Asymptotic Properties of

Nonlinear Feedback Control Systems

A Dissertation

Presented to

The Faculty of the Graduate School of Arts and Sciences

Brandeis University

Department of Mathematics

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

by

Daniel Liberzon

February, 1998

This page is a substitute for the real page 2, the signature page. I am inserting it so that subsequent page numbers will appear on the correct side in the top margin.

ACKNOWLEDGEMENTS

First of all, I would like to express my deep gratitude to my advisor, Professor Roger W. Brockett of Harvard University. His boundless knowledge has made it possible for me to write this thesis, and his unfailing enthusiasm has made it impossible not to write it. I consider myself extremely lucky to have had the opportunity to work with him.

I wish to acknowledge the help of two other people that has directly contributed to this dissertation. Mark Adler has always been willing to share his insights with me, and I am indebted to him for many illuminating discussions of various areas of mathematics related to this thesis. Susan Parker has provided helpful comments on my writing, and has communicated to me her excitement about teaching, which has been of great value in getting through my moments of frustration.

My sincere thanks go to everyone at the Mathematics Department of Brandeis University for creating such a comfortable and stimulating atmosphere. I wish all of you the best of luck in surviving the difficult times.

I dedicate this thesis to my family, who started me out on this journey and have always been supportive.

ABSTRACT

Asymptotic Properties of Nonlinear Feedback Control Systems

A dissertation presented to the Faculty of

the Graduate School of Arts and Sciences of Brandeis

University, Waltham, Massachusetts

by Daniel Liberzon

We study asymptotic behaviour of nonlinear feedback control systems, both deterministic and stochastic. Of particular interest is the case of quantized feedback, i.e., when the nonlinearity takes the form of a specific piecewise constant function. In the context of deterministic linear control systems with quantized measurements, we show how quantized feedback can be used to asymptotically stabilize the system (Chapter II).

For systems perturbed by white noise, we address the question of existence of steady-state probability distributions. In the linear case, the solution to the

Fokker-Planck equation which describes the evolution of the probability density is well known. In particular, one has an expression for the steady-state probability density, which is an eigenfunction of the Fokker-Planck operator with eigenvalue zero. We show that other eigenvalues and eigenfunctions of the Fokker-Planck operator associated with a linear system can also be directly computed (Chapter III). In the nonlinear case, the situation is more complicated. We describe a class of nonlinear feedback systems for which explicit formulae for the steady-state probability densities can be found, and give two interpretations of this result, one related to certain concepts from statistical thermodynamics, and the other related to Lur'e problem of absolute stability (Chapter IV). We demonstrate how the solutions obtained here can be used to treat a number of stochastic optimal control problems (Chapter V).

CONTENTS

Chapter I

BACKGROUND

1. Stability and control of dynamical systems. This section is an introduction to those basic notions and results from the theory of stability and control of dynamical systems which will be used explicitly in the sequel. General concepts from the theory of ordinary differential equations, such as the standard existence and uniqueness results, are not discussed here, the reader's familiarity with them being assumed.

1.1. Liapunov functions and stability. Consider a general system of differential equations

$$\dot{x} = f(x) \tag{I.1}$$

where $x \in \mathbb{R}^n$ (this is a shorthand notation for $\dot{x}_i = f_i(x_1, \ldots, x_n),\ i = 1, \ldots, n$). Let us assume that the origin is an equilibrium: $f(0) = 0$. Let $S(R)$ be the spherical region of radius $R > 0$ about the origin: $S(R) = \{x \in \mathbb{R}^n : \|x\| \leq R\}$. Following LaSalle [30], we will say that the origin is a *stable* equilibrium for the system (I.1) if for every R there exists an $r > 0$ such that any solution starting in $S(r)$ remains in $S(R)$ for all future time. If, in addition, there exists an $R_0 > 0$ such that every solution starting in $S(R_0)$ approaches 0 as $t \to \infty$, the origin is called *asymptotically stable*. Moreover, if *all* solutions tend to 0 as $t \to \infty$, we will say that the origin is *globally asymptotically stable.*

There are many stability theorems within the framework of what is called Liapunov's second method. For our purposes the following one is adequate.

Theorem 1 ([30, Theorem 5]). *Assume that there exists a function $V \in C^1(\mathbb{R}^n)$ such that $V(0) = 0$, $V(x) > 0$ for all $x \neq 0$, $V(x) \to \infty$ as $\|x\| \to \infty$, and the derivative of V along solutions of (I.1) as given by*

$$\dot{V}(x) = \langle \text{grad} V(x), f(x) \rangle$$

satisfies $\dot{V}(x) \leq 0$ for all x. If $\dot{V}$ is not identically zero along any solution other than the origin, then the origin is a globally asymptotically stable equilibrium for the system (I.1).

The most tractable systems from the stability veiwpoint are linear time-invariant systems of the form

$$\dot{x} = Ax. \tag{I.2}$$

For such systems, the concepts of asymptotic stability and global asymptotic stability coincide, and both amount to the property that the eigenvalues of the matrix A must have negative real parts. It is not hard to construct a *Liapunov function* (i.e., a function satisfying the hypothesis of Theorem 1) in this case. Namely, take a positive definite matrix M, and define the matrix $Q = \int_0^\infty e^{A^T t} M e^{At} dt$. Letting $V(x) = x^T Q x$, we obtain a Liapunov function whose derivative along solutions of (I.2) is given by $\frac{d}{dt} x^T Q x = -x^T M x < 0$ if $x \neq 0$.

1.2. Control systems.

DEFINITION ([10]). Let X be a finite-dimensional manifold, and let $\pi : E \to X$ be a vector bundle over X of rank m. Let TX denote the tangent bundle of X, and let π^*TX denote the pullback of TX over E. Regarding (x, u) as a point in E, we can describe a section of the pullback π^*TX locally as a function $f(x, u)$ which, for each (x, u), specifies a velocity $\dot{x} = f(x, u)$. We will call any such section a *control system.*

Thus finite-dimensional, continuous-time control systems have as their description in local coordinates

$$\dot{x} = f(x, u). \tag{I.3}$$

A basic problem in control theory is that of *controllability.* Namely, given the system (I.3) and an initial state $x_0 = x(t_0)$, one would like to find a control function $u(t)$ that will steer the system to a specified destination point x_1 at a given time $t_1 > t_0$. If this is always possible, the system is called *controllable.* This problem has an explicit solution for linear time-invariant systems of the form

$$\dot{x} = Ax + Bu \tag{I.4}$$

where $x \in \mathbb{R}^n$ (state space) and $u \in \mathbb{R}^m$ (control space). The variation of constants formula gives

$$x(t) = e^{At}x_0 + \int_0^t e^{A(t-\sigma)}Bu(\sigma)d\sigma.$$

This expression may be regarded as defining a functional taking piecewise continuous control functions u into $\mathbb{R}^n$. If this functional is of full rank, then we can steer the system (I.4) from any given point to any other point. The following theorem makes this condition precise.

Theorem 2. *If the $n \times mn$ matrix $(B, AB, \ldots, A^{n-1}B)$ (horizontal arrangement) is of rank n, then the system (I.4) is controllable.*

The above rank condition is known as the *controllability condition.* If it is satisfied, we will say that (A, B) is a *controllable pair.*

Now, consider the system

$$\dot{x} = Ax,$$
$$y = Cx \tag{I.5}$$

where $x \in \mathbb{R}^n$ (state space) and $y \in \mathbb{R}^p$ (output space). If it is possible to determine the initial state $x(0)$ uniquely from the knowledge of the output $y(t)$ over some time interval $[0, t]$, then the system (I.5) is called *observable.* The following result can be considered as dual to Theorem 2.

Theorem 3. *If the $np \times n$ matrix $(C; CA; \ldots; CA^{n-1})$ (vertical arrangement) is of rank n, then the system (I.5) is observable.*

The above results are due to R. E. Kalman and can be found, e.g., in [4].

1.3. Feedback stabilization. Consider the control system (I.3) together with the output equation

$$y = g(x).$$

While the controllability problem consists in driving a particular solution, e.g., to zero, one may be concerned with a more general problem of finding a universal control u that makes the origin globally asymptotically stable, so that all solutions tend to 0 as $t \to \infty$. The usual approach is to design a stabilizing control u that depends not simply on time t, but rather on current measurements of the system's behavior. According as u is to be a function of the state x (*state feedback*) or a function of the output y (*output feedback*), we are led to the problems of *state feedback stabilization* and *output feedback stabilization*, respectively.

Let us consider both problems in the context of linear time-invariant systems. The state feedback stabilization problem for (I.4) is in fact closely related to controllability. Changing the basis, we can rewrite the system equations as

$$\frac{d}{dt}\begin{pmatrix} x_u \\ x_l \end{pmatrix} = \begin{pmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{pmatrix}\begin{pmatrix} x_u \\ x_l \end{pmatrix} + \begin{pmatrix} B_u \\ 0 \end{pmatrix} u$$

where (A_{11}, B_u) is a controllable pair [38]. Clearly, the eigenvalues of A_{22} must have negative real parts if x_l is to go to 0 as $t \to \infty$. On the other hand, one can show that there always exists a matrix K such that $A_{11} - B_u K$ has all its eigenvalues in the open left half-plane. It follows that the linear state feedback control law defined by $u = -Kx$ makes the origin an asymptotically stable equilibrium for (I.4).

The above discussion can be summarized as follows [38].

Theorem 4. *The null solution of (I.4) is stabilizable by linear state feedback if and only if all the modes associated with the eigenvalues with non-negative real parts are controllable.*

The solution to the problem of stabilizing a linear time-invariant system with output feedback is contained in the widely known *Nyquist criterion* (see, e.g., [35]). Consider the single-input, single-output system

$$\begin{aligned} \dot{x} &= Ax + bu, \\ y &= c^T x \end{aligned} \tag{I.6}$$

where $x, b, c \in \mathbb{R}^n$ and A is an $n \times n$ matrix. The *Nyquist locus* is defined to be the image of the imaginary axis $\{s \in \mathbb{C} : \operatorname{Re} s = 0\}$ under the mapping $s \mapsto g(s) = c^T(Is - A)^{-1}b$. Suppose that $g(s)$ has ν poles in the right half-plane $\{s : \operatorname{Re} s > 0\}$ and no poles on the imaginary axis. Choose a number $k \in \mathbb{R}$.

Theorem 5 (Nyquist criterion). *If the Nyquist locus does not intersect the point $-1/k$ and encircles it precisely ν times in the counterclockwise direction, then the closed-loop system obtained by letting $u = -ky$ is asymptotically stable.*

Due to its simple geometrical interpretation, this theorem actually allows one to find all values of k that stabilize (I.6). The number k is usually referred to as the *feedback gain*. A natural, though less constructive, generalization of the Nyquist criterion to the multivariable case was obtained by Brockett and Byrnes in [9].

2. Stochastic differential equations. This section is an introduction to the basic notions and results from the theory of stochastic differential equations. The motivation for studying differential equations with random disturbances comes from the fact that they often provide a more accurate description of a physical process. As a typical example, when modelling the motion of an airplane by such equations, we can take into account the effects of the wind, temperature changes, etc. Stochastic differential equations are also used on Wall Street in models describing random fluctuations of the stock market.

Arguably, the mathematical theory that arises is more technically involved than the one that corresponds to the ordinary differential equations. On the other hand, available methods of analysis are complemented by those of the theory of probability and stochastic processes.

The general approach here is basically adapted from [13]. For a more rigorous treatment, including the standard existence and uniqueness results, see [24], [27].

2.1. Poisson counters. By a *Poisson counter of rate* λ we will understand a stochastic process $N(t)$ which consecutively takes on the integer values $0, 1, 2, \ldots$ according to the probability law

$$\mathcal{P}(N(t) = n) = \frac{(\lambda t)^n}{n!} e^{-\lambda t}.$$

We have $\mathcal{E} N(t) = \lambda t$, so that the average number of jumps of $N(t)$ per unit time equals λ.

We would like to give a precise meaning to the differential equation

$$dx(t) = f(x(t), t)dt + g(x(t), t)dN(t) \tag{I.7}$$

which can also be rewritten in the integral form

$$x(t) = x(0) + \int_0^t f(x(\sigma), \sigma)d\sigma + \int_0^t g(x(\sigma), \sigma)dN(\sigma).$$

We will always assume that the functions f and g satisfy a global Lipschitz condition, namely, for some $K > 0$ and all x_1 and x_2 we have

$$|f(x_2, t) - f(x_1, t)| + |g(x_2, t) - g(x_1, t)| \leq K|x_2 - x_1|.$$

DEFINITION. A function $x(t)$ is a *solution in the Itô sense* of (I.7) if

1) While N is constant, $x(t)$ satisfies the ordinary differential equation $\dot{x}(t) = f(x, t)$.

2) $x(t)$ is continuous from the left.

3) If N jumps at time t_1, then $\lim_{t \to t_1^+} x(t) = \lim_{t \to t_1^-} x(t) + g(\lim_{t \to t_1^-} x(t), t_1)$.

In other words, the stochastic integral is interpreted as the limit of the Riemann sums with $g(x, t)$ evaluated at the left endpoints, so that each jump occurs "in advance" of any change in x (unlike in the ordinary calculus, the exact choice of the subdivision points is important: choosing middle points instead results in a significantly different calculus [37]).

EXAMPLE 1. Let $x(t)$ be the work that remains to be done by a server where customers arrive at rate λ and get served at rate μ. This situation is quite adequately described by the Itô equation

$$dx = -\mu f_+(x)dt + dN$$

where N is a Poisson counter of rate λ and

$$f_+(x) = \begin{cases} 1 & \text{if } x > 0 \\ 0 & \text{if } x \leq 0 \end{cases}$$

2.2. The Wiener process (Brownian motion). Consider the Itô equation

$$dx = \frac{1}{\sqrt{\lambda}}(dN_1 - dN_{-1}),$$

where $x(0) = 0$ and N_1 and N_{-1} are independent Poisson counters of rate $\lambda/2$. Solving this equation and making λ go to $+\infty$, we obtain a process whose moments are those associated with the gaussian density of mean 0 and variance t. In fact, a stochastic process $w(t)$ known as the *standard Wiener process*, or *Brownian motion*, can be defined by the following requirements:

1) $w(0) = 0$

2) $w(t) - w(\tau)$ is a random variable whose probability distribution depends only on $|t-\tau|$, and if, moreover, $[t,\tau] \cap [s,\sigma] = \emptyset$, then $w(\tau) - w(t)$ and $w(\sigma) - w(s)$ are independent (the property of *orthogonal increments*).

3)

$$\mathcal{E}(w(t) - w(\tau))^2 = |t - \tau|. \tag{I.8}$$

4) $w(t)$ is continuous with probability 1.

We will abbreviate $\lim_{\lambda\to+\infty}\frac{1}{\sqrt{\lambda}}(dN_1 - dN_{-1})$ as dw and consider stochastic differential equations of the form

$$dx = f(x)dt + g(x)dw \tag{I.9}$$

or, in a less rigorous but more convenient notation,

$$\dot{x} = f(x) + g(x)\dot{w}.$$

The process $\frac{dw}{dt} = \dot{w}$ formally does not exist, even though it is usually referred to as *white noise*.

If several independent Wiener processes are at hand, we arrive at the equation

$$\dot{x} = f(x) + \sum_{i=1}^{m} g_i(x)\dot{w}_i$$

or, equivalently,

$$\dot{x} = f(x) + G(x)\dot{w}. \tag{I.10}$$

In the most general setting, we may assume that in (I.10) we have x, $f(x) \in \mathbb{R}^n$, $G(x)$ is an $n \times m$ matrix, and $\dot{w}$ is an m-dimensional white noise.

2.3. The Itô rule. Given a twice continuously differentiable function ψ, the equation (I.10) implies

$$\begin{aligned} d\psi = \langle\frac{\partial\psi}{\partial x}, f(x)\rangle dt &+ \sum_{i=1}^{m}\lim_{\lambda\to+\infty}\big[\psi(x + \frac{1}{\sqrt{\lambda}}g_i(x)) - \psi(x)\big]dN_i \\ &+ \sum_{i=1}^{m}\lim_{\lambda\to+\infty}\big[\psi(x - \frac{1}{\sqrt{\lambda}}g_i(x)) - \psi(x)\big]dN_{-i}. \end{aligned}$$

Expanding $\psi(x)$ in a Taylor series and exploring the limit as $\lambda \to +\infty$, K. Itô derived the following formula:

$$d\psi(x) = \langle \frac{\partial \psi}{\partial x}, f(x) \rangle dt + \sum_{i=1}^{m} \langle \frac{\partial \psi}{\partial x}, g_i(x) \rangle dw_i$$
$$+ \frac{1}{2} \sum_{i=1}^{m} \langle g_i(x), \frac{\partial^2 \psi(x)}{\partial x^2} g_i(x) \rangle dt. \tag{I.11}$$

This is known as the *Itô rule for Brownian motion.* The presence of the second-order term in (I.10) is the main feature of the stochastic calculus. It is justified by the fact that, due to the condition (I.8), the term $(dw)^2$ is of the same order as dt.

2.4. Expectations. The rule for computing expectations that goes along with the equation (I.9) is simply that $\mathcal{E} g(x) dw = 0$. This is true because $\mathcal{E}(N_1 - N_{-1}) = 0$, and the solution after a jump is defined in terms of the value before the jump. Thus associated with (I.10) we have the equation for expectations

$$\frac{d}{dt} \mathcal{E} x = \mathcal{E} f(x).$$

Similarly, for (I.11) we have

$$\frac{d}{dt} \mathcal{E} \psi(x) = \mathcal{E} \langle \frac{\partial \psi}{\partial x}, f(x) \rangle + \frac{1}{2} \mathcal{E} \sum_{i=1}^{m} \langle g_i(x), \frac{\partial^2 \psi(x)}{\partial x^2} g_i(x) \rangle. \tag{I.12}$$

2.5. The Fokker-Planck equation. We are interested in obtaining a (deterministic) partial differential equation that describes the evolution of the probability density $\rho(t, x)$ associated with (I.10). Note that for $\dfrac{d}{dt} \mathcal{E} \psi(x)$ we have, on one hand, the formula (I.12) above, and on the other hand,

$$\frac{d}{dt}\mathcal{E}\psi = \frac{d}{dt}\int_{\mathbb{R}^n} \psi(x)\rho(t,x)dx = \int_{\mathbb{R}^n} \psi(x)\frac{\partial \rho(t,x)}{\partial t}dx.$$

Putting these facts together and integrating by parts, we can derive the following partial differential equation:

$$\frac{\partial \rho}{\partial t} = -\langle \frac{\partial}{\partial x}, f(x)\rho\rangle + \frac{1}{2}\sum_{j,k=1}^{n} \frac{\partial^2}{\partial x_j \partial x_k}\left[(GG^T)_{jk}\rho\right].$$

This is the *Fokker-Planck equation* for the evolution of the probability density. The differential operator on the right-hand side is denoted by L and called the *Fokker-Planck operator* associated with (I.10).

DEFINITION. A solution $\rho(x)$ of the equation $L\rho = 0$ is called a *steady-state probability density* for (I.10).

2.6. Linear equations. We will now illustrate in the linear setting the theory developed above.

EXAMPLE 2.

$$\dot{x} = -ax + \dot{w}, \quad a > 0.$$

The equation $L\rho = ax\rho_x + a\rho + \frac{1}{2}\rho_{xx} = 0$ has a solution $\rho(x) = \sqrt{\frac{a}{\pi}}e^{-ax^2}$. This is a (unique) steady-state probability density.

More is true. We can actually solve the Fokker-Planck equation

$$\frac{\partial \rho}{\partial t} = ax\rho_x + a\rho + \frac{1}{2}\rho_{xx}$$

with the initial distribution $\rho(0,x) = \delta(x - x_0)$ (Dirac's delta function centered at x_0) to obtain

$$\rho(t,x) = \frac{\sqrt{a}}{\sqrt{\pi(1-e^{-2at})}} e^{-a(x-x_0e^{-at})^2/(1-e^{-2at})}.$$

This is a gaussian density with mean $x_0 e^{-at}$ and variance $\frac{1}{2a}(1-e^{-2at})$. In the limit as $t \to +\infty$ these approach 0 and $\frac{1}{2a}$, respectively.

For a general $\rho(0,x)$ we may then apply the formula

$$\rho(t,x) = \int_{-\infty}^{+\infty} \frac{\sqrt{a}}{\sqrt{\pi(1-e^{-2at})}} e^{-a(x-ve^{-at})^2/(1-e^{-2at})} \rho(0,v)dv$$

which shows that, regardless of the initial distribution, the density approaches steady state as $t \to +\infty$.

We would like to generalize the above discussion to the n-dimensional system

$$dx = Axdt + Bdw$$

where A and B are matrices of dimensions $n \times n$ and $n \times m$, respectively. In the context of the linear-gaussian theory, it suffices to compute the mean and the variance to determine the probability distribution uniquely. We have the equation for the mean $\frac{d}{dt}\mathcal{E}x = A\mathcal{E}x$ which imples $\mathcal{E}x(t) = e^{At}\mathcal{E}x(0) \underset{t\to+\infty}{\to} 0$ if all eigenvalues of A have negative real parts.

Since by the Itô rule and the expectation rule $\frac{d}{dt}\mathcal{E}xx^T = A\mathcal{E}xx^T + \mathcal{E}xx^T A^T + BB^T$, the equation for the *steady-state variance* $Q = \lim_{t\to+\infty} \mathcal{E}xx^T$ takes the form

$$AQ + QA^T + BB^T = 0. \tag{I.13}$$

It is well known that (I.13) has a solution $Q = \int_0^{+\infty} e^{At} B B^T e^{A^T t} dt$. When is this matrix positive definite? For an arbitrary vector $\mu \in \mathbb{R}^n$ we have $\mu^T Q \mu = \int_0^{+\infty} \mu^T e^{At} B B^T e^{A^T t} \mu dt = 0$ if and only if $\mu^T e^{At} B = 0 \; \forall t$ if and only if $\mu^T B = 0$, $\mu^T AB = 0$, $\mu^T A^2 B = 0$, $\ldots$ (by Taylor series). This will not happen if

$$\operatorname{rank}(B, AB, \ldots, A^{n-1}B) = n$$

(by the Cayley-Hamilton theorem, it suffices to consider only the first $(n-1)$ powers of A). This is precisely the controllability condition of Section 1.

As we know from Section 1, the controllability condition implies that we can steer the corresponding deterministic control system (I.4) from any given point to any other point. In the stochastic context, controllability provides a suitable mixing condition for the noise which, when combined with asymptotic stability of the nonrandom part of the system, guarantees the existence of a nondegenerate gaussian steady-state probability density. This is just one of many examples of the fruitful interplay between ideas from ordinary and stochastic systems theory.

Chapter II

QUANTIZED FEEDBACK STABILIZATION

1. Introduction. This chapter deals with quantized feedback stabilization problems for linear time-invariant control systems. A quantizer, as defined here, acts as a functional that maps a real-valued function into a piecewise constant function taking on a finite set of values. Given a system that is stabilizable by linear time-invariant feedback, the problem under consideration is to find a quantized feedback control law that stabilizes the system. Problems of this kind arise, for example, when the output measurements to be used for feedback are transmitted via a digital communication channel.

A standard assumption in the literature on quantized control is that one is given a fixed quantizer (see, among other sources, [18], [20], [22], and [34]). Here we adopt a different point of view. Namely, we treat the number of values of the quantizer as being fixed *a priori*, but we allow ourselves to alter other quantization parameters while the system evolves. This approach enables us to achieve asymptotic stability, a property that cannot be obtained with the schemes previously investigated.

We now introduce some notation and give a definition that makes the above concepts precise. We will let $\|x\|$ denote the standard euclidean norm of $x \in \mathbb{R}^n$.

We will also use the *maximum norm* on $\mathbb{R}^n$ defined by $\|x\|_\infty = \max\{|x_i| : 1 \le i \le n\}$, as well as the induced *infinity norm* on $\mathbb{R}^{n\times n}$ defined by $\|A\|_\infty = \max\{\sum_{j=1}^n |A_{ij}| : 1 \le i \le n\}$. The set of nonnegative integers will be denoted by $\mathbb{Z}_+$. We will denote by I_A the *indicator* of a set $A \in \mathbb{R}$, i.e.,

$$I_A(t) = \begin{cases} 1 & \text{if } t \in A \\ 0 & \text{if } t \notin A \end{cases}$$

We find it convenient to use the following *floor* function: $\lfloor x \rfloor = \max\{k \in \mathbb{Z} : k < x\}$. Functions denoted by capital letters F, G and H are assumed to be piecewise continuous in all their arguments.

