Towards a Perfect Reconstruction Theory

Olivier Rioul
LTCI Télécom Paris
Institut Polytechnique de Paris
91120 Palaiseau, France
olivier.rioul@telecom-paris.fr

Antoine Souloumiac
CEA, LIST
Université Paris-Saclay
91120 Palaiseau, France
antoine.souloumiac@cea.fr

***Abstract*—This paper presents some preliminary considerations on the general problems of missing/complementary information and perfect reconstruction, with the hope to attract the attention of the signal processing researchers that the rigorous derivation of a general theory of reconstruction should be desirable and possible.**

I. Problem Statement

There are numerous examples in signal processing where one argues in terms of an informal notion of information, such as: “This signal contains enough information to provide this or that”; or on the contrary, “information about this or that is missing in the data to achieve something”, or: “This signal has the required information to go from some low resolution to a higher resolution”, etc. The aim of this paper is to try to formalize mathematically such intuitive notions.

A. Examples

Such arguments are generally related to some notion of “perfect reconstruction” of some data. Let us illustrate with some specific examples when considering a discrete-time sequence of real or complex-valued samples X.

- If we take the modulus $|X|$, then we say that the phase information is missing. In other words, the phase signal ϕ is needed to perfectly reconstruct the original signal by multiplication $X = |X|e^{i\phi}$.
- If we filter X with some half-band low-pass filter: $Y = h * X$, then one loses information about the high frequencies in the signal, but adding the half-band high-pass filtered signal $g * X$ may yield to perfect reconstruction (just as in a perfect reconstruction filter bank). Here reconstruction occurs by addition instead of multiplication as in the preceding example.
- The wavelet coefficients corresponding to a given scale contain the necessary information to go from one coarse resolution to the next finer level of resolution. This example is similar to the preceding one.
- If we downsample a discrete-time signal X by, say, discarding every other sample, then some information is lost. In general, interpolating the resulting signal is not enough to achieve perfect reconstruction of X.
- If we hide some information X by shuffling, masking, or applying any type of one-way function, then, in principle, no information is lost because this transformation is revertible—even though such a reconstruction can be hard to carry out.

B. Formalization

Formally, one can define such problems using *deterministic functional* relationships: Let $X_1 = f_1(X)$ and $X_2 = f_2(X)$ be some deterministic functions of X. We say that X_2 contains the *missing information* to go from X_1 to X if there exists some deterministic function f such that $X = f(X_1, X_2)$. We then say that X is *perfectly reconstructed* from X_1 and X_2. This statement can of course be generalized to n signal components $X_1, X_2, \dots, X_n$.

In this statement, there are two different subproblems: (i) determine whether such a reconstruction function f exists; (ii) find the simplest possible f to achieve reconstruction efficiently. In this paper, we mainly focus on the first aspect (i). As a result, the problem is invariant by any bijective transformation on any signal X. For example, any encoded (reversible) version of X is a signal of a different form, but is completely equivalent to X as far as information is concerned:

Definition 1: We say that X and Y are *equivalent* (represent the *same information*) if there exists a bijective function ϕ such that $Y = \phi(X)$.

It is easily seen that this is an equivalence relation (in a certain considered set of signals). We may write formally $X \equiv Y$ in this case.

Definition 2: We say that Y has *less information* than X if there exists a (not necessarily bijective) function f such that $Y = f(X)$. We write $Y \leqq X$.

Clearly, $Y \leqq X$ and $X \leqq Y$ is equivalent to $X \equiv Y$.

C. Link With Information Theory

When speaking about information, one generally refers to Shannon’s theory of information [9]. In this theory, signal samples X are modeled as random variables or vectors. Shannon’s theory is based on the definition of informational quantities such as (absolute) entropy $H(X)$ or mutual information $I(X; Y)$. Such quantities are nonnegative numbers that *measure* information. A nice property is that they are compatible with the above notion of equivalence. Thus $X \equiv Y \implies H(X) = H(Y)$ since $H(\phi(X)) = H(X)$ for any bijective function ϕ. Similarly, $X \equiv X'$ and $Y \equiv Y'$ imply $I(X; Y) = I(X'; Y')$. Also, $X \leqq Y \implies H(X) \leq H(Y)$.

