Proceedings of the 45th IEEE Conference on Decision & Control
Manchester Grand Hyatt Hotel
San Diego, CA, USA, December 13-15, 2006

Combinatorial Optimization Approach to Coarse Control Quantization

Srinivasa Salapaka
Mechanical Sciences and Engineering, UI Urbana-Champaign, IL 61801

Abstract— **In this paper we consider the problem of stabilization of a single input linear discrete time invariant system where the control can take values from a *countable set*. The problem definition includes the design of a quantization scheme that will produce the coarsest such set while guaranteeing stability. This problem is cast as a *static* combinatorial resource allocation or *coverage* problem where in the control values are viewed as resources which have to cover the state space. A relaxed notion of stability is used to represent the *coverage* cost function. The resulting quantizer on applying Deterministic Annealing algorithm is logarithmic and explicitly gives the partition of the state space.**

I. Introduction

Recently there has been a growing interest in the seemingly unrelated areas of coarse quantization [1], [2], coverage control, mobile sensing networks, and motion coordination algorithms [3], [4]. These problems each with different and unrelated goals, in fact have some fundamental common attributes. All these areas, either directly or not, bring together the concepts from information theory and control theory. For example, [1] studies the aspect of control with minimum information while [3] studies the coverage control problem for mobile sensing networks with a dynamically changing environment with a distributed communication and computation architecture. Another striking similarity in these areas is that after disregarding the details, they aim to solve similar optimization problems - in fact, they try to obtain; (1) an optimal partition of the underlying domain, and (2) an optimal assignment of values from a finite or a countable set to each cell of the partition. The differences in these problems come from having different conditions of optimality and constraints. The coarse quantization problem consists of obtaining a partition of the state space and allocating a control value to each cell in such a way to obtain the *coarsest* such partition while maintaining the stability of the underlying system. Similarly the coverage control problem consists of obtaining an optimal placement and tuning of sensors via *optimal* partitioning of the underlying space. It should however be emphasized that the quantization problem in [1], [2] is a *stabilization* problem pertaining to a *dynamic* system and the quantizer is obtained using tools from systems theory where it is obtained as a solution to a special linear quadratic regulator (LQR) problem.

Even though these formulations are relatively recent in the control theory, these optimal partitioning-assignment (also called combinatorial resource allocation) problems come up in various forms and are studied in different areas such as minimum distortion problem in data compression [5], facility location problems [6], optimal quadrature rules and discretization of partial differential equations [7], pattern recognition [8], neural networks [9], and clustering analysis [10]. These problems are non convex and computationally complex. It has been well documented (e.g. [11]) that most of them suffer from poor local minima that riddle the cost surface. A variety of heuristic approaches have been proposed to tackle this difficulty, and they range from repeated optimization with different initialization, and heuristics to good initialization, to heuristic rules for cluster splits and merges. In this context, the Deterministic Annealing (DA) algorithm developed in the data compression literature [12] offers two important features: (1) ability to avoid many poor local optima and (2) has a relatively faster convergence rate. It formulates an effective energy function parameterized by a (pseudo) temperature variable and this function is deterministically optimized at successively reduced temperatures.

In this paper, we view the coarse quantization problem introduced in [1], [2] as a *static* resource allocation problem. The problem seeks a quantization scheme, i.e., a partition of the state space and the allocation of control values to each cell in the partition so that a cost function that reflects a relaxed notion of quadratic stability is minimized. We use the DA algorithm to solve this quantization problem. In this approach, a probability distribution function is defined on the space of quantizers that has the maximum entropy for a given average (expected) value of the cost function. The control values are then calculated by finding the *most probable* values under this distribution. This algorithm results in quantizers that have same qualitative features as the coarsest quantizers from the dynamic theory approach presented in [1], [2]. Since the coarseness of partition is not explicitly included in the cost function, this viewpoint aptly captures the relationship between the coarseness (and therefore 'informational content') of a quantizer to the entropy of probability measures on the space of qunatizers. It suggests that finding coarsest quantizer under the constraint of stability is equivalent to finding the probability measure with the maximum entropy under the same stability constraint.

The rest of the paper is organized as follows. Section II gives the problem formulation where we first describe the quantization problem considered in [1], [2]. Then we present the resource allocation problem and describe the DA algorithm. In Section III, we present the quantization problem in the resource allocation framework and solve for the quantizer. Section IV discusses some properties of the quantizer and analyzes the results. We conclude with some observations and state on-going and future research directions in Section V.

II. Problem Formulation

We consider the problem of stabilizing a linear time invariant discrete-time system with a possibly countable number of fixed control values to be determined. This problem was first formulated and addressed in [1]. Their approach based on systems theory resulted in some fundamental insights about control quantization schemes. In this paper, we study a variant of the same problem where our approach is based on maximum entropy principle and we view the problem as a resource allocation task. This new viewpoint gives us new insights into the problem while at the same time reproduces the results qualitatively from the earlier approach. We first present here the problem formulation in [1].

