International Journal of Applied Mathematical Sciences.
ISSN 0973-0176 Vol.3 No.2 (2006), pp. 141-165

http://www.gbspublisher.com/ijams.htm

Rate Distortion Theory for General Sources With Potential Application to Image Compression

Farzad Rezaei, N.U. Ahmed[1] **and Charalambos D. Charalambous**[2]
School of Information Technology and Engineering
University of Ottawa
Ottawa, CANADA K1N 6N5
E-mail: frezaei@site.uottawa.ca, ahmed@site.uottawa.ca,
E-mail:chadcha@ucy.ac.cy

Abstract

This paper introduces a general framework for dealing with the subject of rate distortion or source coding with fidelity criterion, in Polish spaces. Previous well known results for abstract alphabets are extended to the more general case in which marginal measures on the reproduction space may not be absolutely continuous with respect to the optimal marginal measure. In this situation, the chain rule for Radon-Nikodym derivatives does not hold and hence standard techniques do not apply. This problem is resolved by developing a variational approach on the space of measure valued functions. It is also shown that the compactness assumption on the reproduction space can be relaxed. Moreover, the question of existence of a solution to the implicit equation of optimal distribution is addressed by proving a fixed point theorem. In the discrete alphabet case, the existence of such a solution follows from the Blahut algorithm. However in the abstract alphabet case, more analysis is needed. This is formulated as a fixed point problem on a suitable locally convex topological vector space. Existence of solution of the original problem is proved by establishing the existence of a fixed point.

AMS subject classification: 62B10,94A34, 94A17.
Keywords and Phrases: Rate distortion theory, Banach spaces, Polish spaces, Finitely additive measures, Lower semi-continuity, Fixed point theorem.

---

[1]Also with the Department of Mathematics, University of Ottawa.

[2]E-mail: chadcha@ucy.ac.cy. Also with the Department of Electrical and Computer Engineering, University of Cyprus, 75 Kallipoleos Avenue, Nicosia, CYPRUS.

This work was supported by the European Commission under the project ICCCSYSTEMS.

1 Introduction

The conventional rate distortion formulation for finite-alphabet sources and also continuous sources has been studied thoroughly in the literature [1], [2], [13] and [15]. For a recent survey of the theory of rate distortion see [9]. In this paper, we consider the problem of characterization of rate distortion function in Polish spaces for abstract sources. The formulation of rate distortion function for abstract alphabets has been studied by Csiszár [3]. The question of existence of solution in Polish spaces under some continuity assumptions on the distortion function and compactness of the reproduction space, was resolved under the topology of weak convergence. The formulation in [3] is based on two important assumptions, namely, 1) compactness of the reproduction space, 2) absolute continuity of all marginal distributions with respect to the optimal marginal distribution. The compactness assumption is crucial in [3] in order to formulate the problem using countably additive measures and to show existence of minimizing measure using tightness arguments and Prohorov theorem. The absolute continuity assumption is vital throughout the derivations for the optimal solution, because it enables one to apply chain rule of Radon-Nikodym derivatives. Under these assumptions, the optimal solution is derived and it is given by

$$q^*(x, dy) = \frac{e^{s\rho(x,y)}\nu^*(dy)}{\int_{\hat{A}} e^{s\rho(x,z)}\nu^*(dz)} \tag{1}$$

where ρ is the distortion function, q^* is the optimal conditional distribution, ν^* is the optimal marginal distribution and $\hat{A}$ is the reproduction space. Next, we elaborate more on the above assumptions and the restrictions imposed by them on the formulation of rate distortion problems.

Compactness. Indeed the compactness assumption is too restrictive. In some applications such as compression using wavelet transforms [8], the reproduction space is the linear span of an infinite basis, which is not compact. More discussion about the issue of compactness in this application can be found in Section 5. In order to relax the compactness assumption, one has to use a weaker topology. In this paper, the weak* topology is considered. This topology has recently been used in the literature in the context of channel capacity for continuous alphabets [12]. Using this topological framework, the compactness assumption on the reproduction space is relaxed. However, in order to formulate the problem in the weak* topology, identifying the appropriate dual Banach spaces is crucial This will be discussed in Section 2.

Absolute Continuity. Csiszár [3] uses an assumption which implies that all marginal distributions on the reproduction space are absolutely continuous with respect to the marginal distribution corresponding to the optimal solution. This crucial assumption, enables one to use the chain rule for Radon-Nikodym derivatives [[3], Lemmas 1.3-1.4, p60]. In general, such an assumption may not be valid, hence necessary condition of optimality must be derived using calculus of variations on suitable space of functions with values in the space of measures. One of the main results of this paper, is to show that the mutual information is Gateaux differentiable with respect to the conditional distribution defined on the reproduction space. This property relaxes the absolute continuity assumption. The Gateaux differentiability is then employed to derive necessary and sufficient conditions for optimality. The forms of the optimal distributions are the same as in the abstract alphabet case in [3].

Implicit form of the minimizer. One of the fundamental issues that is yet to be addressed, is whether the nonlinear equation in (1) has a solution. For the case of finite alphabet, the analog of (1), has a solution due to the Blahut algorithm [2], because in the limit the algorithm leads to an equation like (1). It should be pointed out that in this problem we are essentially dealing with two different existence problems. The first one is concerned with the existence of a conditional measure on the reproduction space, which minimizes the mutual information. However, it does not provide us with any specific minimizing candidate. Such a candidate is found later using the variational method in Section 4 (or chain rule and inequalties as in [3]). This candidate does not have an explicit form, rather it is an implicit equation describing a nonlinear relationship between the conditional and marginal distributions on the reproduction space. Such an equation may not have solutions in general, and therefore the second existence problem appears, i.e., existence of a solution to the nonlinear equation in (1). If this equation does not have a solution, then the minimizing measure exists but does not have the form given in that equation.

There is an extension of the Blahut algorithm in the general case, in which $\lim_{n\to\infty} I(\mu; q_n) = R(D)$, where $\{q_n\}$ is a sequence constructed via the alternating minimization method. Moreover, one can find a measure q_0 such that $q_n \to q_0$ weakly, and by the lower semi-continuity of mutual information, $I(\mu; q_0) = R(D)$ [[6], p.217-218] . However, unlike the finite dimensional case, it is not possible to show that q_0 and its marginal ν_0 will satisfy (1), since the convergence $q_n \to q_0$ is not in the strong topology [3]. To the best of the authors knowledge, the issue of existence of solution to (1) has not been addressed in the literature. In this paper, existence of a solution to the implicit nonlinear equation (1), is proved using Tihonov Fixed Point theorem which holds for locally convex topological vector spaces. Without this existence result, the necessary conditions of optimality do not hold, since in that proof the solution is an implicit one, and if the implicit relationship does not have a solution, then the optimality condition would be meaningless.

Source coding theorem with fidelity criterion for abstract sources has been addressed in many papers. For separable metric spaces results in this direction can be found in [4]. This result can be applied to the set up considered in this paper. Alternative approaches based on Large Deviation techniques are found in

[3] $q_n \to q_0$ strongly iff $q_n(x, F) \to q_0(x, F)$ for all x and all F in the σ-algebra on the reproduction space.

[7], while methods based on generalized AEP (asymptotic equipartition property) are given in [11]. A source coding theorem for stationary source is given in [5].

The rest of the paper is organized as follows. In Section 2, the general rate distortion problem is formulated and appropriate function spaces are introduced. The question of existence of solution of the problem is treated in Section 3. Moreover, within the same section, the equivalence of the constrained and unconstrained problems is established. In other words, the conditions required for using Lagrange multipliers method are rigorously justified and proved. Necessary conditions of optimality are presented in Section 4. In the same section, existence of solution to a fixed point problem arising from the necessary conditions is presented. In Section 5, we have provided some examples of concrete Polish spaces of practical interest.

2 Problem Formulation

In this section, the appropriate topologies and function spaces are identified and the weak* compactness of the constrained set is shown. Let $(A, \mathcal{A})$ be a measurable space denoting the source space with $\mathcal{A}$ being the sigma algebra of subsets of the set A generated by closed sets. Similarly let $(\hat{A}, \hat{\mathcal{A}})$ be another measurable space denoting the reproduction space with $\hat{\mathcal{A}}$ being the sigma algebra of subsets of the set $\hat{A}$. The reproduction space may be a proper subset of the source space $\hat{A} \subseteq A$.

