CAUSAL RATE DISTORTION FUNCTION AND RELATIONS TO FILTERING THEORY

PHOTIOS A. STAVROU* AND CHARALAMBOS D. CHARALAMBOUS†

Key words. causal rate distortion function (RDF), realization, causal filter, weak*-convergence, optimal reconstruction kernel

AMS subject classifications. 28A33, 46E15, 60G07

EXTENDED ABSTRACT. Shannon's information theory for reliable communication evolved over the years without much emphasis on real-time realizability or causality imposed on the communication sub-systems. In particular, the classical rate distortion function (RDF) for source data compression deals with the characterization of the optimal reconstruction conditional distribution subject to a fidelity criterion [1, 3], without regard for realizability.
On the other hand, filtering theory is developed by imposing real-time realizability on estimators with respect to measurement data. Although, both reliable communication and filtering (state estimation for control) are concerned with reconstruction of processes, the main underlying assumptions characterizing them are different.

In this paper, the intersection of rate distortion function (RDF) and realizable filtering theory is discussed by invoking the additional assumption that the reconstruction kernel is realizable via causal operations, while the optimal causal reconstruction kernel is derived. Consequently, the connection between causal RDF, its characterization via the optimal reconstruction kernel, and realizable filtering theory are established under very general conditions on the source (including Markov sources). The fundamental advantage of the new filtering approach based on causal RDF, is the ability to ensure average or probabilistic bounds on the estimation error, which is a non-trivial task when dealing with Bayesian filtering techniques.

The first relation between information theory and filtering via distortion rate function is discussed by R. S. Bucy in [2], by carrying out the computation of a realizable distortion rate function with square criteria for two samples of the Ornstein-Uhlenbeck Gaussian process. Related work on realizable rate distortion theory is pursued by A. K. Gorbunov and M. S. Pinsker in [5, 6]. Specifically, [5] discussed non-anticipative or causal RDF for general stationary processes and establishes existence of the infinite horizon limit while [6] computes a closed-form expression for non-anticipative or causal RDF (called ϵ-entropy) for stationary Gaussian processes using power spectral methods.
Next,we give a high level discussion on Bayesian filtering theory and we present some aspects of the problem and results pursued in this paper. Consider a discrete-time process $X^n \triangleq \{X_0, X_1, \ldots, X_n\} \in \mathcal{X}_{0,n} \triangleq \times_{i=0}^{n}\mathcal{X}_i$, and its reconstruction $Y^n \triangleq \{Y_0, Y_1, \ldots, Y_n\} \in \mathcal{Y}_{0,n} \triangleq \times_{i=0}^{n}\mathcal{Y}_i$, where $\mathcal{X}_i$ and $\mathcal{Y}_i$ are Polish spaces (complete separable metric spaces). The objective is to reconstruct X^n by Y^n causally subject to a distortion or fidelity criterion.

In classical filtering, one is given a mathematical model that generates the process

*Ph.D. student at ECE Department, University of Cyprus, Green Park, Aglantzias 91, P.O. Box 20537, 1687, Nicosia, Cyprus (photios.stavrou@ucy.ac.cy).

†Professor at ECE Department, University of Cyprus, Green Park, Aglantzias 91, P.O. Box 20537, 1687, Nicosia, Cyprus (chadcha@ucy.ac.cy).

X^n, $\{P_{X_i|X^{i-1}}(dx_i|x^{i-1}) : i = 0, 1, \ldots, n\}$ often induced via discrete-time recursive dynamics, a mathematical model that generates observed data obtained from sensors, say, Z^n, $\{P_{Z_i|Z^{i-1},X^i}\ (dz_i|z^{i-1}, x^i) : i = 0, 1, \ldots, n\}$ while Y^n are the causal estimates of some function of the process X^n based on the observed data Z^n. Note that for a memoryless channel that generates the observation sequence $\{Z_i :\ i = 0, 1, \ldots, n\}$ then $P_{Z_i|Z^{i-1},X^i}(dz_i|z^{i-1}, x^i) = P_{Z_i|X_i}(dz_i|x_i) - a.s.,\ i = 0, 1, \ldots, n$. In Bayesian estimation one is interested in causal estimators of some function $\Phi : \mathcal{X}_n \longmapsto \mathbb{R}$, $Y_n \triangleq \Phi(X_n)$ based on the observed data $Z^n \triangleq \{Z_0, Z_1, \ldots, Z_n\}$. With respect to minimizing the least-squares error pay-off the best estimate of $\Phi(X_i)$ given Z^{i-1}, denoted by $\widehat{\Phi}(X_i)$ is given by the conditional mean

$$\widehat{\Phi}(X_i) \triangleq \mathbb{E}\Big\{\Phi(X_i)|Z^{i-1}\Big\} = \int_{\mathcal{X}_i} \Phi(x) P_{X_i|Z^{i-1}}(dx|z^{i-1}),\ i = 0, 1, \ldots, n$$

