Variable-Length Stop-Feedback Codes With Finite Optimal Decoding Times for BI-AWGN Channels

Recep Yavas*

Joint work with Hengjie Yang, Victoria Kostina*, and Richard Wesel

University of California, Los Angeles (UCLA)
*California Institute of Technology (Caltech)

2022 IEEE International Symposium on Information Theory (ISIT)

Espoo, Finland, June 2022

Outline

1. Variable-Length Stop-Feedback (VLSF) Codes with Finite Decoding Times
2. Tight Approximations to $\mathbb{P}[\iota(X^n; Y^n) \geq \gamma]$
3. Gap-Constrained Sequential Differential Optimization (SDO)
4. Summary

1. Variable-Length Stop-Feedback (VLSF) Codes with Finite Decoding Times
2. Tight Approximations to $\mathbb{P}[\imath(X^n; Y^n) \geq \gamma]$
3. Gap-Constrained Sequential Differential Optimization (SDO)
4. Summary

Feedback is Useful

Bad news: Feedback does not improve the capacity of a memoryless channel.

Variable-length transmission:

- Simplify coding schemes: e.g., Horstein scheme, posterior matching.
- Achieve better error exponents: [Burnashev, 1976]. For any $R \in [0, C]$,

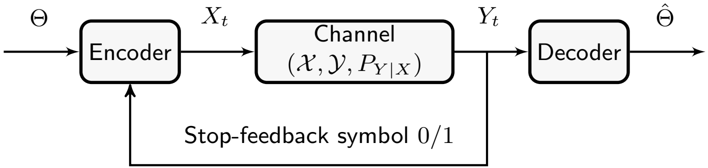
$$E(R) \triangleq \lim_{\epsilon \to 0} \frac{\log \frac{1}{\epsilon}}{\mathbb{E}[\tau_\epsilon^*]} = C_1 \left(1 - \frac{R}{C}\right). \tag{1}$$

- Achieve universality: LT codes (or fountain codes) [Luby, 2002]
- Improve first- and second-order coding rates: [Polyanskiy *et al.*, 2011]

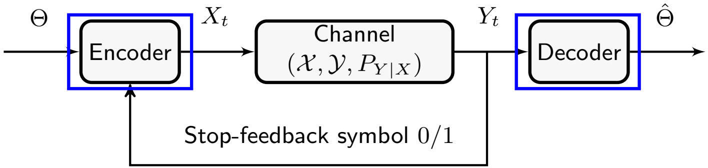
Fixed-length transmission:

- Achieve better error exponents: e.g., Schalkwijk–Kailath scheme.
- Improve second-order coding rate for compound-dispersion DMCs: [Wagner, 2020].

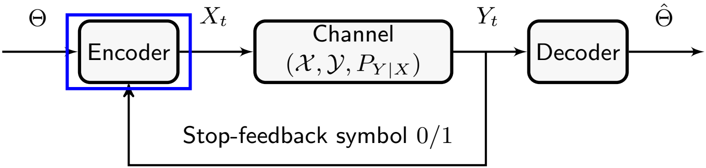
VLSF Codes with m Decoding Times



Given $l > 0$, $n_1^m \in \mathbb{N}_+^m$ with $n_1 < n_2 < \cdots < n_m$, $M \in \mathbb{N}_+$, $\epsilon \in (0, 1)$, we want to specify an (l, n_1^m, M, ϵ) VLSF code.



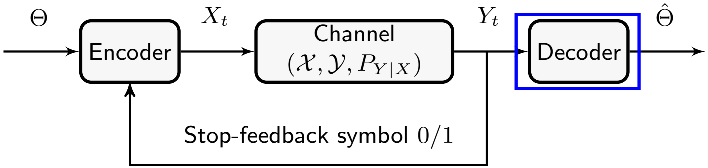
Codebook $U \in \mathcal{U}$: designed and fixed before transmission (common randomness)



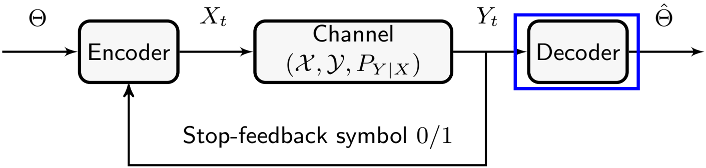
Encoding function $e_t : \mathcal{U} \times [M] \to \mathcal{X}$:

$$X_t = e_t(U, \Theta), \quad t \in \mathbb{N}_+$$

where $\Theta \sim \mathrm{Unif}\,([M])$.



Decoding function $g_t : \mathcal{U} \times \mathcal{Y}^t \to [M]$: providing the best estimate of Θ at time t, $t \in \{n_1, n_2, \ldots, n_m\}$.



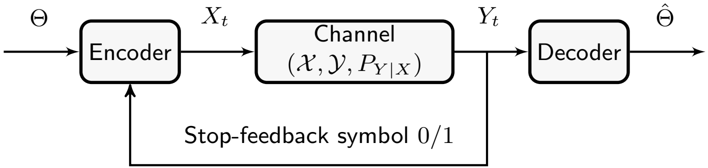
Stopping time $\tau \in \{n_i\}_{i=1}^m$: a function of filtration generated by $\{U, Y^{n_i}\}_{i=1}^m$ and must satisfy $\mathbb{E}[\tau] \leq l$.

