Some Problems Are Easier With Feedback

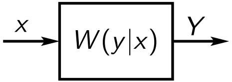
Amos Lapidoth
ETH Zurich

Information and Control in Networks, Lund, Oct. 17–19, 2012.

October 18, 2012

Joint work with Christoph Bunte.

A Discrete Memoryless Channel



- $\mathcal{X}$—input alphabet (finite).
- $\mathcal{Y}$—output alphabet (finite).
- $W(y|x)$—channel law.
- Channel is memoryless: $\Theta_1, \ldots, \Theta_n$ are IID and

$$Y_k = g\big(x_k, \Theta_k\big), k = 1, \ldots, n.$$

Encoders with or without Feedback

- $\mathcal{M} = \{1, \ldots, M\}$—message set.
- n—blocklength.
- R—rate, i.e., $n^{-1} \log M$.

- Blocklength-n encoder without feedback:

$$f : \mathcal{M} \to \mathcal{X}^n,$$

with the m-th message transmitted as

$$\mathbf{x}(m) = f(m) = \big(x_1(m), \ldots, x_n(m)\big).$$

- Blocklength-n encoder with feedback:

$$f_1 : \mathcal{M} \to \mathcal{X},\ f_2 : \mathcal{M} \times \mathcal{Y} \to \mathcal{X}, \ldots, f_n : \mathcal{M} \times \mathcal{Y}^{n-1} \to \mathcal{X},$$

with the m-th message transmitted as

$$\begin{aligned} \mathbf{x}(m) &= \Big(f_1(m), f_2(m, Y_1), \ldots, f_n(m, Y^{n-1})\Big) \\ &= \Big(x_1(m), x_2(m, Y_1), \ldots, x_n(m, Y^{n-1})\Big). \end{aligned}$$

Decoders, Errors, and Erasure

A decoder ϕ is a mapping

$$\phi\colon \mathcal{Y}^n \to \mathcal{M} \cup \{?\}.$$

Success is when $\phi(\mathbf{Y}) = m$.

Two failure modes:

- An erasure is when $\phi(\mathbf{Y}) = ?$.
- An error is when $\phi(\mathbf{Y}) \in \mathcal{M} \setminus \{m\}$.

Channel Capacity

The channel capacity C is the supremum of achievable rates, where a rate R is said to be achievable if for every $\epsilon > 0$ we can find a sufficiently large positive integer n_0 such that for all blocklengths n exceeding n_0 there exists a rate-R blocklength-n encoder f and and a decoder ϕ such that

$$\Pr\bigl(\phi(\mathbf{Y}) \in \mathcal{M}\setminus\{m\} \bigm| M = m\bigr) + \Pr\bigl(\phi(\mathbf{Y}) = ? \bigm| M = m\bigr) < \epsilon, \ m \in \mathcal{M}.$$

We allow both errors and erasures but with small probability.

In the presence of feedback it is denoted C_{FB}.

Zero-Error Capacity

The Zero-Error Capacity C_0 is the supremum of achievable rates with

$$\Pr\big(\phi(\mathbf{Y}) \in \mathcal{M} \setminus \{m\} \,\big|\, M = m\big) + \Pr\big(\phi(\mathbf{Y}) = ? \,\big|\, M = m\big) = 0, \;\; m \in \mathcal{M}.$$

We allow neither erasures nor errors

In the presence of feedback it is denoted $C_{0,\mathrm{FB}}$.

Erasures-Only Capacity

The Erasures-Only Capacity $C_{\text{e-o}}$ is the supremum of achievable rates with

$$\Pr\bigl(\phi(\mathbf{Y}) \in \mathcal{M} \setminus \{m\} \bigm| M = m\bigr) = 0, \quad m \in \mathcal{M}$$

and

$$\Pr\bigl(\phi(\mathbf{Y}) = ? \bigm| M = m\bigr) < \epsilon, \quad m \in \mathcal{M}.$$

We do not allow errors, but we do allow erasures
(with small probability).

In the presence of feedback it is denoted $C_{\text{e-o,FB}}$.

