Chapter 6
Design of Information Channels for Optimization and Stabilization in Networked Control

Serdar Yüksel

6.1 Introduction and the Information Structure Design Problem

In stochastic control, typically a partial observation model/channel is given and one looks for a control policy for optimization or stabilization. Consider a single-agent dynamical system described by the discrete-time equations

$$x_{t+1} = f(x_t, u_t, w_t), \tag{6.1}$$

$$y_t = g(x_t, v_t), \quad t \geq 0, \tag{6.2}$$

for (Borel measurable) functions f, g, with $\{w_t\}$ being an independent and identically distributed (i.i.d.) system noise process and $\{v_t\}$ an i.i.d. measurement disturbance process, which are independent of x_0 and each other. Here, $x_t \in \mathbb{X}$, $y_t \in \mathbb{Y}$, $u_t \in \mathbb{U}$, where we assume that these spaces are Borel subsets of finite dimensional Euclidean spaces.

In (6.2), we can view g as inducing a measurement channel Q, which is a stochastic kernel or a regular conditional probability measure from $\mathbb{X}$ to $\mathbb{Y}$ in the sense that $Q(\cdot|x)$ is a probability measure on the (Borel) σ-algebra $\mathcal{B}(\mathbb{Y})$ on $\mathbb{Y}$ for every $x \in \mathbb{X}$, and $Q(A|\cdot) : \mathbb{X} \to [0, 1]$ is a Borel measurable function for every $A \in \mathcal{B}(\mathbb{Y})$.

In networked control systems, the observation channel described above itself is also subject to design. In a more general setting, we can shape the channel input by coding and decoding. This chapter is concerned with design and optimization of such channels.

We will consider a controlled Markov model given by (6.1). The observation channel model is described as follows: This system is connected over a noisy channel with a finite capacity to a controller, as shown in Fig. 6.1. The controller has

S. Yüksel (✉)
Department of Mathematics and Statistics, Queen's University, Kingston, Ontario, Canada, K7L 3N6
e-mail: yuksel@mast.queensu.ca

G. Como et al. (eds.), *Information and Control in Networks*, Lecture Notes in Control and Information Sciences 450,
DOI 10.1007/978-3-319-02150-8_6,

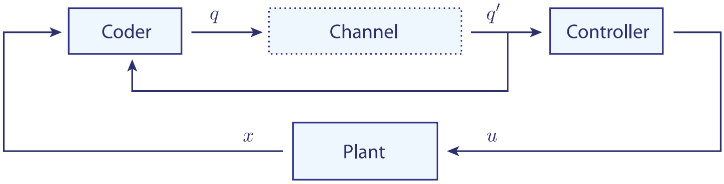

Fig. 6.1 Control over a noisy channel with feedback. The quantizer and the channel encoder form the coder in the figure

access to the information it has received through the channel. A quantizer maps the source symbols, state values, to corresponding channel inputs. The quantizer outputs are transmitted through a channel, after passing through a channel encoder. We assume that the channel is a discrete channel with input alphabet $\mathcal{M}$ and output alphabet $\mathcal{M}'$. Hence, the channel maps $q \in \mathcal{M}$ to channel outputs $q' \in \mathcal{M}'$ probabilistically so that $P(q'|q)$ is a stochastic kernel. Further probabilistic properties can be imposed on the channels depending on the particular application.

We refer by a *Composite Coding Policy* Π^{comp}, a sequence of functions $\{Q_t^{\text{comp}}, t \geq 0\}$ which are causal such that the quantization output (channel input) at time t, $q_t \in \mathcal{M}$, under Π^{comp} is generated by a function of its local information, that is, a mapping measurable on the sigma-algebra generated by

$$\mathcal{I}_t^e = \left\{x_{[0,t]}, q'_{[0,t-1]}\right\}$$

to a finite set $\mathcal{M}$, the quantization output alphabet given by

$$\mathcal{M} := \{1, 2, \dots, M\},$$

for $0 \leq t \leq T-1$ and $i = 1, 2$. Here, we have the notation for $t \geq 1$:

$$x_{[0,t-1]} = \{x_s, 0 \leq s \leq t-1\}.$$

The receiver/controller, upon receiving the information from the encoders, generates its decision at time t, also causally: An admissible causal controller policy is a sequence of functions $\gamma = \{\gamma_t\}$ such that

$$\gamma_t : \mathcal{M}'^{t+1} \to \mathbb{R}^m, \quad t \geq 0,$$

so that $u_t = \gamma_t(q'_{[0,t]})$.

We call such encoding and control policies, *causal* or *admissible*.

Two problems will be considered.

Problem P1: Value and Design of Information Channels for Optimization Given a controlled dynamical system (6.1), find solutions to minimization problem

$$\inf_{\Pi^{\text{comp}},\gamma} E_P^{\Pi^{\text{comp}},\gamma}\left[\sum_{t=0}^{T-1} c(x_t, u_t)\right], \tag{6.3}$$

over the set of all admissible coding and control policies, given $c : \mathbb{X} \times \mathbb{U} \to \mathbb{R}_+$, a cost function.

Problem P2: Value of Information Channels for Stabilization The second problem concerns stabilization. In this setting, we replace (6.1) with an n-dimensional linear system of the form

$$x_{t+1} = Ax_t + Bu_t + w_t, \quad t \geq 0 \tag{6.4}$$

where x_t is the state at time t, u_t is the control input, the initial state x_0 is a zero-mean second order random variable, and $\{w_t\}$ is a sequence of zero-mean i.i.d. Gaussian random variables, also independent of x_0. We assume that the system is open-loop unstable and controllable, that is, at least one eigenvalue has magnitude greater than 1.

The stabilization problem is as follows: Given a system of the form (6.4) controlled over a channel, find the set of channels $\mathcal{Q}$ for which there exists a policy (both control and encoding) such that $\{x_t\}$ is stable. Stochastic stability notions will be ergodicity and existence of finite moments, to be specified later.

The literature on such problems is rather long and references will be cited as they are particularly relevant. We refer the reader to [56, 63], and [58] for a detailed literature review.

6.2 Problem P1: Channel Design for Optimization

In this section, we consider the optimization problem. We will first consider a single state problem and investigate topological properties of measurement channels.

6.2.1 Measurement Channels as Information Structures

6.2.1.1 Topological Characterization of Measurement Channels

Let, as in (6.2), g induce a stochastic kernel Q, P be the probability measure on the initial state, and PQ denote the joint distribution induced on $(\mathbb{X} \times \mathbb{Y}, \mathcal{B}(\mathbb{X} \times \mathbb{Y}))$ by channel Q with input distribution P via

$$PQ(A) = \int_A Q(dy|x)P(dx), \quad A \in \mathcal{B}(\mathbb{X} \times \mathbb{Y}).$$

We adopt the convention that given a probability measure μ, the notation $z \sim \mu$ means that z is a random variable with distribution μ.

Consider the following cost function:

$$J(P, Q, \gamma) = E_P^{Q,\gamma}\left[\sum_{t=0}^{T-1} c(x_t, u_t)\right], \tag{6.5}$$

over the set of all admissible policies γ, where $c : \mathbb{X} \times \mathbb{U} \to \mathbb{R}$ is a Borel measurable stagewise cost (loss) function and $E_P^{Q,\gamma}$ denotes the expectation with initial state probability measure given by P, under policy γ and given channel Q.

Here, we have $\mathbb{X} = \mathbb{R}^n$ and $\mathbb{Y} = \mathbb{R}^m$, and $\mathcal{Q}$ denotes the set of all measurement channels (stochastic kernels) with input space $\mathbb{X}$ and output space $\mathbb{Y}$.

Let $\{\mu_n,\ n \in \mathbb{N}\}$ be a sequence in $\mathcal{P}(\mathbb{R}^n)$, where $\mathcal{P}(\mathbb{R}^n)$ is the set of probability measures on $\mathbb{R}^n$. Recall that $\{\mu_n\}$ is said to converge to $\mu \in \mathcal{P}(\mathbb{R}^n)$ *weakly* [5] if

$$\int_{\mathbb{R}^n} c(x)\mu_n(dx) \to \int_{\mathbb{R}^n} c(x)\mu(dx)$$

for every continuous and bounded $c : \mathbb{R}^n \to \mathbb{R}$. The sequence $\{\mu_n\}$ is said to converge to $\mu \in \mathcal{P}(\mathbb{R}^n)$ *setwise* if

$$\mu_n(A) \to \mu(A), \quad \text{for all } A \in \mathcal{B}\big(\mathbb{R}^n\big)$$

For two probability measures $\mu, \nu \in \mathcal{P}(\mathbb{R}^n)$, the *total variation* metric is given by

$$\begin{aligned} \|\mu - \nu\|_{TV} &:= 2 \sup_{B \in \mathcal{B}(\mathbb{R}^n)} \big|\mu(B) - \nu(B)\big| \\ &= \sup_{f:\, \|f\|_\infty \le 1} \left| \int f(x)\mu(dx) - \int f(x)\nu(dx) \right|, \end{aligned}$$

where the infimum is over all measurable real f such that $\|f\|_\infty = \sup_{x \in \mathbb{R}^n} |f(x)| \le 1$. A sequence $\{\mu_n\}$ is said to converge to $\mu \in \mathcal{P}(\mathbb{R}^n)$ in total variation if $\|\mu_n - \mu\|_{TV} \to 0$.

These three convergence notions are in increasing order of strength: convergence in total variation implies setwise convergence, which in turn implies weak convergence.

Given these definitions, we have the following.

Definition 6.1 (Convergence of Channels [63])

(i) A sequence of channels $\{Q_n\}$ converges to a channel Q *weakly at input* P if $PQ_n \to PQ$ weakly.
(ii) A sequence of channels $\{Q_n\}$ converges to a channel Q *setwise at input* P if $PQ_n \to PQ$ setwise, i.e., if $PQ_n(A) \to PQ(A)$ for all Borel sets $A \subset \mathbb{X} \times \mathbb{Y}$.
(iii) A sequence of channels $\{Q_n\}$ converges to a channel Q in *total variation at input* P if $PQ_n \to PQ$ in total variation, i.e., if $\|PQ_n - PQ\|_{TV} \to 0$.

If we introduce the equivalence relation $Q \equiv Q'$ if and only if $PQ = PQ'$, $Q, Q' \in \mathcal{Q}$, then the convergence notions in Definition 6.1 only induce the corresponding topologies on the resulting equivalence classes in $\mathcal{Q}$, instead of $\mathcal{Q}$. Let

$$J(P, Q) := \inf_{\gamma} E_P^{Q,\gamma} \left[\sum_{t=0}^{T-1} c\big(x_t, \gamma_t(y_{[0,t]})\big) \right].$$

In the following, we will discuss the following problems.

Continuity on the space of measurement channels (stochastic kernels): Suppose that $\{Q_n, n \in \mathbb{N}\}$ is a sequence of communication channels converging in some sense to a channel Q. Then the question we ask is when does $Q_n \to Q$ imply

$$\inf_{\gamma \in \Gamma} J(P, Q_n, \gamma) \to \inf_{\gamma \in \Gamma} J(P, Q, \gamma)?$$

Existence of optimal measurement channels and quantizers: Let $\mathcal{Q}$ be a set of communication channels. A second question we ask is when do there exist minimizing and maximizing channels for the optimization problems

$$\inf_{Q \in \mathcal{Q}} \inf_{\gamma} E_P^{Q,\gamma}\left[\sum_{t=0}^{T-1} c(x_t, u_t)\right] \quad \text{and} \quad \sup_{Q \in \mathcal{Q}} \inf_{\gamma} E_P^{Q,\gamma}\left[\sum_{t=0}^{T-1} c(x_t, u_t)\right]. \tag{6.6}$$

If solutions to these problems exist, are they unique?

Before proceeding further, however, we will obtain in the next section a structural result on such optimization problems.

6.2.1.2 Concavity of the Measurement Channel Design Problem and Blackwell's Comparison of Information Structures

We first present the following concavity results.

Theorem 6.1 [61] *Let $T = 1$ and let the integral $\int c(x, \gamma(y)) PQ(dx, dy)$ exist for all $\gamma \in \boldsymbol{\Gamma}$ and $Q \in \mathcal{Q}$. Then, the function*

$$J(P, Q) = \inf_{\gamma \in \boldsymbol{\Gamma}} E_P^{Q,\gamma}\big[c(x, u)\big]$$

is concave in Q.