Final decision: $\hat{\Theta} = g_\tau(Y^\tau)$

τ also has to satisfy

$$P_e \triangleq \mathbb{P}[\Theta \neq \hat{\Theta}] \leq \epsilon.$$

VLSF Codes with m Decoding Times



Goal: Determine $l^*(m, M, \epsilon) \triangleq \min\{l : \exists (l, n_1^m, M, \epsilon) \text{ VLSF code}\}$

The normal PDF and CDF:

$$\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}, \quad \Phi(x) = \int_{-\infty}^{x} \phi(t) dt, \qquad Q(x) = 1 - \Phi(x)$$

BI-AWGN channel:

- Input alphabet $\mathcal{X} = \{-\sqrt{P}, \sqrt{P}\}$, output alphabet $\mathcal{Y} = \mathbb{R}$
- Channel law: $\mathsf{P}_{Y|X}(y|x) = \phi(y - x)$

The BI-AWGN Channel and Information Density

The normal PDF and CDF:

$$\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}, \quad \Phi(x) = \int_{-\infty}^{x} \phi(t)dt, \qquad Q(x) = 1 - \Phi(x)$$

BI-AWGN channel:

- Input alphabet $\mathcal{X} = \{-\sqrt{P}, \sqrt{P}\}$, output alphabet $\mathcal{Y} = \mathbb{R}$
- Channel law: $\mathsf{P}_{Y|X}(y|x) = \phi(y - x)$

Information density:

$$\iota(x^n; y^n) \triangleq \log \frac{\mathsf{P}_{Y^n|X^n}(y^n|x^n)}{\mathsf{P}_{Y^n}(y^n)} \tag{2}$$

- If $\mathsf{P}_{X^n} = \prod_{i=1}^{n} \mathsf{P}_{X_i}$ and the channel is memoryless, $\iota(x^n; y^n) = \sum_{i=1}^{n} \iota(x_i; y_i)$.
- Let $\mathsf{P}_X = \mathsf{P}_X^*$, the unique capacity-achieving distribution. The channel capacity and dispersion are

$$C \triangleq \mathbb{E}_{\mathsf{P}_X^* \mathsf{P}_{Y|X}}[\iota(X; Y)], \tag{3}$$

$$V \triangleq \mathrm{var}_{\mathsf{P}_X^* \mathsf{P}_{Y|X}}[\iota(X; Y)]. \tag{4}$$

Theorem 1 (Polyanskiy *et al.*, 2011)

Fix $M \in \mathbb{N}_+$ *and* $\epsilon \in (0, 1/2)$. *There exists an* $(l, \mathbb{N}, M, \epsilon)$ *variable-length stop-feedback (VLSF) code for DMC with*

$$l \leq \frac{\log(M-1)}{C} + \frac{\log \frac{1}{\epsilon}}{C} + \frac{a_0}{C} \tag{5}$$

where $a_0 \triangleq \sup_{x \in \mathcal{X}, y \in \mathcal{Y}} \iota(x; y)$.

Y. Polyanskiy *et al.*, "Feedback in the non-asymptotic regime," *IEEE Trans. Inf. Theory*, Aug. 2011.

An Achievability Bound for (l, n_1^m, M, ϵ) VLSF Code

Theorem 2 (Yavas *et al.*, 2021)

Fix a constant $\gamma > 0$, *integer-valued decoding times* $n_1 < n_2 < \cdots < n_m$, *and a memoryless channel* $(\mathcal{X}, \mathcal{Y}, \mathsf{P}_{Y|X})$. *For any* $l > 0$ *and* $\epsilon \in (0,1)$, *there exists an* (l, n_1^m, M, ϵ') *VLSF code with*

$$l \leq n_m + \sum_{i=1}^{m-1} (n_i - n_{i+1}) \mathbb{P}\left[\bigcup_{j=1}^{i} \{\iota(X^{n_j}; Y^{n_j}) \geq \gamma\} \right], \tag{6}$$

$$\epsilon' \leq 1 - \mathbb{P}[\iota(X^{n_m}; Y^{n_m}) \geq \gamma] + (M-1)2^{-\gamma}, \tag{7}$$

where $\mathsf{P}_{X^{n_m}}$ *is the product of distributions of* m *subvectors of lengths* $n_i - n_{i-1}$, $i \in [m]$, *i.e.*,

$$\mathsf{P}_{X^{n_m}}(x_1^{n_m}) = \prod_{i=1}^{m} \mathsf{P}_{X_{n_{i-1}+1}^{n_i}}\left(x_{n_{i-1}+1}^{n_i}\right). \tag{8}$$

R. Yavas *et al.*, "Variable-length feedback codes with several decoding times for the Gaussian channel," *IEEE Int. Sym. Inf. Theory (ISIT)*, Jul. 2021.

