Inspecting NOMA from Information Theory

The Art of Dealing with Interference

Weijia Zheng [1]

Department of Information Engineering
The Chinese University of Hong Kong

wjzheng@link.cuhk.edu.hk

March 25, 2024

[1] I mainly take reference to: Chapter 5, "Multiple Access Techniques for 5G Wireless Networks and Beyond". Springer Publishing Company, Incorporated, by M. Vaezi, Z. Ding, and H. V. Poor. 2018.

Overview

Overview

Multiple Access, OMA and NOMA

Multiple access is of crucial importance of cellular communication systems. From 1G to 4G, different multiple access schemes are adopted.

Multiple access methods used from 1G to 4G

- Frequency division multiple access (FDMA, for 1G)
- Time division multiple access (TDMA, for 2G)
- Code division multiple access (CDMA, for 3G)
- Orthogonal frequency division multiple access (OFDMA, for 4G)

One common idea: different users have orthogonal signals. Hence these above methods are usually considered as orthogonal multiple access (OMA).

Multiple Access, OMA and NOMA

Multiple access is of crucial importance of cellular communication systems. From 1G to 4G, different multiple access schemes are adopted.

Multiple access methods used from 1G to 4G

- Frequency division multiple access (FDMA, for 1G)
- Time division multiple access (TDMA, for 2G)
- Code division multiple access (CDMA, for 3G)
- Orthogonal frequency division multiple access (OFDMA, for 4G)

One common idea: different users have orthogonal signals. Hence these above methods are usually considered as orthogonal multiple access (OMA).

The term "NOMA" is rather vague

Different people preceive the "non-orthogonality" in NOMA via different ways. We present some of the definitions then.

3 Definitions of NOMA

Linear Transform in Decoding

This view requires OMA schemes to be: signals of different users can be separated into orthogonal subspaces using a linear transform.
And, any scheme does not meet such condition $\implies$ NOMA.

3 Definitions of NOMA

Linear Transform in Decoding

This view requires OMA schemes to be: signals of different users can be separated into orthogonal subspaces using a linear transform.
And, any scheme does not meet such condition $\Longrightarrow$ NOMA.

Superposition coding & Succesive interference cancellation (SC-SIC)

Many consider NOMA to be equivalent to schemes that use superposition coding at transmitter and successive interference cancellation at receiver (SC-SIC). Aka, SC-SIC = NOMA.
Given that SC-SIC achieves capacity of degraded downlink/broadcast channel, and the capacity of an uplink channel (MAC) can be regarded as doing SIC at the decoder.

3 Definitions of NOMA

Linear Transform in Decoding

This view requires OMA schemes to be: signals of different users can be separated into orthogonal subspaces using a linear transform.
And, any scheme does not meet such condition $\implies$ NOMA.

Superposition coding & Succesive interference cancellation (SC-SIC)

Many consider NOMA to be equivalent to schemes that use superposition coding at transmitter and successive interference cancellation at receiver (SC-SIC). Aka, SC-SIC = NOMA.
Given that SC-SIC achieves capacity of degraded downlink/broadcast channel, and the capacity of an uplink channel (MAC) can be regarded as doing SIC at the decoder.

Information Theory Proving Mindset & Technique

NOMA may refer to any technique allowing concurrent transmission is over the same resources in time/frequency/code. This achieves a better rate region when compared to orthogonalizing of one or some of the resources. Under such viewpoint, SC-SIC $\subsetneq$ NOMA

Overview

Beyond orthogonal schemes

We hope NOMA "to reap the benefits promised by information theory for the downlink and uplink transmission of wireless systems, modelled by the broadcast channel (BC) and multiple access channel (MAC)."

Capacity regions of degraded BC and MAC have been established several decades ago (which we will see later), and concurrent non-orthogonal schemes are a must to achieve capacity.

Beyond orthogonal schemes

We hope NOMA "to reap the benefits promised by information theory for the downlink and uplink transmission of wireless systems, modelled by the broadcast channel (BC) and multiple access channel (MAC)."

Capacity regions of degraded BC and MAC have been established several decades ago (which we will see later), and concurrent non-orthogonal schemes are a must to achieve capacity.

Then... why from 1G to 4G we insisted on OMA?

This was mainly to avoid inter-user interference cancellation, which would have resulted in unacceptably complex receivers.

Some schemes of this type (avoiding interference cancellation), e.g., treat-interference-as-noise (TIN) are more practical.

Beyond orthogonal schemes

We hope NOMA "to reap the benefits promised by information theory for the downlink and uplink transmission of wireless systems, modelled by the broadcast channel (BC) and multiple access channel (MAC)."
Capacity regions of degraded BC and MAC have been established several decades ago (which we will see later), and concurrent non-orthogonal schemes are a must to achieve capacity.