Computing C, C_0, and $C_{\text{e-o}}$

Shannon'48:

$$C = \max I(X;Y),$$

where the maximum is over all input distributions.

Computing C, C_0, and $C_{\text{e-o}}$

Shannon'48:

$$C = \max I(X;Y),$$

where the maximum is over all input distributions.

C_0 is unknown.

We do know that C_0 is positive if, and only if, we can find $x, x' \in \mathcal{X}$ that are not confusable, i.e., for all $y \in \mathcal{Y}$

$$W(y|x) \cdot W(y|x') = 0.$$

(When one is positive the other is zero.)

Computing C, C_0, and $C_{\text{e-o}}$

Shannon'48:

$$C = \max I(X;Y),$$

where the maximum is over all input distributions.

C_0 is unknown.

We do know that C_0 is positive if, and only if, we can find $x, x' \in \mathcal{X}$ that are not confusable, i.e., for all $y \in \mathcal{Y}$

$$W(y|x) \cdot W(y|x') = 0.$$

(When one is positive the other is zero.)

$C_{\text{e-o}}$ is unknown.

We do know that $C_{\text{e-o}}$ is positive if, and only if, we can find $x, x' \in \mathcal{X}$ and some $y \in \mathcal{Y}$ such that

$$W(y|x) > 0 \text{ and } W(y|x') = 0.$$

Some y is reachable from x but not from x'.

C_0 and the Adjacency Graph

- We say that x and x' are confusable if for some $y \in \mathcal{Y}$ both $W(y|x)$ and $W(y|x')$ are positive.
- The adjacency graph has vertices $\mathcal{X}$, and x and x' are connected by an edge if they are confusable.
- C_0 is determined by this graph. Only the zeros of $W(\cdot|\cdot)$ matter.

When is C_0 Positive?

- $C_0 > 0$ if, and only if, $\exists x, x'$ that are not confusable.

When is C_0 Positive?

- $C_0 > 0$ if, and only if, $\exists x, x'$ that are not confusable.
- Sufficiency: Use x and x' to send a bit per channel-use.
- Necessity: Assume that the condition is not met. Suppose both $\mathbf{x} = (x_1, \ldots, x_n)$ and $\mathbf{x}' = (x'_1, \ldots, x'_n)$ are codewords. Since x_k and x'_k are confusable, there exists an output y^*_k that is reachable from both. The output $(y^*_1, \ldots, y^*_n)$ cannot be decoded with zero probability of error.

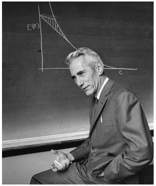
$C_{0,\mathrm{FB}}$

Shannon'56:

$$C_{0,\mathrm{FB}} = \begin{cases} 0 & \text{if } C_0 = 0 \\ \log \frac{1}{\rho} & \text{otherwise} \end{cases}$$

where

$$\rho = \min_{Q} \max_{y \in \mathcal{Y}} \sum_{x : W(y|x) > 0} Q(x).$$



Claude Shannon
(1916–2001)

The Converse when $C_0 > 0$

Let $f_1, \ldots, f_n$ be given.
Will exhibit outputs $y_1^*, \ldots, y_n^*$ that can be produced by at least $\rho^n \# \mathcal{M}$ messages.

The Converse when $C_0 > 0$ Contd.

$$\mathcal{M}_x \triangleq \{m \in \mathcal{M} : f_1(m) = x\}, \quad x \in \mathcal{X},$$

$$P_1(x) \triangleq \frac{\#\mathcal{M}_x}{\#\mathcal{M}}, \quad x \in \mathcal{X}.$$

Choose y_1^* as the argmax in

$$\rho = \min_Q \max_{y \in \mathcal{Y}} \sum_{x : W(y|x) > 0} Q(x) \le \max_{y \in \mathcal{Y}} \sum_{x : W(y|x) > 0} P_1(x)$$

$$= \sum_{x : W(y_1^*|x) > 0} P_1(x) = \frac{1}{\#\mathcal{M}} \sum_{x : W(y_1^*|x) > 0} \#\mathcal{M}_x$$

$$= \frac{1}{\#\mathcal{M}} \# \left(\bigcup_{x : W(y_1^*|x) > 0} \mathcal{M}_x \right) = \frac{\#\mathcal{M}^{(1)}}{\#\mathcal{M}}.$$

