Recent Results on Input-Constrained Erasure Channels
—A Case Study for Markov Approximation

Yonglong Li and Guangyue Han

The University of California, San Diego and The University of Hong Kong

July, 2017@NUS
Memoryless Channels

Channel transitions are characterized by time-invariant transition probabilities \( \{ p(y|x) \} \). Channel inputs are independent and identically distributed. Representative examples include (memoryless) binary symmetric channels, binary erasure channels and Gaussian channels.
Memoryless Channels

- Channel transitions are characterized by time-invariant transition probabilities \( \{p(y|x)\} \).
Memoryless Channels

- Channel transitions are characterized by time-invariant transition probabilities \(\{p(y|x)\}\).
- Channel inputs are independent and identically distributed.
Memoryless Channels

- Channel transitions are characterized by time-invariant transition probabilities \( \{ p(y|x) \} \).
- Channel inputs are independent and identically distributed.
- Representative examples include (memoryless) binary symmetric channels, binary erasure channels and Gaussian channels.
Capacity of Memoryless Channels

Shannon's channel coding theorem

\[ C = \sup_{p(x)} I(X;Y) = \sup_{p(x)} -\sum_{x,y} p(x,y) \log p(x,y) / p(x) p(y) . \]
Capacity of Memoryless Channels

Shannon’s channel coding theorem

\[ C = \sup_{p(x)} I(X; Y) \]

\[ = \sup_{p(x)} \sum_{x,y} p(x, y) \log \frac{p(x, y)}{p(x)p(y)}. \]
Capacity of Memoryless Channels

Shannon’s channel coding theorem

\[
C = \sup_{p(x)} I(X; Y) \\
\quad = \sup_{p(x)} \left( - \sum_{x, y} p(x, y) \log \frac{p(x, y)}{p(x)p(y)} \right).
\]

The Blahut-Arimoto algorithm (BAA)

Fig. 1. Capacity algorithm.
Memory Channels

- Channel transitions are characterized by probabilities \( p(y_i | x_{i1}, y_{i-11}, s_i) \), where channel outputs are possibly dependent on previous/current channel inputs and previous outputs and current channel state; for example, inter-symbol interference channels, flash memory channels, Gilbert-Elliot channels.

- Channel inputs may have to satisfy certain constraints which necessitate dependence among channel inputs; for example, \((d, k)\)-RLL constraints, more generally, finite-type constraints.

- Such channels are widely used in a variety of real-life applications, including magnetic and optical recording, solid state drives, communications over band-limited channels with inter-symbol interference.
Memory Channels

- Channel transitions are characterized by probabilities \( \{p(y_i|x_1^i, y_1^{i-1}, s_i)\} \), where channel outputs are possibly dependent on previous/current channel inputs and previous outputs and current channel state; for example, inter-symbol interference channels, flash memory channels, Gilbert-Elliot channels.
Memory Channels

Channel transitions are characterized by probabilities \( \{p(y_i|x_i^i, y_1^{i-1}, s_i)\} \), where channel outputs are possibly dependent on previous/current channel inputs and previous outputs and current channel state; for example, inter-symbol interference channels, flash memory channels, Gilbert-Elliot channels.

Channel inputs may have to satisfy certain constraints which necessitate dependence among channel inputs; for example, \((d, k)\)-RLL constraints, more generally, finite-type constraints.
Memory Channels

- Channel transitions are characterized by probabilities $\{p(y_i|x_{1}^{i}, y_{1}^{i-1}, s_i)\}$, where channel outputs are possibly dependent on previous/current channel inputs and previous outputs and current channel state; for example, inter-symbol interference channels, flash memory channels, Gilbert-Elliot channels.

- Channel inputs may have to satisfy certain constraints which necessitate dependence among channel inputs; for example, $(d, k)$-RLL constraints, more generally, finite-type constraints.

- Such channels are widely used in a variety of real-life applications, including magnetic and optical recording, solid state drives, communications over band-limited channels with inter-symbol interference.
Capacity of Memory Channels
Capacity of Memory Channels

Despite a great deal of efforts by Zehavi and Wolf [1988], Mushkin and Bar-David [1989], Shamai and Kofman [1990], Goldsmith and Varaiya [1996], Arnold, Loeliger, Vontobel, Kavcic and Zeng [2006], Holliday, Goldsmith, and Glynn [2006], Vontobel, Kavcic, Arnold and Loeliger [2008], Pfister [2011], Permuter, Asnani and Weissman [2013], Han [2015], ...
Capacity of Memory Channels
The Markov Approximation Scheme

It is widely believed that for generic memory (multi-user) channels, memory among channel inputs are necessary for the purpose of achieving capacity (region; by Chandra's talk).