Proof For $\alpha \in [0, 1]$ and $Q', Q'' \in \mathcal{Q}$, let $Q = \alpha Q' + (1-\alpha)Q'' \in \mathcal{Q}$, i.e.,

$$Q(A|x) = \alpha Q'(A|x) + (1-\alpha)Q''(A|x)$$

for all $A \in \mathcal{B}(\mathbb{Y})$ and $x \in \mathbb{X}$. Noting that $PQ = \alpha PQ' + (1-\alpha)PQ''$, we have

$$\begin{aligned} J(P, Q) = J\big(P, \alpha Q' + (1-\alpha)Q''\big) &= \inf_{\gamma \in \boldsymbol{\Gamma}} E_P^{Q,\gamma}\big[c(x, u)\big] \\ &= \inf_{\gamma \in \boldsymbol{\Gamma}} \int c\big(x, \gamma(y)\big) PQ(dx, dy) \\ &= \inf_{\gamma \in \boldsymbol{\Gamma}} \bigg(\alpha \int c\big(x, \gamma(y)\big) PQ'(dx, dy) \end{aligned}$$

$$
\begin{aligned}
&+ (1-\alpha) \int c\bigl(x, \gamma(y)\bigr) P Q''(dx, dy)\biggr) \\
&\geq \inf_{\gamma \in \boldsymbol{\Gamma}} \biggl(\alpha \int c\bigl(x, \gamma(y)\bigr) P Q'(dx, dy)\biggr) \\
&\quad + \inf_{\gamma \in \boldsymbol{\Gamma}} \biggl((1-\alpha) \int c\bigl(x, \gamma(y)\bigr) P Q''(dx, dy)\biggr) \\
&= \alpha J\bigl(P, Q'\bigr) + (1-\alpha) J\bigl(P, Q''\bigr),
\end{aligned}
\tag{6.7}
$$

proving that $J(P, Q)$ is concave in Q. □

Proposition 6.1 [61] *The function*

$$
V(P) := \inf_{u \in \mathbb{U}} \int c(x, u) P(dx),
$$

is concave in P, *under the assumption that c is measurable and bounded.*

We will use the preceding observation to revisit a classical result in statistical decision theory and comparison of experiments, due to David Blackwell [4]. In a single decision maker setup, we refer to the probability space induced on $\mathbb{X} \times \mathbb{Y}$ as an information structure.

Definition 6.2 An information structure induced by some channel Q_2 is weakly stochastically degraded with respect to another one, Q_1, if there exists a channel Q' on $\mathbb{Y} \times \mathbb{Y}$ such that

$$
Q_2(B|x) = \int_{\mathbb{Y}} Q'(B|y) Q_1(dy|x), \quad B \in \mathcal{B}(\mathbb{Y}), \ x \in \mathbb{X}.
$$

We have the following.

Theorem 6.2 (Blackwell [4]) *If Q_2 is weakly stochastically degraded with respect to Q_1, then the information structure induced by channel Q_1 is more informative with respect to the one induced by channel Q_2 in the sense that*

$$
\inf_{\gamma} E_P^{Q_2, \gamma}\bigl[c(x, u)\bigr] \geq \inf_{\gamma} E_P^{Q_1, \gamma}\bigl[c(x, u)\bigr],
$$

for all measurable and bounded cost functions c.

Proof The proof follows from [61]. Let $(x, y^1) \sim P Q_1$, y^2 be such that $\Pr(y^2 \in B | x = x, y^1 = y) = Q'(B|y)$ for all $B \in \mathcal{B}(\mathbb{Y})$, $y^1 \in \mathbb{Y}$, and $x \in \mathbb{X}$. Then x, y^1, and y^2 form a Markov chain in this order, and therefore $P(dy^2|y^1, x) = P(dy^2|y^1)$ and $P(x|dy^2, y^1) = P(x|y^1)$. Thus we have

$$
J(P, Q_2) = \int V\bigl(P\bigl(\cdot|y^2\bigr)\bigr) P\bigl(dy^2\bigr)
$$

$$
\begin{aligned}
&= \int V\Big(\int P(\cdot|y^1)P(dy^1|y^2)\Big)P(dy^2) \\
&\geq \int\Big(\int P(dy^1|y^2)V(P(\cdot|y^1))\Big)P(dy^2) \\
&= \int V(P(\cdot|y^1))\Big(\int P(dy^1|y^2)P(dy^2)\Big) \\
&= \int V(P(\cdot|y^1))P(dy^1) = J(P, Q_1),
\end{aligned}
$$

where in arriving at the inequality, we used Proposition 6.1 and Jensen's inequality. □

Remark 6.1 When $\mathbb{X}$ is finite, Blackwell showed that the above condition also has a converse theorem if P has positive measure on each element of $\mathbb{X}$: For an information structure to be more informative, weak stochastic degradedness is a necessary condition. For Polish $\mathbb{X}$ and $\mathbb{Y}$, the converse result holds under further technical conditions on the stochastic kernels (information structures), see [6] and [10].

The comparison argument applies also for the case $T > 1$.

Theorem 6.3 [60] *For the multi-stage problem* (6.5), *if Q_2 is weakly stochastically degraded with respect to Q_1, then the information structure induced by channel Q_1 is more informative with respect to the one induced by channel Q_2 in the sense that for all measurable and bounded cost functions c in* (6.5)

$$
J(P, Q_1) \leq J(P, Q_2).
$$

Remark 6.2 Blackwell's informativeness provides a partial order in the space of measurement channels; that is, not every pair of channels can be compared. We will later see that, if the goal is not the minimization of a cost function, but that of stochastic stabilization in an appropriate sense, then one can obtain a total order on the space of channels.

6.2.1.3 Single Stage: Continuity of the Optimal Cost in Channels

In this section, we study continuity properties under total variation, setwise convergence, and weak convergence, for the single-stage case. Thus, we investigate the continuity of the functional

$$
\begin{aligned}
J(P, Q) &= \inf_{\gamma} E_P^{Q,\gamma}\big[c(x_0, u_0)\big] \\
&= \inf_{\gamma \in \boldsymbol{\Gamma}} \int_{\mathbb{X}\times\mathbb{Y}} c\big(x, \gamma(y)\big)Q(dy|x)P(dx) \qquad (6.8)
\end{aligned}
$$

in the channel $Q \in \mathcal{Q}$, where $\boldsymbol{\Gamma}$ is the collection of all Borel measurable functions mapping $\mathbb{Y}$ into $\mathbb{U}$. Note that γ is an admissible one-stage control policy. As before, $\mathcal{Q}$ denotes the set of all channels with input space $\mathbb{X}$ and output space $\mathbb{Y}$.

Our results in this section as well as subsequent sections in this chapter will utilize one or more of the assumptions on the cost function c and the (Borel) set $\mathbb{U} \subset \mathbb{R}^k$:

Assumption 6.1

A1. The function $c : \mathbb{X} \times \mathbb{U} \to \mathbb{R}$ is non-negative, bounded, and continuous on $\mathbb{X} \times \mathbb{U}$.
A2. The function $c : \mathbb{X} \times \mathbb{U} \to \mathbb{R}$ is non-negative, measurable, and bounded.
A3. The function $c : \mathbb{X} \times \mathbb{U} \to \mathbb{R}$ is non-negative, measurable, bounded, and continuous on $\mathbb{U}$ for every $x \in \mathbb{X}$.
A4. $\mathbb{U}$ is a compact set.

Before proceeding further, we look for conditions under which an optimal control policy exists, i.e., when the infimum in $\inf_\gamma E_P^{Q,\gamma}[c(x,u)]$ is a minimum.

Theorem 6.4 [63] *Suppose assumptions* A3 *and* A4 *hold. Then, there exists an optimal control policy for any channel Q.*

Theorem 6.5 [63]

(a) *J defined in (6.8) is not continuous under setwise or weak convergence even for continuous and bounded cost functions c.*
(b) *Suppose that c is continuous and bounded on $\mathbb{X} \times \mathbb{U}$, $\mathbb{U}$ is compact, and $\mathbb{U}$ is convex. If $\{Q_n\}$ is a sequence of channels converging weakly at input P to a channel Q, then J satisfies $\limsup_{n\to\infty} J(P, Q_n) \le J(P, Q)$, that is, $J(P, Q)$ is upper semi-continuous under weak convergence.*
(c) *If c is bounded, measurable, then J is sequentially upper semi-continuous on $\mathcal{Q}$ under setwise convergence.*

We have continuity under the stronger notion of total variation.

Theorem 6.6 [63] *Under Assumption* A2, *the optimal cost $J(P, Q)$ is continuous on the set of communication channels $\mathcal{Q}$ under the topology of total variation.*

Thus, total variation, although a strong metric, is useful in establishing continuity. This will be useful in our analysis to follow for the existence of optimal quantization/coding policies.

In [63], (sequential) compactness conditions for a set of communication channels have been established. Given the continuity conditions, these may be used to identify conditions for the existence of best and worst channels for (6.6) when $T = 1$.

6.2.2 Quantizers as a Class of Channels

In this section, we consider the problem of convergence and optimization of quantizers.

We start with the definition of a quantizer.

Definition 6.3 An M-cell *vector quantizer*, Q, is a (Borel) measurable mapping from a subset of $\mathbb{X} = \mathbb{R}^n$ to the finite set $\{1, 2, \ldots, M\}$, characterized by a measurable partition $\{B_1, B_2, \ldots, B_M\}$ such that $B_i = \{x : Q(x) = i\}$ for $i = 1, \ldots, M$. The B_i's are called the cells (or bins) of Q.

We allow for the possibility that some of the cells of the quantizer are empty. Traditionally, in source coding theory, a quantizer is a mapping $Q : \mathbb{R}^n \to \mathbb{R}^n$ with a finite range. Thus q is defined by a partition and a reconstruction value in $\mathbb{R}^n$ for each cell in the partition. That is, for given cells $\{B_1, \ldots, B_M\}$ and reconstruction values $\{q^1, \ldots, q^M\} \subset \mathbb{R}^n$, we have $Q(x) = q^i$ if and only if $x \in B_i$. In the definition above, we do not include the reconstruction values.

A quantizer Q with cells $\{B_1, \ldots, B_M\}$ can also be characterized as a stochastic kernel Q from $\mathbb{X}$ to $\{1, \ldots, M\}$ defined by

$$Q(i|x) = 1_{\{x \in B_i\}}, \quad i = 1, \ldots, M,$$

so that $Q(x) = \sum_{i=1}^{M} q^i Q(i|x)$. We denote by $\mathcal{Q}_D(M)$ the space of all M-cell quantizers represented in the channel form. In addition, we let $\mathcal{Q}(M)$ denote the set of (Borel) stochastic kernels from $\mathbb{X}$ to $\{1, \ldots, M\}$, i.e., $Q \in \mathcal{Q}(M)$ if and only if $Q(\cdot|x)$ is probability distribution on $\{1, \ldots, M\}$ for all $x \in \mathbb{X}$, and $Q(i|\cdot)$ is Borel measurable for all $i = 1, \ldots, M$. Note that $\mathcal{Q}_D(M) \subset \mathcal{Q}(M)$. We note also that elements of $\mathcal{Q}(M)$ are sometimes referred to as random quantizers.

Consider the set of probability measures

$$\Theta := \big\{\zeta \in P\big(\mathbb{R}^n \times \mathcal{M}\big) : \zeta = PQ, Q \in \mathcal{Q}\big\},$$

on $\mathbb{R}^n \times \mathcal{M}$ having fixed input marginal P, equipped with weak topology. This set is the (Borel measurable) set of the extreme points on the set of probability measures on $\mathbb{R}^n \times \mathcal{M}$ with a fixed input marginal P [9]. Borel measurability of Θ follows from [40] since set of probability measures on $\mathbb{R}^n \times \mathcal{M}$ with a fixed input marginal P is a convex and compact set in a complete separable metric space, and therefore, the set of its extreme points is Borel measurable.

