Two-terminal source coding with common sum reconstruction

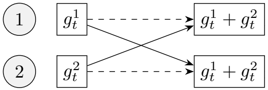
Tharindu Adikari and Stark Draper
University of Toronto

ISIT
Espoo, Finland

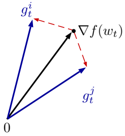
28 June 2022

Motivation

Some algorithms require distributed sum (or mean) computation.



Basic building block



Two-terminal stochastic gradient descent (SGD).

Two terminals want to update

$$w_{t+1} = w_t - \eta_t(g_t^i + g_t^j)/2.$$

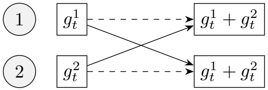
(1) Need only $g_t^i + g_t^j$, no need to recover g_t^i and g_t^j separately.

(2) g_t^i and g_t^j are correlated.

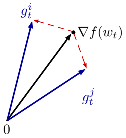
(3) Two terminals must recover identical sums to (remain) synchronized.

Motivation

Some algorithms require distributed sum (or mean) computation.



Basic building block



Two-terminal stochastic gradient descent (SGD).

Two terminals want to update

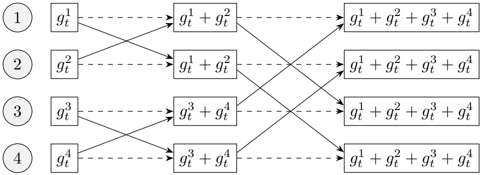
$$w_{t+1} = w_t - \eta_t(g_t^i + g_t^j)/2.$$

(1) Need only $g_t^i + g_t^j$, no need to recover g_t^i and g_t^j separately.

(2) g_t^i and g_t^j are correlated.

(3) Two terminals must recover identical sums to (remain) synchronized.

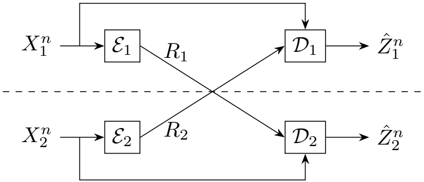
Algorithms exist for extending to more than 2 terminals.



- E.g., Butterfly all-reduce.
- Recursively apply the basic building block.
- 4 terminals take $\log 4 = 2$ rounds for sum computation.
- Other examples: power iteration, k-means.

Two-terminal source coding with "Common Sum Reconstruction" (CSR)

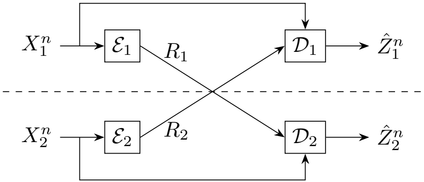
Problem setup:



- Two decoders each want to form two estimates ($\hat{Z}_1^n$ and $\hat{Z}_2^n$) of the sum $Z^n = X_1^n + X_2^n$.
- Want $\Pr(\hat{Z}_1^n \neq \hat{Z}_2^n)$ to be small.
- The decoders may use X_1^n and X_2^n as side information at $\mathcal{D}_1$ and $\mathcal{D}_2$.

Two-terminal source coding with "Common Sum Reconstruction" (CSR)

Problem setup:



- Two decoders each want to form two estimates ($\hat{Z}_1^n$ and $\hat{Z}_2^n$) of the sum $Z^n = X_1^n + X_2^n$.
- Want $\Pr(\hat{Z}_1^n \neq \hat{Z}_2^n)$ to be small.
- The decoders may use X_1^n and X_2^n as side information at $\mathcal{D}_1$ and $\mathcal{D}_2$.

CSR-achievability:

- A scalar triple (R_1, R_2, D) is CSR-achievable if there exists a coding scheme such that
  - $\mathbb{E}[d(Z^n, \hat{Z}_1^n)] \leq D$ and $\mathbb{E}[d(Z^n, \hat{Z}_2^n)] \leq D$.
  - $\Pr(\hat{Z}_1^n \neq \hat{Z}_2^n)$ can be made arbitrarily small.
- $\mathcal{R}_{\text{CSR}} \subset \mathbb{R}^3$ is the set of all CSR-achievable triples.

Our contributions

Focus on binary/hamming case.

- $X_1, X_2 \in \mathbb{F}_2$, and $d(\cdot,\cdot)$ is Hamming distortion.
- $X_1 + X_2$ is modulo-two sum.
- (X_1, X_2) is a doubly symmetric binary source (DSBS).

X_1 \ X_2	0	1
0	$\frac{(1-p)}{2}$	$\frac{p}{2}$
1	$\frac{p}{2}$	$\frac{(1-p)}{2}$

Joint distribution of a DSBS(p).

Develop two inner bounds to $\mathcal{R}_{\text{CSR}}$ (achievability results).