Assume $q : A \times \hat{\mathcal{A}} \to [0, 1]$ is a mapping satisfying the following two properties:
1) For every $x \in A$, the set function $q(x, \cdot)$ is a probability measure (possibly finitely additive) on $\mathcal{A}$.
2) For every $F \in \hat{\mathcal{A}}$, the function $q(\cdot, F)$ is $\mathcal{A}$-measurable.
Any such map q is called a stochastic kernel or transition probability. Let Q denote the class of all such stochastic kernels.
Let $\mathcal{M}_1(A)$ denote the space of probability measures (possibly finitely additive) on the source space A and let $\mu \in \mathcal{M}_1(A)$ be fixed. For the given pair $\{q, \mu\}$ we may introduce three probability measures as follows:

(P1): the joint (or compound) probability measure $P \in \mathcal{M}_1(A \times \hat{A})$ given by

$$P(G) = (\mu \otimes q)(G) = \int_A q(x, G_x)\mu(dx) \qquad \forall G \in \mathcal{A} \times \hat{\mathcal{A}}$$

where $\otimes$ denotes the convolution, G_x is the x-section of G defined by $G_x = \{y \in \hat{A} : (x, y) \in G\}$;

(P2): the marginal probability measure $\nu \in \mathcal{M}_1(\hat{A})$ corresponding to $q \in Q$ is given by

$$\nu(F) = P(A \times F) = \int_A q(x, (A \times F)_x)\mu(dx) = \int_A q(x, F)\mu(dx) \qquad \forall F \in \hat{\mathcal{A}};$$

and finally

(P3): the product measure π of μ and ν is given by

$$\pi(G) = (\mu \times \nu)(G) = \int_A \nu(G_x)\mu(dx) \qquad \forall G \in \mathcal{A} \times \hat{\mathcal{A}}.$$

Let $\rho : A \times \hat{A} \to [0, \infty)$ be a $\mathcal{A} \times \hat{\mathcal{A}}$-measurable function. For each $D \in [0, \infty)$, define the set $Q(D)$ as

$$Q(D) = \{q \in Q : \int_A \int_{\hat{A}} \rho(x, y)q(x, dy)\mu(dx) \leq D\}.$$

For the given ρ and μ we assume that $Q(D)$ is nonempty. Necessary and sufficient conditions for this are given in the sequel.

In the following definition, the mutual information is defined using Radon-Nikodym derivatives. An alternative way is to define it by taking supremum over finite partitions over the product space [14](p.89 and p.122). This alternative definition is used in the appendix to show Gateaux differentiability of the mutual information.

Definition 2.1 (Rate Distortion Function) *Corresponding to any pair $(\mu, q) \in M_1(A) \times Q$, the relative entropy of the associated joint probability measure P with respect to the product measure $\pi = \mu \times \nu$ of its marginals is called the mutual information which is denoted by*

$$I(\mu; q) \equiv H(P||\pi) = \int_{A \times \hat{A}} \log\Big(\frac{q(x, dy)}{\nu(dy)}\Big) q(x, dy)\mu(dx),$$

where the argument of the logarithmic function denotes the Radon-Nikodym derivative of $q(x, \cdot)$ with respect to its marginal $\nu(\cdot)$. The rate distortion function is then given by,

$$R(D) \equiv \inf_{q \in Q(D)} I(\mu; q). \tag{2}$$

Remark 2.2 *The interpretation of $R(D)$ is that it specifies the minimum amount of information required to reproduce the source with average distortion (fidelity) equal or less than D.*

Throughout the rest of the paper we assume that the function $I_\mu(q) \equiv I(\mu; q)$ is well defined for each $q \in Q$. This of course requires that for every $q \in Q$, $q(x, \cdot)$ has Radon-Nikodym derivative with respect to its marginal $\nu(\cdot)$. Necessary and sufficient conditions for existence of RND for finitely additive measures can be found in [24]. Next we use a theorem known as chain rule for relative entropy.

Lemma 2.3 (Chain Rule) Let $\mathcal{X}$ and $\mathcal{Y}$ be polish [4] spaces and α and β two probability measures on $\mathcal{X} \times \mathcal{Y}$. Let α_1 and β_1 denote the first marginals of α and β (i.e., $\alpha_1(E) = \alpha(E \times \mathcal{Y})$ and $\beta_1(E) = \beta(E \times \mathcal{Y})$, $\forall E \in \mathcal{B}(\mathcal{X})$) respectively, and

[4] A complete separable metric space

$\alpha_2(x, dy)$ and $\beta_2(x, dy)$ the stochastic kernels on $\mathcal{Y}$ given $\mathcal{X}$ such that the following decompositions hold, $\alpha = \alpha_1 \otimes \alpha_2$ and $\beta = \beta_1 \otimes \beta_2$. Then the function mapping $x \mapsto H(\alpha_2(x, \cdot)||\beta_2(x, \cdot))$ is measurable and the relative entropy of α relative to β is given by

$$H(\alpha||\beta) = H(\alpha_1||\beta_1) + \int_{\mathcal{X}} H(\alpha_2(x, \cdot)||\beta_2(x, \cdot)))\alpha_1(dx)$$

Proof. See [[21], Theorem B.2.1, p 401].

By applying the chain rule to $H(P||\pi)$ as defined earlier, we find that

$$H(P||\pi) = H(\mu||\mu) + \int_A H(q(x, \cdot)||\nu(\cdot))\mu(dx) = \int_A H(q(x, \cdot)||\nu(\cdot))\mu(dx)$$

and hence we have

$$H(P||\pi) = \int_A \int_{\hat{A}} \log\Big(\frac{q(x, dy)}{\nu(dy)}\Big)q(x, dy)\mu(dx)$$

which coincides with the expression given in Definition 2.1. Assuming that $Q(D) \equiv Q_\rho(D)$ is nonempty, it follows from the Definition 2.1 that the rate distortion function is given by

$$R(D) = \inf_{q \in Q(D)} I_\mu(q) = \inf_{q \in Q(D)} \int_A \int_{\hat{A}} \log\Big(\frac{q(x, dy)}{\nu(dy)}\Big)q(x, dy)\mu(dx). \tag{3}$$

Next we introduce the appropriate topologies and function spaces used in this paper. Throughout the rest of the paper we assume that both A and $\hat{A}$ are Polish spaces (complete separable metric spaces) and so normal topological spaces. In fact, it is not necessary to restrict to Polish spaces; it suffices if both A and $\hat{A}$ are regular topological spaces. For applications however (see section 6), it is necessary to use metrizable spaces and hence Polish spaces are good choices. Let $BC(\hat{A})$ denote the vector space of bounded continuous real valued functions defined on the Polish space $\hat{A}$. Furnished with the sup norm topology, this is a Banach space. Let $\big(BC(\hat{A})\big)^*$ denote its topological dual. It is known [[16], IV.6.2, p 262] that $\big(BC(\hat{A})\big)^*$ is isometrically isomorphic to the Banach space of finitely additive regular bounded measures on $\hat{\mathcal{A}}$. Denote this by $M_{rba}(\hat{A})$ and let $\Pi_{rba}(\hat{A}) \subset M_{rba}(\hat{A})$ denote the set of regular bounded finitely additive probability measures on $\hat{A}$. Clearly if $\hat{A}$ is compact, then $\big(BC(\hat{A})\big)^*$ will be the space of countably additive measures, as in [3]. Now consider the space $L_1(\mu, BC(\hat{A}))$, i.e. the space of all μ integrable functions defined on A with values in $BC(\hat{A})$. In other words, for each $\phi \in L_1(\mu, BC(\hat{A}))$ we have

$$\parallel \phi \parallel_\mu \equiv \int_A ||\phi(x)(\cdot)||_{BC(\hat{A})}\mu(dx) < \infty.$$

With respect to this norm topology this is a Banach space. Since the Banach spaces $BC(\hat{A})$ and its dual $M_{rba}(\hat{A})$ do not satisfy RNP(Radon Nikodym property), the dual of $L_1(\mu, BC(\hat{A}))$ is not $L_\infty(\mu, M_{rba}(\hat{A}))$. However, it follows from the theory of "lifting" [[20],Theorem 7, p94, Theorem 9, p97] that the dual of the above space is $L^w_\infty(\mu, M_{rba}(\hat{A}))$, i.e., the space of all $M_{rba}(\hat{A})$ valued functions $\{q\}$ which are weak star measurable in the sense that for each $\phi \in BC(\hat{A})$, $x \longrightarrow q_x(\phi) \equiv \int_{\hat{A}} \phi(z)q(x,dz)$ is μ measurable and μ-essentially bounded. Now define the admissible set as follows

$$Q_{ad} \equiv L^w_\infty(\mu, \Pi_{rba}(\hat{A})) \subset L^w_\infty(\mu, M_{rba}(\hat{A})).$$

In other words, Q_{ad} is the unit sphere in the space $L^w_\infty(\mu, M_{rba}(\hat{A}))$. Clearly, for each $\phi \in L_1(\mu, BC(\hat{A}))$ we may define a linear functional on $L^w_\infty(\mu, M_{rba}(\hat{A}))$ by

$$\ell_\phi(q) = \int_A \Big(\int_{\hat{A}} \phi(x,y) q(x,dy) \Big) \mu(dx).$$

This is certainly a bounded linear w^*-continuous functional on $L^w_\infty(\mu, M_{rba}(\hat{\mathcal{A}}))$. Now let $\rho : A \times \hat{A} \longrightarrow [0,\infty]$ be any $\mathcal{A} \times \hat{\mathcal{A}}$ measurable (cost) function from the class $L_1(\mu, BC(\hat{A}))$ and introduce the constraint set

$$Q(D) = \{q \in Q_{ad};\ \ \ell_\rho(q) \le D\}.$$

We assume that this set is nonempty. It is clear that it is convex, bounded and w^*-closed and hence it is w^*-compact (as a w^*-closed subset of the w^*-compact set Q_{ad}). Compactness of Q_{ad} follows from the Alaoglu theorem [[16],Theorem V.4.2, 424] see also [22]].
There are certain important cases in which ρ may not be bounded, for instance when ρ is a metric of a metric space. We state a lemma which is crucial to extend the w^*-closedness property of $Q(D)$ to those cases.

