Error Exponents for Decentralized Detection in Feedback Architectures

Wee Peng Tay[1] John Tsitsiklis[2]

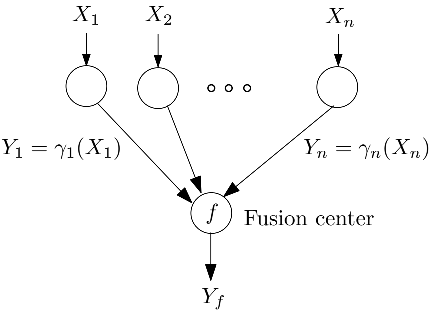
[1]Nanyang Technological University

[2]Massachusetts Institute of Technology

ICASSP 2011

Outline

Decentralized Detection

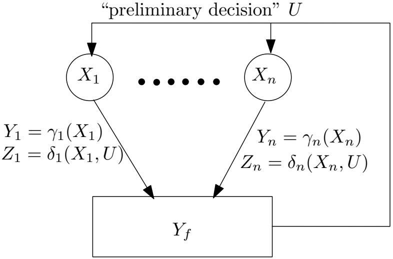


- Parallel configuration
- Binary hypothesis testing:
  $H_0 : X_k \sim \mathbb{P}_0$
  $H_1 : X_k \sim \mathbb{P}_1$
- $\gamma_k : \mathcal{X} \mapsto \{0, \ldots, d-1\}$

- Find optimal strategy $\{\gamma_i\}$ to minimize error criteria.
- Widely studied by many authors, e.g. Tenney, Sandell, Varshney, Tong, etc.
- X_k assumed to be independent under both hypotheses.
- Without independence assumption, NP hard (Tsitsiklis & Athans, 85)

Feedback in Decentralized Detection

- Some/all sensors have information about some/other sensors' messages.
- Form messages based on own observation and extra information. Example:



- Sensor messages are now NOT independent.
- What can we say about the optimal error exponent and strategies?

Optimality without Feedback

- Find best strategy $\{(\gamma_i, \delta_i)\}_{i=1}^n$ to minimize $P_n(error)$.
- Using likelihood ratio quantizers are optimal.
- Alhakeem & Varshney 96 provided PBPO solutions for a feedback architecture with memory.
- No closed form, needs numerical computation.

Difficult problem even if there is no feedback.

- Solve a system of coupled equations, as many equations as there are thresholds!
- Optimal thresholds can be different even with i.i.d. assumption.
- Analytically intractable for large n
  $\Rightarrow$ Detection performance of most networks unknown, comparison of different networks impossible.

Optimality without Feedback

- Find best strategy $\{(\gamma_i, \delta_i)\}_{i=1}^{n}$ to minimize $P_n(error)$.
- Using likelihood ratio quantizers are optimal.
- Alhakeem & Varshney 96 provided PBPO solutions for a feedback architecture with memory.
- No closed form, needs numerical computation.

Difficult problem even if there is no feedback.

- Solve a system of coupled equations, as many equations as there are thresholds!
- Optimal thresholds can be different even with i.i.d. assumption.
- Analytically intractable for large n
  $\Rightarrow$ Detection performance of most networks unknown, comparison of different networks impossible.

Bayesian Asymptotics

- Maximize rate of error probability decay
- error exponent $= \limsup_{n\to\infty} \frac{1}{n} \log P_n(error)$

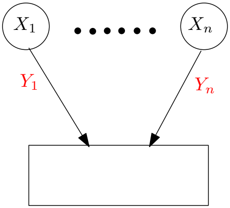
For parallel configuration (Tsitsiklis 88), for i.i.d. observations,

$$\mathcal{E}_p^* = \limsup_{n\to\infty} \frac{1}{n} \log \inf_{\{\gamma_i\}} P_n(error) = \inf_{\gamma} \min_{s\in[0,1]} \log \mathbb{E}_0 \left[\left(\frac{d\mathbb{P}_1^\gamma}{d\mathbb{P}_0^\gamma} \right)^s \right]$$

Let all sensors use γ^* that minimizes Chernoff exponent.