For non-linear problems, Bayesian filtering is often addressed via the conditional distribution $\{P_{X_i|Z^{i-1}}(dx_i|z^{i-1}) : i = 0, 1, \ldots, n\}$ or its unnormalized versions which satisfy discrete-recursions [4], and forms a sufficient statistic for the filtering problem. The classical Kalman Filter is a well-known example for which the optimal reconstruction $\widehat{X_i} = \mathbb{E}[X_i|Z^{i-1}]$, $i = 0, 1, \ldots, n$, is the conditional mean which minimizes the average least-squares estimation error. Thus, in classical filtering theory both models which generate the unobserved and observed processes, X^n and Z^n, respectively, are given á priori. Fig. 0.1 illustrates the cascade block diagram of the filtering problem.

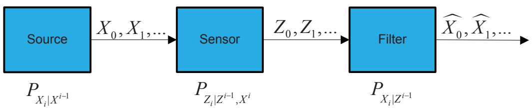

Fig. 0.1. *Block Diagram of Filtering Problem*

In causal rate distortion theory one is given the process X^n, which induces $\{P_{X_i|X^{i-1}}(dx_i|x^{i-1}) :\ i = 0, 1, \ldots, n\}$ (or the observation process in case this is observed via noisy measurements), and determines the causal reconstruction conditional distribution $\{P_{Y_i|Y^{i-1},X^i}(dy_i|y^{i-1}, x^i) :\ i = 0, 1, \ldots, n\}$ which minimizes the mutual information between X^n and Y^n subject to a distortion or fidelity constraint, via a causal (realizability) constraint. The filter $\{Y_i :\ i = 0, 1, \ldots, n\}$ of $\{X_i :\ i = 0, 1, \ldots, n\}$ is found by realizing the reconstruction distribution $\{P_{Y_i|Y^{i-1},X^i}(dy_i|y^{i-1}, x^i) :\ i = 0, 1, \ldots, n\}$ via a cascade of sub-systems as shown in Fig. 0.2. The point to

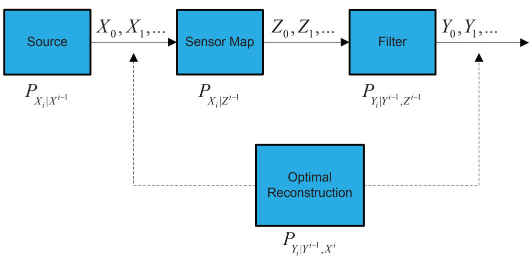

Fig. 0.2. *Block Diagram of Filtering via Causal Rate Distortion Function*

be made here is that the auxiliary random sequence $\{Z_0, Z_1, \ldots\}$ which is the analog of

sensor measurements is identified during the realization of the optimal reconstruction distribution $\{P_{Y_i|Y^{i-1},X^i}(dy_i|y^{i-1},x^i):\ i=0,1,\ldots,n\}$. Thus in Bayesian estimation, the sensor map is given á priori, while in causal rate distortion theory, this map is identified during the realization of the optimal reconstruction distribution.
The distortion function or fidelity constraint between x^n and its reconstruction y^n, is a measurable function defined by

$$d_{0,n}:\mathcal{X}_{0,n}\times\mathcal{Y}_{0,n}\mapsto[0,\infty],\ \ d_{0,n}(x^n,y^n)\triangleq\sum_{i=0}^{n}\rho_{0,i}(x^i,y^i)$$

The mutual information between X^n and Y^n, for a given distribution $P_{X^n}(dx^n)$, and conditional distribution $P_{Y^n|X^n}(dy^n|x^n)$, is defined by

$$(0.1)\quad I(X^n;Y^n)\triangleq\int_{\mathcal{X}_{0,n}\times\mathcal{Y}_{0,n}}\log\Big(\frac{P_{Y^n|X^n}(dy^n|x^n)}{P_{Y^n}(dy^n)}\Big)P_{Y^n|X^n}(dy^n|x^n)\otimes P_{X^n}(dx^n)$$