Then... why from 1G to 4G we insisted on OMA?

This was mainly to avoid inter-user interference cancellation, which would have resulted in unacceptably complex receivers.
Some schemes of this type (avoiding interference cancellation), e.g., treat-interference-as-noise (TIN) are more practical.

Okay, then... is it a good timing to use NOMA now?

If we are really going into an era of IoT, then yes.
"Amazon alone has $> 10^8$ devices already, and networks connects them operate in (congested) ISM bands (900MHz 2.4GHz)." Overhead of orthogonal schemes will be a nightmare, given such massive connectivity.

Overview

In the following, no matter for multiple access channel, broadcast channel or interference channel, we will mainly focus on Gaussian channels without MIMO.

But we do will mention some more general channel types for completeness.

Definition (Gaussian capacity function)

We define the Gaussian capacity function $\mathcal{C}(x), x \geqslant 0$ as

$$\mathcal{C}(x) \coloneqq \frac{1}{2}\log(1+x).$$

This gives the capacity of a point-to-point Gaussian channel, more precisely, if

$$Y = gX + Z$$

with $Z \sim N(0,1)$ and $\mathbb{E}X^2 \leqslant P$. Then $\mathcal{C}(\frac{g^2P}{1}) = \mathcal{C}(\mathrm{SNR}) = \frac{1}{2}\log(1+\mathrm{SNR})$ gives you the channel capacity.

Multiple Access Channel, Uplink ($K \geqslant 2$ senders, 1 receiver)

MAC is well-understood. Its general capacity is known. Most studies are done in 1970 $\sim$ 1985.

2-sender Gaussian MAC (single letter)

$$Y = g_1 X_1 + g_2 X_2 + Z,$$

where $Z \sim N(0,1)$ is noise, g_1, g_2 are channel gains. $\mathbb{E}[X_i^2] \leqslant P$ is power constraint.

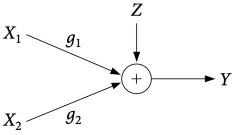


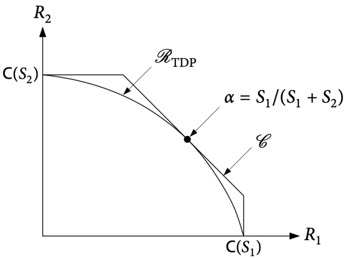
Figure 4.6. Gaussian multiple access channel.

K-sender GMAC Capacity (Cover, 1974)

The capacity region of the K-user Gaussian MAC is the set of all nonnegative $(R_1, \ldots, R_K)$ such that

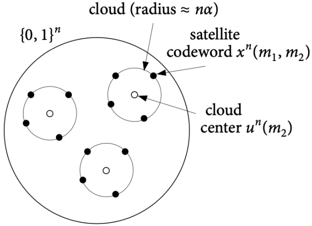
$$\sum_{j \in S} R_j \leqslant \mathcal{C}\Big(\sum_{j \in \mathcal{J}} S_j\Big), \forall \mathcal{J} \subset [K]$$

where $S_j = g_j^2 P$ is received SNR for user j. Remember that P is power constraint.



The corner points of the Gaussian MAC capacity region can be achieved using successive cancellation decoding.

Broadcast Channel, Downlink (1 sender, $K \geqslant 2$ receivers)



Theorem (2Rx Superposition Coding Inner Bound)

If $\exists$ some pmf $p(u,x)$ such that rate pairs $(R_1, R_2) \in \mathbb{R}_+^2$ satisfying

$$R_1 < I(X; Y_1|U),\ R_2 < I(U; Y_2),\ R_1 + R_2 < I(X; Y_1).$$

Then (R_1, R_2) is achievable.

In general, we do not know the capacity of BC.

However, superposition coding scheme attains capacity for some BCs. In particular, Gaussian BC is of a class called "degraded broadcast channels", whose capacity is known to be achieved by superposition coding.

Types of (discrete memoryless) Broadcast Channels

$$\underbrace{\text{Degraded}}_{\text{SC: capacity}} \subsetneq \underbrace{\text{Less Noisy}}_{\substack{\text{2,3Rx: SC is capacity} \\ \text{believe true for } \geqslant 4\text{Rx}}} \subsetneq \underbrace{\text{More Capable}}_{\substack{\text{2Rx: SC is capacity (El Gamal, 1978)} \\ \text{3Rx: SC strictly suboptimal (Xia \& Nair, 2012)}}} \subset \underbrace{\ldots}_{\text{know little}}$$

Gaussian BC

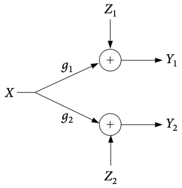


Figure 5.8. Gaussian broadcast channel.