Evidently, X and Y can have the same quantity of information as measured by entropy, yet they may not represent the same information (they may even be completely independent). In other words, the reverse implication $H(X) = H(Y) \implies X \equiv Y$ does not hold. For this reason, Shannon himself [10], [11] defined the "true" information contained in X as X itself, modulo any reversible transformation:

Definition 3: The *"true" information* contained in X is the equivalence class of X for the equivalence relation $\equiv$.

We may, therefore, identify X with its equivalence class and write $X = Y$ instead of $X \equiv Y$. Also, the relation $Y \leqq X$ is compatible with the equivalence relation and we simply write $Y \leq X$. It is easily seen that it constitutes a *partial order* (reflexive, transitive, and antisymmetric) on the considered set of signals.

D. Missing (Complementary) Information and Perfect Reconstruction

With the above notations, the problems studied in this paper can be formally written as follows:

1) The *missing information* problem: *Given $X_1 \leq X$, how can a missing information $X_2 \leq X$ be determined such that $X \leq (X_1, X_2)$ (hence $X = (X_1, X_2)$)?*
2) The *perfect reconstruction* problem: *Under which condition on $X_1 \leq X$ and $X_2 \leq X$ can we reconstruct $X \leq (X_1, X_2)$ (hence $X = (X_1, X_2)$)?*

Notice that the pair (X_1, X_2) can be seen as the *supremum* of X_1 and X_2, denoted by $X_1 \vee X_2$, for the considered partial order $\leq$. Indeed,

$$(X_1 \leq X \text{ and } X_2 \leq X) \iff X_1 \vee X_2 \leq X. \qquad (1)$$

We may interpret $X_1 \vee X_2$ as their *total* information. Similarly, the *infimum* of X_1 and X_2, denoted by $X_1 \wedge X_2$, for the considered partial order $\leq$, is defined by

$$(X_1 \geq X \text{ and } X_2 \geq X) \iff X_1 \wedge X_2 \geq X. \qquad (2)$$

Shannon used the term *common information* [11] for $X_1 \wedge X_2$: it is the greatest information contained in both X_1 and X_2.

It is clear that the *missing information* problem, as it is stated, is trivial since one can obviously choose $X_2 = X$ to achieve perfect reconstruction. To avoid such a trivial solution, it is important to impose that X_2 is not redundant with respect to X_1 (does not share any information with X_2) in the following sense:

Definition 4: We say that X_1 and X_2 are *not redundant* if $X_1 \wedge X_2 = 0$, where 0 is any constant (deterministic random variable). We also write $X_1 \perp X_2$.

In other words, X_1 and X_2 share no common information: any function of both X_1 and X_2 is constant.

Remark 1: If X_1, X_2 are statistically *independent* (noted $X_1 \perp\!\!\!\perp X_2$) then they are not redundant: $X_1 \perp\!\!\!\perp X_2 \implies X_1 \perp X_2$ since as is easily seen, any quantity that is determined by both X_1 and X_2 is necessarily constant.

The *missing information* problem becomes the

$1'$) *Complementary information* problem: *Given $X_1 \leq X$, how can the complementary information $X_2 \leq X$ be determined such that $X_1 \wedge X_2 = 0$ and $X_1 \vee X_2 = X$?*

It would also be possible to allow for some (small) common information between X_1 and X_2 in another variant $1''$) of missing information problem. The issue is then how to quantity the common information. We shall not address this version of the missing information problem here.

E. Continuous vs. Discrete Variables

Suppose X takes values in a continuum set of reals. Then equivalence $X = Y$ requires an infinite precision of the values of the signal. This is a problem because the corresponding information measure (absolute entropy) $H(X)$ is necessarily *infinite*, since the number of bits needed to represent X is infinite.