Lemma 6.1 [63] *The set of quantizers $\mathcal{Q}_D(M)$ is setwise sequentially precompact at any input P.*

Proof The proof follows from the interpretation above viewing a quantizer as a channel. In particular, a majorizing finite measure ν is obtained by defining $\nu = P \times \lambda$, where λ is the counting measure on $\{1, \ldots, M\}$ (note that $\nu(\mathbb{R}^n \times$

$\{1, \ldots, M\}) = M$). Then for any measurable $B \subset \mathbb{R}^n$ and $i = 1, \ldots, M$, we have $\nu(B \times \{i\}) = P(B)\lambda(\{i\}) = P(B)$ and thus

$$PQ\big(B \times \{i\}\big) = P(B \cap B_i) \leq P(B) = \nu\big(B \times \{i\}\big).$$

Since any measurable $D \subset \mathbb{X} \times \{1, \ldots, M\}$ can be written as the disjoint union of the sets $D_i \times \{i\}$, $i = 1, \ldots, M$, with $D_i = \{x \in \mathcal{X} : (x, i) \in D\}$, the above implies $PQ(D) \leq \nu(D)$ and this domination leads to precompactness under setwise convergence (see [5, Theorem 4.7.25]). □

The following lemma provides a useful result.

Lemma 6.2 [63] *A sequence $\{Q_n\}$ in $\mathcal{Q}(M)$ converges to a Q in $\mathcal{Q}(M)$ setwise at input P if and only if*

$$\int_A Q_n(i|x)P(dx) \to \int_A Q(i|x)P(dx) \quad \textit{for all } A \in \mathcal{B}(\mathbb{X}) \textit{ and } i = 1, \ldots, M.$$

Proof The lemma follows by noticing that for any $Q \in \mathcal{Q}(M)$ and measurable $D \subset \mathbb{X} \times \{1, \ldots, M\}$,

$$PQ(D) = \int_D Q(dy|x)P(dx) = \sum_{i=1}^{M} \int_{D_i} Q(i|x)P(dx)$$

where $D_i = \{x \in \mathcal{X} : (x, i) \in D\}$. □

However, unfortunately, the space of quantizers $\mathcal{Q}_D(M)$ is not closed under setwise (and hence, weak) convergence, see [63] for an example. This will lead us to consider further restrictions in the class of quantizers considered below.

In the following, we show that an optimal channel can be replaced with an optimal quantizer without any loss in performance.

Proposition 6.2 [63] *For any $Q \in \mathcal{Q}(M)$, there exists a $Q' \in \mathcal{Q}_D(M)$ with $J(P, Q') \leq J(P, Q)$. If there exists an optimal channel in $\mathcal{Q}(M)$, then there is a quantizer in $\mathcal{Q}_D(M)$ that is optimal.*

Proof For a policy $\gamma : \{1, \ldots, M\} \to \mathbb{U} = \mathbb{X}$ (with finite cost) define for all i,

$$\bar{B}_i = \big\{x : c\big(x, \gamma(i)\big) \leq c\big(x, \gamma(j)\big), j = 1, \ldots, M\big\}.$$

Letting $B_1 = \bar{B}_1$ and $B_i = \bar{B}_i \setminus \bigcup_{j=1}^{i-1} B_j$, $i = 2, \ldots, M$, we obtain a partition $\{B_1, \ldots, B_M\}$ and a corresponding quantizer $Q' \in \mathcal{Q}_D(M)$. Then $E_P^{Q',\gamma}[c(x, u)] \leq E_P^{Q,\gamma}[c(x, u)]$ for any $Q \in \mathcal{Q}(M)$. □

The following shows that setwise convergence of quantizers implies convergence under total variation.

Theorem 6.7 [63] *Let $\{Q_n\}$ be a sequence of quantizers in $\mathcal{Q}_D(M)$ which converges to a quantizer $Q \in \mathcal{Q}_D(M)$ setwise at P. Then, the convergence is also under total variation at P.*

Combined with Lemma 6.2, this theorem will be used to establish existence of optimal quantizers.

Now, assume $Q \in \mathcal{Q}_D(M)$ with cells $B_1, \dots, B_M$, each of which is a convex subset of $\mathbb{R}^n$. By the separating hyperplane theorem [24], there exist pairs of complementary closed half spaces $\{(H_{i,j}, H_{j,i}) : 1 \le i, j \le M, i \ne j\}$ such that for all $i = 1, \dots, M$,

$$B_i \subset \bigcap_{j \ne i} H_{i,j}.$$

Each $\bar{B}_i := \bigcap_{j \ne i} H_{i,j}$ is a closed convex polytope and by the absolute continuity of P one has $P(\bar{B}_i \setminus B_i) = 0$ for all $i = 1, \dots, M$. One can thus obtain a (P-a.s.) representation of Q by the $M(M-1)/2$ hyperplanes $h_{i,j} = H_{i,j} \cap H_{j,i}$.

Let $\mathcal{Q}_C(M)$ denote the collection of M-cell quantizers with convex cells and consider a sequence $\{Q_n\}$ in $\mathcal{Q}_C(M)$. It can be shown (see the proof of Theorem 1 in [20]) that using an appropriate parametrization of the separating hyperplanes, a subsequence Q_{n_k} can be found which converges to a $Q \in \mathcal{Q}_C(M)$ in the sense that $P(B_i^{n_k} \triangle B_i) \to 0$ for all $i = 1, \dots, M$, where the $B_i^{n_k}$ and the B_i are the cells of Q_{n_k} and Q, respectively.

In the following, we consider quantizers with convex codecells and an input distribution that is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^n$ [20]. We note that such quantizers are commonly used in practice; for cost functions of the form $c(x,u) = \|x-u\|^2$ for $x, u \in \mathbb{R}^n$, the cells of optimal quantizers, if they exist, will be convex by Lloyd–Max conditions of optimality; see [20] for further results on convexity of bins for entropy constrained quantization problems.

Theorem 6.8 [63] *The set $\mathcal{Q}_C(M)$ is compact under total variation at any input measure P that is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^n$.*

We can now state an existence result for optimal quantization.

Theorem 6.9 [63] *Let P admit a density function and suppose the goal is to find the best quantizer Q with M cells minimizing $J(P,Q) = \inf_\gamma E_P^{Q,\gamma} c(x,u)$ under Assumption* A2, *where Q is restricted to $\mathcal{Q}_C(M)$. Then an optimal quantizer exists.*

Remark 6.3 Regarding existence results, there have been few studies in the literature in addition to [63]. The authors of [1] and [41] have considered nearest neighbor encoding/decoding rules for norm based distortion measures. The L_2-norm leads to convex codecells for optimal design. We also note that the convexity assumption as well as the atomlessness property of the input measure can be relaxed in a class of settings, see [1] and Remark 4.9 in [60].

6.2.3 *The Multi-stage Case*

6.2.3.1 Static Channel/Coding

We now consider the general stochastic control problem in (6.5) with T stages. It should be noted that the effect of a control policy applied at any given time-stage presents itself in two ways, in the cost incurred at the given time-stage and the effect on the process distribution (and the evolution of the controller's uncertainty on the state) at future time-stages. This is known as the dual effect of control [3]. The next theorem shows the continuity of the optimal cost in the measurement channel under some regularity conditions.

Definition 6.4 A sequence of channels $\{Q_n\}$ converges to a channel Q *uniformly* in total variation if

$$\lim_{n\to\infty} \sup_{x\in\mathbb{X}} \left\| Q_n(\cdot|x) - Q(\cdot|x) \right\|_{TV} = 0.$$

Note that in the special but important case of additive measurement channels, uniform convergence in total variation is equivalent to the weaker condition that $Q_n(\cdot|x) \to Q(\cdot|x)$ in total variation for each x. When the additive noise is absolutely continuous with respect to the Lebesgue measure, uniform convergence in total variation is equivalent to requiring that the noise density corresponding to Q_n converges in the L_1-sense to the density corresponding to Q. For example, if the noise density is estimated from n independent observations using any of the L_1-consistent density estimates described in, e.g., [15], then the resulting Q_n will converge (with probability one) uniformly in total variation [63].

Theorem 6.10 [63] *Consider the cost function* (6.5) *with arbitrary* $T \in \mathbb{N}$. *Suppose Assumption* A2 *holds. Then, the optimization problem is continuous in the observation channel in the sense that if* $\{Q_n\}$ *is a sequence of channels converging to* Q *uniformly in total variation, then*

$$\lim_{n\to\infty} J(P, Q_n) = J(P, Q).$$

We obtained the continuity of the optimal cost on the space of channels equipped with a more stringent notion for convergence in total variation. This result and its proof indicate that further technical complications arise in multi-stage problems. Likewise, upper semi-continuity under weak convergence and setwise convergence require more stringent uniformity assumptions. On the other hand, the concavity property applies directly to the multi-stage case. That is, $J(P, Q)$ is concave in the space of channels; the proof of this result follows that of Theorem 6.1.

One further interesting problem regarding the multi-stage case is to consider adaptive observation channels. For example, one may aim to design optimal adaptive quantizers for a control problem. We consider this next.

6.2.3.2 Dynamic Channel and Optimal Vector Quantization

We consider a causal encoding problem where a sensor encodes an observed source to a receiver with zero-delay. Consider the source in (6.1). The source $\{x_t\}$ to be encoded is an $\mathbb{R}^n$-valued Markov process. The encoder encodes (quantizes) its information $\{x_t\}$ and transmits it to a receiver over a discrete noiseless channel with common input and output alphabet $\mathcal{M} := \{1, 2, \dots, M\}$, where M is a positive integer, i.e., the encoder quantizes its information.

As in (6.5), for a finite horizon setting the goal is to minimize the cost

$$J_{\pi_0}\big(\Pi^{\mathrm{comp}}, \gamma, T\big) := E_{\pi_0}^{\Pi^{\mathrm{comp}},\gamma}\left[\frac{1}{T}\sum_{t=0}^{T-1} c_0(x_t, u_t)\right], \tag{6.9}$$

for some $T \geq 1$, where $c_0 : \mathbb{R}^n \times \mathbb{U} \to \mathbb{R}_+$ is a (measurable) cost function and $E_{\pi_0}^{\Pi}[\cdot]$ denotes the expectation with initial state distribution π_0 and under the composite quantization policy Π^{comp} and receiver policy γ.

There are various structural results for such problems, primarily for control-free sources; see [25, 27, 49, 52, 53, 58] among others. In the following, we consider the case with control, which have been considered for finite-alphabet source and control action spaces in [51] and [27]. The result essentially follows from Witsenhausen [53].

Theorem 6.11 [57] *For the finite horizon problem, any* causal composite quantization policy *can be replaced without any loss in performance by one which, at time $t = 1, \dots, T-1$, only uses, x_t and $q_{[0,t-1]}$, with the original control policy unaltered.*

Hereafter, let $\mathcal{P}(\mathbb{X})$ denote the space of probability measures on $\mathbb{X}$ endowed with weak convergence. Given a composite quantization policy Π^{comp}, let $\pi_t \in \mathcal{P}(\mathbb{R}^n)$ be the conditional probability measure defined by

$$\pi_t(A) := P(x_t \in A | q_{[0,t-1]})$$

for any Borel set A. Walrand and Varaiya [52] considered sources taking values in a finite set, and obtained the essence of the following result. For control-free sources, the result appears in [58] for $\mathbb{R}^n$-valued sources.

Theorem 6.12 [57] *For a finite horizon problem, any* causal composite quantization policy *can be replaced, without any loss in performance, by one which at any time $t = 1, \dots, T-1$ only uses the conditional probability $P(dx_{t-1}|q_{[0,t-1]})$ and the state x_t. This can be expressed as a quantization policy which only uses (π_t, t) to generate a quantizer $Q_t : \mathbb{R}^n \to \mathcal{M}$, where the quantizer Q_t uses x_t to generate the quantization output as $q_t = Q_t(x_t)$ at time t.*

For any quantization policy in Π_W and any $T \geq 1$ we have

$$\inf_{\gamma} J_{\pi_0}\big(\Pi^{\text{comp}}, \gamma, T\big) = E_{\pi_0}^{\Pi^{\text{comp}}}\left[\frac{1}{T}\sum_{t=0}^{T-1} c(\pi_t, Q_t)\right],$$

where

$$c(\pi_t, Q_t) = \sum_{i=1}^{M} \inf_{u\in\mathbb{U}} \int_{Q_t^{-1}(i)} \pi_t(dx)c_0(x,u).$$

In the following, we consider the existence problem. However, to facilitate the analysis we will take the source to be control-free, and assume further structure on the source process. We have the following assumptions in the source $\{x_t\}$ and the cost function.