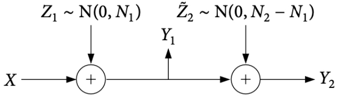


Figure 5.9. Corresponding physically degraded Gaussian BC.

Theorem (Capacity region of 2Rx Gaussian BC = SC-SIC)

The capacity region of Gaussian BC above is the set of rate pairs (R_1, R_2) *such that* $R_1 \leqslant \mathcal{C}(\alpha S_1)$, $R_2 \leqslant \mathcal{C}(\frac{(1-\alpha)S_2}{\alpha S_1 + 1})$, *for some* $\alpha \in [0,1]$, *where* $\mathcal{C}(x)$ *is the Gaussian capacity function.* S_i *denotes the signal-to-noise ratio of user* i.

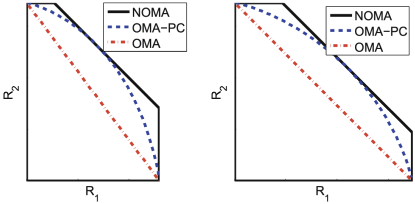
Theorem (Capacity region of $K \geqslant 2$-users Gaussian BC = SC-SIC)

The capacity region of K*-user Gaussian BC is the set of rate pairs* $(R_1, ..., R_K)$ *such that*

$$R_k \leqslant \mathcal{C}\left(\frac{\beta_k S_k}{1 + \sum_{j=1}^{k-1} \beta_j \gamma_k}\right),$$

for some $\alpha \in [0,1]$, *where* $\mathcal{C}(x)$ *is the Gaussian capacity function.*

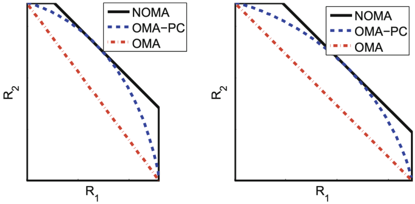
OMA vs. NOMA in Gaussian MAC and Gaussian BC



(a) $P_1 = 5$, $P_2 = 10$ (b) $P_1 = 10$, $P_2 = 10$

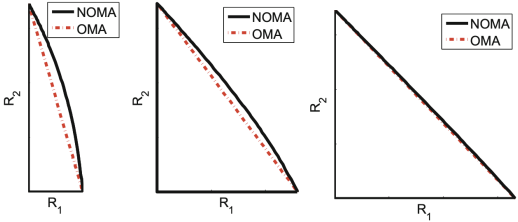
Fig. 5.2 Best achievable regions by OMA and NOMA in the two-user MAC (uplink) for different values of the two users' transmit powers P_1 and P_2

OMA vs. NOMA in Gaussian MAC and Gaussian BC



(a) $P_1 = 5$, $P_2 = 10$ (b) $P_1 = 10$, $P_2 = 10$

Fig. 5.2 Best achievable regions by OMA and NOMA in the two-user MAC (uplink) for different values of the two users' transmit powers P_1 and P_2



(a) $P_1 = 1$, $P_2 = 10$ (b) $P_1 = 5$, $P_2 = 10$ (c) $P_1 = 9$, $P_2 = 10$

Fig. 5.3 Best achievable regions by OMA and NOMA in the BC (download) for different values of P_1 and P_2

Overview

Interference Channel ($\geqslant$ 2 senders, $\geqslant$ 2 receivers)

Unlike previous MAC and BC, we really know little about interference channel, even for 2Tx-2Rx setting the capacity is unknown. Hence we focus on 2Tx-2Rx.

Even for Gaussian 2Tx-2Rx IC, capacity not known in general.

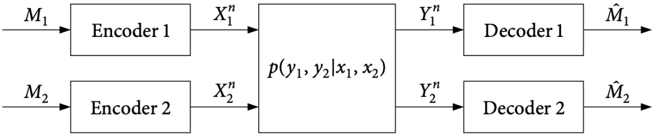


Figure: Two sender–receiver pair DM-IC.

Recall our theme: different schemes are... all about dealing interference. Let's see some examples now.

Some inner bounds and outer bounds

Fix interference (individual capacity outer bound)

Maximal achievable individual rates are

$$R_1 = C_1 \leqslant \max_{p(x_1),x_2} I(X_1; Y_1|X_2 = x_2), R_2 = C_2 \leqslant \max_{p(x_2),x_1} I(X_2; Y_2|X_2 = x_1).$$

Avoid interference (time-sharing OMA inner bound)

$\forall \omega \in [0,1]$, $R_1 \leqslant \omega C_1$ and $R_2 \leqslant (1-\omega)C_2$.