On the other hand, the differential entropy $h(X)$ of a *continuous* random variable is not invariant by a reversible function: $h(\phi(X)) \neq h(X)$ in general for bijective ϕ, and $h(X)$ can even take negative values. Therefore it is not an adequate measure of information. In fact, one can establish the link between the differential entropy $h(X)$ and the (absolute or genuine) entropy $H(X)$ as follows. Let X_Δ be a quantized version of X with regularly spaced discrete values $\{0, \pm\Delta, \pm 2\Delta, \ldots\}$. Then under some mild conditions [1], [8],

$$H(X_\Delta) + \log \Delta \to h(X) \qquad (3)$$

as $\Delta \to 0$. Thus in the limiting case $\Delta = 0$, $H(X) = +\infty$ while $h(X)$ can be finite. For discrete $X = X_\Delta$, however, it can be checked that $h(X_\Delta) = -\infty$.

For all these reasons, one is naturally led to restrict the complementary information and perfect reconstruction problems to the case of discrete variables. When given continuous variables X, one can either

- solve the problem only for discrete variables X_Δ within a given tolerance Δ on the values taken by the signal samples;
- or solve the problem for discrete variables X_Δ for arbitrarily small $\Delta > 0$, for which the limit of solutions for $\Delta \to 0$ would constitute a solution for the original continuous case.

In the remainder of this paper, we only consider discrete random variables. Based of the information theoretic study of the lattice of information in [3], Section II solves the complementary information problem as stated by Shannon in [11], following the concept of communication class due to Gács and Körner [5]. Section III reviews the developments in [3] to give a necessary condition for perfect reconstruction and provides some solutions to the perfect reconstruction problem. Section IV concludes.

II. Finding the Complementary Information

A. Common Information

First, we need to show the existence of common information $X \wedge Y$. Such a common information is both a function of X

and of Y, of the form $f(X)=g(Y)$. As a consequence, any pair (x,y) with nonzero probability $\mathbb{P}(X=x,Y=y)>0$ will be such that the two values $f(x)$ and $g(y)$ coincide. More generally, we adopt the following definition:

Definition 5: Two values x and y *communicate*: $x\sim y$ or $y\sim x$, if there exists a "path" $xy_1x_1y_2\cdots y_nx_ny$ in which all transitions are of non zero probability: $\mathbb{P}(X=x,Y=y_1)>0$, $\mathbb{P}(Y=y_1,X=x_1)>0$, $\ldots$, $\mathbb{P}(X=x_n,Y=y)>0$.

Strictly speaking, the relation $x\sim y$ is not an equivalence relation because x and y do not belong to the same set of possible values of X and Y, respectively. However, by concatenating paths non zero transition probabilities, it is easily seen to be transitive in the sense that $x_1\sim y_1$, $y_1\sim x_2$ and $x_2\sim y_2$ imply $x_1\sim y_2$.

Definition 6 (Communication Class [5]): The *communication class* $C(x,y)$ as the set of all (x',y') such that $x'\sim y$ and $x\sim y'$.

By transitivity, two classes are either equal or distinct. Therefore, the distinct communication classes partition the set of all values (x,y) for which $\mathbb{P}(X=x)>0$ and $\mathbb{P}(Y=y)>0$.

We may identify any communication class C to its characteristic function $1_{(x,y)\in C}$ so that $C(X,Y)$ is seen as a binary random variable. In fact, this is the *common information* of X and Y:

Theorem 1 (Common Information): $X\wedge Y=C(X,Y)$.

Proof: If $Z=f(X)=g(Y)$ a.s. then Z is constant for each pair (x,y) such that $x\sim y$. Thus Z is a function of $C(X,Y)$. ■

Remark 2: An intuitive illustration of the concept of common information is as follows. Consider the stochastic matrix $\mathbb{P}(X=x,Y=y)$ which, after adequate permutations of rows/columns, has the "block diagonal" form

$$\mathbb{P}_{X,Y}=\begin{pmatrix} C_1 & & & & & \\ & C_2 & \mathbf{0} & & & \\ & & \ddots & & & \\ \mathbf{0} & & & C_k & & \\ & & & & 0 & \\ & & & & & \ddots \\ & & & & & & 0 \end{pmatrix} \qquad (4)$$

where the number k of blocks is maximal. The k rectangular matrices then represent the k different equivalence classes, and the probability $\mathbb{P}(C(X,Y)=i)$ is the sum of all entries in block C_i. Thus the effective computation of the common information is equivalent to the non-trivial identification of a "block diagonal" structure in a stochastic matrix. The algorithm proposed in [3] has quadratic complexity in the total number of possible values of X and Y.

