Sign-Magnitude Decomposition of Mutual Information with Polarization Effect in Digital Identification

VOLOSHYNOVSKYY, Svyatoslav, et al.

Abstract

Content identification based on digital fingerprinting attracts a lot of attention in different emerging applications. In this paper, we consider digital identification based on the sign-magnitude decomposition of fingerprint codewords and analyze the achievable rates for each component. We introduce a channel splitting approach and reveal certain interesting phenomena related to channel polarization. It is demonstrated that under certain conditions almost all rate in the sign channel is concentrated in reliable components, this can be of interest for complexity and security in various content identification applications. The envisioned extensions cover applications where the input and output alphabets of the channel are different at the encoding and decoding stages. Additionally, the reduction of the input data dimensionality at the encoding/enrollment stage can increase the cryptographic protection in terms of privacy leakage and simplify the decoding algorithms in biometric applications.

Reference


DOI: 10.1109/ITW.2011.6089485

Available at:
http://archive-ouverte.unige.ch/unige:47622

Disclaimer: layout of this document may differ from the published version.
Sign-Magnitude Decomposition of Mutual Information with Polarization Effect in Digital Identification

Svyatoslav Voloshynovskiy, Taras Holotyak, Oleksiy Koval, Fokko Beekhof and Farzad Farhadvzadeh
Department of Computer Science, University of Geneva, Switzerland

Abstract—Content identification based on digital fingerprinting attracts a lot of attention in different emerging applications. In this paper, we consider digital identification based on the sign-magnitude decomposition of fingerprint codewords and analyze the achievable rates for each component. We introduce a channel magnitude decomposition of fingerprint codewords and analyze the decoding algorithms in biometric applications. Cryptographic protection in terms of privacy leakage and simplify dimensionality at the encoding/enrollment stage can increase the and decoding stages. Additionally, the reduction of the input data output alphabets of the channel are different at the encoding and security in various content identification applications. The tagging) and security (copyright protection, broadcast moni-
toring, etc.). Most identification techniques are based on digital variables, X, corresponding small letters x and x to denote the realizations of scalar and vector random variables, respectively, i.e., x = {x(1), x(2),..., x(N)}. xS is used to denote the sign of x and xM the magnitude of x. We use X ~ f(x) to indicate that a continuous random variable X follows fX(x) and X ~ p(x) to characterize discrete random variables. h(·), H(·) and H2(·) denote differential entropy, entropy and binary entropy, respectively, while I(·) defines pair-wise mutual information.

II. A SIGN-MAGNITUDE DECOMPOSITION OF MUTUAL INFORMATION

A. General decomposition

Consider a memoryless source X with some symmetric distribution and a discrete memoryless channel (DMC) f(y|x) = ∏N i=1 f(yi|xi) with input xi = xSi · xMi and output yi = ySi · yMi, where xSi = sign(xi) and ySi = sign(yi) are the sign components and xMi = |xi| and yMi = |yi| are the magnitudes, xS ∈ XS, yS ∈ YA (XS, YA = {−1, +1}); xM ∈ XM, yM ∈ YM (XM, YM = R+).

Proposition 1 (sign-magnitude decomposition). The mutual information between channel input and output can be decomposed as shown in Fig. 1 and yields:

\[ I(X; Y) = I(X_S; Y_M) + I(X_S; Y_S | X_M) + H(Y_S | X_S, Y_M) - H(Y_S | X_S, X_M) \]

Proof: The expansion (a) in (1) follows from the chain rule for mutual information [4] that yields:

\[ I(X; Y) = I(X_S, X_M; Y_S, Y_M) = I(X_S; Y_S, Y_M | X_M) + I(X_M; Y_M | X_S, Y_S) \]

I. INTRODUCTION

Identification systems are widely used in various emerging applications ranging from identification of physical objects and people to multimedia management (content filtering, content tagging) and security (copyright protection, broadcast monitoring, etc.). Most identification techniques are based on digital fingerprints, which represent a short, robust and distinctive content description. In this case, all operations are performed on the fingerprint instead of on the original large and privacy-sensitive data, thus allowing the introduction of crypto-based security into the analog and noisy world [1].