Given a positive integer M and a nonnegative real number Δ, we define the *quantizer* $q_\Delta : \mathbb{R} \to \mathbb{Z}$ with *sensitivity* Δ and *saturation value* M by the formula

$$q_\Delta(x) = \begin{cases} M & \text{if } x > (M+1/2)\Delta \\ -M & \text{if } x \le -(M+1/2)\Delta \\ \left\lfloor \dfrac{x}{\Delta} + \dfrac{1}{2} \right\rfloor & \text{if } -(M+1/2)\Delta < x \le (M+1/2)\Delta \end{cases}$$

Thus on the interval $((k-1/2)\Delta, (k+1/2)\Delta]$, where $k \in \mathbb{Z}$ and $-M \le k \le M$, the function q_Δ takes on the value k. We will omit the subscript Δ to simplify the notation. Suppose that we have n quantizers $q_i : \mathbb{R} \to \mathbb{Z}$ with sensitivities Δ_i and the same saturation value M $(i = 1, \ldots, n)$. We define the quantizer $q : \mathbb{R}^n \to \mathbb{Z}^n$ with sensitivity $(\Delta_1, \ldots, \Delta_n)$ and saturation value M as follows: $q(x) = (q_1(x_1), \ldots, q_n(x_n))$, where $(x_1, \ldots, x_n)$ are the coordinates of x relative to a fixed orthonormal basis in $\mathbb{R}^n$. If all q_i's have the same sensitivity Δ, we will call q a *uniform* quantizer with sensitivity Δ. The above notation is similar to the one used by Delchamps in [20], but an essential feature that makes our definition different is that the set of values taken by the quantizer here is finite rather than countable. In fact, we are especially interested in situations where M is small. For example, we will consider the case when $M = 1$. The corresponding quantizer can be thought of as describing a sensor which determines whether the temperature of a certain object is "normal", "too high", or "too low".

The approach to be used here is based on the hypothesis that it is possible to change the sensitivity (but not the saturation value) of the quantizer on the basis of available quantized measurements. For example, given a temperature sensor with limited capability of the kind mentioned above, this would simply amount to changing the threshold values. In the continuous-time case, we will describe the evolution of Δ with time by an equation of the form

$$\Delta(t) = G(\lfloor t/\tau \rfloor, q(x(\lfloor t/\tau \rfloor \tau)), \Delta(\lfloor t/\tau \rfloor \tau))$$

where τ is a fixed positive number. The above equation defines a "strictly causal" function Δ that is continuous from the left and maintains a constant value on each interval $(k\tau, k\tau + \tau]$, $k \in \mathbb{Z}_+$. In the control policies considered below, such an equation for Δ will be coupled with the given linear system. This results in a *hybrid system* that might take the form

$$\begin{aligned} \dot{x}(t) &= Ax(t) + Bu(t, q(x(t)), \Delta(t)), \\ \Delta(t) &= G(\lfloor t/\tau \rfloor, q(x(\lfloor t/\tau \rfloor \tau)), \Delta(\lfloor t/\tau \rfloor \tau)). \end{aligned} \tag{II.1}$$

This system falls into the general framework for hybrid systems presented in [12]. Clearly, for every initial condition $(x(0), \Delta(0))$ there exists a unique solution trajectory. The system (II.1), as well as all other systems of differential-difference equations considered in this chapter, is of *hereditary type*, and as such is covered by the theory of hereditary systems developed in [26].

The control policy will usually be divided into two stages. First, since the initial state is unknown, we will have to "zoom out", i.e., increase Δ until we can adequately measure the state of the system. Second, we will "zoom in", i.e., decrease Δ in such a way as to drive the state to 0. In essence, our goal is to demonstrate that if a linear system can be stabilized by linear time-invariant feedback, then it can also be stabilized by quantized feedback with the help of the approach described above.

The outline of this chapter is as follows. In Section 2 we develop techniques for stabilizing continuous-time linear systems with quantized state feedback. In Section 3 we present analogous results for discrete-time systems. Section 4 deals with quantized output feedback stabilization. In Section 5 we describe control strategies that involve state observation.

2. Quantized state feedback stabilization: continuous time. This section deals with state feedback stabilization problems for the continuous-time linear system

$$\dot{x} = Ax + Bu \tag{II.2}$$

where $x \in \mathbb{R}^n$, $u \in \mathbb{R}^m$, and A and B are matrices if suitable dimensions. If (II.2) is controllable in the unstable modes, then there exists a matrix K such that all eigenvalues of $A - BK$ have negative real parts (see [38, Section 6.3]). In this case it seems logical to try to implement a quantized state feedback control law of the form $u = -K\Delta q(x)$, where q is a uniform quantizer with sensitivity Δ.

Theorem 1. *Suppose that all eigenvalues of $A - BK$ have negative real parts. Then there exists a control policy of the form*

$$\Delta(t) = G(\lfloor t/\tau \rfloor, q(x(\lfloor t/\tau \rfloor \tau)), \Delta(\lfloor t/\tau \rfloor \tau)),$$

$$u(t) = -KI_{[k_0\tau,+\infty)}(t)\Delta(t)q(x(t))$$

where q is a uniform quantizer with sensitivity $\Delta(t)$, such that the solutions of the system

$$\dot{x}(t) = Ax(t) - BKI_{[k_0\tau,+\infty)}(t)\Delta(t)q(x(t)), \quad x(0) \text{ arbitrary}$$

$$\Delta(t) = G(\lfloor t/\tau \rfloor, q(x(\lfloor t/\tau \rfloor \tau)), \Delta(\lfloor t/\tau \rfloor \tau)), \quad \Delta(0) = 0$$

approach 0 as $t \to +\infty$.

Proof. Consider the system

$$\dot{x} = Ax - BK\Delta q(x)$$

which we can also write as

$$\dot{x} = (A - BK)x + BKs(x) \tag{II.3}$$

thus displaying the "error" vector $s(x) = x - \Delta q(x)$. When

$$\|x\|_\infty \le \Big(M + \frac{1}{2}\Big)\Delta \tag{II.4}$$

we have $\|s(x)\| \leq \Delta\sqrt{n}/2$. We will let $\lambda_{min}(P)$ and $\lambda_{max}(P)$ denote the smallest and the largest eigenvalue of a symmetric matrix P, respectively. Recall that by standard Liapunov stability theory there exist positive definite symmetric matrices Q and D such that $(A-BK)^T Q + Q(A-BK) = -D$. Whenever (II.4) holds, the derivative of $x^T Qx$ along the solutions of (II.3) is given by

$$\begin{aligned}\frac{d}{dt}x^T Qx &= -x^T Dx + 2x^T QBKs(x)\\ &\leq -\lambda_{min}(D)\|x\|^2 + 2\|x\|\|QBK\|\Delta\sqrt{n}/2\\ &= -\|x\|(\lambda_{min}(D)\|x\| - \|QBK\|\Delta\sqrt{n}).\end{aligned} \tag{II.5}$$

The last expression is negative outside the ball $\{x : \|x\| \leq \Theta\Delta\sqrt{n}\}$, where $\Theta = \|QBK\|/\lambda_{min}(D)$. In what follows we will use the simple facts that the radius of the ball inscribed in an ellipsoid $\{x : x^T Qx \leq \gamma^2\}$ equals $\gamma/\sqrt{\lambda_{max}(Q)}$ and the radius of the ball circumscribed about the same ellipsoid equals $\gamma/\sqrt{\lambda_{min}(Q)}$. Fix an arbitrary $\epsilon > 0$. Define the *scaling factor* Ω by the formula

$$\Omega = \left(\Theta\sqrt{n} + \epsilon\right)\frac{\lambda_{max}(Q)}{\lambda_{min}(Q)}\left(M + \frac{1}{2}\right)^{-1}$$

and take the saturation value M of q to be large enough so that $\Omega < 1$. Define

$$\tau = \frac{(M+1/2)^2\lambda_{min}(Q) - (\Theta\sqrt{n}+\epsilon)^2\lambda_{max}(Q)}{(\Theta\sqrt{n}+\epsilon)\lambda_{min}(D)\epsilon}.$$

Since $\Omega < 1$, it is easy to see that $\tau > 0$. Set the control to 0 and let $\Delta(0) = 0$. Increase Δ fast enough relative to the rate of growth of $\|e^{At}\|$, e.g., let $\Delta(k\tau) = k\tau e^{\|A\|k\tau}$ for $k \in \mathbb{Z}_+$. Then $\exists k > 0$ such that

$$\|x(k\tau)\| \leq \Delta(k\tau)\left(\left(M + \frac{1}{2}\right)\sqrt{\frac{\lambda_{min}(Q)}{\lambda_{max}(Q)}} - \sqrt{n}\right)$$

hence

$$\|q(x(k\tau))\| \leq \Big(M + \frac{1}{2}\Big)\sqrt{\frac{\lambda_{min}(Q)}{\lambda_{max}(Q)}} - \frac{\sqrt{n}}{2}.$$

We define

$$k_0 = \min\Big\{k \geq 1 : \|q(x(k\tau))\| \leq \Big(M + \frac{1}{2}\Big)\sqrt{\frac{\lambda_{min}(Q)}{\lambda_{max}(Q)}} - \frac{\sqrt{n}}{2}\Big\}$$

which implies

$$\|x(k_0\tau)\| \leq \Delta(k_0\tau)\Big(M + \frac{1}{2}\Big)\sqrt{\frac{\lambda_{min}(Q)}{\lambda_{max}(Q)}}.$$

Therefore,

$$x^T(k_0\tau)Qx(k_0\tau) \leq (\Delta(k_0\tau))^2\Big(M + \frac{1}{2}\Big)^2\lambda_{min}(Q). \tag{II.6}$$

Observe that (II.4) holds with $\Delta = \Delta(k_0\tau)$ for any x that belongs to the ellipsoid

$$R_1 = \Big\{x : x^T Qx \leq (\Delta(k_0\tau))^2\Big(M + \frac{1}{2}\Big)^2\lambda_{min}(Q)\Big\}.$$

Since $\Omega < 1$, it follows that if we let $u(t) = -K\Delta(t)q(x(t))$ with $\Delta(t) = \Delta(k_0\tau)$ for $k_0\tau < t \leq k_0\tau + \tau$, then x will not leave R_1. We claim that

$$\|x(k_0\tau + \tau)\| \leq \Delta(k_0\tau)(\Theta\sqrt{n} + \epsilon)\sqrt{\frac{\lambda_{max}(Q)}{\lambda_{min}(Q)}}. \tag{II.7}$$

Suppose that (II.7) is not true. Then

$$x^T(k_0\tau + \tau)Qx(k_0\tau + \tau) > (\Delta(k_0\tau))^2(\Theta\sqrt{n} + \epsilon)^2\lambda_{max}(Q) \tag{II.8}$$

and therefore

$$\|x(t)\| > \Delta(k_0\tau)(\Theta\sqrt{n} + \epsilon) \quad \text{for all } t \in [k_0\tau, k_0\tau + \tau]. \tag{II.9}$$

But (II.5) and (II.9) imply that for $k_0\tau \le t \le k_0\tau + \tau$ we have

$$\frac{d}{dt} x^T Q x \le -(\Delta(k_0\tau))^2(\Theta\sqrt{n} + \epsilon)\lambda_{min}(D)\epsilon.$$

Comparing the last inequality with (II.6) and (II.8), we arrive at a contradiction, which establishes the validity of (II.7).

The inequality (II.7) guarantees that

$$x^T(k_0\tau + \tau)Qx(k_0\tau + \tau) \le (\Omega\Delta(k_0\tau))^2\Big(M + \frac{1}{2}\Big)^2\lambda_{min}(Q).$$

Thus we let $u(t) = -K\Delta(t)q(x(t))$ with $\Delta(t) = \Omega\Delta(k_0\tau)$ for $k_0\tau + \tau < t \le k_0\tau + 2\tau$, which yields

$$x^T(k_0\tau + 2\tau)Qx(k_0\tau + 2\tau) \le (\Omega^2\Delta(k_0\tau))^2\Big(M + \frac{1}{2}\Big)^2\lambda_{min}(Q).$$

Similarly, we let $\Delta(t) = \Omega^2\Delta(k_0\tau)$ for $k_0\tau + 2\tau < t \le k_0\tau + 3\tau$. Repeating this procedure, we obtain the desired control policy: $\Delta(t) \to 0$ as $t \to +\infty$, and by the above analysis the same is true for $x(t)$. □

Remark. The above control policy calls for Δ taking on a countable set of values. In some applications there may only be a finite set of possible values for Δ (for example, if the values of Δ have to be passed through a quantizer with fixed sensitivity). Adjusting our control policy to this case, we would only obtain stability and not global asymptotic stability claimed in Theorem 1.

The control policy described above uses the so-called *dwell-time switching logic* [31] in the sense that the value of Δ is held constant on time intervals of fixed length τ. Another possibility is to change Δ every time $\|q(x(t))\|$ becomes smaller than or equal to a certain prescribed value. To demonstrate how this alternative method works, we will use it in proving the discrete-time counterpart of Theorem 1 (Theorem 3 in the next section). The main advantage of the dwell-time switching approach is that it can also be applied to quantized output feedback stabilization problems (cf. Sections 4 and 5 below). In specific applications, one could compare the effectiveness of these two methods with respect to various performance characteristics, such as the speed of convergence of solution trajectories to zero (time-optimality), or the number of times the control function changes its value per unit interval of time ("minimum attention control"—cf. [16]).

EXAMPLE 1. Consider the equation

$$\dot{x} = ax - k\Delta q(x)$$

where $x \in \mathbb{R}$ and $k > a > 0$. In this case we have $\Theta = 3/2$. In fact, if $k > 3a/2$ and $|x(0)| \leq (M + 1/2)\Delta$, it is possible to show that x eventually enters the interval $[-\Delta/2, \Delta/2]$. When that happens, we can scale Δ according to the rule $\Delta \mapsto \dfrac{\Delta/2}{M+1/2}$.

We will now address in passing the issue of time sampling in the context of Example 1. Namely, suppose that we are not measuring $q(x(t))$ continuously, but instead we are sampling the process at times $0, \delta, 2\delta, \ldots$, where $\delta > 0$ is the *sampling period.* This results in the equation

$$\dot{x}(t) = ax(t) - k\Delta q(\lfloor x/\delta \rfloor \delta).$$

Do we still have asymptotic stability? The answer is yes, provided that *no overshooting occurs*. Namely, we have to make sure that if, say, $x(0) < 0$, then $x(t)$ remains negative for all time. This can be done by means of a simple calculation. Suppose that the sampling is performed at $t = 0$ and $x(0) = -\Delta/2$ (the most "dangerous" case). Then for $t \in (0, \delta]$ we have $\dot{x}(t) = ax(t) + k\Delta$, therefore the solution is given by $x(t) = \left(\frac{k\Delta}{a} - \frac{\Delta}{2}\right)e^{at} - \frac{k\Delta}{a}$. We see that $x(t)$ will become zero if $t = \frac{1}{a}\ln\frac{2k}{2k-a}$, thus in fact it will never happen if $\delta < \frac{1}{a}\ln\frac{2k}{2k-a}$ (i.e., if sampling is performed frequently enough). It is important to notice that this last number does not depend on Δ, so we can still change the sensitivity in the way described above.

As we are about to see, it is possible to stabilize the system (II.2) with quantized state feedback even if the saturation value M of the quantizer is substantially smaller than that required in the proof of Theorem 1. In fact, let us show that we can achieve global asymptotic stability using a (non-uniform) quantizer q with $M = 1$.

Theorem 2. *Suppose that all eigenvalues of $A - BK$ have negative real parts. Then there exists a control policy of the form*

$$\Delta_i(t) = G_i(\lfloor t/\tau \rfloor, q(x(\lfloor t/\tau \rfloor \tau))), \quad i = 1, \ldots, n,$$

$$u(t) = H(t - \lfloor t/\tau \rfloor \tau, q(x(\lfloor t/\tau \rfloor \tau))),$$

where q is a quantizer with sensitivity $(\Delta_1(t), \ldots, \Delta_n(t))$ and saturation value 1, *such that the solutions of the system*

$$\dot{x}(t) = Ax(t) + BH(t - \lfloor t/\tau \rfloor \tau, q(x(\lfloor t/\tau \rfloor \tau))), \quad x(0) \ \textit{arbitrary}$$

$$\Delta_i(t) = G_i(\lfloor t/\tau \rfloor, q(x(\lfloor t/\tau \rfloor \tau))), \quad \Delta(0) = 0$$

approach 0 *as* $t \to +\infty$.

Proof. Fix a number $\epsilon \in (0,1)$. Since $\|I\|_\infty = 1$, we can find a number $\tau > 0$ such that $\|e^{At}\|_\infty < 1+\epsilon$ for all $t \in [0,\tau]$. If we let $u(t) = 0$ and $\Delta_i(k\tau) = k\tau e^{\|A\|k\tau}$ for $k \in \mathbb{Z}_+$, then there exists a well-defined number $k_0 = \min\{k \geq 1 : q(x(k\tau)) = 0\}$. We have $\|x(k_0\tau)\|_\infty \leq E_0$, where $E_0 = k_0\tau e^{\|A\|k_0\tau}/2$. Thus 0 can be viewed as an estimate of $x(k_0\tau)$ with the estimation error whose maximum norm is at most E_0. Our goal is to construct a sequence of state estimates with estimation errors approaching 0 as $t \to +\infty$.

For $k_0\tau < t \leq k_0\tau + \tau$, let $u(t) = 0$. This gives $x(t) = e^{A(t-k_0\tau)}x(k_0\tau)$, hence $\|x(t)\|_\infty \leq (1+\epsilon)E_0$. The quantized measurement $q(x(k_0\tau+\tau))$ with $\Delta_i(k_0\tau+\tau) = (1+\epsilon)E_0$ singles out a rectilinear box with edges at most $(1+\epsilon)E_0$ which contains $x(k_0\tau+\tau)$. Denoting the center of this box by $\bar{x}(k_0\tau+\tau)$, we see that

$$\|x(k_0\tau+\tau) - \bar{x}(k_0\tau+\tau)\|_\infty \leq (1+\epsilon)E_0/2.$$

For $k_0\tau+\tau < t \leq k_0\tau+2\tau$, let $u(t) = -Ke^{(A-BK)(t-k_0\tau-\tau)}\bar{x}(k_0\tau+\tau)$. The variation of constants formula gives

$$\begin{aligned}
x(t) &= e^{A(t-k_0\tau-\tau)}x(k_0\tau+\tau) \\
&- \int_{k_0\tau+\tau}^{t} e^{A(t-\sigma)}BKe^{(A-BK)(\sigma-k_0\tau-\tau)}\bar{x}(k_0\tau+\tau)d\sigma \\
&= e^{A(t-k_0\tau-\tau)}x(k_0\tau+\tau) \\
&- e^{At}e^{-(A-BK)(k_0\tau+\tau)}\bar{x}(k_0\tau+\tau)\int_{k_0\tau+\tau}^{t} BKe^{-BK\sigma}d\sigma \\
&= e^{A(t-k_0\tau-\tau)}(x(k_0\tau+\tau) \\
&- \bar{x}(k_0\tau+\tau)) + e^{(A-BK)(t-k_0\tau-\tau)}\bar{x}(k_0\tau+\tau)
\end{aligned}$$

hence $\|x(t) - e^{(A-BK)(t-k_0\tau-\tau)}\bar{x}(k_0\tau+\tau)\|_\infty \leq (1+\epsilon)^2 E_0/2$. The quantized measurement $q(x(k_0\tau+2\tau))$ with

$$\Delta_i(k_0\tau+2\tau) = \begin{cases} 2e^{(A-BK)\tau}(\bar{x}(k_0\tau+\tau))_i & \text{if } (\bar{x}(k_0\tau+\tau))_i \neq 0 \\ (1+\epsilon)^2 E_0/2 & \text{if } (\bar{x}(k_0\tau+\tau))_i = 0 \end{cases}$$

singles out a rectilinear box with edges at most $(1+\epsilon)^2 E_0/2$ which contains $x(k_0\tau+2\tau)$. Denoting the center of this box by $\bar{x}(k_0\tau+2\tau)$, we see that

$$\|x(k_0\tau+2\tau) - \bar{x}(k_0\tau+2\tau)\|_\infty \leq (1+\epsilon)^2 E_0/4.$$

For $k_0\tau+2\tau < t \leq k_0\tau+3\tau$, let $u(t) = -Ke^{(A-BK)(t-k_0\tau-2\tau)}\bar{x}(k_0\tau+2\tau)$. Proceeding in this fashion, we obtain a piecewise continuous control function u such that $\|u(t)+Kx(t)\|_\infty \to 0$ as $t \to +\infty$. The same statement is therefore true for the euclidean norm $\|u(t)+Kx(t)\|$. Repeating the argument used in the proof of Theorem 1, we see that $x(t) \to 0$ as $t \to +\infty$. □

Remark. Again, if the set of possible values for Δ_i is finite, global asymptotic stability is replaced by stability (cf. Section 4 below).

3. Quantized state feedback stabilization: discrete time. In this section we will establish counterparts of Theorems 1 and 2 for the discrete-time system in $\mathbb{R}^n$

$$x(k+1) = Ax(k) + Bu(k). \tag{II.10}$$

Theorem 3. *Suppose that all eigenvalues of $A - BK$ lie inside the unit circle. Then there exists a control policy of the form*

$$\Delta(k+1) = G(k, q(x(k)), \Delta(k)),$$
$$u(k) = -KI_{[k_0,+\infty)}(k)\Delta(k)q(x(k))$$

where q is a uniform quantizer with sensitivity $\Delta(k)$, such that the solutions of the system

$$x(k+1) = Ax(k) - BKI_{[k_0,+\infty)}(k)\Delta(k)q(x(k)), \quad x(0) \text{ arbitrary}$$
$$\Delta(k+1) = G(k, q(x(k)), \Delta(k)), \quad \Delta(0) = 0$$

approach 0 as $k \to +\infty$.

Proof. Consider the system

$$x(k+1) = Ax(k) - BK\Delta q(x(k))$$

which we can also write as

$$x(k+1) = (A - BK)x(k) + BKs(x(k)) \tag{II.11}$$

with $s(x) = x - \Delta q(x)$ as before. If (II.4) holds, then we have $\|s(x)\| \leq \Delta\sqrt{n}/2$. By standard Liapunov stability theory for discrete-time linear systems, there exist positive definite symmetric matrices Q and D such that $(A - BK)^T Q(A - BK) - Q = -D$. For the solutions of (II.11) this implies

$$
\begin{aligned}
&x^T(k+1)Qx(k+1) - x^T(k)Qx(k) \\
&= -x^T(k)Dx(k) + 2x^T(k)(A-BK)^T QBKs(x(k)) \\
&+ s^T(x(k))K^T B^T QBKs(x(k)) \\
&\leq -\lambda_{min}(D)\|x(k)\|^2 + 2\|x(k)\|\|(A-BK)^T QBK\|\Delta\sqrt{n}/2 \\
&+ \|K^T B^T QBK\|\Delta^2 n/4.
\end{aligned}
$$

The last expression is negative ouside the ball $\{x : \|x\| \leq \Theta\Delta\sqrt{n}\}$, where

$$
\begin{aligned}
\Theta = \frac{1}{2\lambda_{min}(D)} \Big(&\|(A-BK)^T QBK\| \\
&+ \sqrt{\|(A-BK)^T QBK\|^2 + \lambda_{min}(D)\|K^T B^T QBK\|} \Big)
\end{aligned}
$$

(cf. [20, Proposition 2.3]). Define the scaling factor Ω by the formula

$$
\Omega = \left(\Theta\sqrt{\frac{n\lambda_{max}(Q)}{\lambda_{min}(Q)}} + \sqrt{n} \right) \sqrt{\frac{\lambda_{max}(Q)}{\lambda_{min}(Q)}} \left(M + \frac{1}{2} \right)^{-1}
$$

and take the saturation value M of q to be large enough so that $\Omega < 1$. If we let $u(k) = 0$ and $\Delta(k) = k\|A\|^k$ for $k \in \mathbb{Z}_+$, then there exists a well-defined number

$$
k_0 = \min\left\{ k \geq 1 : \|q(x(k))\| \leq \left(M + \frac{1}{2} \right) \sqrt{\frac{\lambda_{min}(Q)}{\lambda_{max}(Q)}} - \frac{\sqrt{n}}{2} \right\}.
$$

We have

$$
\|x(k_0)\| \leq \Delta(k_0) \left(M + \frac{1}{2} \right) \sqrt{\frac{\lambda_{min}(Q)}{\lambda_{max}(Q)}}.
$$

Therefore, $x(k_0)$ belongs to the ellipsoid

$$R_1 = \Big\{x : x^T Qx \leq (\Delta(k_0))^2 \Big(M + \frac{1}{2}\Big)^2 \lambda_{min}(Q)\Big\}.$$

Observe that (II.4) holds with $\Delta = \Delta(k_0)$ for all $x \in R_1$. Since $\Omega < 1$, it follows that if we let $u(k) = -K\Delta(k)q(x(k))$ with $\Delta(k) = \Delta(k_0)$ for $k \geq k_0$, then x will never leave R_1. Moreover, x will eventually enter the ellipsoid

$$R_0 = \{x : x^T Qx \leq (\Delta(k_0))^2 \Theta^2 n\lambda_{max}(Q)\}.$$

Thus we can define

$$k_1 = \min\Big\{k \geq k_0 + 1 : \|q(x(k))\| \leq \Theta\sqrt{\frac{n\lambda_{max}(Q)}{\lambda_{min}(Q)}} + \frac{\sqrt{n}}{2}\Big\}$$

which implies

$$\|x(k_1)\| \leq \Delta(k_0)\Big(\Theta\sqrt{\frac{n\lambda_{max}(Q)}{\lambda_{min}(Q)}} + \sqrt{n}\Big).$$

When $k = k_1$, change the sensitivity to $\Delta(k_1) = \Omega\Delta(k_0)$. Arguing as before, we can show that if we let $u(k) = -K\Delta(k)q(x(k))$ with $\Delta(k) = \Delta(k_1)$ for $k \geq k_1$, then (II.4) will still hold, and there exists a well-defined number

$$k_2 = \min\Big\{k \geq k_1 + 1 : \|q(x(k))\| \leq \Theta\sqrt{\frac{n\lambda_{max}(Q)}{\lambda_{min}(Q)}} + \frac{\sqrt{n}}{2}\Big\}.$$

When $k = k_2$, change the sensitivity to $\Delta(k_2) = \Omega\Delta(k_1)$. Repeating this procedure, we obtain a sequence $\Delta(k_0), \Delta(k_1), \Delta(k_2), \Delta(k_3), \cdots \to 0$. We see that $x(k) \to 0$ as $k \to +\infty$. $\square$

Our analogue of Theorem 2 for the discrete-time case contains one additional assumption which means, loosely speaking, that the state of the uncontrolled system $x(k+1) = Ax(k)$ is "not too unstable".

