Technische Universität München
Professur für
Leitungsgebundene Übertragungstechnik
Prof. Dr.-Ing. Norbert Hanik

Master's Thesis

Thesis Title

Vorgelegt von:
Student Name

München, Month Year

Betreut von:
Dipl.-Ing. Advisor

Master's Thesis an der
Professur für Leitungsgebundene Übertragungstechnik (LÜT)
der Technischen Universität München (TUM)
Titel : Thesis Title
Autor : Student Name

Student Name
Streetname Number
ZIP München
your.email@addre.ss

Ich versichere hiermit wahrheitsgemäß, die Arbeit bis auf die dem Aufgabensteller bereits bekannte Hilfe selbständig angefertigt, alle benutzten Hilfsmittel vollständig und genau angegeben und alles kenntlich gemacht zu haben, was aus Arbeiten anderer unverändert oder mit Abänderung entnommen wurde.

München, xx.xx.20xx
....

Ort, Datum (Student Name)

Contents

Abstract

The abstract comes here.

1 Introduction

The constant demand for higher capacity digital links has motivated the development of communication schemes which approach closer and closer the analytical limits of the channel capacity. This search for optimality has depicted a logical path. Sending a single bit per time-frequency slot is inefficient according to the definition of Channel capacity. Therefore, higher-order modulations like amplitude-shift keying (ASK) or quadrature amplitude modulation (QAM) were developed to improve the efficiency. Under these modulation schemes, the receiver handles more than two signal points per real dimension. The set of signal points is known as constellation. However, these schemes make use of uniform probability distributions for the occurrence of the constellation points leading to a constant-width gap to capacity. Once these schemes were adopted and additional capacity was required, research tackled the root cause of the constant gap, the usage of uniform distributions.

To overcome this gap constellation shaping techniques can be applied. Shaping a constellation means optimizing the distribution of the signal to be transmitted. Furthermore, this optimization unfolds into the improvement of the constellation points' location or their occurrence probability. The first is known as geometric shaping while the latter is known as probabilistic shaping. In both cases, the goal is to maximize the mutual information $I(X;Y)$ of the channel input X and output Y by optimizing the constellation. This optimization problem arises from the definition of channel capacity C:

$$C = \max_{p(X)} I(X;Y) \tag{1.1}$$

Currently, the optimal $p(x)$ has only be found for specific channels such as the AWGN as knowledge of the channel distribution $p(y|x)$ is required. Still, solving 1.1 can become intractable despite knowing $p(y|x)$.

Here is where deep-learning can be applied to find constellations which maximize $I(X;Y)$, without analytical knowledge of the channel. As shown in [OH17], the complete communication systems can be interpreted as an autoencoder. This approach tackles the physical layer by optimizing the end-to-end performance instead of the performance of the individual components by carefully choosing the loss function. Geometric shaping under the autoencoder framework has been already performed in [OH17], [JEY+18]. Furthermore, geometric shaping and probabilistic shaping have been jointly implemented in [SAAH19], [AH20] and [AC21]. The results show that the joint application of probabilistic and geometric shaping outperform the PS-QAM scheme from [?] and approach the limit to within 0.1 dB in the AWGN channel. Yet the existence of multiple and apparently different architectures is intriguing. Consequently, in this work we study the differences in the implementations proposed by [SAAH19] and [AC21].

The rest of this document is organized as follows. Chapter 2, Preliminaries, presents the fundamentals of classical Probabilistic constellation systems as well as of autoencoder systems. Chapter 3, presents an implementation of both architectures with Pytorch [PGM+19] and analyses their main differences. Chapter 4 wraps-up this work.

2 Preliminaries

2.1 Probabilistic Constellation Shaping

We begin with an important result from Information theory. Under a second-moment constraint, also known as power constraint, the probability distribution which maximizes the differential entropy is the Gaussian distribution. We thus have

$$h(X) \leq \frac{1}{2}\log\left(2\pi e\sigma^2\right) \tag{2.2}$$

where $\sigma^2 = \mathbb{E}[X^2]$, and with equality if and only if X is Gaussian-distributed. More generally in the multi-dimensional case we have

$$h(\underline{X}) \leq \frac{1}{2}\log\left((2\pi e)^n|\mathbf{Q}_{\underline{X}}|\right) \tag{2.3}$$

where we have considered a random column vector $\underline{X}$ of dimension n, mean $\mathbb{E}[\underline{X}] = \underline{m}$ and covariance matrix

$$\mathbf{Q}_{\underline{X}} = \mathbb{E}[(\underline{X}-\underline{m})(\underline{X}-\underline{m})^{\intercal}] \tag{2.4}$$

