Optimal Randomization in Quantizer Design with Marginal Constraint

Naci Saldi

Queen's University

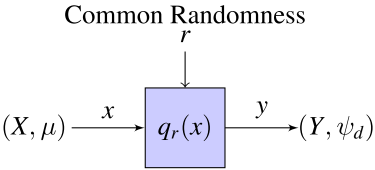
October 2012

Outline

- Informal definition of the problem.
- Representation of the quantizers as probability measures.
- Definition of the randomization scheme.
- Parametrization of the quantizer set.
- Existence of the minimizer for the fixed output marginal constraint case.
- Definition of the problem with relaxed output marginal constraint.
- Optimality of the set of finite randomizations for the relaxed problem.

Motivation

- In this work, we consider the optimal randomized quantization problem with a constraint on the marginal distribution of the output, i.e.



  where X and Y are Polish spaces (complete, separable metric space) and $q_r(x)$ is M-point quantizer.

- Recall that M-point quantizer $q(\cdot)$ is a measurable function from X to Y whose range cardinality is at most M.
- r is the common randomness between the encoder and the decoder.
- First, we have to define the randomization appropriately.

Definitions and Notation

- Let X denote quantizer's input space and Y denote its output space.
- Let $P(X \times Y)$ denote the set of probability measures on the product space $X \times Y$.
- Let μ and ψ_d be fixed probability measures on X and Y respectively.
- Yuksel and Linder in [1] and Borkar in [2] characterize the quantizers as a stochastic kernels between X and Y as follows:

$$Q(dy|x) = \delta_{q(x)}(dy)$$

where $\delta_{q(x)}(\cdot)$ is Dirac measure at $q(x)$.

- With this point of view, we can define the following subset of $\mathcal{P}(X \times Y)$ which is called quantizer set:

$$\Gamma_Q(M) = \{\upsilon \in \mathcal{P}(X \times Y) : \upsilon(dx, dy) = \mu(dx)Q(dy \mid x)$$
$$\text{where } Q(dy \mid x) = 1_{\{q(x) \in dy\}} \text{ s.t. } q(x) \text{ is a } M\text{-point quantizer } \}$$

- Randomly picking a quantizer equivalent to putting a probability measure on $\Gamma_Q(M)$ and each probability measure on $\Gamma_Q(M)$ corresponds to different randomization scheme.
- We have to prove the measurability of $\Gamma_Q(M)$ in $\mathcal{P}(X \times Y)$ in terms of some σ-algebra.

- We will work with the weak topology on $\mathcal{P}(X \times Y)$ and the Borel σ-algebra generated by this topology.

Definition (Weak Convergence and Topology)

A sequence of probability measures $\{\upsilon_n\}$ in $\mathcal{P}(X \times Y)$converges weakly to υ in $\mathcal{P}(X \times Y)$ if

$$\lim_{n\to\infty} \int h\, \upsilon_n = \int h\, \upsilon \text{ for every } h \text{ in } \mathcal{C}_b(X \times Y).$$

Correspondingly, the weak topology on $\mathcal{P}(X \times Y)$ is defined as the weakest topology on $\mathcal{P}(X \times Y)$ for which all functionals $\upsilon \mapsto \int h\, \upsilon, h \in \mathcal{C}_b(X \times Y)$ are continuous.

Measurability of Quantizer Set

- The following proposition can be found in Borkar et al. [3] or in Borkar [2], as an application of Choquet theorem [4].

Proposition (1)

Let X be a Polish space and let Y be a compact Polish space. Define the following subset of $\mathcal{P}(X \times Y)$:

$$\Gamma_\mu = \{ \upsilon \in \mathcal{P}(X \times Y) : \upsilon(A \times Y) = \mu(A) \text{ for all } A \in B(X) \}$$

where μ is a fix probability measure on X and let Γ_E denote extreme points of Γ_μ. Then Γ_μ is convex and compact in the weak topology. Furthermore, Γ_E is a Borel set in the weak topology.

Lemma (1)

Let X be a Polish space and Y be a compact Polish space. Then $\Gamma_Q(M)$ is a Borel set in the weak topology.

- From Proposition 1 and Lemma 1, we have the following theorem.

Theorem (1)

Let X be a Polish space and let Y be a σ-compact Polish space. Then $\Gamma_Q(M)$ is Borel subset of $\mathcal{P}(X \times Y)$ in the weak topology.