Theorem 4. *Suppose that all eigenvalues of $A - BK$ lie inside the unit circle. Suppose also that $\|A\|_\infty \le 2$. Then there exists a control policy of the form*

$$\Delta_i(k+1) = G_i(k, q(x(k))), \quad i = 1, \dots, n$$
$$u(k+1) = H(q(x(k)))$$

where q is a quantizer with sensitivity $(\Delta_1(k), \dots, \Delta_n(k))$ and saturation value 1, such that the solutions of the system

$$x(k+1) = Ax(k) + BH(q(x(k))), \quad x(0) \text{ arbitrary}$$
$$\Delta_i(k+1) = G_i(k, q(x(k))), \quad \Delta(0) = 0$$

approach 0 as $k \to +\infty$.

Proof. If we let $u(k) = 0$ and $\Delta_i(k) = k\|A\|^k$ for $k \in \mathbb{Z}_+$, then there exists $k_0 = \min\{k \ge 1 : q(x(k)) = 0\}$. We have $\|x(k_0)\|_\infty \le E_0$, where $E_0 = k_0\|A\|^{k_0}/2$. We will construct a sequence of state estimates $\bar{x}(k)$ such that $\|x(k) - \bar{x}(k)\|_\infty \to 0$ as $k \to +\infty$.

Let $u(k_0) = 0$. Then $\|x(k_0+1)\|_\infty = \|Ax(k_0)\|_\infty \le \|A\|_\infty E_0$. The quantized measurement $q(x(k_0+1))$ with $\Delta_i(k_0+1) = \|A\|_\infty E_0$ singles out a rectilinear box with edges at most $\|A\|_\infty E_0$ which contains $x(k_0+1)$. Denoting the center of this box by $\bar{x}(k_0+1)$, we obtain

$$\|x(k_0+1) - \bar{x}(k_0+1)\|_\infty \le \|A\|_\infty E_0/2.$$

Next, let $u(k_0+1) = -K\bar{x}(k_0+1)$. We have $x(k_0+2) = Ax(k_0+1) - BK\bar{x}(k_0+$

1) hence $\|x(k_0+2)-(A-BK)\bar{x}(k_0+1)\|_\infty \le \|A\|_\infty^2 E_0/2$. The quantized measurement $q(x(k_0+2))$ with

$$\Delta_i(k_0+2) = \begin{cases} 2(A-BK)(\bar{x}(k_0+1))_i & \text{if } (\bar{x}(k_0+1))_i \neq 0 \\ \|A\|_\infty^2 E_0/2 & \text{if } (\bar{x}(k_0+1))_i = 0 \end{cases}$$

singles out a rectilinear box with edges at most $\|A\|_\infty^2 E_0/2$ which contains $x(k_0+2)$. Denoting the center of this box by $\bar{x}(k_0+2)$, we obtain

$$\|x(k_0+2)-\bar{x}(k_0+2)\|_\infty \le \|A\|_\infty^2 E_0/4.$$

Next, let $u(k_0+2) = -K\bar{x}(k_0+2)$. Repeating the above procedure for each k, we obtain a control function u such that $\|u(k)+Kx(k)\|_\infty \to 0$ as $k \to +\infty$. The statement of the theorem follows. $\square$

4. Quantized output feedback stabilization. We now turn to the problem of stabilizing the system

$$\begin{aligned} \dot{x} &= Ax + Bu, \\ y &= Cx \end{aligned} \tag{II.12}$$

with quantized measurements of the output. Assume (without loss of generality) that $\|C\|_\infty \le 1$. Suppose that there exists a matrix K such that all eigenvalues of $A-BKC$ have negative real parts (see [9] for conditions under which this is the case). If $y \in \mathbb{R}^p$ with $p < n$, then the initial stage of the stabilizing control policy of Theorem 1 cannot be implemented. For this reason, in the next theorem the asymptotic stability is local.

Theorem 5. *Suppose that all eigenvalues of $A - BKC$ have negative real parts. For any number $E_0 > 0$, there exists a control policy of the form*

$$\Delta(t) = G(\Delta(\lfloor t/\tau \rfloor \tau)),$$

$$u(t) = -K\Delta(t)q(y(t))$$

where q is a uniform quantizer with sensitivity $\Delta(t)$, such that the solutions of the system

$$\dot{x}(t) = Ax(t) - BK\Delta(t)q(Cx(t)), \quad \|x(0)\| \le E_0$$

$$\Delta(t) = G(\Delta(\lfloor t/\tau \rfloor \tau)), \quad \Delta(0) = 0$$

approach 0 as $t \to +\infty$.

Proof. As before, there exist positive definite symmetric matrices Q and D such that $(A - BKC)^T Q + Q(A - BKC) = -D$. Fix an arbitrary $\epsilon > 0$. Define Θ, Ω and τ as in the proof of Theorem 1, and take the saturation value M of q to be large enough so that $\Omega < 1$. Let $u = -K\Delta q(Cx)$, where q is a uniform quantizer with sensitivity Δ. We can write (II.12) as

$$\dot{x} = Ax - BKq(Cx) = (A - BKC)x + BKs(Cx) \tag{II.13}$$

where $s(Cx) = Cx - \Delta q(Cx)$. Whenever $\|y\|_\infty \le (M + 1/2)\Delta$, the upper bound given by (II.5) is valid for the derivative of $x^T Q x$ along the solutions of (II.13).

Suppose that $\|x(0)\| \le E_0$. Let us choose the initial sensitivity $\Delta(0)$ large enough to have

$$E_0 \le \Delta(0)\Big(M + \frac{1}{2}\Big)\sqrt{\frac{\lambda_{min}(Q)}{\lambda_{max}(Q)}}$$

hence

$$x^T(0)Qx(0) \le (\Delta(0))^2\Big(M + \frac{1}{2}\Big)^2 \lambda_{min}(Q).$$

This is completely analogous to (II.6) (with $k_0 = 0$), and the rest of the proof carries over from Theorem 1 with obvious minor modifications. $\square$

The discrete-time case can be treated similarly. We now turn our attention to quantizers with saturation value 1. We will describe a quantized output feedback control policy that yields asymptotic stability. It takes the form of a differential equation which makes the sensitivity change fast enough so as to dominate the dynamics of the underlying linear system. For simplicity, let us consider the single-input, single-output system

$$\begin{aligned}\dot{x} &= Ax + bu,\\ y &= c^T x\end{aligned} \qquad \text{(II.14)}$$

(generalization to the case of (II.12) is straightforward). Suppose that there exists a feedback gain k such that all eigenvalues of $A - bkc^T$ have negative real parts. All such gains k can be found by using the Nyquist criterion (see Section 1 of Chapter I). Without loss of generality we may assume that $\|c\| \le 1$.

Theorem 6. *Suppose that all eigenvalues of $A - bkc^T$ have negative real parts. For any number $E_0 > 0$, there exists a control policy of the form*

$$\begin{aligned}\dot{\Delta}(t) &= G(q(y(t))),\\ u(t) &= -k\Delta(t)q(y(t))/2\end{aligned}$$

where q is a quantizer with sensitivity $\Delta(t)$ and saturation value 1, such that the solutions of the system

$$\dot{x}(t) = Ax(t) - bk\Delta(t)q(c^T x(t))/2, \quad \|x(0)\| \le E_0$$

$$\dot{\Delta}(t) = G(q(y(t))), \quad \Delta(0) = 0$$

(interpreted in the Filippov's sense) approach 0 *as* $t \to +\infty$.

Proof. We know that there exist positive definite symmetric matrices Q and D such that $(A - bkc^T)^T Q + Q(A - bkc^T) = -D$. Now, consider the system

$$\dot{x} = Ax - bk\Delta q(c^T x)/2, \quad \|x(0)\| \le E_0$$

$$\dot{\Delta} = (2|q(c^T x)| - 1)\beta, \quad \Delta(0) = 0 \tag{II.15}$$

where $\beta = 2(\|A\| + \|b\|k)E + 2$ and $E = \max\left(\frac{\|Qbk\|2E_0}{\lambda_{min}(D)}, E_0\right)\sqrt{\frac{\lambda_{max}(Q)}{\lambda_{min}(Q)}}$. The right-hand side of (II.15) is discontinuous when $|c^T x| = \Delta/2$, and we will interpret solutions of (II.15) in the Filippov's sense [23].

We have $||c^T x(0)| - \Delta(0)/2| \le E_0$, and $\frac{d}{dt}(||c^T x(t)| - \Delta(t)/2|) \le -1$ for $t \ge 0$ as long as $|c^T x(t)| \ne \Delta(t)/2$ and $\|x(t)\| \le E$. But from the analysis of Section 2 it follows that the solution trajectory will never leave the region $\{x : \|x\| \le E\}$. This means that $\exists t_0 > 0$ such that $|c^T x(t_0)| = \Delta(t_0)/2$. Now it is not hard to show that the solution trajectory stays on the discontinuity locus $\{(x, \Delta) : |c^T x| = \Delta/2\}$ for all $t \ge t_0$. It remains to use the fact that the system $\dot{x} = (A - bkc^T)x$ is asymptotically stable. $\square$

Given an *a priori* upper bound E_0 on the norm of $x(0)$, we can use the control strategy described in Theorem 6 to drive the state to 0. A drawback of the above solution is that Δ is changing continuously, which might be undesirable in some applications. As the most restrictive case, consider a situation where Δ can only take on values in the set $\{-N\sigma, (-N+1)\sigma, \ldots, -\sigma, 0, \sigma, \ldots, (N-1)\sigma, N\sigma\}$, where σ is a fixed positive real number and N is a fixed positive integer. Then we can replace (II.15) by a hybrid system that takes the form

$$\begin{aligned} \dot{x}(t) &= Ax(t) - bk\Delta(t)q(c^T x(t))/2, \quad \|x(0)\| \leq E_0 \\ \Delta(t) &= \Delta(\lfloor t/\tau \rfloor \tau) + (2|q(c^T x(t))| - 1)\sigma, \quad \Delta(0) = 0 \end{aligned} \tag{II.16}$$

where $\tau = \sigma(2(\|A\| + \|b\|k)N\sigma + 2)^{-1}$. Note that Δ changes its value by σ every τ units of time. Using our earlier developments, one can easily establish the following stability property of (II.16).

Proposition 7. *Let Q, D, σ and N be as in the foregoing. Suppose that N is large enough and $E_0 \in \mathbb{R}$ is small enough so that*

$$\max\left(\frac{\|Qbk\|(2E_0 + \sigma)}{\lambda_{min}(D)}, E_0\right)\sqrt{\frac{\lambda_{max}(Q)}{\lambda_{min}(Q)}} \leq \sigma N.$$

Then solutions of the system (II.16) eventually enter the region

$$R_0 = \left\{x : x^T Q x \leq \left(\frac{\|Qbk\|2\sigma}{\lambda_{min}(D)}\right)^2 \lambda_{max}(Q)\right\}$$

and stay there for all future time.

One could also consider a situation where σ can take on different values from a certain finite set. This would increase the domain of stability for (II.16) and make the attracting invariant region R_0 smaller.

5. Observability and quantized feedback stabilization. We will now show that, by employing more sophisticated techniques than those presented in Section 4, it is possible to design a quantized output feedback control that makes *all* solutions of the system (II.12) approach 0. At the beginning of Section 4 we made the assumption that (II.12) is stabilizable by linear time-invariant output feedback, which implies that (II.12) is observable in the unstable modes (see [38, Section 6.4]). This property is known as *detectability*. Since we are concerned with feedback stabilization, it makes sense to assume that (A, C) is actually an observable pair. As it turns out, the hypothesis that (II.12) is stabilizable by linear time-invariant output feedback can then be relaxed considerably. Namely, we will only require that (II.12) be stabilizable by linear time-invariant *state* feedback. This means that there exists a matrix K such that all eigenvalues of $A - BK$ have negative real parts, but there might not exist any K such that all eigenvalues of $A - BKC$ have negative real parts.

Since only quantized measurements of the output, and not of the state, are available, we have to develop a method for constructing state estimates which we will denote by $\bar{x}(t)$. We will need to use the history of past quantized output measurements, in contrast with the simpler techniques of the previous sections.

The evolution of $\bar{x}(t)$ will thus be described by a Volterra integral equation.

Theorem 8. *Suppose that (A, C) is an observable pair, and that all eigenvalues of $A - BK$ have negative real parts. Then there exists a control policy of the form*

$$\bar{x}(t) = \iiint\limits_{t-\tau \leq s_i \leq t,\ i=1,2,3} F(t, q(y(s_1)), \Delta(s_2), u(s_3)) ds_1 ds_2 ds_3,$$
$$\Delta(t) = G(\lfloor t/\tau \rfloor, \bar{x}(\lfloor t/\tau \rfloor \tau), q(x(\lfloor t/\tau \rfloor \tau)), \Delta(\lfloor t/\tau \rfloor \tau)),$$
$$u(t) = H(t, \bar{x}(t))$$

where q is a uniform quantizer with sensitivity $\Delta(t)$, such that the solutions of the system

$$\dot{x}(t) = Ax(t) + BH(t, \bar{x}(t)), \quad x(0) \text{ arbitrary}$$
$$\Delta(t) = G(\lfloor t/\tau \rfloor, \bar{x}(\lfloor t/\tau \rfloor \tau), q(x(\lfloor t/\tau \rfloor \tau)), \Delta(\lfloor t/\tau \rfloor \tau)), \quad \Delta(0) = 0$$

approach 0 as $t \to +\infty$.

Proof. Let Q, D and Θ be as in the proof of Theorem 1. Define

$$\Omega = \Psi\left(\Theta\sqrt{n} + \epsilon\right) \frac{\lambda_{max}(Q)}{\lambda_{min}(Q)} \left(M + \frac{1}{2}\right)^{-1}$$

where

$$\Psi = \|W^{-1}\|_\infty \tau (1 + \epsilon)^2 \|C^T\|_\infty.$$

Take the saturation value M of q to be large enough so that $\Omega < 1$. Define

$$\tau = \frac{(M + 1/2)^2 \lambda_{min}(Q) - \Psi^2(\Theta\sqrt{n} + \epsilon)^2 \lambda_{max}(Q)}{\Psi^2(\Theta\sqrt{n} + \epsilon)\lambda_{min}(D)\epsilon}.$$

Let $\epsilon > 0$ be such that $\|e^{At}\|_\infty < 1+\epsilon$ and $\|e^{A^Tt}\|_\infty < 1+\epsilon$ for all $t \in [0,\tau]$. If we let $u(t) = 0$ and $\Delta(t) = \lfloor t/\tau \rfloor \tau e^{\|A\| \lfloor t/\tau \rfloor \tau}$, then $\exists k_0 > 0$ such that

$$\|q(y(t))\| \leq \Big(M + \frac{1}{2}\Big)\sqrt{\frac{\lambda_{min}(Q)}{\lambda_{max}(Q)}} - \frac{\sqrt{n}}{2} \qquad \text{for all } t \in [k_0\tau, k_0\tau + \tau].$$

We have

$$\int_{k_0\tau}^{k_0\tau+\tau} e^{A^T(t-k_0\tau)} C^T y(t) dt = W x(k_0\tau)$$

where the *observability Gramian*

$$W = \int_0^\tau e^{A^Tt} C^T C e^{At} dt$$

is a matrix of rank n. Define

$$\bar{x}(k_0\tau) = W^{-1} \int_{k_0\tau}^{k_0\tau+\tau} e^{A^T(t-k_0\tau)} C^T \Delta(k_0\tau) q(y(t)) dt.$$

We have

$$\|x(k_0\tau) - \bar{x}(k_0\tau)\|_\infty \leq \|W^{-1}\|_\infty \tau (1+\epsilon) \|C^T\|_\infty \Delta(k_0\tau)/2.$$

Denoting $e^{A\tau}\bar{x}(k_0\tau)$ by $\bar{x}(k_0\tau + \tau)$, we obtain

$$\|x(k_0\tau + \tau) - \bar{x}(k_0\tau + \tau)\|_\infty \leq \Psi \Delta(k_0\tau)/2.$$

Let $u(k_0\tau + \tau) = -K\bar{x}(k_0\tau + \tau)$. Choose $\Delta(k_0\tau + \tau) \geq \Delta(k_0\tau)$ such that

$$\bar{x}(k_0\tau + \tau) \leq \Delta(k_0\tau + \tau)\Big(M + \frac{1}{2}\Big)\sqrt{\frac{\lambda_{min}(Q)}{\lambda_{max}(Q)}} - \Psi \Delta(k_0\tau)\frac{\sqrt{n}}{2}$$

and let $\Delta(t) = \Delta(k_0\tau + \tau)$ for $k_0\tau + \tau \leq t < k_0\tau + 2\tau$.

Now, take any $t_1 \in (k_0\tau, k_0\tau + \tau]$, and assume that $u(t)$ has already been defined for $t < t_1 + \tau$. We have

$$\int_{t_1}^{t_1+\tau} e^{A^T(t-t_1)} C^T y(t) dt = W x(t_1) + v_1$$

where v_1 is a known vector that depends on $u(t)$ for $t_1 \leq t < t_1 + \tau$. Letting

$$\bar{x}(t_1) = W^{-1}\left(\int_{t_1}^{t_1+\tau} e^{A^T(t-t_1)} C^T \Delta(t) q(y(t)) dt - v_1 \right)$$

we obtain

$$\|x(t_1) - \bar{x}(t_1)\|_\infty \leq \|W^{-1}\|_\infty \tau(1+\epsilon)\|C^T\|_\infty \Delta(k_0\tau + \tau)/2.$$

Denoting by $\bar{x}(t_1+\tau)$ the solution of (II.2) with $x(t_1) = \bar{x}(t_1)$ and $u(t) = -K\bar{x}(t)$ for $t_1 \leq t < t_1 + \tau$, we have

$$\|x(t_1+\tau) - \bar{x}(t_1+\tau)\|_\infty \leq \Psi\Delta(k_0\tau + \tau)/2.$$

We then let $u(t_1+\tau) = -K\bar{x}(t_1+\tau)$. We thus obtain a control function u such that for all $t \in [k_0\tau + \tau, k_0\tau + 2\tau]$ we have

$$\|u(t) + Kx(t)\|_\infty \leq \|K\|_\infty \Psi\Delta(k_0\tau + \tau)/2.$$

As in the proof of Theorem 1, we can show that

$$\|x(k_0\tau + 2\tau)\| \leq \Psi\Delta(k_0\tau + \tau)(\Theta\sqrt{n} + \epsilon)\sqrt{\frac{\lambda_{max}(Q)}{\lambda_{min}(Q)}}.$$

Thus we let $\Delta(t) = \Omega\Delta(k_0\tau+\tau)$ for $k_0\tau+2\tau \leq t < k_0\tau+3\tau$, $\Delta(t) = \Omega\Delta(k_0\tau+2\tau)$ for $k_0\tau + 3\tau \leq t < k_0\tau + 4\tau$, etc., which gives $x(t) \to 0$ as needed. $\square$

Remark. Once the state estimate $\bar{x}(k_0\tau+\tau)$ has been generated, we can in fact switch to the control policy described in Theorem 5 (since $\|x(k_0\tau+\tau)\| \le E_0$ for an appropriate value of E_0 that depends on $\bar{x}(k_0\tau+\tau)$, $\Delta(k_0\tau)$, and Ψ). Another possibility, suggested by A. Stephen Morse, is to replace $\bar{x}(t)$ by the output of a Luenberger dynamic observer [38] for $t \ge k_0\tau+\tau$.

The next theorem is a counterpart of Theorem 8 for the discrete-time linear system

$$\begin{aligned} x(k+1) &= Ax(k) + Bu(k), \\ y(k) &= Cx(k) \end{aligned} \tag{II.17}$$

Theorem 9. *Suppose that (A, C) is an observable pair, and that all eigenvalues of $A - BK$ lie inside the unit circle. Then there exists a control policy of the form*

$$\begin{aligned} \bar{x}(k) &= \sum_{i,j,l=0}^{n-1} F_{ijl}(q(y(k-i)), \Delta(k-j), u(k-l)), \\ \Delta(k+1) &= G(k, \bar{x}(k), q(x(k)), \Delta(k)), \\ u(k) &= H(k, \bar{x}(k)) \end{aligned}$$

where q is a uniform quantizer with sensitivity $\Delta(k)$, such that the solutions of the system

$$\begin{aligned} x(k+1) &= Ax(k) + BH(k, \bar{x}(k)), \quad x(0) \text{ arbitrary} \\ \Delta(k+1) &= G(k, \bar{x}(k), q(x(k)), \Delta(k)), \quad \Delta(0) = 0 \end{aligned}$$

approach 0 *as* $t \to +\infty$.

Proof. Let Q, D and Θ be as in the proof of Theorem 3. Define

$$\Omega = \left(\Psi\Theta\sqrt{\frac{n\lambda_{max}(Q)}{\lambda_{min}(Q)}} + \sqrt{n}\right)\sqrt{\frac{\lambda_{max}(Q)}{\lambda_{min}(Q)}}\left(M + \frac{1}{2}\right)^{-1}$$

where

$$\Psi = \|A^{n-1}\|_\infty \|W^{-1}\|_\infty.$$

Take the saturation value M of q to be large enough so that $\Omega < 1$. If we let $u(k) = 0$ and $\Delta(k) = k\|A\|^k$ for $k \in \mathbb{Z}_+$, then $\exists k_0 > 0$ such that

$$\|q(y(k))\| \leq \left(M + \frac{1}{2}\right)\sqrt{\frac{\lambda_{min}(Q)}{\lambda_{max}(Q)}} - \frac{\sqrt{n}}{2} \qquad \text{for } k_0 \leq k \leq k_0 + n - 1.$$

We have $(y(k_0); \ldots; y(k_0 + n - 1)) = Wx(k_0)$, where $W = (C; CA; \ldots; CA^{n-1})$ is a matrix of rank n (we use semicolons to indicate the vertical arrangement in a matrix or a vector). Define

$$\bar{x}(k_0) = W^{-1}(\Delta(k_0)q(y(k_0)); \ldots; \Delta(k_0 + n - 1)q(y(k_0 + n - 1)))$$

which gives

$$\|x(k_0) - \bar{x}(k_0)\|_\infty \leq \|W^{-1}\|_\infty \Delta(k_0 + n - 1)/2.$$

Now, if $\bar{x}(k_0 + n - 1)$ is the state of the system at time $k_0 + n - 1$ subject to the condition $x(k_0) = \bar{x}(k_0)$, then

$$\|x(k_0 + n - 1) - \bar{x}(k_0 + n - 1)\|_\infty \leq \Psi\Delta(k_0 + n - 1)/2.$$

Let $u(k_0+n-1) = -K\bar{x}(k_0+n-1)$. Choose $\Delta(k_0+n) \geq \Delta(k_0+n-1)$ such that

$$\begin{aligned}\|(A-BK)\bar{x}(k_0+n-1)\| \\ \leq \Delta(k_0+n)\Big(M+\frac{1}{2}\Big)\sqrt{\frac{\lambda_{min}(Q)}{\lambda_{max}(Q)}} - \|A\|\Psi\Delta(k_0+n-1)\frac{\sqrt{n}}{2}\end{aligned}$$

and let $\Delta(k) = \Delta(k_0+n)$ for $k \geq k_0+n$.