The idea is that instead of maximizing the mutual information rate over all stationary processes, one can maximize the mutual information rate over all $m$-th order Markov processes to obtain the $m$-th order Markov capacity.

Under suitable assumptions (see, e.g., Chen and Siegel [2008]), when $m$ tends to infinity, the corresponding sequence of Markov capacities will converge to the memory channel capacity.
The Markov Approximation Scheme

- It is widely believed that for generic memory (multi-user) channels, memory among channel inputs are necessary for the purpose of achieving capacity (region; by Chandra’s talk).
The Markov Approximation Scheme

- It is widely believed that for generic memory (multi-user) channels, memory among channel inputs are necessary for the purpose of achieving capacity (region; by Chandra’s talk).
- The idea is that instead of maximizing the mutual information rate over all stationary processes, one can maximize the mutual information rate over all $m$-th order Markov processes to obtain the $m$-th order Markov capacity.
The Markov Approximation Scheme

- It is widely believed that for generic memory (multi-user) channels, memory among channel inputs are necessary for the purpose of achieving capacity (region; by Chandra’s talk).
- The idea is that instead of maximizing the mutual information rate over all stationary processes, one can maximize the mutual information rate over all $m$-th order Markov processes to obtain the $m$-th order Markov capacity.
- Under suitable assumptions (see, e.g., Chen and Siegel [2008]), when $m$ tends to infinity, the corresponding sequence of Markov capacities will converge to the memory channel capacity.
Input-Constrained Erasure Channels

- Let $F$ be a set of forbidden words over \{1, 2, ..., K\} and $S$ be the constraint with respect to $F$, consisting of all the words, each of which does not contain any element in $F$ as a contiguous subsequence.

- When $K = 2$ and $F = \{22\}$, $S$ is the set of all the words, which does not contain "22". For example, "121", "212" ∈ $S$ but not "122".

This constraint is the so-called (1, ∞)-run length limited constraint, which is a widely-used constraint in data storage.
Let $\mathcal{F}$ be a set of forbidden words over $\{1, 2, \cdots, K\}$ and $S$ be the constraint with respect to $\mathcal{F}$, consisting of all the words, each of which does not contain any element in $\mathcal{F}$ as a contiguous subsequence.
Input-Constrained Erasure Channels

Let $\mathcal{F}$ be a set of forbidden words over $\{1, 2, \cdots, K\}$ and $\mathcal{S}$ be the constraint with respect to $\mathcal{F}$, consisting of all the words, each of which does not contain any element in $\mathcal{F}$ as a contiguous subsequence.

When $K = 2$ and $\mathcal{F} = \{22\}$, $\mathcal{S}$ is the set of all the words, which does not contain “22”. For example, “121”, “212” $\in \mathcal{S}$ but not “122”.
Input-Constrained Erasure Channels

Let $\mathcal{F}$ be a set of forbidden words over $\{1, 2, \cdots, K\}$ and $S$ be the constraint with respect to $\mathcal{F}$, consisting of all the words, each of which does not contain any element in $\mathcal{F}$ as a contiguous subsequence.

When $K = 2$ and $\mathcal{F} = \{22\}$, $S$ is the set of all the words, which does not contain “22”. For example, “121”, “212” $\in S$ but not “122”.

This constraint is the so-called $(1, \infty)$-run length limited constraint, which is a widely-used constraint in data storage.
Input-Constrained Erasure Channels

An erasure channel is the channel whose input is either received correctly or erased. It can be characterized by the following input-output equation:

\[ Y_n = X_n \cdot E_n, \]

where \( \{X_n\} \) takes on values from \( \{1, 2, \ldots, K\} \), \( \{E_n\} \) is a binary stationary process with erasure rate \( \varepsilon = P(E_n = 0) \), and 0 output is interpreted as an erasure.

We will consider input-constrained erasure channels, whose input \( \{X_n\} \) is supported on \( S \) (with respect to the set \( F \) of forbidden words).
Input-Constrained Erasure Channels

- An erasure channel is the channel whose input is either received correctly or erased.
An erasure channel is the channel whose input is either received correctly or erased.