and equality in 2.3 if only if $\underline{X}$ are jointly Gaussian.
Lets now consider an Additive white gaussian noise (AWGN) channel with input X, of zero mean and variance P, noise Z and output Y; i.e. $Y = X + Z$. Furthermore, the capacity of the AWGN is

$$\begin{align}
C(P) &= \max_{P_X:\mathbb{E}[X^2]\leq P} \mathbb{I}(X;Y) \tag{2.5}\\
&= \max_{P_X:\mathbb{E}[X^2]\leq P} [h(Y) - h(Y|X)] \tag{2.6}\\
&= \frac{1}{2}\log(2\pi e(P+N)) - \frac{1}{2}\log(2\pi eN) \tag{2.7}\\
&= \frac{1}{2}\log\left(1+\frac{P}{N}\right). \tag{2.8}
\end{align}$$

We can analyse the mutual information

$$\mathbb{I}(X;Y) = h(Y) - h(Y|X) \tag{2.9}$$

in two parts, the differential entropy of the output and the conditional differential entropy of the output given the input. We expand the second term as

$$\begin{aligned} h(Y|X) &= h(Y - X|X) && (2.10)\\ &= h(Z|X) && (2.11)\\ &= h(Z) = \frac{1}{2}\log\left(2\pi e\sigma^2\right) && (2.12)\end{aligned}$$

and observe that the term $h(Y|X)$ does not depend on how X is distributed. In contrast, $h(y)$ does depend on how X is distributed by

$$p_Y(y) = \int_{-\infty}^{\infty} p_X(X)p_Z(y-x)\,dx = (p_X \star p_Z)(y). \tag{2.13}$$

To circumvent the fact that it is difficult to find a closed-form expression of $h(Y)$, we denote the Gaussian distribution as p_G and make use of the information divergence as

$$h(Y) \overset{\text{(a)}}{=} h(Y_G) - \mathbb{D}(p_Y \| p_G) \tag{2.14}$$

where (a) arises from the fact that $\mathbb{X}(p_X\|p_G) = h(Y_G)$ if and only if p_X has zero mean and variance P as p_G. 2.14 is very useful as it allows us to express the differential entropy of the output in terms of the difference between p_G and any other distribution by means of the cross entropy. Now we can rewrite 2.9 as

$$\begin{aligned} \mathbb{I}(X;Y) &= h(Y) - h(Y|X) && (2.15)\\ &= h(Y) - h(Z) && (2.16)\\ &= h(Y_G) - \mathbb{D}(p_Y\|p_G) - h(Z) && (2.17)\\ &= [h(Y_G) - h(Z)] - \mathbb{D}(p_Y\|p_G) && (2.18)\\ &= C(P/\sigma^2) - \mathbb{D}(p_Y\|p_G) && (2.19)\end{aligned}$$

This last result indicates that the loss of MI when using a distribution P_X different than P_G is the informational divergence $\mathbb{D}(p_Y\|p_G)$.

We would like now to understand how far a uniform distribution is from 2.2. To do this we will follow the approach presented in [Bö18] to lower bound the MI. Start by defining X_U as a uniformly distributed input on the interval $[-A, A]$ where A is carefully chosen so that $\mathbb{E}[{X_U}^2] = P$. The corresponding output is Y_U and we proceed

$$\begin{aligned} \mathbb{I}(X_U; Y_U) &= C(snr) - \mathbb{D}(p_{Y_U} \| p_{Y_G}) && (2.20) \\ &\geq C(snr) - \mathbb{D}(p_{X_U} \| p_{X_G}) && (2.21) \\ &= C(snr) - [h(X_G) - h(X_U)] && (2.22) \\ &= C(snr) - \frac{1}{2} \log_2 \left(\frac{\pi e}{6} \right). && (2.23) \end{aligned}$$

3 contribution

3.1 Notation

In the following we will use the notation:

$$\mathbb{I}\left(X, Y; D, P_M, C_M\right)$$

which expresses the mutual information between X and Y and D, P_M, C_M, separated by a semi-colon, are the trainable parameters of the system. The parameters can be seen as additional input to a function.