- Inner bound 1 – Based on Steinberg's common reconstruction (CR) problem[1].
- Inner bound 2 – Based on the lossy version of Körner-Marton's modulo-two sum (LKM) problem[2].

Develop an outer bound to $\mathcal{R}_{\text{CSR}}$ (a converse result).

- Based on Wyner-Ziv's source coding with side information problem[3].

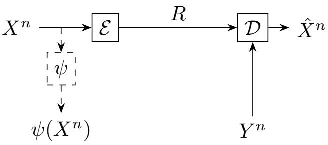
[1] Yossef Steinberg. "Coding and common reconstruction". In: *IEEE Trans. Inform. Theory* 55.11 (2009), pp. 4995–5010

[2] Janos Körner and Katalin Marton. "How to encode the modulo-two sum of binary sources". In: *IEEE Trans. Inform. Theory* 25.2 (1979), pp. 219–221

[3] Aaron Wyner and Jacob Ziv. "The rate-distortion function for source coding with side information at the decoder". In: *IEEE Trans. Inform. Theory* 22.1 (1976), pp. 1–10

Inner bound 1 – Based on Steinberg's CR problem

Steinberg's CR problem

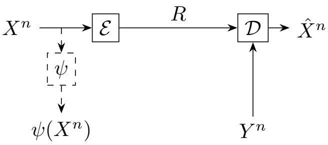


- Similar setup to source coding with side information. Want $\psi(X^n) = \hat{X}^n$ w.h.p. in addition.
- Achievable rate distortion region is known.

$$R_{\mathsf{CR}}(D) = \begin{cases} H(p * D) - H(D) & \text{if } 0 \leq D \leq \frac{1}{2}. \\ 0 & \text{if } \frac{1}{2} \leq D. \end{cases}.$$

Inner bound 1 – Based on Steinberg's CR problem

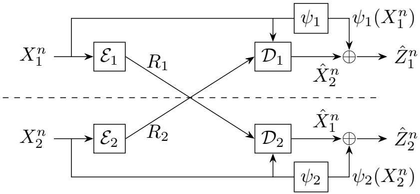
Steinberg's CR problem



- Similar setup to source coding with side information. Want $\psi(X^n) = \hat{X}^n$ w.h.p. in addition.
- Achievable rate distortion region is known.

$$R_{\mathsf{CR}}(D) = \begin{cases} H(p * D) - H(D) & \text{if } 0 \le D \le \frac{1}{2}. \\ 0 & \text{if } \frac{1}{2} \le D. \end{cases}.$$

Application to CSR problem



$\hat{Z}_1^n = \psi_1(X_1^n) + \hat{X}_2^n$, $\hat{Z}_2^n = \hat{X}_1^n + \psi_2(X_2^n)$

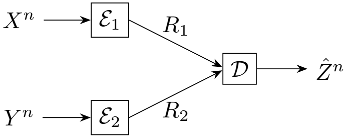
- Employ two parallel Steinberg systems.

$$\mathcal{R}_{\mathsf{A}} = \left\{ (R_1, R_2, D) \;\middle|\; \begin{array}{l} 0 \le D_1, D_2 \le \frac{1}{2}, \\ R_1 \ge R_{\mathsf{CR}}(D_1), \\ R_2 \ge R_{\mathsf{CR}}(D_2), \\ D \ge D_1 * D_2 \end{array} \right\},$$

$$\mathcal{R}_{\mathsf{B}} = \{(R_1, R_2, D) \mid R_1 \ge 0, R_2 \ge 0, D \ge p\}.$$

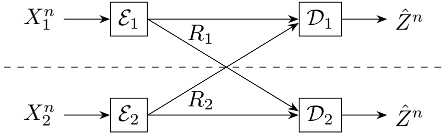
- Then, convex hull $\mathrm{conv}(\mathcal{R}_{\mathsf{A}} \cup \mathcal{R}_{\mathsf{B}}) \subseteq \mathcal{R}_{\mathsf{CSR}}$.

Lossy Körner-Marton problem



$\mathcal{D}$ estimates the sum $Z^n = X^n + Y^n$ as $\hat{Z}^n$.

Application of LKM to CSR problem

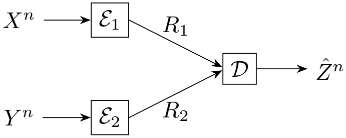


Both $\mathcal{D}_1$ and $\mathcal{D}_2$ are identical LKM decoders.

[4] Sung Hoon Lim et al. "Towards an Algebraic Network Information Theory: Distributed Lossy Computation of Linear Functions". In: *Proc. Int. Symp. Inform. Theory*. IEEE. 2019, pp. 1827–1831

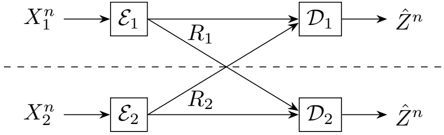
Inner bound 2 – Based on Lossy Körner-Marton (LKM) problem

Lossy Körner-Marton problem



$\mathcal{D}$ estimates the sum $Z^n = X^n + Y^n$ as $\hat{Z}^n$.