The Converse when $C_0 > 0$ Contd.

$$\mathcal{M}_x^{(1)} \triangleq \left\{m \in \mathcal{M}^{(1)} : f_2(m, y_1^*) = x\right\}, \quad x \in \mathcal{X},$$

$$P_2(x) \triangleq \frac{\# \mathcal{M}_x^{(1)}}{\# \mathcal{M}^{(1)}}, \quad x \in \mathcal{X}.$$

Choose y_2^* as the argmax in

$$\rho = \min_Q \max_{y \in \mathcal{Y}} \sum_{x: W(y|x) > 0} Q(x) \geq \max_{y \in \mathcal{Y}} \sum_{x: W(y|x) > 0} P_2(x)$$

$$= \sum_{x: W(y_2^*|x) > 0} P_2(x) = \frac{1}{\# \mathcal{M}^{(1)}} \sum_{x: W(y_2^*|x) > 0} \# \mathcal{M}_x^{(1)}$$

$$= \frac{1}{\# \mathcal{M}^{(1)}} \# \left(\bigcup_{x: W(y_2^*|x) > 0} \mathcal{M}_x \right) = \frac{\# \mathcal{M}^{(2)}}{\# \mathcal{M}^{(1)}}.$$

The Converse when $C_0 > 0$ Contd.

After n steps we obtain

$$\frac{\#\mathcal{M}^{(1)}}{\#\mathcal{M}}, \frac{\#\mathcal{M}^{(2)}}{\#\mathcal{M}^{(1)}}, \ldots, \frac{\#\mathcal{M}^{(n)}}{\#\mathcal{M}^{(n-1)}} \geq \rho.$$

Consequently,

$$\frac{\#\mathcal{M}^{(n)}}{\#\mathcal{M}} \geq \rho^n.$$

But for zero error we must have $\#\mathcal{M}^{(n)} = 1$, so

$$\#\mathcal{M} \leq \rho^{-n},$$

i.e.,

$$\frac{1}{n} \log \#\mathcal{M} \leq \log \frac{1}{\rho}.$$

A Simple Upper Bound on ρ

If C_0 is positive, then

$$\rho \le 1 - \frac{1}{\# \mathcal{X}}, \quad C_0 > 0.$$

Pf: Recall that

$$\rho = \min_{Q} \max_{y \in \mathcal{Y}} \sum_{x : W(y|x) > 0} Q(x).$$

Now choose Q (perhaps suboptimally) to be uniform

$$\begin{aligned} \rho &\le \max_{y \in \mathcal{Y}} \sum_{x : W(y|x) > 0} \frac{1}{\# \mathcal{X}} \\ &= \frac{1}{\# \mathcal{X}} \max_{y \in \mathcal{Y}} \# \Big\{ x : W(y|x) > 0 \Big\} \\ &\le \frac{1}{\# \mathcal{X}} \big(\# \mathcal{X} - 1 \big), \end{aligned}$$

because if there were a y with $W(y|x)$ positive for all x then C_0 would be zero.

Achievability

Assume $C_0 > 0$. Let P^* achieve ρ:

$$\rho = \min_Q \max_y \sum_{x: W(y|x)>0} Q(x) = \max_y \sum_{x: W(y|x)>0} P^*(x).$$

By choosing ceilings/floors judiciously, we find nonnegative integers $\{m_x\}_{x\in\mathcal{X}}$ s.t.

$$\frac{m_x}{\#\mathcal{M}} = P^*(x) \pm \frac{1}{\#\mathcal{M}}, \quad x \in \mathcal{X}.$$

Choose m_x of the messages in $\mathcal{M}$ to result in x_1 being x:

$$\frac{\#\mathcal{M}_x}{\#\mathcal{M}} = P^*(x) \pm \frac{1}{\#\mathcal{M}}, \quad x \in \mathcal{X}.$$

Achievability Contd.