By relaxing $\mathbb{P}\left[\bigcup_{j=1}^{i}\{\iota(X^{n_j};Y^{n_j})\geq\gamma\}\right]$ to $\mathbb{P}[\iota(X^{n_i};Y^{n_i})\geq\gamma]$, define

$$N(\gamma,n_1^m)\triangleq n_m+\sum_{i=1}^{m-1}(n_i-n_{i+1})\mathbb{P}[\iota(X^{n_i};Y^{n_i})\geq\gamma], \tag{9}$$

$$\begin{aligned}\mathcal{F}_m(\gamma,M,\epsilon)\triangleq\{n_1^m\in\mathbb{R}_+^m: n_{i+1}-n_i\geq 1,\forall i\in[m-1];\\ \mathbb{P}[\iota(X^{n_m};Y^{n_m})\geq\gamma]\geq 1-\epsilon+(M-1)2^{-\gamma}\}.\end{aligned} \tag{10}$$

By relaxing $\mathbb{P}\left[\bigcup_{j=1}^{i}\{\iota(X^{n_j};Y^{n_j})\geq\gamma\}\right]$ to $\mathbb{P}[\iota(X^{n_i};Y^{n_i})\geq\gamma]$, define

$$N(\gamma, n_1^m) \triangleq n_m + \sum_{i=1}^{m-1}(n_i - n_{i+1})\mathbb{P}[\iota(X^{n_i};Y^{n_i})\geq\gamma], \tag{9}$$

$$\begin{aligned}\mathcal{F}_m(\gamma, M, \epsilon) \triangleq \{n_1^m \in \mathbb{R}_+^m : n_{i+1} - n_i \geq 1, \forall i \in [m-1]; \\ \mathbb{P}[\iota(X^{n_m};Y^{n_m})\geq\gamma] \geq 1 - \epsilon + (M-1)2^{-\gamma}\}.\end{aligned} \tag{10}$$

Integer program: for a given $m \in \mathbb{N}_+$, $M \in \mathbb{N}_+$, $\epsilon \in (0,1)$, and $\gamma \geq \log\frac{M-1}{\epsilon}$,

$$\begin{aligned} \min_{n_1^m} \quad & N(\gamma, n_1^m) \\ \text{s.t.} \quad & n_1^m \in \mathcal{F}_m(\gamma, M, \epsilon) \\ & n_1^m \in \mathbb{N}_+^m. \end{aligned} \tag{11}$$

By relaxing $\mathbb{P}\left[\bigcup_{j=1}^{i}\{\iota(X^{n_j};Y^{n_j}) \geq \gamma\}\right]$ to $\mathbb{P}[\iota(X^{n_i};Y^{n_i}) \geq \gamma]$, define

$$N(\gamma, n_1^m) \triangleq n_m + \sum_{i=1}^{m-1}(n_i - n_{i+1})\mathbb{P}[\iota(X^{n_i};Y^{n_i}) \geq \gamma], \tag{9}$$

$$\begin{aligned}\mathcal{F}_m(\gamma, M,\epsilon) \triangleq \{&n_1^m \in \mathbb{R}_+^m : n_{i+1} - n_i \geq 1, \forall i \in [m-1]; \\ &\mathbb{P}[\iota(X^{n_m};Y^{n_m}) \geq \gamma] \geq 1 - \epsilon + (M-1)2^{-\gamma}\}.\end{aligned} \tag{10}$$

Integer program: for a given $m \in \mathbb{N}_+$, $M \in \mathbb{N}_+$, $\epsilon \in (0,1)$, and $\gamma \geq \log\frac{M-1}{\epsilon}$,

$$\begin{aligned}\min_{n_1^m} \quad & N(\gamma, n_1^m) \\ \text{s.t.} \quad & n_1^m \in \mathcal{F}_m(\gamma, M, \epsilon) \\ & n_1^m \in \mathbb{N}_+^m.\end{aligned} \tag{11}$$

Two-step minimization: $\min_\gamma \min_{n_1^m} N(\gamma, n_1^m)$

Outline

Revisit to the Central Limit Theorem (CLT)

For BI-AWGN channel,

- $\iota(X;Y) = 1 - \log\left(1 + e^{-2XY}\right)$ is continuous.
- $\iota(X^n;Y^n)$ is a sum of i.i.d. $\iota(X;Y)$.

H. Wang *et al.*, "An information density approach to analyzing and optimizing incremental redundancy with feedback", *IEEE Int. Sym. Inf. Theory (ISIT)*, Jun. 2017.

Revisit to the Central Limit Theorem (CLT)

For BI-AWGN channel,

- $\iota(X;Y) = 1 - \log\left(1 + e^{-2XY}\right)$ is continuous.
- $\iota(X^n;Y^n)$ is a sum of i.i.d. $\iota(X;Y)$.

CLT: If $W_1, W_2, \ldots, W_n$ i.i.d. with zero mean, variance σ^2, define the standardized sum

$$S_n \triangleq \frac{\sum_{i=1}^n W_i}{\sigma\sqrt{n}}. \tag{12}$$

Then, $\mathbb{P}[S_n \leq x] \to \Phi(x)$ as $n \to \infty$.

H. Wang *et al.*, "An information density approach to analyzing and optimizing incremental redundancy with feedback", *IEEE Int. Sym. Inf. Theory (ISIT)*, Jun. 2017.

Revisit to the Central Limit Theorem (CLT)

For BI-AWGN channel,

- $\iota(X;Y) = 1 - \log\left(1 + e^{-2XY}\right)$ is continuous.
- $\iota(X^n;Y^n)$ is a sum of i.i.d. $\iota(X;Y)$.

CLT: If $W_1, W_2, \ldots, W_n$ i.i.d. with zero mean, variance σ^2, define the standardized sum

$$S_n \triangleq \frac{\sum_{i=1}^n W_i}{\sigma\sqrt{n}}. \tag{12}$$

Then, $\mathbb{P}[S_n \leq x] \to \Phi(x)$ as $n \to \infty$.