A. The Quantization Problem:

The system is assumed unstable, stabilizable and governed by the following equation:

$$x(k+1) = Ax(k) + Bu(x_k), \tag{1}$$

where $x(k) \in \mathbb{R}^n \stackrel{\triangle}{=} \Omega$, k denotes the discrete time, $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times 1}$ and $u(x_k)$ belongs to a countable set $\mathcal{U} = \{u_i \in \mathbb{R}, i \in \mathbb{Z}\}$. Note that in this difference equation, the set $\mathcal{U}$ of control values is a countable set and not the entire set of real numbers. Had $\mathcal{U}$ included the whole of $\mathbb{R}$, and since the system is stabilizable and linear, it would have been possible to find a static linear function $u(x) = k^T x, k \in \mathbb{R}^n$ such that any quadratic function of the form $x^T Px$, where P is positive definite, becomes a (control) Lyapunov function for the closed loop system. The goal is to design a countable set $\mathcal{U}$ and find a function $f : \Omega \to \mathcal{U}$, that will result in a *stable* closed loop system with the prescription for $u(x_k)$. More specifically, we want to solve the following problem:

For a given quadratic Lyapunov function $V(x) = x^T Px, P > 0$, find a set $\mathcal{U} = \{u_i \in \mathbb{R} : i \in \mathbb{Z}\}$ and a *quantizer* $f : \Omega \to \mathcal{U}$ such that $\Delta V(x) = V(Ax+Bf(x)) - V(x) < 0$.

Note that to each quantizer one can associate a partition of the state space Ω into countably many cells such that on each cell the quantizer takes a unique value. Also in the above formulation, the only restriction on the control set $\mathcal{U}$ is that it is countable - and therefore there exist many (non unique) quantizers that stabilize the system (1). For example, once we have a stabilizing quantizer, we can easily construct a series of other stabilizing quantizers that provide *finer* partitions of the state space. This motivates defining a measure of *coarseness* of partitions and to find a quantizer that gives the coarsest partition of the state space. The coarseness of a stabilizing quantizer g (for a given Lyapunov function $V(x)$) is defined in terms of the *quantization density* η_g given by

$$\eta_g = \limsup_{\epsilon\to 0} \frac{\#g(\epsilon)}{\log \frac{1}{\epsilon}},$$

where $\#g(\epsilon)$ denotes the number of elements in $\mathcal{U}$ that are in the interval $[\epsilon, 1/\epsilon]$. A quantizer f is the *coarsest* for $V(x)$ if it has the smallest quantization density in the set of all stabilizing quantizers. One of the main contributions of [1] is the characterization of the coarsest quantizer where they show that the quadratic Lyapunov functions induce a countable logarithmic quantization of control values and of the system state-space.

B. Resource Allocation Problem and Deterministic Annealing (DA) Algorithm

The determination of a quantizer in the above formulation is equivalent to finding the associated partition of the state space and allocation of control values to each cell in the partition. More specifically, the quantization problem described above can be viewed as a problem of finding a partition $\{R_i\}$ of the state space Ω and assigning to each cell R_i a control value u_i such that $V(Ax + Bu_i) < V(x)$ for all x in R_i. Note that, in this viewpoint, finding the coarsest quantizer is equivalent to a *static* optimization problem (as it is independent of the time variable k in the difference equation (1)). This optimization problem can be viewed as a combinatorial resource allocation problem. The remaining part of this section is organized as follows: We will first introduce and discuss a combinatorial resource allocation problem. This problem is known to be a computationally hard problem and there is a suit of algorithms that address it. We then discuss the Deterministic Annealing algorithm based on Maximum Entropy Principle that solves a relaxed formulation of the resource allocation problem. Then we will reinterpret the quantization problem in terms of the relaxed formulation and solve for the quantizer.

In its prototypical form, the problem of selecting resource locations for the purpose of coverage of a set of sites can be described as:

Given a domain Ω, find the set of M resource locations r_j that solves the following minimization problem

$$\min_{r_j,\ 1\le j\le M} \int_\Omega \left\{ \min_{1\le j\le M} \rho(x, r_j) dx \right\} \tag{2}$$

Here $\rho(x, r_j)$ represents an appropriate *distance* metric between the resource r_j from the site x. In simpler terms, M resource locations are sought such that the average distance of sites x to their nearest resource locations is minimized. This formulation in the context of facility location problems, for example, can pertain to finding facility locations (r_j) so that the average distance of a person (x) in an area to the nearest facility is minimized.

