Channel Polarization Through the Lens of Blackwell Measures

N. Goela, M. Raginsky – Fall 2020

Summary of Theoretical Results

N. Goela and M. Raginsky, **Channel Polarization Through the Lens of Blackwell Measures**,
IEEE Transactions on Information Theory, vol. 66, no. 10, pp. 6222-6241, October 2020.

Real-Valued Functionals of BICs

Any measurable function $f: [0,1] \Rightarrow \mathbb{R}$ induces a **channel functional** $\mathbf{I}_f(\mathbf{W})$ on the collection of binary-input, memoryless, channels (BICs). We will focus on discrete BICs.

$$\mathrm{I}_f(W) = \int_{[0,1]} f\,(\mathrm{dm}_W) = \mathbf{E}[f(S)]$$

$$S \sim \mathrm{m}_W$$

m_W denotes the Blackwell measure of channel W

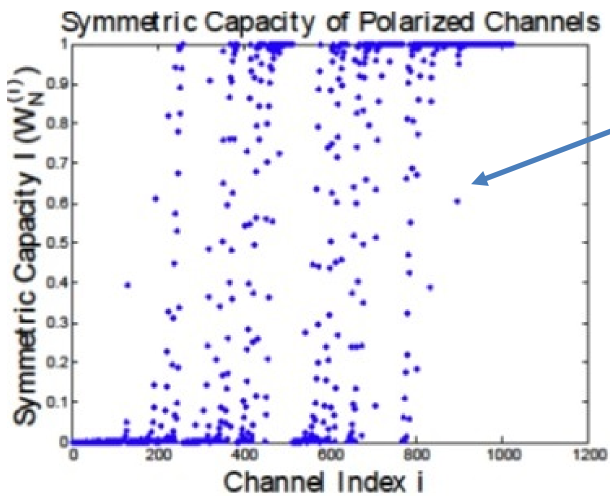
https://en.wikipedia.org/wiki/David_Blackwell

Real-Valued Functionals of BICs

Measurable Function $f:[0,1] \Rightarrow \mathbb{R}$	**Functional** $\mathrm{I}_f(W)$	**Description**
$f(s) = 1 - h_2(s)$	$I(W)$	Mutual Information
$f(s) = s(1+\log_2 s)^r + \bar{s}(1+\log_2 \bar{s})^r$	$M_r(W)$	Moments of Information Density
$f(s) = 2\sqrt{s(1-s)}$	$Z(W)$	Bhattacharyya Parameter
$f(s) = 2s^\alpha(1-s)^{1-\alpha}$	$\mathrm{H}_\alpha(W)$	Hellinger Affinity
$f(s) = 2^{-\rho}\left(s^{\frac{1}{1+\rho}} + (1-s)^{\frac{1}{1+\rho}}\right)^{1+\rho}$	$e^{-E_0(\rho,W)}$	Gallager E_0 function
$f(s) = (\lambda \wedge \bar{\lambda}) - ((2\bar{\lambda}s) \wedge (2\bar{s}\lambda))$	$B_\lambda(W)$	Bayesian Information Gain
$f(s) = \|2s - 1\|$	$1 - 2P_{e,ML}(W)$	$1-2$(Prob of ML Decoding Error)
$f(s) = (2s-1)^2$	$\rho^2_{max}(W)$	Squared Maximal Correlation

Polarization of Channels

After several steps of polarization, transformed channels become either **“perfect”** channels with capacity nearly 1 bit or completely **opaque channels** with capacity nearly 0 bits.



$\mathbf{N} = 2^{10}$

Each dot represents a communication channel!

[1] *E. Arikan*, **Channel Polarization: A Method for Constructing Capacity-Achieving Codes for Symmetric Binary-Input Memoryless Channels**, IEEE Transactions on Information Theory, vol. 55, no. 7, pp. 3051--3073, July 2009.

