

## A NOVEL INFORMATION TRANSMISSION PROBLEM AND ITS OPTIMAL SOLUTION\*

ERIC BACH<sup>†</sup> AND JIN-YI CAI<sup>‡</sup>

**Abstract.** We propose and study a new information transmission problem motivated by today's internet. Suppose a real number needs to be transmitted in a network. This real number may represent data or control and pricing information of the network. We propose a new transmission model in which the real number is encoded using Bernoulli trials. This differs from the traditional framework of Shannon's information theory. We propose a natural criterion for the quality of an encoding scheme. Choosing the best encoding reduces to a problem in the calculus of variations, which we solve rigorously. In particular, we show there is a unique optimal encoding, and give an explicit formula for it.

We also solve the problem in a more general setting in which there is prior information about the real number, or a desire to weight errors for different values non-uniformly.

Our tools come mainly from real analysis and measure-theoretic probability. We also explore a connection to classical mechanics.

**1. Introduction.** In Shannon's information theory and the theory of error correcting codes, the following communication model is basic. Two parties  $A$  and  $B$  share a line of transmission, on which one can send an *ordered* sequence of bits. The receiver gets another *ordered* sequence of bits, possibly corrupted. While this corruption can change, omit, or locally transpose bits, by and large the order of the bits is kept intact. (Most work has focused on the so-called discrete memoryless channel, in which only bit changes are allowed. The model of [15] allows arbitrary changes, but only on fixed-length blocks.) Of course this model was very much motivated by the teletype networks of Shannon's day.

With today's internet, one might revisit this model. When a message is sent from one node to another, it has no fixed path. Abstractly, one might imagine a model in which symbols are being sent in a highly parallel and non-deterministic fashion with no particular fixed route. The receiver receives these symbols in some probabilistic sense but in no particular order.

Suppose we still consider sending bit sequences. Then if arbitrary re-orderings are allowed, then only the cardinality, or what amounts to the same thing, the fraction of 1's observed, will matter. Furthermore, if some omissions occur probabilistically then even this fraction is only meaningful approximately. Thus, with arbitrary re-ordering

---

\*An extended abstract of this paper appeared in the 16th International Symposium of Fundamentals of Computation Theory (FCT), Budapest, Hungary, 2007 [4].

<sup>†</sup>Supported by NSF CCF-0523680 and a Vilas Research Associate Award. Computer Sciences Department, University of Wisconsin, Madison, WI 53706.

<sup>‡</sup>Supported by NSF CCF-0511679 and CCF-0830488. Computer Sciences Department, University of Wisconsin, Madison, WI 53706. E-mail: jyc@cs.wisc.edu

of the bits, it severely restricts the ways by which information may be meaningfully conveyed.

Instead of sending bit sequences, what about sending a real number holistically? Let's consider the following new model of information transmission. Two parties  $A$  and  $B$  have access to a one-way communication medium, and  $A$  wishes to transmit a real number  $x$  to  $B$ . The medium may transmit signals, with some probabilistic error, in large multiplicity but in no particular order. By normalizing we assume  $0 \leq x \leq 1$ , and think of  $x$  as a probability. Communication is done by the following process. Party  $A$  can send a large number of i.i.d. samples from a Bernoulli distribution to  $B$ . The receiver observes these bits and estimates  $x$ . (The Bernoulli distribution, on the samples generated a priori, accounts for the probabilistic nature of errors and losses of signals due to the communication medium.)

The new information transmission problem is the following. We may think of the Bernoulli random variable as an "encoding" of  $x$ , through its mean value. Then what does it mean to be a good encoding scheme? How do we evaluate encoding strategies, and is there an optimal one? We note that  $x$  is only transmitted completely in the limit, so the answers must be asymptotic.

Although abstract, this problem is motivated by concrete current research in computer networking. As is familiar, messages are broken up into small packets which are then sent more or less independently along different routes. These routes can vary with time of day, system load, etc., so the network must maintain and transmit information about their quality.

We can think of a particular route as consisting of  $\ell$  links, say  $v_{i-1} \rightarrow v_i$  for  $i = 1, \dots, \ell$ . Each link has an associated number  $p_i$ ,  $0 \leq p_i \leq 1$ . For example,  $p_i$  could be a normalized cost or a measure of congestion for using the  $i$ -th link. The network can determine through observation the average  $x = (\sum_i p_i)/\ell$  for a particular route, allowing the routing protocol to take this into account so as to avoid congestion.

To allow efficient estimation of this average, researchers have investigated the possibility of using current packet designs, which already specify a bit called the Explicit Congestion Notification (ECN) bit. Each link on a route may set this bit to 0 or 1 as it sees fit, for every packet it handles. This bit then gets transmitted to the next link, which may be reset again. Recently, networking researchers have focused on a class of protocols using ECN (so-called *one-bit protocols*), which can be defined mathematically as follows. The link  $v_{i-1} \rightarrow v_i$  receives a bit  $X_{i-1} \in \{0, 1\}$  from the previous link; based on  $X_{i-1}$  and  $p_i$  it uses randomization to produce  $X_i$ . The last node can observe  $X_\ell$  many times and combine these observations to produce an estimate for  $x$ .

Several protocols of this type appear in the literature [2, 12]. What they have in common is that the expected value of  $X_\ell$  is some function  $f$  of the average  $x$ . The observer then tries to infer  $x$  from the observed approximate value of  $f(x)$ . This is

an example of our new model of information transmission, in that, one produces a collection of 0-1 random variables all with the expected value equal to some function  $f$  of some number  $x$ . The receiver observes these 0-1 random variables, in no particular order and with probabilistic losses and delays. From an observational record the receiver tries to infer  $x$ .

Since one can imagine more general schemes using this idea, there is no reason to expect developments to stop with ECN. For inspiration, we look to Shannon, who did not waste time optimizing teletype codes, but rather went on to study general methods of symbolic encoding. It is compelling, therefore, to develop a theory applicable to all of the more general schemes, and ask if there is any choice of  $f$  that is in some sense optimal. In this paper, we answer this question affirmatively, under conditions on  $f$  that are as general as could be desired.

**2. The Formalized Problem and a Guide to its Solution.** Initially,  $A$  and  $B$  agree on a transformation function  $f$ . To send  $x \in [0, 1]$ , to  $B$ , the transmitter  $A$  generates random bits, which are i.i.d. 0-1 random variables with expected value  $y = f(x)$ . The receiver  $B$  gets  $n$  of these, say  $Y_1, \dots, Y_n$ , and uses  $f^{-1}(\frac{1}{n} \sum_{i=1}^n Y_i)$  to estimate  $x$ . For this to work,  $f$  must be strictly monotonic, say increasing. Also,  $f$  should map 0 to 0 and 1 to 1, to avoid loss of bandwidth.

We now outline our criterion for evaluating  $f$ , and justify its choice. Let  $g = f^{-1}$  and  $\bar{Y} = n^{-1} \sum_{i=1}^n Y_i$ . If  $g$  is smooth, then by the strong law of large numbers,  $g(\bar{Y}) \rightarrow x$ , *a.e.* We expect  $g(\bar{Y}) - x$  to be  $\Theta(n^{-1/2})$ , so the natural measure for the error is  $\mathbf{E}[n(g(\bar{Y}) - x)^2]$ . By the mean value theorem, we should have  $n(g(\bar{Y}) - x)^2 \approx g'(y)^2[n(\bar{Y} - y)^2]$ , and  $\mathbf{E}[n(\bar{Y} - y)^2] = y(1 - y)$ . Thus, we expect

$$(1) \quad \mathbf{E}[n(g(\bar{Y}) - x)^2] \rightarrow g'(y)^2 y(1 - y).$$

Written in terms of  $f$ , this is  $\frac{f(x)(1-f(x))}{f'(x)^2}$ . Thus, we should try to minimize

$$(2) \quad \int_0^1 \frac{f(x)(1-f(x))}{f'(x)^2} dx,$$

over a suitable class of functions  $f$ . The optimal choice turns out to be

$$(3) \quad f = \frac{1 - \cos(\pi x)}{2}.$$

In particular, the optimal choice is *not* the identity function, as one might naively suppose. Nor is the naive choice even close: its value of (2) exceeds the optimum by about 64%.

In the rest of this paper, we carry out this argument in a rigorous way. The interchange of limits and integration is not trivial, because we want it to hold for the optimum curve, for which the integrand is unbounded. Also, we derive the optimal curve using Calculus of Variations by the Euler-Lagrange equation. But as is typical

with Calculus of Variations, this derivation only *suggests* optimality. (Euler's mathematics was brilliant, but it was from a time when the standard of mathematical rigor was different.) It is notorious with such problems, as typified by the Dirichlet problem [9, p. 119], the hard part is to *prove* optimality. We will do this by an independent argument, under very general conditions on the curve. Our tools come mainly from real analysis and measure-theoretic probability, in particular Lebesgue's convergence theorems, Fatou's lemma, and uniform integrability.