Remark 3: From the above considerations, every pair X,Y has a infimum $X\wedge Y$ and a supremum $X\vee Y$. Therefore, the considered set of signals forms a *lattice*, which Shannon called "the lattice of information" (LoI) [11]. See [3] for a detailed study of the mathematical properties of the LoI.

B. Complementary Information

The existence of the complementary information was stated by Shannon [11] without proof. We present a simple, explicit construction.

Theorem 2 (Complementary Information): If $X_1\leq X$, there exists a complementary information X_2 such that $X_1\vee X_2=X$ and $X_1\wedge X_2=0$.

Proof: Since $X_1\leq X$, we simply have $X_1=X_1\wedge X=C(X_1,X)$. Thus, a given class $C(X_1,X)=x_1$ has only one value $X_1=x_1$ per class, corresponding in general to several values of X, say, $x_1^{x_1},x_2^{x_1},\ldots,x_{k_{x_1}}^{x_1}$. Now let $X_2\in\{1,\ldots,k_{X_1}\}$ be the unique index such that $X=X_{X_2}^{X_1}$. By construction, $X_2\leq X_1\vee X=X$, and since $X_1\leq X$, one also has $X_1\vee X_2\leq X$. But the formula $X=X_{X_2}^{X_1}$ shows that $X\leq X_1\vee X_2$, hence equality $X_1\vee X_2=X$ holds. Finally, the value $X_2=1$ connects each pair (x_1,x_2), so there is only one class according to (X_1,X_2), i.e. $X_1\wedge X_2=0$. ■

Note that the complementary information X_2 is *not* uniquely determined by X_1 and X. In the above construction, it depends on how the values of X are indexed by the class $X_1=x_1$.

Remark 4: This construction can be visualized on the stochastic tensor of (X,X_1,X_2) described in Fig. 1. The algorithm proposed in [3] has again quadratic complexity in the total number of values of X_1 and X.

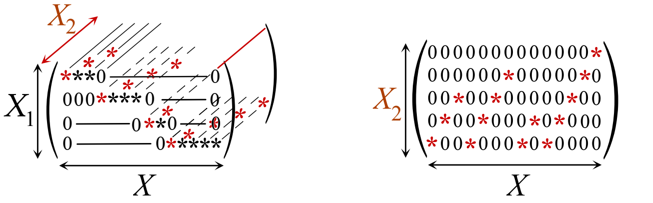

Fig. 1: Construction of the complementary information X_2 allowing to pass from X_1 to X. The stochastic tensor of (X,X_1,X_2) representing $\mathbb{P}_{X,X_1,X_2}$ has nonzero entries marked in red. The distribution $\mathbb{P}_{X,X_2}$ of (X,X_2) is obtained by marginalizing the tensor on the X_1 axis.

Remark 5: There is a striking resemblance between the complementary information problem and the "functional representation lemma" which has been used recently in network coding (see Appendix B pp. 626–627 of [4]). Indeed, Theorem 2 can be rewritten as

$$\forall\, X_1\leq X,\quad \exists X_2\perp X_1\quad \text{s.t.}\quad X=X_1\vee X_2 \qquad (5)$$

while the "functional representation lemma" [4] writes

$$\forall\, X_1,X,\quad \exists X_2\perp\!\!\!\perp X_1\quad \text{s.t.}\quad X\leq X_1\vee X_2. \qquad (6)$$

One of these two statements cannot be deduced from the other in general, because as we have seen above, $X_1\perp\!\!\!\perp X_2\implies X_1\perp X_2$ on one hand, and the second statement has a weaker hypothesis and conclusion on the other hand.

III. Conditions for Perfect Reconstruction

We base our derivation on the consideration of two entropic distances compatible with the partial order $X\leq Y$.