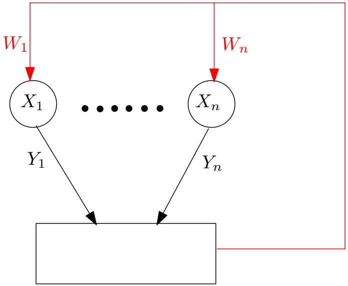
Can we get similar "nice" results if there is feedback?

Feedback: Two-message Architecture



- Sensor k sends $Y_k = \gamma_k(X_k)$.
- Fusion center feedbacks $W_k = f_k(Y_1, \ldots, Y_{k-1}, Y_{k+1}, \ldots, Y_n)$ to sensor k.
- Sensor k forms new message $Z_k = \delta_k(X_k, W_k)$.
- Fusion center makes final decision $Y_f = \gamma_f(Y_1, \ldots, Y_n, Z_1, \ldots, Z_n)$.
- Full feedback: $W_k = (Y_1, \ldots, Y_{k-1}, Y_{k+1}, \ldots, Y_n)$
- Restricted feedback: $W_k \in \{0, \ldots, d-1\}$. Assume fusion center retains $Y_1, \ldots, Y_n$.

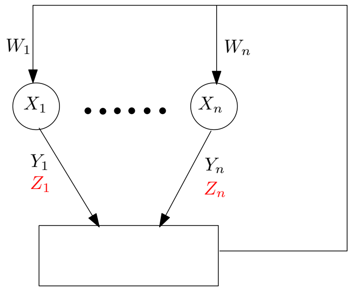
Feedback: Two-message Architecture



- Sensor k sends $Y_k = \gamma_k(X_k)$.
- Fusion center feedbacks $W_k = f_k(Y_1, \ldots, Y_{k-1}, Y_{k+1}, \ldots, Y_n)$ to sensor k.
- Sensor k forms new message $Z_k = \delta_k(X_k, W_k)$.

- Fusion center makes final decision $Y_f = \gamma_f(Y_1, \ldots, Y_n, Z_1, \ldots, Z_n)$.
- Full feedback: $W_k = (Y_1, \ldots, Y_{k-1}, Y_{k+1}, \ldots, Y_n)$
- Restricted feedback: $W_k \in \{0, \ldots, d-1\}$. Assume fusion center retains $Y_1, \ldots, Y_n$.

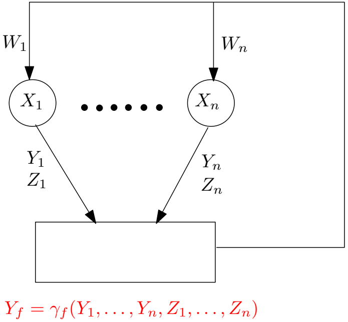
Feedback: Two-message Architecture



- Sensor k sends $Y_k = \gamma_k(X_k)$.
- Fusion center feedbacks $W_k = f_k(Y_1, \ldots, Y_{k-1}, Y_{k+1}, \ldots, Y_n)$ to sensor k.
- Sensor k forms new message $Z_k = \delta_k(X_k, W_k)$.

- Fusion center makes final decision $Y_f = \gamma_f(Y_1, \ldots, Y_n, Z_1, \ldots, Z_n)$.
- Full feedback: $W_k = (Y_1, \ldots, Y_{k-1}, Y_{k+1}, \ldots, Y_n)$
- Restricted feedback: $W_k \in \{0, \ldots, d-1\}$. Assume fusion center retains $Y_1, \ldots, Y_n$.

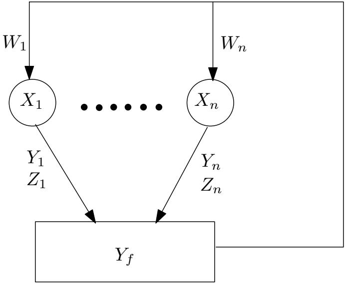
Feedback: Two-message Architecture



- Sensor k sends $Y_k = \gamma_k(X_k)$.
- Fusion center feedbacks $W_k = f_k(Y_1, \ldots, Y_{k-1}, Y_{k+1}, \ldots, Y_n)$ to sensor k.
- Sensor k forms new message $Z_k = \delta_k(X_k, W_k)$.

