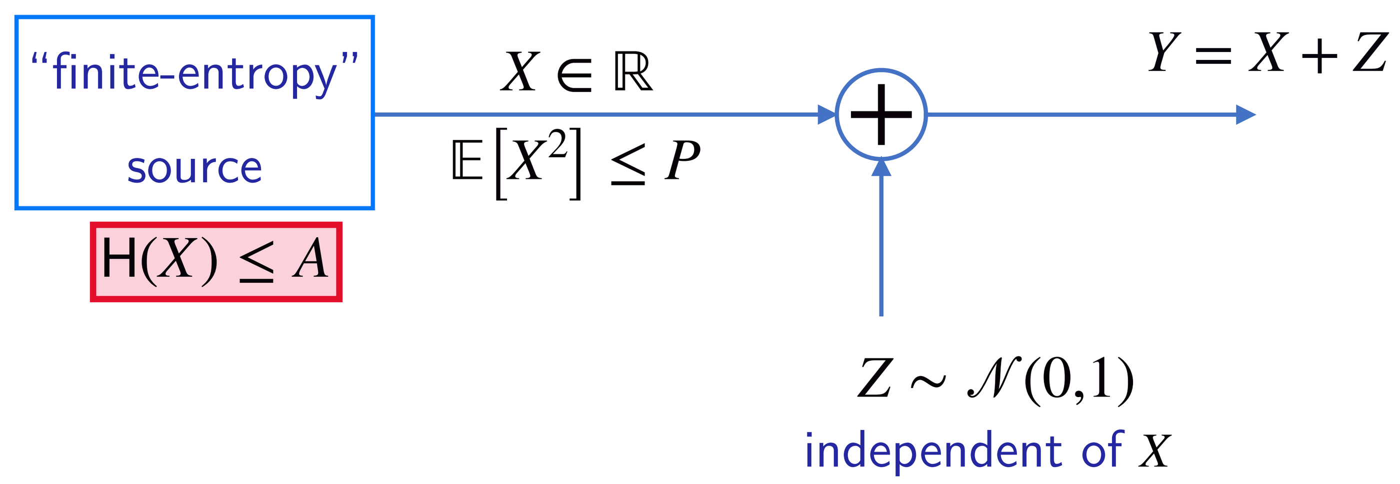
Additive Gaussian Channels with an Input-Entropy Constraint

a.k.a. a reason to study discrete distributions on continuous alphabets

Adway Girish [EPFL], Emre Telatar [EPFL], Shlomo Shamai [Technion]

EPFL Information Processing Group



$$\mathsf{C_H}(A,P) = \sup_{\mathsf{P}_X:\ \mathbb{E}[X^2]\le P,\ \boxed{\mathsf{H}(X)\le A}} \mathrm{I}(X;Y)$$