- This theorem enables us to endow $\Gamma_Q(M)$ with a probability measure. Hence, we can define the randomized quantizer set as follows:

$$\Gamma_R(M) = \{\upsilon \in \mathcal{P}(X \times Y) : \upsilon(dx, dy) = \int_{\Gamma_Q(M)} \bar{\upsilon}(dx, dy) P(d\bar{\upsilon}) \text{ where } P \in \mathcal{P}(\Gamma_Q(M))\}$$

Parametrization with Unit Interval

- We parameterize $\Gamma_Q(M)$ with unit interval.
- A well known isomorphism theorem states that all uncountable Borel spaces are isomorphic to each other.
- Since both $\Gamma_Q(M)$ and unit interval are uncountable Borel spaces, $\exists$ function g between unit interval and $\Gamma_Q(M)$ s.t. g is 1-1, measurable with measurable inverse.
- Let us write g as $g(r) = \upsilon^r(dx, dy)$. Then, we can write the elements in $\Gamma_R(M)$ as follows:

$$\upsilon(dx, dy) = \int_{\Gamma_Q(M)} \bar{\upsilon}(dx, dy) P(d\bar{\upsilon}) = \int_{[0,1]} \upsilon^r(dx, dy) \widetilde{P}(dr)$$

  where $\widetilde{P}(A) = P(\{\bar{\upsilon} : g^{-1}(\bar{\upsilon}) \in A\})$.
- Based on this isomorphism, the following fact can be proved:
  - $q(r, x) := q_r(x)$ ($\upsilon^r(dx, dy) = \mu(dx)\delta_{q_r(x)}(dy)$) is a measurable function such that $q(r, \cdot)$ is a M-point quantizer for all r.

Definition of the Problem

- Recall that $\Gamma_R(M)$ is defined as follows:

$$\Gamma_R(M) = \{\upsilon \in \mathcal{P}(X \times Y) : \upsilon(dx, dy) = \int_{\Gamma_Q(M)} \bar{\upsilon}(dx, dy) P(d\bar{\upsilon}) \text{ where } P \in \mathcal{P}(\Gamma_Q(M))\}$$

or equivalently

$$= \{\upsilon \in \mathcal{P}(X \times Y) : \upsilon(dx, dy) = \int_{[0,1]} \upsilon^r(dx, dy) P(dr) \ , \upsilon^r(dx, dy) = g(r), P \in \mathcal{P}([0, 1])\}.$$

- Define the following subset of $\mathcal{P}(X \times Y)$:

$$\Gamma_{\mu\psi_d} = \{\upsilon \in \mathcal{P}(X \times Y) : \upsilon(dx, Y) = \mu(dx), \upsilon(X, dy) = \psi_d(dy)\}.$$

where ψ_d is a fixed probability measure on Y.

- Define the following subset of $\Gamma_R(M)$:

$$\begin{aligned} \Gamma_R^{\psi_d}(M) &= \{\upsilon \in \Gamma_R(M) : \upsilon(X, dy) = \psi_d(dy)\} \\ &= \Gamma_R(M) \cap \Gamma_{\mu\psi_d}. \end{aligned}$$

Definition of the Problem

- We will optimize over $\Gamma_R^{\psi_d}(M)$.
- We can define average distortion function as a functional on $\mathcal{P}(X \times Y)$:

$$L(\upsilon) = \int_{X \times Y} c(x, y)\upsilon(dx, dy).$$

where $c(x, y)$ is a continuous and non-negative function on $X \times Y$.

- Optimal randomized quantization with marginal constraint problem can be written in the following form:

$$(P_1) \quad \inf_{\upsilon \in \Gamma_R^{\psi_d}(M)} L(\upsilon).$$

Existence of the Minimizer

Lemma (2)

$L(\upsilon(dx,dy)) = \int_{X\times Y} c(x,y)\upsilon(dx,dy)$ is lower semi-continuous on $\mathcal{P}(X\times Y)$ under weak convergence, i.e.

$$\liminf_{n\to\infty} \int_{X\times Y} c(x,y)\upsilon_n(dx,dy) \geq \int_{X\times Y} c(x,y)\upsilon(dx,dy)$$

as $\upsilon_n \to \upsilon$ weakly.

- If we can prove the compactness of $\Gamma_R^{\psi_d}(M)$, then we are done.
- Instead, we show the compactness of some subset of $\Gamma_R^{\psi_d}(M)$ which is an optimal class for this problem.

- First, we show that randomization can be restricted to a certain subset of $\Gamma_Q(M)$.
- Then, we prove the compactness of the optimal class which is the randomization of this subset.
- To construct such a subset we use some results from optimal transport theory.