- Fusion center makes final decision $Y_f = \gamma_f(Y_1, \ldots, Y_n, Z_1, \ldots, Z_n)$.
- Full feedback: $W_k = (Y_1, \ldots, Y_{k-1}, Y_{k+1}, \ldots, Y_n)$
- Restricted feedback: $W_k \in \{0, \ldots, d-1\}$. Assume fusion center retains $Y_1, \ldots, Y_n$.

Feedback: Two-message Architecture



- Sensor k sends $Y_k = \gamma_k(X_k)$.
- Fusion center feedbacks $W_k = f_k(Y_1, \ldots, Y_{k-1}, Y_{k+1}, \ldots, Y_n)$ to sensor k.
- Sensor k forms new message $Z_k = \delta_k(X_k, W_k)$.

- Fusion center makes final decision $Y_f = \gamma_f(Y_1, \ldots, Y_n, Z_1, \ldots, Z_n)$.
- Full feedback: $W_k = (Y_1, \ldots, Y_{k-1}, Y_{k+1}, \ldots, Y_n)$
- Restricted feedback: $W_k \in \{0, \ldots, d-1\}$. Assume fusion center retains $Y_1, \ldots, Y_n$.

Error Exponents

Proposition

$\mathcal{E}^*_{two}(full) = \mathcal{E}^*_{two}(restricted) = \mathcal{E}^*_p$, *where*

$$\mathcal{E}^*_p = \inf_{\gamma,\delta} \min_{s\in[0,1]} \log \mathbb{E}_0 \left[(\ell_{10}(\gamma(X_1), \delta(X_1)))^s \right].$$

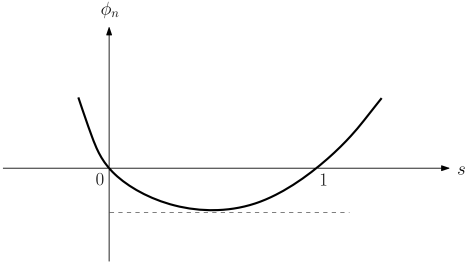
Proof I.

- Let log MGF $\phi_n(s) = \log \mathbb{E}_0 \left[\exp(s\mathcal{L}^{(n)}_{10}) \right]$.
- $\mathcal{E}^*_{two}(full) \leq \mathcal{E}^*_{two}(restricted) \leq \mathcal{E}^*_p \;\Rightarrow\;$ suffices to show $\mathcal{E}^*_{two}(full) \geq \mathcal{E}^*_p$.

Let $s^*_n = \arg\min_{s\in(0,1)} \phi_n(s)$. $\therefore \phi'_n(s^*_n) = 0 \; \forall n$.
From Cramèr bound,

$$\max_{j=0,1} P_{n,j} \geq \frac{1}{4} \exp\left(\phi_n(s^*_n) - \sqrt{2\phi''_n(s^*_n)} \right)$$

Bounds on MGF



Proposition

(i) *For all* $s \in [0, 1]$, $n\mathbb{E}_0 [\log \ell_{10}(X_1)] \leq \phi_n'(s) \leq n\mathbb{E}_1 [\log \ell_{10}(X_1)]$.

(ii) *Let* $s_n \in (0, 1)$ *s.t.* $\phi_n'(s_n) = t$. *Then, there exists a constant* C *such that for all* n, *we have* $\phi_n''(s_n) \leq nC$.

Bounds on MGF

Uniform bounds on ϕ_n, ϕ'_n and ϕ''_n can be obtained by bounding with the corresponding quantities involving the unquantized versions.

Lemma

Suppose $\varphi : (0, \infty) \mapsto \mathbb{R}$ *is a convex function. Then for any function* γ*, we have*

$$\mathbb{E}_j \left[\varphi \left(\ell_{ij}(\gamma(X)) \right) \right] \leq \mathbb{E}_j \left[\varphi \left(\ell_{ij}(X) \right) \right].$$

Proof.