Compound MAC inner bound "Decode all". Capacity for strong IC

$\forall Q$ with $p(q,x_1,x_2) = p(q)p(x_1|q)p(x_2|q)$ satisfying

$$R_1 \leqslant I(X_1, X_2; Y_1|, Q), \quad R_2 \leqslant I(X_1, X_2; Y_2|Q),$$

$$R_1 + R_2 \leqslant \min\{I(X_1, X_2; Y_1|Q), I(X_1, X_2; Y_2|Q)\}$$

Suffer interference (TIN IB) "Decode none": Conjectured capacity for (very) weak IC

$\forall Q$ with $p(q,x_1,x_2) = p(q)p(x_1|q)p(x_2|q)$ satisfying

$$R_1 \leqslant I(X_1; Y_1|Q), \quad R_2 \leqslant I(X_2; Y_2|Q).$$

Han-Kobayashi Inner Bound (1981)

Theorem (Han+Kobayashi inner bound). A rate pair (R_1, R_2) is achievable for a DM-IC $(\mathcal{X}_1 \times \mathcal{X}_2, p(y_1, y_2|x_1, x_2), \mathcal{Y}_1 \times \mathcal{Y}_2)$ if it satisfies

$$
\begin{aligned}
R_1 &\le I(X_1; Y_1|U_2, Q) && (1)\\
R_2 &\le I(X_2; Y_2|U_1, Q) && (2)\\
R_1 + R_2 &\le I(X_1, U_2; Y_1|Q) + I(X_2; Y_2|U_1, U_2, Q) && (3)\\
R_1 + R_2 &\le I(X_1; Y_1|U_1, U_2, Q) + I(X_2, U_1; Y_2|Q) && (4)\\
R_1 + R_2 &\le I(X_1, U_2; Y_1|U_1, Q) + I(X_2, U_1; Y_2|U_2, Q) && (5)\\
2R_1 + R_2 &\le I(X_1, U_2; Y_1|Q) + I(X_1; Y_1|U_1, U_2, Q) + I(X_2, U_1; Y_2|U_2, Q) && (6)\\
R_1 + 2R_2 &\le I(X_2, U_1; Y_2|Q) + I(X_2; Y_2|U_1, U_2, Q) + I(X_1, U_2; Y_1|U_1, Q) && (7)
\end{aligned}
$$

for some $p(q, u_1, u_2, x_1, x_2) = p(q)p(u_1, x_1|q)p(u_2, x_2|q)$ where $|\mathcal{U}_1| \le |\mathcal{X}_1| + 4$, $|\mathcal{U}_2| \le |\mathcal{X}_2| + 4$, and $|\mathcal{Q}| \le 7$.

Remark (HK inner bound: "decode some & suffer the rest")

In the Han-Kobayashi (HK) scheme, each user split its message to be sent into two submessages of smaller rates and power. These are known as private and common messages. Common message is intended to be decoded only at the respective receiver, the private can be decoded at both receivers.
The rationale behind this coding scheme is to decode part of the interference (the common message) and treat the rest as noise.

Examples show that HKIB cannot attain capacity in general. (Nair et al., 2015).
HKIB is only $\frac{1}{2}$ bit from capacity of 2Tx-2Rx Gaussian IC. (Etkin, Tse, Wang, 2008)

Overview

Conclusion and Summary

Why we need NOMA now? Because we want to approach channel capacity promised by information theory.

From theory's side, NOMA is the default mindset. Because as we saw, OMA schemes are (almost) always suboptimal, and we know this for decades. From practice side, SC-SIC gives channel capacity for Gaussian MAC and Gaussian BC.

According to information theory, we know MAC the most, BC the second and we do not know much about IC. (MAC, BC can be seen as a special case of IC).

So, we would try NOMA schemes on MAC and BC first. This is indeed part of the focus of topics like machine-typed comm, massive connectivity, grant-free comm, etc..

How to deal with interference from another cell in practice maybe not in the near future. If we do NOMA well enough intra-cell and just suffer interference from other cells, this will be aleady very good.

Thank You!

References

M. Vaezi, Z. Ding, and H. V. Poor, Chapter 5 of "Multiple Access Techniques for 5G Wireless Networks and Beyond". Springer Publishing Company, Incorporated, 2018.

Natasha Devroye, "The interference channel", online resource (slides), 2016

Abbas El Gamal and Young-Han Kim, "Network Information Theory" Cambridge University Press, USA, 2012.

L. Xia, Chapter 3, 4 of "On Tightness of Several Achievable Rate Regions in Network Information Theory", 2016.

Peng Xu, Zhiguo Ding, Xuchu Dai and H. Vincent Poor, "NOMA: An Information Theoretic Perspective", arXiv, 2015.