Definition

Probability measure P on X is said to be c-continuous if it satisfies

$$P(\{x : c(x,a) - c(x,b) = k\}) = 0$$

for all $a, b \in Y$, $a \neq b$, and for all $k \in \Re$.

- We have the following assumptions to prove the existence of the minimizer:
  (a) μ is c-continuous.
  (b) Y is compact.

- Observe that each quantizer induces a probability measure on Y whose support cardinality is at most M.
- Let $\mathcal{P}_M(Y)$ denote the set of probability measures on Y which are induced by M-point quantizers.
- We are achieving a given distribution on Y by randomization of $\Gamma_Q(M)$ which is essentially equivalent to randomization of $\mathcal{P}_M(Y)$.
- We can construct an equivalence class among probability measures in $\Gamma_Q(M)$ based on their second marginals, i.e.

$$\upsilon_1(dx,dy) \sim \upsilon_2(dx,dy) \text{ if } \upsilon_1(X,dy) = \upsilon_2(X,dy).$$

- If we can find optimal elements in each equivalence class, then these elements form an optimal set for the randomization.

- Let $\psi \in \mathcal{P}_M(Y)$, then finding optimal elements in each equivalence class is essentially equivalent to the optimal mass transfer problem with marginals μ and ψ.
- The following fact is due to the optimal mass transport theory: If the probability measure μ on X is c-continuous, then there exists a unique optimal element in each equivalence class [5, Cuesta-Albertos et al.].
- Let $\Gamma_{opt}(M)$ be the collection of these optimal elements.
- $\Gamma_{opt}(M)$ is the optimal subset of $\Gamma_Q(M)$ for the randomization.
- In the rest of this section, the set, on which the randomization is applied, is $\Gamma_{opt}(M)$ instead of $\Gamma_Q(M)$.

- If Y is compact, then we can conclude the compactness of $\Gamma_{opt}(M)$ under the following assumption:

$$\text{(c)} \quad \int_{X \times Y} c(x,y)\upsilon(dx,dy) < \infty \text{ for all } \upsilon \in \Gamma_{opt}(M)[6, \mathit{Villani}].$$

- Let $\Gamma_{Ropt}(M)$ denote the randomization of $\Gamma_{opt}(M)$.
- Hence, the original problem $(P1)$ reduces to the following one:

$$\text{(P2)} \quad \inf_{\upsilon \in \Gamma_{Ropt}^{\psi_d}(M)} \int c(x,y)\upsilon(dx,dy)$$

- To show the existence of the minimizer, it is enough to prove compactness of the set $\Gamma_{Ropt}^{\psi_d}(M)$ which is equivalent to proving the compactness of $\Gamma_{Ropt}(M)$ since $\Gamma_{\mu\psi_d}$ is already compact.

- Let us define the following mapping between $\mathcal{P}(\Gamma_{opt}(M))$ and $\Gamma_{Ropt}(M)$:

$$s(P) = \int_{\Gamma_{opt}(M)} \upsilon(dx, dy) P(d\upsilon).$$

- $s(\cdot)$ is continuous.
- Since the set of probability measures on compact sets is compact in the weak topology, $\mathcal{P}(\Gamma_{opt}(M))$ is also compact.
- Hence, the compactness of $\Gamma_{Ropt}(M)$ implies the compactness of the set $\Gamma_{Ropt}^{\psi_d}(M)$.

Theorem (2)

There exists a minimizer for the following problem:

$$\inf_{\upsilon \in \Gamma_R^{\psi_d}} \int c(x, y) \upsilon(dx, dy)$$

if the assumptions (a), (b) and (c) are satisfied.

Approximation with Finite Randomization

- Since the randomization should be common both to the decoder and the encoder, infinite randomization may not be practical and realistic.
- However, if the desired probability measure ψ_d on Y is continuous, then we must apply infinite randomization to achieve this.
- Hence, we should relax the fixed output marginal constraint in order to get more realistic optimal randomization schemes (i.e. finite randomization).

- Now, we will consider the following relaxed minimization problem:

$$\text{(P3)} \quad \inf_{\upsilon \in M^{\delta}_{\psi_d}} \int_{X \times Y} c(x,y)\upsilon(dx,dy)$$

where $M^{\delta}_{\psi_d} = \{\upsilon \in \Gamma_R(M) : \upsilon(X,dy) \in B(\psi_d,\delta)\}$ and $B(\psi_d,\delta)$ is a ball in $\mathcal{P}(Y)$ with center ψ_d and radius δ in terms of Prokhorov metric which metrizes the weak topology.