Gaussian Model [Wang *et al.*, 2017]: for n sufficiently large,

$$\mathbb{P}[\iota(X^n;Y^n) \geq \gamma] \approx Q\left(\frac{\gamma - nC}{\sqrt{nV}}\right) \tag{13}$$

H. Wang *et al.*, "An information density approach to analyzing and optimizing incremental redundancy with feedback", *IEEE Int. Sym. Inf. Theory (ISIT)*, Jun. 2017.

For BI-AWGN channel,

- $\iota(X;Y) = 1 - \log\left(1 + e^{-2XY}\right)$ is continuous.
- $\iota(X^n;Y^n)$ is a sum of i.i.d. $\iota(X;Y)$.

CLT: If $W_1, W_2, \ldots, W_n$ i.i.d. with zero mean, variance σ^2, define the standardized sum

$$S_n \triangleq \frac{\sum_{i=1}^n W_i}{\sigma\sqrt{n}}. \tag{12}$$

Then, $\mathbb{P}[S_n \le x] \to \Phi(x)$ as $n \to \infty$.

Gaussian Model [Wang *et al.*, 2017]: for n sufficiently large,

$$\mathbb{P}[\iota(X^n;Y^n) \ge \gamma] \approx Q\left(\frac{\gamma - nC}{\sqrt{nV}}\right) \tag{13}$$

Question: What if n is small?

H. Wang *et al.*, "An information density approach to analyzing and optimizing incremental redundancy with feedback", *IEEE Int. Sym. Inf. Theory (ISIT)*, Jun. 2017.

Where Does CLT Come From?

Consider the characteristic function of W,

$$\chi_W(t) \triangleq \mathbb{E}\left[e^{\mathrm{i}tW}\right]. \tag{14}$$

Where Does CLT Come From?

Consider the characteristic function of W,

$$\chi_W(t) \triangleq \mathbb{E}\left[e^{\mathrm{i}tW}\right]. \tag{14}$$

We want to obtain a Taylor series for $\chi_W(t)$ using

$$\chi_W(t) = \exp\left(\kappa_1(\mathrm{i}t) + \frac{1}{2!}\kappa_2(\mathrm{i}t)^2 + \frac{1}{3!}\kappa_3(\mathrm{i}t)^3 + \cdots\right). \tag{15}$$

From $\kappa_1 = 0$, $\kappa_2 = \sigma^2$ ($\{\kappa_i\}_{i=1}^{\infty}$ known as cumulants),

Where Does CLT Come From?

Consider the characteristic function of W,

$$\chi_W(t) \triangleq \mathbb{E}\left[e^{\mathrm{i}tW}\right]. \tag{14}$$

We want to obtain a Taylor series for $\chi_W(t)$ using

$$\chi_W(t) = \exp\left(\kappa_1(\mathrm{i}t) + \frac{1}{2!}\kappa_2(\mathrm{i}t)^2 + \frac{1}{3!}\kappa_3(\mathrm{i}t)^3 + \cdots\right). \tag{15}$$

From $\kappa_1 = 0$, $\kappa_2 = \sigma^2$ ($\{\kappa_i\}_{i=1}^{\infty}$ known as cumulants),

$$\chi_{S_n}(t) = \mathbb{E}\left[\exp\left(\mathrm{i}\frac{tn^{-\frac{1}{2}}}{\sigma}\sum_{i=1}^{n} W_i\right)\right] = \left(\chi_W\left(\frac{tn^{-\frac{1}{2}}}{\sigma}\right)\right)^n$$

Where Does CLT Come From?

Consider the characteristic function of W,

$$\chi_W(t) \triangleq \mathbb{E}\left[e^{\mathrm{i}tW}\right]. \tag{14}$$

We want to obtain a Taylor series for $\chi_W(t)$ using

$$\chi_W(t) = \exp\left(\kappa_1(\mathrm{i}t) + \frac{1}{2!}\kappa_2(\mathrm{i}t)^2 + \frac{1}{3!}\kappa_3(\mathrm{i}t)^3 + \cdots\right). \tag{15}$$

From $\kappa_1 = 0$, $\kappa_2 = \sigma^2$ ($\{\kappa_i\}_{i=1}^{\infty}$ known as cumulants),

$$\begin{aligned}\chi_{S_n}(t) &= \mathbb{E}\left[\exp\left(\mathrm{i}\frac{tn^{-\frac{1}{2}}}{\sigma}\sum_{i=1}^{n} W_i\right)\right] = \left(\chi_W\left(\frac{tn^{-\frac{1}{2}}}{\sigma}\right)\right)^n \\ &= \exp\left[n\left(\kappa_1\left(\frac{\mathrm{i}t}{\sigma}\right)n^{-\frac{1}{2}} + \frac{1}{2}\kappa_2\left(\frac{\mathrm{i}t}{\sigma}\right)^2 n^{-1} + \frac{1}{6}\kappa_3\left(\frac{\mathrm{i}t}{\sigma}\right)^3 n^{-\frac{3}{2}} + \cdots\right)\right]\end{aligned}$$

Where Does CLT Come From?