Assumption 6.2

(i) The evolution of the Markov source $\{x_t\}$ is given by

$$x_{t+1} = f(x_t) + w_t, \quad t \geq 0 \tag{6.10}$$

where $\{w_t\}$ is an independent and identically distributed zero-mean Gaussian vector noise sequence and $f : \mathbb{R}^n \to \mathbb{R}^n$ is measurable.

(ii) $\mathbb{U}$ is compact and $c_0 : \mathbb{R}^n \times \mathbb{U} \to \mathbb{R}_+$ is bounded and continuous.

(iii) The initial condition x_0 is zero-mean Gaussian.

We note that the class of quantization policies which admit the structure suggested in Theorem 6.12 is an important one. We henceforth define:

$$\begin{aligned}\Pi_W := \big\{\Pi^{\text{comp}} = \big\{Q_t^{\text{comp}}, t \geq 0\big\} : \exists \Upsilon_t : \mathcal{P}(\mathbb{X}) \to \mathcal{Q} \\ Q_t^{\text{comp}}\big(\mathcal{I}_t^e\big) = \big(\Upsilon_t(\pi_t)\big)(x_t), \forall \mathcal{I}_t^e\big\},\end{aligned} \tag{6.11}$$

to represent this class of policies. For a policy in this class, properties of conditional probability lead to the following expression for $\pi_t(dx)$:

$$\frac{\int \pi_{t-1}(dx_{t-1}) P(q_{t-1}|\pi_{t-1}, x_{t-1}) P(x_t \in dx|x_{t-1})}{\int\int \pi_{t-1}(dx_{t-1}) P(q_{t-1}|\pi_{t-1}, x_{t-1}) P(x_t \in dx|x_{t-1})}.$$

Here, $P(q_{t-1}|\pi_{t-1}, x_{t-1})$ is determined by the quantizer policy. The following follows from the proof of Theorem 2.5 of [58].

Theorem 6.13 *The sequence of conditional measures and quantizers* $\{(\pi_t, Q_t)\}$ *form a controlled Markov process in* $\mathcal{P}(\mathbb{R}^n) \times \mathcal{Q}$.

Theorem 6.14 *Under Assumption 6.2, an optimal receiver policy exists.*

Proof At any given time an optimal receiver will minimize $\int P(dx_t|q_{[0,t]})c(x_t, u_t)$. The existence of a minimizer then follows from Theorem 3.1 in [63]. □

Let Π_W^C be the set of coding policies in Π_W with quantizers having convex codecells (that is, $Q_t \in \mathcal{Q}_C(M)$). We have the following result on the existence of optimal quantization policies.

Theorem 6.15 [62] *For any $T \geq 1$, under Assumption 6.2, there exists a policy in Π_W^C such that*

$$\inf_{\Pi^{\text{comp}} \in \Pi_W^C} \inf_{\gamma} J_{\pi_0}\big(\Pi^{\text{comp}}, \gamma, T\big) \tag{6.12}$$

is achieved. Letting $J_T^T(\cdot) = 0$ and

$$J_0^T(\pi_0) := \min_{\Pi^{\text{comp}} \in \Pi_W^C, \gamma} J_{\pi_0}\big(\Pi^{\text{comp}}, \gamma, T\big),$$

the dynamic programming recursion

$$T J_t^T(\pi_t) = \min_{Q \in \mathcal{Q}_C(M)} \big(c(\pi_t, Q_t) + T E\big[J_{t+1}^T(\pi_{t+1})|\pi_t, Q_t\big]\big)$$

holds for all $t = 0, 1, \ldots, T-1$.

We note that also for optimal multi-stage vector quantization, [8] has obtained existence results for an infinite horizon setup with discounted costs under a uniform boundedness assumption on the reconstruction levels.

6.2.3.3 The Linear Quadratic Gaussian (LQG) Case

There is a large literature on jointly optimal quantization for the LQG problem dating back to early 1960s (see, for example, [23] and [13]). References [2, 7, 17, 18, 31, 38, 48], and [57] considered the optimal LQG quantization and control, with various results on the optimality or the lack of optimality of the separation principle.

For controlled Markov sources, in the context of Linear Quadratic Gaussian (LQG) systems, existence of optimal policies has been established in [57] and [60], where it has been shown that without any loss, the control actions can be decoupled from the performance of the quantization policies, and a result similar to Theorem 6.15 for linear systems driven by Gaussian noise leads to the existence of an optimal quantization policy.

Structural results with control have also been studied by Walrand and Varaiya [51] in the context of finite control and action spaces and by Mahajan and Teneketzis [27] for control over noisy channels, also for finite state-action space settings.

6.2.3.4 Case with Noisy Channels with Noiseless Feedback

The results presented in this section apply also to coding over discrete memoryless (noisy) channels (DMCs) with feedback. The equivalent results of Theorems 6.11 and 6.12 apply with q'_t terms replacing q_t, if q'_t is the output of a DMC at time t, as we state in the following.

In this context, let $\pi_t \in \mathcal{P}(\mathbb{X})$ to be the regular conditional probability measure given by $\pi_t(\cdot) = P(x_t \in \cdot | q'_{[0,t-1]})$, where q'_t is the channel output when the input is q_t. That is, $\pi_t(A) = P(x_t \in A | q'_{[0,t-1]})$, $A \in \mathcal{B}(\mathbb{X})$.

Theorem 6.16 [60] *Any composite encoding policy can be replaced, without any loss in performance, by one which only uses x_t and $q'_{[0,t-1]}$ at time $t \geq 1$ to generate the channel input q_t.*

Theorem 6.17 [60] *Any composite quantization policy can be replaced, without any loss in performance, by one which only uses the conditional probability measure $\pi_t(\cdot) = P(x_t \in \cdot | q'_{[0,t-1]})$, the state x_t, and the time information t, at time $t \geq 1$ to generate the channel input q_t.*

Remark 6.4 When there is no feedback from the controller, or when there is noisy feedback, the analysis requires a Markov chain construction in a larger state space under certain conditions on the memory update rules at the decoder. We refer the reader to [26, 49], and [25] for a class of such settings.

6.3 Problem P2: Characterization of Information Channels for Stabilization

In this section, we consider the stabilization problem over communication channels. The goal will be to identify conditions so that the controlled state is stochastically stable in the sense that

- $\{x_t\}$ is asymptotically mean stationary (AMS) and satisfies the requirements of Birkhoff's sample path ergodic theorem. This may also include the condition that the controlled (and possibly sampled) state and encoder parameters have a unique invariant probability measure.
- $\lim_{T\to\infty} \frac{1}{T} \sum_{t=0}^{T-1} |x_t|^2$ exists and is finite almost surely (this will be referred to as quadratic stability).

There is a very large literature on this problem. Particularly related references include [11, 28–30, 32, 37, 42, 43, 45, 46, 54, 55, 59]. In the context of discrete channels, many of these papers considered a bounded noise assumption, except notably [30, 36, 37, 55, 64], and [56]. We refer the reader to [32] and [60] for a detailed literature review.

In this section, we will present a proof program developed in [55] and [64] for stochastic stabilization of Markov chains with event-driven samplings applied to networked control. Toward this end, we will first review few results from the theory of Markov chains. However, we first will establish fundamental bounds on information requirements for stabilization.

6.3.1 Fundamental Lower Bounds for Stabilization

We consider a scalar LTI discrete-time system and then later in Sect. 6.3.6 the multi-dimensional case. Here, the scalar system is described by

$$x_{t+1} = ax_t + bu_t + w_t, \quad t \geq 0 \tag{6.13}$$

where x_t is the state at time t, u_t is the control input, the initial state x_0 is a zero-mean second order random variable, and $\{w_t\}$ is a sequence of zero-mean i.i.d. Gaussian random variables, also independent of x_0. We assume that the system is open-loop unstable and controllable, that is, $|a| \geq 1$ and $b \neq 0$. This system is connected over a noisy channel with a finite capacity to a controller, as shown in Fig. 6.1, with the information structures described in Sect. 6.1.

We consider first memoryless noisy channels (in the following definitions, we assume feedback is not present; minor adjustments can be made to capture the case with feedback).

Definition 6.5 A Discrete Memoryless Channel (DMC) is characterized by a discrete input alphabet $\mathcal{M}$, a discrete output alphabet $\mathcal{M}'$, and a conditional probability mass function $P(q'|q)$, from $\mathcal{M} \times \mathcal{M}'$ to $\mathbb{R}$ which satisfies the following. Let $q_{[0,n]} \in \mathcal{M}^{n+1}$ be a sequence of input symbols, let $q'_{[0,n]} \in \mathcal{M}'^{n+1}$ be a sequence of output symbols, where $q_k \in \mathcal{M}$ and $q'_k \in \mathcal{M}'$ for all k and let P^{n+1}_{DMC} denote the joint mass function on the $(n+1)$-tuple input and output spaces. It follows that $P^{n+1}_{\mathrm{DMC}}(q'_{[0,n]}|q_{[0,n]}) = \prod_{k=0}^{n} P_{\mathrm{DMC}}(q'_k|q_k)$, $\forall q_{[0,n]} \in \mathcal{M}^{n+1}$, $q'_{[0,n]} \in \mathcal{M}'^{n+1}$, where q_k, q'_k denote the kth component of the vectors $q_{[0,n]}, q'_{[0,n]}$, respectively.

Channels can also have memory. We state the following for both discrete and continuous-alphabet channels.

Definition 6.6 A discrete channel (continuous channel) with memory is characterized by a sequence of discrete (continuous) input alphabets $\mathcal{M}^{n+1}$, discrete (continuous) output alphabets $\mathcal{M}^{n+1}$, and a sequence of regular conditional probability measures $P_n(dq'_{[0,n]}|q_{[0,n]})$, from $\mathcal{M}^{n+1}$ to $\mathcal{M}'^{n+1}$.

In this chapter, while considering discrete channels, we will assume channels with finite alphabets.

Remark 6.5 Another setting involves continuous-alphabet channels. Such channels can be regarded as limits of discrete-channels: Note that the mutual information for real valued random variables x, y is defined as

$$I(x;y) := \sup_{Q_1,Q_2} I\big(Q_1(x);Q_2(y)\big),$$

where Q_1 and Q_2 are quantizers with finitely many bins (see Chap. 5 in [19]). As a consequence, the discussion for discrete channels applies for continuous alphabet channels. On the other hand, the Gaussian channel is a very special channel which needs to be considered in its own right, especially in the context of linear quadratic Gaussian (LQG) systems and problems. A companion chapter deals with such channels, see [65], as well as [60].

Theorem 6.18 [56] *Suppose that a linear plant given as in* (6.13) *controlled over a DMC*, *under some admissible coding and controller policies*, *satisfies the condition*

$$\liminf_{T\to\infty} \frac{1}{T} h(x_T) \le 0, \tag{6.14}$$

where h denotes the entropy function. Then, *the channel capacity C must satisfy*

$$C \ge \log_2\big(|a|\big).$$

Remark 6.6 Condition (6.14) is a weak one. For example, a stochastic process whose second moment grows subexponentially in time, namely,

$$\liminf_{T\to\infty} \frac{\log(E[x_T^2])}{T} \le 0,$$

satisfies this condition.

We now present a supporting result due to Matveev.

Proposition 6.3 [30] *Suppose that a linear plant given as in* (6.13) *is controlled over a DMC. If*

$$C < \log_2\big(|a|\big),$$

then

$$\limsup_{T\to\infty} P\big(|x_T| \le b(T)\big) \le \frac{C}{\log_2(|a|)},$$

for all $b(T) > 0$ such that $\lim_{T\to\infty} \frac{1}{T}\log_2(b(T)) = 0$

We note that similar characterizations have also been considered in [29, 43], and [32], for systems driven by bounded noise.