$\ell_{ij}(\gamma(X)) = \mathbb{E}_j[\ell_{ij}(X) \mid \mathscr{F}^\gamma]$, apply Jensen's inequality. □

"KL divergence after quantization is at most that before quantization"

Error Exponents

Proof II.

$$\max_{j=0,1} P_{n,j} \geq \frac{1}{4} \exp\left(\phi_n(s_n^*) - \sqrt{2\phi_n''(s_n^*)}\right)$$
$$\geq \exp(\phi_n(s_n^*) - C\sqrt{n})$$

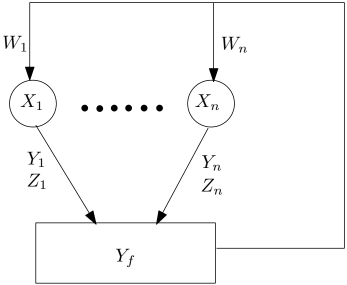
Need to show $\phi_n(s_n^*) \geq n\mathcal{E}_p^*$ ... can be done by conditioning on the first messages $Y_1, \ldots, Y_n$.

$$\limsup_{n\to\infty} \frac{1}{n} \log P_e = \limsup_{n\to\infty} \frac{1}{n} \log \max_{j=0,1} P_{n,j} \geq \mathcal{E}_p^*.$$

□

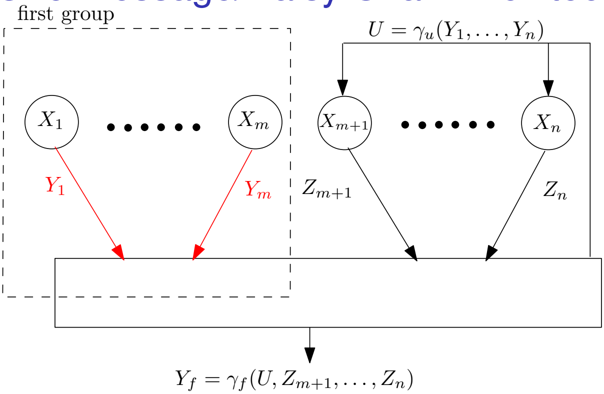
Limited Memory

- We assumed the fusion center retains memory of $Y_1, \ldots, Y_n$. What happens if it has limited memory and can only retain a compressed version?



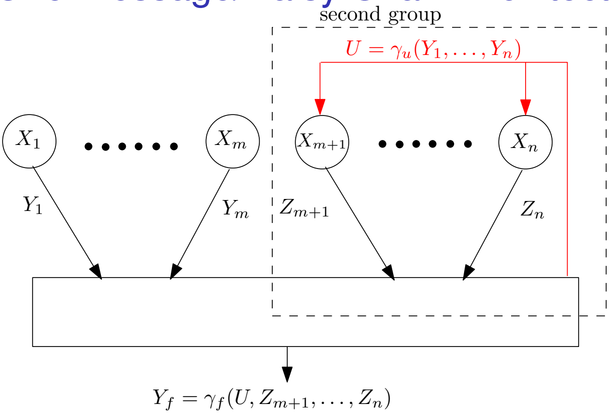
- Apparently, a difficult problem ...
  Shalaby & Papmarcou '92 uses a method of types approach, but assumes finite alphabet observations, identical quantization functions, constraint on feedback messages etc.