After observing y_1, the survivor set $\mathcal{M}^{(1)}$ is

$$\mathcal{M}^{(1)} = \bigcup_{x:W(y_1|x)>0} \mathcal{M}_x.$$

Its cardinality is upper-bouned by:

$$\begin{aligned}
\#\,\mathcal{M}^{(1)} &= \sum_{x:W(y_1|x)>0} \#\,\mathcal{M}_x \\
&\leq \sum_{x:W(y_1|x)>0} \left(\#\,\mathcal{M}P^*(x) + 1\right) \\
&\leq \max_y \sum_{x:W(y|x)>0} \left(\#\,\mathcal{M}P^*(x) + 1\right) \\
&= \#\,\mathcal{M}\rho + \max_y \#\left\{x : W(y|x) > 0\right\} \\
&\leq \#\,\mathcal{M}\rho + \left(\#\,\mathcal{X} - 1\right),
\end{aligned}$$

where the last line is b/c $C_0 > 0$, so no y is reachable from all x's.

Achievability Contd.

After observing y_1, we choose the sets

$$\mathcal{M}_x^{(1)} \triangleq \left\{ m \in \mathcal{M}^{(1)} : f_2(m, y_1) = x \right\}, \quad x \in \mathcal{X}$$

so that

$$\frac{\# \mathcal{M}_x^{(1)}}{\# \mathcal{M}^{(1)}} = P^*(x) \pm \frac{1}{\# \mathcal{M}^{(1)}}, \quad x \in \mathcal{X}.$$

After observing y_2 the survivor set $\mathcal{M}^{(2)}$ is

$$\mathcal{M}^{(2)} = \bigcup_{x : W(y_2|x) > 0} \mathcal{M}_x^{(1)}.$$

Its cardinality is upper-bounded by:

$$\begin{aligned} \# \mathcal{M}^{(2)} &\leq \# \mathcal{M}^{(1)} \rho + \left(\# \mathcal{X} - 1 \right) \\ &\leq \# \mathcal{M} \rho^2 + (\rho + 1)\left(\# \mathcal{X} - 1 \right) \end{aligned}$$

Achievability Contd.

After observing $y_1, \ldots, y_k$, the survivor set $\mathcal{M}^{(k)}$ satisfies

$$
\begin{aligned}
\#\,\mathcal{M}^{(k)} &\le \#\,\mathcal{M}\rho^k + (\rho^{k-1} + \rho^{k-2} + \cdots + 1)\big(\#\,\mathcal{X} - 1\big) \\
&\le \#\,\mathcal{M}\rho^k + \frac{1}{1-\rho}\big(\#\,\mathcal{X} - 1\big) \\
&\le \#\,\mathcal{M}\rho^k + \#\,\mathcal{X}\big(\#\,\mathcal{X} - 1\big)
\end{aligned}
$$

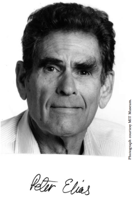
Thus, if $\#\,\mathcal{M} = \lfloor \rho^{-n} \rfloor$ we can reduce the survivor set to a singleton in $n + \big\lceil \log_2\big(1 + \#\,\mathcal{X}(\#\,\mathcal{X} - 1)\big)\big\rceil$ channel uses for a total rate of

$$
\frac{\log\lfloor \rho^{-n} \rfloor}{n + \big\lceil \log_2\big(1 + \#\,\mathcal{X}(\#\,\mathcal{X} - 1)\big)\big\rceil} \to \log \frac{1}{\rho}.
$$

$C_{0,\mathrm{FB}}$ Can Exceed C_0

For some channels

$$C_{0,\mathrm{FB}} > C_0$$

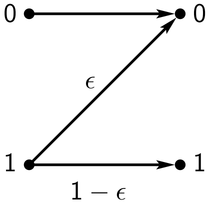


Peter Elias
(1923–2001)

Acknowledgement

I am indebted to Peter Elias for first pointing out that a feedback link could increase the zero-error capacity, as well as for several suggestions that were helpful in the proof of Theorem 7.