Lemma 2.4 *Let $A, \hat{A}$ be two Polish spaces and $\rho : A \times \hat{A} \longrightarrow [0,\infty]$, is measurable, nonnegative, extended real valued function and also $y \rightarrow \rho(x,y)$ is continuous on $\hat{A}$, for μ-almost all $x \in A$. For any $D \in [0,\infty)$, introduce the set*

$$Q_\rho(D) \equiv \{q \in Q_{ad} : \ell_\rho(q) \equiv \int_A \Big(\int_{\hat{A}} \rho(x,y) q(x,dy) \Big) \mu(dx) \le D\}$$

and suppose it is nonempty. Then $Q_\rho(D)$ is a bounded w^* closed convex subset of Q_{ad} and hence w^* compact.
Proof. Clearly the set $Q_\rho(D)$ is bounded and convex. We prove that it is weak star closed. Let $\{q_\alpha\} \in Q_\rho(D) \subset Q_{ad}$ be a net. Since Q_{ad} is weak star compact, there exists a subnet of the net $\{q_\alpha\}$, relabeled as the original net, and an element

$q \in Q_{ad}$ such that $q_\alpha \xrightarrow{w^*} q$. [5] We must show that $q \in Q_\rho(D)$. Considering the sequence $\{\rho_k \equiv \rho \wedge k, k \in N\}$, which are bounded measurable functions (continuous in the second argument), it follows from the weak star convergence of the net $\{q_\alpha\}$ to q that

$$\int_A \Big(\int_{\hat{A}} \rho_k(x,y) q(x,dy) \Big) \mu(dx) = \lim_\alpha \int_A \Big(\int_{\hat{A}} \rho_k(x,y) q_\alpha(x,dy) \Big) \mu(dx) \tag{4}$$

for each $k \in N$. Since ρ is non-negative and $\rho_k \uparrow \rho$ as $k \to \infty$ and $q_\alpha \in Q_\rho(D)$, we have

$$\lim_\alpha \int_A \Big(\int_{\hat{A}} \rho_k(x,y) q_\alpha(x,dy) \Big) \mu(dx) \leq \lim_\alpha \int_A \Big(\int_{\hat{A}} \rho(x,y) q_\alpha(x,dy) \Big) \mu(dx) \leq D. \tag{5}$$

Combining (4) and (5) we arrive at the following inequality

$$\int_A \Big(\int_{\hat{A}} \rho_k(x,y) q(x,dy) \Big) \mu(dx) \leq D,$$

which is valid for all $k \in N$. Since $\rho_k \uparrow \rho$ and they are nonnegative, it follows from Lebesgue monotone convergence theorem and nonnegativity of stochastic kernels that

$$\int_A \Big(\int_{\hat{A}} \rho(x,y) q(x,dy) \Big) \mu(dx) \leq D.$$

This shows that the weak star limit $q \in Q_\rho(D)$ and hence we have proved that the set $Q_\rho(D)$ is a weak star closed subset of Q_{ad}. Being a weak star closed subset of a weak star compact set, it is weak star compact. This completes the proof. •

3 Existence of Solutions

In this section we study the question of existence of solution to the rate distortion problem (2) as stated in the preceding section. The methodology is based on the classical lower semi continuity of relative entropy and compactness of the set $Q(D)$. This approach is also used in [3] under stronger topologies and compactness assumption for the reproduction space. However, here we do not require compactness of $\hat{A}$, and hence the method of [3] is not applicable. In addition, our approach covers both countably additive and finitely additive measures.

Following this we demonstrate the equivalence of constrained and unconstrained problems which is used later to develop necessary conditions Since ρ is fixed, for convenience of notation, from now on we set $Q_\rho(D) = Q(D)$.

---

[5] i.e. $\left| \int_A \int_{\hat{A}} \phi(x,y) q_\alpha(x,dy) \mu(dx) - \int_A \int_{\hat{A}} \phi(x,y) q(x,dy) \mu(dx) \right| \to 0$ for any $\phi \in L_1(\mu; BC(\hat{A}))$.

Theorem 3.1 *Suppose the assumptions of Lemma 2.4 hold. Then the problem* $R(D) = \inf_{q \in Q(D)} I_\mu(q)$ *has a minimum.*

Proof. First we prove that $q \longrightarrow I_\mu(q)$ is weak star lower semi continuous. Let $\{q_\alpha\}$ be a net from Q_{ad} and suppose it is weak star convergent to q. Define the net $P_\alpha \in \Pi_{rba}(A \times \hat{A})$ given by the convolution product $P_\alpha \equiv \mu \otimes q_\alpha$. Take any $\varphi \in BC(A \times \hat{A})$ and consider the expression

$$\int_{A \times \hat{A}} \varphi(x, y) P_\alpha(dx \times dy) \equiv \int_{A \times \hat{A}} \varphi(x, y) q_\alpha(x, dy)\mu(dx).$$

Since $q_\alpha \xrightarrow{w^*} q$ in $L_\infty^w(\mu, \Pi_{rba}(\hat{A}))$, it is clear from the above expression that

$$P_\alpha \xrightarrow{w^*} P \equiv \mu \otimes q \ \text{ in } \ \Pi_{rba}(A \times \hat{A}).$$

Similarly one can easily verify that the net of the product measures $\{\pi_\alpha\}$ converges to the product measure π,

$$\pi_\alpha \equiv \nu_\alpha \times \mu \xrightarrow{w^*} \pi \equiv \nu \times \mu,$$

where $\{\nu_\alpha\}$ are the marginals of $\{P_\alpha\}$ on $\hat{A}$ and ν is its weak star limit. Now we use the lower semi continuity property of relative entropy [[21], Lemma 1.4.3, p36]. Examining the proof in [21] one can easily verify that the same procedure holds true not only for countably additive measures but also for finitely additive ones. Using this fact we conclude that

$$H(P||\pi) \leq \underline{\lim} H(P_\alpha||\pi_\alpha).$$

By definition (2.1), this is equivalent to

$$I_\mu(q) \leq \underline{\lim}_\alpha I_\mu(q_\alpha).$$

This proves weak star lower semi continuity of $I_\mu(\cdot)$ on Q_{ad}. We have already observed in Lemma 2.4 that the set $Q(D) \equiv Q_\rho(D)$ is weak* compact, and we have just seen that I_μ is w^*-lower semi continuous. Hence $I_\mu(q)$ attains its infimum on $Q(D)$. So there exists a $q^* \in Q(D)$ such that $R(D) = I_\mu(q^*)$. •

Remark 3.2 *The measures considered here are finitely additive. However, if* $\{A, \hat{A}\}$ *are compact Polish spaces, these measures are countably additive. In the noncompact case, we may use Stone-Cěck compactification turning* A *and* $\hat{A}$ *into compact Hausdorff spaces [19]* βA *and* $\beta\hat{A}$*, respectively. Using these spaces one can extend the finitely additive measures to countably additive ones. In this case one may replace* $L_\infty^w(\mu, M_{rba}(\hat{A}))$ *by* $L_\infty^w(\mu, M_{rca}(\beta\hat{A}))$.

In summary, we have shown that the rate distortion problem as stated in this paper has a solution. The next step will be to show the equivalence of constrained and unconstrained optimization problems and develop necessary conditions of optimality.

Theorem 3.3 *Suppose $\rho : A \times \hat{A} \longrightarrow \overline{R}_0 \equiv [0,\infty]$ is continuous in the second argument and the set $\Gamma \equiv \{(x,y) \in A \times \hat{A} : \rho(x,y) < D\}$ is nonempty. Then the constrained problem as stated in Theorem 3.1, is equivalent to an unconstrained problem as stated below:*

$$\begin{aligned}\inf_{q\in Q(D)} I_\mu(q) &= \max_{z\geq 0} \inf_{q\in Q(D)} \{I_\mu(q) + zG(q)\} \\ &= \max_{z\geq 0} \inf_{q\in Q(D)} \Big\{I_\mu(q) + z\Big(\int_A\int_{\hat{A}} \rho(x,y)q(x,dy)\mu(dx) - D\Big)\Big\}.\end{aligned}$$

Further the infimum occurs on the boundary of the set $Q(D)$.

Proof. Our proof is based on Lagrange Duality theorem [17], Theorem 1, p224. We choose $X \equiv L^w_\infty(\mu, M_{rba}(\hat{A}))$ which is clearly a vector space. For the set Ω the natural choice is the set $\Omega = Q_{ad} \equiv L^w_\infty(\mu, \Pi_{rba}(\hat{A}) \subseteq X$. Define

$$G(q) = \ell_\rho(q) - D \equiv \int_A\Big(\int_{\hat{A}} \rho(x,y)q(x,dy)\Big)\mu(dx) - D, \qquad q \in L^w_\infty(\mu, M_{rba}(\hat{A})).$$