Theorem 6.19 [60] *Suppose that a linear plant given as in* (6.13) *is controlled over a DMC. If, under some causal encoding and controller policy, the state process is AMS, then the channel capacity C must satisfy*

$$C \geq \log_2\bigl(|a|\bigr).$$

In the following, we will observe that the condition $C \geq \log_2(|a|)$ in Theorems 6.18 and 6.19 is almost sufficient as well for stability in the AMS sense. Furthermore, the result applies to multi-dimensional systems. Toward this goal, we first discuss the erasure channel with feedback (which includes the noiseless channel as a special case), and then consider more general DMCs, followed by a class of channels with memory. We will also investigate quadratic stability. We discuss an essential ingredient in the proof program next.

6.3.2 Stochastic Stability and Random-Time State-Dependent Drift Approach

Let $\boldsymbol{X} = \{x_t, t \geq 0\}$ denote a Markov chain with state space $\mathbb{X}$. Assume that the state space is a complete, separable, metric space, whose Borel σ-field is denoted $\mathcal{B}(\mathbb{X})$. Let the transition probability be denoted by P, so that for any $x \in \mathbb{X}$, $A \in \mathcal{B}(\mathbb{X})$, the probability of moving from x to A in one step is given by $P(x_{t+1} \in A \mid x_t = x) = P(x, A)$. The n-step transitions are obtained via composition in the usual way, $P(X_{t+n} \in A \mid X_t = x) = P^n(x, A)$, for any $n \geq 1$. The transition law acts on measurable functions $f : \mathbb{X} \to \mathbb{R}$ and measures μ on $\mathcal{B}(\mathbb{X})$ via

$$Pf(x) := \int_{\mathbb{X}} P(x, dy) f(y), \quad x \in \mathbb{X},$$

$$\mu P(A) := \int_{\mathbb{X}} \mu(dx) P(x, A), \quad A \in \mathcal{B}(\mathbb{X}).$$

A probability measure π on $\mathcal{B}(\mathbb{X})$ is called invariant if $\pi P = \pi$. That is,

$$\int \pi(dx) P(x, A) = \pi(A), \quad A \in \mathcal{B}(\mathbb{X}).$$

For any initial probability measure ν on $\mathcal{B}(\mathbb{X})$ we can construct a stochastic process with transition law P, and satisfying $x_0 \sim \nu$. We let P_ν denote the resulting probability measure on the sample space, with the usual convention $\nu = \delta_x$ when the initial state is $x \in \mathbb{X}$. When $\nu = \pi$, then the resulting process is stationary. A comprehensive treatment of Markov chains can be found in [35].

Throughout this subsection, the sequence of stopping times $\{\mathcal{T}_i : i \in \mathbb{N}_+\}$ is assumed to be non-decreasing, with $\mathcal{T}_0 = 0$, measurable on the filtration generated by the state process. Additional assumptions are made in the results that follow.

Before proceeding further, we note that a set $A \subset \mathbb{X}$ is μ-small on $(\mathbb{X}, \mathcal{B}(\mathbb{X}))$ if for some n, and some positive measure μ, $P^n(x, B) \geq \mu(B), \forall x \in A$, and $B \in \mathcal{B}(\mathbb{X})$. A small set leads to the construction of an accessible atom and to an invariant probability measure [35]. In many practical settings, compact sets are small; sufficient conditions on when a compact set is small has been presented in [60] and [35].

Theorem 6.20 [64] *Suppose that* $\mathbf{X}$ *is a* φ*-irreducible and aperiodic Markov chain. Suppose moreover that there are functions* $V: \mathbb{X} \to (0, \infty)$, $\delta: \mathbb{X} \to [1, \infty)$, $f: \mathbb{X} \to [1, \infty)$, *a small set* C *and a constant* $b \in \mathbb{R}$, *such that the following hold*:

$$E\big[V(\phi_{\mathcal{T}_{z+1}}) \bigm| \mathcal{F}_{\mathcal{T}_z}\big] \leq V(\phi_{\mathcal{T}_z}) - \delta(\phi_{\mathcal{T}_z}) + b1_{\{\phi_{\mathcal{T}_z} \in C\}},$$
$$E\left[\sum_{k=\mathcal{T}_z}^{\mathcal{T}_{z+1}-1} f(\phi_k) \mid \mathcal{F}_{\mathcal{T}_z}\right] \leq \delta(\phi_{\mathcal{T}_z}), \quad z \geq 0. \tag{6.15}$$

Then the following hold:

(i) ϕ *is positive Harris recurrent, with unique invariant distribution* π
(ii) $\pi(f) := \int f(\phi)\pi(d\phi) < \infty$
(iii) *For any function* g *that is bounded by* f, *in the sense that* $\sup_\phi |g(\phi)|/f(\phi) < \infty$, *we have convergence of moments in the mean, and the Law of Large Numbers holds*:

$$\lim_{t\to\infty} E_\phi\big[g(\phi_t)\big] = \pi(g),$$
$$\lim_{N\to\infty} \frac{1}{N} \sum_{t=0}^{N-1} g(\phi_t) = \pi(g) \quad a.s., \ \phi \in \mathbb{X}.$$

This theorem will be important for the stability analysis to follow.

6.3.3 Noiseless and Erasure Channels

We begin with erasure channels (which contain discrete noiseless channels as a special case), before discussing more general noisy channels. Before discussing the multi-dimensional case in Sect. 6.3.3.2, we first discuss the scalar version described by (6.13).

The details of the erasure channel are specified as follows: The channel source consists of state values from $\mathbb{R}$. The source output is, as before, quantized. We consider the following uniform quantizer class. A *modified uniform quantizer* Q_K^Δ : $\mathbb{R} \to \mathbb{R}$ with step size Δ and $K + 1$ (with K even) number of bins satisfies the

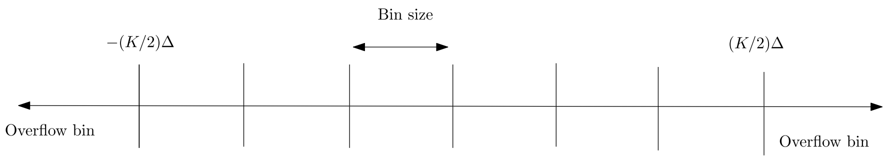

Fig. 6.2 A modified uniform quantizer. There is a single overflow bin

following for $k = 1, 2 \ldots, K$ (see Fig. 6.2):

$$Q_K^{\Delta}(x) = \begin{cases} (k - \frac{1}{2}(K+1))\Delta & \text{if } x \in [(k-1-\frac{1}{2}K)\Delta, (k-\frac{1}{2}K)\Delta), \\ (\frac{1}{2}(K-1))\Delta & \text{if } x = \frac{1}{2}K\Delta, \\ 0 & \text{if } x \notin [-\frac{1}{2}K\Delta, \frac{1}{2}K\Delta]. \end{cases} \tag{6.16}$$

where we have $\mathcal{M} = \{1, 2, \ldots, K+1\}$. The quantizer–decoder mapping thus described corresponds to a uniform quantizer with bin size Δ. The interval $[-K/2, K/2]$ is termed the *granular region* of the quantizer, and $\mathbb{R} \setminus [-K/2, K/2]$ is named the *overflow region* of the quantizer (see Fig. 6.2). We will refer to this quantizer as a *modified uniform quantizer*, since the overflow region is assigned a single bin.

The quantizer outputs are transmitted through a memoryless erasure channel, after being subjected to a bijective mapping, which is performed by the channel encoder. The channel encoder maps the quantizer output symbols to corresponding channel inputs $q \in \mathcal{M} := \{1, 2, \ldots, K+1\}$. A channel encoder at time t, denoted by $\mathcal{E}_t$, maps the quantizer outputs to $\mathcal{M}$ such that $\mathcal{E}_t(Q_t(x_t)) = q_t \in \mathcal{M}$.

The controller/decoder has access to noisy versions of the encoder outputs for each time, which we denote by $\{q'\} \in \mathcal{M} \cup \{e\}$, with e denoting the erasure symbol, generated according to a probability distribution for every fixed $q \in \mathcal{M}$. The channel transition probabilities are given by

$$P\big(q' = i | q = i\big) = p, \qquad P\big(q' = e | q = i\big) = 1 - p, \quad i \in \mathcal{M}.$$

At each time $t \geq 0$, the controller/decoder applies a mapping $\mathcal{D}_t : \mathcal{M} \cup \{e\} \to \mathbb{R}$, given by

$$\mathcal{D}_t\big(q_t'\big) = \mathcal{E}_t^{-1}\big(q_t'\big) \times 1_{\{q_t' \neq e\}} + 0 \times 1_{\{q_t' = e\}}.$$

Let $\{\Upsilon_t\}$ denote a binary sequence of i.i.d. random variables, representing the erasure process in the channel, where the event $\Upsilon_t = 1$ indicates that the signal is transmitted with no error through the channel at time t. Let $p = E[\Upsilon_t]$ denote the probability of success in transmission.

The following key assumptions are imposed throughout this section: Given $K \geq 2$ introduced in the definition of the quantizer, define the *rate variable*s

$$R := \log_2(K+1), \qquad R' = \log_2(K). \tag{6.17}$$

We fix positive scalars δ, α satisfying

$$|a|2^{-R'} < \alpha < 1 \tag{6.18}$$

and

$$\alpha\big(|a| + \delta\big)^{p^{-1}-1} < 1. \tag{6.19}$$

We consider the following update rules. For $t \in \mathbb{Z}_+$ and with $\Delta_0 \in \mathbb{R}$ selected arbitrarily, consider

$$\begin{aligned} u_t &= -\frac{a}{b}\hat{x}_t, \\ \hat{x}_t &= \mathcal{D}_t\big(q'_t\big) = \Upsilon_t Q_K^{\Delta_t}(x_t), \\ \Delta_{t+1} &= \Delta_t \bar{Q}\left(\Delta_t, \left|\frac{x_t}{\Delta_t 2^{R'-1}}\right|, \Upsilon_t\right). \end{aligned} \tag{6.20}$$

Here, $\bar{Q} : \mathbb{R} \times \mathbb{R} \times \{0, 1\} \to \mathbb{R}$ is defined below, where $L > 0$ is a constant;

$$\begin{aligned} &\bar{Q}(\Delta, h, p) = |a| + \delta \quad \text{if } |h| > 1, \text{ or } p = 0, \\ &\bar{Q}(\Delta, h, p) = \alpha \quad \text{if } 0 \le |h| \le 1,\ p = 1,\ \Delta > L, \\ &\bar{Q}(\Delta, h, p) = 1 \quad \text{if } 0 \le |h| \le 1,\ p = 1,\ \Delta \le L. \end{aligned}$$

The update equations above imply that

$$\Delta_t \ge L\alpha =: L'. \tag{6.21}$$

Without any loss of generality, we assume that $L' \ge 1$.

We note that given the channel output $q'_t \neq e$, the controller can simultaneously deduce the realization of Υ_t and the event $\{|h_t| > 1\}$, where $h_t := \frac{x_t}{\Delta_t 2^{R'-1}}$. This is due to the fact that if the channel output is not the erasure symbol, the controller knows that the signal is received with no error. If $q'_t = e$, however, then the controller applies 0 as its control input and enlarges the bin size of the quantizer. As depicted in Fig. 6.1, the encoder has access to channel outputs, that is, there is noiseless feedback.

Lemma 6.3 *Under* (6.20), *the process* (x_t, Δ_t) *is a Markov chain.*

Proof The system's state evolution can be expressed

$$x_{t+1} = ax_t - a\hat{x}_t + w_t,$$

where $\hat{x}_t = \Upsilon_t Q_K^{\Delta_t}(x_t)$. It follows that the process (x_t, Δ_t) evolves as a nonlinear state space model:

$$\begin{aligned} x_{t+1} &= a\big(x_t - \Upsilon_t Q_K^{\Delta_t}(x_t)\big) + w_t, \\ \Delta_{t+1} &= \Delta_t \bar{Q}\left(\Delta_t, \left|\frac{x_t}{2^{R'-1}\Delta_t}\right|, \Upsilon_t\right). \end{aligned} \tag{6.22}$$

in which (w_t, Υ_t) is i.i.d. Thus, the pair (x_t, Δ_t) forms a Markov chain. □

Let for a Borel set S, $\tau_S = \inf(k > 0 : (x_k, \Delta_k) \in S)$ and $E_{x,\Delta}$, $P_{(x,\Delta)}$ denote the expectation and probabilities conditioned on $(x_0, \Delta_0) = (x, \Delta)$.