From coding perspectives, the identification problem can be considered as a form of communication with random codewords [2], which can be further implemented based on bounded distance decoding [3]. In contrast to the high attractiveness of binary data representations for memory storage, complexity, security and privacy, an important fraction of information is neglected once the data are binarized. It is demonstrated that soft information extracted from noisy observations can enhance the overall system performance, complexity and privacy [3].

In this paper, we extend this consideration to a general decomposition of real signals into sign and magnitude components in a random projection domain. Then, we investigate the basic decomposition of mutual information between channel input and output into four terms, where two terms are dominating with one degree of freedom for each channel, and two mixed channels that cover the exchange of information across the channels. Such a decomposition makes it possible to use multistage decoding for random codewords as well as to reveal some interesting phenomena linking this consideration with rate concentration or channel polarization, where all rates can be concentrated in several components. The results are applied to Gaussian random data and an additive white Gaussian noise (AWGN) channel in the projected domain after random projections with a random sensing matrix. This decomposition might reveal some interesting insights into the analysis and design of future identification systems.

Notations. We use capital letters to denote scalar random variables X, bold capital letters to denote vector random variables X, corresponding small letters x and x to denote the realizations of scalar and vector random variables, respectively, i.e., x = {x(1), x(2), ..., x(N)}. xS is used to denote the sign of x and xM the magnitude of x. We use X ~ f(x) to indicate that a continuous random variable X follows fX(x) and X ~ p(x) to characterize discrete random variables. h(·), H(·) and H2(·) denote differential entropy, entropy and binary entropy, respectively, while I(·) defines pair-wise mutual information.
\[
I(X_S; X_M; Y_S, Y_M)
= I(X_M; Y_S, Y_M) + I(X_S; Y_S, Y_M|X_M)
= I(X_M; Y_M) + I(X_S; Y_S|X_M)
\]
\[
\text{magnitude term, } R_M \quad \text{sign term, conditioned on } X_M, R_{\|y_M}
+ I(X_M; Y_S|Y_M) + I(X_S; Y_M|X_M, Y_M).
\]

(cross-term 1) = 0 \quad \text{(cross-term 2) = 0}

(2)

Remark 1. Under the proper knowledge of BSC state \(x_M\), the mutual information \(I(X; Y)\) can be decomposed and achieved by independent processing of the magnitude and sign channels. Alternatively, one can apply multistage decoding (MSD) by first decoding the magnitude \(x_M\) and then the sign. If for some technical or security reasons \(x_M\) is unavailable and the channel state \(x_M\) is approximated by \(y_M\), one observes a rate loss proportional to the entropy of the mismatch between the exact channel state and its approximation.

Remark 2. The decomposition (1) can be very helpful to understand “soft” fingerprinting schemes, where only binary information is stored, but soft information about bit reliabilities is extracted from the noisy magnitudes [3].

In some applications such as biometrics and content identification, where the identification is based on random codebooks, the mutual information between channel input and output defines the identification capacity \(C_{id} = I(X; Y)\) [2]. The difference with the design of data transmission codebook consists in the absence of the maximization with respect to the input distribution \(f(x)\). In the following Sections we will consider the sign-magnitude decomposition of mutual information \(I(X; Y)\) and analyze some interesting dependencies for Gaussian input \(X\) and AWGN channel.

B. Decomposition of AWGN identification channel

In this section we apply decomposition (1) to the Gaussian input \(X\) and AWGN channel. Besides the great importance of the Gaussian case, it can be demonstrated that any i.i.d. pdf \(f(x)\) or correlated signals following a Gauss-Markov model of the first order can be transformed into an approximately i.i.d. Gaussian signal of lower dimensionality using random projections [5]. Therefore, under these conditions we assume that \(X \sim \mathcal{N}(0, \sigma_X^2)\) and \(Y = X + Z\), where \(Z \sim \mathcal{N}(0, \sigma_Z^2)\).