It can be characterized by the following input-output equation:

\[ Y_n = X_n \cdot E_n, \]

where \( \{X_n\} \) takes on values from \( \{1, 2, \ldots, K\} \), \( \{E_n\} \) is a binary stationary process with erasure rate \( \varepsilon \triangleq P(E_n = 0) \), and 0 output is interpreted as an erasure.
An erasure channel is the channel whose input is either received correctly or erased.

It can be characterized by the following input-output equation:

\[ Y_n = X_n \cdot E_n, \]

where \( \{X_n\} \) takes on values from \( \{1, 2, \ldots, K\} \), \( \{E_n\} \) is a binary stationary process with erasure rate \( \varepsilon \triangleq P(E_n = 0) \), and 0 output is interpreted as an erasure.

We will consider input-constrained erasure channels, whose input \( \{X_n\} \) is supported on \( S \) (with respect to the set \( \mathcal{F} \) of forbidden words).
Capacity of Input-Constrained Erasure Channels

The Shannon capacity of the channel can be computed as

$$C(S, \varepsilon) = \sup I(X; Y)$$

where the supremum is taken over all the stationary processes $X$ supported on $S$.

The $m$-th order Markov capacity is defined as

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

where the maximum is taken over all $m$-th order Markov processes supported on $S$. 

Yonglong Li and Guangyue Han

The University of California, San Diego and The University of Hong Kong
The Shannon capacity of the channel can be computed as

\[ C(S, \varepsilon) = \sup I(X; Y), \]

where the supremum is taken over all the stationary processes \( X \) supported on \( S \).
Capacity of Input-Constrained Erasure Channels

The Shannon capacity of the channel can be computed as

\[ C(S, \varepsilon) = \sup I(X; Y), \]

where the supremum is taken over all the stationary processes \( X \) supported on \( S \).

The \( m \)-th order Markov capacity is defined as

\[ C^{(m)}(S, \varepsilon) = \max I(X; Y), \]

where the maximum is taken over all \( m \)-th order Markov processes supported on \( S \).
Capacity of Input-Constrained Erasure Channels

- When $\varepsilon = 0$, $C(S, 0)$ may be referred to as the noiseless constrained capacity of $S$, which has an explicit formula [Parry 1964].

- When $\varepsilon > 0$, $C(S, \varepsilon)$ may be referred to as noisy constrained capacity, for which there is no explicit formula for $C(S, \varepsilon)$.

- The Markov approximation scheme says that to compute $C(S, \varepsilon)$, one can use $C(m(S, \varepsilon))$ to approximate it.

- When passing a Markov process through the erasure channel, the output is a hidden Markov process, whose entropy rate is extremely difficult to compute.

- However, many tractable mathematical properties of the Markovian input processes also pass through the channel.
Capacity of Input-Constrained Erasure Channels

- When $\varepsilon = 0$, $C(S, 0)$ may be referred to as the noiseless constrained capacity of $S$, which has an explicit formula [Parry 1964].
- When $\varepsilon > 0$, $C(S, \varepsilon)$ may be referred to as noisy constrained capacity, for which there is no explicit formula for $C(S, \varepsilon)$. 

The Markov approximation scheme says that to compute $C(S, \varepsilon)$, one can use $C(m) (S, \varepsilon)$ to approximate it.

When passing a Markov process through the erasure channel, the output is a hidden Markov process, whose entropy rate is extremely difficult to compute. However, many tractable mathematical properties of the Markovian input processes also pass through the channel.
Capacity of Input-Constrained Erasure Channels

- When $\varepsilon = 0$, $C(S, 0)$ may be referred to as the noiseless constrained capacity of $S$, which has an explicit formula [Parry 1964].
- When $\varepsilon > 0$, $C(S, \varepsilon)$ may be referred to as noisy constrained capacity, for which there is no explicit formula for $C(S, \varepsilon)$.
- The Markov approximation scheme says that to compute $C(S, \varepsilon)$, one can use $C^{(m)}(S, \varepsilon)$ to approximate it.
Capacity of Input-Constrained Erasure Channels

- When $\varepsilon = 0$, $C(S, 0)$ may be referred to as the **noiseless constrained capacity** of $S$, which has an explicit formula [Parry 1964].
- When $\varepsilon > 0$, $C(S, \varepsilon)$ may be referred to as **noisy constrained capacity**, for which there is no explicit formula for $C(S, \varepsilon)$.
- The Markov approximation scheme says that to compute $C(S, \varepsilon)$, one can use $C^{(m)}(S, \varepsilon)$ to approximate it.
- When passing a Markov process through the erasure channel, the output is a **hidden Markov process**, whose entropy rate is extremely difficult to compute.
Capacity of Input-Constrained Erasure Channels