One-Step Channel Polarization

Definition

$$\mathrm{I}_f(W_1 \boxast W_2) \le \mathrm{I}_f(W_1) \wedge \mathrm{I}_f(W_2) \le \mathrm{I}_f(W_1) \vee \mathrm{I}_f(W_2) \le \mathrm{I}_f(W_1 \circledast W_2)$$

Theorem 5

For arbitrary **symmetric BICs** W_1 , W_2, the above phenomenon of one-step polarization holds **for all convex-**∪ $f: [0,1] \Rightarrow \mathbb{R}$

Several Different Proofs

(1) Follows as a consequence of the Blackwell-Sherman-Stein theorem.

(2) Follows from the definitions of one-step polarized channels, Jensen's inequality, and symmetry.

(3) Equivalent proofs of Blackwell ordering of channels in the literature of density evolution.

Martingale Conditions

Random Channel Polarization Process: $\{\mathrm{I}_f(W_n)\}_{n=0}^{n=\infty}$

Consider any class of BICs closed under the polar transforms, such that $\mathrm{I}_f(\cdot)$ is bounded. If one of the following conditions is true for all pairs of arbitrary BICs (W_1, W_2) in the class, then the polarization process $\{\mathrm{I}_f(W_n)\}_{n=0}^{n=\infty}$ is a MG, sub-MG, or super-MG:

$$\mathrm{I}_f(W_1 \boxast W_2) + \mathrm{I}_f(W_1 \circledast W_2) \quad = \quad \mathrm{I}_f(W_1) + \mathrm{I}_f(W_2)$$

$\geq$ (sub-MG)

$\leq$ (super-MG)

Martingale Conditions

Theorem 6 – Martingales

For the class of **symmetric BICs**, the martingale conditions hold **if and only if** they hold on the **class of BSCs**. The function $f: [0,1] \Rightarrow \mathbb{R}$ may be non-convex.

Corollary 4 – Functional Inequalities

If $f(s) = f(\bar{s})$, then it suffices to show for all $p, q \in \left[0, \frac{1}{2}\right]$:

$$f(p \star q) + (1 - p \star q) f\left(\frac{pq}{1 - p \star q}\right) + (p \star q) f\left(\frac{\bar{p}q}{p \star q}\right) \; \boxed{\begin{matrix} = \\ \geq \\ \leq \end{matrix}} \; \begin{matrix} f(p) + f(q) \\ \text{(sub-MG)} \\ \text{(super-MG)} \end{matrix}$$

Note if $f(s) = f(\bar{s})$, $\quad f\left(\frac{\bar{p}q}{p \star q}\right) = f\left(\frac{p\bar{q}}{p \star q}\right)$

Proof

Exploit theorem by Land and Huber establishing the channel decomposition of symmetric BICs into compound “mixture” of BSCs.

Example – Maximal Correlation

Definition

$$\rho_m(X;Y) = \max_{(f(X),g(Y))\in\mathcal{S}} \mathbb{E}\left(f(X)g(Y)\right)$$

$\mathcal{S}$ is the collection of real-valed random variables such that
$\mathbb{E}f(X) = \mathbb{E}g(Y) = 0$
$\mathbb{E}f^2(X) = \mathbb{E}g^2(Y) = 1$

If either X or Y is binary-valued, then

$$\rho_m^2(X;Y) = \left[\sum_{x,y} \frac{(P_{XY}(x,y))^2}{P_X(x)P_Y(y)}\right] - 1$$

Define $\rho_m^2(W) = \rho_m^2(X;Y)$ for uniform input distribution:
$P_X(x) = \frac{1}{2}$ *and* $P_{Y|X} = W$

Example – Maximal Correlation

If $f(s) = (2s-1)^2$, then $\mathrm{I}_f(W) = \rho_m^2(W)$

Theorem 5 applies for all symmetric BICs since f is **convex-**∪

$$\mathrm{I}_f(W_1 \boxast W_2) \le \mathrm{I}_f(W_1) \wedge \mathrm{I}_f(W_2) \le \mathrm{I}_f(W_1) \vee \mathrm{I}_f(W_2) \le \mathrm{I}_f(W_1 \circledast W_2)$$