Consider the characteristic function of W,

$$\chi_W(t) \triangleq \mathbb{E}\left[e^{\mathrm{i}tW}\right]. \tag{14}$$

We want to obtain a Taylor series for $\chi_W(t)$ using

$$\chi_W(t) = \exp\left(\kappa_1(\mathrm{i}t) + \frac{1}{2!}\kappa_2(\mathrm{i}t)^2 + \frac{1}{3!}\kappa_3(\mathrm{i}t)^3 + \cdots\right). \tag{15}$$

From $\kappa_1 = 0$, $\kappa_2 = \sigma^2$ ($\{\kappa_i\}_{i=1}^{\infty}$ known as cumulants),

$$\begin{aligned}
\chi_{S_n}(t) &= \mathbb{E}\left[\exp\left(\mathrm{i}\frac{tn^{-\frac{1}{2}}}{\sigma}\sum_{i=1}^{n} W_i\right)\right] = \left(\chi_W\left(\frac{tn^{-\frac{1}{2}}}{\sigma}\right)\right)^n \\
&= \exp\left[n\left(\kappa_1\left(\frac{\mathrm{i}t}{\sigma}\right)n^{-\frac{1}{2}} + \frac{1}{2}\kappa_2\left(\frac{\mathrm{i}t}{\sigma}\right)^2 n^{-1} + \frac{1}{6}\kappa_3\left(\frac{\mathrm{i}t}{\sigma}\right)^3 n^{-\frac{3}{2}} + \cdots\right)\right] \\
&= e^{-\frac{t^2}{2}}\exp\left[\frac{1}{6}\kappa_3\left(\frac{\mathrm{i}t}{\sigma}\right)^3 n^{-\frac{1}{2}} + \cdots\right]
\end{aligned}$$

Where Does CLT Come From?

Consider the characteristic function of W,

$$\chi_W(t) \triangleq \mathbb{E}\left[e^{\mathrm{i}tW}\right]. \tag{14}$$

We want to obtain a Taylor series for $\chi_W(t)$ using

$$\chi_W(t) = \exp\left(\kappa_1(\mathrm{i}t) + \frac{1}{2!}\kappa_2(\mathrm{i}t)^2 + \frac{1}{3!}\kappa_3(\mathrm{i}t)^3 + \cdots\right). \tag{15}$$

From $\kappa_1 = 0$, $\kappa_2 = \sigma^2$ ($\{\kappa_i\}_{i=1}^{\infty}$ known as cumulants),

$$\begin{aligned}
\chi_{S_n}(t) &= \mathbb{E}\left[\exp\left(\mathrm{i}\frac{tn^{-\frac{1}{2}}}{\sigma}\sum_{i=1}^{n} W_i\right)\right] = \left(\chi_W\left(\frac{tn^{-\frac{1}{2}}}{\sigma}\right)\right)^n \\
&= \exp\left[n\left(\kappa_1\left(\frac{\mathrm{i}t}{\sigma}\right)n^{-\frac{1}{2}} + \frac{1}{2}\kappa_2\left(\frac{\mathrm{i}t}{\sigma}\right)^2 n^{-1} + \frac{1}{6}\kappa_3\left(\frac{\mathrm{i}t}{\sigma}\right)^3 n^{-\frac{3}{2}} + \cdots\right)\right] \\
&= e^{-\frac{t^2}{2}}\exp\left[\frac{1}{6}\kappa_3\left(\frac{\mathrm{i}t}{\sigma}\right)^3 n^{-\frac{1}{2}} + \cdots\right] \\
&\to e^{-\frac{t^2}{2}} \quad (\text{as } n \to \infty) \quad (\text{The characteristic function for } \mathcal{N}(0,1))
\end{aligned}$$

Edgeworth Expansion

Applying Taylor series to $\exp(x)$, $\chi_{S_n}(t)$ can be written as

$$\chi_{S_n}(t) = e^{-\frac{t^2}{2}} \exp\left(\sum_{j=1}^{\infty} \frac{\kappa_{j+2}(\mathrm{i}t)^{j+2}}{(j+2)!\sigma^{j+2}} n^{-\frac{j}{2}}\right) = e^{-\frac{t^2}{2}} \left(1 + \sum_{j=1}^{\infty} n^{-\frac{j}{2}} r_j(\mathrm{i}t)\right) \tag{16}$$

Edgeworth Expansion

Applying Taylor series to $\exp(x)$, $\chi_{S_n}(t)$ can be written as

$$\chi_{S_n}(t) = e^{-\frac{t^2}{2}} \exp\left(\sum_{j=1}^{\infty} \frac{\kappa_{j+2}(\mathrm{i}t)^{j+2}}{(j+2)!\sigma^{j+2}} n^{-\frac{j}{2}}\right) = e^{-\frac{t^2}{2}} \left(1 + \sum_{j=1}^{\infty} n^{-\frac{j}{2}} r_j(\mathrm{i}t)\right) \quad (16)$$

Applying inverse Fourier transform to $\chi_{S_n}(t)$ and integrating, we obtain

$$\mathbb{P}[S_n \leq x] = \Phi(x) + \phi(x) \sum_{j=1}^{\infty} n^{-\frac{j}{2}} p_j(x) \quad \text{(Edgeworth series)} \quad (17)$$

where $p_j(x)$ requires cumulants $\kappa_3, \kappa_4, \ldots, \kappa_{j+2}$.