The Z-Channel



- If $y = 1$, then x must be 1.

$$\mathcal{L}(\mathbf{y}) = \{ m \in \mathcal{M} : x_i(m) = 1 \text{ whenever } y_i = 1 \}.$$

- If $\# \mathcal{L}(\mathbf{y}) = 1$, we can decode error-free. Otherwise we must declare an erasure.

For the Z-Channel $C_{\text{e-o}} = C$

- Let the received sequence $\mathbf{y} = y_1, \ldots, y_n$ have ν_1 ones.
- Assume each codeword has n_0 zeros and $n_1 = n - n_0$ ones.

$$p\big(\mathbf{y}|\mathbf{x}(m)\big) = (1-\epsilon)^{\nu_1}\,\epsilon^{n_1-\nu_1}, \quad m \in \mathcal{L}(\mathbf{y}).$$

- All the messages in $\mathcal{L}(\mathbf{y})$ have the same likelihood.
- The erasures-only decoder is identical to an ML decoder that declares a failure if there are ties.
- Since constant-composition codes with an ML decoder that declares a failure in the case of ties achieve capacity

$$C_{\text{e-o}} = C.$$

- In particular,

$$C_{\text{e-o}} > 0 \text{ whenever } \epsilon < 1.$$

When is $C_{\text{e-o}} > 0$?

$$\Big(C_{\text{e-o}} > 0\Big) \Leftrightarrow \Big(\exists x, x', y : W(y|x) > 0 \text{ and } W(y|x') = 0\Big).$$

When is $C_{\text{e-o}} > 0$?

$$\Big(C_{\text{e-o}} > 0\Big) \Leftrightarrow \Big(\exists x, x', y : W(y|x) > 0 \text{ and } W(y|x') = 0\Big).$$

Necessity: If every reachable y is reachable from all inputs, then no output sequence can be decoded error-free.

Sufficiency

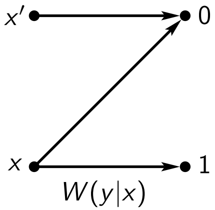
Let x, x', y be as above:

$$W(y|x) > 0 \text{ and } W(y|x') = 0.$$

- Use only x, x', and define

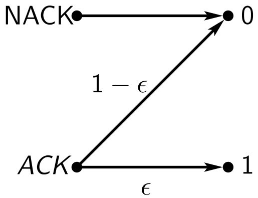
$$\tilde{Y} = \begin{cases} 0 & \text{if } Y \neq y \\ 1 & \text{otherwise.} \end{cases}$$

- This induces a Z-channel



- And for this Z-channel $C_{\text{e-o}} = C > 0$.

The Z-Channel Is Very Useful



- To send NACK send $0, 0, \ldots, 0$ (ν times).
- To send ACK send $1, 1, \ldots, 1$ (ν times).
- If 1 is received at least once, declare "ACK". Otherwise, "NACK".

With this approach

$$\Pr(\text{"ACK"} \mid \text{NACK}) = 0,$$
$$\Pr(\text{"NACK"} \mid \text{ACK}) \leq (1-\epsilon)^{\nu}.$$

$C_{\text{e-o,FB}}$

Bunte & AL: We don't know $C_{\text{e-o}}$, but we do know $C_{\text{e-o,FB}}$!

$$C_{\text{e-o,FB}} = \begin{cases} 0 & \text{if } C_{\text{e-o}} = 0 \\ C & \text{otherwise.} \end{cases}$$

$C_{\text{e-o,FB}}$

Bunte & AL: We don't know $C_{\text{e-o}}$, but we do know $C_{\text{e-o,FB}}$!

$$C_{\text{e-o,FB}} = \begin{cases} 0 & \text{if } C_{\text{e-o}} = 0 \\ C & \text{otherwise.} \end{cases}$$

The proof that

$$\left(C_{\text{e-o}} = 0 \right) \Rightarrow \left(C_{\text{e-o,FB}} = 0 \right)$$

is straightforward: if $C_{\text{e-o}} = 0$, then every reachable y is reachable from all x's, and no output sequence can be decoded also in the presence of feedback.