Application of LKM to CSR problem



Both $\mathcal{D}_1$ and $\mathcal{D}_2$ are identical LKM decoders.

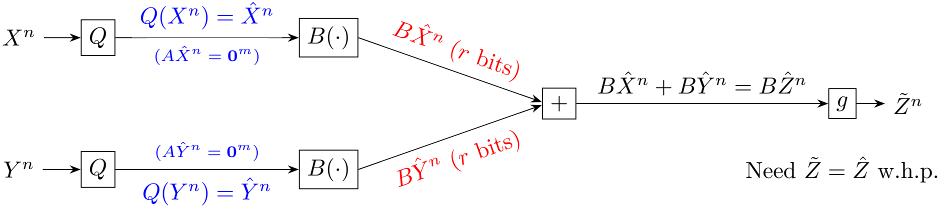
- A scalar triple (R_1, R_2, D) is LKM-achievable if there exists a coding scheme such that $\mathbb{E}[d(Z^n, \hat{Z}^n)] \leq D$.
- $\mathcal{R}_{\text{LKM}} \subset \mathbb{R}^3$ is the set of all LKM-achievable triples. Note $\mathcal{R}_{\text{LKM}} \subseteq \mathcal{R}_{\text{CSR}}$.
- Lossless version ($D = 0$): Tight bound is known (Körner and Marton 1979).
- Lossy version ($D > 0$): Joint typicality-based encoding/decoding to obtain an inner bound[4].

[4] Sung Hoon Lim et al. "Towards an Algebraic Network Information Theory: Distributed Lossy Computation of Linear Functions". In: *Proc. Int. Symp. Inform. Theory*. IEEE. 2019, pp. 1827–1831

Claim:
All triples in $\{(R_1, R_2, D) \mid R_1, R_2 \geq H(p * q * q) - H(q), D \geq q * q, 0 \leq q \leq \frac{1}{2}\}$ are LKM-achievable.

Proof sketch:

- $U \in \mathbb{F}_2^{m \times n}$: parity check matrix of a 'good' linear source code with avg. distortion q.
- $V \in \mathbb{F}_2^{r \times n}$: parity check matrix of a 'good' linear channel code for BSC($p * q * q$).



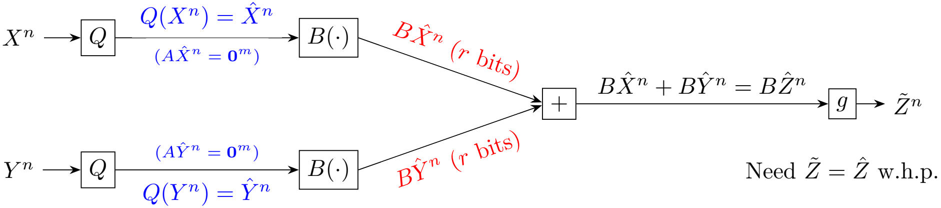
An achievability result for LKM problem

Claim:

All triples in $\{(R_1, R_2, D) \mid R_1, R_2 \geq H(p*q*q) - H(q), D \geq q*q, 0 \leq q \leq \frac{1}{2}\}$ are LKM-achievable.

Proof sketch:

- $U \in \mathbb{F}_2^{m\times n}$: parity check matrix of a 'good' linear source code with avg. distortion q.
- $V \in \mathbb{F}_2^{r\times n}$: parity check matrix of a 'good' linear channel code for BSC($p*q*q$).

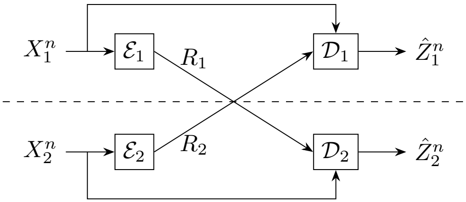


- Can show that we only need $r > H(p*q*q) - H(q)$.
- Resulting average distortion $\mathbb{E}[d(Z, \hat{Z})] \leq q*q$. (Recall $Z = X + Y$, $\hat{Z}^n = \hat{X}^n + \hat{Y}^n$.)

Outer bound with Wyner-Ziv problem

Waive the common reconstruction (CR) constraint in CSR

- CSR problem reduces to *two* source coding with side information (Wyner-Ziv) problems.
- This relaxation yields an outer bound (a converse result).



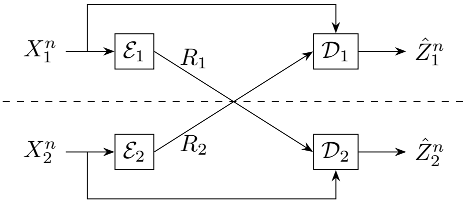
$\hat{Z}_1^n = X_1^n + \hat{X}_2^n$ (use X_1^n as side info. to decode $\hat{X}_2^n$).