Dependence on A, fix $P=1$, $\mathsf{C_H}(A) = \mathsf{C_H}(A,1)$

Dependence on P

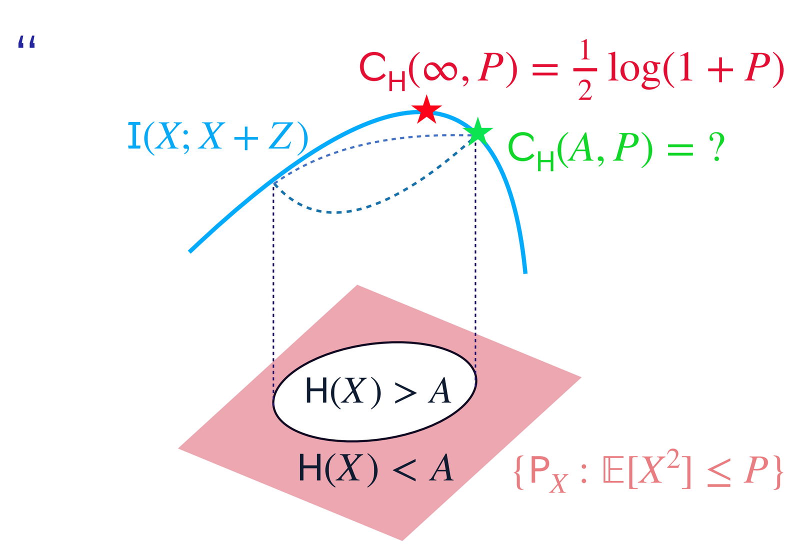
Motivation

Capacity-achieving distributions for continuous alphabet channels are sometimes surprisingly discrete (e.g., amplitude-constrained Gaussian channel [1], fading channel [2]);

what happens if they are forced to be discrete?

