

Capacity-Achieving Probabilistic Shaping

Georg Böcherer

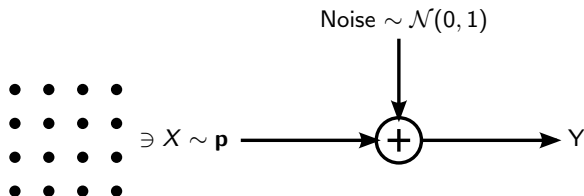
Institute for Theoretical Information Technology
RWTH Aachen University

Parts I-III of a talk given December 15, 2011
at the Institute for Communications Engineering, TU Munich

Part I:
Motivation

Motivation

M-QAM AWGN channel



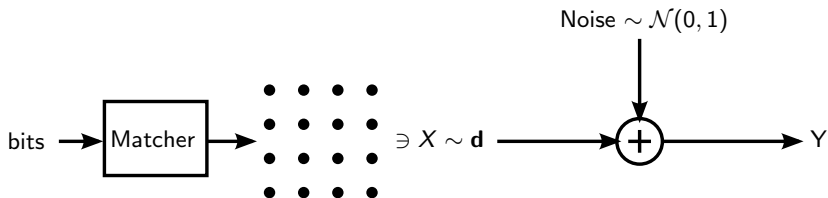
- ▶ Power vector \mathbf{w} , $w_i = |x_i|^2$
- ▶ Power constraint: $\mathbf{w}^T \mathbf{p} \leq E$
- ▶ Capacity is given by

$$C = \max_{\mathbf{p}: \mathbf{w}^T \mathbf{p} \leq E} \mathbb{I}(X; Y)$$

- ▶ Achieved if $X \sim \mathbf{p}^*$

Motivation

Channel matcher



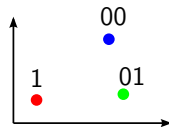
- ▶ Matcher maps iid equiprobable bits to signal points
- ▶ Matcher generates signal point pmf $X \sim \mathbf{d}$
- ▶ Generated pmf \mathbf{d} should be 'close' to \mathbf{p}^*

Motivation

Prefix-free matcher

- ▶ Signal points $\{\bullet, \bullet, \bullet\}$, full prefix-free code $\{1, 00, 01\}$

- ▶ Prefix-free matcher:



- ▶ Matching: $10001101 \dots \rightarrow \bullet \bullet \bullet \bullet \dots$

- ▶ Generated pmf \mathbf{d}

$$d_{\bullet} = 2^{-1}, \quad d_{\bullet} = 2^{-2}, \quad d_{\bullet} = 2^{-2}$$

Motivation

Which prefix-free matcher should we use?

Intuition: generated pmf \mathbf{d} should be 'close' to \mathbf{p}^*

- ▶ Kschischang & Pasupathy (1993): use Huffman code of \mathbf{p}^*
- ▶ $\mathbf{d}_{\text{HUFFMAN}}$ minimizes $\mathbb{D}(\mathbf{p}^* \parallel \mathbf{d}) = \sum_i p_i^* \log \frac{p_i^*}{d_i}$

Two questions:

- ▶ Why minimize $\mathbb{D}(\mathbf{p}^* \parallel \mathbf{d})$?
- ▶ Can we achieve capacity by generating signal points blockwise?

Part II:
Using a 'wrong' pmf

Using a 'wrong' pmf

Objective

Question: what is the decrease of mutual information when using some pmf \mathbf{p} instead of the capacity achieving pmf \mathbf{p}^* ?

Using a 'wrong' pmf

Notation

- ▶ DMC with transition matrix \mathbf{H}
- ▶ Input pmf \mathbf{p}
- ▶ Output pmf $\mathbf{r} = \mathbf{H}\mathbf{p}$
- ▶ Mutual information between input and output $\mathbb{I}(\mathbf{p})$
- ▶ Capacity $C = \max_{\mathbf{p}} \mathbb{I}(\mathbf{p})$, achieved by \mathbf{p}^* .

Using a 'wrong' pmf

- ▶ Optimal pmf \mathbf{p}^* , pmf \mathbf{p} .
- ▶ Output pmfs $\mathbf{r}^* = \mathbf{H}\mathbf{p}^*$, $\mathbf{r} = \mathbf{H}\mathbf{p}$.
- ▶ Assumption: $p_i = 0$ whenever $p_i^* = 0$.