1) Magnitude term: The first term of both decompositions (2) and (4) is \(R_M = I(X_M; Y_M) = h(Y_M) - h(Y_M|X_M)\). The differential entropy \(h(Y_M)\) is \(1/2 \log \left[1/(2 \pi e (\sigma_X^2 + \sigma_Z^2))\right] \approx 1/2 \log \left[2 \pi e (\sigma_X^2 + \sigma_Z^2)\right] \sim -1\) in the entropy of the half-normal distribution, where “–1” reflects the absence of the sign information. The conditional term can be rewritten as:

\[
h(Y_M|X_M) = h(Z) - H(Y_S|X_S, X_M)
\]

(6)

that follows from the decomposition:

\[
h(Z) = h(Y|x) = h(Y_S, Y_M|X_S, X_M)
= h(Y_M|X_S, X_M) + H(Y_S|X_S, X_M, Y_M)
= h(Y_M|X_M) + H(Y_S|X_S, X_M).
\]

(7)

with differential entropy \(h(Z)\) that corresponds to the entropy of the event related to the mismatch of the signs of \(Y_S\) and \(X_S\) under \(X_M\) that can be developed as:

\[
H(Y_S|X_S, X_M) = \int_{x_M} H(Y_S|X_S, x_M) f(x_M) dx_M
\]

\[
\approx E_f[f(x_M)] H(Y_S|X_S, x_M)
\]

\[
\approx E_f[f(x_M)] H_2(p|Y_S \neq X_S|x_M),
\]

(8)
where \( f(x_M) = \frac{2}{\sqrt{2\pi}\sigma^2_x} \exp\left[-\frac{x_M^2}{2\sigma^2_x}\right] \) and
\[
P_{b|x_M} = \Pr[Y_S \neq X_S | x_M]
= \Pr[|Y_S| = -1 \mid X_S = +1, x_M] \Pr[X_S = +1] + \Pr[|Y_S| = +1 \mid X_S = -1, x_M] \Pr[X_S = -1]
= \Pr[|Y_S| = -1 \mid X_S = +1, x_M] + \Pr[|Y_S| = +1 \mid X_S = -1, x_M]
= \int_0^\infty p(y|x_M)dy
= \int_0^\infty \frac{1}{\sqrt{2\pi}\sigma^2_y} \exp\left[-\frac{(y - x_M)^2}{2\sigma^2_y}\right] dy
= Q\left(\frac{x_M}{\sigma_y}\right).
\] (9)

Substituting (9) into (8) yields:
\[
H(Y_S|X_S, X_M) = \int_{X_M} H_2\left[Q\left(\frac{x_M}{\sigma_Z}\right)\right] f(x_M)dx_M.
\] (10)

Alternatively, one can find the conditional entropy (6) as:
\[
h(Y_S(X_M)|X_M) = -\int_{X_M} I_{Y_M}(x_M) \log_2 (f(y_M|x_M)) dy_M dx_M,
\] (11)
where the conditional pdf \( f(y_M|x_M) = \frac{1}{\sqrt{2\pi\sigma^2_{y_M}}} \exp\left[-\frac{(y_M - x_M)^2}{2\sigma^2_{y_M}}\right] \) is characterized by a folded normal distribution.

Remark 3. The conditional entropy \( h(Y_M|X_M) \) converges to the entropy of \( h(Z) \) for the regime of small degradations, when the term \( H(Y_S|X_S, X_M) \) tends to zero in (6). For high degradations, the term \( H(Y_S|X_S, X_M) \) tends to 1 and \( h(Y_M|X_M) \) converges to the entropy of channel noise magnitude \( h(Z_M) = h(Z) - 1 \). Denoting the signal-to-noise ratio as \( SNR = 10\log_{10}\frac{\sigma^2_Y}{\sigma^2_{N0}} \), we present the behavior of \( h(Y_M|X_M), h(Z), \) and \( h(Z_M) \) in Fig. 2. The conditional entropy \( h(Y_M|X_M) \) was computed according to (6) and (11) by numerical integration.