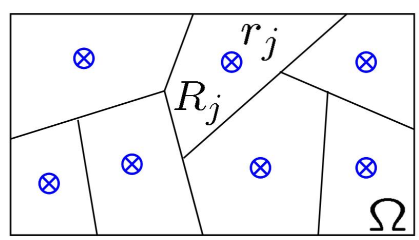

Fig. 1. A schematic of the static problem in which the set of sites is partitioned into cells R_j and to each R_j a resource location r_j ascribed so that the coverage cost function is minimized

Alternatively, this problem can also be formulated as finding a partition of the domain Ω into M cells R_j (see Figure 1) and assign to each cell R_j a resource location r_j such that the following cost function is minimized

$$\sum_{j=1}^{M} \int_{R_j} \rho(x, r_j) dx. \tag{3}$$

The DA algorithm [13], [14] is suited for this purpose since it is specifically designed to avoid local minima. This algorithm views the resource allocation problem in a probabilistic setting by ascribing a probability distribution (by using the maximum entropy principle) on the space of partitions and resource locations. It then finds the optimal resource allocation by finding the *most probable* resource locations. In this way it relaxes the resource allocation problem by finding weighted averages over the space of partitions.

More specifically, for every partition $R = \{R_j\}$ of Ω and a set $r = \{r_j\}$ of resource locations, the cost in (3) is rewritten as

$$E(R, r) = \int_\Omega \sum_j \chi_{xj} \rho(x, r_j) dx, \tag{4}$$

where χ_{xj} represents an indicator function where $\chi_{xj} = 1$ if $x \in R_j$ and 0 otherwise. If we assume a probability distribution $P(R, r)$ on the space

$$\mathcal{I} = \{(R, r) \text{ s.t. } R \text{ is a partition of } \Omega \text{ and } r \in \mathbb{R}^M\},$$

then the average cost function over this space of partitions and resource locations is given by

$$\sum P(R, r) E(R, r) \tag{5}$$

where the summation (equivalently an integral with an appropriate measure [13]) is over all the possible elements in $\mathcal{I}$. Since we do not make any assumption about the distribution on $\mathcal{I}$, we apply the Maximum Entropy Principle; i.e., of all the probability distributions that yield a given average cost, we choose one that maximizes entropy. As is well known the probability distribution that maximizes entropy under the constraint (5) is given by [15]

$$P(R, r) = \frac{e^{-\beta E(R,r)}}{\sum_{R', r'} e^{-\beta E(R', r')}},$$

where β is a constant that is determined by the value of the average cost function. The most probable resource locations r are found by maximizing the marginal distribution $P(r)$ given by

$$P(r) = \sum_R P(R, r). \tag{6}$$

To compute $P(r)$ we derive from (4) that

$$\sum_R e^{-\beta E(R,r)} = \prod_x \sum_k e^{-\beta \rho(x, r_k)} \triangleq Z,$$

where Z is called the partition function (borrowing the notation from statistical physics). If we define the *Free Energy* F by

$$F \triangleq -\frac{1}{\beta} \log Z = -\frac{1}{\beta} \int_\Omega \log \sum_k e^{-\beta \rho(x, r_j)} dx,$$

then the marginal probability in (6) becomes

$$P(r) = \frac{Z}{\sum_{r'} Z} = \frac{e^{-\beta F}}{\sum_{r'} e^{-\beta F}}. \tag{7}$$

It is evident from (7) that the most probable resource location vector r is one that minimizes the Free Energy

$$F = -\frac{1}{\beta} \int_\Omega \log \sum_k e^{-\beta \rho(x, r_j)} dx. \tag{8}$$

In this way the DA algorithm relaxes the minimization problem of (2) to that of minimization of the Free Energy function F given by (8).

III. Quantization Problem as a Resource Allocation Problem

As described earlier, the quantization problem can be viewed as a resource allocation problem where a partition of the state space Ω is sought and placement of the control values $\{u_j\}$ in each cell. In order to capture the notion of stability in the quantization problem, we define the following distance function

$$\rho(x, u_j) = \frac{\|Ax + Bu_j\|_P^2}{\|x\|_P^2} \tag{9}$$

for the DA algorithm where $V(x) = \|x\|_P^2 \triangleq x^T P x$ is the Lyapunov function in the definition of the quantization problem. The Free Energy in the DA algorithm for this problem with this distance function is given by

$$F_1 = -\frac{1}{\beta} \int_\Omega \log \sum_k e^{-\beta \frac{\|Ax + Bu_j\|_P^2}{\|x\|_P^2}} dx. \tag{10}$$

To ensure that control effort is restricted and also guarantee the coarseness of the partitions, we augment the cost function with the term F_2 given by

$$F_2 = \gamma \sum_i \int_{\Omega_a} \frac{\|Bu_i\|_P^2}{\|x\|_P^2} \text{`} 1 p(u_i|x) dx$$

where

$$p(u_j|x) = \frac{e^{-\beta \|Ax + Bu_j\|^2 / \|x\|_P^2}}{\sum_i e^{-\beta \|Ax + Bu_i\|^2 / \|x\|_P^2}}.$$