Proposition 3.11¹: Mutual information achieved by \mathbf{p} is

$$\mathbb{I}(\mathbf{p}) = C - \mathbb{D}(\mathbf{r} \parallel \mathbf{r}^*)$$

¹All proposition are from my submitted PhD thesis [Capacity-Achieving Probabilistic Channels for Noisy and Noiseless Channels](#). Send me an email to get an electronic copy.

Using a 'wrong' pmf

Sketch of the proof:

- ▶ KKT condition for optimal \mathbf{p}^* :

$$\sum_j h_{ji} \log \frac{h_{ji}}{r_j^*} \leq C, \quad \text{with equality if } p_i^* > 0.$$

Now:

$$\begin{aligned} \mathbb{I}(\mathbf{p}) &= \sum_i p_i \sum_j h_{ji} \log \frac{h_{ji} r_j^*}{r_j r_j^*} \\ &= \sum_i p_i \sum_j \underbrace{h_{ji} \log \frac{h_{ji}}{r_j^*}}_{=C \text{ if } p_i > 0} - \mathbb{D}(\mathbf{r} \parallel \mathbf{r}^*) \\ &= C - \mathbb{D}(\mathbf{r} \parallel \mathbf{r}^*). \end{aligned}$$

Equality in last line because of assumption $p_i = 0$ whenever $p_i^* = 0$.

Using a 'wrong' pmf

Penalty bound

- ▶ By log-sum inequality:

$$\mathbb{D}(\mathbf{r} \parallel \mathbf{r}^*) \leq \mathbb{D}(\mathbf{p} \parallel \mathbf{p}^*).$$

- ▶ Bounds for mutual information achieved by \mathbf{p} :

$$C \geq \mathbb{I}(\mathbf{p}) \geq C - \mathbb{D}(\mathbf{p} \parallel \mathbf{p}^*)$$

- ▶ Bounds for mutual information loss ('shaping gap'):

$$0 \leq C - \mathbb{I}(\mathbf{p}) \leq \mathbb{D}(\mathbf{p} \parallel \mathbf{p}^*)$$

minimize $\mathbb{D}(\mathbf{p} \parallel \mathbf{p}^*)$

[not $\mathbb{D}(\mathbf{p}^* \parallel \mathbf{p})$]

Using a 'wrong' pmf

DMC Variations

Capacity	Mutual information	Reference
$C := \max_{\mathbf{p}} \mathbb{I}(\mathbf{p})$	$C - \mathbb{D}(\mathbf{r} \parallel \mathbf{r}^*)$	Prop. 3.11
$C_{\text{puc}} := \max_{\mathbf{p}} \frac{\mathbb{I}(\mathbf{p})}{\mathbf{w}^T \mathbf{p}}$	$C_{\text{puc}} - \frac{\mathbb{D}(\mathbf{r} \parallel \mathbf{r}^*)}{\mathbf{w}^T \mathbf{p}}$	Prop. 4.7
$C(E) := \max_{\mathbf{p}: \mathbf{w}^T \mathbf{p} \leq E} \mathbb{I}(\mathbf{p})$	$C(E) + \frac{\partial C(E)}{\partial E} (\mathbf{w}^T \mathbf{p} - E) - \mathbb{D}(\mathbf{r} \parallel \mathbf{r}^*)$	Prop. 5.10

Summary

- ▶ The shaping gap that results from using \mathbf{p} instead of \mathbf{p}^* is upper bounded by $\mathbb{D}(\mathbf{p} \parallel \mathbf{p}^*)$.

Part III:
Geometric Huffman coding

Geometric Huffman coding

Objective

- ▶ Prefix-free matcher generates pmf \mathbf{d}
- ▶ **Objective:** find prefix-free matcher that minimizes

$$\mathbb{D}(\mathbf{d} \parallel \mathbf{p}^*)$$

Geometric Huffman coding

Target pmf \mathbf{p}^* with $p_1^* \geq p_2^* \geq \dots \geq p_n^*$.