Fig. 2. Behavior of conditional entropy \( h(Y_M|X_M) \) with respect to \( h(Z) \) and \( h(Z_M) \).

2) Sign term: In the decomposition (1), there are two sign terms conditioned on \( X_M \) and \( Y_M \). The sign term in the decomposition (1a) is:
\[
I(X_S; Y_S|X_M) = H(Y_S|X_M) - H(Y_S|X_S, X_M)
= H(Y_S) - H(Y_S|X_S, X_M).
\] (12)
The entropy \( H(Y_S) = 1 \) due to the symmetry of the Gaussian pdf of \( Y \); the second term is given by (10). Thus, (12) yields:
\[
R_{S|x_M} = I(X_S; Y_S|X_M)
= 1 - \int_{X_M} H_2\left[Q\left(\frac{x_M}{\sigma_Z}\right)\right] f(x_M)dx_M.
\] (13)
The sign term in the decomposition (1b) is:
\[
I(Y_S; Y_S|Y_M) = H(Y_S|Y_M) - H(Y_S|X_S, Y_M)
= 1 - H(Y_S|X_S, Y_M).
\] (14)
The term \( H(Y_S|X_S, Y_M) \) corresponds to the mismatch of the signs between input and output under \( Y_M \) and similar to (8) can be rewritten as:
\[
H(Y_S|X_S, Y_M) = \mathbb{E}_{f(y_M)}[H(Y_S|X_S, Y_M)]
= \mathbb{E}_{f(y_M)}[H_2(Pr[Y_S \neq X_S|y_M])]
= \mathbb{E}_{f(y_M)}[H_2(P_{b|y_M})],
\] (15)

where
\[
P_{b|y_M} = \Pr[|Y_S| \neq X_S|Y_M]
= \mathbb{E}_{f(x_M|y_M)}[Pr[Y_S \neq X_S|x_M]]
= \mathbb{E}_{f(x_M|y_M)}[Q\left(\frac{x_M}{\sigma_Z}\right)]
= \int_{X_M} Q\left(\frac{x_M}{\sigma_Z}\right) f(x_M|y_M)dx_M.
\] (16)

\[
\text{Remark 4. In the case of no channel state information (CSI) about the state } x_M \text{ of channel } p(y|s|x_M), \text{ the rate is:}
\]
\[
R_{S|\emptyset} = I(X_S; Y_S|\emptyset) = 1 - H_2(P_b),
\] (18)
\[
\text{where } P_b = \mathbb{E}_{f(x_M)}[P_{b|x_M}] \text{ corresponds to the average probability of bit error.}
\]

This situation corresponds to identification based on hard fingerprints, where only the sign information is used for decoding. We present the resulting rates \( R_M \) and \( R_{S|x_M} \) with respect to the capacity of the AWGN channel denoted as \( C_{id} = \frac{1}{2} \log_2 \left(1 + \frac{\sigma^2_Y}{\sigma^2_{N0}}\right) \) in Fig.3a and \( R_{S|x_M}, R_{S|y_M} \) and \( R_{S|\emptyset} \) in Fig.3b.

Remark 5 (Sign-magnitude decomposition of the mutual information for AWGN). The decomposition of \( C_{id} \) for the AWGN channel results into the rates \( R_M \) and \( R_{S|x_M} \) that in sum coincide with the capacity of the AWGN channel. At high \( SNR \), the achievable rate \( R_{S|x_M} \) converges to 1 bit/channel.
use that exactly corresponds to the difference between \( C_{id} \) and \( R_M \).