Thus the objective function is given by

$$F = F_1 + F_2. \tag{11}$$

For easier exposition, we view the same system in new coordinates where we have one of the axis is collinear with $k = (B^T P B)^{-1} B^T P A$. Consider the change of coordinates $x = \Gamma \mu$ where $\Gamma = [\gamma_1 \ \gamma_2 \ \cdots \ \gamma_n]$ is an orthonormal matrix ($\Gamma \Gamma^T = \Gamma^T \Gamma = I$) and $\gamma_1 = \frac{k}{\sqrt{k^T k}}$. In these coordinates, F_1 is

$$F_1 = -\frac{1}{\beta} \int_\Omega \log \sum_i e^{-\beta \|B\|_P^2 \|k\|^2 \left((\mu_1 - v_i)^2 + g(\mu) \right) / \|\mu\|_{\bar{P}}^2} d\mu,$$

where $v_i = -u_i / \|k\|$, $\bar{P} = \Gamma^T P \Gamma$ and $g(\mu) = \|A \Gamma \mu\|_P^2 - (\|B\|_P \|k\| \mu_1)^2$. Since $g(\mu)$ is not dependent on $\{v_i\}$, minimizing the Free energy is same as minimizing the function

$$\hat{F}_1 = -\frac{1}{\beta} \int_\Omega \log \sum_i e^{-\beta \|B\|_P^2 \|k\|^2 (\mu_1 - v_i)^2 / \|\mu\|_{\bar{P}}^2} d\mu. \tag{12}$$

We consider another change of coordinates $\eta_1 = \frac{1}{\mu_1},\ \eta_i = \frac{\mu_i}{\mu_1}$ for further simplifying the above expression. The determinant of the Jacobian matrix $\frac{\partial \mu}{\partial \eta} = \eta^{1-n}$ and the cost function becomes

$$\hat{F}_1 = -\frac{1}{\beta}\int_{\Omega} \log \sum_i e^{-\beta \|B\|_P^2 \|k\|^2 v_i^2 (\eta_1 - \frac{1}{v_i})^2/\|\tilde{\eta}\|_P^2} \eta_1^{1-n} d\eta, \quad (13)$$

where $\tilde{\eta} = [1\ \eta_2\ \eta_3\ \cdots\ \eta_n]^T$. If we define $\eta_b \triangleq [\eta_2\ \cdots\ \eta_n]$ and decompose accordingly the domain $\Omega = \Omega_a \times \Omega_b$ and compute the integral in the above equation with respect to η_b coordinates the (see appendix for a note on the domain Ω and this computation), $\hat{F}_1$ reduces to

$$\hat{F}_1 = -\frac{c_1}{\beta}\int_{\Omega_a} \log \sum_i e^{-\beta \|B\|_P^2 \|k\|^2 v_i^2 (\eta_1 - \frac{1}{v_i})^2} \eta_1^{1-n} d\eta_1, \quad (14)$$

for some constant $c_1 > 0$.

Similarly, by making same coordinate changes and manipulations the term F_2 in the objective function becomes

$$\hat{F}_2 = c_1\gamma \sum_i \int_{\Omega_a} e^{-\beta \|B\|_P^2\ \|k\|^2 v_i^2 p(v_i|\eta_1)} \eta^{1-n} d\eta_1.$$

The cost function F to be minimized is given by

$$F = \hat{F}_1 + \hat{F}_2.$$

Thus control values $\{v_j\}$ is found by minimizing F, i.e. by setting $\frac{\partial F}{\partial v_j} = 0$. This yields

$$\int_{\Omega_a} p(v_j|\eta_1)((1+\gamma)\eta_1 v_j - 1)\eta_1^{-n} d\eta_1 + I_2 = 0 \quad (15)$$

where $p(v_j/\eta_1)$ is the Gibbs distribution

$$p(v_j|\eta_1) = \frac{e^{-\beta \|B\|_P^2 \|k\|^2 v_j^2 (\eta_1 - \frac{1}{v_j})^2}}{\sum_i e^{-\beta \|B\|_P^2 \|k\|^2 v_i^2 (\eta_1 - \frac{1}{v_i})^2}},$$

and

$$I_2 = 2\beta\gamma c_2 \int_{\Omega_a} \left(\sum_i p(v_i \eta_1) p(v_j \eta_1) v_i^2 - p(v_j|\eta_1) v_j^2 \right) (\eta_1 v_j - 1)\eta_1^{-n} d\eta_1$$

for some $c_2 > 0$. In fact, it can be shown that $I_2 \to 0$ as $\beta \to \infty$ (see appendix). On solving Equation(15) as $\beta \to \infty$ gives the sequence $\{v_i\}$ given by

$$v_i = \rho^i v_0, \quad (16)$$

where $\rho = \frac{1+r}{1-r}$ where ν satisfies $\nu = tanh((1+\gamma)\nu)$ (for the case $n = 1$. The analysis is similar for $n > 1$ and the proof is not shown in this paper for $n > 1$ where ρ is given by a different formula. See appendix for details). The term v_0 can be chosen to be 1 without any loss of generality as emphasized in the discussion in the next section.