Similarly, $(y(k_0+1);\ldots;y(k_0+n)) = Wx(k_0+1)+v_1$, where v_1 is a known vector that depends on $u(k_0+1),\ldots,u(k_0+n-1)$. Define

$$\bar{x}(k_0+1) = W^{-1}[(\Delta(k_0+1)q(y(k_0+1));\ldots;\Delta(k_0+n)q(y(k_0+n))) - v_1]$$

which gives

$$\|x(k_0+1)-\bar{x}(k_0+1)\|_\infty \leq \|W^{-1}\|_\infty\Delta(k_0+n)/2.$$

Therefore, if $\bar{x}(k_0+n)$ is the state of the system at time k_0+n subject to the condition $x(k_0+1) = \bar{x}(k_0+1)$, then

$$\|x(k_0+n)-\bar{x}(k_0+n)\|_\infty \leq \Psi\Delta(k_0+n)/2.$$

We now let $u(k_0+n) = -K\bar{x}(k_0+n)$. Proceeding in this fashion, we obtain a control function u such that for all $k \geq k_0+n-1$ we have

$$\|u(k)+Kx(k)\|_\infty \leq \|K\|_\infty\Psi\Delta(k_0+n)/2.$$

The rest of the proof is essentially the same as that of Theorem 3. $\square$

Chapter III

SPECTRAL THEORY OF FOKKER-PLANCK OPERATORS

1. Introduction. In this chapter we investigate the spectral properties of certain families of linear second-order differential operators arising from Fokker-Planck operators associated with linear stochastic differential equations. These operators are defined on a dense subspace of $L^2(\mathbb{R}^n)$. To clarify the exposition, we start with the one-dimensional case. Motivated by the explicit form of the steady-state probability density, we modify the classical Hermite polynomials by introducing one additional parameter σ (in our context, σ corresponds to the steady-state variance). This construction leads to an orthonormal basis for $L^2(\mathbb{R})$ with respect to which the operators take a particularly transparent form. The representation thus obtained allows us to compute their eigenvalues and eigenfunctions directly. We then show that many of the features of this analysis carry over to the multidimensional case and enable us to obtain information about eigenvalues of Fokker-Planck operators in a more general setting.

Moreover, we observe that the operator families parameterized by σ can be described by differential equations on an operator Lie algebra which take the so-called *double bracket form* $\dfrac{dL}{d\sigma} = [L, [L, M]]$. The study of differential equations in the double bracket form on finite-dimensional Lie algebras was initiated in [2] and [11] in connection with integrable gradient flows and numerical algorithms. It was shown, in particular, that such equations give rise to isospectral flows. In this chapter we present what seems to be a new framework in which double bracket equations appear. We show that the corresponding flows on an operator Lie algebra preserve at least part of the spectrum (actually, the entire spectrum in the case when the operators are self-adjoint). We also demonstrate that this fact is supported by probabilistic intuition.

This chapter is organized as follows. In Section 2 we construct an orthonormal basis in $L^2(\mathbb{R})$ using generalized Hermite polynomials. In Section 3 we introduce the Fokker-Planck operators and compute their eigenvalues. In Section 4 we use the gauge transformation technique to arrive at a family of self-adjoint operators and then undertake a similar treatment of those, discovering that the analysis is simplified significantly. In Section 5 we discuss double bracket differential equations on an operator Lie algebra and indicate connections with some known results on completely integrable gradient flows. Finally, in Section 6 we study Fokker-Planck operators in the multidimensional case, giving generalizations of the results obtained for scalar equations.

2. Orthonormal bases in $L^2(\mathbb{R})$. It is well known [29, p.121] that the *Hermite functions*

$$u_k(x) = h_k(x)e^{-x^2/2}, \quad k = 0, 1, \ldots \tag{III.1}$$

where $h_k(x) = \dfrac{1}{\pi^{1/4}\sqrt{2^k k!}}\, e^{x^2}\, \dfrac{d^k e^{-x^2}}{dx^k}$ are the Hermite polynomials, form an orthonormal basis for $L^2(\mathbb{R})$. We consider here the *generalized Hermite polynomials*

$$h_k(x,\sigma) = \frac{\sqrt{\sigma^k}}{(\sigma\pi)^{1/4}\sqrt{2^k k!}} e^{x^2/\sigma} \frac{d^k e^{-x^2/\sigma}}{dx^k}, \qquad k = 0, 1, \dots$$

where $\sigma > 0$ is a real parameter, and introduce the *generalized Hermite functions*

$$u_k(x,\sigma) = h_k(x,\sigma)e^{-x^2/2\sigma} = c_k(\sigma)e^{x^2/2\sigma} \frac{d^k e^{-x^2/\sigma}}{dx^k} \tag{III.2}$$

with constants $c_k(\sigma)$ given by the relations

$$c_k(\sigma) = \frac{\sqrt{\sigma^k}}{(\sigma\pi)^{1/4}\sqrt{2^k k!}}. \tag{III.3}$$

The functions (III.2) reduce to those given by (III.1) for $\sigma = 1$. Various modifications of the classical Hermite polynomials, analogous to (and more general than) the one considered here, can be found in the literature [17], [25].

Lemma 1. *For any $\sigma > 0$, the functions (III.2) form an orthonormal basis for $L^2(\mathbb{R})$.*

Proof. Consider the inner product

$$< u_k(x,\sigma), u_l(x,\sigma) >= c_k(\sigma)c_l(\sigma) \int_{-\infty}^{+\infty} e^{x^2/\sigma} \frac{d^k e^{-x^2/\sigma}}{dx^k} \frac{d^l e^{-x^2/\sigma}}{dx^l} dx.$$

Let us make the change of variable $x = \sqrt{\sigma}\, y$. We have $x^2/\sigma = y^2$, $dx = \sqrt{\sigma} dy$, $\dfrac{d^k e^{-x^2/\sigma}}{dx^k} = \dfrac{1}{(\sqrt{\sigma})^k} \dfrac{d^k e^{-y^2}}{dy^k}$, so that the above inner product becomes

$$< u_k(x,\sigma), u_l(x,\sigma) >= \frac{\sqrt{\sigma}}{\sqrt{\sigma^{k+l}}} c_k(\sigma)c_l(\sigma) \int_{-\infty}^{+\infty} e^{y^2} \frac{d^k e^{-y^2}}{dy^k} \frac{d^l e^{-y^2}}{dy^l} dy.$$

It is easy to see now that all the terms containing σ cancel out by virtue of (III.3). The statement of the lemma follows from the fact that the functions (III.1) form an orthonormal basis for $L^2(\mathbb{R})$. $\square$

3. Fokker-Planck operators. Let us consider the linear stochastic differential equation in the Itô sense

$$dx = -ax dt + b dw, \quad a > 0 \tag{III.4}$$

where $x \in \mathbb{R}$ and w is a standard Wiener process. The equation for the steady-state probability density that corresponds to (III.4) is $L(a,b)\rho(x) = 0$, where

$$L(a,b)\rho = \frac{b^2}{2}\rho_{xx} + ax\rho_x + a\rho \tag{III.5}$$

and ρ_x and ρ_{xx} denote the first and the second derivatives of ρ, respectively. The operator $L(a,b)$ is the Fokker-Planck operator associated with (III.4), whose domain will be specified below. Define

$$\sigma = \frac{b^2}{2a}.$$

The steady-state probability density is then given by $\rho(x) = Ne^{-x^2/2\sigma}$, where $N > 0$ is a normalization constant. Dividing the Fokker-Planck operator $L(a,b)$ by a, we are led to studying a one-parameter family of differential operators, L_σ, defined by

$$L_\sigma \rho = \frac{1}{a}L(a, \sqrt{2a\sigma})\rho = \sigma\rho_{xx} + x\rho_x + \rho\,, \quad \sigma > 0. \tag{III.6}$$

More precisely, it is easy to see that L_σ is well defined on $C^2(\mathbb{R})$ by (III.6) and $L_\sigma u_k(x,\sigma) \in L^2(\mathbb{R})$ for each k. We then define L_σ to be the minimal closed linear operator such that $L_\sigma\rho$ is given by (III.6) whenever $\rho \in C^2(\mathbb{R})$. We thus obtain an operator $L_\sigma \colon \mathcal{D}_{L_\sigma} \to L^2(\mathbb{R})$, where $\mathcal{D}_{L_\sigma}$ is a dense subspace of $L^2(\mathbb{R})$. Throughout the chapter, unless specified otherwise, all differential operators are to be interpreted in the above sense.

We have already mentioned that $L_\sigma u_0(x,\sigma) = L_\sigma c_0 e^{-x^2/2\sigma} = 0$, i.e., $e^{-x^2/2\sigma}$ is an eigenfunction with the eigenvalue zero. To investigate the spectrum of the operator L_σ, it seems natural to use the basis given by (III.2) (with the same value of σ).

Theorem 2. *The spectrum of the operator $L_\sigma \colon \mathcal{D}_{L_\sigma} \to L^2(\mathbb{R})$ is independent of σ. For any $\sigma > 0$, the eigenvalues of L_σ are all numbers in the half-plane $\{\lambda \in \mathbb{C} : \operatorname{Re}\lambda < 1/2\}$.*

Proof. We have

$$
\begin{aligned}
L_\sigma u_k(x,\sigma) &= c_k(\sigma)\Big[\sigma \frac{d^2}{dx^2}\Big(e^{x^2/2\sigma}\frac{d^k e^{-x^2/\sigma}}{dx^k}\Big) + \frac{d}{dx}\Big(x e^{x^2/2\sigma}\frac{d^k e^{-x^2/\sigma}}{dx^k}\Big)\Big] \\
&= c_k(\sigma)\Big[\sigma \frac{d}{dx}\Big(\frac{x}{\sigma} e^{x^2/2\sigma}\frac{d^k e^{-x^2/\sigma}}{dx^k}\Big) + \sigma \frac{d}{dx} e^{x^2/2\sigma}\frac{d^{k+1} e^{-x^2/\sigma}}{dx^{k+1}} \\
&\quad + \frac{d}{dx}\Big(x e^{x^2/2\sigma}\frac{d^k e^{-x^2/\sigma}}{dx^k}\Big)\Big] \\
&= c_k(\sigma)\Big[2\frac{d}{dx}\Big(x e^{x^2/2\sigma}\frac{d^k e^{-x^2/\sigma}}{dx^k}\Big) + \sigma\frac{d}{dx}\Big(e^{x^2/2\sigma}\frac{d^{k+1} e^{-x^2/\sigma}}{dx^{k+1}}\Big)\Big] \\
&= c_k(\sigma)\Big[2 e^{x^2/2\sigma}\frac{d^k e^{-x^2/\sigma}}{dx^k} + \frac{2x^2}{\sigma} e^{x^2/2\sigma}\frac{d^k e^{-x^2/\sigma}}{dx^k} \\
&\quad + 2x e^{x^2/2\sigma}\frac{d^{k+1} e^{-x^2/\sigma}}{dx^{k+1}} + x e^{x^2/2\sigma}\frac{d^{k+1} e^{-x^2/\sigma}}{dx^{k+1}} + \sigma e^{x^2/2\sigma}\frac{d^{k+2} e^{-x^2/\sigma}}{dx^{k+2}}\Big] \\
&= c_k(\sigma)\Big[\Big(2 + \frac{2x^2}{\sigma}\Big) e^{x^2/2\sigma}\frac{d^k e^{-x^2/\sigma}}{dx^k} + 3x e^{x^2/2\sigma}\frac{d^{k+1} e^{-x^2/\sigma}}{dx^{k+1}} \\
&\quad + \sigma e^{x^2/2\sigma}\frac{d^{k+2} e^{-x^2/\sigma}}{dx^{k+2}}\Big].
\end{aligned}
\tag{III.7}
$$

We introduce the notation $d_k(x,\sigma) = e^{x^2/2\sigma}\dfrac{d^k e^{-x^2/\sigma}}{dx^k}$, so that (III.7) becomes

$$L_\sigma u_k = c_k\Big(\Big(2 + \frac{2x^2}{\sigma}\Big)d_k + 3x d_{k+1} + \sigma d_{k+2}\Big). \tag{III.8}$$

To obtain recurrence relations on d_k, notice that

$$d_{k+1} = e^{x^2/2\sigma}\frac{d^k}{dx^k}\frac{de^{-x^2/\sigma}}{dx} = e^{x^2/2\sigma}\frac{d^k}{dx^k}\Big(-\frac{2x}{\sigma}e^{-x^2/\sigma}\Big)$$

which by Newton's binomial formula becomes

$$d_{k+1} = e^{x^2/2\sigma}\Big(-\frac{2x}{\sigma}\frac{d^k e^{-x^2/\sigma}}{dx^k} - \frac{2k}{\sigma}\frac{d^{k-1}e^{-x^2/\sigma}}{dx^{k-1}}\Big)$$

or, in the new notation,

$$d_{k+1} = -\frac{2x}{\sigma}d_k - \frac{2k}{\sigma}d_{k-1}\,. \tag{III.9}$$

From (III.9) we obtain

$$x d_k = -\frac{\sigma}{2}d_{k+1} - k d_{k-1}, \quad k = 1, 2, \ldots \tag{III.10}$$

and also (multiplying both sides of (III.9) by x and then using (III.10))

$$\begin{aligned}\frac{2x^2}{\sigma}d_k &= -x d_{k+1} - \frac{2k}{\sigma}x d_{k-1}\\ &= \frac{\sigma}{2}d_{k+2} + (k+1)d_k + k d_k + \frac{2}{\sigma}k(k-1)d_{k-2}\\ &= (2k+1)d_k + \frac{\sigma}{2}d_{k+2} + \frac{2k(k-1)}{\sigma}d_{k-2}. \end{aligned}\tag{III.11}$$

Combining (III.8)–(III.11) gives

$$\begin{aligned}L_\sigma u_k &= c_k\Big(2d_k + (2k+1)d_k + \frac{\sigma}{2}d_{k+2} + \frac{2k(k-1)}{\sigma}d_{k-2}\\ &\quad - \frac{3\sigma}{2}d_{k+2} - 3(k+1)d_k + \sigma d_{k+2}\Big)\\ &= c_k\Big(-k d_k + \frac{2k(k-1)}{\sigma}d_{k-2}\Big)\end{aligned}\tag{III.12}$$

and we see that the terms containing d_{k+2} disappear. Moreover, notice that since

$$c_k = \frac{\sqrt{\sigma^k}}{(\sigma\pi)^{1/4}\sqrt{2^k k!}}$$

and

$$c_{k-2} = \frac{\sqrt{\sigma^{k-2}}}{(\sigma\pi)^{1/4}\sqrt{2^{k-2}(k-2)!}}$$

we have

$$\frac{2k(k-1)}{\sigma}\, c_k = \sqrt{k(k-1)}\, c_{k-2}. \tag{III.13}$$

The formulae (III.12) and (III.13) imply that with respect to the basis (III.2) the operator L_σ takes the upper triangular form as given by

$$L_\sigma u_k(x,\sigma) = -k u_k(x,\sigma) + \sqrt{k(k-1)}u_{k-2}(x,\sigma). \tag{III.14}$$

From (III.14) it immediately follows that the spectrum of L_σ is independent of σ. Moreover, it is easy to see that the nonpositive integers are eigenvalues of L_σ. The corresponding eigenfunctions are finite linear combinations of the basis elements u_k and thus belong to $C^\infty(\mathbb{R}) \cap L^2(\mathbb{R})$. They do not, however, form a complete set of eigenfunctions. The formula (III.14) implies that the existence of an eigenfunction of L_σ with an eigenvalue λ is equivalent to the absolute convergence of at least one of the series

$$\sum_{n=1}^{+\infty} \frac{\lambda^2(\lambda+2)^2\cdots(\lambda+2n-2)^2}{(2n)!}$$

and

$$\sum_{n=1}^{+\infty} \frac{(\lambda+1)^2(\lambda+3)^2\cdots(\lambda+2n-1)^2}{(2n+1)!}.$$

Using Gauss' test for convergence (see, e.g., [33]), one can show in a straightforward manner that each series converges if $\mathrm{Re}\lambda < 1/2$ and diverges if $\mathrm{Re}\lambda \geq 1/2$. □

We can gain more insight into the spectral properties of the operator L_σ from its probabilistic interpretation. Recall that L_σ was defined in terms of the Fokker-Planck operator $L(a, b)$ associated with the equation (III.4) via the formula (III.6). It follows from Theorem 2 that the eigenvalues of $L(a, b)$ are all numbers in the half-plane $\{\lambda \in \mathbb{C} : \mathrm{Re}\,\lambda < a/2\}$. However, it is well known that any eigenfunction of $L(a, b)$ that is nonnegative and belongs to $L^1(\mathbb{R})$ must be proportional to $e^{-x^2/2\sigma}$ (which is the steady-state probability density that corresponds to the eigenvalue zero).

The fact that the spectrum of $L(a, b)$ does not depend on the noise coefficient b should not be surprising if we notice that we can change b by simply rescaling x; i.e., substituting $y = px$ in (III.4) for an arbitrary $p \in \mathbb{R}$ gives $\dot{y} = -ay + pb\dot{w}$. Thus Theorem 2 states, in particular, that the spectrum of the Fokker-Planck operator associated with (III.4) is not affected by linear changes of variable.

4. Self-adjoint operators. The analysis of the operators L_σ is complicated by the fact that they are not self-adjoint. There is a standard technique that consists in applying a gauge transformation to convert our operators to self-adjoint ones. In our case, write $\rho = vf$, where the function v is to be fixed. We have

$$L_\sigma(vf) = \sigma v_{xx} f + 2\sigma v_x f_x + \sigma v f_{xx} + x v_x f + x v f_x + v f.$$

We see that in order for the first-order derivatives to disappear, v must satisfy the equation $v_x = -\dfrac{x}{2\sigma}v$. Letting

$$v = e^{-x^2/4\sigma}$$

we obtain $v^{-1}L_\sigma(vf) = \sigma f_{xx} + (\frac{1}{2} - \frac{x^2}{4\sigma})f$.

Motivated by the above discussion, we define a new operator family, T_σ, by the formula

$$T_\sigma \rho = \sigma \rho_{xx} + \Big(\frac{1}{2} - \frac{x^2}{4\sigma}\Big)\rho, \quad \sigma > 0. \tag{III.15}$$

For any positive σ, the operator T_σ is closed and self-adjoint, its domain being a dense subspace $\mathcal{D}_{T_\sigma}$ of $L^2(\mathbb{R})$ (defined as in Section 3).

It is well known that Hermite polynomials appear frequently in expressions for eigenfunctions of self-adjoint linear second-order differential operators. The next theorem shows that the eigenfunctions of T_σ are given by the generalized Hermite functions (III.2). For $\sigma = 1/2$, the statement reduces to a standard result on the classical Hermite functions [1, p.256].

Theorem 3. *For any $\sigma > 0$, the spectrum of the operator $T_\sigma \colon \mathcal{D}_{T_\sigma} \to L^2(\mathbb{R})$ consists of the nonpositive integers, all of which are eigenvalues. The corresponding eigenfunctions are the functions $u_k(x, 2\sigma)$, i.e., $T_\sigma u_k(x, 2\sigma) = -k\, u_k(x, 2\sigma)$.*

Proof. For $\rho = e^{x^2/4\sigma}\dfrac{d^k e^{-x^2/2\sigma}}{dx^k}$ we have

$$\begin{aligned}
T_\sigma\rho &= \sigma\frac{d}{dx}\Big(\frac{x}{2\sigma}e^{x^2/4\sigma}\frac{d^k e^{-x^2/2\sigma}}{dx^k} + e^{x^2/4\sigma}\frac{d^{k+1}e^{-x^2/2\sigma}}{dx^{k+1}}\Big)\\
&+\Big(\frac{1}{2}-\frac{x^2}{4\sigma}\Big)e^{x^2/4\sigma}\frac{d^k e^{-x^2/2\sigma}}{dx^k}\\
&=\sigma\Big(\frac{1}{2\sigma}e^{x^2/4\sigma}\frac{d^k e^{-x^2/2\sigma}}{dx^k} + \frac{x^2}{4\sigma^2}e^{x^2/4\sigma}\frac{d^k e^{-x^2/2\sigma}}{dx^k}\\
&+\frac{x}{\sigma}e^{x^2/4\sigma}\frac{d^{k+1}e^{-x^2/2\sigma}}{dx^{k+1}} + e^{x^2/4\sigma}\frac{d^{k+2}e^{-x^2/2\sigma}}{dx^{k+2}}\Big)\\
&+\Big(\frac{1}{2}-\frac{x^2}{4\sigma}\Big)e^{x^2/4\sigma}\frac{d^k e^{-x^2/2\sigma}}{dx^k}\\
&=e^{x^2/4\sigma}\frac{d^k e^{-x^2/2\sigma}}{dx^k} + xe^{x^2/4\sigma}\frac{d^{k+1}e^{-x^2/2\sigma}}{dx^{k+1}} + \sigma e^{x^2/4\sigma}\frac{d^{k+2}e^{-x^2/2\sigma}}{dx^{k+2}}
\end{aligned}$$

which in our previous notation becomes

$$T_\sigma d_k(x,2\sigma) = d_k(x,2\sigma) + xd_{k+1}(x,2\sigma) + \sigma d_{k+2}(x,2\sigma). \tag{III.16}$$

Replacing σ by 2σ in (III.10) and substituting into (III.16), we arrive at

$$\begin{aligned}
T_\sigma d_k(x,2\sigma) &= d_k(x,2\sigma) - \sigma d_{k+2}(x,2\sigma)\\
&- (k+1)d_k(x,2\sigma) + \sigma d_{k+2}(x,2\sigma)\\
&= -k\, d_k(x,2\sigma).
\end{aligned}$$

From this the statement of the theorem is an easy step. □

As a consequence, each function of the form $\alpha e^{-x^2/4\sigma}u_k(x,2\sigma) = \alpha c_k\dfrac{d^k e^{-x^2/2\sigma}}{dx^k}$ with $\alpha\in\mathbb{R}$ is an eigenfunction of the original operator L_σ that corresponds to the eigenvalue $-k$. The rest of the eigenfunctions of L_σ belong to the space $\{\rho : e^{x^2/4\sigma}\rho \notin L^2(\mathbb{R})\}$.

5. Double bracket equations. Let $A_2(\mathbb{R})$ be the space of functions $\rho\colon \mathbb{R} \to \mathbb{R}$ such that ρ_x exists and is absolutely continuous. Consider the operators P_1, P_2, P_3, and P_4 acting on the space

$$\mathcal{D} = \{\rho \in L^2(\mathbb{R}) \cap A_2(\mathbb{R}) : \rho_{xx}, x\rho_x, x^2\rho \in L^2(\mathbb{R})\}$$

that are defined as follows:

$$P_1\rho = \rho_{xx}, \quad P_2\rho = x\rho_x, \quad P_3\rho = x^2\rho, \quad P_4\rho = \rho.$$

It is easy to verify that the linear span of the above operators is closed under commutation with respect to the usual Lie bracket $[P_i, P_j] = P_iP_j - P_jP_i$. We will let $\mathfrak{g}$ denote the operator Lie algebra spanned by P_i, $i = 1, 2, 3, 4$. Such Lie algebras and their representations have been studied in the context of quantum mechanics and, more recently, estimation theory [8].

Observe that L_σ and T_σ can be realized as operators in $\mathfrak{g}$. More precisely, let us denote by $L(\sigma)$ and $T(\sigma)$ the restrictions of L_σ and T_σ to $\mathcal{D}$. Theorem 3 implies that $T(\sigma)$, $0 < \sigma < +\infty$, is an isospectral family of operators in $\mathfrak{g}$. In fact, for any $\sigma > 0$ the spectrum of $T(\sigma)$ consists of eigenvalues which are the nonpositive integers. As we know from Theorem 2, the nonpositive integers are also eigenvalues of $L(\sigma)$ for each $\sigma > 0$ (because the corresponding eigenfunctions of L_σ belong to $\mathcal{D}$).