A. Shannon and Rajski distances

It is well-known [1] that the conditional entropy $H(X|Y)$ vanishes only when X is a deterministic function of Y:

$$X \leq Y \iff H(X|Y) = 0 \tag{7}$$

Since $X = Y \iff (X \leq Y \text{ and } X \geq Y)$, it suffices that $H(X|Y) + H(Y|X) = 0$ in order for X and Y to be equivalent: $X = Y$. Shannon [11] noted that the distance

$$D(X,Y) = H(X|Y) + H(Y|X) \tag{8}$$

satisfies a triangular inequality, and, therefore, defines a metric (this is a classical exercise, see e.g., [2]).

Interestingly, Rajski [7] noted that when X and Y are non-deterministic, by normalizing $D(X,Y)$ by the joint entropy $H(X,Y) > 0$, we obtain another distance metric taking values in $[0,1]$:

$$d(X,Y) = \frac{D(X,Y)}{H(X,Y)} \tag{9}$$

with the convention $d(0,0) = 0$. A simple proof of the triangular inequality is given in [6].

Definition 7: We say that X, Y, Z (in this order) are *aligned* for Shannon's distance D if the triangular inequality is met with equality $D(X,Y) + D(Y,Z) = D(X,Z)$, and similarly for the Rajski distance d.

From the derivation of the triangular inequalities for D and d, the following characterizations are easily established (see [3] for a detailed proof):

Lemma 1: X, Y and Z are aligned w.r.t. D if and only if $X - Y - Z$ is a Markov chain and $Y \leq X \vee Z$.

X, Y and Z are aligned w.r.t. d if and only if they satisfy the stronger condition that either $X = Y$, $Y = Z$, or $Y = X \vee Z$.

B. A Geometric Interpretation

Recall that we aim at perfectly reconstructing X from a finite number of components $X_1, X_2, \ldots, X_n$, which are defined as deterministic functions of X. For convenience in notations, let $\langle X \rangle$ be the set of *all possible* deterministic functions of X. It is easily seen that $\langle X \rangle$ is itself a sublattice of the LoI, the lattice generated by X.

Intuitively, the elements in some sequence X_1, $X_1 \vee X_2$, $\ldots, X_1 \vee X_2 \vee \ldots \vee X_n$ are getting closer and closer to X (with respect to D or d). Perfect reconstruction holds precisely when the final distance to X is zero. The following lemmas quantify how these distances to X can decrease.

We first need the following analog of Apollonius's theorem in geometry:

Lemma 2: For any $Y, Z \in \langle X \rangle$,

$$D(X, Y \vee Z) = \frac{D(X,Y) + D(X,Z) - D(Y,Z)}{2}. \tag{10}$$

This can also be written as

$$D(X,Y) + D(X,Z) = D(Y,Z) + 2D(X, Y \vee Z). \tag{11}$$

This is illustrated in Fig. 2.

Proof: Since $Y, Z \in \langle X \rangle$, we simply have $D(X,Y) = H(X) - H(Y)$, $D(X,Z) = H(X) - H(Z)$ and $D(X, Y \vee Z) = H(X) - H(Y,Z)$. Hence $D(X,Y) + D(X,Z) - 2D(X, Y \vee Z) = 2H(Y,Z) - H(Y) - H(Z) = H(Z|Y) + H(Y|Z) = D(Y,Z)$. ■

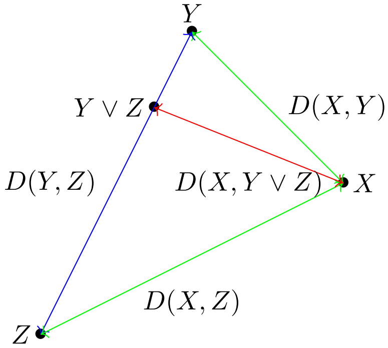

Fig. 2: Graphical representation of Apollonius's theorem (Lemma 2).

From Lemma 2 we derive the following

Lemma 3: For any $Y, Z \in \langle X \rangle$,

$$d(X,Y) + d(X,Z) \leq d(X, Y \vee Z) + 1 \tag{12}$$

with equality if and only if Y and Z are independent.