Theorem 6 and Corollary 4 apply for class of symmetric BICs

$$\mathrm{I}_f(W_1 \boxast W_2) + \mathrm{I}_f(W_1 \circledast W_2) \le \mathrm{I}_f(W_1) + \mathrm{I}_f(W_2)$$ **(super-MG)**

Example – Maximal Correlation

Theorem 7

For the class of **symmetric BICs**, corresponding to the channel functional $\mathrm{I}_f(W) = \rho_m^2(W)$, the channel polarization process is a super-martingale.

Proof

Apply Corollary 4.

For $f(s) = (2s-1)^2$, the functional inequality holds for all $p, q \in \left[0, \frac{1}{2}\right]$:

$$f(p \star q) + (1 - p \star q) f\left(\frac{pq}{1 - p \star q}\right) + (p \star q) f\left(\frac{\bar{p}q}{p \star q}\right) \boxed{\leq} f(p) + f(q)$$

$$RHS(p,q) - LHS(p,q) = \frac{4pq\bar{p}\bar{q}(2p-1)^2(2q-1)^2}{(1 - p \star q)(p \star q)} \geq 0$$

Results for Class of Symmetric BICs

Possible to generalize to arbitrary BICs

Measurable Function $f: [0,1] \Rightarrow \mathbb{R}$	**Functional** $\mathrm{I}_f(W)$	**Polarization Processes** $\{\mathrm{I}_f(W_n)\}_{n=0}^{n=\infty}$
$f(s) = 1 - h_2(s)$	$I(W)$	Analytical: MG **[1]**
$f(s) = s(1 + \log_2 s)^r + \bar{s}(1 + \log_2 \bar{s})^r$ For $r = 2$, $f(s)$ is non-convex!	$M_r(W)$	Numerical: $M_2(W)$ super-MG Analytical: $V(W)$ super-MG **[3]**
$f(s) = 2\sqrt{s(1-s)}$	$Z(W)$	Analytical: super-MG **[1]**
$f(s) = 2s^{\alpha}(1-s)^{1-\alpha}$	$\mathrm{H}_{\alpha}(W)$	Numerical: super-MG, $\alpha \in (0,1)$
$f(s) = 2^{-\rho}\left(s^{\frac{1}{1+\rho}} + (1-s)^{\frac{1}{1+\rho}}\right)^{1+\rho}$	$e^{-E_0(\rho,W)}$	Numerical: $e^{-E_0(\rho,W)}$ super-MG Analytical: $E_0(\rho, W)$ sub-MG **[2]**
$f(s) = (\lambda \wedge \bar{\lambda}) - ((2\bar{\lambda}s) \wedge (2\bar{s}\lambda))$	$B_{\lambda}(W)$	Analytical: $B_{1/3}(W)$ neither sub-MG nor super-MG
$f(s) = \lvert 2s - 1 \rvert$	$1 - 2P_{e,ML}(W)$	Analytical: super-MG
$f(s) = (2s-1)^2$	$\rho_{max}^2(W)$	Analytical: super-MG

References Cited

[1] *E. Arikan*, **Channel Polarization: A Method for Constructing Capacity-Achieving Codes for Symmetric Binary-Input Memoryless Channels**, IEEE Transactions on Information Theory, vol. 55, no. 7, pp. 3051--3073, July 2009.

[2] The following paper established that the random process associated to Gallager's reliability function is a submartingale: *M. Alsan and E. Telatar*, **Polarization improves $\boldsymbol{E_0}$**, IEEE Transactions on Information Theory, vol. 60, no. 5, pp. 2714--2719, May 2014.

[3] The second moment of information density is related to the concept of channel dispersion, or varentropy as introduced generally in: *E. Arikan*, **Varentropy decreases under the polar transform**, IEEE Transactions on Information Theory, vol. 62, no. 6, pp. 3390--3400, June 2016.