In this section we show that the families of operators $L(\sigma)$ and $T(\sigma)$ correspond to integral curves of differential equations in the double bracket form on $\mathfrak{g}$.

Proposition 4. *Let N be an operator in $\mathfrak{g}$ defined by $N\rho = \frac{1}{2}\rho_{xx}$. Then $T(\sigma)$, $0 < \sigma < +\infty$, is a solution of the differential equation*

$$\frac{dT}{d\sigma} = [T, [T, N]]. \tag{III.17}$$

Proof. On one hand,

$$\frac{dT(\sigma)}{d\sigma}\rho = \rho_{xx} + \frac{x^2}{4\sigma^2}\rho.$$

On the other hand,

$$\begin{aligned}
[T(\sigma), N]\rho = T(\sigma)N\rho - NT(\sigma)\rho &= \sigma\frac{\partial^2}{\partial x^2}\Big(\frac{1}{2}\rho_{xx}\Big) \\
&+ \Big(\frac{1}{4} - \frac{x^2}{8\sigma}\Big)\rho_{xx} - \frac{1}{2}\frac{\partial^2}{\partial x^2}\Big(\sigma\rho_{xx} + \Big(\frac{1}{2} - \frac{x^2}{4\sigma}\Big)\rho\Big) \\
&= \frac{x}{2\sigma}\rho_x + \frac{1}{4\sigma}\rho.
\end{aligned}$$

Further,

$$\begin{aligned}
[T(\sigma), [T(\sigma), N]]\rho = \sigma\frac{\partial^2}{\partial x^2}\Big(\frac{x}{2\sigma}\rho_x + \frac{\rho}{4\sigma}\Big) &+ \Big(\frac{1}{2} - \frac{x^2}{4\sigma}\Big)\Big(\frac{x}{2\sigma}\rho_x + \frac{\rho}{4\sigma}\Big) \\
- \frac{x}{2\sigma}\frac{\partial}{\partial x}\Big(\sigma\rho_{xx} + \Big(\frac{1}{2} - \frac{x^2}{4\sigma}\Big)\rho\Big) &- \frac{1}{4\sigma}\Big(\sigma\rho_{xx} + \Big(\frac{1}{2} - \frac{x^2}{4\sigma}\Big)\rho\Big) \\
= \rho_{xx} + \frac{x^2}{4\sigma^2}\rho &
\end{aligned}$$

which proves (III.17). □

In the same direct fashion one can establish an analogous property of the operators $L(\sigma)$.

Proposition 5. *Let M be an operator in $\mathfrak{g}$ defined by $M\rho = \frac{1}{4}\rho_{xx}$. Then $L(\sigma)$, $0 < \sigma < +\infty$, is a solution of the differential equation*

$$\frac{dL}{d\sigma} = [L, [L, M]]. \tag{III.18}$$

To each of the operators $L(\sigma)$ there corresponds the steady-state probability density ρ_σ which satisfies the equation $L(\sigma)\rho_\sigma(x) = 0$. The flow (III.18) on the operator Lie algebra $\mathfrak{g}$ thus induces a flow on the manifold of gaussian probability densities. For example, making the change of variable $\sigma = e^t$, we obtain a particular case of the gradient flow of gaussians described by Nakamura in [32].

We point out an interesting analogy between the results of Propositions 4 and 5 and the sorting algorithms described in [11]. If N is a real diagonal matrix with unrepeated eigenvalues, and if $H(0)$ is a suitably chosen symmetric matrix, then the solution of the double bracket equation $\dot{H} = [H, [H, N]]$ approaches a diagonal matrix $H(\infty)$ such that the diagonal elements of $H(\infty)$ and N are similarly ordered; since $H(\infty)$ is diagonal, it commutes with N. For large positive values of σ, the "principal term" of the operators $L(\sigma)$ and $T(\sigma)$ is $\sigma \dfrac{d^2}{dx^2}$, which is proportional to both M and N and thus commutes with them. This "operator sorting" property of the double bracket equations (III.17) and (III.18) is to be investigated further.

In a recent paper [21], a hierarchy of double bracket equations on a space of (formal) pseudo-differential operators is introduced and studied. It is shown that some of the algebraic properties of these equations are similar to those of Lax equations arising in KP theory.

6. Fokker-Planck operators in $L^2(\mathbb{R}^n)$. Consider the system of linear stochastic differential equations

$$dx = Axdt + Bdw\,, \quad x \in \mathbb{R}^n,\ w = \begin{pmatrix} w_1 \\ w_2 \\ \cdots \\ w_m \end{pmatrix} \tag{III.19}$$

where each w_i is a standard Wiener process, and A and B are matrices of suitable dimensions. Recall that *separable functions*, i.e., functions that can be expressed as products $\rho_1(x_1)\cdots\rho_n(x_n)$, span a dense subspace of $L^2(\mathbb{R}^n)$. Thus we can construct an orthonormal basis for $L^2(\mathbb{R}^n)$ by taking products of the functions (III.2) for each variable. The analysis of the previous sections now directly generalizes to those linear stochastic systems in $\mathbb{R}^n$ whose equations are completely decoupled. In this case, the Fokker-Planck operator decomposes into a sum of Fokker-Planck operators of the kind considered above for each variable. The results of Section 3 then imply, in particular, that the sums of the eigenvalues of the matrix A are eigenvalues of the corresponding Fokker-Planck operator, and that the corresponding eigenfunctions belong to the space $C^\infty(\mathbb{R}^n) \cap L^2(\mathbb{R}^n)$. Although the analysis in the general multidimensional case is more complicated than in the scalar case, results that parallel most of our earlier developments can be obtained.

As in the scalar case, the spectrum of the Fokker-Planck operator associated with (III.19) is invariant under linear coordinate transformations. Indeed, let P be a nonsingular $n \times n$ matrix, and consider the linear change of coordinates $y = Px$. Applying the Itô rule (see Chapter I), we obtain the system

$$dy = PAP^{-1}y\,dt + PB\,dw. \tag{III.20}$$

Let us denote the Fokker-Planck operators associated with (III.19) and (III.20) by L_n and $\overline{L}_n$, respectively. We will consider them as being closed operators defined on dense subspaces $\mathcal{D}_{L_n}$ and $\mathcal{D}_{\overline{L}_n}$ of $L^2(\mathbb{R}^n)$ (cf. Section 3).

Proposition 6. *The following diagram is commutative:*

$$\begin{tikzcd}
\mathcal{D}_{L_n} \arrow[r, "L_n"] \arrow[d, "y=Px"'] & L^2(\mathbb{R}^n) \arrow[d, "y=Px"] \\
\mathcal{D}_{\overline{L}_n} \arrow[r, "\overline{L}_n"] & L^2(\mathbb{R}^n)
\end{tikzcd}$$

Proof. We have the following expressions for L_n and $\overline{L}_n$:

$$L_n\rho = \frac{1}{2}\sum_{i,j=1}^{n}(BB^T)_{ij}\,\rho_{x_ix_j} - \sum_{i,j=1}^{n} A_{ij}x_j\rho_{x_i} - \mathrm{tr}A\cdot\rho, \tag{III.21}$$

$$\overline{L}_n\rho = \frac{1}{2}\sum_{i,j=1}^{n}(PBB^TP^T)_{ij}\,\rho_{y_iy_j} - \sum_{i,j=1}^{n}(PAP^{-1})_{ij}y_j\rho_{y_i} - \mathrm{tr}PAP^{-1}\cdot\rho. \tag{III.22}$$

Further,

$$\rho_{y_i} = \sum_{m=1}^{n}\rho_{x_m}(P^{-1})_{mi}$$

and

$$\rho_{y_iy_j} = \sum_{m,l=1}^{n}(P^{-1})_{mi}(P^{-1})_{lj}\,\rho_{x_mx_l}\,.$$

Now one can easily verify that (III.21) and (III.22) match term by term. $\square$

The same conclusion in fact holds for more general classes of systems, not necessarily linear. However, the linearity of the coordinate transformation in the above discussion is essential, since we relied on the absence of the second-order term in the Itô rule. From Proposition 6 it follows, in particular, that L_n and $\overline{L}_n$ have the same spectrum.

From this point on for the rest of the chapter, let us make the following two assumptions with regard to the system (III.19).

a) The eigenvalues of A have negative real parts.

b) (A,B) is a controllable pair.

Under these assumptions, the steady-state variance equation (I.13) associated with (III.19) has a positive definite solution Q. After an appropriate change of coordinates in $\mathbb{R}^n$ we can have $Q = \frac{1}{2}I$, so that $A = \Omega - BB^T$ with Ω skew-symmetric. By the above remarks, such a coordinate transformation does not change the eigenvalues of the Fokker-Planck operator L_n defined by (III.21). The steady-state probability density associated with (III.19) then becomes $\bar{\rho}(x) = Ne^{-x^T x}$, $N > 0$, and this is an eigenfunction that corresponds to the eigenvalue zero of the Fokker-Planck operator.

Next, let us determine all eigenfunctions of L_n that take the form

$$\rho(x) = (h_1 x_1 + \cdots + h_n x_n)\bar{\rho}(x) = h^T x \bar{\rho}(x), \quad h \in \mathbb{R}^n. \tag{III.23}$$

Lemma 7. *Suppose that $A = \Omega - BB^T$, where $\Omega = -\Omega^T$. Then the function (III.23) is an eigenfunction of the operator L_n with eigenvalue λ if and only if h is an eigenvector of the matrix A with the same eigenvalue λ.*

Proof. Let ρ be of the form (III.23). Taking into account that $L_n\bar{\rho} = 0$, we have

$$\begin{aligned}
L_n\rho &= \sum_{i,j=1}^{n} (BB^T)_{ij}(h)_i(-2x_j)\bar{\rho} - \sum_{i,j} A_{ij}x_j(h)_i\bar{\rho} \\
&= -\bar{\rho}\sum_{i,j=1}^{n} (A_{ji} + 2(BB^T)_{ji})(h)_j x_i \\
&= \bar{\rho}\sum_{i,j=1}^{n} (\Omega_{ij} - (BB^T)_{ij})(h)_j x_i = \sum_{i=1}^{n} (Ah)_i x_i\bar{\rho}
\end{aligned}$$

and this obviously equals $\lambda\rho = \lambda\sum_i h_i x_i\bar{\rho}$ if and only if $Ah = \lambda h$. $\square$

Denote by $h_1, \ldots, h_k$ the eigenvectors of A and by $\lambda_1, \ldots, \lambda_k$ the corresponding eigenvalues ($k \leq n$). Now, let us see how L_n acts on functions of the form

$$\rho(x) = \bar{\rho}(x)\prod_{m\in J} h_m^T x \tag{III.24}$$

where the product is taken over some index set J whose elements are (not necessarily distinct) positive integers no greater than n. We have

$$\begin{aligned}
L_n(\bar{\rho}\prod_{m\in J} h_m^T x) &= \sum_{i,j=1}^{n}\sum_{m,l\in J} (BB^T)_{ij}(h_m)_i(h_l)_j\bar{\rho}\prod_{p\in J\setminus\{m,l\}} h_p^T x \\
&+ \sum_{i,j=1}^{n}\sum_{m\in J} (BB^T)_{ij}(h_m)_i\bar{\rho}_{x_j}\prod_{p\in J\setminus\{m\}} h_p^T x \\
&- \sum_{i,j=1}^{n}\sum_{m\in J} A_{ij}x_j(h_m)_i\bar{\rho}\prod_{p\in J\setminus\{m\}} h_p^T x \\
&= \sum_{i,j=1}^{n}\sum_{m,l\in J} (BB^T)_{ij}(h_m)_i(h_l)_j\bar{\rho}\prod_{p\in J\setminus\{m,l\}} h_p^T x + \Big(\sum_{m\in J}\lambda_m\Big)\bar{\rho}\prod_{m\in J} h_m^T x.
\end{aligned}$$

In the above calculation we have used the key fact that

$$L_n(h_m^T x\bar{\rho}) = \sum_{i,j=1}^{n} (BB^T)_{ij}(h_m)_i\bar{\rho}_{x_j} - \sum_{i,j=1}^{n} A_{ij}x_j(h_m)_i\bar{\rho} = \lambda_m h_m^T x\bar{\rho}.$$

Thus functions of the form (III.24) for various index sets J form an invariant subspace under the action of L_n. It is not hard to see that $\sum_{m\in J}\lambda_m$ are eigenvalues of L_n. The corresponding eigenfunctions are finite linear combinations of functions of the form (III.24). Summarizing, we have the following theorem.

Theorem 8. *The sums of the eigenvalues of the matrix A are eigenvalues of the Fokker-Planck operator $L_n\colon \mathcal{D}_{L_n} \to L^2(\mathbb{R}^n)$. The corresponding eigenfunctions belong to the space $C^\infty(\mathbb{R}^n)\cap L^2(\mathbb{R}^n)$.*

Theorem 8 can probably be best appreciated in the following context. It is well known and easy to show that there are $N_n^p = \binom{n+p-1}{p}$ linearly independent monomials of degree p in n variables of the form $x_1^{p_1}\ldots x_n^{p_n}$, where $\sum_{i=1}^n p_i = p$ and $p_i \geq 0$. The linear differential equation

$$\dot{x} = Ax, \quad x\in\mathbb{R}^n$$

gives rise to the equation

$$\frac{d}{dt}x^{[p]} = A_{[p]}x^{[p]}, \quad x^{[p]} \in \mathbb{R}^{N_n^p}.$$

One of the basic properties of the matrix $A_{[p]}$ defined in this way is that its eigenvalues are the p-term sums of the eigenvalues of A. As is shown in [6], the matrices $A_{[p]}$ are directly related to the pth moment equations for the system (III.19).

Theorem 8 shows that the situation in the infinite-dimensional case is similar in the following sense. Associated with (III.19) we have the Fokker-Planck equation for the probability density

$$\frac{\partial \rho(t,x)}{\partial t} = L_n \rho(t,x).$$

The operator L_n is well defined on a dense subspace of $L^2(\mathbb{R}^n)$. We know that the basis elements in $L^2(\mathbb{R}^n)$ can be taken to be polynomials of an arbitrary degree multiplied by gaussians, and we have shown that the sums (with an arbitrary number of terms) of the eigenvalues of A are eigenvalues of the operator L_n.

In view of the results of Section 4, it would be interesting to obtain conditions under which it is possible to convert the Fokker-Planck operator L_n associated with (III.19) to a self-adjoint operator by means of an appropriate gauge transformation. The following theorem provides such conditions, as well as an explicit formula for the function v to be used.

Proposition 9. *Suppose that the matrix B is nondegenerate and that we have*

$$ABB^T = BB^T A^T. \tag{III.25}$$

If the function v is defined by the formula

$$v = e^{x^T (BB^T)^{-1} Ax/2} \tag{III.26}$$

then the operator T_n given by

$$T_n \rho = v^{-1} L_n (v\rho)$$

is self-adjoint.

Proof. The first-order terms in the expression for T_n are

$$\frac{1}{2}\sum_{j,k=1}^{n}(BB^T)_{jk}(v_{x_j}\frac{\partial}{\partial x_k}+v_{x_k}\frac{\partial}{\partial x_j})-\sum_{i,j=1}^{n}A_{ij}x_j v\frac{\partial}{\partial x_i}.$$

We see that the coefficient of $\dfrac{\partial}{\partial x_i}$ is

$$\sum_{j=1}^{n}(BB^T)_{ij}v_{x_j}-\sum_{j=1}^{n}A_{ij}x_j v$$

and we need this to be zero for each i. This is equivalent to having

$$(BB^T)\operatorname{grad} v = Ax\, v$$

or

$$\operatorname{grad} v = (BB^T)^{-1}Ax\, v. \tag{III.27}$$

Therefore, we must have

$$\begin{aligned} v_{x_i x_j} &= \frac{\partial}{\partial x_i}\Big[\sum_{k=1}^{n}((BB^T)^{-1}A)_{jk}x_k\, v\Big] \\ &= ((BB^T)^{-1}A)_{ji}\, v + \sum_{k=1}^{n}((BB^T)^{-1}A)_{jk}x_k\, v_{x_i} \\ &= ((BB^T)^{-1}A)_{ji}\, v + \sum_{k=1}^{n}((BB^T)^{-1}A)_{jk}x_k \sum_{l=1}^{n}((BB^T)^{-1}A)_{il}x_l\, v. \end{aligned}$$

The compatibility conditions $v_{x_i x_j} = v_{x_j x_i}$ now imply that the matrix $(BB^T)^{-1}A$ has to be symmetric:

$$(BB^T)^{-1}A = A^T(BB^T)^{-1}.$$

Multiplying both sides of this formula by BB^T, we arrive at (III.25). It is straighforward to show that the function given by (III.26) satisfies the equation (III.27). $\square$

Let us switch to coordinates in which $Q = \sigma I$ for some $\sigma > 0$ (in the language of statistical thermodynamics, these are coordinates in which the *equipartition of energy* property holds, and σ is the *temperature* of the system—cf. Chapter IV, Section 5). Then $(A + A^T)\sigma = -BB^T$, and (III.25) can be rewritten as $A^2 = (A^T)^2$. This last condition is satisfied, for example, if A is symmetric. In this case, (III.26) becomes $v = e^{-x^T x/4\sigma}$, which is a constant multiple of the square root of the steady-state probability density. This is in accordance with our earlier results for the one-dimensional case.

Let us denote by ∇ the gradient with respect to the metric on $\mathbb{R}^n$ given by $G = (BB^T)^{-1}$. In other words, given a function $\phi\colon \mathbb{R}^n \to \mathbb{R}$, we define the vector $\nabla\phi$ by $(\nabla\phi)_i = \sum_{j=1}^n (BB^T)_{ij}\phi_{x_j}$. Assume that $Q = \sigma I$ and $A = A^T$, and let $\phi(x) = \frac{1}{4\sigma}x^T x$. Then we have

$$(\nabla\phi)_i = 2\sigma \sum_{j=1}^n A_{ij}\phi_{x_j} = \sum_{j=1}^n A_{ij}x_j$$

so that the system (III.19) can be rewritten as

$$dx = -\nabla\phi(x)\,dt + B\,dw. \tag{III.28}$$

Systems of the general form (III.28) and the corresponding steady-state probability densities were studied in [15] (see also Chapter IV). Under the present assumptions, the Fokker-Planck operator associated with (III.19) takes the form

$$L_{n,\sigma}\rho = -\sigma \sum_{i,j=1}^n A_{ij}\rho_{x_i x_j} - \sum_{i,j=1}^n A_{ij}x_j\rho_{x_i} - \mathrm{tr}A \cdot \rho.$$

Using Proposition 9, we can also construct a self-adjoint operator which in this case is given by

$$T_{n,\sigma}\rho = -\sigma \sum_{i,j=1}^{n} A_{ij}\rho_{x_i x_j} - \Big(\frac{1}{2}\,\mathrm{tr}A - \frac{1}{4\sigma}\sum_{i,j=1}^{n} A_{ij}x_i x_j\Big)\rho.$$

As we did throughout the chapter, we will consider the above expression as defining a closed operator acting on a dense subspace of $L^2(\mathbb{R}^n)$ which we denote by $\mathcal{D}_{T_{n,\sigma}}$. The following result is to be considered as a generalization of Theorem 3.

Theorem 10. *For any $\sigma > 0$, the spectrum of the operator $T_{n,\sigma}\colon \mathcal{D}_{T_{n,\sigma}} \to L^2(\mathbb{R}^n)$ consists of eigenvalues which are the sums of the eigenvalues of the matrix A.*

Proof. The matrix A has n real negative eigenvalues $\lambda_1, \ldots, \lambda_n$. There exists an orthogonal matrix Θ such that $\Theta A \Theta^T = D$, where $D = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$. Making the change of variables $y = \Theta x$, we obtain an operator $\overline{T}_{n,\sigma}$ given by

$$\overline{T}_{n,\sigma}\rho = -\sigma \sum_{i=1}^{n} \lambda_i \rho_{y_i y_i} - \Big(\frac{1}{2}\sum_{i=1}^{n}\lambda_i - \frac{1}{4\sigma}\sum_{i=1}^{n}\lambda_i y_i^2\Big)\rho.$$

The spectrum of $\overline{T}_{n,\sigma}$ is the same as that of $T_{n,\sigma}$. We have $\overline{T}_{n,\sigma} = -\sum_{i=1}^{n}\lambda_i T_{\sigma,y_i}$, where T_{σ,y_i} are the operators considered in Section 4 for each variable (cf. remarks made at the beginning of Section 6). To complete the proof, recall that by Theorem 3 the eigenvalues of T_{σ,y_i} are the nonpositive integers. The eigenfunctions of $\overline{T}_{n,\sigma}$ are given by the products of the functions $u_k(y_i, 2\sigma)$ for each variable. $\square$

Multiplying the eigenfunctions of $T_{n,\sigma}$ by $e^{-x^T x/4\sigma}$, we obtain eigenfunctions of $L_{n,\sigma}$ with the same eigenvalues. The rest of the eigenfunctions of $L_{n,\sigma}$ belong to the space $\{\rho : e^{x^T x/4\sigma}\rho \notin L^2(\mathbb{R}^n)\}$.

We would like to generalize to the multidimensional case the results of Section 5 on double bracket equations. To this end, let $A_2(\mathbb{R}^n)$ be a space of twice (weakly) differentiable functions on $\mathbb{R}^n$ (in fact, we can simply let $A_2(\mathbb{R}^n) = C^2(\mathbb{R}^n)$). Consider the operators $P_{1,i,j}$, $P_{2,i,j}$, $P_{3,i,j}$, and P_4 acting on the space

$$\mathcal{D}_n = \{\rho \in L^2(\mathbb{R}^n) \cap A_2(\mathbb{R}^n) : \rho_{x_i x_j}, x_i \rho_{x_j}, x_i x_j \rho \in L^2(\mathbb{R}^n)\ \forall i, j = 1, 2, \ldots, n\}$$

that are defined as follows:

$$P_{1,i,j}\rho = \rho_{x_i x_j}, \quad P_{2,i,j}\rho = x_i \rho_{x_j}, \quad P_{3,i,j}\rho = x_i x_j \rho, \quad P_4\rho = \rho.$$

These operators span a Lie algebra which we denote by $\mathfrak{g}_n$. For each $\sigma > 0$, let $L_n(\sigma)$ and $T_n(\sigma)$ denote the restrictions of $L_{n,\sigma}$ and $T_{n,\sigma}$ to $\mathfrak{g}_n$. Theorem 10 implies that $T_n(\sigma)$, $0 < \sigma < +\infty$, is an isospectral family of operators in $\mathfrak{g}_n$. The two propositions below state that the evolution of $L_n(\sigma)$ and $T_n(\sigma)$ with respect to σ can be described by double bracket differential equations.

Proposition 11. *Let M be an operator in $\mathfrak{g}_n$ defined by*

$$M\rho = -\frac{1}{4} \sum_{i,j=1}^{n} (A^{-1})_{ij} \rho_{x_i x_j}.$$

Then $L_n(\sigma)$, $0 < \sigma < +\infty$, is a solution of the differential equation

$$\frac{dL}{d\sigma} = [L, [L, M]].$$

Proposition 12. *Let N be an operator in $\mathfrak{g}_n$ defined by*

$$N\rho = -\frac{1}{2}\sum_{i,j=1}^{n}(A^{-1})_{ij}\rho_{x_i x_j}.$$

Then $T_n(\sigma)$, $0 < \sigma < +\infty$, is a solution of the differential equation

$$\frac{dT}{d\sigma} = [T, [T, N]].$$

We omit the proofs which are straightforward. When $n = 1$, the above equations reduce to those given in Section 5.

Finally, notice that, given the system (III.19), we can *always* find a basis in which $Q = \sigma I$ and $A = A^T$, if A is allowed to depend on time. First, switch to a basis in which $Q = \sigma I$. Note that this property is invariant under orthogonal coordinate transformations. Let $\Omega = \frac{1}{2}(A + A^T)$. Substituting $y = e^{-\Omega t}x$ into (III.19), we obtain

$$dy = \frac{1}{2}e^{-\Omega t}(A + A^T)e^{\Omega t}y\,dt + e^{-\Omega t}B\,dw. \tag{III.29}$$

Propositions 11 and 12 apply to the system (III.29) without any changes (except that now $T_n(\sigma)$ will also depend on t).