Proposition 6.4 [64] *If* (6.18)–(6.19) *hold, then there exists a compact set $A \times B \subset \mathbb{R}^2$ satisfying the recurrence condition*

$$\sup_{(x,\Delta)\in A\times B} E_{x,\Delta}[\tau_{A\times B}] < \infty$$

and the recurrence condition $P_{(x,\Delta)}(\tau_{A\times B} < \infty) = 1$ for any admissible (x, Δ).

A result on the existence and uniqueness of an invariant probability measure is the following. It basically establishes irreducibility and aperiodicity, which leads to positive Harris recurrence, by Proposition 6.4.

Theorem 6.21 [64] *For an adaptive quantizer satisfying* (6.18)–(6.19), *suppose that the quantizer bin sizes are such that their base-2 logarithms are integer multiples of some scalar s, and $\log_2(\bar{Q}(\cdot))$ takes values in integer multiples of s. Then the process (x_t, Δ_t) forms a positive Harris recurrent Markov chain. If the integers taken are relatively prime (that is they share no common divisors except for 1), then the invariant probability measure is independent of the value of the integer multiplying s.*

We note that the (Shannon) capacity of such an erasure channel is given by $\log_2(K + 1)p$ [12]. From (6.18)–(6.19), the following is obtained.

Theorem 6.22 *If $\log_2(K)p > \log_2(|a|)$, then α, δ exist such that Theorem 6.21 is satisfied.*

Remark 6.7 Thus, the Shannon capacity of the erasure channel is an almost sufficient condition for the positive Harris recurrence of the state and the quantizer process. We will see that under a more generalized interpretation of stationarity, this result applies to a large class of memoryless channels and a class of channels with memory as to be seen later in this chapter (see Theorem 6.27): There is a direct relationship between the existence of a stationary measure and the Shannon capacity of the channel used in the system.

Under slightly stronger conditions we obtain a finite second moment:

Theorem 6.23 [64] *Suppose that the assumptions of Theorem 6.21 hold, and in addition the following bound holds*:

$$a^2\left(1 - p + \frac{p}{(2^R - 1)^2}\right) < 1. \tag{6.23}$$

Then, for each initial condition, $\lim_{t\to\infty} E[x_t^2] = E_\pi[x_0^2] < \infty$.

Remark 6.8 We note from Minero et al. [36] that a necessary condition for mean square stability is $a^2(1 - p + \frac{p}{(2^R)^2}) < 1$. Thus, the sufficiency condition in Theorem 6.23 almost meets this bound except for the additional symbol sent for the under-zoom events. We note that the average rates can be made arbitrarily close to zero by sampling the control system with larger periods. Such a relaxation of the sampling period, however, would lead to a process which is not Markov, yet n-ergodic, quadratically stable, and asymptotic mean stationary (AMS).

6.3.3.1 Connections with Random-Time Drift Criteria

We point out the connection of the results above with random-time drift criteria in Theorem 6.20.

By Lemma 6.3, the process (x_t, Δ_t) forms a Markov chain. Now, in the model considered, the controller can receive meaningful information regarding the state of the system when two events occur concurrently: the channel carries information with no error, and the source lies in the granular region of the quantizer, that is, $x_t \in [-\frac{1}{2}K\Delta_t, \frac{1}{2}K\Delta_t)$ and $\Upsilon_t = 1$. The times at which both of these events occur form an increasing sequence of random stopping times, defined as

$$\mathcal{T}_0 = 0, \qquad \mathcal{T}_{z+1} = \inf\bigl\{k > \mathcal{T}_z : |h_k| \le 1, \Upsilon_k = 1\bigr\}, \quad z \in \mathbb{N}.$$

We can apply Theorem 6.20 for these stopping times. These are the times when information reaches the controller regarding the value of the state when the state is in the granular region of the quantizer. The following lemma is key:

Lemma 6.4 *The discrete probability measure* $P(\mathcal{T}_{z+1} - \mathcal{T}_z = k \mid x_{\mathcal{T}_z}, \Delta_{\mathcal{T}_z})$ *has the upper bound*

$$P(\mathcal{T}_{z+1} - \mathcal{T}_z \ge k \mid x_{\mathcal{T}_z}, \Delta_{\mathcal{T}_z}) \le (1 - p)^{k-1} + G_k(\Delta_{\mathcal{T}_z}),$$

where $G_k(\Delta_{\mathcal{T}_z}) \to 0$ *as* $\Delta_{\mathcal{T}_z} \to \infty$ *uniformly in* $x_{\mathcal{T}_z}$.

In view of Lemma 6.20, first without an irreducibility assumption, we can establish recurrence of the set $C_x \times C_h$ by defining a Lyapunov function of the form $V(x_t, \Delta_t) = \frac{1}{2}\log_2(\Delta^2) + B_0$ for some $B_0 > 0$. One can establish the irreducibility of the Markov chain by imposing a countability condition on the set of admissible bin sizes. A similar discussion, with a quadratic Lyapunov function, applies for finite moment analysis.

6.3.3.2 The Multi-dimensional Case

The result for the scalar problem has a natural counterpart in the multi-dimensional setting. Consider the linear system described by

$$x_{t+1} = Ax_t + Bu_t + Gw_t, \tag{6.24}$$

where $x_t \in \mathbb{R}^N$ is the state at time t, $u_t \in \mathbb{R}^m$ is the control input, and $\{w_t\}$ is a sequence of zero-mean i.i.d. $\mathbb{R}^d$-valued Gaussian random vectors. Here A is the square system matrix with at least one eigenvalue greater than or equal to 1 in magnitude, that is, the system is open-loop unstable. Furthermore, (A, B) and (A, G) are controllable pairs. We also assume at this point that the eigenvalues are real, even though the extension to the complex case is primarily technical. Without any loss of generality, we assume A to be in Jordan form. Because of this, we allow w_t to have correlated components, that is, the correlation matrix $E[w_t w_t^T]$ is not necessarily diagonal. We also assume that B is invertible (if B is not invertible, a sampled system can be made to have an invertible control matrix, with a periodic scheme with period at most n).

We restrict the analysis to noiseless channel in this section. The scheme proposed in the previous section is also applicable to the multi-dimensional setup. Stabilizability for the diagonalizable case immediately follows from the discussion for scalar systems, since the analysis for the scalar case is applicable to each of the subsystems along each of the eigenvectors. The possibly correlated noise components will lead to the recurrence analysis discussed earlier. For such a setup, the stopping times can be arranged to be identical for each modes, for the case when the quantizer captures all the state components. Once this is satisfied, the drift conditions will be obtained. The non-diagonalizable Jordan case, however, is more involved, as we discuss now.

Consider the following system:

$$\begin{bmatrix} x_{t+1}^1 \\ x_{t+1}^2 \end{bmatrix} = \begin{bmatrix} \lambda & 1 \\ 0 & \lambda \end{bmatrix} \begin{bmatrix} x_t^1 \\ x_t^2 \end{bmatrix} + B \begin{bmatrix} u_t^1 \\ u_t^2 \end{bmatrix} + \begin{bmatrix} w_t^1 \\ w_t^2 \end{bmatrix}. \tag{6.25}$$

The approach entails quantizing the components in the system according to the adaptive quantization rule provided earlier for scalar systems: For $i = 1, 2$, let $R' = R_i' = \log_2(2^{R_i} - 1) = \log_2(K_i)$ (that is, the same rate is used for quantizing the components with the same eigenvalue). For $t \geq 0$ and with $\Delta_0^1, \Delta_0^2 \in \mathbb{R}$, con-

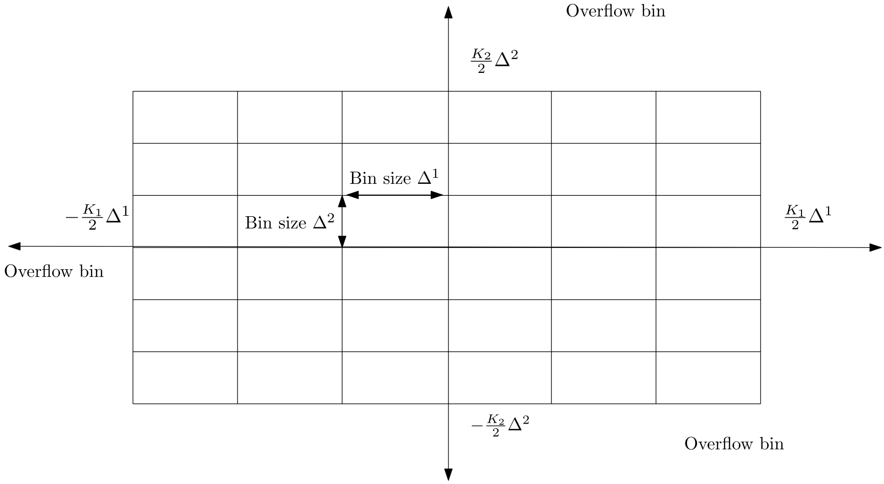

Fig. 6.3 A uniform vector quantizer. There is a single overflow bin

sider:

$$u_t = -B^{-1}A\hat{x}_t,$$

$$\begin{bmatrix} \hat{x}_t^1 \\ \hat{x}_t^2 \end{bmatrix} = \begin{bmatrix} Q_{K_1}^{\Delta_t^1}(x_t^1) \\ Q_{K_2}^{\Delta_t^2}(x_t^2) \end{bmatrix}, \tag{6.26}$$

$$\Delta_{t+1}^1 = \Delta_t^1 \bar{Q}\left(\left|h_t^1\right|, \left|h_t^2\right|, \Delta_t^1\right), \qquad \Delta_{t+1}^2 = \Delta_t^2 \bar{Q}\left(\left|h_t^1\right|, \left|h_t^2\right|, \Delta_t^2\right), \tag{6.27}$$

with, for $i = 1, 2$, $\delta^i, \varepsilon^i, \eta^i > 0$, $\eta^i < \varepsilon^i$ and $L^i > 0$ such that

$$\bar{Q}(x, y, \Delta) = |\lambda| + \delta^i \quad \text{if } |x| > 1, \text{ or } |y| > 1,$$

$$\bar{Q}(x, y, \Delta) = \frac{|\lambda|}{2^{R_i'} - \eta^i} \quad \text{if } 0 \le |x| \le 1, \ |y| \le 1, \ \Delta^i > L^i,$$

$$\bar{Q}(x, y, \Delta) = 1 \quad \text{if } 0 \le |x| \le 1, \ |y| \le 1, \ \Delta^i \le L^i.$$

Note that the above imply that $\Delta_t^i \ge L^i \frac{|\lambda|}{2^{R_i'} - \eta^i} =: L'^i$. We also assume that for some sufficiently large η_Δ, $\Delta_0^1 = \eta_\Delta \Delta_0^2$, which leads to the result that $\Delta_t^1 = \eta_\Delta \Delta_t^2$ for all $t \ge 0$. See Fig. 6.3 for a depiction of the quantizer used at a particular time. The sequence of stopping times is now defined as follows:

$$\mathcal{T}_0 = 0, \qquad \mathcal{T}_{z+1} = \inf\left\{k > \mathcal{T}_z : \left|h_k^i\right| \le 1, i \in \{1, 2, \ldots, n\}\right\}, \quad z \in \mathbb{Z}_+,$$

where $h_k^i = \frac{x_t^i}{\Delta_t^i 2^{R_i'-1}}$. Here Δ^i is the bin size of the quantizer in the direction of the eigenvector x^i, with rate R_i'.

With this approach, the drift criterion applies almost identically as it does for the scalar case.

Theorem 6.24 [21, 60] *Consider the multi-dimensional system* (6.24). *If the system is controlled over a discrete-noiseless channel with capacity*

$$C > \sum_{|\lambda_i|>1} \log_2\big(|\lambda_i|\big),$$

there exists a stabilizing scheme leading to a Markov chain with a bounded second moment in the sense that $\limsup_{t\to\infty} E[|x_t|_2^2] < \infty$.

Extensions of such settings also apply to systems with decentralized multiple sensors. We refer the reader to [22] and [60].