These results give full characterization of the quantizer. The quantizer is defined by assigning the control value $u_j = -\|k\| v_j$ to the cell R_j of the partition, i.e.

$$f : \Omega \to \mathcal{U}, f(x) = u_j.$$

This scheme finds a direction (given by $k = A^T PB/B^T PB$) and a partition of this axis that completely determines the partition on the entire state space Ω. A schematic of this quantization scheme is shown in Figure 2. Plot (a) shows the partition of Ω consists of parallel sections R_i that cut the μ_1 axis in a logarithmic way. The plot shown here depicts only the $n = 2$ case for easier exposition. In higher dimensions too, the partition of Ω is completely determined by Gibbs distributions defined on the μ_1 axis. The partition contains $n-1$ dimensional parallel planes that are perpendicular to the μ_1 axis and pass through the partition points $\nu_i = \nu^i u_0$ on the μ_1 axis. These results can be translated to the x-axis after noting that μ_1 axis is parallel to $k = A^T PB/B^T PB$. The corresponding partition of Ω in the x coordinates is shown in plot (b).

IV. Analysis of the quantizer and observations

The quantizer obtained from the approach based on the DA algorithm has many common features with the coarsest quantizer obtained in [1]. The foremost is of course the logarithmic partition of the state space. In both these methods, the direction given by $k = A^T PB/(B^T PB)$ plays an important role in determining the quantizer. The partition of Ω is completely determined by a logarithmic partition of this axis. Furthermore in both the methods, the control values (i.e. the set $\mathcal{U}$) is completely determined by this partition on the k axis. This logarithmic partition reinforces the intuitive argument that when the system is far from the origin, we do not need precise knowledge of the control and therefore use imprecise controls to steer the system in the right direction. Another property of this quantizer is that it is symmetric about the origin, i.e. $f(x) = -f(-x)$. This can be derived from the fact that if you change the coordinates from x to $-x$ and u_j to $-u_j$ in the integral in equation (11), the equation remains unchanged. This property was assumed in the for the coarse quantizer in [1]. Another property that results from the structure of Equation (11) is that if $\{\{R_i\},\mathcal{U}\}$ is a partition and control set pair that determines the quantizer $f : \Omega \to \mathcal{U}$, then the quantizer $p : \Omega \to c\mathcal{U}$ described by $p(x) = cu_i, x \in cR_i$ for any $c > 0$ achieves the same value for the Free energy function as the quantizer f. This can be seen that from minimization problem (11) which is invariant if we replace u_j by cu_j. This property implies that there is no loss of generality in considering quantizers that take value value 1, i.e. $u_i = 1$ for some i.

The main difference between the quantizers in these two approaches are that the notion of stability in our approach is captured through a distance function that is averaged over the whole state space. This makes the analysis somewhat easier as it leads to a static optimization problem which is not dependent on the stability properties of matrix A. Accordingly, the resulting partitions in the quantization schemes are simple parallel sections. This is in contrast to the more stringent analysis required with the regular notion of stability where it is important to specify the stable portions of the state space that take the quantization value 0.

One of the new features of this new approach gives is that it brings in the notion of entropy of probability measures defined on the space of quantizers (defined in terms partitions and quantization values). The deterministic annealing algorithm did solve an optimization problem for a probability measure that maximized the entropy for a given average

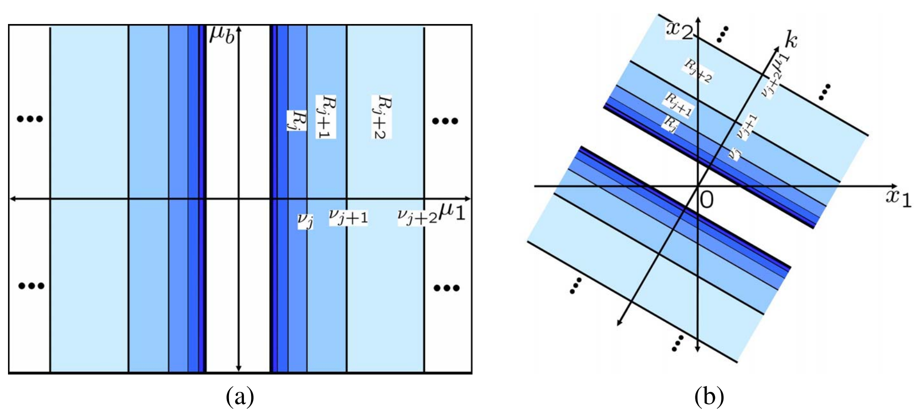

Fig. 2. A schematic of the quantizer where the partition is given by $\{R_i\}$ which are completely determined by the partition of the μ_1 (or k) axis. (a) the partition when Ω is in the μ coordinates, (b) Ω in the original (x) coordinates.