- When $\varepsilon = 0$, $C(S, 0)$ may be referred to as the **noiseless constrained capacity** of $S$, which has an explicit formula [Parry 1964].
- When $\varepsilon > 0$, $C(S, \varepsilon)$ may be referred to as **noisy constrained capacity**, for which there is no explicit formula for $C(S, \varepsilon)$.
- The Markov approximation scheme says that to compute $C(S, \varepsilon)$, one can use $C^{(m)}(S, \varepsilon)$ to **approximate** it.
- When passing a Markov process through the erasure channel, the output is a **hidden Markov process**, whose entropy rate is **extremely** difficult to compute.
- However, many **tractable** mathematical properties of the Markovian input processes also pass through the channel.
Main Results

- An “explicit” formula for the mutual information rate of input-constrained erasure channels with Markovian inputs.
- Asymptotics of the capacity of input-constrained erasure channels.
- Concavity of mutual information rate for some special erasure channels with first-order input Markov processes, and computation and asymptotics of the capacity of such channels.
- Effect of feedback to input-constrained erasure channels.
Theorem (Li and Han 2016)

If \( \{E_n\} \) is stationary and \( \{X_n\} \) is an \( m \)-th order Markov process, then

\[
I(X; Y) = \sum_{k=0}^{\infty} b(k - 1, m) \sum_{t=0}^{\infty} \sum_{\{i_1, \ldots, i_t\} \in B_2(k - 1, t)} H(X_0 | X_{i_1}, X_{i_1 + 1} - k - m) \times P(E_{|A(k, i_1)} = 1, E_{\bar{A}(k, i_1)} = 0).
\]

Proof. It follows from the “forgetting” property of \( Y \) and some tedious yet straightforward computations.
An “Explicit” Formula for $I(X; Y)$

Theorem (Li and Han 2016)

If $\{E_n\}$ is stationary and $\{X_n\}$ is an $m$-th order Markov process, then

$$I(X; Y) = \sum_{k=0}^{\infty} \sum_{t=0}^{b(k-1,m)} \sum_{\{i_{i_1}^t\} \in B_2(k-1,t)} H(X_0|X_{i_1^t}, X_{-k-1}^{-k-m}) \\ \times P(E_A(k,i_1^t) = 1, E_{\bar{A}}(k,i_1^t) = 0).$$

Proof. It follows from the “forgetting” property of $Y$ and some tedious yet straightforward computations.
An “Explicit” Formula for $I(X; Y)$

Theorem (Li and Han 2016)

If $\{E_n\}$ is stationary and $\{X_n\}$ is an $m$-th order Markov process, then

$$I(X; Y) = \sum_{k=0}^{\infty} \sum_{t=0}^{b(k-1,m)} \sum_{\{i_t^i\} \in B_2(k-1,t)} H(X_0|X_{i_t^i}, X_{-k-1}^{-k-m}) \times P(E_A(k,i_t^i) = 1, E_{\bar{A}}(k,i_t^i) = 0).$$

Proof.
It follows from the “forgetting” property of $Y$ and some tedious yet straightforward computations.

Yonglong Li and Guangyue Han
The University of California, San Diego and The University of Hong Kong
Corollary (Li and Han 2016)

If \( \{E_n\} \) is i.i.d. and \( \{X_n\} \) is an \( m \)-th order Markov process, then

\[
I(X; Y) = (1 - \varepsilon)^{m+1} \sum_{k=0}^{\infty} \sum_{t=0}^{b(k-1,m)} a(k, t)(1 - \varepsilon)^t \varepsilon^{k-t},
\]

where

\[
a(k, t) = \sum_{\{i_1...i_t\} \in B_2(k-1,t)} H(X_0|X_{i_1^t}, X_{-k-1}^{k-m}).
\]
Corollary (Li and Han 2016)

If \( \{E_n\} \) is i.i.d. and \( \{X_n\} \) is an \( m \)-th order Markov process, then

\[
I(X; Y) = (1 - \varepsilon)^{m+1} \sum_{k=0}^{\infty} b(k-1,m) \sum_{t=0}^{b(k-1,m)} a(k, t)(1 - \varepsilon)^t \varepsilon^{k-t},
\]

where

\[
a(k, t) = \sum_{\{i_1...i_t\} \in B_2(k-1,t)} H(X_0|X_{i_1^t}, X_{-k-1}^{k-m}).
\]