- ▶ **Huffman source coding:** recursively construct a full prefix-free code with the updating rule

$$p' = p_{n-1}^* + p_n^*.$$

- ▶ **Geometric Huffman coding:** recursively construct a full prefix-free code with the updating rule

$$p' = \begin{cases} 2\sqrt{p_n^* p_{n-1}^*} & \text{if } p_{n-1}^* < 4p_n^* \\ p_{n-1}^* & \text{if } p_{n-1}^* \geq 4p_n^*. \end{cases}$$

Geometric Huffman coding

Proposition 3.3: GHC constructs the prefix-free matcher that minimizes $\mathbb{D}(\mathbf{d} \parallel \mathbf{p}^*)$.

Geometric Huffman coding

Asymptotic behavior

- ▶ k symbols iid $\sim \mathbf{p}^*$
- ▶ joint pmf \mathbf{p}^{*k}
- ▶ $\mathbf{d}_k = \text{GHC}(\mathbf{p}^{*k})$

Proposition 3.8: As blocklength k approaches infinity:

$$\frac{\mathbb{D}(\mathbf{d}_k \| \mathbf{p}^{*k})}{k} \rightarrow 0$$

Geometric Huffman coding

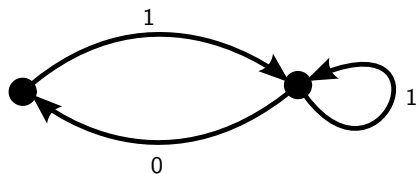
Variations of GHC

Objective	Algorithm	Reference
minimize $\mathbb{D}(\mathbf{d} \mathbf{p})$	GHC	Algorithm 1
minimize $\frac{\mathbb{D}(\mathbf{d} \mathbf{p})}{\mathbf{w}^T \mathbf{d}}$	normalized GHC	Algorithm 3
minimize $\mathbb{D}(\mathbf{d} \mathbf{p})$ $\mathbf{d}:\mathbf{w}^T \mathbf{d} \leq E$	cost constrained GHC	Algorithm 5

Geometric Huffman coding

Application: Finite state noiseless channels

Finite state noiseless channel



Memoryless representation



Proposition 6.10:

- ▶ Every finite state noiseless channel has a memoryless representation of same capacity.
- ▶ Normalized GHC achieves capacity of any memoryless discrete noiseless channel.

Geometric Huffman coding

Open Problem: Channel matching vs. source coding

- ▶ Is Huffman coding asymptotically a good channel matcher, i.e.,

$$\frac{\mathbb{D}(\mathbf{d}_k^{\text{HUFFMAN}} \parallel \mathbf{p}^{*k})}{k} \xrightarrow{k \rightarrow \infty} 0?$$

- ▶ Is GHC asymptotically a good source code, i.e.,

$$\frac{\mathbb{D}(\mathbf{p}^{*k} \parallel \mathbf{d}_k^{\text{GHC}})}{k} \xrightarrow{k \rightarrow \infty} 0?$$

Special case considered in [Kerpez \(1991\)](#), proof very rough/incomplete/wrong(?)

Summary

- ▶ The shaping gap that results from using \mathbf{p} instead of \mathbf{p}^* is upper bounded by $\mathbb{D}(\mathbf{p} \parallel \mathbf{p}^*)$.
- ▶ Geometric Huffman coding minimizes $\mathbb{D}(\mathbf{d} \parallel \mathbf{p}^*)$ over all pmfs \mathbf{d} that can be generated by a prefix-free matcher.
- ▶ $\frac{\mathbb{D}(\mathbf{d}_k^{\text{GHC}} \parallel \mathbf{p}^{*k})}{k}$ vanishes as blocklength $k \rightarrow \infty$

References

- G. Böcherer, “Capacity-Achieving Probabilistic Shaping for Noisy and Noiseless Channels,” *submitted PhD thesis*.
- F. Kschischang and S. Pasupathy, “Optimal nonuniform signaling for Gaussian channels,” *IEEE Trans. Inf. Theory*, vol. 39., no. 3, pp. 913–929, 1993.
- K. Kerpez, “Runlength codes from source codes,” *IEEE Trans. Inf. Theory*, vol. 37, no. 3, pp 682–687, 1991.