The rest of the paper is organized as follows. In Section 3 we prove the point-wise (almost everywhere) convergence stated in (1). Next we show that the limit of its average (the integration over possible values of  $x$ ) is given by (2), for the particular choice (3). In Section 4 we prove that (3) actually minimizes (2). The more difficult proof of a general limit theorem is given in Section 5, where we show that the average of (1) has a limit, for a wide class of transformations. In Section 6 we treat these problems in a more general setting in which the receiver has prior information about  $x$ , or wishes to weight errors for different  $x$  differently. In Section 7 we connect our variational problems to classical mechanics. In Section 8 we discuss open problems.

### 3. Two Convergence Theorems.

**3.1. Notation.** We call  $f$  *admissible* if  $f \in C[0, 1]$  (continuous), and is strictly increasing, with  $f(0) = 0$  and  $f(1) = 1$ . Let  $g = f^{-1}$  be its inverse function (also admissible). Since  $f$  and  $g$  are increasing,  $f'$  and  $g'$  exist a.e. [14]. Whenever  $f'(x) \neq 0$ , then at  $y = f(x)$ ,  $g'(y)$  exists and  $g'(y) = 1/f'(x)$ . If  $f'(x) = 0$ , we say  $g'$  has a *singularity* at  $y$ .

Our class of functions is the natural one to consider, for  $f$  can only be computable if it is continuous, as is well known [8].

Let  $Y_1, Y_2, \dots, Y_n$  be i.i.d. 0-1 random variables with  $\Pr[Y_i = 1] = y$ , and let  $\bar{Y} = \frac{\sum_{i=1}^n Y_i}{n}$  be their sample mean. We also let  $\hat{Y} = \left( \frac{\sum_{i=1}^n (Y_i - y)}{\sqrt{n}} \right)^2$ , so that  $n(\bar{Y} - y)^2 = \hat{Y}$ . Note that  $0 \leq \hat{Y} \leq n$ .

We will find it convenient to use measure theory notation. Accordingly, let  $\Omega = \{0, 1\}^n$ , with the measure  $\mu$  induced by  $n$  Bernoulli trials with success probability  $y$ . Then, for example,  $\mathbf{E}[\hat{Y}] = \int_{\Omega} \hat{Y} d\mu = \mathbf{Var}(Y_1) = y(1 - y)$ .

For a choice of  $f$  as above, it will be convenient to let

$$(4) \quad F_n(y) = \int_{\Omega} n(g(\bar{Y}) - g(y))^2 d\mu,$$

and  $\alpha = n(g(\bar{Y}) - g(y))^2$ . Since  $\alpha \leq n$ , we have  $0 \leq F_n(y) \leq n$ .

### 3.2. A Pointwise Convergence Theorem.

**THEOREM 1.** *Let  $f$  be admissible and  $0 < y < 1$ . If  $g'(y)$  exists, we have  $\lim_{n \rightarrow \infty} F_n(y) = (g'(y))^2 y(1 - y)$ , where  $F_n$  is given by (4). Therefore, the convergence*

is almost everywhere (a.e.).

*Proof.* The proof is easiest when  $g'$  is continuous in an interval around  $y$ , so we assume this first. Then, for any  $\epsilon > 0$ , there exists a  $\delta > 0$  such that if  $|y' - y| \leq \delta$  then  $|(g'(y'))^2 - (g'(y))^2| \leq \frac{\epsilon}{2y(1-y)}$ . For this  $\delta$ , let  $B_\delta = \{\omega \in \Omega \mid |\bar{Y} - y| > \delta\}$ .

Since  $\int_\Omega \hat{Y} \, d\mu = y(1-y)$ , we have  $F_n(y) - (g'(y))^2 y(1-y) = I_1 + I_2 + I_3$ , where

$$I_1 = \int_{\Omega - B_\delta} [\alpha - (g'(y))^2 \hat{Y}] \, d\mu; \quad I_2 = \int_{B_\delta} \alpha \, d\mu; \quad I_3 = - \int_{B_\delta} (g'(y))^2 \hat{Y} \, d\mu.$$

We will estimate these three integrals separately.

For  $I_1$ , by the mean value theorem (MVT), there exists some  $\xi = \xi(y, \bar{Y})$  which lies between  $y$  and  $\bar{Y}$ , such that  $g(\bar{Y}) - g(y) = g'(\xi)(\bar{Y} - y)$ . Thus,  $\alpha = (g'(\xi))^2 \hat{Y}$ . Note that  $n(\bar{Y} - y)^2 = \hat{Y}$ , and on  $\Omega - B_\delta$ ,  $|\xi - y| \leq \delta$ , we have  $|(g'(\xi))^2 - (g'(y))^2| \leq \frac{\epsilon}{2y(1-y)}$ . It follows that

$$\begin{aligned} |I_1| &\leq \int_{\Omega - B_\delta} |(g'(\xi))^2 - (g'(y))^2| \hat{Y} \, d\mu \\ &\leq \frac{\epsilon}{2y(1-y)} \int_{\Omega - B_\delta} \hat{Y} \, d\mu \\ &\leq \frac{\epsilon}{2y(1-y)} \int_\Omega \hat{Y} \, d\mu \\ &= \frac{\epsilon}{2}. \end{aligned}$$

By the Chernoff bound [3],  $\mu(B_\delta) < 2e^{-2\delta^2 n}$ , so  $|I_2| \leq n \int_{B_\delta} d\mu = n\mu(B_\delta) < 2ne^{-2\delta^2 n}$ , and since  $\hat{Y} \leq n$ , we have  $|I_3| \leq (g'(y))^2 n \int_{B_\delta} d\mu \leq 2n(g'(y))^2 e^{-2\delta^2 n}$ .

Combining these three estimates, we get

$$|F_n(y) - (g'(y))^2 y(1-y)| = |I_1 + I_2 + I_3| \leq \frac{\epsilon}{2} + 2ne^{-2\delta^2 n}(1 + (g'(y))^2) < \epsilon,$$

for sufficiently large  $n$ . Since  $\epsilon$  was arbitrary, we get Theorem 1.

We indicate briefly how to modify this proof to work at any  $y \neq 0, 1$  where  $g'(y)$  exists. Only  $I_1$  needs to be reconsidered. Suppose first that  $g'(y) > 0$ . Then there is a  $\delta > 0$  such that for  $\bar{Y}$  within  $\delta$  of  $y$  (but not equal to  $y$ ), we have

$$\frac{g(\bar{Y}) - g(y)}{\bar{Y} - y} = g'(y)(1 + \eta),$$

with  $|\eta| \leq \epsilon/(6g'(y)^2 y(1-y))$ , and  $|\eta| \leq 1$ . Then (even allowing  $\bar{Y} = y$ ),

$$\alpha = n(g(\bar{Y}) - g(y))^2 = ng'(y)^2(\bar{Y} - y)^2(1 + \eta)^2.$$

Plug this into  $I_1$ , and expand  $(1 + \eta)^2$ . The main terms will cancel, and we can estimate  $\eta^2$  by  $|\eta|$ , we find

$$|I_1| \leq \int_\Omega \frac{\epsilon}{2y(1-y)} \hat{Y} = \frac{\epsilon}{2}.$$

We handle  $g' = 0$  similarly, but with

$$\frac{g(\bar{Y}) - g(y)}{\bar{Y} - y} = \eta,$$

where  $\eta^2 \leq \epsilon/(2y(1-y))$ . The case  $g' < 0$  is forbidden by monotonicity.  $\square$

**3.3. Convergence for the Optimal Transformation.** Our information transmission problem is concerned with minimizing the limit of

$$\int_0^1 \int_{\Omega} n(g(\bar{Y}) - x)^2 d\mu dx,$$

for an unknown function  $y = f(x)$ , where  $g = f^{-1}$ . Assuming the relevant integrals exist, we can write this entirely in terms of its inverse function  $g$ ,

$$\int_0^1 g'(y) \int_{\Omega} n(g(\bar{Y}) - g(y))^2 d\mu dy.$$

In this section, we evaluate the limit of this for the optimal  $f$ . A corresponding theorem for general  $f$  was stated in [2], and proved in [1]. This result, however, assumed  $g'(y)$  to be continuous on  $[0, 1]$ , and in particular bounded on this interval. While adequate for the class of functions realizable in the on-line setting for the ECN bit in a network, this assumption is not satisfied by our optimal function  $f$ . In particular, our particular  $g'(y)$  is unbounded near 0 and 1, making the resulting proof much more difficult. A proof for the general case is provided in Section 6.

In the remainder of this section, we let  $f(x) = (1 - \cos \pi x)/2$ . We note that  $f$  is smooth and strictly increasing. Its inverse function  $g(y)$  is continuously differentiable except at 0 and 1. Explicitly,

$$(5) \quad (g'(y))^2 = \frac{1}{\pi^2 y(1-y)};$$

this has a pole of order 1 at  $y = 0$  and  $y = 1$ . Let  $\tilde{F}_n(y) = g'(y)F_n(y)$ .