In particular, if \( \{X_n\} \) is a first-order Markov chain, then

\[
I(X; Y) = (1 - \varepsilon)^2 \sum_{k=0}^{\infty} H(X_0|X_{-k-1}) \varepsilon^k.
\]
Asymptotics of the Capacity

Assume that \( \{E_n\} \) is i.i.d. and \( S \) is a finite-type constraint of topological order \( m \). Then,

\[
C(S, \varepsilon) = C(S, 0) - \left\{ (m + 1) H(\hat{X}_0 | \hat{X}_{-1} - m) - m \sum_{i=1}^{m} H(\hat{X}_0 | \hat{X}_{-1} - i + 1, \hat{X}_{-i} - i - m) \right\} \varepsilon + O(\varepsilon^2).
\]

Moreover, for any \( n \geq m \),

\[
C(n)(S, \varepsilon) \text{ is of the same asymptotic form as above, namely,}
\]

\[
C(n)(S, \varepsilon) = C(S, 0) - \left\{ (m + 1) H(\hat{X}_0 | \hat{X}_{-1} - m) - m \sum_{i=1}^{m} H(\hat{X}_0 | \hat{X}_{-1} - i + 1, \hat{X}_{-i} - i - m) \right\} \varepsilon + O(\varepsilon^2).
\]

Proof. Use the “explicit” formula and some convexity analysis.
Asymptotics of the Capacity

Theorem (Li and Han 2016)

Assume that \( \{ E_n \} \) is i.i.d. and \( S \) is a finite-type constraint of topological order \( m \). Then,

\[
C(S, \varepsilon) = C(S, 0) - \left\{ (m + 1)H(\hat{X}_0|\hat{X}_{-m}^{-1}) - \sum_{i=1}^{m} H(\hat{X}_0|\hat{X}_{-i+1}^{-1}, \hat{X}_{-i-m}^{-1}) \right\} \varepsilon + O(\varepsilon^2).
\]
Asymptotics of the Capacity

Theorem (Li and Han 2016)

Assume that \( \{E_n\} \) is i.i.d. and \( S \) is a finite-type constraint of topological order \( m \). Then,

\[
C(S, \varepsilon) = C(S, 0) - \left\{ (m + 1)H(\hat{X}_0 | \hat{X}_{-m}^{-1}) - \sum_{i=1}^{m} H(\hat{X}_0 | \hat{X}_{i+1}^{-1}, \hat{X}_{i-m}^{-1}) \right\} \varepsilon + O(\varepsilon^2).
\]

Moreover, for any \( n \geq m \), \( C^{(n)}(S, \varepsilon) \) is of the same asymptotic form as above, namely,

\[
C^{(n)}(S, \varepsilon) = C(S, 0) - \left\{ (m + 1)H(\hat{X}_0 | \hat{X}_{-m}^{-1}) - \sum_{i=1}^{m} H(\hat{X}_0 | \hat{X}_{i+1}^{-1}, \hat{X}_{i-m}^{-1}) \right\} \varepsilon + O(\varepsilon^2).
\]
Asymptotics of the Capacity

Theorem (Li and Han 2016)

Assume that \( \{E_n\} \) is i.i.d. and \( S \) is a finite-type constraint of topological order \( m \). Then,

\[
C(S, \varepsilon) = C(S, 0) - \left\{ (m + 1)H(\hat{X}_0|\hat{X}_{-m}^{-1}) - \sum_{i=1}^{m} H(\hat{X}_0|\hat{X}_{-i+1}^{-1}, \hat{X}_{-i-m}^{-1}) \right\} \varepsilon + O(\varepsilon^2).
\]

Moreover, for any \( n \geq m \), \( C^{(n)}(S, \varepsilon) \) is of the same asymptotic form as above, namely,

\[
C^{(n)}(S, \varepsilon) = C(S, 0) - \left\{ (m + 1)H(\hat{X}_0|\hat{X}_{-m}^{-1}) - \sum_{i=1}^{m} H(\hat{X}_0|\hat{X}_{-i+1}^{-1}, \hat{X}_{-i-m}^{-1}) \right\} \varepsilon + O(\varepsilon^2).
\]

Proof.