Chapter IV

STEADY-STATE PROBABILITY DISTRIBUTIONS

1. Introduction. The study of linear systems excited by white noise of constant intensity is greatly facilitated by the fact that one has an explicit formula for the solution of the Fokker-Planck equation which describes the evolution of the probability density. For nonlinear systems the situation is quite different. Namely, not only are the transient solutions difficult to find, but even the steady-state solutions are hard to characterize. In this chapter we show that for a certain class of nonlinear systems the steady-state densities can be found explicitly. These systems correspond to feedback systems of Lur'e type, and can also be thought of as a generalization of the monotemperaturic systems introduced in [7].

We consider Itô stochastic systems of the form

$$\begin{aligned} dx &= Axdt + Gdw + budt \\ y &= c^T x \end{aligned} \tag{IV.1}$$

where $x, b, c \in \mathbb{R}^n$, $u, y \in \mathbb{R}$, w is an m-dimensional Wiener process, and A and G are matrices of appropriate dimensions. Letting $u = f(y)$, where f is a piecewise continuous function, we obtain the system

$$dx = Axdt + Gdw + bf(c^T x)dt. \tag{IV.2}$$

We will be particularly interested in the case when $u = q(y)$, where q is a quantizer with sensitivity Δ and saturation value M as defined in the Introduction to Chapter II. This leads to the system

$$dx = Axdt + Gdw + bq(c^T x)dt. \tag{IV.3}$$

Unlike in Chapter II, we will fix Δ, but allow the possibility that the quantizer q has a countable set of values. In other words, we will sometimes let $M = +\infty$ so that $q(x) = \left\lfloor \frac{x}{\Delta} + \frac{1}{2} \right\rfloor$.

Unless explicitly stated otherwise, we make the following two assumptions (as in Chapter III).

a) All eigenvalues of A have negative real parts.

b) (A,G) is a controllable pair.

The problem under consideration is that of existence of a steady-state probability distribution for the process $x(t)$ defined by either (IV.2) or (IV.3). Let us first recall what happens in the linear case ($b = 0$). As we know from Section 2 of Chapter I, the steady-state probability density is

$$\rho(x) = \frac{1}{\sqrt{(2\pi)^n \det Q}} e^{-\frac{1}{2}x^T Q^{-1} x}$$

where Q is the positive definite symmetric steady-state variance matrix satisfying the equation

$$AQ + QA^T + GG^T = 0. \tag{IV.4}$$

In this chapter we will be concerned with extending this result to the non-linear case. In Section 2 we formulate a condition on the parameters of the system (IV.3) which enables us to obtain an explicit formula for a steady-state probability density. This condition will be called the *compatibility condition*, and systems for which it is satisfied will be called *compatible*. In Section 3 we extend this result to the more general case of (IV.2). We give some examples in Section 4.

A physical interpretation of compatibility is given in Section 5 in terms of certain concepts from statistical thermodynamics. Namely, systems that are monotemperaturic in the sense of [7] turn out to be compatible. This shows more clearly the place that compatible systems occupy among all systems of the form (IV.2). In Section 6 we single out a class of systems for which the compatibility condition takes a particularly transparent form, and show how the steady-state probability densities are related to Liapunov functions for deterministic nonlinear feedback systems. This makes contact with the Lur'e problem of absolute stability.

In Section 7 we discuss convergence of the probability distribution associated with (IV.2) to steady state, and indicate connections with existing literature on stochastic methods for function minimization. In Section 8 we consider systems with unstable nonrandom part, making connections with the material of Chapter II. Finally, in Section 9 we use the work of Zakai on the existence of steady-state probability distributions to obtain bounds on second moments for certain non-compatible systems. This provides a natural generalization of the results of

Section 5 to a class of systems whose temperature is not constant but rather varies along a certain sufficiently small interval.

2. Compatible systems and piecewise gaussian steady-state probability densities. Let us denote by Q the solution of the equation (IV.4). We will say that the system (IV.2) is *compatible* if the following *compatibility condition* is satisfied:

$$b = \lambda AQc \quad \text{for some } \lambda > 0. \tag{IV.5}$$

Consider the system (IV.3) with q being a quantizer with sensitivity $\Delta > 0$ and saturation value M ($1 \leq M \leq +\infty$). For $k \in \mathbb{Z}$, define the interval $J_k = ((k-1/2)\Delta, (k+1/2)\Delta]$. It is not hard to see that the function defined by

$$\rho(x) = Ne^{-\frac{1}{2}(x+A^{-1}bk)^T Q^{-1}(x+A^{-1}bk)+d_k} \quad \text{if } c^T x \in J_k \tag{IV.6}$$

with arbitrary constants N and d_k satisfies the equation for a steady-state probability density associated with (IV.3) almost everywhere. This function is piecewise gaussian. Clearly, f $x \in \mathbb{R}$, we can always determine particular values of d_k so as to make ρ continuous. This is not always true in the multidimensional case. As we now show, the compatibility condition (IV.5) is precisely what makes it possible to obtain a continuous steady-state probability density.

Theorem 1. *If the compatibility condition (IV.5) is satisfied, then the process described by the system (IV.3) admits a steady-state probability distribution with a continuous piecewise gaussian density.*

Proof. Define

$$\rho(x) = Ne^{-\frac{1}{2}x^TQ^{-1}x-\lambda c^Txq(c^Tx)} \tag{IV.7}$$

where N is fixed by the requirement that $\int_{\mathbb{R}^n} \rho(x)dx = 1$. Such a normalization constant N exists since $\lambda > 0$. In light of (IV.5), the function (IV.7) is a special case of (IV.6) with $d_k = \dfrac{1}{2}b^T(A^{-1})^TQ^{-1}A^{-1}bk^2$. The statement of the theorem follows. □

Note that our reasoning did not depend on particular values of Δ or M. In fact, a proof for an arbitrary piecewise constant function q would proceed along the same lines. This suggests that our techniques are amenable to generalizations, and in the next section we will treat a larger class of nonlinear feedback controls.

The statement of Theorem 1 can be generalized as follows: *the process described by the system*

$$\begin{aligned} dx &= Axdt + Gdw + \sum_{i=1}^{m} b_i q_i(c_i^Tx)dt \\ y_1 &= c_1^Tx \\ &\dots \\ y_m &= c_m^Tx \end{aligned} \tag{IV.8}$$

admits a piecewise gaussian steady-state probability density if $b_i = \lambda_i AQc_i$, $\lambda_i > 0$ $(i = 1, \dots, m)$. In this case, the steady-state density takes the form

$$\rho(x) = Ne^{-\frac{1}{2}x^TQ^{-1}x-\sum_{i=1}^m \lambda_i c_i^Txq_i(c_i^Tx)}.$$

Thus for the quantized state feedback system

$$dx = Axdt + Gdw + Bq(x)dt \tag{IV.9}$$

where q is a uniform quantizer, we need

$$B^{(i)} = \lambda_i (AQ)^{(i)} \quad \text{for some } \lambda_i > 0 \ (i = 1, \ldots, n) \tag{IV.10}$$

where $P^{(i)}$ stands for the ith column of a matrix P. The formula (IV.10) can be rewritten as

$$B = AQD, \quad D = \operatorname{diag}(\lambda_1, \ldots, \lambda_n).$$

A simple observation concerning the system (IV.9) is contained in the following statement.

Proposition 2. *If the condition (IV.10) holds with all $\lambda_i \neq 0$, then (A, B) is a controllable pair.*

Proof. In view of (IV.10), we need to evaluate

$$\begin{aligned} &\operatorname{rank}\{(AQ)^{(1)}, \ldots, (AQ)^{(n)}, A(AQ)^{(1)}, \ldots, A(AQ)^{(n)}, \\ &\quad \ldots, A^{n-1}(AQ)^{(1)}, \ldots, A^{n-1}(AQ)^{(n)}\} \\ &= \operatorname{rank}\{AQ^{(1)}, \ldots, AQ^{(n)}, \ldots, A^n Q^{(1)}, \ldots, A^n Q^{(n)}\}. \end{aligned}$$

By the Cayley-Hamilton theorem we have $A^n = \alpha_{n-1}A^{n-1} + \cdots + \alpha_1 A + \alpha_0$ with $\alpha_i \in \mathbb{R}$ and $|\alpha_0| = |\det A| \neq 0$ by the assumption a). We conclude that

$$\begin{aligned} &\operatorname{span}\{AQ^{(i)}, A^2Q^{(i)}, \ldots, A^nQ^{(i)}\} \\ &= \operatorname{span}\{AQ^{(i)}, \ldots, A^{n-1}Q^{(i)}, \alpha_{n-1}A^{n-1}Q^{(i)} + \cdots + \alpha_1 AQ^{(i)} + \alpha_0 Q^{(i)}\} \end{aligned}$$

for each i, and this contains $Q^{(i)}$. The last calculation and the fact that Q is positive definite hence nonsingular prove the proposition. $\square$

3. More general nonlinear feedback case. Consider the system (IV.2), and assume that the compatibility condition (IV.5) holds. Let us look for a steady-state probability density taking the form

$$\rho(x) = Ne^{-\frac{1}{2}x^TQ^{-1}x-\lambda F(c^Tx)} \tag{IV.11}$$

where $F(z) = \int_0^z f(v)dv$ and $N > 0$ is a normalization constant. Note that (IV.11) includes (IV.7) is a special case.

Theorem 3. *If the compatibility condition (IV.5) is satisfied, then the function ρ given by (IV.11) is a steady-state probability density for the process described by (IV.2) whenever f is such that $\rho \in L^1(\mathbb{R}^n)$.*

Remark. The last requirement is not fulfilled automatically. However, ρ does belong to $L^1(\mathbb{R}^n)$ if, for example, $xf(x) \geq 0$ for all x.

Proof. The Fokker-Planck operator L corresponding to (IV.2) is defined by the formula

$$\begin{aligned} L\rho = &-\big(\mathrm{tr}A + \sum_{i=1}^n b_ic_if'(c^Tx)\big)\rho \\ &-\big(\sum_{i,j=1}^n A_{ij}x_j + \sum_{i=1}^n b_if(c^Tx)\big)\rho_{x_i} + \frac{1}{2}\sum_{j,k=1}^n (GG^T)_{jk}\rho_{x_jx_k}. \end{aligned}$$

Let us evaluate the above expression when ρ is given by (IV.11). Combining the terms and making use of (IV.5), we obtain

$$L\rho = \rho\big[(-\mathrm{tr}A - \frac{1}{2}\mathrm{tr}GG^TQ^{-1}) + (\lambda^2(f(c^Tx))^2 - \lambda f'(c^Tx))c^T(\frac{1}{2}GG^T + AQ)c$$
$$+ x^T(Q^{-1}A + \frac{1}{2}Q^{-1}GG^TQ^{-1})x + \lambda f(c^Tx)c^T(A + GG^TQ^{-1} + QA^TQ^{-1})x\big].$$

Now it is easy to see that all the terms equal zero because of (IV.4). □

We can switch to new coordinates in which $Q = TI$ for some $T \in \mathbb{R}$. The structure of compatible systems is then revealed by the following statement.

Corollary 4. *If $\Omega = -\Omega^T$ and $\lambda > 0$, then the system*

$$dx = (\Omega - \frac{1}{2T}GG^T)xdt + Gdw + \lambda T(\Omega - \frac{1}{2T}GG^T)cf(c^Tx)dt \qquad \text{(IV.12)}$$

is compatible.

4. Examples. It is insightful to see how solutions to certain known problems are captured as special cases of Theorem 3.

EXAMPLE 1. Consider the system

$$dx_i = f_i(x_1, \dots, x_n)dt + gdw_i, \quad i = 1, \dots, n \qquad \text{(IV.13)}$$

where w_i's are independent scalar Wiener processes. We will call such a system *gradient* if there exists a function $\phi(x_1, \dots, x_n)$ such that $f_i = -\dfrac{\partial \phi}{\partial x_i}$. It is not difficult to verify that a steady-state density is then given by the formula

$$\rho(x) = Ne^{-2\phi(x)/g^2}$$

(whenever $\rho \in L^1(\mathbb{R}^n)$). For a justification of the existence of this steady-state density for (IV.13), see Appendix I.

In fact, (IV.13) belongs to a general class of systems that take the form

$$\dot{x} = -\nabla\phi(x) + G\dot{w}$$

where the gradient ∇ is computed with respect to the Riemannian metric given by $(GG^T)^{-1}$. For a detailed study of such systems, extended also to degenerate diffusions, see [15]. An arbitrary compatible system will possess, in addition to gradient terms, certain "skew-symmetric" terms which do not change the steady-state probability distribution (more precisely, these come from vector fields of divergence zero that are everywhere tangential to the equiprobable surfaces). In fact, all compatible systems naturally fall into the framework of [15] for the case of $\mathbb{R}^n$ with a constant metric. A special class of such systems in $\mathbb{R}^2$ has been described by Rueda in [36].

Example 2. The Newton's second law for a nonlinear spring in a viscous fluid in the presence of random external forces may be expressed by the second-order equation

$$\ddot{x} + a\dot{x} + f(x) = \dot{w} \tag{IV.14}$$

where $\dot{w} = \dfrac{dw}{dt}$ is white noise and $a > 0$. The total energy of the system is $\dfrac{1}{2}\dot{x}^2 + F(x)$, and a steady-state probability density is

$$\rho(x, \dot{x}) = Ne^{-2a(\frac{1}{2}\dot{x}^2 + F(x))}. \tag{IV.15}$$

This formula reflects the fact that the levels of equal energy are also the levels of equal probability in steady state, and ensures that the fluctuation introduced by the presence of white noise and the energy dissipation due to the damping term $a\dot{x}$ eventually neutralize each other. See [14] for a generalization of these ideas and an application to function minimization using simulated annealing.

To see how this example fits into the above framework, consider the following auxiliary system:

$$\frac{d}{dt}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} -\epsilon & 1 \\ 0 & -a \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} 0 \\ 1 \end{pmatrix}\dot{w} + \begin{pmatrix} 0 \\ -1 \end{pmatrix} f(x), \quad \epsilon > 0. \tag{IV.16}$$

One can check that the compatibility condition (IV.5) holds for (IV.16) with the proportionality constant $\lambda = 2(a+\epsilon) \to 2a$ as $\epsilon \to 0$, which reveals the meaning of the constant $2a$ in the formula (IV.15). If we compute the steady-state density for (IV.16) using the formula (IV.11), and then take the limit as $\epsilon \to 0$, we arrive precisely at (IV.15).

EXAMPLE 3. As another physical example, consider the circuit shown below.

Electrical circuit of Example 3.

Suppose that its elements are a linear inductor with inductance L, a noisy resistor in the Nyquist-Johnson form, and a nonlinear capacitor. The Kirchhoff's voltage law reads

$$V_R + V_C + V_L = 0.$$

For the inductor we have

$$V_L = L\dot{I}.$$

The Nyquist-Johnson resistor model gives

$$V_R = RI + \sqrt{R}\,\dot{w}$$

(this reflects the fluctuation-dissipation equality for an approprate value of the temperature, namely, $1/2$). Finally, for the nonlinear capacitor we have a voltage-charge relationship of the form

$$V_C = f(q_c).$$

Therefore, letting $x_1 = q_c$ (charge) and $x_2 = I$ (current), we obtain the system

$$\frac{d}{dt}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 0 & -\frac{R}{L} \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \begin{pmatrix} 0 \\ -\frac{\sqrt{R}}{L} \end{pmatrix}\dot{w} + \begin{pmatrix} 0 \\ -\frac{1}{L} \end{pmatrix} f(x_1). \tag{IV.17}$$

The equation for a steady-state density is

$$\frac{R}{2L^2}\frac{\partial^2 \rho}{\partial x_2 x_2} + \frac{1}{L}(Rx_2 + f(x_1))\frac{\partial \rho}{\partial x_2} - x_2\frac{\partial \rho}{\partial x_1} + \frac{R}{L}\rho = 0.$$

The energy of the inductor is $\int_0^t IV_L dt = \int_0^t LIdI = \dfrac{L}{2}x_2^2$. The energy of the capacitor is $\int_0^t IV_C = \int_0^t f(x_1)dx_1 = F(x_1)$. One can verify that a steady-state density is given by

$$\rho(x_1, x_2) = Ne^{-2(\frac{L}{2}x_2^2 + F(x_1))}$$

(c.f. Example 2). Notice that f automatically satisfies the inequality $xf(x) \geq 0$ because it expresses the voltage-charge relationship in the capacitor, hence $\rho \in L^1(\mathbb{R}^n)$. The system (IV.17) is compatible in the same limiting sense as (IV.14).

Alternatively, we could represent the nonlinear law for the capacitor as a linear one plus a perturbation, which amounts to letting $f(x) = \frac{1}{C}\, x + g(x)$ for some function g and a positive constant C. Further, we could switch to canonical coordinates in which the equipartition of energy theorem holds. Namely, if we scale the variables: $x_1 \mapsto \frac{1}{\sqrt{C}}\, x_1,\ x_2 \mapsto \sqrt{L}\, x_2$, then the steady-state variance matrix Q becomes $\frac{1}{2}I$, and compatibility can be easily verified using Corollary 4. Moreover, $\lambda = 2$ regardless of the numerical characteristics of the circuit elements. The same method would also work for Example 2.

The steady-state probability densities of Examples 2 and 3 are in fact unique. This is a consequence of a result by Zakai [41] for a class of systems that includes (IV.14) and (IV.17) as special cases (see Theorem 5 below).

5. Compatibility from the statistical thermodynamics viewpoint. We are now in position to give an interpretation of the compatibility property on physical grounds. It involves systems that describe the behavior of electrical networks with noisy resistors in Nyquist-Johnson form. If all the resistors are of the same temperature T, the system is called *monotemperaturic*. This concept was first mathematically defined in [7], where the authors give a canonical representation for such systems in the form

$$
\begin{aligned}
\dot{x} &= (\Omega - \frac{1}{2T}GG^T)x + G\dot{w} + Du, \\
\dot{y} &= -D^T x - Fu + \sqrt{2TF}\dot{\nu}
\end{aligned}
\tag{IV.18}
$$

Here $\Omega = -\Omega^T$, $F = F^T$, and $\dot{w}$ and $\dot{\nu}$ are independent white noise processes. The steady-state variance for (IV.18) is $Q = TI$, so we can say that in steady state all the modes possess equal energy. This property is sometimes referred to as the *equipartition of energy* property.

In the present framework, certain types of circuits with nonlinear capacitors or inductors are described by equations of the form (IV.2). We claim that by closing the feedback loop in (IV.18) we can obtain a compatible system. Indeed, let $u = y + f(y)$ (assuming single-input, single-output case, otherwise do it for each pair (u_i, y_i)). This yields

$$
\begin{aligned}
\frac{d}{dt}\begin{pmatrix} x \\ y \end{pmatrix} &= \begin{pmatrix} \Omega - \frac{1}{2T}GG^T & D \\ -D^T & -F \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} \\
&\quad + \begin{pmatrix} G & 0 \\ 0 & \sqrt{2TF} \end{pmatrix}\begin{pmatrix} \dot{w} \\ \dot{\nu} \end{pmatrix} + \begin{pmatrix} D \\ -F \end{pmatrix} f(y).
\end{aligned}
\tag{IV.19}
$$

We have thus obtained a system that takes the form (IV.12) described in Corollary 4. Summarizing, we can say that compatibility can be thought of as a natural property of monotemperaturic systems with nonlinear reactances. Notice that (IV.2) is more general than (IV.19) since the noise matrix of (IV.2) does not necessarily take a block diagonal form.

EXAMPLE 4. Consider the circuit shown below.

Electrical circuit of Example 4.

Let us assume for simplicity that all the resistance and capacitance values are equal to 1, and that the temperature of all the resistors is 1/2. Suppose also that the inductor is nonlinear. Then we have: $V_1 + V_3 + V_4 - V_2 = 0$, $\dot{V}_1 = I_1$, $\dot{V}_2 = -I_3$, $V_1 = I_2 - I_1 + \dot{w}_1$, $V_2 = I_3 - I_2 + \dot{w}_2$, $V_4 = I_2 + \dot{w}_3$, and $I_2 = \phi + f(\phi)$ where ϕ is the flux in the inductor. The open-loop equations are

$$\frac{d}{dt}\begin{pmatrix} V_1 \\ V_2 \end{pmatrix} = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}\begin{pmatrix} V_1 \\ V_2 \end{pmatrix} + \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} \dot{w}_1 \\ \dot{w}_2 \end{pmatrix} + \begin{pmatrix} 1 \\ -1 \end{pmatrix} I_2,$$
$$\dot{\phi} = (\,-1, \quad 1\,)\begin{pmatrix} V_1 \\ V_2 \end{pmatrix} - I_2 + \dot{w}_3.$$

The closed-loop equations are

$$\frac{d}{dt}\begin{pmatrix} V_1 \\ V_2 \\ \phi \end{pmatrix} = \begin{pmatrix} -1 & 0 & 1 \\ 0 & -1 & -1 \\ -1 & 1 & -1 \end{pmatrix}\begin{pmatrix} V_1 \\ V_2 \\ \phi \end{pmatrix} + \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}\begin{pmatrix} \dot{w}_1 \\ \dot{w}_2 \\ w_3 \end{pmatrix} + \begin{pmatrix} 1 \\ -1 \\ -1 \end{pmatrix} f(\phi).$$

6. Steady-state probability densities and Liapunov functions. Consider a system excited by a scalar white noise

$$\dot{x} = Ax + g\dot{w} + bf(c^T x).$$

Since by the assumptions made in the Introduction (A, g) is a controllable pair, in the appropriate basis the linear part of the system takes the standard *controllable companion form*, so that we have

$$\frac{d}{dt}\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ -p_0 & -p_1 & -p_2 & \dots & -p_{n-1} \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}\dot{w} + bf(c^T x).$$

If, moreover, b and g are proportional then, multiplying f by a scalar if necessary, we arrive at

$$p(D)x + f(c(D)x) = \dot{w}. \tag{IV.20}$$

Here $D = \dfrac{d}{dt}$, and $p(D) = D^n + p_{n-1}D^{n-1} + \cdots + p_1 D + p_0$ and $c(D) = c_{n-1}D^{n-1} + \cdots + c_1 D + c_0$ are polynomials. The class of systems thus constructed includes (IV.14) and (IV.17) as special cases, and is of considerable interest despite the special form of (IV.20) (see, e.g., [41]).

In this section we will be concerned with formulating conditions on the polynomials p and c under which (IV.20) is compatible. We will adopt certain results from [3] regarding the Lur'e problem of absolute stability for the deterministic counterpart of (IV.20)

$$p(D)x + f(c(D)x) = 0. \tag{IV.21}$$

We will assume that $xf(x) > 0$ for all x except $x = 0$, and that either the equation $p(D)x = 0$ is asymptotically stable or $p(D) = Dh(D)$ with $h(D)x = 0$ asymptotically stable. Denoting $Dc(D)$ by $m(D)$, assume also that the function $\dfrac{m(s)}{p(s)}$ is positive real. Then we can apply the classical *factorization lemma* to conclude that there exists a unique polynomial $r(s)$ with real positive coefficients and no zeros in the right half plane, such that $\mathrm{Ev}p(s)m(-s) = r(s)r(-s) = r^+(s)r^-(s)$ (here Ev stands for the even part of a polynomial). We construct a Liapunov function for (IV.21) as follows:

$$V(x) = \int_0^x [p(D)x\, m(D)x - (r^-(D)x)^2]dt + F(c(D)x) \tag{IV.22}$$

where $\dfrac{\partial F(c(D)x)}{\partial x^{(i)}} = c_i f(c(D)x)$ as before. In [3] the problem of absolute stability for (IV.21) is investigated with the aid of the function (IV.22). It can be shown that V is well-defined, positive definite, and that its derivative along the solutions of (IV.21) is given by

$$\dot{V} = -(r^-(D)x)^2. \tag{IV.23}$$

Obtaining the Liapunov function (IV.22) is a matter of multiplying both sides of (IV.21) by $m(D)x$, integrating by parts, and completing a square if necessary. In many situations (cf. Example 2) such a Liapunov function arises naturally as the total energy of the system.

Now, given a polynomial $p(D)$, let us choose $c(D)$ by setting each c_i to p_{i+1} if i is even and to 0 if i is odd. Notice that $m(D)$ is then simply the odd part of $p(D)$, and therefore $\mathrm{Ev}p(D)m(-D) = m(D)m(-D)$, so we see that $\dfrac{m(s)}{p(s)}$ is positive real and $r^-(D) = m(-D)$. We can also take $m(D)$ to be a constant multiple of $\mathrm{Odd}p(D)$ as in (IV.14), which amounts to a simple modification of the nonlinearity. We will now use the Liapunov function (IV.22) to arrive at a steady-state density for the stochastic system (IV.20).