**THEOREM 2.** For  $f(x) = (1 - \cos \pi x)/2$ , we have

$$\lim_{n \rightarrow \infty} \int_0^1 \tilde{F}_n(y) dy = \int_0^1 \lim_{n \rightarrow \infty} \tilde{F}_n(y) dy = \int_0^1 (g'(y))^3 y(1-y) dy.$$

*Proof.* Observe that there is a symmetry between the first and the second half of the interval, by the map  $y \mapsto 1 - y$ , and therefore we will only need to evaluate  $\lim_{n \rightarrow \infty} \int_0^{1/2} \tilde{F}_n(y) dy$ .

Let  $\delta_n = \frac{8 \log n}{n}$ . Then

$$(6) \quad \int_0^{1/2} \tilde{F}_n(y) dy = \int_0^{\delta_n} \tilde{F}_n(y) dy + \int_0^{1/2} F_n^*(y) dy,$$

where  $F_n^*(y) = \tilde{F}_n(y)\mathbf{1}_{[\delta_n, 1/2]}$ , and  $\mathbf{1}$  denotes the indicator function. Our strategy will be to prove that the first term has the limit 0, and use Lebesgue's dominated convergence theorem to evaluate the limit of the second.

Let  $y < \delta_n$ . As  $F_n(y)$  is itself an integral, we may (as with Gaul) divide it into three parts:

$$(7) \quad F_n(y) = \int_{\bar{Y} \leq y} \alpha \, d\mu + \int_{y < \bar{Y} \leq 1/2} \alpha \, d\mu + \int_{\bar{Y} > 1/2} \alpha \, d\mu.$$

We will show that the contributions of each part in the integral  $\int_0^{\delta_n} \tilde{F}_n(y) \, dy$  goes to 0.

If  $\bar{Y} \leq y$ , by the monotonicity of  $g$  we get  $(g(\bar{Y}) - g(y))^2 \leq (g(y))^2 = x^2$ . It is easy to check by elementary calculus that  $1 - \cos t \geq t^2/4$  for  $0 \leq t \leq \pi/3$ . then  $y = f(x) = (1 - \cos \pi x)/2 \geq \frac{\pi^2}{8}x^2$ , for  $0 \leq x \leq 1/3$ . It follows that, for  $0 \leq y \leq 1/4$ ,

$$\int_{\bar{Y} \leq y} \alpha \, d\mu \leq nx^2 \int_{\Omega} d\mu = nx^2 \leq \frac{8ny}{\pi^2}.$$

So, there is a  $c > 0$  such that for  $\bar{Y} \leq y$  and sufficiently large  $n$ ,

$$(8) \quad \int_0^{\delta_n} g'(y) \int_{\bar{Y} \leq y} \alpha \, d\mu \, dy \leq cn \int_0^{\delta_n} \sqrt{y} \, dy = \frac{2c}{3} n \delta_n^{3/2} \longrightarrow 0.$$

For  $y < \bar{Y} \leq 1/2$ , by MVT, there exists some  $\xi = \xi(y, \bar{Y})$  such that  $g(\bar{Y}) - g(y) = g'(\xi)(\bar{Y} - y)$ , satisfying  $y \leq \xi \leq \bar{Y} \leq 1/2$ . By the explicit formula for  $g'$  we have  $(g'(\xi))^2 \leq \frac{2}{\pi^2 y}$ . Thus

$$\int_{y < \bar{Y} \leq 1/2} \alpha \, d\mu \leq \frac{2}{\pi^2 y} \int_{\Omega} \hat{Y} \, d\mu \leq \frac{2}{\pi^2}.$$

Then

$$(9) \quad \int_0^{\delta_n} g'(y) \int_{y < \bar{Y} \leq 1/2} \alpha \, d\mu \, dy \leq \frac{2g(\delta_n)}{\pi^2} \longrightarrow 0.$$

Finally we treat  $\bar{Y} > 1/2$ . From the Chernoff bound, we have

$$\int_{\bar{Y} > 1/2} \alpha \, d\mu \leq n\mu(\bar{Y} > 1/2) < ne^{-n/8}.$$

Therefore

$$(10) \quad \int_0^{\delta_n} g'(y) \int_{\bar{Y} > 1/2} \alpha \, d\mu \, dy < ne^{-n/8} \int_0^{\delta_n} g'(y) \, dy = ne^{-n/8} g(\delta_n) \longrightarrow 0.$$

Combining (8)–(10) with (7), we get  $\lim_{n \rightarrow \infty} \int_0^{\delta_n} \tilde{F}_n(y) \, dy = 0$ .

We now consider the second integral in (6). Our first goal is to bound  $F_n(y)$  independently of  $n$  on  $\delta_n \leq y \leq 1/2$ .

Let  $B$  denote the event that  $[\bar{Y} < y/2$  or  $\bar{Y} > 3/4]$ . Inspired by King Solomon, we now divide  $F_n$  into two:

$$F_n(y) = \int_B \alpha \, d\mu + \int_{B^c} \alpha \, d\mu.$$

By the Chernoff bound [3], and  $y \geq \delta_n$ ,

$$\mu(B) < e^{-yn/8} + e^{-n/8} < 2/n.$$

It follows that

$$(11) \quad \int_B \alpha \, d\mu \leq n\mu(B) < 2.$$

On  $B^c$ , by the mean value theorem (MVT), there exists some  $\xi = \xi(y, \bar{Y})$  which lies between  $y$  and  $\bar{Y}$ , such that  $g(\bar{Y}) - g(y) = g'(\xi)(\bar{Y} - y)$ . Therefore  $\alpha = (g'(\xi))^2 \hat{Y}$ . Since  $\bar{Y} \in B^c$ , we have  $y/2 \leq \bar{Y} \leq 3/4$ . Combining this with  $y \leq 1/2$ , we get  $y/2 \leq \xi \leq 3/4$ . Using this in (5), we see that  $(g'(\xi))^2 \leq \frac{8}{\pi^2 y}$ . Then

$$(12) \quad \int_{B^c} \alpha \, d\mu \leq \frac{8}{\pi^2 y} \int_{B^c} \hat{Y} \, d\mu \leq \frac{8}{\pi^2 y} \int_{\Omega} \hat{Y} \, d\mu = \frac{8(1-y)}{\pi^2} \leq \frac{8}{\pi^2}.$$

From (11) and (12) we see that for  $y \geq \delta_n$ ,  $F_n(y) \leq \frac{8}{\pi^2} + 2 < 3$ . This implies that

$$|F_n^*| \leq 3g'(y),$$

and since  $g'$  is integrable on  $[0, 1/2]$  (near 0,  $g'$  is of order  $1/\sqrt{y}$ ) we can apply dominated convergence to get

$$\lim_{n \rightarrow \infty} \int_0^{1/2} F_n^*(y) \, dy = \int_0^{1/2} \lim_{n \rightarrow \infty} \tilde{F}_n(y) \, dy = \int_0^{1/2} (g'(y))^3 y(1-y) \, dy.$$

□

**4. Deriving the Optimal Transformation.** We consider the following optimization problem. Let

$$I_y = \int_0^1 \frac{y(1-y)}{(y')^2} dx.$$

We seek a smooth increasing function  $y$ , satisfying the boundary conditions  $y(0) = 0$  and  $y(1) = 1$ , that minimizes  $I_y$ . (Note that we are now letting  $y$  stand for a function, instead of a value.)

We use the calculus of variations to get a guess for  $y$ . Form the Euler-Lagrange equation

$$\frac{\partial L}{\partial y} - \frac{d}{dx} \frac{\partial L}{\partial y'} = 0,$$

with  $L(y, y') = y(1-y)/(y')^2$ . Then,  $y(x) = (1 - \cos \pi x)/2$  is a solution matching the boundary conditions, for which  $I_y = 1/\pi^2$ . (Integrability is to be expected here, since  $L$  did not involve  $x$  explicitly. See [5].)

More work is needed to prove this is optimal. Recall that  $y$  is *admissible* if it is in  $C[0, 1]$ , strictly increasing, with  $y(0) = 0$ , and  $y(1) = 1$ .



THEOREM 3. For any admissible function  $y$ , we have

$$\int_0^1 \frac{y(1-y)}{(y')^2} dx \geq \frac{1}{\pi^2},$$

with equality iff  $y = (1 - \cos \pi x)/2$ . The case where the integral is infinite is not excluded.

*Proof.* Define a new admissible function  $\theta$  by  $y(x) = (1 - \cos \pi\theta(x))/2$ . Since  $\theta$  increases,  $\theta'$  exists a.e., and at any point  $x$  of differentiability,  $y'(x) = \frac{\pi}{2} \sin \pi\theta(x) \cdot \theta'(x)$ , by the mean value theorem. Also  $y(x) \neq 0, 1$  except for  $x = 0, 1$ , so we have (using  $\sin^2 + \cos^2 = 1$ )  $\frac{(y')^2}{y(1-y)} = \pi^2 \theta'$ , a.e.