Difficulties

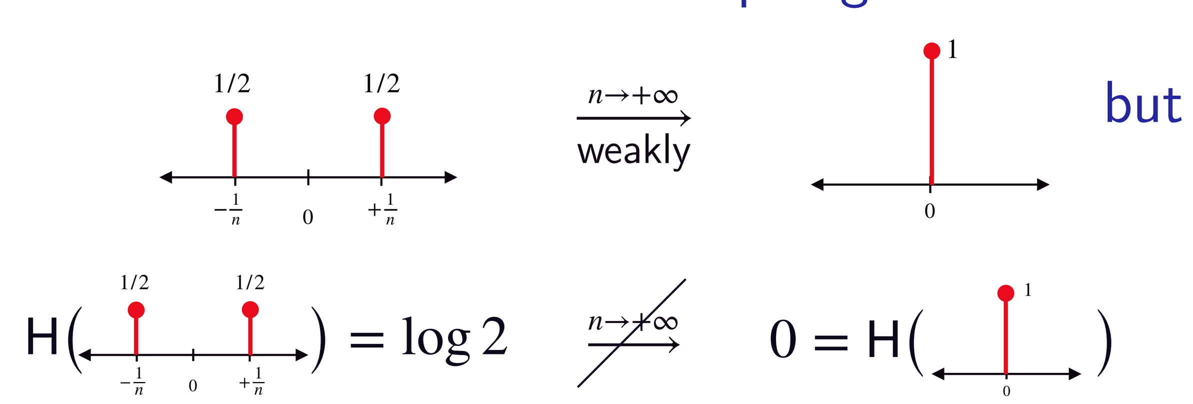
1. Non-convex optimization problem



2. Discrete distributions over continuous alphabets?

3. H is not continuous w.r.t. usual topologies on distributions,

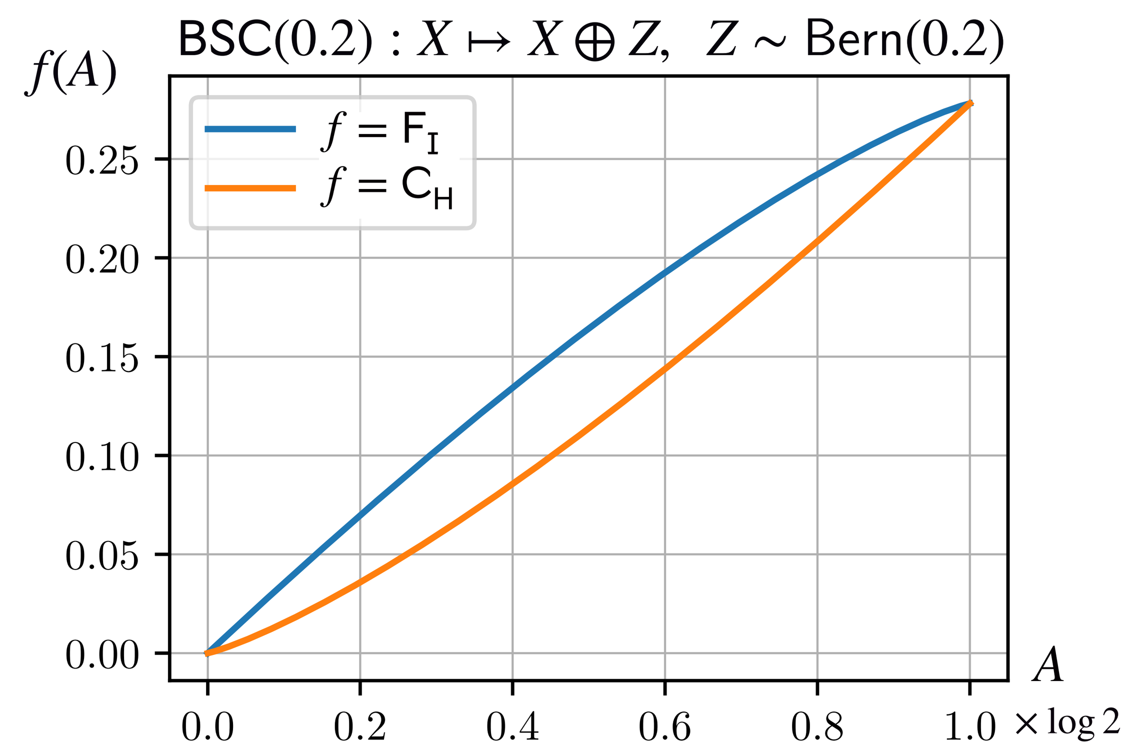
e.g.,



Connections to SDPI and $\mathsf{F_I}$-curves [3]

$$\mathsf{F_I}(A) = \sup_{\boxed{\mathsf{P}_{XW}}:\ \mathbb{E}[X^2]\le 1,\ \mathrm{I}(X;W)\le A} \mathrm{I}(W;Y)$$

$\mathsf{F_I}(A) \ge \mathsf{C_H}(A)$ for general channels, usually strict, e.g.,



But for such power-constrained Gaussian channels, it is known that $W = X$ is optimal (in $\mathsf{F_I}$ as above) asymptotically as $A \to 0^+, \infty$, with the correct rate [3]!

Add cardinality constraint

$$\mathsf{C}^n_{\mathsf{H}}(A,P) = \sup_{\mathsf{P}_X:\ \mathbb{E}[X^2]\le P,\ \mathsf{H}(X)\le A,\ \boxed{|\mathsf{supp}(\mathsf{P}_X)| = n}} \mathrm{I}(X;Y)$$