Theorem 5. *Suppose that either $p(D)x = 0$ is asymptotically stable or $p(D) = Dh(D)$ with $h(D)x = 0$ asymptotically stable, and that*

$$c(D) = \frac{1}{aD}\operatorname{Odd} p(D) \tag{IV.24}$$

with $a > 0$. Then the function

$$\rho(x) = Ne^{-2aV(x)} \tag{IV.25}$$

is a unique steady-state probability density associated with the system (IV.20).

Proof. Let us consider the case when n is even. Using our definition of $c(D)$, we easily see that the equation for a steady-state probability density can be written as

$$\dot{\rho} = p_{n-1}\rho + \frac{1}{2}\frac{\partial^2 \rho}{\partial x_{n-1}^2} \tag{IV.26}$$

where $\dot{\rho}$ stands for the derivative of $\rho(x)$ along the solutions of the deterministic system (IV.21). Using (IV.23) and (IV.25) we can write (IV.26) as

$$2a(r^-(D)x)^2 = p_{n-1} - ac_{n-2} + 2a^2(c_{n-2}x_{n-1} + \cdots + c_0x_1)^2 = 2(a\,m(D)x)^2$$

and this is true by hypothesis.

The case of odd n is treated similarly: (IV.26) is replaced by

$$\dot{\rho} = (p_{n-1} + \frac{1}{a}f'(c(D)x))\rho + \frac{1}{2}\frac{\partial^2 \rho}{\partial x_{n-1}^2}$$

and we arrive at the equation

$$a(r^-(D)x)^2 = (p(D)x + f(c(D)x) - a\,m(D)x)^2$$

to verify which it remains to use (IV.21) again.

The uniqueness of the steady-state density (IV.25) follows from the work of Zakai [41, Theorems 3 and 4]. □

The formula (IV.24) provides a more concrete interpretation of the compatibility condition (IV.5) applied to systems of the form (IV.20). From the results of Section 3 it follows that if we want (IV.20) to be compatible, the choice of $c(D)$ is unique up to a constant. See also Appendix II.

We would like to establish a connection between the steady-state probability densities and Liapunov functions in a more general setting. To this end, consider the system (IV.2) together with the deterministic one

$$\dot{x} = Ax + bf(c^T x). \tag{IV.27}$$

If the compatibility condition (IV.5) holds, then (IV.27) is equivalent to

$$Q^{-1}A^{-1}\dot{x} = Q^{-1}x + \lambda c f(c^T x).$$

The above analysis and the formula (IV.11) suggest considering the function

$$V(x) = \frac{1}{2}x^T Q^{-1} x + \lambda F(c^T x)$$

where $\dfrac{\partial F}{\partial x_i} = c_i f(c^T x)$. This function is positive definite, and its derivative along the solutions of (IV.27) is

$$\begin{aligned}\dot{V} &= \sum_i \frac{\partial V}{\partial x_i}\dot{x}_i = \sum_i (\sum_j (Q^{-1})_{ij} x_j + \lambda c_i f(c^T x))\dot{x}_i \\ &= \sum_{i,j} \dot{x}_i (Q^{-1}A^{-1})_{ij}\dot{x}_j = \dot{x}^T (AQ)^{-1}\dot{x} = -\frac{1}{2}\dot{x}^T (AQ)^{-1} G G^T ((AQ)^{-1})^T \dot{x}\end{aligned}$$

by virtue of (IV.4). The last expression is obviously negative semidefinite, which means that V is a Liapunov function for (IV.27). Thus we see that the steady-state probability densities obtained in Section 3 are closely related to Liapunov functions for the corresponding nonrandom systems.

7. Convergence to steady state. It is well known that the time-varying solution of the Fokker-Planck equation

$$\frac{\partial \rho}{\partial t} = L\rho \tag{IV.28}$$

takes the form

$$\rho(t, x) = \sum_{i=0}^{+\infty} p_i(t) e^{\lambda_i t} g_i(x)$$

where λ_i's are the eigenvalues of L, p_i's are polynomials in t, and g_i's are functions of x. Thus, eigenvalues of L provide information about the convergence of the stochastic process to steady state (for a discussion along these lines and examples, see [5]). In the paper by Holley et al. [28], and more recently in [14] and [15], Fokker-Planck operators and their spectral properties were studied with the view towards applications to function minimization. The methods of this section are based on the techniques employed in those references.

Consider the function

$$\phi = \frac{1}{4} x^T Q^{-1} x + \frac{1}{2} \lambda F(c^T x)$$

(cf. Example 1). Define the vector $\nabla\phi$ by

$$(\nabla\phi)_i = \sum_{j=1}^{n} (GG^T)_{ij} \phi_{x_j}.$$

In view of (IV.4) and (IV.5) we have

$$\nabla\phi = -\frac{1}{2}(A + QA^TQ^{-1})x - \frac{1}{2}(I + QA^TQ^{-1}A^{-1})bf(c^Tx).$$

The Fokker-Planck operator associated with the system (IV.2) can be written as

$$L = L_{\text{grad}} + L_{\text{skew}}$$

where

$$\begin{aligned} L_{\text{grad}}\rho &= \sum_{i=1}^{n} \frac{\partial}{\partial x_i}((\nabla\phi)_i\rho) + \frac{1}{2}\sum_{i,j=1}^{n}(GG^T)_{ij}\rho_{x_ix_j} \\ &= \sum_{i,j=1}^{n}(GG^T)_{ij}[\frac{\partial}{\partial x_i}(\phi_{x_j}\rho) + \frac{1}{2}\rho_{x_ix_j}] \\ &= \frac{1}{2}\sum_{i,j=1}^{n}(GG^T)_{ij}\frac{\partial}{\partial x_i}(e^{-2\phi}\frac{\partial}{\partial x_j}e^{2\phi}\rho) \end{aligned}$$

and

$$\begin{aligned} L_{\text{skew}}\rho = &\frac{1}{2}\sum_{i,j=1}^{n}\frac{\partial}{\partial x_i}[(QA^TQ^{-1} - A)_{ij}x_j\rho] \\ &+ \frac{1}{2}\sum_{i,j=1}^{n}\frac{\partial}{\partial x_i}[(QA^TQ^{-1}A^{-1} - I)_{ij}b_jf(c^Tx)\rho]. \end{aligned}$$

Define a *gauge transform* $\widetilde{P}$ of an operator P by $\widetilde{P}\rho = e^{2\phi}P(e^{-2\phi}\rho)$. We have

$$\begin{aligned} \int_{\mathbb{R}^n}\rho\widetilde{L}_{\text{grad}}\rho e^{-2\phi}dx &= \int_{\mathbb{R}^n}\frac{1}{2}\sum_{i,j=1}^{n}(GG^T)_{ij}\rho\frac{\partial}{\partial x_i}(e^{-2\phi}\rho_{x_j})dx \\ &= -\int_{\mathbb{R}^n}\frac{1}{2}\sum_{i,j=1}^{n}(GG^T)_{ij}\rho_{x_i}\rho_{x_j}e^{-2\phi}dx \leq 0. \end{aligned}$$

In the above calculation we used integration by parts. Now consider

$$\begin{aligned}
\int_{\mathbb{R}^n} \rho \widetilde{L}_{\text{skew}} \rho e^{-2\phi} dx &= \int_{\mathbb{R}^n} \frac{1}{2} \sum_{i,j=1}^{n} \rho \frac{\partial}{\partial x_i} [(QA^TQ^{-1} - A)_{ij} x_j e^{-2\phi} \rho] dx \\
&+ \int_{\mathbb{R}^n} \frac{1}{2} \sum_{i,j=1}^{n} \rho \frac{\partial}{\partial x_i} [(QA^TQ^{-1}A^{-1} - I)_{ij} b_j f(c^T x) e^{-2\phi} \rho] dx \\
&= \int_{\mathbb{R}^n} \frac{1}{2} \sum_{i,j=1}^{n} \rho [(QA^TQ^{-1} - A)_{ij} x_j] e^{-2\phi} \rho_{x_i} dx \\
&+ \int_{\mathbb{R}^n} \frac{1}{2} \sum_{i,j=1}^{n} \rho [(QA^TQ^{-1}A^{-1} - I)_{ij} b_j f(c^T x)] e^{-2\phi} \rho_{x_i} dx \\
&+ \int_{\mathbb{R}^n} \rho^2 L_{\text{skew}} e^{-2\phi} dx. \qquad \text{(IV.29)}
\end{aligned}$$

Since both L and L_{grad} annihilate the steady-state density $e^{-2\phi}$, it follows that $L_{\text{skew}} e^{-2\phi} = 0$ and so the last integral in the formula (IV.29) is zero. Using integration by parts, we can rewrite the first two integrals in (IV.29) as

$$\begin{aligned}
&- \int_{\mathbb{R}^n} \frac{1}{2} \sum_{i,j=1}^{n} \rho \frac{\partial}{\partial x_i} [(QA^TQ^{-1} - A)_{ij} x_j e^{-2\phi} \rho] dx \\
&- \int_{\mathbb{R}^n} \frac{1}{2} \sum_{i,j=1}^{n} \rho \frac{\partial}{\partial x_i} [(QA^TQ^{-1}A^{-1} - I)_{ij} b_j f(c^T x) e^{-2\phi} \rho] dx
\end{aligned}$$

which implies that

$$\int_{\mathbb{R}^n} \rho \widetilde{L}_{\text{skew}} \rho e^{-2\phi} dx = - \int_{\mathbb{R}^n} \rho \widetilde{L}_{\text{skew}} \rho e^{-2\phi} dx = 0.$$

Putting all the above calculations together, we conclude that

$$\int_{\mathbb{R}^n} \rho \widetilde{L} \rho e^{-2\phi} dx \leq 0.$$

Therefore, all eigenvalues of $\widetilde{L}$ with eigenfunctions in the space $\{\rho : e^{-\phi}\rho \in L^2(\mathbb{R}^n)\}$ are nonpositive. This implies that all eigenvalues of L with eigenfunctions in the space $\{\rho : e^{\phi}\rho \in L^2(\mathbb{R}^n)\}$ are nonpositive (cf. Section 4 of Chapter

III). Thus it follows that the steady-state density is a stable equilibrium for (IV.28). To obtain specific information about its domain of attraction and the speed of convergence to steady state, one needs to carry out a more detailed spectral analysis of the Fokker-Planck operator L.

Given the system (IV.2), let us define the *spectral gap* to be

$$\lambda_f = \min_\rho \Big\{ \Big(-\int_{\mathbb{R}^n} \rho \widetilde{L} \rho e^{-2\phi} dx \Big) : \int_{\mathbb{R}^n} \rho^2 e^{-2\phi} dx = 1, \ \int_{\mathbb{R}^n} \rho e^{-2\phi} dx = 0 \Big\}.$$

The problem of estimating the spectral gap plays a role in the theory of simulated annealing [28]. In view of the results obtained in Chapter III, it might be useful to relate λ_f to the value of the spectral gap in the case when f is linear. We now show that this can be done, at least in some cases. As an example, consider the familiar spring equation (IV.14) and the corresponding steady-state density (IV.15). The spectral gap is $\lambda_{f(x)} = \min_\rho \{ \frac{1}{2} \int_{\mathbb{R}^2} (\rho_y)^2 e^{-y^2 - 2F(x)} dx\, dy \}$ subject to the conditions $\int_{\mathbb{R}^2} \rho^2 e^{-y^2-2F(x)} dx\, dy = 1$ and $\int_{\mathbb{R}^2} \rho e^{-y^2-2F(x)} dx\, dy = 0$. Note that if a function ρ is such that the last condition is satisfied, and if we define $\psi(x,y) = \rho(x,y) e^{\alpha x^2 - 2F(x)}$ for a certain $\alpha \in \mathbb{R}$, then $\int_{\mathbb{R}^2} \psi e^{-y^2-\alpha x^2} dx\, dy = 0$. Now, assume that there exist positive numbers α and ϵ such that $\alpha x^2 \leq 2F(x) \leq \alpha x^2 + \epsilon$ for all x (geometrically, this means that the graphs of $f(x)$ and αx must be close enough to each other so that the area between them is finite). Then it is not hard to see that

$$\begin{aligned} \frac{\int_{\mathbb{R}^2} (\psi_y)^2 e^{-y^2-\alpha x^2} dx\, dy}{(\int_{\mathbb{R}^2} \psi^2 e^{-y^2-\alpha x^2} dx\, dy)^2} &= \frac{\int_{\mathbb{R}^2} (\rho_y)^2 e^{\alpha x^2 - 2F(x)} e^{-y^2-2F(x)} dx\, dy}{(\int_{\mathbb{R}^2} \rho^2 e^{\alpha x^2-2F(x)} e^{-y^2-2F(x)} dx\, dy)^2} \\ &\leq e^{2\epsilon} \int_{\mathbb{R}^2} (\rho_y)^2 e^{-y^2-2F(x)} dx\, dy \leq e^{2\epsilon} \lambda_{f(x)} \end{aligned}$$

which implies that $e^{2\epsilon}\lambda_{f(x)} \geq \lambda_{\alpha x}$. The ease with which we could make the above estimate is due to the fact that in $\widetilde{L}_{\mathrm{grad}}$ there is no differentiation with respect to x, the argument of the nonlinearity. Thus, similar estimates hold for any system (IV.2) such that $G_{(i)}$ (the ith row of G) is zero whenever $c_i \neq 0$.

Regarding convergence to the steady-state density (IV.25) for the system (IV.20), some results can be found in the literature. Namely, according to Theorem 3 of [41], for $T \to +\infty$ and all $x_0 \in \mathbb{R}^n$ we have

$$\mathcal{P}\left\{\frac{1}{T}\int_0^T f(x(t))dt \to \int_{\mathbb{R}^n} f(x)\rho_{ss}dx \,\middle|\, x(0) = x_0\right\} = 1$$

where ρ_{ss} is the steady-state density and f is any real-valued function integrable with respect to the measure $\rho_{ss}dx$.

8. Compatible stabilization. We will now indicate a connection between the above material on steady-state probability distributions and the quantized feedback stabilization problems considered in Chapter II. Consider the system

$$dx = Axdt + Gdw - BK\Delta q(x)dt \tag{IV.30}$$

where q is a uniform quantizer with a countable set of values. Let us drop the stability assumption a) of the Introduction, and suppose instead that all eigenvalues of $A - BK$ have negative real parts. Rewrite (IV.30) as

$$dx = (A - BK)xdt + Gdw + BKs(x)dt \tag{IV.31}$$

where $s(x) = x - \Delta q(x)$. We have

$$\|s(x)\|_\infty \leq \Delta/2. \tag{IV.32}$$

Now, assume that the system (IV.31) is compatible in the sense that

$$BK = AQD, \quad D = \mathrm{diag}(\lambda_1, \ldots, \lambda_n). \tag{IV.33}$$

Since (IV.31) satisfies the assumption a), we can apply Theorem 3 to obtain a steady-state probability density which will automatically belong to $L^1(\mathbb{R}^n)$ by virtue of (IV.32). A possible graph of such a density is sketched in the figure.

A steady-state probability density for the system (IV.31).

An interesting question that arises in this context is the following: given the matrices A, G, and B, is it possible to find a matrix K that makes (IV.31) compatible with asymptotically stable nonrandom part? We are not aware of a general answer to this question. What stands in the way of a simultaneous treatment of asymptotic stability and compatibility is the fact that the steady-state variance matrix associated with $dx = (A - BK)xdt + Gdw$ depends on the choice of K, and is at the same time needed to verify compatibility. In order to circumvent this difficulty, one might try to develop a procedure which deals with the issues of asymptotic stability and compatibility separately.

For instance, even if not all eigenvalues of A have negative real parts, we might still be able to find a symmetric matrix Q that satisfies the equation (IV.4). In general, this Q need not be positive definite. In view of (IV.33), one must then restrict the search for K to matrices such that BK takes the form AQD with D diagonal. This would then yield a system of the form (IV.31) that possesses a steady-state probability distribution by construction.

The above "compatible stabilization" problem is also meaningful for quantized output feedback systems of the form (IV.8). The simplest example (single-input, single-output case) is served by the familiar n-th order equation

$$p(D)x + u = \dot{w} \tag{IV.34}$$

where the control u has to take the form $u = kq(c(D)x)$. We know that a necessary and sufficient condition for (IV.34) to be compatible is given by (IV.24). Our choice of the output $c(D)x$ being thus fixed up to a constant, we can rewrite (IV.34) as

$$p(D)x + kc(D)x - ks(c(D)x) = \dot{w}.$$

Therefore, all we need to do now is choose k that makes the system $p(D)x + kc(D)x = 0$ asymptotically stable. All such values of k can be determined by a straighforward application of the Nyquist criterion.

9. Non-compatible systems and temperature bounds. Compatibility is not a necessary condition for the existence of a steady-state probability distribution. In the previous sections we have discussed systems for which explicit formulae for steady-state probability densities can be found. It would be interesting to try to develop a perturbation theory that would allow us to obtain specific information about steady-state probability distributions for those systems that are not compatible.

Consider the system (IV.2), and denote by L the corresponding Fokker-Planck operator and by L^* its adjoint. Suppose that we have a nonnegative, twice continuously differentiable function $V(x)$ in $\mathbb{R}^n$, which is dominated by a polynomial. Theorem 2 of [41] can now be formulated as follows: *If there exist numbers $R < \infty$ and $k > 0$ such that $L^*V(x) \leq -k$ for all x satisfying $\|x\| > R$, then the process defined by (IV.2) admits a steady-state probability distribution.* We can apply this Liapunov-like criterion to establish the existence of steady-state probability distributions at the expense of having to abandon constructive proofs and explicit formulae.

As is well known, if the assumption a) of the Introduction holds, then there exists a positive quadratic function $V(x) = x^TCx$ whose derivative along the solutions of $\dot{x} = Ax$ is $-x^TDx$, with D symmetric positive definite. For the system (IV.2) this implies that outside the ball of radius R

$$\begin{aligned} L^*V(x) &= -x^TDx + 2x^TCbf(c^Tx) + \mathrm{tr}(CGG^T) \\ &\leq -\lambda_{min}(D)\|x\|^2 + \max(0, \beta\lambda_{max}(Cbc^T + cb^TC^T)\|x\|^2) + \mathrm{tr}(CGG^T) \end{aligned}$$

providing that

$$\alpha x^2 \leq x f(x) \leq \beta x^2 \quad \text{for } \|x\| > R \text{ and some } \beta > \alpha > 0. \tag{IV.35}$$

Here λ_{min} and λ_{max} stand for the minimal and the maximal eigenvalue, respectively. Now we see that for β small enough there exists a steady-state probability distribution. Observe that the same conclusion holds if $\beta - \alpha$ is small enough, as long as suitable stability conditions are satisfied. Namely, we just have to write $f(c^T x) = (a - \epsilon) c^T x + g(c^T x)$ with a small $\epsilon > 0$, and require that all eigenvalues of $A + (a - \epsilon) b c^T$ have negative real parts.

Example 5. Consider the (non-compatible) second-order system

$$\ddot{x} + f(\dot{x}) + x = \dot{w}. \tag{IV.36}$$

Assuming that (IV.35) holds, we can recast (IV.36) as $\ddot{x} + \epsilon \dot{x} + g(\dot{x}) + x = \dot{w}$ with $0 < \epsilon < \alpha$. In the above notation, take $C = \begin{pmatrix} a & 1 \\ 1 & a \end{pmatrix}$, where $a > 0$. This gives $Cbc^T + cb^T C^T = \begin{pmatrix} 0 & -1 \\ -1 & -a \end{pmatrix}$ with $\lambda_{max}(Cbc^T + cb^T C^T) = \sqrt{a^2+1} - a$, and $D = \begin{pmatrix} 2 & \epsilon \\ \epsilon & 2(a\epsilon - 1) \end{pmatrix}$. Now, let $\epsilon \to 0$ and $a \to \infty$ in such a way as to have $a\epsilon \to \infty$. Then obviously $\lambda_{min}(D) \to 2$ and $\lambda_{max}(Cbc^T + cb^T C^T) \to 0$, which proves the existence of a steady-state probability distribution for all β.

In certain cases it is possible to obtain explicit bounds for the steady-state variance. The corresponding class of systems serves as a natural generalization of monotemperaturic ones. The following observation regarding linear systems provides some motivation for such analysis. Consider the system

$$\dot{x} = (\Omega(t) - G(t) G^T(t)) x + F(t) \dot{w} \tag{IV.37}$$

and assume that there exist two positive numbers T_1 and T_2 (which may be viewed as temperature bounds) such that

$$2T_1 G(t)G(t)^T \leq F(t)F(t)^T \leq 2T_2 G(t)G(t)^T$$

for all t. Here $\Omega(t)$ is a skew-symmetric matrix. The variance equation associated with (IV.37) is

$$\dot{Q} = (\Omega - GG^T)Q + Q(-\Omega - GG^T) + FF^T$$

which implies

$$\frac{d}{dt}(T_2 I - Q) = (\Omega - GG^T)(T_2 I - Q) + (T_2 I - Q)(-\Omega - GG^T) + 2T_2 GG^T - FF^T$$

and

$$\frac{d}{dt}(Q - T_1 I) = (\Omega - GG^T)(Q - T_1 I) + (Q - T_1 I)(-\Omega - GG^T) + FF^T - 2T_1 GG^T.$$

From this it follows that if $T_1 I \leq Q(0) \leq T_2 I$, then both $T_2 I - Q$ and $Q - T_1 I$ are nonnegative definite matrices, i.e., $T_1 I \leq Q(t) \leq T_2 I$ for all t, as one may have expected .

We now single out a class of nonlinear feedback systems perturbed by white noise for which similar bounds can be obtained for certain second moments in steady state. These are single-input, single-output systems of the form

$$\dot{x} = \Omega x - bf(b^T x) + b\dot{w} \tag{IV.38}$$

where $\Omega = -\Omega^T$ (Ω is now time-independent). Let us assume that the condition (IV.35) holds. In this case it may be interpreted as saying that the temperature of the system (IV.38) is between $\frac{1}{2\beta}$ and $\frac{1}{2\alpha}$. Rewriting (IV.38) as

$$\dot{x} = (\Omega - bb^T)x - bg(b^T x) + b\dot{w}$$

one easily verifies that it is not compatible unless $\Omega b = 0$. However, notice that

$$\frac{d}{dt}\mathcal{E}x^T x = -2\mathcal{E}b^T x f(b^T x) + b^T b.$$

Assume for simplicity that $\|b\| = 1$. Providing that the steady-state probability distribution exists, we deduce that in steady state $\mathcal{E}b^T x f(b^T x) = 1/2$ and therefore

$$\frac{1}{2\beta} \leq \mathcal{E}(b^T x)^2 \leq \frac{1}{2\alpha}.$$

These are natural bounds imposed by the temperature. The equation (IV.36) may serve as a simple example.

Chapter V

STOCHASTIC OPTIMAL CONTROL

1. Introduction. Consider the quantized output feedback system (IV.3), where the quantizer q has a fixed sensitivity $\Delta > 0$ and saturation value 1:

$$q(x) = \begin{cases} 1, & \text{if } x > \Delta/2 \\ 0, & \text{if } -\Delta/2 < x \leq \Delta/2 \\ -1, & \text{if } x \leq -\Delta/2 \end{cases}$$

Let us assume that the compatibility condition (IV.5) holds. In Chapter IV we have proved the existence of a piecewise gaussian steady-state probability density of the form (IV.7). We have also shown that, at least for some initial probability distributions, $\rho(t,x)$ approaches steady state, and that for an important class of systems (given by (IV.20)) the steady-state probability distribution is unique.

In this chapter we study certain steady-state optimal control problems associated with the system (IV.3). We consider several performance criteria that are formulated in terms of the steady-state probability density. In view of the above results, solutions to such problems provide information about the behaviour of the system (IV.3) for large times t, when the difference between the probability distribution $\rho(t,x)$ and the steady-state one can be thought of as being negligible.

Throughout the chapter, we will let $\mathcal{E}$ denote the expectation with respect to the steady-state probability density. We will refer to the two hyperplanes $\{x : c^T x = \pm\Delta/2\}$ as *switching hyperplanes.* We will let ∇ denote the gradient computed with respect to the standard euclidean metric on $\mathbb{R}^n$.

In Section 2 we treat several optimal control problems of the quadratic-gaussian type. In Section 3 we consider another performance criterion which is related to the number of switching hyperplane crossings per unit time.