Proof: Observe that $D(Y,Z) + D(X, Y \vee Z) = H(Y|Z) + H(Z|Y) + H(X|Y \vee Z) \leq H(Y) + H(Z|Y) + H(X|Y,Z) = H(Y,Z) + H(X|Y,Z) = H(X,Y,Z) = H(X)$ since $Y, Z \in \langle X \rangle$, with equality iff Y and Z are independent. Now by Lemma 2, $D(X,Y) + D(X,Z) = D(Y,Z) + 2D(X, Y \vee Z) \leq D(X, Y \vee Z) + H(X)$. Dividing by $H(X) = H(X,Y) = H(X,Z) = H(X,Y,Z)$ yields the announced inequality. ■

C. A Necessary Condition for Perfect Reconstruction

Theorem 3 (Necessary Condition for Perfect Reconstruction): Let X be a random variable and let $X_1, X_2, \ldots, X_n \in \langle X \rangle$. If perfect reconstruction is possible: $X = X_1 \vee X_2 \vee \cdots \vee X_n$, then

$$\sum_{i=1}^{n} d(X, X_i) \leq n - 1 \tag{13}$$

with equality iff $X_1, X_2, \ldots, X_n$ are independent.

The first occurence of such a problem (for $n = 2$) is given in Exercise 6 of the French textbook [8].

Proof: By repeated use of Lemma 3, each joining operation of two components in the sum—e.g., passing from $d(X, X_i) + d(X, X_j)$ to $d(X, X_i \vee X_j)$—decreases this sum by at most 1. Thus,

$$\begin{aligned} \sum_{i=1}^{n} d(X, X_i) &\leq \sum_{i=1}^{n-2} d(X, X_i) + d(X, X_{n-1} \vee X_n) + 1 \\ &\leq \sum_{i=1}^{n-3} d(X, X_i) + d(X, X_{n-2} \vee X_{n-1} \vee X_n) + 2 \\ &\;\vdots \\ &\leq d(X, X_1 \vee X_2 \vee \cdots \vee X_n) + n - 1 = n - 1. \end{aligned} \tag{14}$$

Equality holds iff all the above $n-1$ inequalities are equalities. By the equality condition of Lemma 3, this means by induction that X_1 is independent from $X_2 \vee \cdots \vee X_n$, where X_2 is independent from $X_3 \vee \cdots \vee X_n$, and so on until X_{n-1} is independent from X_n. Overall this is equivalent to saying that all components $X_1, X_2, \dots, X_n$ are mutually independent. ■

In practice, Theorem 3 gives an *impossibility* condition for perfect reconstruction of the random variable X from components $X_1, X_2, \dots, X_n$. Indeed, if the latter are such that

$$\sum_{i=1}^{n} d(X, X_i) > n-1 \tag{15}$$

then perfect reconstruction is *impossible*, however complex the reconstruction function f might have been. In other words, $X < X_1 \vee X_2 \vee \cdots \vee X_n$, information was lost by processing.

That perfect reconstruction is impossible does not mean that it would never be possible to deduce one particular value of X from some particular values of $X_1, X_2, \dots, X_n$. It means that such a deduction is not possible in general, for every possible values taken by $X_1, X_2, \dots, X_n$. In other words, there is at least one set of values $X_1 = x_1$, $X_2 = x_2$, ..., $X_n = x_n$ for which X cannot be reconstructed unambiguously.

D. A Sufficient Condition for Perfect Reconstruction

For independent components $X_1, X_2, \dots, X_n$ (with no redundant information between them), the necessary condition of Theorem 3 becomes also a sufficient condition:

Theorem 4 (Sufficient Condition for Perfect Reconstruction): Let X be a random variable and let $X_1, X_2, \dots, X_n \in \langle X \rangle$ be *independent*. If inequality (13) holds, then it necessarily holds with equality:

$$\sum_{i=1}^{n} d(X, X_i) = n-1 \tag{16}$$

and perfect reconstruction is possible: $X = X_1 \vee X_2 \vee \cdots \vee X_n$.