(expected) value of the cost function. This scheme yielded a quantizer that is qualitatively the same as the coarsest quantizer that guarantees stability (as obtained in [1]). This suggests that there is equivalence (in the sense described above) between the objective of finding the coarsest quantizer that guarantees stability and finding the maximum entropy probability measure (i.e. invoking the Maximum Entropy Principle) under a similar stability constraint.

There are many possible directions of research that arise from this approach. One immediate direction is to generalize these results to multi input case. The preliminary investigation does show that much of the analysis for the one input case can be generalized to higher dimensions. However, the underlying subspace (akin to the the μ_1 or k axis) that determines the partition of the entire state space is multidimensional and its part of ongoing work to come up with a reasonable parametrization of the partition on this subspace.

In this paper, we have just considered the control quantization scheme for stabilization that approximates state feedback stabilization. Similar quantization problems that approximate state estimation and output feedback schemes have been carried out in [1]. Generalization of our approach to these problems is a part of the ongoing work.

V. Conclusions

In this paper, we have developed a new approach for the design of quantizers for single input discrete time systems. This approach views the quantization problem as a static combinatorial resource allocation problem. This approach gives some new insights into the quantization schemes where it suggests relationship between the problem of finding the probability measure with maximum entropy on the space of quantizers and that of finding coarsest quantizers that guarantee stability. We have shown that the quadratic Lyapunov functions for these systems induce a logarithmic quantization of controls and the state space. These results match with results in [1].

Appendix

Theorem 1: The first mean value theorem for integration states

If $G : [a, b] \to \mathbb{R}$ is a continuous function, then there exists a number x in (a, b) such that

$$G(x)(b-a) = \int_a^b G(t)dt.$$

Proof: A direct consequence of the intermediate value theorem for continuous functions.

Simplification of Equation (13) to Equation (14)

Note that if we write the matrix $\bar{P} = \begin{pmatrix} p_{11} & p_{b1}^T \\ p_{b1} & P_{bb} \end{pmatrix}$, then $\|\tilde{\eta}\|^2_{\bar{P}} = [1\ \eta_b^T]^T \bar{P} [1\ \eta_b^T]^T$ can be rewritten as $\|\hat{\eta} + q\|^2 + s^2$ where $\hat{\eta} = P_{bb}^{\frac{1}{2}} \eta_b \in \mathbb{R}^{n-1}$, $q = P_{bb}^{-\frac{1}{2}} P_{b1}$ and $s^2 = p_{11} - \|q\|^2$. Now consider the change of coordinates from η_b to $(r, \theta) \stackrel{\triangle}{=} (r, \theta_1, \theta_2, \cdots, \theta_{n-2}) \in \mathbb{R} \times [0\ 2\pi]^{n-2}$ given by

$$\begin{array}{rcl} \|\hat{\eta}\| & = & r \\ \hat{\eta}_1 + q_1 & = & r\cos\theta_1 \\ \hat{\eta}_2 + q_2 & = & r\sin\theta_1\cos\theta_2 \\ & \vdots & \\ \hat{\eta}_{n-2} + q_{n-2} & = & r\sin\theta_1\sin\theta_2\cdots\sin\theta_{n-3}\cos\theta_{n-2} \\ \hat{\eta}_{n-1} + q_{n-1} & = & r\sin\theta_1\sin\theta_2\cdots\sin\theta_{n-3}\sin\theta_{n-2}. \end{array} .$$

The determinant of the jacobian $\frac{\partial \eta}{\partial(r,\theta)}$ is of the form $r^{n-1}h(\theta)det(P_{bb}^{-\frac{1}{2}})$. So $\hat{F}_1$ is given by the integral

$$\hat{F}_1 = -\frac{C}{\beta}\int_\Omega \log\sum_i e^{-\beta\|B\|_P^2\|k\|^2 v_i^2(\eta_1 - \frac{1}{v_i})^2/(r^2+s^2)} \eta_1^{1-n} dr d\eta_1,$$

where $C = det(P_{bb}^{-\frac{1}{2}})\int h(\theta)d\theta$. From the mean value theorem (see Theorem 1) on this integral after substituting r by $\tan\omega$, there exists an $\bar{\omega}$ where the integral is equal to

$$\hat{F}_1 = -\frac{C\pi}{2\beta}\int_\Omega \log\sum_i e^{-\beta\|B\|_P^2\|k\|^2 v_i^2(\eta_1 - \frac{1}{v_i})^2/(\bar{r}^2+s^2)} \eta_1^{1-n} d\eta_1,$$