It is clear that $G(\cdot)$ is a convex mapping from $L^w_\infty(\mu, M_{rba}(\hat{A}))$ into the real line with the natural ordering $(\Re, \prec) \equiv Z$. Also recall that $q \longrightarrow I_\mu(q)$ is convex and well defined on Ω and that, by Theorem 3.1, $\inf\{I_\mu(q), q \in Q(D)\}$ exists and is finite. Thus, according to the Lagrange duality theorem referred to above, it suffices to show that there exists a $q_1 \in \Omega$ such that

$$G(q_1) = \int_A\{\int_{\hat{A}} \rho(x,y)q_1(x,dy)\}\mu(dx) - D < 0.$$

Introduce the sets $A_1 \equiv \{x \in A : \Gamma_x \neq \emptyset\}$ and $A_0 \equiv A \setminus A_1$, with Γ_x denoting the x-section of Γ. Define the measure valued function q_1 as follows

$$q_1(x,\Gamma_x) = 0, \;\; \forall \;\; x \in A_0; \;\; q_1(x,\hat{A}) = 1, \;\; \forall \;\; x \in A$$

$$0 \leq q_1(x,B) \leq 1, B \subset \Gamma_x, \;\; q_1(x,\Gamma_x) = 1, \;\; \forall \;\; x \in A_1$$

where $B \in \hat{\mathcal{A}}$. Since by hypothesis $\Gamma \neq \emptyset$ we have $\mu(A_1) > 0$ and thus the kernel q_1 is well defined and it belongs to $L^w_\infty(\mu, \Pi_{rba}(\hat{A}))$. Using this kernel in the expression for $\ell_\rho(q)$, one can easily verify that $\ell_\rho(q_1) < D$ and hence $G(q_1) < 0$. Then, by the Lagrange Duality theory, we arrive at the conclusion of the theorem as stated. Also it follows from the same duality theory that if the infimum is achieved by some $q^* \in L^w_\infty(\mu, \Pi_{rba}(\hat{A}))$, then

$$z\Big(\int_A\int_{\hat{A}} \rho(x,y)q^*(x,dy)\mu(dx) - D\Big) = 0.$$

In other words, for non-zero $z \in [0,\infty)$, solution occurs on the boundary. This completes the proof. •

Remark 3.4. In the next section it will be shown that the case $z = 0$ is trivial since it leads to a situation where the optimum q is independent of $x \in A$

4 Necessary Conditions of Optimality

The following result is essential in deriving the necessary conditions of optimality. Since we did not assume absolute continuity of marginal measures as in [3], an alternative method based on calculus of variations on the space of measures is developed to find the necessary conditions of optimality.

Theorem 4.1 *Suppose $I_\mu(q) \equiv I(\mu; q)$ is well defined for every $q \in L^w_\infty(\mu, \Pi_{rba}(\hat{A}))$ possibly taking values from the set $[0, \infty]$. Then $q \to I_\mu(q)$ is Gateaux differentiable at every point in $L^w_\infty(\mu, \Pi_{rba}(\hat{A}))$, and the Gateaux derivative at the point q_0 in the direction $q - q_0$ is given by*

$$\delta I_\mu(q_0, q - q_0) = \int_A \int_{\hat{A}} \log\left(\frac{q_0(x, dy)}{\nu_0(dy)}\right)(q - q_0)(x, dy)\mu(dx)$$

where ν_0 is the marginal measure on $\hat{\mathcal{A}}$ corresponding to q_0.

Proof. See Appendix.

In view of Theorem 3.3, the constrained problem defined by (3) can be reformulated using Lagrange multipliers as follows:

$$R(D) = \inf_{q \in Q_{ad}} \{I_\mu(q) - s(\ell_\rho(q) - D)\}, \tag{6}$$

where I_μ and ℓ_ρ are as defined in the preceding section and $s \in (-\infty, 0]$ is the Lagrange multiplier. Note that Q_{ad} is a proper subset of the vector space $L^w_\infty(\mu, M_{rba}(\hat{A}))$. Thus the problem (6) is not completely free of constraints. To obtain a fully unconstrained problem we must introduce yet another Lagrange multiplier so that we have an optimization problem on the vector space $L^w_\infty(\mu, M_{rba}(\hat{A}))$ without constraints. This is presented in the proof of the following theorem.

Theorem 4.2 *The infimum in problem (6) is attained at $q^* \in L^w_\infty(\mu, \Pi_{rba}(\hat{A}))$ given by*

$$q^*(x, F) = \frac{\int_F e^{s\rho(x,y)}\nu^*(dy)}{\int_{\hat{A}} e^{s\rho(x,z)}\nu^*(dz)}, \qquad s \le 0 \tag{7}$$

where $F \in \hat{\mathcal{A}}$ and ν^* is the marginal of $P^* = \mu \otimes q^*$ on $\hat{A}$. The corresponding rate distortion function has the following form

$$R(D) = sD - \int_A \log\Big(\int_{\hat{A}} e^{s\rho(x,y)}\nu^*(dy)\Big)\mu(dx).$$

If $R(D) > 0$ then $s < 0$ and hence

$$\int_A \int_{\hat{A}} \rho(x, y)q^*(x, dy)\mu(dx) = D.$$

Proof. First note that $s \leq 0$ comes from the equivalence of constrained and unconstrained problems in Theorem 3.3. Using a pair of Lagrange multipliers $\{s, \lambda(\cdot)\}$, introduce the extended cost functional $J_{\mu,D}(q)$ as follows:

$$J_{\mu,D}(q) = I_\mu(q) - s(\ell_\rho(q) - D) + \int_A \lambda(x)\Big(\int_{\hat{A}} q(x,dy) - 1\Big)\mu(dx).$$

This is a fully unconstrained problem defined on the vector space $L_\infty^w(\mu, M_{rba}(\hat{A}))$. Since both μ and D are fixed for the given problem, for simplicity of notation we may suppress them and set $J_{\mu,D}(q) \equiv J(q)$. By use of the same technique as in Theorem 4.1, we can show that the Gateaux derivative of J on $L_\infty^w(\mu, M_{rba}(\hat{A}))$ at any point q^* in the direction $q - q^*$ is given by the following expression,

$$\begin{aligned}
\delta J(q^*; q - q^*) &= \int_A \int_{\hat{A}} \log\Big(\frac{q^*(x,dy)}{\nu^*(dy)}\Big)(q-q^*)(x,dy)\mu(dx) - s\int_A\int_{\hat{A}} \rho(x,y)(q-q^*)(x,dy)\mu(dx) \\
&\quad + \int_A\int_{\hat{A}} \lambda(x)(q-q^*)(x,dy)\mu(dx) \\
&= \int_A\int_{\hat{A}} \log\Big(e^{\lambda(x) - s\rho(x,y)}\frac{q^*(x,dy)}{\nu^*(dy)}\Big)(q-q^*)(x,dy)\mu(dx), \quad \forall q \in L_\infty^w(\mu, M_{rba}(\hat{A})).
\end{aligned}$$

Since $J(q)$ is convex in q, it follows from the basic principle of calculus of variation that a necessary and sufficient condition for q^* to be the minimizer is that $\delta J(q^*; q - q^*) = 0$ for all $q \in L_\infty^w(\mu, M_{rba}(\hat{A}))$. Since this equality holds for all $q \in L_\infty^w(\mu, M_{rba}(\hat{A}))$ (which is a linear vector space), the corresponding Gateaux gradient must vanish which requires that q^* must satisfy the following identity,

$$\frac{q^*(x,dy)}{\nu^*(dy)} = e^{-\lambda(x) + s\rho(x,y)}. \tag{8}$$

This result and the required constraint $q^* \in L_\infty^w(\mu, \Pi_{rba}(\hat{A}))$, imply that $\lambda(x)$ must satisfy the relation $\lambda(x) = \log(\int_{\hat{A}} e^{s\rho(x,y)}\nu^*(dy))$. Hence q^* is given by the following expression,

$$q^*(x,F) = \frac{\int_F e^{s\rho(x,y)}\nu^*(dy)}{\int_{\hat{A}} e^{s\rho(x,z)}\nu^*(dz)}, \qquad \forall\ x \in A \text{ and } F \in \hat{\mathcal{A}}.$$

Since $s \leq 0$ and $\rho \geq 0$, it is evident that $q^* \in L_\infty^w(\mu, \Pi_{rba}(\hat{A}))$. Substituting this in the expression for $J(q)$ we obtain

$$R(D) = J(q^*) = sD - \int_A \log\Big(\int_{\hat{A}} e^{s\rho(x,y)}\nu^*(dy)\Big)\mu(dx). \tag{9}$$

Clearly, for $s = 0$, we have $R(D) = 0$ and $q^*(x,F) = \nu^*(F)$ for μ-almost all $x \in A$. This is trivial and so we must have $s < 0$. Thus, it follows from the result of Thoerem 3.3, that the solution occurs on the boundary of $Q(D)$ giving

$$\int_A\int_{\hat{A}} \rho(x,y)q^*(x,dy)\mu(dx) = D$$

for some $s < 0$. We denote the corresponding value of s by s^*. This completes the proof of all the necessary (and sufficient) conditions for optimality as stated in the theorem. •

Remark 4.3 *It should be pointed out that the solution given in Theorem 4.2, is a nonlinear non-trivial relationship between q^* and ν^* Equation (7) is essentially a new nonlinear equation connecting those two distributions, which in turn is required to be solved. Such an equation may not have solutions in general, i.e., the form of q^* is just a candidate for the optimal minimizing measure. So far from Theorem 3.1, we already know that there exists a minimizing measure. If (7) does not have a solution, then the minimizing measure exists but does not have the form given in that equation. In the finite alphabet case, Blahut algorithm proves existence of a pair (q^*, ν^*) which satisfies the above equation. However in the abstract alphabet case, this can not be deduced from the extension of Blahut algorithm, since the convergence of measures in the extended algorithm isnot in the strong topology [6]. Hence the question of existence of a pair (q^*, ν^*) satisfying (7) should be investigated.*

Note that the expression (7) giving the optimal solution as stated in Theorem 4.2 is not explicit. This is only an implicit relationship between q^* and its marginal ν^*. So this gives rise to a fixed point problem. We must prove that this fixed point problem has a solution. In view of the expression (7), for fixed but arbitrary $s < 0$, we define the operator T_s as follows:

$$T_s(q)(x,F) \equiv \frac{\int_F e^{s\rho(x,y)}\nu(dy)}{\int_{\hat{A}} e^{s\rho(x,z)}\nu(dz)} = \frac{\int_F e^{s\rho(x,y)} \int_A q(\xi,dy)\mu(d\xi)}{\int_{\hat{A}} e^{s\rho(x,z)} \int_A q(\xi,dz)\mu(d\xi)} \qquad x \in A, F \in \hat{\mathcal{A}} (10)$$

Clearly the operator T_s maps $X \equiv L^w_\infty(\mu, M_{rba}(\hat{A}))$ into the unit ball $B_1(X)$ of X. We prove that it has a fixed point in $Q_{ad} \equiv L^w_\infty(\mu, \Pi_{rba}(\hat{A})) \subset B_1(X)$. Define

$$\rho(x,\hat{A}) \equiv \inf\{\rho(x,y), y \in \hat{A}\}, x \in A.$$

Theorem 4.4 *Suppose the triple $\{\hat{A}, \rho, \mu\}$ satisfy the following inequality*

$$\int_A \rho(x,\hat{A})\mu(dx) \le D,$$

and further, for μ almost all $x \in A$, $y \longrightarrow \rho(x,y)$ is continuous on $\hat{A}$. Then there exists an $\hat{s} \in (-\infty, 0]$ such that for each $s \in (-\infty, \hat{s}]$, T_s has a fixed point in $Q(D) \subset X$.

