribution of inputs, which achieves the channel capacity, is a discrete distribution with finite mass points for temporal and rate coding under a reasonable assumption. In this draft, we show the details of the proof.

1. Introduction

Neurons transfer information to other neurons in the form of spike trains. Although precise control of spike timing is important for reliable information transfer, a lot of studies revealed that spike patterns are noisy. When a communication channel is noisy, the rate at which information can be transmitted reliably through the channel is limited. The upper bound on the rate is the channel capacity [3]. We have studied the capacity of a single neuron [1, 2] for two types of coding, temporal and rate coding. The temporal coding uses the inter-spike intervals (ISIs) to code information while the rate coding uses the number of spikes in a fixed interval.

The channel model is deeply related to the noise of ISIs. Many works reported [4–6] that the statistical properties of ISIs are similar to the gamma distribution. We employ this model, and ISIs are modelled with a gamma distribution. The capacity is defined as the supremum of mutual information over possible input distributions. We pose a natural assumption on the input distributions, and under the assumption, we proved the capacity of each coding is achieved by a discrete distribution which has only finite mass points[1, 2]. Although the proof for each coding shares the steps with other studies of information theory [7–10]. the neuron channel is special and we have to prove each step. In this draft, we provide the details of the proof. Our result shows that the information is maximally transmitted through a single neuron when the inputs to the neuron have only a fixed number of modes.

The problem is formulated mathematically in section 2 and the discreteness for each coding is proved in section 3. Section 4 concludes the paper.

2. Single Neuron Channel

2.1. Communication Channel and Capacity

Let X be the input to a noisy channel and Y be the output. In the following, we assume $X \in \mathcal{X} \subseteq \mathbb{R}$ is a one-dimensional stochastic variable and let $F(\cdot)$ be a cumulative distribution function of X. Communication channel is defined as a stochastic model described as $p(y|x)$ and the mutual information is defined as

$$I(X;Y) = \int_{x\in\mathcal{X}}\int_{y\in\mathcal{Y}} p(y|x)\log\frac{p(y|x)}{p(y)}d\mu(y)dF(x), \quad \text{where} \;\; p(y) = \int_{x\in\mathcal{X}} p(y|x)dF(x). \tag{1}$$

Here, $\mu(y)$ denotes the measure of $y \in \mathcal{Y}$. Since the channel is defined as $p(y|x)$, $I(X;Y)$ is a functional of $F(\cdot)$ and we denote it as $I(F)$.

Let $\mathcal{F}$ be the set of cumulative distribution functions of X. The channel capacity is defined as

$$C = \sup_{F\in\mathcal{F}} I(F). \tag{2}$$

For a noisy channel, one interesting fundamental problem is to compute the capacity C. Another interesting problem is to obtain the distribution, if it exists, which achieves the capacity.

2.2. Single Neuron: Channel and Coding

It has been reported that a gamma distribution is a suitable model to describe the stochastic nature of ISIs [4, 6]. The gamma distribution has two parameters which are the shape parameter κ and the scale parameter θ. From some studies, κ of individual neuron appears to be constant (the value of κ may depends on the type of neuron), while θ changes dynamically over time. Let T denote an ISI, which is a stochastic variable following a gamma distribution, that is, $T \sim \Gamma(\kappa, \theta)$, where $\kappa > 0$ and $\theta > 0$ are the shape and the scale parameter, respectively. We assume κ of each neuron is fixed and known and the distribution of each ISI is independent.

Under these assumption, the scale parameter θ is the only variable parameter which plays the role of input, that is, X in §.2.1. The density function of t is

$$p(t|\theta;\kappa) = \Big(\frac{t^{\kappa-1}}{\theta^{\kappa}}\Big)\frac{\exp[-t/\theta]}{\Gamma(\kappa)}, \quad \kappa, \theta > 0,\; t \geq 0,$$

where we denote it as $p(t|\theta;\kappa)$ to show θ is a stochastic variable and κ is a parameter. The gamma distribution is an exponential family,

$$p(t|\theta;\kappa) = \exp\Big[-\frac{1}{\theta}t + (\kappa-1)\log t - \log\Gamma(\kappa) - \kappa\log\theta\Big]. \tag{3}$$

The conditional entropy becomes

$$H(T\,|\,\theta\,;\kappa) = -\int_0^\infty p(t|\theta;\kappa)\log p(t|\theta;\kappa)dt = \kappa - (\kappa-1)\psi(\kappa) + \log\Gamma(\kappa) + \log\theta,$$

where $\psi(\cdot)$ is the digamma function defined as $\psi(x) = \Gamma'(x)/\Gamma(x)$ for $x > 0$. Next, let us consider the family of all the possible distributions of input θ. Noting that ISI is positive and is not infinite if the neuron is active, it is natural to assume that the average ISI, which depends on θ and κ, is limited between a_0 and b_0, that is,

$$a_0 \leq \overline{T} = \kappa\theta \leq b_0, \quad \text{where} \;\; 0 < a_0 < b_0 < \infty.$$

Thus, θ is bounded in $\Theta(\kappa) = \{\theta \mid a(\kappa) \leq \theta \leq b(\kappa)\}$, where $a(\kappa)$ and $b(\kappa)$ are defined as

$$a(\kappa) = a_0/\kappa, \quad b(\kappa) = b_0/\kappa.$$

Let us define $F(\theta)$ as the cumulative distribution function of θ and $\mathcal{F}$ as the set of all possible $F(\theta)$, that is

$$\mathcal{F} = \big\{F : \mathbb{R} \to [0,1] \mid F(\theta) = 0, (\forall\theta < a),\ F(\theta) = 1, (\forall\theta > b)\big\}. \tag{4}$$

Next, let us consider what is Y, that is, "the output of the channel" of a neuron communication channel. There are mainly two different ideas in neuroscience. One is that Y is ISI, T, itself. This is called "temporal coding." The other is that Y is the rate, which is the number of spikes in fixed time intervals. This is called "rate coding"[1]. How to encode the input θ to the neuron channel depends on which coding is used. For the temporal coding, θ is fixed during the interval t while θ is fixed during Δ for the rate coding.