We may assume that  $\theta'$  is positive a.e. and  $I_y < +\infty$ , as otherwise the theorem is true. Then, by Jensen's inequality,

$$\int_0^1 \frac{y(1-y)}{(y')^2} dx = \frac{1}{\pi^2} \int_0^1 \frac{1}{(\theta')^2} dx \geq \frac{1}{\pi^2 \left( \int_0^1 \theta'(x) dx \right)^2}.$$

(To apply this, we need  $\theta' \in L^1[0, 1]$ , which is true. See [14, Ex. 13, p. 157].)

We have  $\int_0^1 \theta' \leq \theta(1) - \theta(0) = 1$ , with equality iff  $\theta$  is absolutely continuous (AC). (Combine [13, Thm. 2, p. 96] and [14, Thm. 7.18].) This gives the inequality of the theorem.

We may assume that  $\theta$  is AC (otherwise, the inequality is strict). If  $\theta'$  is not constant a.e., then the Jensen inequality is strict and we are done. On the other hand, if the inequality becomes equality, we have to have  $\theta'(x) = c$  a.e. Then,  $\theta(x) = \int_0^x c = cx$ , so  $c = 1$  and  $\theta = x$  (everywhere!), giving the theorem.  $\square$

REMARKS: It is possible that the integral is infinite; this happens, for example, if  $y = x^3$ . Also, without the monotonicity condition, the minimum need not exist. Consider, for example,  $y_n = \sin^2((n+1)\pi x)$ . Then we have  $0 \leq y_n \leq 1$ , with  $y_n(0) = 0$  and  $y_n(1) = 1$ . However,  $\int_0^1 y_n(1-y_n)(y_n')^{-2} dx = 1/(4\pi^2(n+1)^2) \rightarrow 0$ .

THEOREM 4. Let  $f$  be any admissible function. If  $f \neq (1 - \cos \pi x)/2$ , there is a constant  $\delta_f > 0$  with the following property. For sufficiently large  $n$ ,

$$\int_0^1 \mathbf{E}[n(g(\bar{Y}) - y)^2] dx \geq \frac{1}{\pi^2} + \delta_f.$$

*Proof.* By Fatou's lemma [14] and Theorem 1,

$$\liminf_{n \rightarrow \infty} \int_0^1 \mathbf{E}[n(g(\bar{Y}) - y)^2] dx \geq \int_0^1 \lim_{n \rightarrow \infty} \mathbf{E}[n(g(\bar{Y}) - y)^2] dx = \int_0^1 g'(y)^2 y(1-y) dx.$$

But this is strictly greater than the corresponding integral for  $f = (1 - \cos \pi x)/2$ , which is  $1/\pi^2$ .  $\square$

**5. Convergence for General Transformations.** In this section we investigate the convergence properties of  $\int_0^1 \tilde{F}_n(y) dy$ , as  $n \rightarrow \infty$ . We will show that under very general conditions on  $f$ ,  $\lim_{n \rightarrow \infty} \int_0^1 \tilde{F}_n(y) dy$  exists, and

$$\lim_{n \rightarrow \infty} \int_0^1 \tilde{F}_n(y) dy = \int_0^1 \lim_{n \rightarrow \infty} \tilde{F}_n(y) dy = \int_0^1 (g'(y))^3 y(1-y) dy.$$

We will now discuss and enumerate a number of conditions. First we require  $(g'(y))^3 y(1-y) \in L^1[0, 1]$ , i.e., Lebesgue integrable. If  $g'$  has no singularity in  $[0, 1]$ , then this condition is trivially satisfied, since  $f \in C^1[0, 1]$ . We now require that  $g'$  has at most finitely many singularities in  $[0, 1]$ ,

$$0 < y_1 < \dots < y_k < 1$$

where 0 and/or 1 may or may not be a singularity point. If there is any singularity point  $y_i$  (other than 0 or 1), we make two requirements at  $y_i$ : An asymptotic requirement and a monotonicity requirement. First, due to integrability of  $(g')^3$  near  $y_i$ , we require

$$(13) \quad g'(y) = o\left(\left(|y - y_i| \cdot \log \frac{1}{|y - y_i|}\right)^{-1/3}\right),$$

as  $y \rightarrow y_i$ . Second, we require that within some small interval of  $y_i$ ,  $g'(y)$  monotonically increases to  $\infty$ . This second requirement can be replaced by the weaker condition that within some small interval of  $y_i$ ,

$$(14) \quad g'(y) \leq d(y)^{1/3}$$

for some integrable function  $d$ , which monotonically increases to  $\infty$ , as  $y \rightarrow y_i$ . Both requirements (13) and (14) can be simultaneously satisfied by one simpler but slightly more stringent condition:

$$g'(y) = o\left(\left(\left(|y - y_i| \cdot \left(\log \frac{1}{|y - y_i|}\right)^{1+\epsilon}\right)^{-1/3}\right)\right),$$

for some constant  $\epsilon > 0$ , as  $y \rightarrow y_i$ .

If 0 is a singularity of  $g'$ , then due to integrability of  $(g'(y))^3 y$  near 0, we require

$$(15) \quad g'(y) = o\left(y^{-\frac{2-\epsilon}{3}}\right),$$

for some  $\epsilon > 0$ , as  $y \rightarrow 0$ . Similarly if 1 is a singularity of  $g'$ , we require

$$(16) \quad g'(y) = o\left((1-y)^{-\frac{2-\epsilon}{3}}\right),$$

for some  $\epsilon > 0$ , as  $y \rightarrow 1$ .

By choosing a smaller  $\epsilon$  if necessary, we assume the local asymptotic and monotonicity requirements at each  $y_i$ ,  $1 \leq i \leq k$ , are all valid within  $[y_i - \epsilon, y_i + \epsilon] - \{y_i\}$ .

In particular  $g'$  is less than the RHS in (13) and (14). If 0 is also a singularity of  $g'$ , then  $g'(y) < 1/y^{\frac{2-\epsilon}{3}}$  within  $(0, \epsilon)$ . And similarly at 1.

THEOREM 5. *With the above conditions (13), (14), (15) and (16)*

$$\lim_{n \rightarrow \infty} \int_0^1 \tilde{F}_n(y) dy = \int_0^1 \lim_{n \rightarrow \infty} \tilde{F}_n(y) dy = \int_0^1 (g'(y))^3 y(1-y) dy.$$

We will first give an outline of the proof. Denote by  $\delta_n = \sqrt{\frac{\log n}{n}}$ , and  $\delta'_n = \frac{8 \log n}{n}$ . We will assume  $n$  is sufficiently large such that  $\delta_n, \delta'_n < \epsilon/4$ . For now assume both 0 and 1 are singularities of  $g'$ . We will define a sequence of intervals which partition  $[0, 1]$ :

$$[0, 1] = I_0 \cup J_0 \cup K_0 \cup J_1^- \cup I_1 \cup J_1^+ \cup K_1 \cup J_2^- \cup I_2 \cup J_2^+ \cup \dots \cup J_k^- \cup I_k \cup J_k^+ \cup K_k \cup J_{k+1} \cup I_{k+1},$$

where  $I_0 = [0, \delta'_n]$ ,  $J_0 = [\delta'_n, \frac{\epsilon}{2}]$ ,  $K_0 = [\frac{\epsilon}{2}, y_1 - \frac{\epsilon}{2}]$ ; for  $1 \leq i \leq k$ ,  $J_i^- = (y_i - \frac{\epsilon}{2}, y_i - 2\delta_n]$ ,  $I_i = (y_i - 2\delta_n, y_i + 2\delta_n)$ ,  $J_i^+ = [y_i + 2\delta_n, y_i + \frac{\epsilon}{2}]$ ; for  $1 \leq i < k$ ,  $K_i = [y_i + \frac{\epsilon}{2}, y_{i+1} - \frac{\epsilon}{2}]$ ; and finally,  $K_k = [y_k + \frac{\epsilon}{2}, 1 - \frac{\epsilon}{2}]$ ,  $J_{k+1} = (1 - \frac{\epsilon}{2}, 1 - \delta'_n]$ ,  $I_{k+1} = (1 - \delta'_n, 1]$ .

If 0 is not a singularity of  $g'$ , (but there is at least one singularity,) then  $I_0$  and  $J_0$  will disappear, and  $K_0 = [0, z - \frac{\epsilon}{2}]$ , where  $z$  is the smallest singularity. If 1 is not a singularity of  $g'$ , (but there is at least one singularity,) then  $J_{k+1}$  and  $I_{k+1}$  will disappear, and  $K_k = [z' + \frac{\epsilon}{2}, 1]$ , where  $z'$  is the largest singularity. (If there are no singularity at all, then  $k = 0$ , and all intervals disappear, except  $K_0 = [0, 1]$ .)