Use the “explicit” formula and some convexity analysis.
A Special Case (BEC with (1, ∞)-RLL Constraint)

Theorem (Li and Han 2016)

Consider a binary erasure channel with the first-order Markovian input supported on the $(1, ∞)$-RLL constraint $S_0$ parameterized by a transition probability matrix $\Pi = \begin{bmatrix} 1 & -\theta \\ \theta & 1 \end{bmatrix}$.

Then, $I(X;Y) = (1 - \varepsilon)^2 \sum_{k=0}^{\infty} H(X_0 | X_{k-1}) \varepsilon^k$ is concave in $\theta$.

Proof. Establish the concavity of $H(X_0 | X_{k-1})$ for any $k$ through an elementary yet tedious bounding analysis of the Hessian.
A Special Case (BEC with $(1, \infty)$-RLL Constraint)

Theorem (Li and Han 2016)

Consider a binary erasure channel with the first-order Markovian input supported on the $(1, \infty)$-RLL constraint $S_0$ parameterized by a transition probability matrix

$$\Pi = \begin{bmatrix} 1 - \theta & \theta \\ 1 & 0 \end{bmatrix}.$$ 

Then, $I(X; Y) = (1 - \varepsilon)^2 \sum_{k=0}^{\infty} H(X_0|X_{-k-1})\varepsilon^k$ is concave in $\theta$. 

Proof. Establish the concavity of $H(X_0|X_{-k-1})\varepsilon^k$ for any $k$ through an elementary yet tedious bounding analysis of the Hessian.
A Special Case (BEC with \((1, \infty)\)-RLL Constraint)

Theorem (Li and Han 2016)

Consider a binary erasure channel with the first-order Markovian input supported on the \((1, \infty)\)-RLL constraint \(S_0\) parameterized by a transition probability matrix

\[
\Pi = \begin{bmatrix}
1 - \theta & \theta \\
1 & 0
\end{bmatrix}.
\]

Then, \(I(X; Y) = (1 - \varepsilon)^2 \sum_{k=0}^{\infty} H(X_0|X_{-k-1})\varepsilon^k\) is concave in \(\theta\).

Proof.
Establish the concavity of \(H(X_0|X_{-k-1})\) for any \(k\) through an elementary yet tedious bounding analysis of the Hessian. \(\square\)
A Special Case (BEC with \((1, \infty)\)-RLL Constraint)

\[ C(1)(S_0, \varepsilon) \text{ is analytic in } \varepsilon \text{ for } \varepsilon \in [0, 1) \]

with the following Taylor series expansion around \( \varepsilon = 0 \):

\[
C(1)(S_0, \varepsilon) = \sum_{n=0}^{\infty} \left( \frac{d^n G_0(\varepsilon)}{d \varepsilon^n} \bigg|_{\varepsilon=0} + \frac{d^n-1}{n} \left( G_1(\varepsilon) - 2G_0(\varepsilon) \right) \bigg|_{\varepsilon=0} + \sum_{k=2}^{n} \left( \frac{n!}{k!(n-k)!} \right) d^{n-k} \varepsilon^{n-k} \{ G_k(\varepsilon) + G_k - 2G_{k-1}(\varepsilon) \} \bigg|_{\varepsilon=0} \right) \varepsilon^n,
\]

where \( G_k(\varepsilon) = H(X_0 \mid X - k - 1)(\theta_{\text{max}}(\varepsilon)) \).

Proof. Obtain the Taylor series expansion of \( \theta_{\text{max}} \) first using our concavity result and the implicit function theorem.
A Special Case (BEC with \((1, \infty)\)-RLL Constraint)

Theorem (Li and Han 2016)

\( C^{(1)}(S_0, \varepsilon) \) is analytic in \( \varepsilon \) for \( \varepsilon \in [0, 1) \) with the following Taylor series expansion around \( \varepsilon = 0 \):

\[
C^{(1)}(S_0, \varepsilon) = \sum_{n=0}^{\infty} \left( \frac{d^n G_0(\varepsilon)}{d\varepsilon^n} \bigg|_{\varepsilon=0} + \frac{d^{n-1}(G_1(\varepsilon) - 2G_0(\varepsilon))}{d\varepsilon^{n-1}} \bigg|_{\varepsilon=0} ight)
\]

\[
+ \sum_{k=2}^{n} \binom{n}{k} \frac{d^{n-k}}{d\varepsilon^{n-k}} \left\{ (G_k(\varepsilon) + G_{k-2}(\varepsilon) - 2G_{k-1}(\varepsilon)) \right\} \bigg|_{\varepsilon=0} \varepsilon^n,
\]

where \( G_k(\varepsilon) = H(X_0|X_{-k-1})(\theta_{\text{max}}(\varepsilon)) \).
A Special Case (BEC with \((1, \infty)\)-RLL Constraint)