Achievability of $C_{\text{e-o,FB}}$ when $C_{\text{e-o}} > 0$

Phase I:

- Send the message using a blocklength-n encoder $\tilde{f}$ and decoder $\tilde{\phi}$ of rate (nearly) C that have a maximal probability of error smaller than $\delta/2$.
- Form the tentative decision $\tilde{\phi}(Y_1, \ldots, Y_n)$.
- This tentative decision is known to the transmitter via the feedback.

Phase II:

- Send an ACK or NACK ν times with ν large enough so that $\Pr(\text{“NACK”} \,|\, \text{ACK}) < \delta/2$.

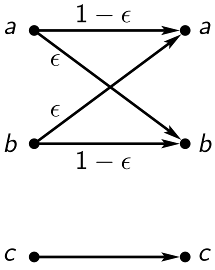
Produce the tentative decision if “ACK”; otherwise an erasure.

Analysis of Two-Phase Scheme

$$\begin{aligned}\Pr\big(\text{error}\,\big|\,M=m\big) &= \Pr\big(\tilde{\phi}(\mathbf{Y})\neq m\,\big|\,M=m\big)\Pr\big(\text{“ACK”}\,\big|\,\text{NACK}\big)\\ &= 0.\end{aligned}$$

$$\begin{aligned}\Pr\big(\text{erasure}\,\big|\,M=m\big) &= \underbrace{\Pr\big(\tilde{\phi}(\mathbf{Y})\neq m\,\big|\,M=m\big)}_{\leq\delta/2}\Pr\big(\text{“NACK”}\,\big|\,\text{NACK}\big)\\ &\quad+\Pr\big(\tilde{\phi}(\mathbf{Y})= m\,\big|\,M=m\big)\underbrace{\Pr\big(\text{“NACK”}\,\big|\,\text{ACK}\big)}_{\leq\delta/2}\\ &\leq\delta.\end{aligned}$$

$C_{\text{e-o,FB}}$ Can Exceed $C_{\text{e-o}}$



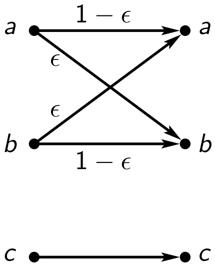
Since $C_{\text{e-o}}$ is positive, $C_{\text{e-o,FB}} = C$, and

$$C_{\text{e-o,FB}} \approx \log 3, \quad \epsilon \ll 1.$$

However, as we next argue,

$$C_{\text{e-o}} = 1.$$

$C_{\text{e-o,FB}}$ Can Exceed $C_{\text{e-o}}$ Contd.



If

$$(y_1, \ldots, y_{i-1}, a, y_{i+1}, \ldots, y_n)$$

is decoded to m, then so can

$$(y_1, \ldots, y_{i-1}, b, y_{i+1}, \ldots, y_n),$$

and this reduces the probability of erasure. Thus, for the purposes of $C_{\text{e-o}}$, we can combine the outputs a and b to a single output $\{a, b\}$. This reduces the size of the output alphabet to 2.

Thank You!

Regular Capacity:

$$C_{\mathsf{FB}} = C.$$

Zero-Error Capacity:

$$C_{0,\mathsf{FB}} = \begin{cases} 0 & \text{if } C_0 = 0 \\ \log \frac{1}{\rho} & \text{otherwise} \end{cases}$$

where

$$\rho = \min_{Q} \max_{y \in \mathcal{Y}} \sum_{x : W(y|x) > 0} Q(x).$$

Erasures-Only Capacity:

$$C_{\mathsf{e\text{-}o},\mathsf{FB}} = \begin{cases} 0 & \text{if } C_{\mathsf{e\text{-}o}} = 0 \\ C & \text{otherwise} \end{cases}.$$