6.3.4 Stochastic Stabilization over Noisy Channels with Noiseless Feedback

In this subsection, we consider discrete noisy channels with noiseless feedback. We first investigate Discrete Memoryless Channels (DMCs).

6.3.4.1 Asymptotic Mean Stationarity and *n*-Ergodicity

The condition $C \geq \log_2(|a|)$ in Theorem 6.18 is almost sufficient for establishing ergodicity and stability, as captured by the following discussion.

Consider the following update algorithm. Let n be a given block length. Consider a class of uniform quantizers, defined by two parameters, with bin size $\Delta > 0$, and an even number $K(n) \geq 2$ (see Fig. 6.1). Define the uniform quantizer as follows: For $k = 1, 2 \ldots, K(n)$,

$$Q_{K(n)}^{\Delta}(x) = \begin{cases} (k - \frac{1}{2}(K(n)+1))\Delta & \text{if } x \in [(k-1-\frac{1}{2}K(n))\Delta, (k-\frac{1}{2}K(n))\Delta), \\ (\frac{1}{2}(K(n)-1))\Delta & \text{if } x = \frac{1}{2}K(n)\Delta, \\ \mathcal{Z} & \text{if } x \notin [-\frac{1}{2}K(n)\Delta, \frac{1}{2}K(n)\Delta], \end{cases}$$

where $\mathcal{Z}$ is the overflow symbol in the quantizer. Let $\{x : Q_{K(n)}^{\Delta}(x) \neq \mathcal{Z}\}$ be the *granular region* of the quantizer.

At every sampling instant $t = kn$, $k = 0, 1, 2, \ldots$, the source coder $\mathcal{E}_t^s$ quantizes output symbols in $\mathbb{R} \cup \{\mathcal{Z}\}$ to a set $\mathcal{M}(n) = \{1, 2, \ldots, K(n)+1\}$. A channel encoder $\mathcal{E}_t^c$ maps the elements in $\mathcal{M}(n)$ to corresponding channel inputs $q_{[kn,(k+1)n-1]} \in$

$\mathcal{M}^n$. For each time $t = kn - 1, k = 1, 2, 3, \ldots$, the channel decoder applies a mapping $\mathcal{D}_{tn} : \mathcal{M}'^n \to \mathcal{M}(n)$ such that

$$c'_{(k+1)n-1} = \mathcal{D}_{kn}\big(q'_{[kn,(k+1)n-1]}\big).$$

Finally, the controller runs an estimator

$$\hat{x}_{kn} = \big(\mathcal{E}^s_{kn}\big)^{-1}\big(c'_{(k+1)n-1}\big) \times 1_{\{c'_{(k+1)n-1} \neq \mathcal{Z}\}} + 0 \times 1_{\{c'_{(k+1)n-1} = \mathcal{Z}\}}.$$

Hence, when the decoder output is the overflow symbol, the estimation output is 0.

As in the previous two chapters, at time kn the bin size Δ_{kn} is taken to be a function of the previous state $\Delta_{(k-1)n}$ and the past n channel outputs. Further, the encoder has access to the previous channel outputs, thus making such a quantizer implementable at both the encoder and the decoder.

With $K(n) > \lceil |a|^n \rceil$, $R = \log_2(K(n) + 1)$, let us introduce $R'(n) = \log_2(K(n))$ and let

$$R'(n) > n \log_2\left(\frac{|a|}{\alpha}\right),$$

for some $\alpha, 0 < \alpha < 1$ and $\delta > 0$. When clear from the context, we will drop the index n in $R'(n)$. We will consider the following update rules in the controller actions and the quantizers. For $t \geq 0$ and with $\Delta_0 > L$ for some $L \in \mathbb{R}_+$, and $\hat{x}_0 \in \mathbb{R}$, consider, for $t = kn$, $k \in \mathbb{N}$,

$$\begin{aligned} u_t &= -1_{\{t=(k+1)n-1\}} \frac{a^n}{b} \hat{x}_{kn}, \\ \Delta_{(k+1)n} &= \Delta_{kn} \bar{Q}\big(\Delta_{kn}, c'_{(k+1)n-1}\big), \end{aligned} \tag{6.28}$$

where c' denotes the decoder output variable. If we use $\delta > 0$ and $L > 0$ such that

$$\begin{aligned} \bar{Q}\big(\Delta, c'\big) &= (|a| + \delta)^n \quad \text{if } c' = \mathcal{Z}, \\ \bar{Q}\big(\Delta, c'\big) &= \alpha^n \quad \text{if } c' \neq \mathcal{Z},\ \Delta \geq L, \\ \bar{Q}\big(\Delta, c'\big) &= 1 \quad \text{if } c' \neq \mathcal{Z},\ \Delta < L, \end{aligned} \tag{6.29}$$

we can show that a recurrent set exists. Note that the above implies that $\Delta_t \geq L\alpha^n =: L'$ for all $t \geq 0$.

Thus, we have three main events: When the decoder output is the overflow symbol, the quantizer is zoomed out (with a coefficient of $(|a| + \delta)^n$). When the decoder output is not the overflow symbol $\mathcal{Z}$, the quantizer is zoomed in (with a coefficient of α^n) if the current bin size is greater than or equal to L, and otherwise the bin size does not change.

In the following, we make the quantizer bin size process space countable and as a result establish the irreducibility of the sampled process (x_{tn}, Δ_{tn}).

Theorem 6.25 [56] *For the existence of a compact coordinate recurrent set, the following is sufficient: The channel capacity C satisfies: $C > \log_2(|a|)$.*

Theorem 6.26 *For an adaptive quantizer satisfying the conditions of Theorem 6.25, suppose that the quantizer bin sizes are such that their logarithms are integer multiples of some scalar s, and $\log_2(\bar{Q}(\cdot))$ takes values in integer multiples of s. Suppose the integers taken are relatively prime (that is they share no common divisors except for 1). Then the sampled process (x_{tn}, Δ_{tn}) forms a positive Harris recurrent Markov chain at sampling times on the space of admissible quantizer bins and state values.*

Theorem 6.27 [56] *Under the conditions of Theorems 6.25 and 6.26, the process $\{x_t, \Delta_t\}$ is n-stationary, n-ergodic, and hence asymptotically mean stationary (AMS).*

Proof Sketch The proof follows from the observation that a positive Harris recurrent Markov chain is recurrent and stationary and that if a sampled process is a positive Harris recurrent Markov chain, and if the intersampling time is fixed, with a time-homogeneous update in the inter-sampling times, then the process is mixing, n-ergodic and n-stationary. □

Remark 6.9 Converse Results for Quadratic Stability For quadratic stability, that is, the condition that $\lim_{T\to\infty} \frac{1}{T}\sum_{t=0}^{T-1} |x_t|^2$, exists and is finite almost surely; more restrictive conditions are needed and Shannon capacity is not sufficient (see [43] and [56]). We note that for erasure channels and noiseless channels, one can obtain tight converse theorems using Theorem 6.18 (see [36] and [64]). For general DMCs, however, a tight converse result on quadratic stabilizability is not yet available. One reason for this is that the error exponents of fixed length block codes with noiseless feedback for general DMCs are not currently known. It is worth noting that the error exponent of DMCs is typically improved with feedback, unlike the capacity of DMCs. Some partial results have been reported in [16] (e.g., the sphere packing upper bound is tight for a class of symmetric channels for rates above a critical rate even with feedback). Related references addressing partial results include [33] and [34] which consider lower bounds on estimation error moments for transmission of a single variable over a noisy channel (in the context of this chapter, this single variable may correspond to the initial state x_0). A further related notion for quadratic stability is the notion of *any-time capacity* introduced by Sahai and Mitter (see [42] and [43]). Further discussion on this topic is available in [60] and [59].

6.3.5 Channels with Memory and Noiseless Feedback

Definition 6.7 Channels are said to be of *Class A* type, if

- They satisfy the following Markov chain condition:

$$q'_t \leftrightarrow q_t, q_{[0,t-1]}, q'_{[0,t-1]} \leftrightarrow \{x_0, w_t, t \geq 0\},$$

  for all $t \geq 0$, and

- Their capacity with feedback is given by

$$C = \lim_{T\to\infty} \max_{\{P(q_t|q_{[0,t-1]},q'_{[0,t-1]}),\quad 0\le t\le T-1\}} \frac{1}{T} I\big(q_{[0,T-1]} \to q'_{[0,T-1]}\big),$$

where the directed mutual information is defined by

$$I\big(q_{[0,T-1]} \to q'_{[0,T-1]}\big) = \sum_{t=1}^{T-1} I\big(q_{[0,t]}; q'_t|q'_{[0,t-1]}\big) + I\big(q_0; q'_0\big).$$

DMCs naturally belong to this class. For DMCs, feedback does not increase the capacity [12]. Such a class also includes finite state stationary Markov channels which are indecomposable [39], and non-Markov channels which satisfy certain symmetry properties [44]. Further examples can be found in [47] and in [14].

Theorem 6.28 [56] *Suppose that a linear plant given by* (6.13) *is controlled over a Class A type noisy channel with feedback. If the channel capacity* (*with feedback*) *is less than* $\log_2(|a|)$, *then* (*i*) *the following condition*

$$\liminf_{T\to\infty} \frac{1}{T} h(x_T) \le 0,$$

cannot be satisfied under any policy, and (*ii*) *the state process cannot be AMS under any policy.*

Remark 6.10 The result above is *negative*, but one can also obtain a positive result: If the channel capacity is greater than $\log_2(|a|)$ and there is a positive error exponent (uniform over all transmitted messages, as in Theorem 14 of [39]), then there exists a coding scheme leading to an AMS state process provided that the channel restarts itself with the transmission of every new block (either independently or as a Markov process). We also note that if the channel is not information stable, then information spectrum methods lead to pessimistic realizations of capacity (known as the lim inf *in probability* of the normalized information density, see [47, 50]).

6.3.6 Higher-Order Plants

The result for the scalar problem has a natural counterpart in the multi-dimensional setting. Consider the linear system described by (6.24). In the following, we assume that all eigenvalues $\{\lambda_i, 1 \le i \le N\}$ of A are unstable, that is, have magnitudes greater than or equal to 1. There is no loss here since if some eigenvalues are stable, by a similarity transformation, the unstable modes can be decoupled from the stable ones and one can instead consider a lower dimensional system; stable modes are already recurrent.

Theorem 6.29 [60] *For such a system controlled over a Class A type noisy channel with feedback, if the channel capacity (with feedback) satisfies*

$$C < \sum_i \log_2(|\lambda_i|),$$

there does not exist a stabilizing coding and control scheme with the property

$$\liminf_{T\to\infty} \frac{1}{T} h(x_T) \leq 0.$$

Proposition 6.5 [60] *For such a system controlled over a Class A type noisy channel with feedback, if*

$$C < \log_2(|A|),$$

then

$$\limsup_{T\to\infty} P(|x_T| \leq b(T)) \leq \frac{C}{\log_2(|A|)} > 0,$$

for all $b(T) > 0$ *such that* $\lim_{T\to\infty} \frac{1}{T} \log_2(b(T)) = 0$.

With this lemma, we state the following.

Theorem 6.30 [60] *Consider such a system controlled over a Class A type noisy channel with feedback. If there exists some encoding and controller policy so that the state process is AMS, then the channel capacity (with feedback) C must satisfy*

$$C \geq \log_2(|A|).$$

For sufficiency, we will assume that A is a diagonalizable matrix (a sufficient condition for which is that its eigenvalues are distinct real).

Theorem 6.31 [56] *Consider a multi-dimensional system with a diagonalizable matrix A. If the Shannon capacity of the DMC used in the controlled system satisfies*

$$C > \sum_{|\lambda_i|>1} \log_2(|\lambda_i|),$$

there exists a stabilizing scheme in the AMS sense.

On achievability of AMS stabilization over channels with memory, the discussions in Remark 6.10 also apply for this setting.

Remark 6.11 Theorem 6.31 can be extended to the case where the matrix A is not diagonalizable, in the same spirit as in Theorem 6.24, by constructing stopping times in view of the coupling between modes sharing a common eigenvalue [60].