One-Message/Daisy Chain Architecture



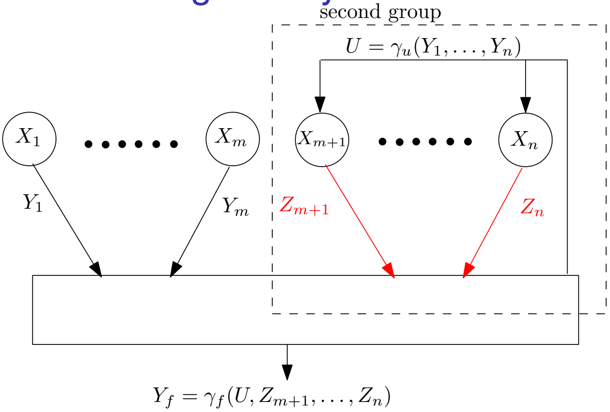
- First group sensor k sends $Y_k = \gamma_k(X_k)$.
- Fusion center feedbacks $U = \gamma_u(Y_1, \ldots, Y_n)$ to all second group sensors.
- Second group sensor k forms message $Z_k = \delta_k(X_k, U)$.
- Fusion center makes final decision $Y_f = \gamma_f(U, Z_{m+1}, \ldots, Z_n)$.

One-Message/Daisy Chain Architecture



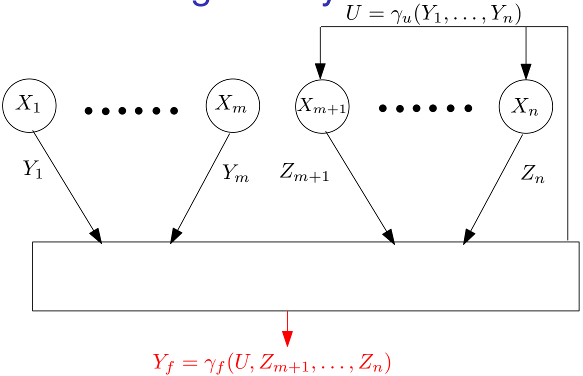
- First group sensor k sends $Y_k = \gamma_k(X_k)$.
- Fusion center feedbacks $U = \gamma_u(Y_1, \ldots, Y_n)$ to all second group sensors.
- Second group sensor k forms message $Z_k = \delta_k(X_k, U)$.
- Fusion center makes final decision $Y_f = \gamma_f(U, Z_{m+1}, \ldots, Z_n)$.

One-Message/Daisy Chain Architecture



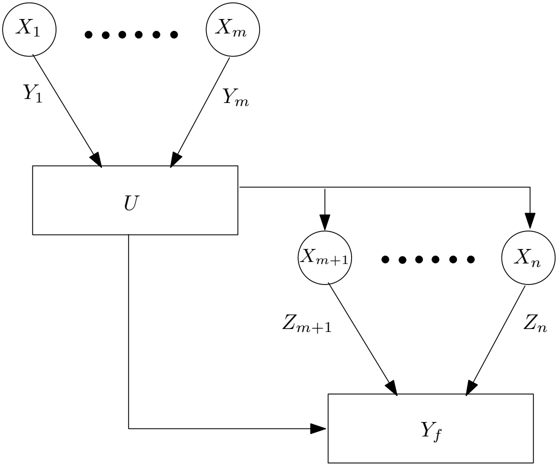
- First group sensor k sends $Y_k = \gamma_k(X_k)$.
- Fusion center feedbacks $U = \gamma_u(Y_1, \ldots, Y_n)$ to all second group sensors.
- Second group sensor k forms message $Z_k = \delta_k(X_k, U)$.
- Fusion center makes final decision $Y_f = \gamma_f(U, Z_{m+1}, \ldots, Z_n)$.

One-Message/Daisy Chain Architecture



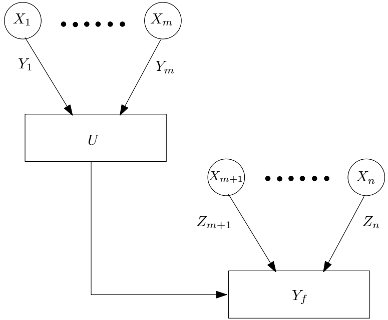
- First group sensor k sends $Y_k = \gamma_k(X_k)$.
- Fusion center feedbacks $U = \gamma_u(Y_1, \ldots, Y_n)$ to all second group sensors.
- Second group sensor k forms message $Z_k = \delta_k(X_k, U)$.
- Fusion center makes final decision $Y_f = \gamma_f(U, Z_{m+1} \ldots, Z_n)$.