Temporal coding In temporal coding, received information is T. For a $F \in \mathcal{F}$, we define the marginal distribution as

$$p(t; F, \kappa) = \int_a^b p(t|\theta; \kappa) dF(\theta) \tag{5}$$

where $p(t|\theta;\kappa)$ is defined in eq.(3). The mutual information of T and θ is defined as

$$I_T(F) = \int_a^b i_T(\theta; F) dF(\theta), \quad \text{where} \quad i_T(\theta; F) = \int_0^\infty p(t|\theta;\kappa) \log \frac{p(t|\theta;\kappa)}{p(t;F,\kappa)} dt. \tag{6}$$

Let us define $g(t; F, \kappa)$ and rewrite $p(t; F, \kappa)$ as follows

$$g(t; F, \kappa) = \int_a^b \frac{\exp[-t/\theta]}{\theta^\kappa} dF(\theta), \quad p(t; F, \kappa) = \frac{t^{\kappa-1}}{\Gamma(\kappa)} g(t; F, \kappa). \tag{7}$$

The mutual information $I_T(F)$ is rewritten as follows

$$I_T(F) = h_T(F;\kappa) - \kappa\ h_{T|\theta}(F;\kappa) - \kappa, \quad \text{where}$$
$$h_T(F;\kappa) = -\int_0^\infty p(t;F,\kappa) \log g(t;F,\kappa) dt, \quad h_{T|\theta}(F;\kappa) = \int_a^b \log\theta dF(\theta).$$

Hence, the capacity per channel use or equivalently per spike is defined as

$$C_T = \sup_{F\in\mathcal{F}} I_T(F) = \sup_{F\in\mathcal{F}} \big\langle h_T(F;\kappa) - \kappa\, h_{T|\theta}(F;\kappa)\big\rangle - \kappa.$$

The capacity C_T and the distribution which achieves C_T will be studied in the next section.

Rate coding In rate coding, a time window is set and the spikes in the interval is counted. Let us denote the interval and the rate as Δ and R, respectively, and define the distribution of R as $p(r|\theta;\kappa,\Delta)$. The form of the distribution of R is shown in the following lemma.

Lemma 1. *The distribution $p(r|\theta;\kappa,\Delta)$ [2] has the following form*

$$p(r|\theta;\kappa,\Delta) = P\big(r\kappa,\ \Delta/\theta\big) - P\big((r+1)\kappa,\ \Delta/\theta\big), \quad r \in \mathbb{Z}^* \ \textit{(nonnegative integers)}, \tag{8}$$

where $P(\alpha, x)$ is the regularized incomplete gamma function

$$P(0,x) = 1, \quad P(\alpha,x) = \frac{1}{\Gamma(\alpha)} \int_0^x t^{\alpha-1} e^{-t} dt, \quad \text{for} \quad \alpha, x > 0.$$

[1] It seems the term "modulation" is more suitable to the above definition. However, we follow the standard usage of the neuroscience community

[2] The same distribution is discussed in [11].

For an $F \in \mathcal{F}$, let us define the following marginal distribution $p(r; F, \kappa, \Delta)$

$$p(r; F, \kappa, \Delta) = \int_a^b p(r|\theta; \kappa, \Delta) dF(\theta).$$

The mutual information of R and θ is defined as

$$I_R(F) = \int_a^b i_R(\theta, F) dF(\theta), \quad \text{where} \quad i_R(\theta, F) = \sum_{r=0}^{\infty} p(r|\theta; \kappa, \Delta) \log \frac{p(r|\theta; \kappa, \Delta)}{p(r; F, \kappa, \Delta)}. \tag{9}$$

Hence, the capacity per channel use or equivalently per Δ is defined as

$$C_R = \sup_{F \in \mathcal{F}} I_R(F).$$

The capacity C_R and the distribution which achieves C_R will be studied in the next section.

3. Theoretical Studies

The cumulative distribution $F \in \mathcal{F}$ is a right-continuous non-decreasing function on a interval Θ. Thus, θ can be a discrete or continuous random variable over Θ. In this section, the capacity achieving distribution of a single neuron channel is proved to be a discrete distribution with finite mass points, for both temporal and rate coding.

For some channels, the capacity achieving distributions have been shown to be discrete under some conditions [7–9, 12, 13]. The neuron channel with temporal coding is different from those and the proof must be provided independently.

3.1. Steps to Prove the Discreteness of the Capacity Achieving Distribution

We first show the common steps of the proof for the discreteness of the capacity achieving distributions. In the following, results of optimization theory and probability theory will be used. Suppose X is a normed linear space. In optimization theory, the space of all bounded linear functionals of X is called the normed dual of X and is denoted X^*. The weak* convergence is defined as follows.

Definition 1. *A sequence $\{x_n^*\}$ in X^* is said to converge weak* to the element x^* if for every $x \in X$, $x_n^*(x) \to x^*(x)$. In this case we write $x_n^*(x) \xrightarrow{w^*} x^*(x)$ ([14], 5.10).*

If X is the real normed linear space of all bounded continuous functions on $\mathbb{R}$, X^* includes the set of all probability measures, and it is clear that "weak convergence" of probability measures is "weak* convergence" on X^*. The following theorem is used to prove the existence and the uniqueness of the capacity achieving distribution.

Theorem 1. *Let J be a weak* continuous real-valued functional on a weak* compact subset S of X^*. Then J is bounded on S and achieves its maximum on S. If S is convex and J is strictly concave, then the following maximum is achieved by a unique x^* in $\mathcal{S}$.*

$$C = \max_{x^* \in \mathcal{S}} J(x^*).$$

Proof. See [14], 5.10, [9] and [12]. □

From the above discussion, $\mathcal{F}$ in eq.(4) is a subset of X^*. It is clear that $\mathcal{F}$ is convex. Thus, if $\mathcal{F}$ is weak* compact and $I_T(F)$ (or $I_R(F)$) is a weak* continuous function on $F \in \mathcal{F}$ and strictly concave in $\mathcal{F}$, the capacity is achieved by a unique distribution F_0 in $\mathcal{F}$. This is the first step of the proof. The following proposition states $\mathcal{F}$ is compact.

Proposition 1. *$\mathcal{F}$ in eq.(4) is compact in the Lévy metric topology.*

Proof. For the proof of compactness, see [7] (proof of proposition 1), the proof is a direct application of the Helly's compactness theorem ([15], section X). $\square$

The Kuhn-Tucker (K-T) condition on the mutual information is used for the next step of the proof. Before showing the condition, let us define the weak differentiability.