where $\bar{r} = \tan\bar{\omega}$. This implies that there exists $c_1 = \frac{C_1\pi(\sec^2\bar{\omega})}{2(\bar{r}^2+s^2)}$ and $\alpha^2 = \frac{1}{\bar{r}^2+s^2} \in \mathbb{R}^+$ such that

$$\hat{F}_1 = -\frac{c_1}{2\beta}\int_\Omega \log\sum_i e^{-\beta\alpha^2\|B\|_P^2\|k\|^2 v_i^2(\eta_1 - \frac{1}{v_i})^2/(\bar{r}^2+s^2)} \eta_1^{1-n} d\eta_1.$$

We absorb the constant α^2 in β without loss of generality and slight abuse of notation in Equation (14) ∎

Lemma 1.1: Let $\{g_i(\beta,x)\},\ i\in\mathbb{Z}$ be a sequence of Gibbs distributions of the form

$$\frac{e^{-\beta(x-w_i)^2/w_i^2}}{\sum_l e^{-\beta(x-w_l)^2/w_l^2}},$$

defined on the domain $\mathcal{D}\overset{\triangle}{=}\{x \text{ such that } \frac{1}{\epsilon}>|x|>\epsilon\}$ where $\epsilon>0$ and $\{w_i\}$ is a monotonically increasing sequence in $\mathcal{D}$. Let the function $G_i(x)$ denote the Gaussian term $e^{-\beta(x-w_i)^2/w_i^2}$ for each i. Let a_{i-1} and a_{i+1} be the values where the Gaussian curves $G_{i-1}(x)$ and $G_{i+1}(x)$ intersect with $G_i(x)$ respectively; i.e. $G_{i-1}(a_{i-1})=G_i(a_{i-1})$ and $G_{i+1}(a_{i+1})=G_i(a_{i+1})$. Let $\chi_i(x)$ be a pulse function defined on $\mathcal{D}$ given by

$$\chi_i(x)=\begin{cases}1 & \text{if } \Omega_i\overset{\triangle}{=}a_{i-1}\le x\le a_{i+1}\\ 0 & \text{elsewhere}\end{cases},$$

then

1) $\lim_{\beta\to\infty}\int_{\mathcal{D}}\beta^s(g_i(\beta,x)-\chi_i(x))h(x)dx=0$,
2) $\lim_{\beta\to\infty}\int_{\mathcal{D}}\beta^s(\sum_i g_i(\beta,x)g_j(\beta,x)v_i^2 - g_j(\beta,x)v_j^2)h(x)dx=0$

for all $s>0$, and functions h which are bounded above by L_1 functions on $\mathcal{D}$.

Proof: Fix j in $\mathbb{N}$. Let $a_{i,j}=\frac{2}{\frac{1}{w_j}+\frac{1}{w_i}}$ denote the point where the Gaussian distributions $G_i(x)$ intersect with the distribution G_j. Then $\Omega_j=(a_{j-1,j}\ a_{j+1,j})$ denotes the interval where $\chi_j(x)$ takes the value 1. Note that from the definition of $g_j(x)$, we have for all x and β, $0\le g_j(\beta,x)\le 1$ and

$$1-g_j(\beta,x)=\frac{\Delta}{1+\Delta}<\Delta \tag{17}$$

where $\Delta=\sum_{i\neq j}e^{-\beta\left(\sigma_{i,j}^2x(x-a_{i,j})\right)}$ and $\sigma_{i,j}=\frac{1}{w_i^2}+\frac{1}{w_j^2}$. Similarly

$$g_j(\beta,x)<\frac{1}{1+\Delta_2}, \tag{18}$$

where $\Delta_2=\sum_{i=-1,1}e^{-\beta\sigma_{i+j,j}^2x(x-a_{i+j,j})}$. Since $\{w_i\}$ is an increasing sequence, we have $\sigma_{i,j}^2x(x-a_{i,j})>0$ for all $x\in\Omega_j$ and $i\neq j$, and $\sigma_{i+j,j}^2x(x-a_{i+j,j})<0$ for $i\in\{-1,1\}$. Therefore from Equations (17) and (18), we have

$$\lim_{\beta\to\infty}\beta^s(g_i(\beta,x)-\chi_i(x))h(x)=0\ a.e.$$

for all $s>0$ and bounded function h. Now let g be a function in $L_1(\mathcal{D})$ such that $|h(x)|<g(x)$ for all x. Then $\beta^s(g(\beta,x)-\chi(x))h(x)$ is bounded above by a L_1 function for all β. Therefore from Lebesgue Dominated Convergence Theorem [16], we have $\lim_{\beta\to\infty}\int_{\mathcal{D}}\beta^s(g_i(\beta,x)-\chi_i(x))h(x)dx=0$ for all $s>0$, and functions h which are bounded above by L_1 functions on $\mathcal{D}$. The second point in the lemma is a direct consequence of the above proof. ∎

Derivation of the quantizer parameter ρ.