- The goal is to show that the set of finitely randomized quantizers is an optimal class for this problem.
- Let $\Gamma_{FR}(M)$ denote the finitely randomized quantizer set.
- Clearly $\Gamma_{FR}(M) \subset \Gamma_R(M)$.
- Hence, we want to show:

$$\inf_{\upsilon \in M^{\delta}_{\psi_d}} \int_{X \times Y} c(x,y)\upsilon(dx,dy) = \inf_{\upsilon \in \Gamma_{FR}(M) \cap M^{\delta}_{\psi_d}} \int_{X \times Y} c(x,y)\upsilon(dx,dy)$$

Lemma (3)

M^{ε}_{ψ} is a open set for any ε and ψ in $\Gamma_R(M)$ in relative topology of weak convergence where $M^{\varepsilon}_{\psi} = \{\upsilon \in \Gamma_R(M) : \upsilon(X, dy) \in B(\psi, \delta)\}$.

- Hence, $M^{\delta}_{\psi_d}$ is an open set in $\Gamma_R(M)$.

- We want to replace any infinite randomization υ in $M^{\delta}_{\psi_d}$ with υ_F in $\Gamma_{FR}(M)$ which is living in some neighborhood $M^{\varepsilon}_{\psi_0} \subset M^{\delta}_{\psi_d}$ of υ and has less distortion than υ.

Lemma (4)

$\Gamma_{FR}(M)$ is dense in $\Gamma_R(M)$, i.e. for any υ in $\Gamma_R(M)$ and for any $\varepsilon > 0$ we can find $\hat{\upsilon}$ in $\Gamma_{FR}(M)$ such that $\hat{\upsilon} \in B(\upsilon, \varepsilon)$.

- Let us define the following subset of $\Gamma_R(M)$:

$$G = \{\upsilon \in \Gamma_R(M) : L(\upsilon) < L(\upsilon_0)\}.$$

- If $L(\cdot)$ is a continuous functional, then G is an open set.

- $L(\cdot)$ is continuous for compact X and Y and is continuous for general X and Y under the following the assumption:

$$\text{(a)} \lim_{A\to\infty} \sup_{\upsilon \in \Gamma_Q(M)} \int c(x,y) 1_{\{c(x,y)\geq A\}} \upsilon(dx,dy) = 0.$$

Lemma (5)

$M^{\varepsilon}_{\psi_0} \cap G$ is a non-empty open set in $\Gamma_R(M)$.

Theorem (3)

Under the assumption (a) or the assumption that X and Y are compact, finite randomization is an optimal class for the problem $(P3)$, i.e.

$$\inf_{\upsilon \in M^{\delta}_{\psi_d}} \int_{X\times Y} c(x,y)\upsilon(dx,dy) = \inf_{\upsilon \in \Gamma_{FR}(M)\cap M^{\delta}_{\psi_d}} \int_{X\times Y} c(x,y)\upsilon(dx,dy)$$

Conclusion

- In this work, we consider optimal randomized quantization with a constraint on the output marginal distribution.
- First, the quantizer set is represented as a set of probability measure on the product space.
- Then, appropriate randomization scheme is defined on this set.
- The existence of the minimizer is proved for the fixed output marginal constrained case under the assumption of compact Y and c-continuous μ on X.
- The problem with relaxed output marginal constraint is investigated.
- It is proved that the set of finite randomizations is an optimal class for the relaxed problem.

- S. Yuksel and T. Linder, "Optimization and convergence of observation channels in stochastic control," *SIAM J. on Control and Optimization*, vol. 50, pp. 864–887, 2012.
- V. Borkar, "White-noise representation in stochastic realization theory," *SIAM Journal of Control and Optimization*, vol. 31, pp. 1093–1102, 1993.
- A. S. V. Borkar, S. Mitter and S. Tatikonda, "Sequential source coding: An optimization viewpoint," *44th IEEE Conference on Decision and Control*, 2005.
- G. Choquet, *Lectures on analysis*.
  W.A. Benjamin, Inc., 1969.
- A. T.-D. J.A. Cuesta-Albertos, "A characterization for the solution of the monge-kantorovich mass transference problem," *Statistics and Probability Letters*, vol. 1, pp. 147–152, 1993.
- C. Villani, *Optimal Transport: Old and New*.
  Springer, 2009.