We will establish the following:

1. For all  $0 \leq i \leq k + 1$ ,

$$(17) \quad \lim_{n \rightarrow \infty} \int_{I_i} \tilde{F}_n(y) dy = 0.$$

2. For every  $0 \leq i \leq k$ ,

$$(18) \quad \lim_{n \rightarrow \infty} \int_{K_i} \tilde{F}_n(y) dy = \int_{K_i} \lim_{n \rightarrow \infty} \tilde{F}_n(y) dy.$$

This will be shown by Lebesgue's dominated convergence theorem.

3. For every  $1 \leq i \leq k$ ,

$$(19) \quad \lim_{n \rightarrow \infty} \int_{J_i^+} \tilde{F}_n(y) dy = \int_{y_i}^{y_i + \frac{\epsilon}{2}} \lim_{n \rightarrow \infty} \tilde{F}_n(y) dy.$$

This will be shown by the property of uniform integrability.

4. For every  $1 \leq i \leq k$ ,

$$(20) \quad \lim_{n \rightarrow \infty} \int_{J_i^-} \tilde{F}_n(y) dy = \int_{y_i - \frac{\epsilon}{2}}^{y_i} \lim_{n \rightarrow \infty} \tilde{F}_n(y) dy.$$

This is symmetric to (20).

5.

$$(21) \quad \lim_{n \rightarrow \infty} \int_{J_0} \tilde{F}_n(y) dy = \int_0^{\frac{\epsilon}{2}} \lim_{n \rightarrow \infty} \tilde{F}_n(y) dy.$$

6.

$$(22) \quad \lim_{n \rightarrow \infty} \int_{J_{k+1}} \tilde{F}_n(y) dy = \int_{1-\frac{\epsilon}{2}}^1 \lim_{n \rightarrow \infty} \tilde{F}_n(y) dy.$$

This is symmetric to (21).

Theorem 5 follows from these claims, which we will establish as a series of Propositions.

PROPOSITION 1. *The limit in (17) for  $i = 0$  holds, namely,*

$$(23) \quad \lim_{n \rightarrow \infty} \int_0^{\delta'_n} \tilde{F}_n(y) dy = 0.$$

*Proof.* We break  $F_n(y) = \int_{\Omega} \alpha d\mu$  into 3 parts

$$F_n(y) = \int_{\bar{Y} \leq y} \alpha d\mu + \int_{y < \bar{Y} \leq \epsilon} \alpha d\mu + \int_{\epsilon < \bar{Y}} \alpha d\mu.$$

For  $\bar{Y} > \epsilon$ , as  $y \leq \delta'_n < \epsilon/2$ , and  $y$  is the expectation of  $\bar{Y}$ ,  $\bar{Y} - y > \epsilon/2$ , we can apply Chernoff bound

$$\mu(\bar{Y} > \epsilon) < e^{-\epsilon^2 n/2}.$$

It follows that  $\int_{\epsilon < \bar{Y}} \alpha d\mu < ne^{-\epsilon^2 n/2}$ , using the trivial bound  $\alpha \leq n$ . Then

$$\begin{aligned} \int_0^{\delta'_n} g'(y) \int_{\epsilon < \bar{Y}} \alpha d\mu dy &\leq ne^{-\epsilon^2 n/2} \int_0^{\delta'_n} g'(y) dy \\ &= ne^{-\epsilon^2 n/2} \int_0^{g(\delta'_n)} dx \\ &\leq ne^{-\epsilon^2 n/2} \rightarrow 0, \end{aligned}$$

as  $n \rightarrow \infty$ .

For  $y < \bar{Y} \leq \epsilon$ , by the mean value theorem (MVT), there exists some  $\xi = \xi(y, \bar{Y})$  which lies between  $y$  and  $\bar{Y}$ , such that  $g(\bar{Y}) - g(y) = g'(\xi)(\bar{Y} - y)$ . Since  $y \leq \xi \leq \bar{Y} \leq \epsilon$ , the upper bound (15) holds, we get

$$(g(\bar{Y}) - g(y))^2 \leq \frac{1}{\xi^{\frac{4-2\epsilon}{3}}} (\bar{Y} - y)^2 \leq \frac{1}{y^{\frac{4-2\epsilon}{3}}} (\bar{Y} - y)^2.$$

Thus,  $\alpha \leq \frac{1}{y^{\frac{4-2\epsilon}{3}}} \hat{Y}$ . It follows that  $\int_{y < \bar{Y} \leq \epsilon} \alpha d\mu \leq \frac{1}{y^{\frac{4-2\epsilon}{3}}} \int_{\Omega} \hat{Y} d\mu = \frac{1-y}{y^{\frac{1-2\epsilon}{3}}} \leq \frac{1}{y^{\frac{1-2\epsilon}{3}}}$ .

Thus, by (15) again,

$$\int_0^{\delta'_n} g'(y) \int_{y < \bar{Y} \leq \epsilon} \alpha d\mu dy \leq \int_0^{\delta'_n} \frac{dy}{y^{1-\epsilon}} \rightarrow 0.$$

For  $\bar{Y} \leq y$ , first we use the monotonicity of  $g$  to get  $(g(\bar{Y}) - g(y))^2 \leq (g(y))^2$ . Then  $\alpha \leq n(g(y))^2$ , and  $\int_{\bar{Y} \leq y} \alpha \, d\mu \leq n(g(y))^2$ .

$$\int_0^{\delta'_n} g'(y) \int_{\bar{Y} \leq y} \alpha \, d\mu \, dy \leq \int_0^{\delta'_n} n(g(y))^2 g'(y) \, dy = \int_0^{g(\delta'_n)} nx^2 \, dx = \frac{n(g(\delta'_n))^3}{3}.$$

By  $y \leq \delta'_n < \epsilon$ , (15) holds, and  $g(y) = \int_0^y g'(t) dt \leq \int_0^y 1/t^{\frac{2-\epsilon}{3}} dt = 3y^{\frac{1+\epsilon}{3}}/(1+\epsilon)$ . Therefore  $(g(\delta'_n))^3 \leq 27(\delta'_n)^{1+\epsilon}$ . It follows that  $n(g(\delta'_n))^3 \rightarrow 0$ .

Combining 3 parts, the Proposition follows.

PROPOSITION 2. *The limit in (17) for  $i = k + 1$  holds, namely,*

$$(24) \quad \lim_{n \rightarrow \infty} \int_{1-\delta'_n}^1 \tilde{F}_n(y) \, dy = 0.$$

*Proof.* This is symmetric to Proposition 1, by  $y \mapsto 1 - y$ .

PROPOSITION 3. *The limit in (17) for  $1 \leq i \leq k$  holds, namely, if  $0 < z < 1$  is a singularity of  $g'$ , then*

$$(25) \quad \lim_{n \rightarrow \infty} \int_{z-2\delta_n}^{z+2\delta_n} \tilde{F}_n(y) \, dy = 0.$$

*Proof.* We break  $F_n(y)$  into two parts  $\int_{\Omega} \alpha \, d\mu = \int_{|\bar{Y}-y| \leq \delta_n} + \int_{|\bar{Y}-y| > \delta_n}$ . By the Chernoff bound,

$$\mu(|\bar{Y} - y| > \delta_n) < 2e^{-2\delta_n^2 n} = 2/n^2.$$

Thus

$$\int_{|\bar{Y}-y| > \delta_n} \alpha \, d\mu \leq 2/n.$$

It follows that

$$\int_{z-2\delta_n}^{z+2\delta_n} g'(y) \int_{|\bar{Y}-y| > \delta_n} \alpha \, d\mu \, dy \leq \frac{2}{n} \int_{g(z-2\delta_n)}^{g(z+2\delta_n)} 1 \, dx \leq \frac{2}{n} \rightarrow 0.$$

For the other part  $\int_{|\bar{Y}-y| \leq \delta_n} \alpha \, d\mu$  we need a lemma.

LEMMA 1. *Let*

$$\Delta_n = \sup_{y, y' \in (z-3\delta_n, z+3\delta_n)} |g(y) - g(y')|.$$

*Then*

$$\Delta_n = o\left(\frac{1}{n^{1/3}}\right),$$

*as  $n \rightarrow \infty$ .*

*Proof.* (of Lemma 1): By (13),  $\forall \epsilon' > 0$ ,  $\exists N$ , such that  $\forall n \geq N$ , and  $\forall t$ , such that  $0 < |t - z| < 3\delta_n$ ,

$$g'(t) < \frac{\epsilon'}{(|t - z| \log \frac{1}{|t-z|})^{1/3}}.$$

Suppose  $y, y' \in (z - 3\delta_n, z + 3\delta_n)$ ,  $y' \leq y$ , by monotonicity of  $g$ ,

$$\begin{aligned} g(y) - g(y') &\leq g(z + 3\delta_n) - g(z - 3\delta_n) \\ &= \int_{z-3\delta_n}^{z+3\delta_n} g'(t) dt \\ &< \epsilon' \int_{z-3\delta_n}^{z+3\delta_n} \frac{dt}{(|t - z| \log \frac{1}{|t-z|})^{1/3}} \\ &= 2\epsilon' \int_0^{3\delta_n} \frac{dt}{(t \log \frac{1}{t})^{1/3}}. \end{aligned}$$

Consider

$$\frac{d}{dt} \left[ t^{2/3} (\log \frac{1}{t})^{-1/3} \right] = \frac{2}{3} \frac{1}{(t \log \frac{1}{t})^{1/3}} \left[ 1 + \frac{1}{2 \log \frac{1}{t}} \right] \geq \frac{2}{3} \frac{1}{(t \log \frac{1}{t})^{1/3}}.$$

This gives us

$$2\epsilon' \int_0^{3\delta_n} \frac{dt}{(t \log \frac{1}{t})^{1/3}} \leq 3\epsilon' \left[ t^{2/3} (\log \frac{1}{t})^{-1/3} \right]_{t=3\delta_n} \leq \frac{c\epsilon'}{n^{1/3}},$$

for some universal constant  $c$ .