Proof. Take $X = L^w_\infty(\mu, M_{rba}(\hat{A}))$, and $K = Q(D) = \{q \in Q_{ad} : \ \ell_\rho(q) \le D\}$. We note that X with respect to its weak star topology is a locally convex topological vector space. Clearly the set K is convex. The assumption related to ρ is necessary for the set K to be nonempty. In fact this condition is also sufficient if the set

$$\{y \in \hat{A} : \rho(x, y) = \rho(x, \hat{A}),\ \forall\ x \in A\}$$

is nonempty. We have shown before that K is compact with respect to the weak star topology. Thus K is a nonempty, compact convex set in the locally convex space X. Given $x \in A$, $q \in X$ and $s \in (-\infty, 0]$, define the measure-valued function $T_s(q)(x, \cdot)$ as follows

$$T_s(q)(x, dy) = \frac{e^{s\rho(x,y)}\nu(dy)}{\int_{\hat{A}} e^{s\rho(x,z)}\nu(dz)}, \qquad \text{where } \nu(F) = \int_A q(x, F)\mu(dx), \quad F \in \hat{\mathcal{A}}\ \ x \in A.$$

It is obvious that $T_s(q)(x, \cdot) \in \Pi_{rba}(\hat{A})$ for each $x \in A$. Hence

$$\mu - ess \sup_{x \in A} ||T_s(q)|| = 1 < \infty, \qquad \forall s \in (-\infty, 0].$$

Clearly, then $T_s(q) \in L^w_\infty(\mu, \Pi_{rba}(\hat{A}))$. We show that for suitable $s \leq 0$, the operator T_s maps K into itself. In other words, $T_s(q)$ satisfies the inequality constraint, $\ell_\rho(T_s(q)) \leq D$ for each $q \in K$. Take any $q \in K \equiv Q(D)$. Then

$$\ell_\rho(T_s(q)) \equiv \int_A \int_{\hat{A}} \rho(x, y)T_s(q)(x, dy)\mu(dx) = \int_A \frac{\int_{\hat{A}} \rho(x, y)e^{s\rho(x,y)}\nu(dy)}{\int_{\hat{A}} e^{s\rho(x,z)}\nu(dz)}\mu(dx). \quad (11)$$

For any finite positive number r define the functions $\rho_r \equiv \rho \wedge r$ and h_r by

$$h_r(s, x) \equiv \frac{\int_{\hat{A}} \rho_r(x, y)e^{s\rho_r(x,y)}\nu(dy)}{\int_{\hat{A}} e^{s\rho_r(x,z)}\nu(dz)}, \quad s \in (-\infty, 0], x \in A.$$

Then consider the function

$$s \longrightarrow \ell_{\rho_r}(T_s(q)) = \int_A h_r(s, x)\mu(dx)$$

on the semi infinite interval $(-\infty, 0]$. Taking the derivative of the function $h_r(s, x)$ with respect to the variable s one can verify that

$$\frac{d}{ds}h_r(s, x) = \frac{d}{ds}\Big(\frac{\int_{\hat{A}} \rho_r(x, y)e^{s\rho_r(x,y)}\nu(dy)}{\int_{\hat{A}} e^{s\rho_r(x,z)}\nu(dz)}\Big)$$
$$= \Bigg(\frac{\Big(\int_{\hat{A}} \rho_r(x, y)^2 e^{s\rho_r(x,y)}\nu(dy)\Big)\Big(\int_{\hat{A}} e^{s\rho_r(x,z)}\nu(dz)\Big) - \Big(\int_{\hat{A}} \rho_r(x, z)e^{s\rho_r(x,z)}\nu(dz)\Big)^2}{\Big(\int_{\hat{A}} e^{s\rho_r(x,z)}\nu(dz)\Big)^2}\Bigg) \geq 0,$$

where non-negativity comes from the Hölder inequality. Since the function ρ_r is positive and bounded above by r, all the integrals in the numerator and denominator of the above expression are finite and positive. Thus for fixed $x \in A$ and $r > 0$, the real valued function

$$s \longrightarrow h_r(s, x)$$

is a nonnegative monotone non decreasing continuous function on $(-\infty, 0)$ having finite limit (from the left) at $s = 0$. Note that as $r \to \infty$,

$$h_r(s,x)) \to h(s,x) \equiv \frac{\int_{\hat{A}} \rho(x,y) e^{s\rho(x,y)} \nu(dy)}{\int_{\hat{A}} e^{s\rho(x,z)} \nu(dz)}$$

for each $s \in (-\infty, 0]$ and $x \in A$. In fact the convergence is uniform on compact subsets of the set $(-\infty, 0)$. Therefore $s \longrightarrow h(s,x)$ is also monotone, nondecreasing, continuous and, since μ is a positive measure, the function

$$s \longrightarrow \int_A h(s,x)\mu(dx) = \ell_\rho(T_s(q)) \equiv H(s)$$

is also a monotone nondecreasing continuous function. From this one can easily verify that

$$H(s) \equiv \int_A \int_{\hat{A}} \rho(x,y) T_s(q)(x,dy)\mu(dx) \leq \int_A \int_{\hat{A}} \rho(x,y)\nu(dy)\mu(dx) = H(0).$$

Using similar approach as given in [[10], property 1, p1097] one can show that

$$\lim_{s \downarrow -\infty} h(s,x) = \nu - ess \inf_{y \in \hat{A}} \rho(x,y). \tag{12}$$

Using this expression it follows from dominated convergence theorem that

$$\lim_{s \to -\infty} H(s) = \int_A \nu - ess \inf_{y \in \hat{A}} \rho(x,y)\mu(dx).$$

Clearly, we have the inequality

$$\int_A \int_{\hat{A}} \rho(x,y) q(x,dy)\mu(dx) \geq \int_A (\nu - ess \inf_{y \in \hat{A}} \rho(x,y))\mu(dx) = \lim_{s \to -\infty} H(s).$$

We have two possible situations: (A): q is an interior point of $Q(D)$. Then it follows from the above expression that $\lim_{s \to -\infty} H(s) < D$ and hence by the monotonicity of $s \to H(s)$, there exists an $\hat{s} \in (-\infty, 0]$ such that $T_s(q) \in Q(D)$ for $s \in (-\infty, \hat{s}]$. (B): q is on the boundary of $Q(D)$. Then $\lim_{s \to -\infty} H(s) \leq D$, and we may face either of two possibilities:
(B-1) If $\lim_{s \to -\infty} H(s) < D$, then again there exists an $\hat{s} \in (-\infty, 0]$ such that $T_s(q) \in Q(D)$ for $s \in (-\infty, \hat{s}]$.
(B-2) If $\lim_{s \to -\infty} H(s) = D$, then we combine this with $\int_A \int_{\hat{A}} \rho(x,y) q(x,dy)\mu(dx) = D$ to deduce the following

$$\int_A \Big\{ \int_{\hat{A}} \rho(x,y) q(x,dy) - (\nu - ess \inf_{y \in \hat{A}} \rho(x,y)) \Big\} \mu(dx) = 0.$$

Since the argument within the parenthesis is non-negative and μ is a positive measure, it follows from this that

$$\int_{\hat{A}} \rho(x,y) q(x,dy) = (\nu - ess \inf_{y \in \hat{A}} \rho(x,y))$$

which leads to the following

$$\rho(x,y) = (\nu - ess \inf_{y\in\hat{A}} \rho(x,y))$$

and this means that $\rho(x,y)$ must be constant in y, ν-a.s., but then for this case it follows from the definition of the operator T_s that $T_s(q)(x,F) = \nu(F)$ for all $x \in A$ and $F \in \hat{\mathcal{A}}$. Then

$$\ell_\rho(T_s(q)) = \int_A \int_{\hat{A}} \rho(x,y)\nu(dy)\mu(dx) = \int_A \rho(x,y)\mu(dx)$$

On the other hand, from the hypothesis that $\lim_{s\to-\infty} H(s) = D$, we have

$$\int_A (\nu - ess \inf_{y\in\hat{A}} \rho(x,y))\mu(dx) = D$$

which leads to $\int_A \rho(x,y)\mu(dx) = D$. Hence $T_s(q) \in Q(D)$.