$\hat{Z}_2^n = \hat{X}_1^n + X_2^n$ (use X_2^n as side info. to decode $\hat{X}_1^n$).

Outer bound with Wyner-Ziv problem

Waive the common reconstruction (CR) constraint in CSR

- CSR problem reduces to *two* source coding with side information (Wyner-Ziv) problems.
- This relaxation yields an outer bound (a converse result).



$\hat{Z}_1^n = X_1^n + \hat{X}_2^n$ (use X_1^n as side info. to decode $\hat{X}_2^n$).

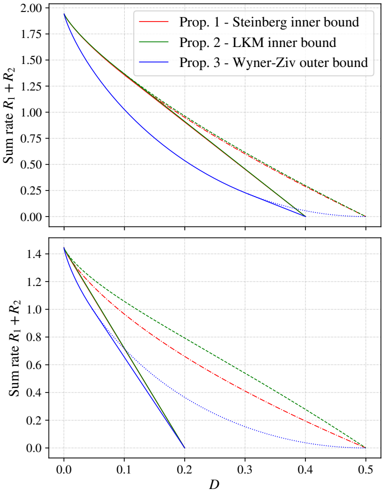
$\hat{Z}_2^n = \hat{X}_1^n + X_2^n$ (use X_2^n as side info. to decode $\hat{X}_1^n$).

- $R_{\mathsf{WZ}}(D) = \text{l.c.e}(g(D))$, the lower convex envelope of $g(D) = \begin{cases} H(p * D) - H(D) & \text{if } 0 \le D < p, \\ 0 & \text{if } p \le D. \end{cases}$
- $\mathcal{R}_{\mathsf{CSR}} \subseteq \{(R_1, R_2, D) \mid R_1, R_2 \ge R_{\mathsf{WZ}}(D), D \ge 0\}$.

Comparison of bounds

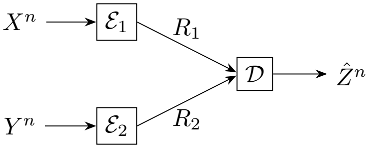
Special case $R_1 = R_2$:

- (X_1, X_2) is a DSBS(p) with $p = 0.4$ (top) and $p = 0.2$ (bottom).
- The straight portion of the solid lines (i.e., the large D portion) is tangent to the corresponding non-solid lines. This results from time-sharing with the zero-sum-rate point $(p, 0)$.
- The inner and outer bounds get tighter as p decreases.
- Steinberg inner bound is better than LKM inner bound.



Open for exploration:

- Find a distribution for (X_1, X_2) where LKM inner bound outperforms Steinberg inner bound.
- Improve inner bounds by correlating the quantization errors.

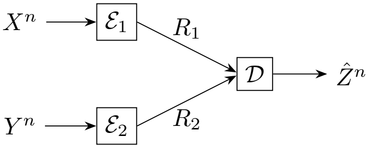


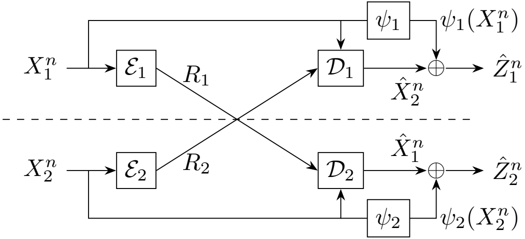
$$\begin{aligned}\hat{Z}^n &= \hat{X}^n + \hat{Y}^n \\ &= (X^n + W_X^n) + (Y^n + W_Y^n) \\ &= X^n + Y^n + (\underbrace{W_X^n + W_Y^n}_{\text{correlate to cancel}}).\end{aligned}$$

Conclusion and next steps

Open for exploration:

- Find a distribution for (X_1, X_2) where LKM inner bound outperforms Steinberg inner bound.
- Improve inner bounds by correlating the quantization errors.





$$
\begin{aligned}
\hat{Z}^n &= \hat{X}^n + \hat{Y}^n \\
&= (X^n + W_X^n) + (Y^n + W_Y^n) \\
&= X^n + Y^n + (\underbrace{W_X^n + W_Y^n}_{\text{correlate to cancel}}).
\end{aligned}
$$

$$
\begin{aligned}
\hat{Z}_1^n &= \psi_1(X_1^n) + \hat{X}_2^n \\
&= \hat{X}_1^n + \hat{X}_2^n \qquad \text{(w.h.p)} \\
&= X_1^n + X_2^n + (\underbrace{W_1^n + W_2^n}_{\text{correlate to cancel}}).
\end{aligned}
$$

Thank you!.