6.4 Conclusion

In this chapter, we considered the optimization of information channels in networked control systems. We made the observation that quantizers can be viewed as a special class of channels and established existence results for optimal quantization and coding policies. Comparison of information channels for optimization has been presented. On stabilization, the relation between ergodicity and Shannon capacity has been discussed.

The value of information channels in optimization and control problems require further analysis. Particularly, further research from the information theory community for optimal non-asymptotic or finite delay coding will lead to useful applications in networked control. Error exponents with fixed block-length and feedback is currently an unresolved problem, which may lead to converse theorems for quadratic stabilization over noisy communication channels.

Acknowledgements This research has been supported by the Natural Sciences and Engineering Research Council of Canada (NSERC). Some of the figures in this chapter have appeared in [64] and [60].

The author is grateful to Giacomo Como, Bo Bernhardsson, and Anders Rantzer for hosting the workshop that took place in Lund University, which led to the publication of this book chapter. Some of the research reported in this chapter are results of the author's collaborations with Tamer Başar, Tamás Linder, Sean Meyn, and Andrew Johnston. Discussions with Giacomo Como, Aditya Mahajan, Nuno Martins, Maxim Raginsky, Anant Sahai, Naci Saldi, Sekhar Tatikonda, Demos Teneketzis, and Tsachy Weissman are gratefully acknowledged.

This research was supported by LCCC—Linnaeus Grant VR 2007-8646, Swedish Research Council.

References

1. Abaya, E.F., Wise, G.L.: Convergence of vector quantizers with applications to optimal quantization. SIAM J. Appl. Math. **44**, 183–189 (1984)
2. Bao, L., Skoglund, M., Johansson, K.H.: Iterative encoder-controller design for feedback control over noisy channels. IEEE Trans. Autom. Control **57**, 265–278 (2011)
3. Bar-Shalom, Y., Tse, E.: Dual effect, certainty equivalence and separation in stochastic control. IEEE Trans. Autom. Control **19**, 494–500 (1974)
4. Blackwell, D.: Equivalent comparison of experiments. Ann. Math. Stat. **24**, 265–272 (1953)
5. Bogachev, V.I.: Measure Theory. Springer, Berlin (2007)
6. Boll, C.: Comparison of experiments in the infinite case. PhD Dissertation, Stanford University (1955)
7. Borkar, V.S., Mitter, S.K.: LQG control with communication constraints. In: Kailath Festschrift. Kluwer Academic, Boston (1997)
8. Borkar, V.S., Mitter, S.K., Tatikonda, S.: Optimal sequential vector quantization of Markov sources. SIAM J. Control Optim. **40**, 135–148 (2001)
9. Borkar, V., Mitter, S., Sahai, A., Tatikonda, S.: Sequential source coding: an optimization viewpoint. In: Proceedings of the IEEE Conference on Decision and Control, pp. 1035–1042 (2005)
10. Cam, L.L.: Comparison of experiments—a short review. In: Ferguson, T., Shapley, L. (eds.) Statistics, Probability and Game Theory Papers in Honor of David Blackwell. IMS Lecture Notes Monograph Ser. (1996)

11. Como, G., Fagnani, F., Zampieri, S.: Anytime reliable transmission of real-valued information through digital noisy channels. SIAM J. Control Optim. **48**, 3903–3924 (2010)
12. Cover, T.M., Thomas, J.A.: Elements of Information Theory. Wiley, New York (1991)
13. Curry, R.E.: Estimation and Control with Quantized Measurements. MIT Press, Cambridge (1969)
14. Dabora, R., Goldsmith, A.: On the capacity of indecomposable finite-state channels with feedback. In: Proceedings of the Allerton Conf. Commun. Control Comput., pp. 1045–1052 (2008)
15. Devroye, L., Györfi, L.: Non-parametric Density Estimation: The L_1 View. Wiley, New York (1985)
16. Dobrushin, R.L.: An asymptotic bound for the probability error of information transmission through a channel without memory using the feedback. Probl. Kibern. **8**, 161–168 (1962)
17. Fischer, T.R.: Optimal quantized control. IEEE Trans. Autom. Control **27**, 996–998 (1982)
18. Fu, M.: Lack of separation principle for quantized linear quadratic Gaussian control. IEEE Trans. Autom. Control **57**, 2385–2390 (2012)
19. Gray, R.M.: Entropy and Information Theory. Springer, New York (1990)
20. György, A., Linder, T.: Codecell convexity in optimal entropy-constrained vector quantization. IEEE Trans. Inf. Theory **49**, 1821–1828 (2003)
21. Johnston, A., Yüksel, S.: Stochastic stabilization of partially observed and multi-sensor systems driven by Gaussian noise under fixed-rate information constraints. IEEE Trans. Autom. Control (2012, under review). arXiv:1209.4365
22. Johnston, A., Yüksel, S.: Stochastic stabilization of partially observed and multi-sensor systems driven by Gaussian noise under fixed-rate information constraints. In: Proceedings of the IEEE Conference on Decision and Control, Hawaii (2012)
23. Lewis, J.B., Tou, J.T.: Optimum sampled-data systems with quantized control signals. IEEE Trans. Ind. Appl. **82**, 229–233 (1965)
24. Luenberger, D.: Linear and Nonlinear Programming. Addison–Wesley, Reading (1984)
25. Mahajan, A., Teneketzis, D.: On the design of globally optimal communication strategies for real-time noisy communication with noisy feedback. IEEE J. Sel. Areas Commun. **26**, 580–595 (2008)
26. Mahajan, A., Teneketzis, D.: Optimal design of sequential real-time communication systems. IEEE Trans. Inf. Theory **55**, 5317–5338 (2009)
27. Mahajan, A., Teneketzis, D.: Optimal performance of networked control systems with non-classical information structures. SIAM J. Control Optim. **48**, 1377–1404 (2009)
28. Martins, N.C., Dahleh, M.A.: Feedback control in the presence of noisy channels: "Bode-like" fundamental limitations of performance. IEEE Trans. Autom. Control **53**, 1604–1615 (2008)
29. Martins, N.C., Dahleh, M.A., Elia, N.: Feedback stabilization of uncertain systems in the presence of a direct link. IEEE Trans. Autom. Control **51**(3), 438–447 (2006)
30. Matveev, A.S.: State estimation via limited capacity noisy communication channels. Math. Control Signals Syst. **20**, 1–35 (2008)
31. Matveev, A.S., Savkin, A.V.: The problem of LQG optimal control via a limited capacity communication channel. Syst. Control Lett. **53**, 51–64 (2004)
32. Matveev, A.S., Savkin, A.V.: Estimation and Control over Communication Networks. Birkhäuser, Boston (2008)
33. Merhav, N.: On optimum parameter modulation-estimation from a large deviations perspective. IEEE Trans. Inf. Theory **58**, 172–182 (2012)
34. Merhav, N.: Exponential error bounds on parameter modulation-estimation for discrete memoryless channels. IEEE Trans. Inf. Theory (under review). arXiv:1212.4649
35. Meyn, S.P., Tweedie, R.: Markov Chains and Stochastic Stability. Springer, London (1993)
36. Minero, P., Franceschetti, M., Dey, S., Nair, G.N.: Data rate theorem for stabilization over time-varying feedback channels. IEEE Trans. Autom. Control **54**(2), 243–255 (2009)
37. Nair, G.N., Evans, R.J.: Stabilizability of stochastic linear systems with finite feedback data rates. SIAM J. Control Optim. **43**, 413–436 (2004)

38. Nair, G.N., Fagnani, F., Zampieri, S., Evans, J.R.: Feedback control under data constraints: an overview. In: Proceedings of the IEEE, pp. 108–137 (2007)
39. Permuter, H.H., Weissman, T., Goldsmith, A.J.: Finite state channels with time-invariant deterministic feedback. IEEE Trans. Inf. Theory **55**(2), 644–662 (2009)
40. Phelps, R.: Lectures on Choquet's Theorem. Van Nostrand, New York (1966)
41. Pollard, D.: Quantization and the method of k-means. IEEE Trans. Inf. Theory **28**, 199–205 (1982)
42. Sahai, A.: Anytime Information Theory. Ph.D. dissertation, Massachusetts Institute of Technology, Cambridge, MA (2001)
43. Sahai, A., Mitter, S.: The necessity and sufficiency of anytime capacity for stabilization of a linear system over a noisy communication link—part I: scalar systems. IEEE Trans. Inf. Theory **52**(8), 3369–3395 (2006)
44. Şen, N., Alajaji, F., Yüksel, S.: Feedback capacity of a class of symmetric finite-state Markov channels. IEEE Trans. Inf. Theory **56**, 4110–4122 (2011)
45. Silva, E.I., Derpich, M.S., Østergaard, J.: A framework for control system design subject to average data-rate constraints. IEEE Trans. Autom. Control **56**, 1886–1899 (2011)
46. Tatikonda, S., Mitter, S.: Control under communication constraints. IEEE Trans. Autom. Control **49**(7), 1056–1068 (2004)
47. Tatikonda, S., Mitter, S.: The capacity of channels with feedback. IEEE Trans. Inf. Theory **55**(1), 323–349 (2009)
48. Tatikonda, S., Sahai, A., Mitter, S.: Stochastic linear control over a communication channels. IEEE Trans. Autom. Control **49**, 1549–1561 (2004)
49. Teneketzis, D.: On the structure of optimal real-time encoders and decoders in noisy communication. IEEE Trans. Inf. Theory **52**, 4017–4035 (2006)
50. Verdú, S., Han, T.S.: A general formula for channel capacity. IEEE Trans. Inf. Theory **40**, 1147–1157 (1994)
51. Walrand, J.C., Varaiya, P.: Causal coding and control of Markov chains. Syst. Control Lett. **3**, 189–192 (1983)
52. Walrand, J.C., Varaiya, P.: Optimal causal coding-decoding problems. IEEE Trans. Inf. Theory **19**, 814–820 (1983)
53. Witsenhausen, H.S.: On the structure of real-time source coders. Bell Syst. Tech. J. **58**, 1437–1451 (1979)
54. Wong, W.S., Brockett, R.W.: Systems with finite communication bandwidth constraints—part ii: stabilization with limited information feedback. IEEE Trans. Autom. Control **42**, 1294–1299 (1997)
55. Yüksel, S.: Stochastic stabilization of noisy linear systems with fixed-rate limited feedback. IEEE Trans. Autom. Control **55**, 2847–2853 (2010)
56. Yüksel, S.: Characterization of information channels for asymptotic mean stationarity and stochastic stability of non-stationary/unstable linear systems. IEEE Trans. Inf. Theory **58**, 6332–6354 (2012)
57. Yüksel, S.: Jointly optimal LQG quantization and control policies for multi-dimensional linear Gaussian sources. In: Proceedings of the Annual Allerton Conference on Communications, Control and Computing, Monticello, IL (2012)
58. Yüksel, S.: On optimal causal coding of partially observed Markov sources in single and multi-terminal settings. IEEE Trans. Inf. Theory **59**, 424–437 (2013)
59. Yüksel, S., Başar, T.: Control over noisy forward and reverse channels. IEEE Trans. Autom. Control **56**, 1014–1029 (2011)
60. Yüksel, S., Başar, T.: Stochastic Networked Control Systems: Stabilization and Optimization Under Information Constraints. Birkhäuser, Boston (2013)
61. Yüksel, S., Linder, T.: Optimization and Convergence of Observation Channels in Stochastic Control, pp. 637–642. American Control Conference, San Francisco (2011)
62. Yüksel, S., Linder, T.: On optimal zero-delay quantization of vector Markov sources. In: Proceedings of the IEEE Conference on Decision and Control, Hawaii (2012)

63. Yüksel, S., Linder, T.: Optimization and convergence of observation channels in stochastic control. SIAM J. Control Optim. **50**, 864–887 (2012)
64. Yüksel, S., Meyn, S.P.: Random-time, state-dependent stochastic drift for Markov chains and application to stochastic stabilization over erasure channels. IEEE Trans. Autom. Control **58**, 47–59 (2013)
65. Zaidi, A.A., Oechtering, T.J., Yüksel, S., Skoglund, M.: Stabilization and control over Gaussian networks. In: Como, G., Bernhardsson, B., Rantzer, A. (eds.) Information and Control in Networks. Springer, Berlin (2013)