It follows that

$$\Delta_n = \sup_{y, y' \in (z-3\delta_n, z+3\delta_n)} |g(y) - g(y')| \leq \frac{c\epsilon'}{n^{1/3}},$$

for all sufficiently large  $n$ . The Lemma is proved.

Using the Lemma, we estimate  $\int_{|\bar{Y}-y| \leq \delta_n} \alpha d\mu$  as follows. For  $y \in (z - 2\delta_n, z + 2\delta_n)$ ,  $\bar{Y} \in (z - 3\delta_n, z + 3\delta_n)$ . Thus,

$$\alpha = n(g(\bar{Y}) - g(y))^2 \leq n\Delta_n^2,$$

and,  $\int_{|\bar{Y}-y| \leq \delta_n} \alpha d\mu \leq n\Delta_n^2$ . Then

$$\int_{z-2\delta_n}^{z+2\delta_n} g'(y) \int_{|\bar{Y}-y| \leq \delta_n} \alpha d\mu dy \leq n\Delta_n^2 (g(z + 2\delta_n) - g(z - 2\delta_n)) \leq n\Delta_n^3.$$

By the Lemma,

$$\lim_{n \rightarrow \infty} \int_{z-2\delta_n}^{z+2\delta_n} g'(y) \int_{|\bar{Y}-y| \leq \delta_n} \alpha d\mu dy = 0.$$

Combining the two parts, we get

$$\lim_{n \rightarrow \infty} \int_{z-2\delta_n}^{z+2\delta_n} \tilde{F}_n(y) dy = 0.$$

This proves Proposition 3.

Next we consider the intervals  $K_0, \dots, K_k$ . Note that, unlike the other intervals, these intervals  $K_i$  are not dependent on  $n$ .

PROPOSITION 4. *The limit in (18) for  $0 \leq i \leq k$  holds, namely, if  $0 \leq z < z' \leq 1$  are two successive singularities of  $g'$ , then*

$$(26) \quad \lim_{n \rightarrow \infty} \int_{z+\frac{\epsilon}{2}}^{z'-\frac{\epsilon}{2}} \tilde{F}_n(y) dy = \int_{z+\frac{\epsilon}{2}}^{z'-\frac{\epsilon}{2}} \lim_{n \rightarrow \infty} \tilde{F}_n(y) dy.$$

If either 0 or 1 (or both) are not singularities of  $g'$ , then a suitable modification should be made to the cases of  $K_0$  and  $K_k$ . It should be clear that the following proof works the same way.

*Proof.* We denote by  $K = [z + \frac{\epsilon}{2}, z' - \frac{\epsilon}{2}]$ . On  $\hat{K} = [z + \frac{\epsilon}{4}, z' - \frac{\epsilon}{4}]$ ,  $g'$  is uniformly continuous, therefore there exists  $N$ , such that  $\forall n \geq N$  and  $\forall y, y' \in \hat{K}$ , if  $|y - y'| \leq \delta_n$ , then  $|(g'(y))^2 - (g'(y'))^2| < 1$ .

If  $y \in K$ , and  $|y - y'| \leq \delta_n$ , then since  $\delta_n < \epsilon/4$ , both  $y, y' \in \hat{K}$ , we have

$$\begin{aligned} \int_{|\bar{Y}-y| \leq \delta_n} \alpha d\mu &= \int_{|\bar{Y}-y| \leq \delta_n} n(g'(\xi))^2 (\bar{Y} - y)^2 d\mu \\ &\leq \int_{|\bar{Y}-y| \leq \delta_n} ((g'(y))^2 + 1) \hat{Y} d\mu \\ &\leq ((g'(y))^2 + 1) \int_{\Omega} \hat{Y} d\mu \\ &= ((g'(y))^2 + 1)y(1 - y), \end{aligned}$$

where  $\xi = \xi(y, \bar{Y})$  lies between  $y$  and  $\bar{Y}$  by MVT.

For  $|\bar{Y} - y| > \delta_n$ , we can again use Chernoff bound,

$$\int_{|\bar{Y}-y| > \delta_n} \alpha d\mu \leq n\mu(|\bar{Y} - y| > \delta_n) \leq 1.$$

Being the sum of these two integrals, it follows that  $F_n(y)$  is dominated by  $[(g'(y))^2 + 1]y(1-y) + 1$ . Then  $\tilde{F}_n(y)$  is dominated by the integrable function  $[(g'(y))^3 + g'(y)]y(1-y) + g'(y)$ . We note that  $0 \leq g'(y) \leq \max\{1, (g'(y))^3\}$ .

Now we can apply Lebesgue's dominated convergence theorem

$$\lim_{n \rightarrow \infty} \int_K \tilde{F}_n(y) dy = \int_K \lim_{n \rightarrow \infty} \tilde{F}_n(y) dy.$$

The Proposition is proved.

Next, we handle the intervals  $J_i^+$ , for  $i = 1, \dots, k$ .

PROPOSITION 5. *The limit in (19) for  $1 \leq i \leq k$  holds, namely, if  $z = y_i$  for some  $i = 1, \dots, k$  is a singularity of  $g'$ , then*

$$(27) \quad \lim_{n \rightarrow \infty} \int_{z+2\delta_n}^{z+\frac{\epsilon}{2}} \tilde{F}_n(y) dy = \int_z^{z+\frac{\epsilon}{2}} \lim_{n \rightarrow \infty} \tilde{F}_n(y) dy.$$

*Proof.* We will use condition (14) that  $(g')^3$  is dominated by some integrable function  $d$  which is monotonic increasing to  $\infty$  in this range.

For  $|\bar{Y} - y| > \delta_n$  again it is easy by Chernoff bound,

$$\int_{|\bar{Y}-y|>\delta_n} \alpha \, d\mu \leq 2/n < 1.$$

For  $|\bar{Y} - y| \leq \delta_n$ , we have  $z < z + \delta_n \leq \bar{Y} < z + \epsilon$ . Thus,  $g'$  exists in that range, and by MVT, there exists  $\xi$  between  $y$  and  $\bar{Y}$ , such that  $\alpha = (g'(\xi))^2 \hat{Y} \leq (d(\xi))^{2/3} \hat{Y}$ . Here  $|\xi - y| \leq \delta_n$ . By the monotonicity of  $d$ ,  $d(\xi) \leq d(y - \delta_n)$ . Thus

$$\int_{|\bar{Y}-y|\leq\delta_n} \alpha \, d\mu \leq (d(y - \delta_n))^{2/3} \int_{\Omega} \hat{Y} \, d\mu \leq (d(y - \delta_n))^{2/3}.$$

It follows that  $F_n(y) = \int_{\Omega} \alpha \, d\mu < (d(y - \delta_n))^{2/3} + 1 < 2(d(y - \delta_n))^{2/3}$ , and  $\tilde{F}_n(y) \leq 2g'(y)(d(y - \delta_n))^{2/3} \leq 2d(y - \delta_n)$ , for any  $y \in [z + 2\delta_n, z + \epsilon/2]$ . Let  $G_n(y) = \tilde{F}_n(y) \cdot \mathbf{1}_{[z+2\delta_n, z+\epsilon/2]}$  be defined on  $[0, 1]$ , where  $\mathbf{1}$  denotes the indicator function, then for all  $a > 0$ ,

$$\{y \mid G_n(y) > a\} \subseteq \{y \in [z + 2\delta_n, z + \epsilon/2] \mid d(y - \delta_n) > a/2\}.$$

So

$$\begin{aligned} \int_0^1 G_n(y) \cdot \mathbf{1}_{(G_n(y) > a)} \, dy &= \int_0^1 \tilde{F}_n(y) \cdot \mathbf{1}_{[z+2\delta_n, z+\epsilon/2]} \cdot \mathbf{1}_{(G_n(y) > a)} \, dy \\ &\leq \int_0^1 2d(y - \delta_n) \cdot \mathbf{1}_{[z+2\delta_n, z+\epsilon/2]} \cdot \mathbf{1}_{(d(y-\delta_n) > a/2)} \, dy \\ &= 2 \int_0^1 d(y) \cdot \mathbf{1}_{[z+\delta_n, z+\epsilon/2-\delta_n]} \cdot \mathbf{1}_{(d(y) > a/2)} \, dy \\ &\leq 2 \int_z^{z+\epsilon/2} d(y) \cdot \mathbf{1}_{(d(y) > a/2)} \, dy. \end{aligned}$$

This last expression is independent of  $n$ . As  $d$  is integrable, this quantity goes to 0 as  $a \rightarrow \infty$ .