Theorem (Li and Han 2016)
\(C^{(1)}(S_0, \varepsilon)\) is analytic in \(\varepsilon\) for \(\varepsilon \in [0, 1)\) with the following Taylor series expansion around \(\varepsilon = 0\):

\[
C^{(1)}(S_0, \varepsilon) = \sum_{n=0}^{\infty} \left( \frac{d^n G_0(\varepsilon)}{d\varepsilon^n} \right|_{\varepsilon=0} + \frac{d^{n-1}(G_1(\varepsilon) - 2G_0(\varepsilon))}{d\varepsilon^{n-1}} \right|_{\varepsilon=0} \\
+ \sum_{k=2}^{n} \binom{n}{k} \frac{d^{n-k}}{d\varepsilon^{n-k}} \{ (G_k(\varepsilon) + G_{k-2}(\varepsilon) - 2G_{k-1}(\varepsilon)) \} \left|_{\varepsilon=0} \right) \varepsilon^n,
\]

where \(G_k(\varepsilon) = H(X_0|X_{-k-1})(\theta_{\text{max}}(\varepsilon))\).

Proof.
Obtain the Taylor series expansion of \(\theta_{\text{max}}\) first using our concavity result and the implicit function theorem.

\(\square\)
A Special Case (BEC with \((1, \infty)\)-RLL Constraint)

The algorithm proposed in [Han 2015] iteratively computes \(\{\theta_n\}\) in the following way:

\[
\theta_{n+1} = \begin{cases} 
\theta_n, & \text{if } \theta_n + a_n g_n b(\theta_n) \in [0, 1] \\
\theta_n + a_n g_n b(\theta_n), & \text{otherwise}
\end{cases}
\]

where \(g_n b(\theta_n)\) is a simulator for \(I'(X; Y)\).

Our concavity result will guarantee the algorithm converges with a proven rate of convergence.
A Special Case (BEC with $(1, \infty)$-RLL Constraint)

The algorithm proposed in [Han 2015] iteratively computes $\{\theta_n\}$ in the following way:

$$\theta_{n+1} = \begin{cases} 
\theta_n, & \text{if } \theta_n + a_n g_{n^b}(\theta_n) \notin [0, 1], \\
\theta_n + a_n g_{n^b}(\theta_n), & \text{otherwise},
\end{cases}$$

where $g_{n^b}(\theta_n)$ is a simulator for $I'(X; Y)$. 
A Special Case (BEC with \((1, \infty)\)-RLL Constraint)

The algorithm proposed in [Han 2015] iteratively computes \(\{\theta_n\}\) in the following way:

\[
\theta_{n+1} = \begin{cases} 
\theta_n, & \text{if } \theta_n + a_n g_n b(\theta_n) \notin [0, 1], \\
\theta_n + a_n g_n b(\theta_n), & \text{otherwise},
\end{cases}
\]

where \(g_n b(\theta_n)\) is a simulator for \(I'(X; Y)\).

Our concavity result will guarantee the algorithm converges with a proven rate of convergence.
A Special Case (BEC with \((1, \infty)\)-RLL Constraint)

\[
\theta_{\text{max}}(\epsilon)
\]

Figure: plot of \(\theta_{\text{max}}\) against \(\epsilon\)

Yonglong Li and Guangyue Han
The University of California, San Diego and The University of Hong Kong
A Special Case (BEC with $(1, \infty)$-RLL Constraint)

Figure: plot of $C^{(1)}(S_0, \varepsilon)$ against $\varepsilon$
Effect of Feedback on Input-Constrained Channels

It follows from our asymptotics result that

\[ C(S_0, \varepsilon) = \log \lambda - 2 \log \frac{1 + \lambda^2}{4} + O(\varepsilon^2), \]

where \( \lambda = \frac{1 + \sqrt{5}}{2}. \)

It has been observed [Sabag, Permuter and Kashyap 2016] that the capacity with feedback is

\[ C_{FB}(S_0, \varepsilon) = \log \lambda - \lambda^2 + 1 \log \lambda \cdot \varepsilon + O(\varepsilon^2), \]