Define the $(n+1)-$fold causal convolution measure

$$(0.2)\qquad \overrightarrow{P}_{Y^n|X^n}(dy^n|x^n)\triangleq\otimes_{i=0}^{n}P_{Y_i|Y^{i-1},X^i}(dy_i|y^{i-1},x^i)-a.s.$$

The realizability constraint for a causal filter is defined by

$$(0.3)\quad \overrightarrow{Q}_{ad}\triangleq\Big\{P_{Y^n|X^n}(dy^n|x^n):\ P_{Y^n|X^n}(dy^n|x^n)=\overrightarrow{P}_{Y^n|X^n}(dy^n|x^n)-a.s.\Big\}$$

The realizability condition (0.3) is necessary, otherwise the connection between filtering and realizable rate distortion theory cannot be established. This is due to the fact that $P_{Y^n|X^n}(dy^n|x^n)=\otimes_{i=0}^{n}P_{Y_i|Y^{i-1},X^n}(dy_i|y^{i-1},x^n)-a.s.$, and hence in general, for each $i=0,1,\ldots,n$, the conditional distribution of Y_i depends on future symbols $\{X_{i+1},X_{i+2},\ldots,X_n\}$ in addition to the past and present symbols $\{Y^{i-1},X^i\}$.
Causal Rate Distortion Function. The causal RDF is defined by

$$(0.4)\qquad R^c_{0,n}(D)\triangleq\inf_{P_{Y^n|X^n}(dy^n|x^n)\in\overrightarrow{Q}_{ad}:\ E\big\{d_{0,n}(X^n,Y^n)\leq D\big\}}I(X^n;Y^n)$$

Note that realizability condition (0.3) is different from the realizability condition in [2], which is defined under the assumption that Y_i is independent of $X^*_{j|i}\triangleq X_j-\mathbb{E}\Big(X_j|X^i\Big),j=i+1,i+2,\ldots,$. The claim here is that realizability condition (0.3) is more natural and applies to processes which are not necessarily Gaussian having square error distortion function. Realizability condition (0.3) implies the causality condition in [6] defined by $X^\infty_{n+1}\leftrightarrow X^n\leftrightarrow Y^n$ forms a Markov chain.
The point to be made regarding (0.4) is that:

$$\begin{aligned}P_{Y^n|X^n}(dy^n|x^n)=\overrightarrow{P}_{Y^n|X^n}(dy^n|x^n)-a.s.&\Longleftrightarrow\\ I(X^n;Y^n)=\int\log\Big(\frac{\overrightarrow{P}_{Y^n|X^n}(dy^n|x^n)}{P_{Y^n}(dy^n)}\Big)&\overrightarrow{P}_{Y^n|X^n}(dy^n|x^n)P_{X^n}(dx^n)\\ \equiv\mathbb{I}(P_{X^n},\overrightarrow{P}_{Y^n|X^n})&\end{aligned}$$

where $\mathbb{I}(P_{X^n},\overrightarrow{P}_{Y^n|X^n})$ points out the functional dependence of $I(X^n;Y^n)$ on $\{P_{X^n},\overrightarrow{P}_{Y^n|X^n}\}$.
Main Results. This paper discussed the connection between realizable or causal RDF and filtering theory. The main results of the above analysis are summarized next.

1) Existence of the causal RDF using the topology of weak*-convergence.
2) Closed-form expression of the optimal reconstruction conditional distribution for stationary processes, which is realizable via causal operations.
3) Realization procedure of the non-linear filter based on the causal RDF.

An extended version of this manuscript can be found in [7].

REFERENCES

[1] T. Berger, *Rate Distortion Theory: A Mathematical Basis for Data Compression*, Englewood Cliffs, NJ: Prentice-Hall, 1971.
[2] R. S. Bucy, *Distortion rate theory and filtering*, IEEE Transactions on Information Theory, 28 (1982), pp. 336–340.
[3] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, John Wiley & Sons, Inc., Hoboken, New Jersey, second ed., 2006.
[4] R. J. Elliott, L. Aggoun, and J. B. Moore, *Hidden Markov Models: Estimation and Control*, Springer-Verlag, Berlin, Heidelberg, New York, 1995.
[5] A. K. Gorbunov and M. S. Pinsker, *Nonanticipatory and prognostic epsilon entropies and message generation rates*, Problems of Information Transmission, 9 (1973), pp. 184–191.
[6] ———, *Asymptotic behavior of nonanticipative epsilon-entropy for Gaussian processes*, Problems of Information Transmission, 27 (1991), pp. 361–365.
[7] P. A. Stavrou and C. D. Charalambous, *Causal rate distortion function and relations to filtering theory*, preprint, abs/1104.0862 (2012).