From the above lemma, I_2 in Equation (15) is identically 0. This reduces the equation to

$$(1+\gamma)v_j=\left(\int_{\frac{1-\nu_2^j}{v_j}}^{\frac{1+\nu_1^j}{v_j}}\eta_1^{-n}d\eta_1\right)\Big/\left(\int_{\frac{1-\nu_2^j}{v_j}}^{\frac{1+\nu_1^j}{v_j}}\eta_1^{-n}d\eta_1\right),$$

where $\Omega_j=\frac{1-\nu_2^j}{v_j}\ \frac{1+\nu_1^j}{v_j}$. For the case when $n=1$ becomes independent of v_j and yields

$$(1+\nu_1^j)e^{-(1+\gamma)(1+\nu_1^j)}=(1-\nu_2^j)e^{-(1+\gamma)(1-\nu_2^j)}.$$

From this implicit equation, we show that ν_1^j and ν_2^j are independent of j and there exists a $\nu>0$ such that $(1+\nu)e^{-(1+\gamma)(1+\nu)}=(1-\nu)e^{-(1+\gamma)(1-\nu)}$. This is equivalent to the condition that $\nu=\tanh((1+\gamma)\nu)$. After noting that $\Omega_j=[\frac{2}{\frac{1}{v_j}+\frac{1}{v_{j-1}}}\ \frac{2}{\frac{1}{v_j}+\frac{1}{v_{j+1}}}]$ yields that

$$\frac{v_{j+1}}{v_j}=\frac{1+\nu}{1-\nu}\overset{\triangle}{=}\rho.$$

A note on the domain Ω

The domain Ω is $\mathbb{R}^n$. However to avoid technical complications in the various coordinate transformations in the analysis, we assume $\Omega=\Omega_a\times\Omega_b$ where $\Omega_a=(\epsilon,\frac{1}{\epsilon})$ and $\Omega_b=\mathbb{R}^{n-1}$. Since the quantizer is symmetric, all arguments hold true for $(-\frac{1}{\epsilon},\epsilon)\times\Omega_b$. The analysis holds true as $\epsilon\to 0$.

References

[1] N. Elia and S. Mitter. Stabilization of Linear Systems with Limited Information. *IEEE Transactions on Automatic Control*, 46(9):1384–1400, September 2001.

[2] S.K. Mitter. Control with limited information. Plenary lecture at International Symposium on Information Theory, Sorento, Italy, 2000.

[3] J. Cortés, S. Martínez, T. Karatas, and F. Bullo. Coverage control for mobile sensing networks. *IEEE Transactions on Robotics and Automation*, 20(2):243–255, 2004.

[4] S. Salapaka and A. Khalak. Constraints on Locational Optimization Problems. *Proceedings of the 42nd IEEE conference on decision and control, Maui*, pages 1741–1746, December 2003.

[5] A. Gersho and R. Gray. *Vector Quantization and Signal Compression*. Kluwer, Boston, Massachusetts, 1st edition, 1991.

[6] Z. Drezner. *Facility Location: A survey of Applications and Methods*. Springer Series in Operations Research. Springer Verlag, New York, 1995.

[7] Q. Du, V. Faber, and M. Gunzburger. Centroidal voronoi tessellations: Applications and algorithms. *SIAM Review*, 41(4):637–676, December 1999.

[8] C.W. Therrien. *Decision, Estimation and Classification: An Introduction to Pattern Recognition and related topics*, volume 14. Wiley, New York, 1st edition, 1989.

[9] S. Haykin. *Neural Networks: A Comprehensive Foundation*. Prentice Hall, Englewoods Cliffs, NJ, 1998.

[10] J.A. Harigan. *Clustering Algorithms*. Wiley, New York, 1975.

[11] R. Gray and E.D. Karnin. Multiple local minima in vector quantizers. *IEEE transactions on Information Theory*, IT-28:256–361, March 1982.

[12] K. Rose. *Deterministic annealing, clustering, and optimization*. PhD thesis, California Institute of Technology, Pasadena, 1991.

[13] K. Rose. Statistical mechanics and phase transitions in clustering. *Physics Review Letters*, 65(8):945–948, 1990.

[14] K. Rose. Deterministic Annealing for Clustering, Compression, Classification, Regression and Related Optimization Problems. *Proceedings of the IEEE*, 86(11):2210–39, November 1998.

[15] A. Papoulis. *Probability and Statistics*. Prentice Hall, Upper Saddle River, New Jersey, 1990.

[16] J.B. Bruckner A.M. Bruckner and B.S. Thomson. *Real Analysis*. Prentice Hall, New Jersey, 1 edition, 1997.