Definition 2. *Let J be a function on a convex set $\mathcal{F}$. Let F_0 be a fixed element of $\mathcal{F}$, and $\eta \in [0,1]$. Suppose there exists a map $J'_{F_0} : \mathcal{F} \to \mathbb{R}$ such that*

$$J'_{F_0}(F) = \lim_{\eta \downarrow 0} \frac{J((1-\eta)F_0 + \eta F) - J(F_0)}{\eta}, \qquad F \in \mathcal{F}.$$

Then J is said to be weakly differentiable in $\mathcal{F}$ at F_0 and $J'_{F_0}(F)$ is the weak derivative in $\mathcal{F}$ at F_0. If J is weakly differentiable in $\mathcal{F}$ at F_0 for all $F \in \mathcal{F}$, J is weakly differentiable in $\mathcal{F}$.

And the K-T condition is described as follows,

Proposition 2. *Assume J is a weakly differentiable, concave functional on a convex set $\mathcal{F}$. If J achieves its maximum on $\mathcal{F}$ at F_0, then a necessary and sufficient condition for F_0 to attain the maximum is to satisfy the following inequality for all $F \in \mathcal{F}$*

$$J'_{F_0}(F) \leq 0.$$

Proof. See Proposition 1 in [7]. $\square$

If $I_T(F)$ (or $I_R(F)$) is weakly differentiable, the K-T condition is derived with the theorem. Finally, the discreteness is proved by deriving a contradiction based on the K-T condition and the assumption that F_0 has infinite mass points as its support. In order to show the discreteness of the capacity achieving distribution for temporal and rate codings, the following properties must be shown.

(i) $I_T(F)$ and $I_R(F)$ are weak* continuous on $\mathcal{F}$ and strictly concave.
(ii) $I_T(F)$ and $I_R(F)$ are weakly differentiable.

Then, the K-T condition is derived and the discreteness will be checked.

3.2. Discreteness of the Capacity Achieving Distribution for Temporal Coding
In this subsection, the capacity achieving distribution for temporal coding is shown to be a discrete distribution with a finite number of points. Let us start with the following lemma.

Lemma 2. *$I_T(F)$ in eq.(6) is a weak* continuous function on $F \in \mathcal{F}$ and strictly concave in $\mathcal{F}$.*

Proof. $I_T(F)$ is weak* continuous if the following relation holds,

$$F_n \overset{w^*}{\to} F \implies I_T(F_n) \to I_T(F), \tag{10}$$

since $I_T(F) = h_T(F;\kappa) - \kappa\, h_{T|\theta}(F;\kappa) - \kappa$, more precisely,

$$F_n \overset{w^*}{\to} F \implies h_T(F_n;\kappa) \to h_T(F;\kappa) \;\text{ and }\; h_{T|\theta}(F_n;\kappa) \to h_{T|\theta}(F;\kappa).$$

$h_{T|\theta}(F_n;\kappa) \to h_{T|\theta}(F;\kappa)$ holds since $h_{T|\theta}(F_n;\kappa) = \int_a^b \log\theta dF_n(\theta)$ and $\log\theta$ is a bounded continuous function for $\theta \in \Theta$.

Next, we show the following equalities

$$\lim_n \; h_T(F_n;\kappa) = -\lim_n \int_0^\infty p(t;F_n,\kappa)\log g(t;F_n,\kappa)dt = -\int_0^\infty \lim_n \; p(t;F_n,\kappa)\log g(t;F_n,\kappa)dt \quad (11)$$

$$= -\int_0^\infty p(t;F,\kappa)\log g(t;F,\kappa)dt = h_T(F;\kappa) \quad (12)$$

The interchange of integral and limit in eq.(11) is justified as follows. From eqs.(5) and (7), $p(t;F,\kappa)$ and $g(t;F,\kappa)$ are bounded as follows.

$$\frac{t^{\kappa-1}}{b^\kappa}\frac{e^{-t/a}}{\Gamma(\kappa)} < p(t;F_n,\kappa) < \frac{t^{\kappa-1}}{a^\kappa}\frac{e^{-t/b}}{\Gamma(\kappa)}, \qquad -\frac{t}{a} - \kappa\log b < \log g(t;F_n,\kappa) < -\frac{t}{b} - \kappa\log a, \quad (13)$$

$p(t;F_n,\kappa)\log g(t;F_n,\kappa)$ is bounded for all F_n with finite A_1 and A_2 as follows

$$\big|p(t;F_n,\kappa)\log g(t;F_n,\kappa)\big| < A_1\, t^{\kappa-1}e^{-t/b} + A_2\, t^\kappa e^{-t/b}. \quad (14)$$

RHS is integrable as

$$\int_0^\infty \Big[A_1\, t^{\kappa-1}e^{-t/b} + A_2\, t^\kappa e^{-t/b}\Big]dt = \Gamma(\kappa)b^\kappa(A_1 + \kappa b A_2).$$

Since eq.(14) is bounded from above with an integrable function, eq.(11) is justified by the Lebesgue dominated convergence theorem. Since $p(t|\theta;\kappa)$ and $\exp[-t/\theta]/\theta^\kappa$ are continuous bounded functions of $\theta \in \Theta$, $p(t;F,\kappa)$ and $g(t;F,\kappa)$ are continuous function on F, hence $p(t;F_n,\kappa)\log g(t;F_n,\kappa)$ is also continuous for every $F_n \in \mathcal{F}$. These arguments justify eq.(12) and eq.(10) is justified.

$I_T(F)$ is also strictly concave following the proof of lemma 2 in [9]. □

Lemma 2 and theorem 1 imply the capacity for temporal coding is achieved by a unique distribution in $\mathcal{F}$. In order to show it is a discrete distribution, the following lemma and corollary are used.