Proof: A closer look at the proof of Theorem 3 shows that we have established (without the perfect reconstruction assumption) the general inequality

$$\sum_{i=1}^{n} d(X, X_i) \leq d(X, X_1 \vee X_2 \vee \cdots \vee X_n) + n - 1 \tag{17}$$

which holds with equality iff $X_1, X_2, \dots, X_n$ are independent. Therefore, by the independence assumption, (13) writes

$$\sum_{i=1}^{n} d(X, X_i) = d(X, X_1 \vee X_2 \vee \cdots \vee X_n) + n - 1 \leq n-1. \tag{18}$$

Since the distance is nonnegative, this necessarily implies that the inequality holds with equality and that $d(X, X_1 \vee X_2 \vee \cdots \vee X_n) = 0$, that is, $X = X_1 \vee X_2 \vee \cdots \vee X_n$. ■

As a simple example consider the following.

Example 1: Consider a nonzero complex-valued discrete random variable X and define the modulus and argument

$$X_1 = |X| \quad \text{and} \quad X_2 = \arg(X). \tag{19}$$

Assume that X is "isotropic" in the sense that X_1 is independent of X_2 and X_2 is uniformly distributed over M possible values. One easily computes $H(X_2) = \log M$, $H(X_1) = H(X) - \log M$, hence $d(X, X_1) + d(X, X_2) = 1$: Inequality (13) is satisfied with equality. Of course, in this trivial example, perfect reconstruction is possible since

$$X = |X| e^{\arg(X)} = X_1 e^{X_2}. \tag{20}$$

More elaborate examples of applications of Theorems 3 and 4 have been shown in [3] for specific discrete problems such as linear transformations over a finite field, integer prime factorization, Chinese remainder theorem, and optimal comparison-based sorting algorithm.

IV. Conclusion

This paper presented some considerations to provide a starting point for a "perfect reconstruction theory" that of course needs to be improved and further investigated along these lines.

In particular, it would be very valuable to illustrate the construction of the complementary information, and the necessary or sufficient conditions for perfect reconstruction by concrete examples in signal processing. This would certainly require the extension of the above geometric considerations to *continuous* random variables.

A closer look at Theorems 3 and 4 shows that they crucially depend on the standard information theoretic assertion that $H(X) \leq \sum H(X_i)$ with equality when X_i are mutually independent. This, of course, does not require all the geometric machinery presented here, but would be easily generalizable to continuous random variables using differential entropy. The construction of a genuine distance applicable to discrete or continuous random variables in order to solve the perfect reconstruction problem is a topic for future investigation.

References

[1] T. Cover and J. Thomas, *Elements of Information Theory*, John Wiley & Sons, 2nd edition, 2006.

[2] I. Csiszár and J. Körner, *Information Theory. Coding Theorems for Discrete Memoryless Systems*. Cambridge University Press, 2nd edition, 2011.

[3] I. Delsol, O. Rioul, J. Béguinot, V. Rabiet, and A. Souloumiac, "An information theoretic condition for perfect reconstruction," Entropy, vol. 26, no. 1, 86, Jan. 2024.

[4] A. El Gamal and Y.-H. Kim, *Network Information Theory*, Cambridge University Press, 2011.

[5] P. Gács and J. Körner, "Common information is far less than mutual information," *Problems of Control and Information Theory*, vol. 2, no. 2, pp. 149–162, Jan. 1973.

[6] Y. Horibe, "A note on entropy metrics," *Information and Control*, vol. 22, no. 4, pp. 403–403, May 1973.

[7] C. Rajski, "A metric space of discrete probability distributions," *Information and Control*, vol. 4, no. 4, pp. 371–377, Dec. 1961.

[8] O. Rioul, *Théorie de l'information et du codage*, Hermes Science - Lavoisier: London, 2007.

[9] C. E. Shannon, "A mathematical theory of communication," *Bell System Technical Journal*, vol. 27, no. 3 & 4, pp. 379–423 & 623–656, July & Oct. 1948.

[10] —, "Some topics on information theory," in *Proc. Int. Congress Math.*, AMS, Ed., vol. II, Aug. 30–Sept. 6, 1950, pp. 262–263.

[11] —, "The lattice theory of information," in *Transactions of the IRE Professional Group on Information Theory*, vol. 1, no. 1, pp. 105–107, Feb. 1953.