Feedback may increase the capacity when the channel input has memory (even if the channel transition does not).
It follows from our asymptotics result that

\[ C(S_0, \varepsilon) = \log \lambda - \frac{2 \log 2}{1 + \lambda^2} \varepsilon + O(\varepsilon^2), \]

where \( \lambda = (1 + \sqrt{5})/2. \)
Effect of Feedback on Input-Constrained Channels

- It follows from our asymptotics result that

\[ C(S_0, \varepsilon) = \log \lambda - \frac{2 \log 2}{1 + \lambda^2 \varepsilon} + O(\varepsilon^2), \]

where \( \lambda = \frac{(1 + \sqrt{5})}{2} \).

- It has been observed [Sabag, Permuter and Kashyap 2016] that the capacity with feedback is

\[ C_{FB}(S_0, \varepsilon) = \log \lambda - \frac{\lambda^2}{\lambda^2 + 1} \log \lambda \cdot \varepsilon + O(\varepsilon^2), \]
It follows from our asymptotics result that

\[ C(S_0, \varepsilon) = \log \lambda - \frac{2 \log 2}{1 + \lambda^2 \varepsilon} + O(\varepsilon^2), \]

where \( \lambda = (1 + \sqrt{5})/2 \).

It has been observed [Sabag, Permuter and Kashyap 2016] that the capacity with feedback is

\[ C_{FB}(S_0, \varepsilon) = \log \lambda - \frac{\lambda^2}{\lambda^2 + 1} \log \lambda \cdot \varepsilon + O(\varepsilon^2), \]

Feedback may increase the capacity when the channel input has memory (even if the channel transition does not).
Concluding Remarks

Why Markov approximation?

The assumption of Markovian inputs may imply tractable mathematical properties such as analyticity, concavity, convergence and asymptotics.

As a general framework, Markov approximation can be applied to elsewhere, such as

- input-constrained BSC [Han and Marcus 2009] [Jacquet and Szpankowski 2010],
- input-constrained finite-state channels [Vontobel, Kavcic, Arnold and Loeliger 2008] [Han 2015],
- flash memory channels [Li, Kavcic and Han 2017].

How far can we go with Markov approximation?

Any other new approach?

Yonglong Li and Guangyue Han
The University of California, San Diego and The University of Hong Kong
Concluding Remarks

- Why Markov approximation?
Concluding Remarks

- Why Markov approximation?
  - The assumption of Markovian inputs may imply tractable mathematical properties such as analyticity, concavity, convergence and asymptotics.
Concluding Remarks

▶ Why Markov approximation?
  ▶ The assumption of Markovian inputs may imply tractable mathematical properties such as analyticity, concavity, convergence and asymptotics.
  ▶ As a general framework, Markov approximation can be applied to elsewhere, such as
    ▶ input-constrained BSC [Han and Marcus 2009] [Jacquet and Szpankowski 2010],
    ▶ input-constrained finite-state channels [Vontobel, Kavcic, Arnold and Loeliger 2008] [Han 2015],
    ▶ flash memory channels [Li, Kavcic and Han 2017].
Concluding Remarks

▶ Why Markov approximation?
  ▶ The assumption of Markovian inputs may imply tractable mathematical properties such as analyticity, concavity, convergence and asymptotics.
  ▶ As a general framework, Markov approximation can be applied to elsewhere, such as
    ▶ input-constrained BSC [Han and Marcus 2009] [Jacquet and Szpankowski 2010],
    ▶ input-constrained finite-state channels [Vontobel, Kavcic, Arnold and Loeliger 2008] [Han 2015],
    ▶ flash memory channels [Li, Kavcic and Han 2017].

▶ How far can we go with Markov approximation?
Concluding Remarks

- Why Markov approximation?
  - The assumption of Markovian inputs may imply tractable mathematical properties such as analyticity, concavity, convergence and asymptotics.
  - As a general framework, Markov approximation can be applied to elsewhere, such as
    - input-constrained BSC [Han and Marcus 2009] [Jacquet and Szpankowski 2010],
    - input-constrained finite-state channels [Vontobel, Kavcic, Arnold and Loeliger 2008] [Han 2015],
    - flash memory channels [Li, Kavcic and Han 2017].

- How far can we go with Markov approximation?
- Any other new approach?
Thank you!