Lemma 3. *$I_T(F)$ in eq.(6) is weakly differentiable in $\mathcal{F}$. The weak derivative at $F_0 \in \mathcal{F}$ has the form*

$$I'_{T,F_0}(F) = \int_a^b i_T(\theta;F_0)dF - I_T(F_0). \quad (15)$$

Proof. Let us define F_η and rewrite $i_T(\theta;F)$ in eq.(6) as follows

$$F_\eta = (1-\eta)F_0 + \eta F, \quad i_T(\theta;F) = -\kappa\log\theta - \int_0^\infty p(t|\theta;\kappa)\log g(t;F,\kappa)dt.$$

Then

$$I_T(F_\eta) - I_T(F_0) = \int_a^b i_T(\theta;F_\eta)dF_\eta - \int_a^b i_T(\theta;F_0)dF_0$$

$$= \eta\left[\int_a^b i_T(\theta;F_\eta)dF - \int_a^b i_T(\theta;F_\eta)dF_0\right] \quad (16)$$

$$+ \int_a^b \big[i_T(\theta;F_\eta) - i_T(\theta;F_0)\big]dF_0. \quad (17)$$

The weak derivative of $I_T(F)$ at F_0 is defined as $I'_{T,F_0}(F) = \lim_{\eta\downarrow 0}\left(I_T(F_\eta) - I_T(F_0)\right)/\eta$. By dividing the term in eq.(16) with η and by taking $\eta \downarrow 0$, it becomes

$$\int_a^b i_T(\theta; F_0)dF - \int_a^b i_T(\theta; F_0)dF_0 = \int_a^b i_T(\theta; F_0)dF - I_T(F_0).$$

By noting $g(t; F_\eta, \kappa) = (1-\eta)\, g(t; F_0, \kappa) + \eta\, g(t; F, \kappa)$, the term in eq.(17) becomes 0. This proves $I_T(F)$ is weakly differentiable. □

Corollary 1. *Let E_0 denote the points of increase of F_0 on $\theta \in \Theta$. F_0 is optimal if and only if*

$$\begin{aligned} i_T(\theta; F_0) &\le I_T(F_0), \quad \forall \theta \in \Theta \\ i_T(\theta; F_0) &= I_T(F_0), \quad \forall \theta \in E_0. \end{aligned} \tag{18}$$

Proof. This is proved following the same steps in [7] (Corollary 1) with eq.(15). □

The main result of this subsection is summarized in the following theorem.

Theorem 2. *Under the constraint $\theta \in \Theta$, the channel capacity of a single neuron channel with temporal coding is achieved by a discrete distribution with a finite number of mass points.*

Proof. The extension of $i_T(\theta; F_0)$ to the complex plain z is analytic for $\operatorname{Re} z > 0$

$$i_T(z; F_0) = -\kappa \log z - \int_0^\infty p(t|z;\kappa) g(t; F_0, \kappa)dt.$$

If E_0 in corollary 1 has infinite points, since Θ is bounded and closed, E_0 has a limit point. Hence, from corollary 1, the identity theorem implies $i_T(z; F_0) = I_T(F_0)$ for the region $\operatorname{Re} z > 0$. This region includes positive real line $\mathbb{R}^+$ and

$$-\int_0^\infty p(t|\theta;\kappa) \log g(t; F_0, \kappa)dt = \kappa \log\theta + I_T(F_0), \quad \theta \in \mathbb{R}^+ \tag{19}$$

is implied. The LHS is bound as follows (eq.(13)).

$$\frac{1}{b}\int_0^\infty t\, p(t|\theta;\kappa)dt + \kappa\log a \le -\int_0^\infty p(t|\theta;\kappa)\log g(t;F_0,\kappa)dt \le \frac{1}{a}\int_0^\infty t\, p(t|\theta;\kappa)dt + \kappa\log b. \tag{20}$$

Since the expectation of T w.r.t $p(t|\theta;\kappa)$ is $\kappa\theta$, eq.(20) shows the LHS of eq.(19) grows linearly with θ. Since the RHS increases only with $\log\theta$, eq.(19) cannot hold for all $\theta \in \mathbb{R}^+$. This is the contradiction and the optimal distribution has a finite number of mass points. □

3.3. Discreteness of the Capacity Achieving Distribution for Rate Coding

The capacity achieving distribution for rate coding is shown to be a discrete distribution with a finite number of points. In [8], it has been proved that the capacity achieving distribution of a Poisson channel under peak and average power constraints is a discrete distribution with a finite point of supports. Since $\theta \in \Theta$ is a peak constraint, this directly proves the case $\kappa = 1$. For $\kappa \ne 1$ further study is needed. First we the following proposition

Proposition 3. *The expectation of R with respect to $p(r|\theta;\kappa,\Delta)$ is finite.*

Proof of proposition 3. The expectation of R is

$$\overline{R}_{\kappa,\Delta_\theta} = \sum_{r=0}^{\infty} r\ p(r|\theta;\kappa,\Delta) = \sum_{r=1}^{\infty} P(r\kappa,\Delta_\theta) = e^{-\Delta_\theta}\sum_{r=1}^{\infty}\sum_{i=0}^{\infty}\frac{\Delta_\theta^{r\kappa+i}}{\Gamma(r\kappa+i+1)}.$$

Since $P(\alpha,x)$ is a strictly decreasing function of α for $\alpha>0$, $x>0$, if $\kappa\geq 1$

$$P(r\kappa,\Delta_\theta)\leq P(r\lfloor\kappa\rfloor,\Delta_\theta),\quad r\in\mathbb{Z}^+,$$

Thus, the upper bound is given as

$$\overline{R}_{\kappa,\Delta_\theta}\leq e^{-\Delta_\theta}\sum_{r=1}^{\infty}\sum_{i=0}^{\infty}\frac{\Delta_\theta^{r\lfloor\kappa\rfloor+i}}{\Gamma(r\lfloor\kappa\rfloor+i+1)}\leq e^{-\Delta_\theta}\sum_{r=1}^{\infty}\sum_{i=0}^{\infty}\frac{\Delta_\theta^{r+i}}{\Gamma(r+i+1)} = \overline{R}_{1,\Delta_\theta} = \Delta_\theta,$$

where $R_{1,\Delta_\theta}=\Delta_\theta$ holds from the fact that $p(r|\theta;1,\Delta_\theta)$ is a Poisson distribution. For $\kappa<1$, $P(r\kappa,\Delta_\theta)\leq P(\lfloor r\kappa\rfloor,\Delta_\theta)$ holds and $\overline{R}_{\kappa,\Delta_\theta}$ is bounded as follows

$$\overline{R}_{\kappa,\Delta_\theta}<e^{-\Delta_\theta}\sum_{r=1}^{\infty}\sum_{i=0}^{\infty}\frac{\Delta_\theta^{\lfloor r\kappa\rfloor+i}}{\Gamma(\lfloor r\kappa\rfloor+i+1)}\leq\lceil\frac{1}{\kappa}\rceil e^{-\Delta_\theta}\sum_{r=1}^{\infty}\sum_{i=0}^{\infty}\frac{\Delta_\theta^{r+i}}{\Gamma(r+i+1)}+\lfloor\frac{1}{\kappa}\rfloor\leq\lceil\frac{1}{\kappa}\rceil(\Delta_\theta+1).$$

□

Next, we show the following lemma.