Thus,  $\sup_n \mathbf{E}[G_n(y) \cdot \mathbf{1}_{(G_n(y) > a)}] \rightarrow 0$  as  $a \rightarrow \infty$ . This is called *uniform integrability* of  $G_n$ . As  $\lim_{n \rightarrow \infty} G_n(y)$  exists pointwise, a.e., we can conclude from the theory of uniform integrability [6, Section 7.10] that

$$\lim_{n \rightarrow \infty} \int_0^1 G_n(y) \, dy = \int_0^1 \lim_{n \rightarrow \infty} G_n(y) \, dy.$$

It is also clear that

$$\lim_{n \rightarrow \infty} G_n(y) = \lim_{n \rightarrow \infty} \tilde{F}_n(y) \cdot \mathbf{1}_{(z, z+\epsilon/2)},$$

and

$$\int_0^1 \lim_{n \rightarrow \infty} G_n(y) \, dy = \int_z^{z+\epsilon/2} \lim_{n \rightarrow \infty} \tilde{F}_n(y) \, dy.$$



Also by definition of  $G_n$ ,

$$\int_0^1 G_n(y) dy = \int_{z+2\delta_n}^{z+\epsilon/2} \tilde{F}_n(y) dy.$$

Proposition 5 is proved.

The next Proposition takes care of  $J_i^-$ , for  $i = 1, \dots, k$ .

PROPOSITION 6. *The limit in (20) for  $1 \leq i \leq k$  holds, namely, if  $z = y_i$  for some  $i = 1, \dots, k$  is a singularity of  $g'$ , then*

$$(28) \quad \lim_{n \rightarrow \infty} \int_{z-\frac{\epsilon}{2}}^{z-2\delta_n} \tilde{F}_n(y) dy = \int_{z-\frac{\epsilon}{2}}^z \lim_{n \rightarrow \infty} \tilde{F}_n(y) dy.$$

*Proof.* This is symmetric to Proposition 5, by  $y \mapsto 1 - y$ .

PROPOSITION 7. *The limit in (21) holds, namely, if 0 is a singularity of  $g'$ , then*

$$(29) \quad \lim_{n \rightarrow \infty} \int_{\delta_n}^{\epsilon/2} \tilde{F}_n(y) dy = \int_0^{\epsilon/2} \lim_{n \rightarrow \infty} \tilde{F}_n(y) dy.$$

*Proof.* We break  $F_n(y) = \int_{\Omega} \alpha d\mu$  into 3 parts,

$$F_n(y) = \int_{\bar{Y} < y/2} + \int_{y/2 \leq \bar{Y} \leq y + \delta_n} + \int_{y + \delta_n < \bar{Y}}.$$

For  $\bar{Y} < y/2$  and  $\bar{Y} > y + \delta_n$ , we use two versions of Chernoff bounds [3],

$$\mu(\bar{Y} < y/2) < e^{-yn/8} < e^{-\delta_n' n/8} = 1/n,$$

and

$$\mu(\bar{Y} > y + \delta_n) < e^{-2\delta_n^2 n} = 1/n^2.$$

Thus,

$$\int_{\bar{Y} < y/2} \alpha d\mu + \int_{y + \delta_n < \bar{Y}} \alpha d\mu \leq n(\mu(\bar{Y} < y/2) + \mu(\bar{Y} > y + \delta_n)) < 1 + 1/n < 2.$$

For  $y/2 \leq \bar{Y} \leq y + \delta_n$ , by MVT, there exists  $\xi$  between  $y$  and  $\bar{Y}$ , such that  $\alpha = (g'(\xi))^2 \hat{Y}$ . We have  $y/2 \leq \xi \leq y + \delta_n \leq \epsilon/2 + \delta_n < \epsilon$ , therefore the upper bound (15) holds at  $\xi$ , and we get  $g'(\xi) \leq 1/\xi^{(2-\epsilon)/3} \leq 1/(y/2)^{(2-\epsilon)/3}$ . Thus,

$$\int_{y/2 \leq \bar{Y} \leq y + \delta_n} \alpha d\mu \leq \frac{4}{y^{(4-2\epsilon)/3}} \int_{\Omega} \hat{Y} d\mu \leq \frac{4}{y^{(1-2\epsilon)/3}}.$$

It follows that, by (15) again,

$$g'(y) \int_{y/2 \leq \bar{Y} \leq y + \delta_n} \alpha d\mu \leq \frac{4}{y^{1-\epsilon}},$$

which is integrable near 0.

Finally,

$$\tilde{F}_n(y) = g'(y) \int_{\Omega} \alpha \, d\mu \leq 2g'(y) + \frac{4}{y^{1-\epsilon}},$$

which is integrable near 0.

Let  $F_n^*(y) = \tilde{F}_n(y) \cdot \mathbf{1}_{[\delta_n, \epsilon/2]}$  be defined on  $[0, 1]$ , then  $F_n^*(y)$  is bounded above by an integrable function on  $[0, 1]$ , and therefore we can apply Lebesgue's dominated convergence theorem.

It follows that

$$\begin{aligned} \lim_{n \rightarrow \infty} \int_0^1 F_n^*(y) \, dy &= \int_0^1 \lim_{n \rightarrow \infty} F_n^*(y) \, dy \\ &= \int_0^1 \left( \lim_{n \rightarrow \infty} \tilde{F}_n(y) \right) \cdot \mathbf{1}_{(0, \epsilon/2)} \, dy \\ &= \int_0^{\epsilon/2} \lim_{n \rightarrow \infty} \tilde{F}_n(y) \, dy. \end{aligned}$$

As clearly  $\int_0^1 F_n^*(y) \, dy = \int_{\delta_n}^{\epsilon/2} \tilde{F}_n(y) \, dy$ , this completes the proof of Proposition 7. Symmetrically, by the map  $y \mapsto 1 - y$  we can prove

PROPOSITION 8. *The limit in (22) holds, namely, if 1 is a singularity of  $g'$ , then*

$$(30) \quad \lim_{n \rightarrow \infty} \int_{1-\epsilon/2}^{1-\delta_n'} \tilde{F}_n(y) \, dy = \int_{1-\epsilon/2}^1 \lim_{n \rightarrow \infty} \tilde{F}_n(y) \, dy.$$

**6. Modeling Prior Information and Non-uniform Penalties.** In this section we generalize our model to let the receiver have prior information about the transmitter's value  $x$ . To convey this information, we use a weight function  $\varphi$  ("prior density" in Bayesian jargon) that we assume differentiable and positive on  $(0, 1)$ .

Such a weight function also allows us to weight errors differently, depending on the value of  $x$ . For example, to send 0 and receive 0.1 might be much worse than to send 0.5 and receive 0.6, and the weight function can reflect this.

We are thus led to the more general problem of choosing an admissible  $y$  to minimize

$$\int_0^1 \frac{w(y)\varphi(x)}{(y')^2} \, dx.$$

For simplicity and clarity of exposition we will assume  $y$  smooth, i.e.  $y \in C^1[0, 1]$ .

Again, we begin with a variational approach. If  $L$  is the integrand, then

$$(31) \quad \frac{\partial L}{\partial y} - \frac{d}{dx} \frac{\partial L}{\partial y'} = 3w'\varphi(y')^{-2} + 2w\varphi'(y')^{-3} - 6w\varphi(y')^{-4}y''.$$

On the other hand,

$$(32) \quad \frac{d}{dx} (w\varphi^\alpha(y')^\beta) = w'\varphi^\alpha(y')^{\beta+1} + \alpha w\varphi^{\alpha-1}\varphi'(y')^\beta + \beta w\varphi^\alpha(y')^{\beta-1}y''.$$

The coefficients of (31) and (32) are proportional provided that  $(3 : 2 : -6) = (1 : \alpha : \beta)$ . Therefore, for  $\alpha = 2/3$  and  $\beta = -2$ , we can put the Euler-Lagrange equation in the form

$$\varphi^{1/3}(y')^{-1} \frac{d}{dx} \left( w\varphi^{2/3}(y')^{-2} \right) = 0.$$

This implies that

$$w(y)\varphi(x)^{2/3} = c(y')^2,$$

for some constant  $c$ . If we take the square root of both sides and then separate variables, we see that

$$(33) \quad \int \varphi^{1/3} dx = c_1 \int \frac{dy}{\sqrt{w(y)}} + c_2.$$

This relation plus the boundary conditions  $y(0) = 0, y(1) = 1$  will determine  $y$ .