One-Message/Daisy Chain Architecture



Types of Feedback

- Full feedback: $U = (Y_1, \ldots, Y_n)$.
- Restricted feedback: $U = \gamma_u(Y_1, \ldots, Y_n) \in \{0, \ldots, d-1\}$
  – fusion center limited memory
- No feedback: $U \equiv 0$ – same as a tree.



- $\mathcal{E}^*_{dc}(full) \leq \mathcal{E}^*_{dc}(restricted) \leq \mathcal{E}^*_{tree}$.

Full Feedback Error Exponent

Proposition

- $\mathcal{E}^*_{dc}(full) = \mathcal{E}^*_{p}$.
- *Feedback not useful - sensors in second group ignores U.*
- *All sensors use the same quantizer.*

Fenchel-Legendre Transform

If receive $\gamma(X_1), \gamma(X_2), \ldots$ at fusion center, and use log likelihood ratio test with threshold $t \in (-\mathrm{D}(\mathbb{P}_0^\gamma || \mathbb{P}_1^\gamma), \mathrm{D}(\mathbb{P}_1^\gamma || \mathbb{P}_0^\gamma))$, then

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}_j(error) = -\Lambda_j^*(\gamma, t),$$

where the rate function is the Fenchel-Legendre Transform of the log moment generating function,

$$\Lambda_j^*(\gamma, t) = \sup_{s\in\mathbb{R}} \left\{ st - \log \mathbb{E}_j \left[e^{s \log \ell_{10}(\gamma(X_1))} \right] \right\}.$$

Restricted Feedback Error Exponent

Let r be the fraction of sensors in first group.

Proposition

$\mathcal{E}_p^* < \mathcal{E}_{dc}^*(restricted) \leq \mathcal{E}_{tree}^*$, where

$$\mathcal{E}_{dc}^* = -(1-r) \sup_{\substack{\gamma,\delta^0,\delta^1 \\ t\in\mathbb{R}}} \min\left\{\Lambda_0^*\left(\delta^0, \frac{r}{1-r}\Lambda_1^*(\gamma,t)\right), \Lambda_1^*\left(\delta^1, -\frac{r}{1-r}\Lambda_0^*(\gamma,t)\right)\right\}$$

Sensors in first group use the same quantizer γ.
Sensors in second group use the same quantizer $\delta(\cdot, U) = \delta^U(\cdot)$.

- Restricted feedback architecture is strictly worse than parallel configuration because fusion center keeps only a summary of the observations of first group of sensors.

Restricted Feedback Error Exponent

- Feedback can improve error exponent, compared to a tree configuration. Specific examples can be constructed.
- Under special conditions, feedback does not improve performance.

Proposition

Suppose δ is optimal second stage sensor quantizer for tree configuration, and $\Lambda_1^*(\delta, t) = \Lambda_0^*(\delta, -t)$ for all t. Then,

$$\mathcal{E}_{dc}^* = \mathcal{E}_{tree}^* = -(1-r)\sup_{\gamma} \Lambda_0^*\left(\delta, \frac{r}{1-r}\Lambda_1^*(\gamma, 0)\right).$$

I.e., no loss in optimality if sensors ignore the feedback message.

Conclusion

- Feedback does not improve error exponent in **binary** hypothesis testing, under various network architectures if fusion center has access to full information.
- Memory at fusion center is cheap, and performance gain due to feedback may not justify increase in communication costs.
- Feedback can improve performance in restricted feedback daisy chain architecture. We provide a characterization for the optimal error exponent.

Open Problem

How about M**-ary** hypothesis testing in Bayesian formulation?
Without feedback: divide sensors into $\binom{M}{2}$ groups, each group uses same quantization function.
With restricted feedback (1-bit):

- Divide hypotheses into two groups. Form preliminary decision and feedbacks to sensors.
- First message: sensors use same quantization function.
- Second message: sensors make use of preliminary decision, divided into $\binom{M/2}{2}$ groups.

Is this optimal??