Lemma 4. *$I_R(F)$ is a weak* continuous function on $F\in\mathcal{F}$ and strictly concave in $\mathcal{F}$.*

Proof. $I_R(F)$ is weak* continuous if the following relation holds,

$$F_n\stackrel{w^*}{\to}F\implies I_R(F_n)\to I_R(F),\tag{21}$$

From the definitions of $I_R(F)$ and $i_R(\theta,F)$ in eq.(9),

$$I_R(F)=\int_a^b i_R(\theta,F)dF(\theta).$$

Since $i_R(\theta,F)$ is a positive continuous function of θ, if it is bounded from above, this is justified from the Helly-Bray theorem. It will be shown separately for $\kappa\geq 1$ and $\kappa<1$.

For $\kappa\geq 1$**:** Since $P(\alpha,\Delta_\theta)$ is a decreasing function of a, the following inequality holds.

$$\begin{aligned}P(r\kappa,\Delta_\theta)-P(r\kappa+\lfloor\kappa\rfloor,\Delta_\theta)\leq p(r|\theta;\kappa,\Delta)\leq P(r\kappa,\Delta_\theta)-P(r\kappa+\lceil\kappa\rceil,\Delta_\theta)\\ e^{-\Delta_\theta}\sum_{i=0}^{\lfloor\kappa\rfloor-1}\frac{\Delta_\theta^{r\kappa+i}}{\Gamma(r\kappa+i+1)}\leq p(r|\theta;\kappa,\Delta)\leq e^{-\Delta_\theta}\sum_{i=0}^{\lceil\kappa\rceil-1}\frac{\Delta_\theta^{r\kappa+i}}{\Gamma(r\kappa+i+1)}.\end{aligned}\tag{22}$$

With the above equation, $p(r;F,\kappa,\Delta)$ is bounded from below as follows

$$p(r;F,\kappa,\Delta)=\int_a^b p(r|\theta;\kappa,\Delta)dF(\theta)>e^{-\Delta_M}\sum_{i=0}^{\lfloor\kappa\rfloor-1}\frac{\Delta_m^{r\kappa+i}}{\Gamma(r\kappa+i+1)}$$

where $\Delta_m = \Delta/b$ and $\Delta_M = \Delta/a$ are the minimum and the maximum of Δ_θ, respectively. Thus,

$$\frac{p(r|\theta;\kappa,\Delta)}{p(r;F,\kappa,\Delta)} < e^{\Delta_M-\Delta_\theta} \frac{\sum_{i=0}^{\lceil\kappa\rceil-1} \frac{\Delta_\theta^{r\kappa+i}}{\Gamma(r\kappa+i+1)}}{\sum_{i=0}^{\lfloor\kappa\rfloor-1} \frac{\Delta_m^{r\kappa+i}}{\Gamma(r\kappa+i+1)}} < B\, e^{\Delta_M-\Delta_\theta}\Big(\frac{\Delta_\theta}{\Delta_m}\Big)^{r\kappa},$$

where B is the following upper bound

$$\frac{\sum_{i=0}^{\lceil\kappa\rceil-1} \frac{\Delta_\theta^{i}}{\Gamma(r\kappa+i+1)}}{\sum_{i=0}^{\lfloor\kappa\rfloor-1} \frac{\Delta_m^{i}}{\Gamma(r\kappa+i+1)}} = \frac{1+\sum_{i=1}^{\lceil\kappa\rceil-1} \Delta_\theta^{i}\frac{\Gamma(r\kappa+1)}{\Gamma(r\kappa+i+1)}}{1+\sum_{i=1}^{\lfloor\kappa\rfloor-1} \Delta_m^{i}\frac{\Gamma(r\kappa+1)}{\Gamma(r\kappa+i+1)}} < 1+\sum_{i=1}^{\lceil\kappa\rceil-1}\Delta_M^i = B.$$

With the result of proposition 3, $i_R(\theta,F)$ is bounded from above

$$i_R(\theta,F) = \sum_{r=0}^{\infty} p(r|\theta;\kappa,\Delta)\log\frac{p(r|\theta;\kappa,\Delta)}{p(r;F,\kappa,\Delta)} < \kappa\Delta_\theta\log\frac{\Delta_\theta}{\Delta_m} - \Delta_\theta + \Delta_M + \log B.$$

For $\kappa < 1$: When $\kappa < 1$, the following relation holds.

$$p(r|\theta;\kappa,\Delta) = P(r\kappa,\Delta_\theta) - P((r+1)\kappa+1,\Delta_\theta) - \frac{e^{-\Delta_\theta}\Delta_\theta^{(r+1)\kappa}}{\Gamma((r+1)\kappa+1)}.$$

Since $P(\alpha,x)$ is a decreasing function, the following relation holds

$$P(r\kappa,\Delta_\theta) - P(r\kappa+1,\Delta_\theta) - \frac{e^{-\Delta_\theta}\Delta_\theta^{(r+1)\kappa}}{\Gamma((r+1)\kappa+1)} < p(r|\theta;\kappa,\Delta) < P(r\kappa,\Delta_\theta) - P(r\kappa+1,\Delta_\theta)$$
$$\frac{e^{-\Delta_\theta}\Delta_\theta^{r\kappa}}{\Gamma(r\kappa+1)}\Big(1-\Delta_\theta^{\kappa}\frac{\Gamma(r\kappa+1)}{\Gamma(r\kappa+\kappa+1)}\Big) < p(r|\theta;\kappa,\Delta) < \frac{e^{-\Delta_\theta}\Delta_\theta^{r\kappa}}{\Gamma(r\kappa+1)}. \tag{23}$$

Above equation gives the following bound of $p(r;F,\kappa,\Delta)$.

$$p(r;F,\kappa,\Delta) = \int_a^b p(r|\theta;\kappa,\Delta)dF(\theta) > \frac{e^{-\Delta_M}\Delta_m^{r\kappa}}{\Gamma(r\kappa+1)}\Big(1-\Delta_M^{\kappa}\frac{\Gamma(r\kappa+1)}{\Gamma(r\kappa+\kappa+1)}\Big).$$