Applying Taylor series to $\exp(x)$, $\chi_{S_n}(t)$ can be written as

$$\chi_{S_n}(t) = e^{-\frac{t^2}{2}} \exp\left(\sum_{j=1}^{\infty} \frac{\kappa_{j+2}(\mathrm{i}t)^{j+2}}{(j+2)!\sigma^{j+2}} n^{-\frac{j}{2}} \right) = e^{-\frac{t^2}{2}} \left(1 + \sum_{j=1}^{\infty} n^{-\frac{j}{2}} r_j(\mathrm{i}t) \right) \tag{16}$$

Applying inverse Fourier transform to $\chi_{S_n}(t)$ and integrating, we obtain

$$\mathbb{P}[S_n \leq x] = \Phi(x) + \phi(x) \sum_{j=1}^{\infty} n^{-\frac{j}{2}} p_j(x) \quad \text{(Edgeworth series)} \tag{17}$$

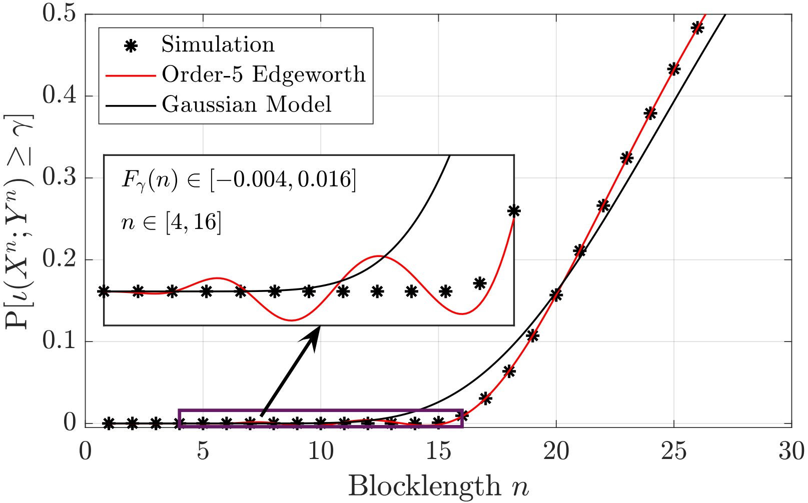
where $p_j(x)$ requires cumulants $\kappa_3, \kappa_4, \dots, \kappa_{j+2}$.

Edgeworth expansion: If $\mathbb{E}[|W|^{s+2}] < \infty$ for some $s \in \mathbb{N}_+$ and $\limsup_{|t|\to\infty} |\chi_W(t)| < 1$ (Cramér's condition), then,

$$\mathbb{P}[S_n \leq x] = \Phi(x) + \phi(x) \sum_{j=1}^{s} n^{-\frac{j}{2}} p_j(x) + o\left(n^{-\frac{s}{2}}\right), \tag{18}$$

where $p_j(x)$ requires cumulants $\kappa_3, \kappa_4, \dots, \kappa_{j+2}$ of W, and $p_1(x) = -\frac{\kappa_3}{6\sigma^3}(x^2 - 1)$.

An Example of Order-5 Edgeworth Expansion



Parameters setup: BI-AWGN channel at 0.2 dB, $\gamma = 13.62$.

Petrov Expansion

Petrov expansion: If $x \geq 0$, $x = o(\sqrt{n})$, and $\mathbb{E}[e^{tW}] < \infty$ for $|t| < H$ for some $H > 0$,

$$\mathbb{P}[S_n \leq x] = 1 - Q(x) \exp\left\{ \frac{x^3}{\sqrt{n}} \Lambda\left(\frac{x}{\sqrt{n}}\right) \right\} \left[1 + O\left(\frac{x+1}{\sqrt{n}}\right) \right], \tag{19}$$

$$\mathbb{P}[S_n \leq -x] = Q(x) \exp\left\{ \frac{-x^3}{\sqrt{n}} \Lambda\left(\frac{-x}{\sqrt{n}}\right) \right\} \left[1 + O\left(\frac{x+1}{\sqrt{n}}\right) \right], \tag{20}$$

where $\Lambda(t) = \sum_{k=0}^{\infty} a_k t^k$ is called the Cramér series.

$$\Lambda^{[3]}(t) = \frac{\kappa_3}{6\sigma^3} + \frac{\kappa_4\kappa_2 - 3\kappa_3^2}{24\sigma^6} t + \frac{\kappa_5\kappa_2^2 - 10\kappa_4\kappa_3\kappa_2 + 15\kappa_3^3}{120\sigma^9} t^2 \tag{21}$$

For BI-AWGN channel: $F_\gamma(n)$: a function to approximate $\mathbb{P}[\iota(X^n;Y^n) \geq \gamma]$.