Then $\mathsf{P}_X \equiv \{(p_i, x_i)\}_{i=1}^n$, $p_i > 0$, x_i distinct

$\longrightarrow$ Constrained optimization over $\mathbb{R}^{2n}$

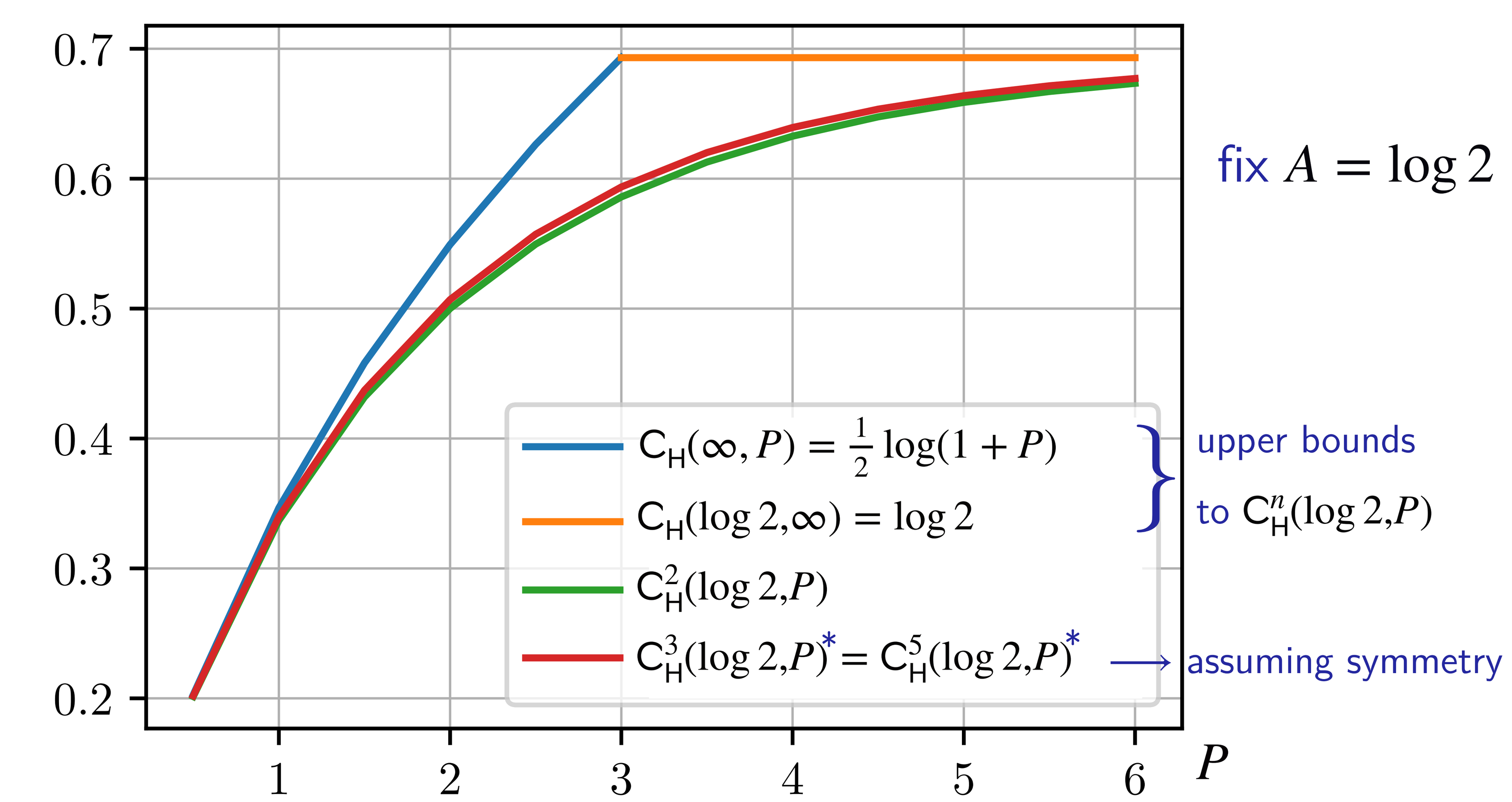
KKT conditions **necessary** for optimality:

there exist $\lambda_1, \lambda_2 \ge 0$ such that

1. $\mathbb{E}\left[Z \log \dfrac{1}{f_Y(x_i + Z; \mathsf{P}_X)}\right] = 2\lambda_2 x_i$

2. $\mathbb{E}\left[\log \dfrac{1}{f_Y(x_i + Z; \mathsf{P}_X)}\right] = \mathsf{C}^n_{\mathsf{H}}(A,P) + \lambda_2 (x_i^2 - P) + \lambda_1 \left(\overbrace{\log \frac{1}{p_i}}^{\text{difficult}} - A\right)$

Initial conjecture: two-point distribution is optimal for small P: False!



References

[1] Smith, J.G., 1971. The information capacity of amplitude-and variance-constrained scalar Gaussian channels. *Information and control*, *18*(3), pp.203-219.

[2] Abou-Faycal, I.C., Trott, M.D. and Shamai, S., 2001. The capacity of discrete-time memoryless Rayleigh-fading channels. *IEEE transactions on Information Theory, 47(4)*, pp.1290-1301.

[3] du Pin Calmon, F., Polyanskiy, Y. and Wu, Y., 2017. Strong data processing inequalities for input-constrained additive noise channels. *IEEE Transactions on Information Theory*, *64*(3), pp.1879-1892.