From the property of the gamma-function, $\Gamma(r\kappa+1)/\Gamma(r\kappa+\kappa+1)$ decreases as r increases for $r > 1/\kappa$, and there exists a finite positive integer $r_0 \geq 1/\kappa$ such that, for all $r \geq r_0$, the following inequality holds for a positive real number C_1.

$$1-\Delta_M^{\kappa}\frac{\Gamma(r\kappa+1)}{\Gamma(r\kappa+\kappa+1)} > C_1.$$

Thus,

$$\frac{p(r|\theta;\kappa,\Delta)}{p(r;F,\kappa,\Delta)} < e^{\Delta_M-\Delta_\theta}\Big(\frac{\Delta_\theta}{\Delta_m}\Big)^{r\kappa}\frac{1}{1-\Delta_M^{\kappa}\frac{\Gamma(r\kappa+1)}{\Gamma(r\kappa+\kappa+1)}} < \frac{e^{\Delta_M-\Delta_\theta}}{C_1}\Big(\frac{\Delta_\theta}{\Delta_m}\Big)^{r\kappa},$$

With the result of proposition 3,

$$S_1 = \sum_{r=r_0}^{\infty} p(r|\theta;\kappa,\Delta)\log\frac{p(r|\theta;\kappa,\Delta)}{p(r;F,\kappa,\Delta)} < \kappa\lceil\frac{1}{\kappa}\rceil(\Delta_\theta+1)\log\frac{\Delta_\theta}{\Delta_m} - \Delta_\theta + \Delta_M - \log C_1,$$

where S_1 is finite. It can be shown that there exists a real number $C_2 > 0$, s.t., $p(r|\theta;\kappa,\Delta) > C_2$ for all $\theta \in \Theta$, $r \in \{0,\cdots,r_0-1\}$ and the following sum is finite,

$$S_2 = \sum_{r=0}^{r_0-1} p(r|\theta;\kappa,\Delta)\log\frac{p(r|\theta;\kappa,\Delta)}{p(r;F,\kappa,\Delta)}.$$

Thus $i_R(\theta,F) = S_1 + S_2$ is bounded from above.

The concavity of $I_R(F)$ can be proved as in [9]. The proof for the strict concavity follows the proof in §7.2 of [8], which is an application of Carleman's theorem [16]. □

Lemma 4 and theorem 1 imply the capacity for rate coding is achieved by a unique F in $\mathcal{F}$.

Lemma 5. *$I_R(F)$ in eq.(9) is weakly differentiable in $\mathcal{F}$. The weak derivative at $F_0 \in \mathcal{F}$ has the form*

$$I'_{R,F_0}(F) = \int_a^b i_R(\theta;F_0)dF - I_R(F_0). \tag{24}$$

Proof. The proof is identical to the proof of lemma 3. □

Corollary 2. *Let E_0 denote the points of increase of F_0 on $\theta \in \Theta$. F_0 is optimal if and only if*

$$\begin{aligned} i_R(\theta;F_0) &\le I_R(F_0), \quad \forall\theta\in\Theta \\ i_R(\theta;F_0) &= I_R(F_0), \quad \forall\theta\in E_0. \end{aligned} \tag{25}$$

Proof. This is proved following the same steps in [7] (Corollary 1) with eq.(24). □

Finally, we prove the capacity achieving distribution is a discrete distribution with a finite number of mass points. We start with the following proposition and corollary.

Proposition 4. *As $x \to \infty$ ($x \in \mathbb{R}^+$), the following equation holds*

$$\lim_{x\to\infty}\frac{\sum_{r=1}^{\infty}P(rm,x)}{x} = \frac{1}{m}, \quad m \in \mathbb{Z}^+. \tag{26}$$

Proof of proposition 4. From proposition 3, $\sum_{r=1}^{\infty}P(rm,x)$ is bounded from above with a linear function of x. Let us define the sum as $S_m(x)$.

$$S_m(x) = \sum_{r=1}^{\infty}P(rm,x) = e^{-x}\sum_{r=1}^{\infty}\sum_{i=0}^{\infty}\frac{x^{rm+i}}{\Gamma(rm+i+1)}.$$

Here, following relation of $P(\alpha,x)$ and the beta function is used

$$P(\alpha,x)=e^{-x}\sum_{i=0}^{\infty}\frac{x^{\alpha+i}}{\Gamma(\alpha+i+1)},\ \ B(\beta,\gamma)=\int_0^1 t^{\beta-1}(1-t)^{\gamma-1}dt=\frac{\Gamma(\beta)\Gamma(\gamma)}{\Gamma(\beta+\gamma)}, \quad \alpha,\beta,\gamma,x>0.$$

It is easily checked that

$$\left(\frac{d}{dx}+1\right)^k S_m(x) = e^{-x}\sum_{r=1}^{\infty}\sum_{i=0}^{\infty}\frac{x^{rm-k+i}}{\Gamma(rm-k+i+1)}, \quad k \in \{0,\cdots,m-1\}.$$

Thus, the following linear differential equation is derived.

$$\sum_{k=0}^{m-1}\Big(\frac{d}{dx}+1\Big)^k S_m(x) = e^{-x}\sum_{r=1}^{\infty}\sum_{i=0}^{\infty}\frac{x^{r+i}}{\Gamma(r+i+1)} = x.$$

Solving the differential equation, the general solution gives the following form of $S_m(x)$

$$S_m(x) = \frac{x}{m} + \sum_{k=1}^{m-1} C_k e^{(-1+\alpha_k)x} - \frac{m-1}{m^2}, \quad \alpha_k = \exp\Big[\frac{2\pi k\sqrt{-1}}{m}\Big]. \tag{27}$$

Since $|\mathrm{Re}\,\alpha_k| < 1$, $\mathrm{Re}(-1+\alpha_k) < 0$ holds for $k \in \{1,\cdots,m-1\}$, and $\lim_{x\to\infty} S_m(x)/x = 1/m$. □

Corollary 3. *As $\theta \downarrow 0$, the expectation of R w.r.t. $p(r|\theta;\kappa,\Delta)$ grows proportional to Δ_θ.*

Proof of corollary 3. Since the expectation $\overline{R}_{\kappa,\Delta_\theta} = \sum_{r=0}^{\infty} r\, p(r|\theta;\kappa,\Delta)$ is bounded as follows

$$\sum_{r=1}^{\infty} P(r\lceil\kappa\rceil,\Delta_\theta) \le \overline{R}_{\kappa,\Delta_\theta} = \sum_{r=1}^{\infty} P(r\kappa,\Delta_\theta) \le \sum_{r=1}^{\infty} P(r\lfloor\kappa\rfloor,\Delta_\theta).$$

From proposition 4, $\sum_{r=1}^{\infty} P(r\lceil\kappa\rceil,\Delta_\theta)$ and $\sum_{r=1}^{\infty} P(r\lfloor\kappa\rfloor,\Delta_\theta)$ grows proportional to Δ_θ. □

Finally, the following theorem shows that the we discretenss of the capacity achieving distribution for rate coding.