3.2 Optimization of trainable parameters

As we have seen in chapter 2, the goal of probabilistic constellation shaping is to maximize the mutual information

$$\max_{D, P_M, C_M} \mathbb{I}\left(X, Y; D, P_M, C_M\right) \overset{\text{(a)}}{=} \mathbb{H}(X) - \mathbb{X}(P_{X|Y} \| Q_{X|Y}; D, P, C) \tag{3.24}$$

where the entropy is maximized when the probabilities of the constellation points follow a Gaussian distribution; and the cross-equivocation is minimum when $Q_{X|Y} = P_{X|Y}$.
Typically, the gradient descent (ascent) allows us to solve the optimization problem by adjusting the trainable parameters as:

$$\theta_{new} = \theta_{old} + \epsilon \frac{\partial}{\partial \theta_{old}} \mathbb{I}\left(X, Y; \theta_{old}\right) \tag{3.25}$$

for all trainable parameters $\theta \in P, C, D$. However, [work in progress...]
We will need to numerically compute the terms in 3.24. For this reason, we will use the following expansions:

$$\mathbb{X}\left(P_{X|Y} \| Q_{X|Y} | Y = b\right) = \sum_{a \in Supp(P_{X|Y}(\cdot|b))} P_{X|Y}(a|b) \log_2(Q_{X|Y}(a|b)) \tag{3.26}$$

$$\mathbb{X}\left(P_{X|Y} \| Q_{X|Y}\right) = \sum_{b \in Supp(P_Y)} P_Y(b) \mathbb{X}\left(P_{X|Y} \| Q_{X|Y} | Y = b\right) \tag{3.27}$$

which combined together and applying Bayes' theorem yields

$$\mathbb{X}\left(P_{X|Y} \| Q_{X|Y}\right) = \sum_{(a,b) \in Supp(P_{XY})} P_X(a) P_{Y|X}(b|a) \log_2(Q_{X|Y}(a|b)). \tag{3.28}$$

Precisely, we must compute the derivative terms of 3.28 with respect to $P_X(j), j \in \mathcal{X}$. This is

$$\frac{\partial}{\partial P_X(j)} \mathbb{X}\left(P_{X|Y} \| Q_{X|Y}\right) = \sum_{b \text{ if } x=j} P_{Y|X}(b|j) \log_2 Q_{X|Y}(j|b) \tag{3.29}$$

$$+ \sum_{(a,b) \in Supp(P_{XY})} P_{XY}(a,b) \frac{\partial}{\partial P_X(j)} \log_2 Q_{X|Y}(a|b) \tag{3.30}$$

which can be rewritten using the expectation operator as

$$\frac{\partial}{\partial P_X(j)} \mathbb{X}\left(P_{X|Y} \| Q_{X|Y}\right) = \mathbb{E}_{Y|X}[\log_2 Q_{X|Y}(j|b) | X = j] \tag{3.31}$$

$$+ \mathbb{E}_{XY}[\frac{\partial}{\partial P_X(j)} \log_2 Q_{X|Y}(a|b)]. \tag{3.32}$$

The terms can now be numerically computed as

$$\mathbb{E}_{Y|X}[\log_2 Q_{X|Y}(j|b) | X = j] \approx \frac{1}{B P_X(j)} \sum_{b \text{ if } x=j} \log_2 Q_{X|Y}(j|b) \tag{3.33}$$

$$\mathbb{E}_{XY}[\frac{\partial}{\partial P_X(j)} \log_2 Q_{X|Y}(a|b)] \approx \frac{1}{B} \sum_{(a,b) \in Supp(P_{XY})} \log_2 Q_{X|Y}(a|b) \tag{3.34}$$

Appendix A

First appendix goes here.

Appendix B

Second appendix goes here.

Bibliography

[AC21] V. Aref and M. Chagnon, “End-to-end learning of joint geometric and probabilistic constellation shaping,” 2021. [Online]. Available: https://arxiv.org/abs/2112.05050

[AH20] F. A. Aoudia and J. Hoydis, “Joint learning of probabilistic and geometric shaping for coded modulation systems,” in *GLOBECOM 2020 - 2020 IEEE Global Communications Conference*, 2020, pp. 1–6.

[Bö18] G. Böcherer, “Principles of coded modulation. [online] available,” 2018. [Online]. Available: http://www.georg-boecherer.de/bocherer2018principles.pdf

[JEY+18] R. T. Jones, T. A. Eriksson, M. P. Yankov, B. J. Puttnam, G. Rademacher, R. S. Luis, and D. Zibar, “Geometric constellation shaping for fiber optic communication systems via end-to-end learning,” 2018. [Online]. Available: https://arxiv.org/abs/1810.00774

[OH17] T. O’Shea and J. Hoydis, “An introduction to deep learning for the physical layer,” *IEEE Transactions on Cognitive Communications and Networking*, vol. 3, no. 4, pp. 563–575, 2017.

[PGM+19] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala, “Pytorch: An imperative style, high-performance deep learning library,” in *Advances in Neural Information Processing Systems 32*, H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, Eds. Curran Associates, Inc., 2019, pp. 8024–8035. [Online]. Available: http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf

[SAAH19] M. Stark, F. Ait Aoudia, and J. Hoydis, “Joint learning of geometric and probabilistic constellation shaping,” in *2019 IEEE Globecom Workshops (GC Wkshps)*, 2019, pp. 1–6.