**Remark 6.** The impact of side information about the BSC state \( x_M \):

- the presence of perfect side information about the BSC state \( x_M \) enhances the rate with respect to the degraded version \( y_M \) at low SNR, where \( y_M \) is less reliable;
- the rate \( R_{S|x_M} \) approaches rate \( R_{S|x_M} \) at high SNR that indicates that the decoding in the sign channel can be performed independently from the magnitude channel;
- blind decoding without any information about \( x_M \), often used in digital fingerprinting, represents about 0.15 bit loss with respect to \( R_{S|x_M} \) (SNR \( \approx 10 \) dB).

### III. CHANNEL SPLITTING AND POLARIZATION

In this Section we introduce a practical model that achieves the above theoretical limits based on *channel splitting*. Along this way, we will demonstrate the effect of rate concentration in a few coefficients that we will refer to as *channel polarization*. The channel splitting and polarization will be demonstrated for the sign channel.

The channel splitting model assumes that the channel input vector \( x_S \) can be transmitted via several BSCs with parameters defined by the state vector \( x_M \) according to the model \( p(y_M|x_S,x_M) \). In the simplest case of 2-channel splitting, two BSCs are considered. The channel splitting is accomplished based on the available side information \( x_M \) or \( y_M \) and can be implemented by thresholding of the magnitude coefficients with a threshold \( T \). Equivalently, the same proportion of coefficients can be chosen based on sorting of the \( N \) magnitude coefficients of \( x_M \) or \( y_M \) and selecting the \( L \) largest ones. The \( L \) bits related to the large magnitude coefficients are considered as those belonging to the strong BSC with the cross-over probability \( P_S \) and the remaining to the weak one characterized by \( P_b \).

**Remark 7.** The crossover probabilities for strong and weak channels based on the perfect CSI (given \( x_M \)) are:

\[
P_{S|x_M} = \frac{1}{P_t} \int_{T_x}^{+\infty} p_{S|x_M}(x_M)dx_M,
\]

\[
P_{W|x_M} = \frac{1}{P_t} \int_{0}^{T_x} p_{W|x_M}(x_M)dx_M,
\]

where \( P_t^S = \int_{T_x}^{+\infty} f(x_M)dx_M \) and \( P_t^W = \int_{0}^{T_x} f(x_M)dx_M \) correspond to probabilities of observing the strong and weak channels, respectively. The corresponding identification rates are:

\[
R_{S|x_M}^S = P_t^S \left[ 1 - H_2 \left( P_{S|x_M} \right) \right],
\]

\[
R_{W|x_M}^W = P_t^W \left[ 1 - H_2 \left( P_{W|x_M} \right) \right],
\]

and the total rate is:

\[
R_{S|x_M}^{Ch} = R_{S|x_M}^S + R_{W|x_M}^W.
\]

**Remark 8.** The crossover probabilities for strong and weak channels based on the degraded CSI (given \( y_M \)) are:

\[
P_{S|y_M}^S = \frac{1}{P_t} \int_{T_y}^{+\infty} p_{S|y_M}(y_M)dy_M,
\]

\[
P_{W|y_M}^W = \frac{1}{P_t} \int_{0}^{T_y} p_{W|y_M}(y_M)dy_M,
\]

where \( T_y \) is selected to satisfy above \( P_t^S = \int_{T_y}^{+\infty} f(y_M)dy_M \) and \( P_t^W = \int_{0}^{T_y} f(y_M)dy_M \). The corresponding identification rates are:

\[
R_{S|y_M}^S = P_t^S \left[ 1 - H_2 \left( P_{S|y_M}^S \right) \right],
\]

\[
R_{W|y_M}^W = P_t^W \left[ 1 - H_2 \left( P_{W|y_M}^W \right) \right],
\]

and the total rate yields:

\[
R_{S|y_M}^{Ch} = R_{S|y_M}^S + R_{W|y_M}^W.
\]