This shows that under any of the conditions, (A) and (B-1) and also the trivial case (B-2), $T_s(q) \in K \equiv Q(D)$ for $s \in (-\infty, \hat{s}]$ whenever $q \in K$. Therefore for such s, the operator T_s maps K into subsets of K. Now we fix $s \in (-\infty, \hat{s}]$ and show that T_s is a continuous operator. Notice that

$$T_s(q)(x,dy) = \frac{\int_A e^{s\rho(x,y)} q(\xi,dy)\mu(d\xi)}{\int_A \int_{\hat{A}} e^{s\rho(x,z)} q(\xi,dz)\mu(d\xi)} \qquad x \in A. \tag{13}$$

Define

$$V_x(q) \equiv \int_{\hat{A}} \int_A e^{s\rho(x,y)} q(\xi,dy)\mu(d\xi),$$
$$U_x(\phi,q) \equiv \int_{\hat{A}} \int_A \phi(x,y) e^{s\rho(x,y)} q(\xi,dy)\mu(d\xi) \ \text{ for } \phi \in L_1(\mu; BC(\hat{A})),$$

and let $q_n \xrightarrow{w*} q$ as $n \to \infty$ in $L^w_\infty(\mu, M_{rba}(\hat{A}))$. Then, for each $\phi \in L_1(\mu, BC(\hat{A}))$, it follows from (13) and the above notations that

$$\begin{aligned} &\Big| \int_A \int_{\hat{A}} \phi(x,y) T_s(q_n)(x,dy)\mu(dx) - \int_A \int_{\hat{A}} \phi(x,y) T_s(q)(x,dy)\mu(dx) \Big| \\ &= \left| \int_A \left\{ \frac{U_x(\phi,q_n)}{V_x(q_n)} - \frac{U_x(\phi,q)}{V_x(q)} \right\} \mu(dx) \right|. \end{aligned} \tag{14}$$

Since $q_n \xrightarrow{w^*} q$ it follows from continuity of ρ over $\hat{A}$ and negativity of s that $\lim_{n\to\infty} V_x(q_n) = V_x(q)$ and $\lim_{n\to\infty} U_x(\phi,q_n) = U_x(\phi,q)$ for every $x \in A$. Hence $\frac{U_x(\phi,q_n)}{V_x(q_n)} - \frac{U_x(\phi,q)}{V_x(q)} \to 0$ as $n \to \infty$ for every $x \in A$. Now notice that

$$\Big| \frac{U_x(\phi,q_n)}{V_x(q_n)} - \frac{U_x(\phi,q)}{V_x(q)} \Big| \le 2||\phi||_{BC(\hat{A})}(x), x \in A.$$

Since $\phi \in L_1(\mu; BC(\hat{A}))$, this shows that the integrand in (12) is dominated by a μ-integrable function. Thus it follows from the Lebesgue Dominated convergence theorem that the expression given by (14) converges to zero as $n \to \infty$. This proves that the map $q \longrightarrow T_s(q)$ is continuous in the weak star topology. Therefore all the conditions of the Tihonov Fixed-Point Theorem, [[18], Corollary 9.6, p452], are satisfied. Hence $T_s(q)$ has a fixed point, i.e. there exists a $q \in Q(D)$ such that $T_s(q) = q$. •

Remark 4.5. Some interesting and useful properties of the rate distortion function R are as stated below.

(P1) $D \to R(D)$ is a convex, non-increasing function of D on $[0, +\infty)$ to $[0, +\infty]$.

(P2) There exists a number $D_{\max} > 0$ such that $R(D) > 0$ for all $D < D_{\max}$ and $R(D) = 0$ for all $D \geq D_{\max}$. The number $D_{\max}$ is given by

$$D_{\max} = \inf_{y \in \hat{A}} \int_A \rho(x, y) \mu(dx).$$

(P3) The solution q^* (with its marginal denoted by ν^*) of the optimization problem (3) must satisfy the following inequality

$$\int_A e^{s^* \rho(x,y)} \beta(x) \mu(dx) \leq 1, \qquad \forall y \in \hat{A}$$

where $\beta(x) = \left(\int_{\hat{A}} e^{s^* \rho(x,y)} \nu^*(dy) \right)^{-1}$.
For proof of these facts the reader is referred to [3].

5 Examples of A and $\hat{A}$ in Image Compression

In this section, we introduce some examples for the spaces used before. Suppose the source space A is a weakly compact subset of a separable Banach space E. The weak topology is metrizable [[16],Theorem V.6.3, 434] and with respect to this metric, A is a compact Polish space. For the reproduction space $\hat{A}$, one may choose any closed subset of A. Since a closed subset of a compact Polish space is also a compact Polish space, $\hat{A}$ is a compact Polish space. In this situation, the source distribution μ belongs to $M_1(A) \equiv \Pi_{rca}(A) \subset M_{rca}(A)$ and, the admissible stochastic kernels will be the set $Q_{ad} \equiv L^w_\infty(\mu, \Pi_{rca}(\hat{A})) \subset L^w_\infty(\mu, M_{rca}(\hat{A}))$, where $\Pi_{rca}(A), \Pi_{rca}(\hat{A})$ are countably additive regular measures rather than finitely additive ones as considered in the general theory. More specifically, one may take A to be the closed unit ball of a separable reflexive Banach space E. This is weakly compact and so the weak topology is metrizable and with respect to this metric topology A is compact, hence a compact Polish space. Recall that for μ to be a regular countably additive measure on a Banach space E (with separable dual E^*) it is necessary and sufficient that the corresponding covariance operator Q_μ be positive and nuclear, that is

$$Q_\mu \geq 0, \quad \text{and} \quad Tr(Q_\mu) \equiv \sum (Q_\mu e_i^*, e_i^*) \equiv \int_E \sum (e_k^*(x-m))^2 \mu(dx) < \infty, \qquad (15)$$

where $\{e_i, e_i^*\}$ is a biorthogonal basis of the pair $\{E, E^*\}$. Here $m \in E$ is the mean given by

$$e^*(m) \equiv \int_E e^*(x)\mu(dx)$$

provided it exists. The operator Q_μ has the form

$$Q_\mu \equiv \sum \lambda_i e_i \otimes e_i, \quad \text{and} \quad TrQ_\mu = \sum \lambda_i < \infty.$$

In practical communication problems involving image compression, an appropriate choice of source and reproduction spaces is important. We present a brief discussion of this here. Let Γ be a bounded measurable subset of R^n ($n = 2$ for image compression) and consider the Lebesgue spaces $\{L_p(\Gamma), 1 < p < \infty\}$. These are separable reflexive Banach spaces with the corresponding duals given by $\{L_q(\Gamma), q = (p/p-1), 1 < p < \infty\}$. Suppose we have $A \equiv B_1(E)$, the closed unit ball of $E \equiv L_p(\Gamma)$, which is weakly compact and so metrizable giving a compact Polish space. Let $\{e_i\}$ be a normalized Schauder basis for E with $\{e_i^*\} \subset E^*$ the associated dual basis. For the reproduction space $\hat{A}$, one may choose

$$\hat{A} \equiv \overline{span}\{e_i, 1 \leq i \leq N\} \cap B_1(E)$$

for any finite positive integer N. Here one natural choice for the distortion measure ρ is

$$\rho(x, y) \equiv \| x - y \| = \left(\int_\Gamma |x(\xi) - y(\xi)|^p d\xi \right)^{1/p}.$$

An approach similar to the above has been used in image compression using wavelet transforms [8], where images are considered as elements of $L_p(\Gamma)$, with $\Gamma \subset \Re^2$ a rectangle. The space of reproduced images is the linear span of a finite family of wavelets.

Remark 5.1 *In case $p = 2$, E is a Hilbert space and in this case the nuclearity of the covariance operator means that*

$$Tr(Q_\mu) \equiv \int_E \| (x - m) \|_E^2 \mu(dx) < \infty. \qquad (16)$$

In other words, the source is a second order random field with mean field $m \in E$ and covariance operator $Q_\mu \in \mathcal{L}_n^+(E)$. Clearly if μ is a Gaussian measure, the pair $\{m, Q_\mu\}$ completely characterizes the source.

Remark 5.2 *If we do not fix the number (N) of elements of the basis used in reproduction a priori, and only require that finitely many of them be used in the compression, then the reproduction space is not finite dimensional. In this case, the reproduction space will not be compact anymore and weak* topology used in Theorem 3.1 should be taken into consideration.*

6 Conclusion

In this paper we have introduced a general framework for dealing with the rate distortion (or source coding) problem with fidelity criterion, in spaces of measure valued functions. The questions of existence of optimal solutions and equivalence of constrained and unconstrained problems have been addressed. This leads to an implicit relation giving rise to a fixed point problem. Existence of a solution to this fixed point problem has been demonstrated. Some examples of source and reproduction spaces are presented illustrating the relevance of the abstract results.