Theorem 3. *Under a bound constraint, the channel capacity of a single neuron channel with the rate coding is achieved by a discrete distribution with a finite number of mass points.*

Proof. The proof follows the same steps of theorem 2. The extension of $i_R(\theta;F)$ to the complex plain z is defined as

$$i_R(z;F) = \sum_{r=0}^{\infty} p(r|z;\kappa,\Delta)\log\frac{p(r|z;\kappa,\Delta)}{p(r;F,\kappa,\Delta)}, \quad p(r|z;\kappa,\Delta) = P(r\kappa,\Delta/z) - P((r+1)\kappa,\Delta/z).$$

Since $P(\alpha,z)$ and $\log z$ is analytic for $\mathrm{Re}\,z>0$, $i_R(z;F_0)$ is analytic for $\mathrm{Re}\,z>0$.

If E_0 in corollary 2 has infinite points, since Θ is bounded and closed, E_0 has a limit point and hence, from eq.(25), the identity theorem implies $i_R(z;F_0) = I_R(F_0)$ for the region $\mathrm{Re}\,z>0$. This region includes positive real line $\mathbb{R}^+$ and

$$\sum_{r=0}^{\infty} p(r|\theta;\kappa,\Delta)\log\frac{p(r|\theta;\kappa,\Delta)}{p(r;F_0,\kappa,\Delta)} = I_R(F_0), \quad \theta\in\mathbb{R}^+ \tag{28}$$

is implied. The proof is completed by deriving a contradiction for eq.(28). The contradiction is derived for $\kappa \ge 1$ and $\kappa < 1$, separately.

For $\kappa \ge 1$: From eq.(22), $p(r;F,\kappa,\Delta)$ is bounded from above as follows

$$p(r;F,\kappa,\Delta) = \int_a^b p(r|\theta;\kappa,\Delta)dF(\theta) < e^{-\Delta_m}\sum_{i=0}^{\lceil\kappa\rceil-1}\frac{\Delta_M^{r\kappa+i}}{\Gamma(r\kappa+i+1)}$$

and

$$\frac{p(r|\theta;\kappa,\Delta)}{p(r;F,\kappa,\Delta)} > e^{\Delta_m-\Delta_\theta}\frac{\sum_{i=0}^{\lfloor\kappa\rfloor-1}\frac{\Delta_\theta^{r\kappa+i}}{\Gamma(r\kappa+i+1)}}{\sum_{i=0}^{\lceil\kappa\rceil-1}\frac{\Delta_M^{r\kappa+i}}{\Gamma(r\kappa+i+1)}} > De^{\Delta_m-\Delta_\theta}\Big(\frac{\Delta_\theta}{\Delta_M}\Big)^{r\kappa},$$

where D is the following lower bound.

$$\frac{\sum_{i=0}^{\lfloor\kappa\rfloor-1}\frac{\Delta_\theta^{i}}{\Gamma(r\kappa+i+1)}}{\sum_{i=0}^{\lceil\kappa\rceil-1}\frac{\Delta_M^{i}}{\Gamma(r\kappa+i+1)}} = \frac{\sum_{i=0}^{\lfloor\kappa\rfloor-1}\Delta_\theta^{i}\frac{\Gamma(r\kappa+1)}{\Gamma(r\kappa+i+1)}}{\sum_{i=0}^{\lceil\kappa\rceil-1}\Delta_M^{i}\frac{\Gamma(r\kappa+1)}{\Gamma(r\kappa+i+1)}} > \frac{1}{1+\sum_{i=1}^{\lceil\kappa\rceil-1}\Delta_M^i} = D.$$

This shows $i_R(\theta,F)$ is bounded from below as

$$i_R(\theta,F) > \kappa\,\overline{R}_{\kappa,\Delta_\theta}\log\frac{\Delta_\theta}{\Delta_M} - \Delta_\theta + \Delta_m + \log D.$$

Since $\overline{R}_{\kappa,\Delta_\theta}$ grows with Δ_θ as $\theta\downarrow 0$, the lower bound of $i_R(\theta,F)$ grows with $\Delta_\theta\log\Delta_\theta$. Thus, $i_R(\theta,F)$ cannot be finite and constant for $\forall\theta\in\mathbb{R}^+$, which brings the contradiction.

For $\kappa<1$: From eq.(23), $p(r;F,\kappa,\Delta)$ is bounded from above as follows

$$p(r;F,\kappa,\Delta) < \frac{e^{-\Delta_m}\Delta_M^{r\kappa}}{\Gamma(r\kappa+1)} \tag{29}$$

Let us denote r as

$$r = r'K+b, \qquad \text{where} \qquad K = \lceil\frac{1}{\kappa}\rceil, \quad r'\in\mathbb{Z}^*, \quad b\in\{0,\cdots,K-1\}.$$

r' and b can be considered as stochastic variables and the following relation holds,

$$p(r|\theta;\kappa,\Delta) = p(r'K+b|\theta;\kappa,\Delta) = q(r'|\theta;\kappa,\Delta)q(b|r',\theta;\kappa,\Delta),$$
$$q(r'|\theta;\kappa,\Delta) = \sum_{b=0}^{K-1}p(r'K+b|\theta;\kappa,\Delta), \quad q(b|r',\theta;\kappa,\Delta) = \frac{p(r'K+b|\theta;\kappa,\Delta)}{q(r'|\theta;\kappa,\Delta)}.$$