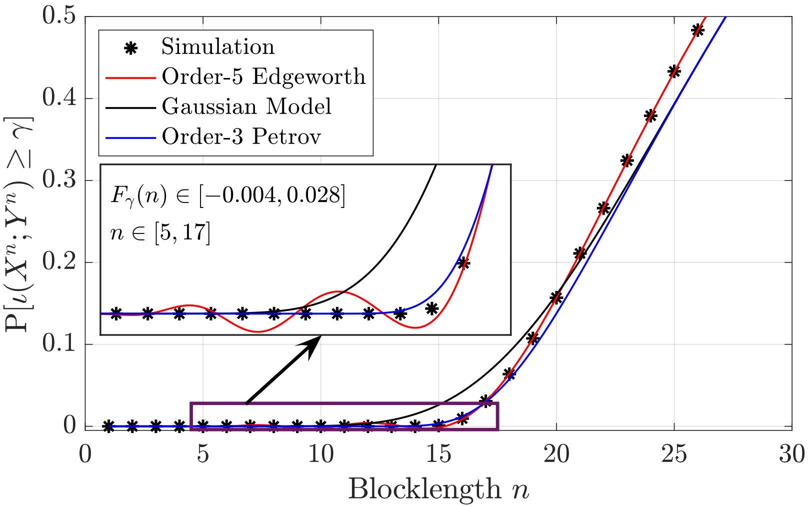
$$F_\gamma(n) = \begin{cases} Q\left(x(n)\right) - \phi(x(n)) \sum_{j=1}^{5} n^{-\frac{j}{2}} p_j(x(n)), \; n > n^* \\ Q(x(n)) \exp\left\{ \frac{x^3(n)}{\sqrt{n}} \Lambda^{[3]} \left(\frac{x(n)}{\sqrt{n}} \right) \right\}, \; 0 \leq n \leq n^*, \end{cases} \tag{22}$$

where

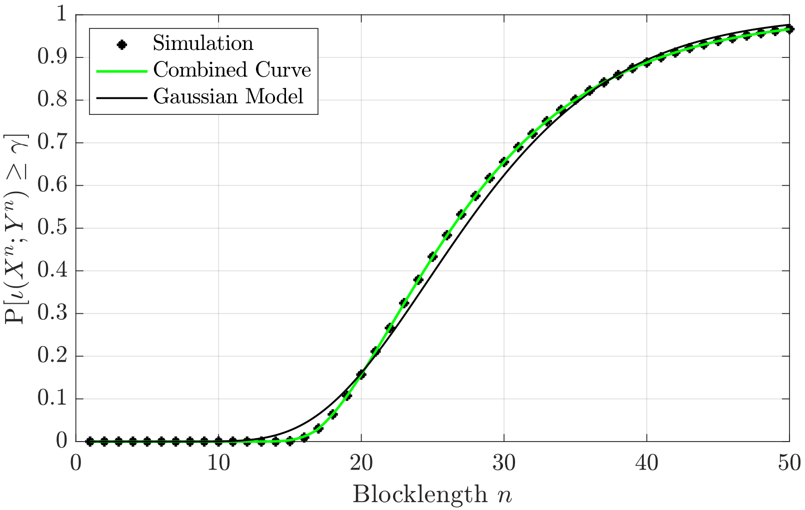
$$x(n) \triangleq \frac{\gamma - nC}{\sqrt{nV}}$$

and n^* is the largest zero of the Edgeworth expansion.

Edgeworth and Petrov Expansions



Parameters setup: BI-AWGN channel at 0.2 dB, $\gamma = 13.62$, $n^* = 16.84$.



Parameters setup: BI-AWGN channel at 0.2 dB, $\gamma = 13.62$, $n^* = 16.84$.

Outline

1. Variable-Length Stop-Feedback (VLSF) Codes with Finite Decoding Times
2. Tight Approximations to $\mathbb{P}[\iota(X^n; Y^n) \geq \gamma]$
3. Gap-Constrained Sequential Differential Optimization (SDO)
4. Summary

Relaxed program: for a given $m \in \mathbb{N}_+$, $M \in \mathbb{N}_+$, $\epsilon \in (0,1)$, and $\gamma \geq \log \frac{M-1}{\epsilon}$,

$$\begin{aligned} \min_{n_1^m} \quad & N(\gamma, n_1^m) \\ \text{s.t.} \quad & n_1^m \in \mathcal{F}_m(\gamma, M, \epsilon) \end{aligned} \tag{23}$$

Theorem 3 (Gap-constrained SDO procedure)

Fix a memoryless channel $(\mathcal{X}, \mathcal{Y}, P_{Y|X})$ *for which* $\iota(X;Y)$ *is continuous and* $\mathbb{P}[\iota(X^n;Y^n) \geq \gamma]$ *is increasing and differentiable. For a given* $m \in \mathbb{N}_+$, $M \in \mathbb{N}_+$, $\epsilon \in (0,1)$, *and* $\gamma \geq \log \frac{M-1}{\epsilon}$, *the optimal real-valued decoding times* $n_1^*, n_2^*, \ldots, n_m^*$ *for the relaxed program* (23) *satisfy*

$$n_m^* = F_\gamma^{-1}\left(1 - \epsilon + (M-1)2^{-\gamma}\right), \tag{24}$$

$$n_{i+1}^* = n_i^* + \max\left\{1, \frac{F_\gamma(n_i^*) - F_\gamma(n_{i-1}^*) - \lambda_{i-1}}{f_\gamma(n_i^*)}\right\}, \tag{25}$$

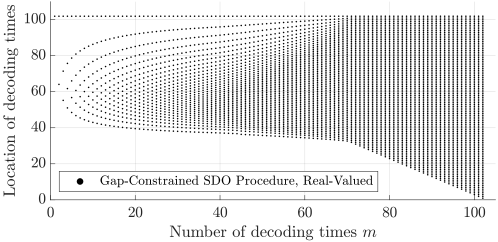
$$\lambda_i = \max\{\lambda_{i-1} + f_\gamma(n_i^*) - F_\gamma(n_i^*) + F_\gamma(n_{i-1}^*), 0\}, \tag{26}$$

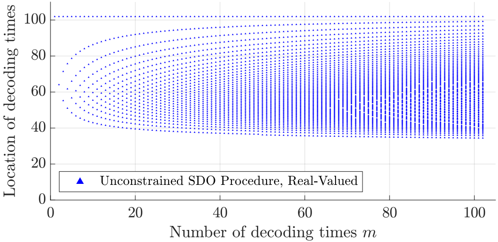
where $i \in [m-1]$, $f_\gamma(n) = \frac{dF_\gamma(n)}{dn}$, $\lambda_0 \triangleq 0$, *and* $n_0^* \triangleq 0$.