7 Appendix

Proof of Theorem 4.1. Denote the marginal measures on $\hat{\mathcal{A}}$ corresponding to q and q_0, by ν and ν_0, respectively. Let $q_1 = q_0 + \epsilon(q - q_0)$ and call its marginal measure ν_1. Consider the following limit,

$$\begin{aligned}
L &= \lim_{\epsilon \downarrow 0} \Big\{ \frac{I(\mu; q_1) - I(\mu; q_0)}{\epsilon} - \int_A \int_{\hat{A}} \log\Big(\frac{q_0(x,dy)}{\nu_0(dy)}\Big)(q - q_0)(x,dy)\mu(dx) \Big\} \\
&= \lim_{\epsilon \downarrow 0} \Big\{ \frac{1}{\epsilon}\Big(\int_A \int_{\hat{A}} \log\Big(\frac{q_1(x,dy)}{\nu_1(dy)}\Big) q_0(x,dy)\mu(dx) - \int_A \int_{\hat{A}} \log\Big(\frac{q_0(x,dy)}{\nu_0(dy)}\Big) q_0(x,dy)\mu(dx)\Big) \\
&+ \int_A \int_{\hat{A}} \log\Big(\frac{q_1(x,dy)}{\nu_1(dy)}\Big)(q - q_0)(x,dy)\mu(dx) - \int_A \int_{\hat{A}} \log\Big(\frac{q_0(x,dy)}{\nu_0(dy)}\Big)(q - q_0)(x,dy)\mu(dx) \Big\}.
\end{aligned} \tag{17}$$

Consider the first two terms of (17). Using a generalized definition of mutual information [14], we have

$$\begin{aligned}
I &= \frac{1}{\epsilon}\Big(\int_A \int_{\hat{A}} \log\Big(\frac{q_1(x,dy)}{\nu_1(dy)}\Big) q_0(x,dy)\mu(dx) - \int_A \int_{\hat{A}} \log\Big(\frac{q_0(x,dy)}{\nu_0(dy)}\Big) q_0(x,dy)\mu(dx)\Big) \\
&= \frac{1}{\epsilon} \int_A \Big(\sup_{P \in \mathcal{P}(\hat{A})} \sum_{E \in P} \log\Big(\frac{q_1(x,E)}{\nu_1(E)}\Big) q_0(x,E) - \sup_{P \in \mathcal{P}(\hat{A})} \sum_{E \in P} \log\Big(\frac{q_0(x,E)}{\nu_0(E)}\Big) q_0(x,E)\Big)\mu(dx)
\end{aligned} \tag{18}$$

where $\mathcal{P}(\hat{A})$ denotes the collection of all finite partitions of $\hat{A}$. Now, given $\delta > 0$ and $x \in A$, there exists a partition P' of $\hat{A}$ such that

$$\sup_{P \in \mathcal{P}(\hat{A})} \sum_{E \in P} \log\Big(\frac{q_1(x,E)}{\nu_1(E)}\Big) q_0(x,E) \leq \sum_{E \in P'} \log\Big(\frac{q_1(x,E)}{\nu_1(E)}\Big) q_0(x,E) + \delta\epsilon.$$

Hence (18) can be written as

$$I \leq \frac{1}{\epsilon}\int_A \Big(\sum_{E \in P'} \log\Big(\frac{q_1(x,E)}{\nu_1(E)}\Big) q_0(x,E) - \sum_{E \in P'} \log\Big(\frac{q_0(x,E)}{\nu_0(E)}\Big) q_0(x,E)\Big)\mu(dx) + \delta \tag{19}$$

In general, q_1 may not be absolutely continuous with respect to q_0 [6]. So there may exist a nontrivial $P_1 \subset P'$ so that we have $q_0(x,E) = 0$ while $q_1(x,E) \neq 0$ for any $E \in P_1$ while restricted to $P' - P_1$, $q_1 << q_0$. Using this decomposition, the inequality in (19) can be written as follows,

$$\begin{aligned} I \leq & \frac{1}{\epsilon}\int_A \Big\{ \sum_{E \in P'-P_1} \log\Big(\frac{q_1(x,E)}{\nu_1(E)}\Big) q_0(x,E) - \sum_{E \in P'-P_1} \log\Big(\frac{q_0(x,E)}{\nu_0(E)}\Big) q_0(x,E)\Big)\mu(dx) \\ & + \frac{1}{\epsilon}\int_A \sum_{E \in P_1} \log\Big(\frac{q_1(x,E)}{\nu_1(E)}\Big) q_0(x,E)\mu(dx) - \frac{1}{\epsilon}\int_A \sum_{E \in P_1} \log\Big(\frac{q_0(x,E)}{\nu_0(E)}\Big) q_0(x,E)\mu(dx) + \delta. \end{aligned}$$

The third term in the above is zero, since $q_0(x,E) = 0$ for any $E \in P_1$. Using the convention $0 \log \frac{0}{0} = 0$, as used in the definition of relative entropy [15], the fourth term is also zero, so we have

$$I \leq \frac{1}{\epsilon}\int_A \Big(\sum_{E \in P'-P_1} \log\Big(\frac{q_1(x,E)}{\nu_1(E)}\Big) q_0(x,E) - \sum_{E \in P'-P_1} \log\Big(\frac{q_0(x,E)}{\nu_0(E)}\Big) q_0(x,E)\Big)\mu(dx) + \delta.$$

Using the definition of ν_0, the above inequality reduces to the following one,

$$I \leq \sum_{E \in P'-P_1} \int_A \frac{1}{\epsilon} \log\Big(\frac{q_1(x,E)}{q_0(x,E)}\Big) q_0(x,E)\mu(dx) - \sum_{E \in P'-P_1} \frac{1}{\epsilon}\nu_0(E) \log\Big(\frac{\nu_1(E)}{\nu_0(E)}\Big) + \delta. \tag{20}$$

Now we have

$$\lim_{\epsilon \downarrow 0} \frac{1}{\epsilon}\Big(\ln q_1(x,E) - \ln q_0(x,E)\Big) = \frac{(q - q_0)(x,E)}{q_0(x,E)}, \quad \forall x \in A. \tag{21}$$

Since $q_1 = q_0 + \epsilon(q - q_0)$, we have $\nu_1 = \nu_0 + \epsilon(\nu - \nu_0)$ and therefore

$$\lim_{\epsilon \downarrow 0} \frac{1}{\epsilon}\Big(\ln \nu_1(E) - \ln \nu_0(E)\Big) = \frac{(\nu - \nu_0)(E)}{\nu_0(E)}. \tag{22}$$

[6]or equivalently, q may not be absolutely continuous with respect to q_0

In order to use (21), we need to show that the integral and the limit as $\epsilon \to 0$ in (20) can be interchanged. Toward this end, consider the following inequalities,

$$\frac{1}{\epsilon}\log\Big(\frac{(1-\epsilon)q_0(x,E)+\epsilon q(x,E)}{q_0(x,E)}\Big)q_0(x,E) \geq \frac{1}{\epsilon}\log(1-\epsilon)q_0(x,E), \tag{23}$$

$$\frac{1}{\epsilon}\log\Big(\frac{(1-\epsilon)q_0(x,E)+\epsilon q(x,E)}{q_0(x,E)}\Big)q_0(x,E) \leq \frac{1}{\epsilon}\log(1-\epsilon).q_0(x,E) + \frac{1}{1-\epsilon}\frac{q(x,E)}{q_0(x,E)}q_0(x,E), \tag{24}$$

in which we have used the classical inequality, $\log x \leq x-1$ where $x>0$. By combining (23) and (24) we arrive at the following inequality,

$$\left|\frac{1}{\epsilon}\log\Big(\frac{(1-\epsilon)q_0(x,E)+\epsilon q(x,E)}{q_0(x,E)}\Big)q_0(x,E)\right| \leq \frac{1}{\epsilon}\log\Big(\frac{1}{1-\epsilon}\Big).q_0(x,E) + \frac{1}{1-\epsilon}q(x,E). \tag{25}$$

Now let $g_\epsilon(x) = \frac{1}{\epsilon}\log\big(\frac{1}{1-\epsilon}\big)q_0(x,E) + \frac{1}{1-\epsilon}q(x,E)$. It can be easily seen that limit and integration can be interchanged for $g_\epsilon(x)$. Thus it follows from (25) and a general form of the Lebesgue Dominated Convergence theorem [23][7] that we can do the same for the left side of (25). Using the results of (21) and (22) and letting $\varepsilon \to 0$ in (20) we obtain

$$\lim_{\epsilon\to 0} I \leq \sum_{E\in P'-P_1}\Big\{\int_A\Big(\frac{q(x,E)-q_0(x,E)}{q_0(x,E)}\Big)q_0(x,E)\mu(dx) - \Big(\frac{\nu(E)-\nu_0(E)}{\nu_0(E)}\Big)\nu_0(E)\Big\} + \delta. \tag{26}$$

Since ν and ν_0 are the marginals of $\mu\otimes q$ and $\mu\otimes q_0$ respectively, finally we have [8]

$$\lim_{\epsilon\to 0} I \leq \delta. \tag{27}$$

Now going back to (18), given $\delta>0$, there exists another partition P'' of $\hat{A}$ such that

$$\sup_{P\in\mathcal{P}(\hat{A})}\sum_{E\in P}\log\Big(\frac{q_0(x,E)}{\nu_0(E)}\Big)q_0(x,E) \leq \sum_{E\in P''}\log\Big(\frac{q_0(x,E)}{\nu_0(E)}\Big)q_0(x,E) + \delta\epsilon.$$

Then (18) can be written as

$$I \geq \frac{1}{\epsilon}\int_A\Big(\sum_{E\in P''}\log\Big(\frac{q_1(x,E)}{q_0(x,E)}\Big)q_0(x,E) - \sum_{E\in P''}\log\Big(\frac{\nu_1(E)}{\nu_0(E)}\Big)q_0(x,E)\Big)\mu(dx) - \delta;$$

[7] If $|f_n(x)| \leq g_n(x)$, $\forall n$, for g_n integrable, $g_n \to g$ and $f_n \to f$ as $n\to\infty$, and if $\lim_{n\to\infty}\int_A g_n(x)\mu(dx) = \int_A g(x)\mu(dx)$, then $\lim_{n\to\infty}\int_A f_n(x)\mu(dx) = \int_A f(x)\mu(dx)$.