Let H_R and $H_{R'}$ be the entropy of R and R', respectively, and $H_{B|R'}$ be the conditional entropy of B given R'. The following relation holds

$$H_R = H_{R'} + H_{B|R'} \le H_{R'} + \log K,$$

which is justified from $0\le H_{B|R'}\le\log K$. With this result,

$$\sum_{r=0}^{\infty}p(r|\theta;\kappa,\Delta)\log p(r|\theta;\kappa,\Delta) \ge \sum_{r'=0}^{\infty}q(r'|\theta;\kappa,\Delta)\log q(r'|\theta;\kappa,\Delta) - \log K.$$

Since $\lfloor\kappa K\rfloor = 1$ holds, the probability $q(r'|\theta;\kappa,\Delta)$ is bounded as follows,

$$q(r'|\theta;\kappa,\Delta) = P(r'K\kappa,\Delta_\theta) - P((r'+1)K\kappa,\Delta_\theta) \ge e^{-\Delta_\theta}\frac{\Delta_\theta^{r'K\kappa}}{\Gamma(r'K\kappa+1)}. \tag{30}$$

With eqs.(29) and (30),

$$\frac{q(r'|\theta;\kappa,\Delta)}{p(r;F,\kappa,\Delta)} = \frac{q(r'|\theta;\kappa,\Delta)}{p(r'K+b;F,\kappa,\Delta)} > e^{\Delta_m-\Delta_\theta}\Big(\frac{\Delta_\theta}{\Delta_M}\Big)^{r'K\kappa}\Delta_M^{-b\kappa}\frac{\Gamma((r'K+b)\kappa+1)}{\Gamma(r'K\kappa+1)}$$
$$\geq E\, e^{\Delta_m-\Delta_\theta}\Big(\frac{\Delta_\theta}{\Delta_M}\Big)^{r'K\kappa},$$

where E is the following lower bound.

$$\Delta_M^{-b\kappa}\frac{\Gamma((r'K+b)\kappa+1)}{\Gamma(r'K\kappa+1)} \geq \min\Big\{1,\ \Delta_M^{-\kappa}\Gamma(\kappa+1),\ \cdots,\ \Delta_M^{-(K-1)\kappa}\Gamma((K-1)\kappa+1)\Big\} = E.$$

This shows $i_R(\theta,F)$ is bounded from below as

$$\begin{aligned}
i_R(\theta,F) &= \sum_{r=0}^{\infty} p(r|\theta;\kappa,\Delta)\log\frac{p(r|\theta;\kappa,\Delta)}{p(r;F,\kappa,\Delta)} \\
&\geq \sum_{r'=0}^{\infty}\sum_{b=0}^{K-1} p(r'K+b|\theta;\kappa,\Delta)\log\frac{q(r'|\theta;\kappa,\Delta)}{p(r'K+b;F,\kappa,\Delta)} - \log K \\
&> \sum_{r'=0}^{\infty} q(r'|\theta;\kappa,\Delta)\log\Big(E\, e^{\Delta_m-\Delta_\theta}\Big(\frac{\Delta_\theta}{\Delta_M}\Big)^{r'K\kappa}\Big) - \log K \\
&= \Big[\sum_{r'=0}^{\infty} r'\, q(r'|\theta;\kappa,\Delta)\Big] K\kappa\log\Big(\frac{\Delta_\theta}{\Delta_M}\Big) - \Delta_\theta + \Delta_m + \log E - \log K.
\end{aligned}$$

Since $\sum_{r'=0}^{\infty} r'\, q(r'|\theta;\kappa,\Delta)$ is equivalent to $\overline{R}_{K\kappa,\Delta_\theta}$, proposition 4 shows that it grows proportional to Δ_θ as $\theta \downarrow 0$. Thus, $i_R(\theta,F)$ is lower bounded with a term which grows with $\Delta_\theta \log\Delta_\theta$ and $i_R(\theta,F)$ cannot be finite and constant for $\forall\theta \in \mathbb{R}^+$, which brings the contradiction.

□

4. Discussion and Conclusion

We considered the channel capacity and the capacity achieving distribution for a single neuron information channel. ISIs are modelled with a gamma distribution and two types of coding, temporal and rate, are considered. We have proved that the channel capacities of a single neuron with temporal and rate coding are achieved with discrete distributions[3].

However, this does not mean the capacity can be achieved biologically. Channel capacity is something similar to the maximum speed indicated in the speed meter of an automobile. Although you will never drive your vehicle with that speed, that speed tells us the potential of the automobile. Channel capacity not only provide the upper limit of the possible information transmission rate, but also describe how good the channel is.

Acknowledgements

This work was supported by the Grant-in-Aid for Scientific Research No. 18079013, MEXT, Japan.

[3] The numerical studies are shown in another paper [1].

References

[1] Ikeda S and Manton J H 2009 *Neural Computation* **21** 1714–1748

[2] Ikeda S and Manton J H 2009 Spiking neuron channel *2009 IEEE International Symposium on Information Theory* (Seoul, Korea) pp 1589–1593

[3] Shannon C E 1948 *The Bell System Technical Journal* **27** 379–423 and 623–656

[4] Baker S N and Lemon R N 2000 *J. Neurophysiol* **84** 1770–1780

[5] Stein R B 1967 *Biophysical Journal* **797-826**

[6] Shinomoto S, Shima K and Tanji J 2003 *Neural Computation* **15** 2823–2842

[7] Smith J G 1971 *Information and Control* **18** 203–219

[8] Shamai (Shitz) S 1990 *IEE Proceedings* **137** 424–430

[9] Abou-Faycal I C, Trott M D and Shamai (Shitz) S 2001 *IEEE Transactions on Information Theory* **47** 1290–1301

[10] Gursoy M C, Poor V and Verdú S 2005 *IEEE transaction on Wireless Communications* **4** 2193–2206

[11] Pawlas Z, Klevanov L B and Prokop M 2008 *Neural Computation* **20** 1325–1343

[12] Gursoy M C, Poor H V and Verdú S 2002 The capacity of the noncoherent Rician fading channel Tech. rep. Princeton University Technical Report

[13] Tchamkerten A 2004 *IEEE Transactions on Information Theory* **50** 2773–2778

[14] Luenberger D G 1969 *Optimization by Vector Space Method* (John Wiley & Sons, Inc.)

[15] Doob J 1994 *Measure Theory* (*Graduate texts in mathematics* vol 143) (Springer-Verlag)

[16] Akhiezer N 1965 *The classical moment problem* (Oliver & Boyd) translated to English by N. Kemmer