**Remark 9.** The total crossover probability \( P_b \) remains the same as for the case of no CSI:

\[
P_b = \int_{X_M} P_{b|x_M}(x_M)dx_M = P_t^S P_{b|x_M}^S + P_t^W P_{b|x_M}^W.
\]

The channel splitting by the selection of the threshold \( T \) can be performed according to the several strategies:

- **Strategy 1:** maximize the total rates \( R_{S|x_M}^{Ch} \) or \( R_{S|y_M}^{Ch} \) to approach upper theoretical limits \( R_{S|x_M} \) or \( R_{S|y_M} \), respectively, that gives optimal values of thresholds \( T_{x,opt} \) and \( T_{y,opt} \) for each \( S/N_R \);
- **Strategy 2:** minimize probabilities \( P_{b|x_M}^S \) or \( P_{b|y_M}^S \) for complexity reasons and privacy amplification [3].

According to strategy 1, the binary channel splitting approaches theoretical performance limits under the optimal threshold selection as shown in Fig. 4a. The remaining gap is easily compensated by more accurate models using more channels in splitting. In fact, two-channel splitting can be considered as the first level decomposition of \( x_M = (x_{M_1},x_{M_2},...,x_{M_L}) \) according to the multilevel coding framework [6]. The optimal thresholds \( T_{x,opt} \) and \( T_{y,opt} \) are shown in Fig. 4b. To exemplify strategy 2 we show in Fig. 5a the pairs of crossover probabilities for strong and weak channels under perfect and degraded CSI that resulted from the strategy 1. Fig. 5b shows the same pairs for the fixed thresholds \( T_x \) and \( T_y \), where one can clearly observe the significant reduction of \( P_{b|x_M}^S \) or \( P_{b|y_M}^S \) that asymptotically goes to zero for \( S/N_R > 15 \) dB.
Remark 10 (Channel polarization). An interesting phenomenon is observed for $T_{x_{opt}}$ and $T_{y_{opt}}$, when practically all rate is concentrated in the strong channel, i.e., $R_{S|x,M} \approx R_{S|y,M}$ and $R_{S|y,M} \approx R_{S|x,M}$ after a certain $SNR$. It means that weak channels can be disregarded from the consideration. Moreover, the positions of the bits belonging to the strong channels can be reliably estimated. We will refer to the effect of rate concentration in strong channels as channel polarization.

The difference to polar codes, where the relationship between bits are created in such a way that conditional entropy is polarized to either 0 or 1, i.e., “weak” or “strong” bits, should be also noticed. This property has a considerable impact on the complexity. In our case, when the codewords are random, the effect of polarization can be achieved by selecting $T > T_{opt}$ (Fig. 5b), when the cross-over probability in strong channel is asymptotically equal to 0. Thus, those bits can be directly considered for identification without decoding. However, an unavoidable price for this option is rate loss in the strong channel. Multichannel splitting with multistage decoding can partially resolve this rate-complexity trade-off, that is out of the scope of this paper. The polarization effect is demonstrated in Fig. 6 as the dependence of achievable rates and crossover probabilities on the threshold $T$ for $SNR = 20 dB$ and $30 dB$.

The described phenomena can be of interest for:

- design of new search algorithms, when the representation of original content is reduced to the vector of signs of length $N$, and the decoder searches for the match of the $L$ most reliable components determined based on the noisy observations;
- joint multistage search in the random Gaussian codebooks, where the search is performed over the magnitude and signs codebooks;
- security and privacy amplification of biometrics and privacy-preserving content identification as the extension to [3].

IV. CONCLUSION

In this paper, we consider sign-magnitude decomposition of mutual information for identification applications. We consider the sign channel and demonstrate the effect of rate concentration under proper channel splitting parameters. This toy model can be of interest for the design of soft multistage decoding algorithms that can trade-off performance and complexity.

ACKNOWLEDGMENT

This paper was partially supported by SNF project 200020-134595. The authors are thankful to V. Balakirsky for his feedback on the early version of this paper.

REFERENCES