When  $w(y) = y(1 - y)$  we can integrate the right hand side and solve for  $y$  to obtain

$$y = \frac{1 - \cos(A\Phi(x) + B)}{2},$$

where  $\Phi(x) = \int_0^x \varphi(t)^{1/3} dt$ . The optimal function will not change if we multiply  $\varphi$  by a constant, so let us normalize  $\varphi$  so that  $\Phi(1) = 1$ . Clearly  $\Phi$  is monotonic, and  $\Phi(0) = 0$ . From the boundary conditions, we get  $A = \pi$  and  $B = 0$ , so

$$y = \frac{1 - \cos(\pi\Phi(x))}{2}.$$

Optimality now can be proved as before. First, for our choice of  $y$  we have

$$\int_0^1 \frac{y(1 - y)\varphi(x)}{(y')^2} dx = \int_0^1 \frac{\varphi(x)}{\pi^2\Phi'(x)^2} dx = \frac{1}{\pi^2} \int_0^1 \varphi(x)^{1/3} dx = \frac{1}{\pi^2}.$$

Now, suppose  $y$  is any other function. Then there is a function  $\theta$ , increasing from 0 to 1 on  $[0,1]$ , for which

$$y = \frac{1 - \cos(\pi\theta(\Phi(x)))}{2}.$$

Then

$$\int_0^1 \frac{y(1 - y)\varphi(x)}{(y')^2} dx = \frac{1}{\pi^2} \int_0^1 \frac{\varphi(x)^{1/3}}{[\theta'(\Phi(x))]^2} dx.$$

Since  $\int_0^1 \varphi^{1/3} = 1$ , we can apply Jensen's inequality to get

$$\int_0^1 \frac{\varphi(x)^{1/3}}{[\theta'(\Phi(x))]^2} dx \geq \left[ \int_0^1 \theta'(\Phi(x))\varphi(x)^{1/3} dx \right]^{-2} = [\theta(1) - \theta(0)]^{-2} = 1.$$

It follows from the considerations above that any admissible  $C^1$  function is optimal with respect to some weight. Indeed, let the equation of the path be  $y = (1 - \cos(\pi\theta(x)))/2$ , where  $\theta$  increases from 0 to 1. Then we may take  $\varphi = (\theta')^3$ .

**7. Connections to Classical Mechanics.** There is an intimate connection between the calculus of variations and classical mechanics, through the work of Lagrange, Jacobi, Hamilton, and many others. (See [11].) Indeed, as remarked by Lanczos [10, p. 170], the language and methods of mechanics can be used on variational problems regardless of their origin. In this appendix we will do this. Although we do not find any new results thereby, we do gain insight into why our variational problems could be solved explicitly.

Consider first the case where there is no prior information. Think of  $y$  as the trajectory of a particle moving from 0 to 1. We want to minimize the path functional

$$\int_0^1 \frac{y(1-y)}{(y')^2} dx.$$

It is clear from this that we want the “velocity”  $y'$  to be small at ends and large in the middle. Suppose we attack this in the most naive and ham-handed way possible, by making the numerator and denominator proportional. That is, let

$$\left(\frac{dy}{dx}\right)^2 = \text{const } y(1-y).$$

This leads to the same equation as before:

$$x = c_1 + c_2 \int \frac{dy}{\sqrt{y(1-y)}}.$$

Why were we so lucky? The integrand  $F$  of (0) has the form

$$F = T(y, y') - V(y),$$

( $V = 0$ ) so  $F$  is the Lagrangian of some system. The corresponding Hamiltonian is

$$H = T(y, y') + V(y) = F.$$

Now,  $\partial H/\partial x = 0$ , since  $H$  doesn't involve  $x$  explicitly, so  $H$  is constant [11, p. 132].

It is also instructive to derive the main differential equation (33) for the general case in a more systematic way. The generalized momentum corresponding to  $y$  is

$$(34) \quad z = y' F_{y'} = -2w\varphi(y')^{-2}$$

So the Hamiltonian (for which  $F$  is the Lagrangian) is

$$H = y' L_{y'} - L = \frac{3}{2} z y'.$$

Using (34) to get a value for  $y'$ , we see that

$$H(x, y, z) = -az^{2/3}w^{1/3}\varphi^{1/3}.$$

Here  $a$  is a positive constant whose value is not important. Knowing  $H$ , we can form the Hamilton-Jacobi equation [11, p. 147]:

$$(35) \quad \frac{\partial S}{\partial x} + H(x, y, \frac{\partial S}{\partial y}) = \frac{\partial S}{\partial x} - a(\frac{\partial S}{\partial y})^{2/3} w^{1/3} \varphi^{1/3} = 0$$

We try a solution of the form  $S_1(x) + S_2(y)$  and discover that

$$\varphi^{-1/3} dS_1/dx = aw(y)^{1/2} dS_2/dy$$

must hold identically, which is only possible if both sides are constants. This implies that

$$S_1 = \int \varphi(t)^{1/3} dt, \quad S_2 = \int \frac{dy}{\sqrt{w(y)}}$$

Reduction of (35) to quadrature can be predicted from its form [5, p. 95].

**8. Open Problems.** A natural generalization to our information transmission problem is to consider a higher dimensional analog of it.

In the problem we have just addressed, there is one real number  $x \in [0, 1]$  that  $A$  wishes to transmit to  $B$ . A natural 2-dimensional version of it is this: We have a point  $x$  on the convex hull  $\Delta$  of  $\{(1, 0, 0), (0, 1, 0), (0, 0, 1)\}$ . That is,  $x = p_1 e_1 + p_2 e_2 + p_3 e_3$ , where  $p_1, p_2, p_3 \geq 0$  and  $p_1 + p_2 + p_3 = 1$ . The transmitter  $A$  can generate i.i.d. random variables with three outcomes, perhaps Red, White, and Blue with probabilities  $q_1, q_2$  and  $q_3$ . Of course,  $(q_1, q_2, q_3) \in \Delta$  as well. Now the transmitter  $A$  and the receiver  $B$  must choose beforehand a transformation  $f$  which maps  $\Delta$  to itself, with an inverse  $g$ . Then, in the same formulation of this paper, what would be the optimal transformation function  $f$ , if one exists?

We don't have any significant results to report, but we can make two remarks. First, the Euler-Lagrange equation is a nonlinear PDE with 95 terms. Second, some of the naive functions from  $\Delta$  to  $\Delta$  are not optimal. However, Dr. Zhengfeng Ji has obtained very nice results in that direction [7].

**9. Acknowledgements.** We thank John Gubner and Jack Lutz for useful comments on this work. We also thank Zhengfeng Ji for sharing his work extending our investigation to higher dimensions.

#### REFERENCES

- [1] J. ADLER, J.-Y. CAI, J.K. SHAPIRO, AND D. TOWSLEY, *Estimate of congestion price using probabilistic packet marking*. Technical Report UM-TR-200223, UMASS-Amherst, 2002. See <http://www-net.cs.umass.edu/~jshapiro/um-tr-2002-23.pdf>.
- [2] J. ADLER, J.-Y. CAI, J.K. SHAPIRO, AND D. TOWSLEY, *Estimate of congestion price using probabilistic packet marking*. In: Proc. INFOCOMM 2003, pages 2068–2078, 2003.
- [3] N. ALON AND J. SPENCER, *The Probabilistic Method*. Wiley-Interscience, 1992.

- [4] ERIC BACH AND JIN-YI CAI, *A Novel Information Transmission Problem and its Optimal Solution*. Fundamentals of Computation Theory, 16th International Symposium (FCT), Budapest, Hungary, 2007. Lecture Notes in Computer Science 4639, pp 64-75. Springer 2007.
- [5] I. M. GELFAND AND S. V. FOMIN, *Calculus of Variations*. Prentice-Hall, 1963.
- [6] G. R. GRIMMETT AND D. R. STIRZAKER, *Probability and Random Processes*. Oxford Univ. Press, 2nd edition, 1992.
- [7] ZHENG FENG JI, *Private communications*. jzf@ios.ac.cn
- [8] K. KO, *Computational Complexity of Real Functions*. Birkhauser, 1991.
- [9] T. W. KÖRNER, *Fourier Analysis*. Cambridge Univ. Press, 1990.
- [10] C. LANCZOS, *The Variational Principles of Mechanics*. Univ. Toronto Press, 3rd edition, 1966.
- [11] L. D. LANDAU AND E. M. LIFSHITZ, *Mechanics*. Pergamon, 3rd edition, 1976.
- [12] S. H. LOW AND D. E. LAPSLEY, *Optimization flow control, I: Basic algorithm and convergence*. IEEE/ACM Transactions on Networking, 7(1999), pp. 861–875.
- [13] H. L. ROYDEN, *Real Analysis*. Macmillan, 1968.
- [14] W. RUDIN, *Real and Complex Analysis*. McGraw-Hill, 1974.
- [15] S. VERDÚ AND T. S. HAN, *A general formula for channel capacity*. IEEE Transactions on Information Theory, 4(1994), pp. 1147–1157.