[8] Note that in general $\sum_{E\in P'-P_1}\nu(E) < 1$ and $\sum_{E\in P'-P_1} q(x,E) < 1$ but $\sum_{E\in P'-P_1}\nu_0(E) = 1$ and $\sum_{E\in P'-P_1} q_0(x,E) = 1$, so the first term on the right hand side of (26) is zero only because ν and ν_0 are marginals of $\mu\otimes q$ and $\mu\otimes q_0$.

and by taking the limit, we arrive at the following result

$$\lim_{\epsilon \to 0} I \geq -\delta. \tag{28}$$

Since $\delta > 0$ is arbitrary, it follows from (27) and (28) that

$$\lim_{\epsilon \to 0} I = 0. \tag{29}$$

Now in (17), the last two terms can be written as follows.

$$\begin{aligned} Y = \int_A \Big(&\sup_{P \in \mathcal{P}(\hat{A})} \sum_{E \in P} \log\Big(\frac{q_1(x,E)}{\nu_1(E)}\Big)(q(x,E) - q_0(x,E)) \\ &- \sup_{P \in \mathcal{P}(\hat{A})} \sum_{E \in P} \log\Big(\frac{q_0(x,E)}{\nu_0(E)}\Big)(q(x,E) - q_0(x,E))\Big)\mu(dx). \end{aligned} \tag{30}$$

For any $\delta > 0$, there exists a partition $P' \in \mathcal{P}(\hat{A})$ such that

$$\sup_{P \in \mathcal{P}(\hat{A})} \sum_{E \in P} \log\Big(\frac{q_1(x,E)}{\nu_1(E)}\Big)(q(x,E) - q_0(x,E)) \leq \sum_{E \in P'} \log\Big(\frac{q_1(x,E)}{\nu_1(E)}\Big)(q(x,E) - q_0(x,E)) + \delta.$$

Then (30) can be upper bounded as follows.

$$\begin{aligned} Y \leq \int_A \Big\{ &\sum_{E \in P'} \log\Big(\frac{q_1(x,E)}{\nu_1(E)}\Big)(q(x,E) - q_0(x,E)) \\ &- \sum_{E \in P'} \log\Big(\frac{q_0(x,E)}{\nu_0(E)}\Big)(q(x,E) - q_0(x,E))\Big\}\mu(dx) + \delta. \end{aligned} \tag{31}$$

In order to interchange the limit and integration, we use inequalities similar to (23) and (24) giving

$$\Big|\log\Big(\frac{q_1(x,E)}{q_0(x,E)}\Big)(q(x,E) - q_0(x,E))\Big| \leq \Big(\log\Big(\frac{1}{1-\epsilon}\Big) + \frac{\epsilon}{1-\epsilon}.\frac{q(x,E)}{q_0(x,E)}\Big).(q(x,E) + q_0(x,E)).$$

Let $f_\epsilon(x) = \Big(\log\Big(\frac{1}{1-\epsilon}\Big) + \frac{\epsilon}{1-\epsilon}.\frac{q(x,E)}{q_0(x,E)}\Big).(q(x,E) + q_0(x,E))$. Then we can obviously interchange the limit and integration for this function. Hence we can do the same in (31) and this leads to

$$\lim_{\epsilon \to 0} Y \leq \delta. \tag{32}$$

Similarly, given $\delta > 0$, there exists a partition $P'' \in \mathcal{P}(\hat{A})$ such that

$$\sup_{P \in \mathcal{P}(\hat{A})} \sum_{E \in P} \log\Big(\frac{q_0(x,E)}{\nu_0(E)}\Big)(q(x,E) - q_0(x,E)) \leq \sum_{E \in P''} \log\Big(\frac{q_0(x,E)}{\nu_0(E)}\Big)(q(x,E) - q_0(x,E)) + \delta.$$

Then (30) can be lower bounded as follows

$$\begin{aligned} Y \geq \quad & \int_A \Big\{ \sum_{E\in P''} \log\Big(\frac{q_1(x,E)}{\nu_1(E)}\Big)(q(x,E)-q_0(x,E)) \\ & - \sum_{E\in P''} \log\Big(\frac{q_0(x,E)}{\nu_0(E)}\Big)(q(x,E)-q_0(x,E))\Big\}\mu(dx) - \delta. \end{aligned}$$

Then by taking limits we obtain

$$\lim_{\epsilon\to 0} Y \geq -\delta \tag{33}$$

By combining (32) and (33) and letting $\delta \to 0$ we arrive at the following result,

$$\lim_{\epsilon\to 0} Y = 0. \tag{34}$$

It follows from (29),(34) and (17) that

$$L = \lim_{\epsilon\to 0}(I+Y) = 0.$$

Thus we may now conclude that $q \longrightarrow I_\mu(q) \equiv I(\mu;q)$ is Gateaux differentiable on Q_{ad} and that its Gateaux differential at $q_0 \in Q_{ad}$ in the direction $q - q_0$ is given by

$$\begin{aligned} I_\mu(q_0, q-q_0) \quad & \equiv \lim_{\epsilon\downarrow 0} \frac{I(\mu; q_0+\epsilon(q-q_0)) - I(\mu;q_0)}{\epsilon} \\ & = \int_A\int_{\hat{A}} \log\Big(\frac{q_0(x,dy)}{\nu_0(dy)}\Big)(q-q_0)(x,dy)\mu(dx). \end{aligned}$$

Moreover, since the form of the initial logarithm function has not changed, $I(\mu;q)$ is continuously Gateaux differentiable. •

References

[1] T. Berger, *Rate Distortion Theory: A Mathematical Basis for Data Compression.* Prentice Hall, Englewood Cliffs, NJ, 1971.

[2] R. E. Blahut, "Computation of Channel Capacity and Rate-Distortion Functions", *IEEE Transactions on Information Theory*, vol. IT-18, pp. 460–473, 1972.

[3] I. Csiszár, "On an extremum problem of information theory", *Studia Scientiarum Mathematicarum Hungarica*, vol.9, pp. 57–71, 1974.

[4] R. M. Gray, L. D. Davisson, “Source coding without the ergodic assumption”, *IEEE Transactions on Information Theory*, vol. IT-20, pp. 502–516, 1974.

[5] M. B. Pursely, R M. Gray, “Source coding theorems for stationary, continuous-time stochastic processes”, Annals of Probability, vol. 5, pp. 966–986, 1977.

[6] I. Csiszár, G. Tusnady, “Information geometry and alternating minimization procedures”, Statistics & Decisions, Supplement Issue No.1, pp. 205–237, 1984.

[7] J. A. Bucklew, “A Large Deviation Theory Proof of the Abstract Alphabet Source Coding Theorem”, *IEEE Transactions on Information Theory*, vol. IT-34, pp. 1081–1083, 1988.

[8] R. A. DeVore, B. Jawerth, and B. J. Lucier, “Image Compression Through Wavelet Transform Coding”, *IEEE Transactions on Information Theory*, vol. IT-38, pp. 719–746, 1992.

[9] T. Berger, “Lossy Source Coding”, *IEEE Transactions on Information Theory*, vol. IT-44, pp. 2693–2723, 1998.

[10] E. Yang and Z. Zhang, “On the Redundancy of Lossy Source Coding with Abstract Alphabets”, *IEEE Transactions on Information Theory*, vol. IT-45, pp. 1092–1110, 1999.

[11] A. Dembo and I. Kontoyiannis, “Source Coding, Large Deviations, and Approximate Pattern Matching”, *IEEE Transactions on Information Theory*, vol. IT-48, pp. 1590–1615, 2002.

[12] M. Fozunbal, S. W. McLaughlin and R. W. Schafer,“Capacity Analysis for Continuous-Alphabet Channels With Side Information, Part I: A General Framework”, *IEEE Transactions on Information Theory*, vol. IT-51, pp. 3075–3085, 2005.

[13] R. E. Blahut, *Principles and Practice of Information Theory.* Addison-Wesley, 1987.

[14] R. M. Gray, *Entropy and Information Theory.* Springer-Verlag, 1990.

[15] T. Cover and J. Thomas, *Elements of Information Theory.* John Wiley & Sons, 1991.

[16] N. Dunford and J. T. Schwartz, *Linear Operators, Part I: General Theory.* Interscience Publishers, Inc., New York, 1958.

[17] D. G. Luenberger, *Optimization by Vector Space Methods.* John Wiley & Sons, 1969.

[18] E. Zeidler, *Nonlinear Functional Analysis and its Applications- Part I: Fixed-Point Theorems.* Springer-Verlag, 1986

[19] J. L. Kelley, *General Topology.* Springer-Verlag, 1975.

[20] A.Ionescu Tulcea & C.Ionescu Tulcea, *Topics in the Theory of Lifting*, Springer Verlag,Berlin, Heidelberg, New York, 1969.

[21] P. Dupuis and R.S. Ellis, *A Weak Convergence Approach to the theory of Large Deviations.* John Wiley & Sons, 1997.

[22] W. Rudin, *Functional analysis.* McGraw-Hill, 1991.

[23] H. L. Royden, *Real Analysis.* Mcmillan, 1988.

[24] H.B.Maynard, *A Radon-Nikodym Theorem for Finitely Additive Bounded Measures,* Pacific Journal of Mathematics, 83(2),1979,pp. 401-413.