Unconstrained SDO procedure:

$$n_{i+1}^* = n_i^* + \frac{F_\gamma(n_i^*) - F_\gamma(n_{i-1}^*)}{f_\gamma(n_i^*)}, \quad i \in [m-1] \tag{27}$$

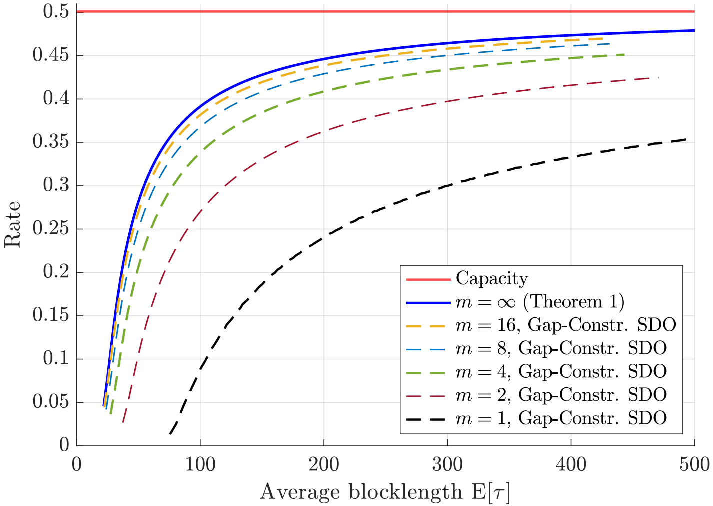
Comparison of Optimal Real-Valued Decoding Times





Parameters setup: BI-AWGN channel at 0.2 dB with $C = 0.5$. $\epsilon = 10^{-2}$, $k = 20$, $\gamma = 27.64$, $n_m^* = 101.97$.

Achievability Bounds for (l, n_1^m, M, ϵ) VLSF Codes over BI-AWGN Channel



Parameters setup: BI-AWGN channel at 0.2 dB with $C = 0.5$. $\epsilon = 10^{-3}$.

Outline

1. Variable-Length Stop-Feedback (VLSF) Codes with Finite Decoding Times
2. Tight Approximations to $\mathbb{P}[\imath(X^n; Y^n) \geq \gamma]$
3. Gap-Constrained Sequential Differential Optimization (SDO)
4. Summary

Summary

1) By leveraging both Edgeworth and Petrov expansions, we develop tight approximations to $\mathbb{P}[\iota(X^n; Y^n) \geq \gamma]$ that is accurate for all blocklengths of interest.

Summary

1) By leveraging both Edgeworth and Petrov expansions, we develop tight approximations to $\mathbb{P}[\iota(X^n; Y^n) \geq \gamma]$ that is accurate for all blocklengths of interest.

2) For the relaxed program, we develop the gap-constrained SDO procedure that solves the optimal real-valued decoding times $n_1^*, \ldots, n_m^*$ satisfying the gap constraint.

Summary

1) By leveraging both Edgeworth and Petrov expansions, we develop tight approximations to $\mathbb{P}[\iota(X^n;Y^n) \geq \gamma]$ that is accurate for all blocklengths of interest.
2) For the relaxed program, we develop the gap-constrained SDO procedure that solves the optimal real-valued decoding times $n_1^*, \ldots, n_m^*$ satisfying the gap constraint.
3) Numerical evaluations show that Polyanskiy's VLSF achievability bound, which assumes $m = \infty$, can be closely approached with a small number of decoding times.

In Our Recent Journal Paper

- We investigated the BSC and BEC cases and confirmed 3) still holds.

H. Yang, R. C. Yavas, V. Kostina, and R. D. Wesel, "Variable-Length Coding for Binary-Input Channels With Limited Stop Feedback," *arXiv: 2205.15399*, May 2022.

In Our Recent Journal Paper

- We investigated the BSC and BEC cases and confirmed 3) still holds.
- We developed the discrete SDO procedure that only requires the tail probability at integer-valued decoding times.

H. Yang, R. C. Yavas, V. Kostina, and R. D. Wesel, "Variable-Length Coding for Binary-Input Channels With Limited Stop Feedback," *arXiv: 2205.15399*, May 2022.

In Our Recent Journal Paper

- We investigated the BSC and BEC cases and confirmed 3) still holds.
- We developed the discrete SDO procedure that only requires the tail probability at integer-valued decoding times.
- For the BEC, we consider a systematic transmission followed by the random linear fountain coding (ST-RLFC). The ST-RLFC scheme provides a new VLSF bound that outperforms the state-of-the-art VLSF bound developed by Devassy.

H. Yang, R. C. Yavas, V. Kostina, and R. D. Wesel, "Variable-Length Coding for Binary-Input Channels With Limited Stop Feedback," *arXiv: 2205.15399*, May 2022.

Thank you!