2. Variance cost and control cost. Let V be the *variance cost* defined by $V = \mathcal{E}(c^T x)^2 = c^T \mathcal{E} x x^T c$. Note that V is invariant under coordinate transformations, and thus represents a reasonable performance criterion which indicates how fast the steady-state density decreases as $\|x\|$ becomes large. In this section we will be interested in the behaviour of V considered as a function of the sensitivity Δ. One would expect that for smaller values of Δ the quantized feedback forces x to move closer to the origin. This idea is formalized in the following statement.

Proposition 1. *The variance cost V is a monotone increasing function of the sensitivity Δ.*

Proof. The proof is geometrical and uses the form of the steady-state probability density explicitly. If we pick two numbers $\Delta_2 > \Delta_1 > 0$, then in the region $\{x : \Delta_1/2 < c^T x < \Delta_2/2\}$ the steady-state densities corresponding to $\Delta = \Delta_1$ and $\Delta = \Delta_2$ take the form

$$\rho_1(x) = N_1 e^{-\frac{1}{2}x^T Q^{-1} x - \lambda c^T x} \tag{V.1}$$

and

$$\rho_2(x) = N_2 e^{-\frac{1}{2}x^T Q^{-1} x} \tag{V.2}$$

respectively, where N_1 and N_2 are normalization constants. Let us compute the derivatives of ρ_1 and ρ_2 in the direction of the vector $A^{-1}b$. Notice that this vector is transversal to the switching hyperplanes since (IV.5) implies that $c^T A^{-1} b = \lambda c^T Q c > 0$. For ρ_1 we have

$$\begin{aligned} \langle \nabla \rho_1(x), A^{-1}b \rangle &= \langle -(Q^{-1}x + \lambda c), A^{-1}b \rangle \rho_1(x) \\ &= -(x^T Q^{-1} A^{-1} b + \lambda c^T Q^{-1} A^{-1} b) \rho_1(x) \\ &= -(\lambda c^T x + \lambda^2 c^T Q c) \rho_1(x). \end{aligned} \tag{V.3}$$

For ρ_2 we have

$$\begin{aligned} \langle \nabla \rho_2(x), A^{-1}b \rangle &= \langle -Q^{-1}x, A^{-1}b \rangle \rho_2(x) \\ &= -x^T Q^{-1} A^{-1} b \rho_2(x) = -\lambda c^T x \rho_2(x). \end{aligned} \tag{V.4}$$

Similar formulae can be obtained for $x \in \{x : -\Delta_2/2 < c^T x < -\Delta_1/2\}$. Now, $\rho_1(x)$ and $\rho_2(x)$ are proportional in the stripe $\{x : |c^T x| < \Delta_1/2\}$. If we assume that $\rho_1(x) \leq \rho_2(x)$ there, we arrive at a contradiction. Indeed, (V.3) and (V.4) and the fact that $\rho_1(x)$ and $\rho_2(x)$ are also proportional in the region $\{x : |c^T x| > \Delta_2/2\}$ would then imply that $\rho_1(x) < \rho_2(x)$ for $|c^T x| > \Delta_1/2$, thus violating the requirement

$$\int_{\mathbb{R}^n} \rho_1(x) dx = \int_{\mathbb{R}^n} \rho_2(x) dx = 1. \tag{V.5}$$

The only remaining possibility is $\rho_1(x) > \rho_2(x)$ for $|c^T x| < \Delta_1/2$. From the above analysis it follows that $\exists \bar{\Delta} \in (\Delta_1, \Delta_2)$ such that $\rho_1(x) = \rho_2(x)$ if $c^T x = \pm \bar{\Delta}/2$, $\rho_1(x) > \rho_2(x)$ if $|c^T x| < \bar{\Delta}/2$, and $\rho_1(x) < \rho_2(x)$ if $|c^T x| > \bar{\Delta}/2$.

In fact, (V.1) and (V.2) imply that $\bar{\Delta} = 2(\ln N_1 - \ln N_2)/\lambda$. Now it is not difficult to conclude that $V(\Delta_1) < V(\Delta_2)$:

$$\begin{aligned}
V(\Delta_1) - V(\Delta_2) &= \int_{\mathbb{R}^n} (c^T x)^2(\rho_1(x) - \rho_2(x))dx \\
&= \int_{|c^T x| < \bar{\Delta}/2} (c^T x)^2(\rho_1(x) - \rho_2(x))dx \\
&+ \int_{|c^T x| > \bar{\Delta}/2} (c^T x)^2(\rho_1(x) - \rho_2(x))dx \\
&< \frac{(\bar{\Delta})^2}{4} \cdot \int_{|c^T x| < \bar{\Delta}/2} (\rho_1(x) - \rho_2(x))dx \\
&+ \frac{(\bar{\Delta})^2}{4} \cdot \int_{|c^T x| > \bar{\Delta}/2} (\rho_1(x) - \rho_2(x))dx \\
&= \frac{(\bar{\Delta})^2}{4} \cdot \int_{\mathbb{R}^n} (\rho_1(x) - \rho_2(x))dx = 0
\end{aligned}$$

by (V.5). $\square$

Relative position of $\rho_1(x)$ and $\rho_2(x)$.

Proposition 1 shows that the optimal control problem that consists in choosing the value of Δ that minimizes V is trivial: the optimal performance is achieved when $\Delta = 0$. The figure above illustrates the proof, and will also be useful in the proof of Proposition 2 to follow (the projection onto the vertical plane through the vector c is shown).

One could also consider the *total variance cost* $\widetilde{V} = \mathcal{E}x^T x$ which is invariant under orthogonal coordinate transformations. It turns out that $\widetilde{V}$ has the same property of being an increasing function of the sensitivity Δ. This fact can be verified with little work, by choosing one of the coordinate vectors to be parallel to c, using Proposition 1, and integrating by parts in the remaining variables. In fact, the variance $\mathcal{E}(d^T x)^2$ is not affected by the feedback if $d \perp c$.

Another performance criterion that we will consider in this section is the *control cost* $U = \mathcal{P}(|c^T x| > \Delta/2) = \mathcal{E}(q(c^T x))^2$. It may be viewed as the proportion of time during which we actively control the system. This results in certain expenses which we would like to keep low. Clearly, the optimal quantized feedback control strategy consists in moving the switching hyperplanes to infinity: $\Delta = +\infty$ gives $U = 0$. In fact, the following is true.

Proposition 2. *The control cost U is a monotone decreasing function of the sensitivity Δ.*

Proof. Let us pick two numbers $\Delta_2 > \Delta_1 > 0$. In the notation of the proof of Proposition 1, we need to compare

$$\int_{|c^T x| > \Delta_1/2} \rho_1(x)dx = \int_{c^T x < -\Delta_1/2} \rho_1(x)dx + \int_{c^T x > \Delta_1/2} \rho_1(x)dx$$

and

$$\int_{|c^T x|>\Delta_2/2} \rho_2(x)dx = \int_{c^T x<-\Delta_1/2} \rho_2(x-d)dx + \int_{c^T x>\Delta_1/2} \rho_2(x+d)dx$$

where

$$d = \frac{\Delta_2 - \Delta_1}{2c^T c} c.$$

For $c^T x > \Delta_1/2$ we have

$$\rho_2(x+d) = N_2 e^{-\frac{1}{2}(x+d)^T Q^{-1}(x+d) - \lambda c^T (x+d)}. \tag{V.6}$$

We know that $\rho_2(x+d) < \rho_1(x)$ along the hyperplane $c^T x = \Delta_1/2$. Let us show that this inequality in fact holds for all x such that $c^T x \geq \Delta_1/2$. Note that the equation $\rho_1(x) = \rho_2(x+d)$ is linear since the quadratic terms in (V.1) and (V.6) are the same. Therefore, the graphs of (V.1) and (V.6), when extended to the whole space, can intersect only along a hyperplane. It remains to determine the location of that hyperplane. Since the graph of (V.6) is obtained from that of (V.1) by vertical stretching and shifting by a negative multiple of the vector c, they must have an intersection point in the region $c^T x < \Delta_1/2$. The hyperplane along which they coincide cannot have any common points with $c^T x = \Delta_1/2$. It follows that whenever $c^T x > \Delta_1/2$ we have $\rho_2(x+d) < \rho_1(x)$. Similarly, we can show that $\rho_2(x-d) < \rho_1(x)$ whenever $c^T x < -\Delta_1/2$. This proves the proposition. $\square$

A nontrivial stochastic optimal control problem consists, for example, in choosing the value of Δ that minimizes a quadratic cost functional of the form $L = \alpha V(\Delta) + \sqrt{1-\alpha^2}U(\Delta)$, where $0 \leq \alpha \leq 1$. Although this problem does not seem to lend itself to an analytic solution, some numerical results can be obtained.

EXAMPLE 1. Consider the equation $dx = -xdt + dw - q(x)dt$, where $x \in \mathbb{R}$ and q is a quantizer with saturation value 1 as before. The diagrams below display, respectively, the optimal sensitivity Δ_{opt} as a function of α, and the optimal cost L_{opt} as a function of α (compared with $L = \alpha/2$ for the uncontrolled system $dx = -xdt + dw$).

Optimal sensitivity.

Optimal cost.

Naturally, the quantized feedback control strategy proves most effective when the weight attached to the control cost is small compared to the weight attached to the variance cost: if $\alpha = 1$, choosing $\Delta = 0$ gives $L \approx 0.18$ as opposed to 0.5 for the uncontrolled system.

Since explicit formulae for the steady-state probability densities are available, a number of other optimization problems can be investigated similarly. For example, one could fix Δ and take the proporionality constant λ that appears in the compatibility condition (IV.5) to be the varying parameter (this is meaningful if we can adjust the norm of the vector b). Then it is not hard to show that both the variance cost V and the control cost U decrease as $||b||$ becomes large.

In all of the optimal control problems considered above, the vector c has been fixed. However, the following problem appears interesting: given the system (IV.3) and a fixed sensitivity Δ, choose c so as to minimize the variance cost V, subject to the constraints $||c|| = 1$ and $\lambda = 1$ (b will then be fixed by (IV.5)). Using the formula (IV.7) and an argument similar to the one given in the proof of Proposition 1, one can show that the optimal c is the eigenvector of Q with the smallest eigenvalue.

3. Attention cost. In this section we look at another criterion for evaluating the performance of the system (IV.3). Every time the solution trajectory crosses one of the switching hyperplanes $c^T x = \pm\Delta/2$, we need to communicate to the controller a request to change the current control policy. This reflects the amount of "attention" needed for implementing the given control law (a similar idea is exploited in [16] in the context of deterministic systems with smooth control functions). Making use of the fact that the steady-state probability density is an even function of x, let us define the *attention cost* to be $C = 2\mathcal{E}C_{\Delta/2}(c^T x(t))$, where $\mathcal{E}C_u(\xi(t))$ stands for the mean number of crossings of a level u per unit time by a scalar stochastic process $\xi(t)$.

Since the expectation is computed with respect to the steady-state probability distribution, we may treat the process $c^T x(t)$ as stationary, assuming that it "has reached steady state". Therefore, we may use the Rice's formula for the mean number of crossings [19]

$$\mathcal{E}C_u(\xi(t)) = \frac{1}{\pi}\sqrt{\frac{-\frac{d^2 r(\tau)}{d\tau^2}\Big|_{\tau=0}}{r(0)}} \cdot e^{-u^2/2r(0)}$$

where $r(\tau) = \mathcal{E}\xi(t)\xi(t+\tau)$ is the autocorrelation function associated with a stationary stochastic process $\xi(t)$. In our case $r(\tau) = \lim_{t\to+\infty} c^T \mathcal{E}x(t)x^T(t+\tau)c$.

Let us first study the following question: when is $\mathcal{E}C_{\Delta/2}(c^T x(t))$ finite? We will need the following easy statement.

Lemma 3. *Assume that* $\frac{d^2 r(\tau)}{d\tau^2}$ *exists in some neighborhood of 0, possibly excluding 0 itself. Then* $\mathcal{E}C_{\Delta/2}(c^T x(t)) < \infty$ *if and only if* $\frac{dr(\tau)}{d\tau}\Big|_{\tau=0} = 0$.

Proof. Since $r(\tau)$ is an even function, it is easy to see that, in order for the second derivative of $r(\tau)$ to exist at the origin, it is necessary and sufficient for the first derivative to be zero at the origin. $\square$

EXAMPLE 2. Consider the equation $dx = -xdt + dw$, where $x \in \mathbb{R}$. We have $\frac{dr(\tau)}{d\tau}\Big|_{\tau=0} = -\mathcal{E}x(t)x(t+\tau)\big|_{\tau=0} = -\mathcal{E}x^2(t) < 0$, so the condition of Lemma 3 is not satisfied.

Let us see whether a control term might help. For the equation

$$dx = -xdt + dw - bq(x)dt \tag{V.7}$$

we have $\frac{dr(\tau)}{d\tau}\Big|_{\tau=0} = -\mathcal{E}x(t)x(t+\tau)\big|_{\tau=0} - b\mathcal{E}x(t)q(x(t+\tau))\big|_{\tau=0} = \mathcal{E}x(t)(-x(t) - bq(x(t)))$. Thus if the deterministic part of (V.7) is to be asymptotically stable, we simply cannot have a finite expected number of crossings of any level. We see that it is in fact a nontrivial task to construct a control system with a finite attention cost.

Consider a general linear system

$$dx = Axdt + Gdw$$

$$y = c^T x$$

where $x \in \mathbb{R}^n$ and $w = (w_1, \ldots, w_n)$ is a vector of independent Wiener processes. We have $\lim_{t\to\infty} \mathcal{E}y(t)y(t+\tau) = \lim_{t\to\infty} c^T\mathcal{E}x(t)x(t+\tau)c = c^TQe^{A^T\tau}c$, where Q is the steady-state variance matrix satisfying the equation (IV.4). Therefore

$$\frac{dr(\tau)}{d\tau}\Big|_{\tau=0} = c^TQA^Tc = \langle c, AQc\rangle. \tag{V.8}$$

Premultiplying (IV.4) by c^T and postmultiplying by c, we obtain

$$2\langle c, AQc\rangle = -\langle G^Tc, G^Tc\rangle. \tag{V.9}$$

Thus the condition of Lemma 3 is satisfied if and only if $G^T c = 0$. In particular, in the single-noise case ($G = g \in \mathbb{R}^n$) we will have a finite attention cost if $g \perp c$, i.e., if the noise vector g is parallel to the switching hyperplanes. On the other hand, if G is a nonsingular $n \times n$ matrix, the attention cost will always be infinite.

The above discussion suggests modifying the equations of Example 2 in the following way: replace $q(x)$ by $q(y)$, where the noise does not enter the equation for y explicitly. Namely, consider the following pair of equations:

$$\begin{aligned} dx &= -x dt + dw, \\ dy &= \beta x dt - \beta y dt \end{aligned} \tag{V.10}$$

where β is a positive constant (so that the nonrandom part of the system is asymptotically stable). Since $g = \binom{1}{0}$ and $c = \binom{0}{1}$ are orthogonal, we have $\mathcal{E} C_u(y) < \infty$ by virtue of (V.8), (V.9), and Lemma 3.

Now, let us replace (V.7) by the system

$$\begin{aligned} dx &= -x dt + dw - bq(y) dt, \\ dy &= \beta x dt - \beta y dt. \end{aligned} \tag{V.11}$$

It is interesting to observe that (V.11) is automatically compatible: $g \perp c$ implies $c \perp AQc$ by (V.9). Therefore, $g \parallel AQc$ since the state space of (V.11) is $\mathbb{R}^2$. Thus the compatibility condition (IV.5) is satisfied whenever $g \parallel b$, which is indeed the case. Notice that we did not have to compute the matrix Q to establish compatibility. In fact, in the case of the general quantized output feedback control system (IV.3) in $\mathbb{R}^n$ compatibility together with $b \perp c$ imply a finite attention cost. If $n = 2$ then $c \perp b$ and $G^T c = 0$ imply both compatibility and a finite attention cost. This interplay between the properties of compatibility and finite attention cost is to be investigated further.

We have shown that the mean number of crossings per unit time associated with (V.10) is finite. Since $q(c^T x)$ is piecewise constant, one might expect that the mean number of crossings for (V.11) would be finite too, although its value may be different. This is indeed true as we now show. The linear theory applied to (V.10) yields the steady-state probability density

$$\rho(x,y) = Ne^{-(\beta+1)x^2+2(\beta+1)xy-\frac{(\beta+1)^2}{\beta}y^2} = Ne^{-(\beta+1)(x-y)^2-\frac{\beta+1}{\beta}y^2}.$$

We have

$$\begin{aligned}
&\frac{d}{d\tau}\Big|_{\tau=0}\mathcal{E}y(t)y(t+\tau) = \beta\mathcal{E}(x(t)-y(t))y(t) \\
&= \beta N \iint_{\mathbb{R}^2}(x-y)ye^{-(\beta+1)(x-y)^2-\frac{\beta+1}{\beta}y^2}dxdy = \beta N \iint_{\mathbb{R}^2} zye^{-(\beta+1)z^2-\frac{\beta+1}{\beta}y^2}dzdy \\
&= \beta N \int_{-\infty}^{\infty} ze^{-(\beta+1)z^2}dz \cdot \int_{-\infty}^{\infty} ye^{-\frac{\beta+1}{\beta}y^2}dy = 0
\end{aligned}$$

where we have used the substitution $z = x - y$. As we have seen in Chapter IV, in the presence of quantized output feedback $q(y)$ we will have an additional term in the exponent, but that term will only depend on y and not on z. Thus the first integral in the last formula will still be zero, and therefore the attention cost for (V.11) will still be finite.

It is not very difficult to compute this attention cost directly using our knowledge of the steady-state density $\rho(x, y)$ associated with (V.11). We have

$$\mathcal{E}y^2 = \frac{\beta}{2(\beta+1)} + b^2\mathcal{P}(|y| > \Delta/2) - \frac{\mu\beta b}{\beta+1} e^{-\dfrac{(\beta+1)\Delta^2}{4\beta}} \sqrt{\frac{\pi}{\beta+1}}$$

where $\mu = \rho(0, 0)$. The attention cost is given by

$$C = \frac{\sqrt{\beta}}{\pi} \sqrt{\frac{\frac{\beta}{2(\beta+1)}}{\mathcal{E}y^2}} \cdot 2\, e^{-\dfrac{\Delta^2}{8\mathcal{E}y^2}}.$$

The picture below shows the graph of the attention cost C as a function of Δ and b (with β taken to be 1). The corresponding axes are labeled d and b, respectively.

Attention cost.

We conclude the chapter by an informal discussion of the properties of C revealed by this picture. Suppose that we fix Δ and look at the behaviour of C

as b varies. We see that C is small for small values of b, because the solution trajectories can stay outside the inner stripe $|c^T x| < \Delta/2$ for long periods of time. As b increases, the trajectories have to return to the inner stripe faster, which results in increasing attention cost. However, as we increase the gain further, the attention cost becomes smaller again and keeps decreasing for bigger values of b. Interestingly, a justification of this latter phenomenon is much less obvious. Solving the deterministic counterpart of (V.10)

$$\dot{x} = -x$$

$$\dot{y} = \beta x - \beta y$$

we have

$$y(t) = \begin{cases} \dfrac{\beta}{\beta - 1}(e^{-t} - e^{-\beta t})x(0) + e^{-\beta t}y(0) & \text{if } \beta \neq 1 \\ te^{-t}x(0) + e^{-t}y(0) & \text{if } \beta = 1 \end{cases} \tag{V.12}$$

Each of the functions given by the first terms on the right-hand side of both formulae in (V.12) has a global maximum for a certain $t > 0$, the maximal value being proportional to $x(0)$. Therefore, a solution trajectory which enters the inner stripe far from the y-axis, especially in the I and III quadrants, is likely to leave it again at a later time. However, if we increase b, solution trajectories in those quadrants tend to enter the inner stripe at points closer to the y-axis, so that they are less likely to leave again, hence the decrease in the cost. The smaller the value of Δ is, the bigger value of b is required to achieve the maximal value of C, and the bigger this maximal value is. The reason for this is also seen from the above analysis. On the other hand, if we fix b and increase Δ, the attention cost becomes smaller, by virtue of the fact that the trajectories spend more time in the inner stripe.

One can employ numerical techniques similar to the ones used above to treat stochastic optimal control problems that arise when the cost functional is taken to be a linear combination of the variance cost V, the control cost U, and the attention cost C.

CONCLUSIONS AND FUTURE WORK

We now briefly summarize the essential developments of Chapters II–V and indicate some directions for future research.

Quantized feedback stabilization (Chapter II). We studied the problem of stabilizing a linear system with saturating quantized feedback. We proposed a new control strategy which consists in changing the sensitivity of the quantizer while the system evolves. The equation that describes the evolution of the sensitivity with time (discrete rather than continuous in most cases) is interconnected with the given system (either continuous or discrete), thus resulting in a hybrid system. When applied to systems that are stabilizable by linear time-invariant feedback, our approach yields global asymptotic stability.

One could try to analyze the effectiveness of the methods proposed here, e.g., from the point of view of time-optimality or the "attention cost" involved. In the recent literature, quantized measurements are often considered as "partial observations" rather than "approximations" of the state of the system. Techniques from ergodic theory of dynamical systems [20] and information theory [22] have proved to be useful in learning the system's behavior in the presence of round-off errors. It would be interesting to look at our developments from this perspective.

Spectral theory of Fokker-Planck operators (Chapter III). We studied certain families of linear second-order differential operators arising from Fokker-Planck operators associated with linear stochastic differential equations. We constructed a basis in the Hilbert space of square-integrable functions using generalized Hermite polynomials, and obtained a representation for these operators from which their eigenvalues and eigenfunctions can be computed. The operator families under consideration can be realized as solutions of differential equations in the so-called *double bracket form* [2], [11]. The time variable of the corresponding flow can be interpreted as the *temperature* of the system. We proved that at least part of the spectrum of the Fokker-Planck operator is preserved under this flow. A more complete understanding is needed of what our results imply about the rate of convergence of the underlying stochastic process to steady state.

Steady-state probability distributions (Chapter IV). We addressed the question of existence of steady-state probability distributions for systems perturbed by white noise. We described a class of nonlinear feedback systems for which an explicit formula for the steady-state probability density can be found. These are systems that satisfy what we called the *compatibility condition.* In the case of quantized feedback, compatibility amounts to the matching of gaussian probability densities along hyperplanes in the state space. Compatible systems include what has been called monotemperaturic systems in earlier work [7]. We also established relationships between the steady-state probability densities and Liapunov functions for the corresponding deterministic systems.

In [41], a (non-constructive) Liapunov-like criterion was given for the existence of a steady-state probability distribution. An important objective for future research is to develop a perturbation theory that would allow one to obtain specific information about steady-state probability densities for systems that are not compatible but satisfy this criterion. Some preliminary results in this direction are reported in Section 9. Other problems that are largely open include the spectral gap problem of Section 7 and the “compatible stabilization” problem of Section 8.

Stochastic optimal control (Chapter V). We discussed stochastic optimal control problems that consist in optimizing the steady-state performance of a compatible quantized feedback system with respect to various performance measures. Arguably, the most interesting and challenging among them is the problem of minimizing the *attention cost*, i.e., the expected number of times the control function changes its value per unit interval of time. The expression for this cost can be obtained via the Rice's formula for the mean number of crossings of a level by a stationary stochastic process [19]. We used numerical techniques to carry out an initial investigation; some analytical tools need to be employed for a more thorough examination of the problem.

Appendix I

In view of Example 1 of Chapter III, it is of interest to show directly that the system (IV.2) is compatible whenever it is gradient. Consider the system

$$\dot{x} = Ax + gI\dot{w} + bf(c^T x)$$

such that $\frac{\partial \phi}{\partial x_i} = -\sum_{j=1}^{n} A_{ij}x_j - b_i f(c^T x)$ for some function $\phi(x)$. Then

$$\frac{\partial \phi}{\partial x_j x_i} = -A_{ij} - b_i c_j f'(c^T x)$$

and this has to be equal to

$$\frac{\partial \phi}{\partial x_i x_j} = -A_{ji} - b_j c_i f'(c^T x).$$

It is now readily seen that A must be symmetric, and that we must have $b = \mu c$ for some $\mu \in \mathbb{R}$. Now, notice that the steady-state variance equation (IV.4) amounts to $AQ + QA^T = -g^2 I$. Since A is nondegenerate by assumption, $Q = -\frac{g^2}{2}A^{-1}$ satisfies this equation, resulting in $AQ = -\frac{g^2}{2}I$. But then it follows that $\mu c = -\frac{2}{g^2}\mu AQc$, so that letting $\lambda = -\frac{2}{g^2}\mu$ we obtain $b = \lambda AQc$ as needed.

Appendix II

For the sake of completeness, let us see why, if (IV.24) holds, the system (IV.20) is indeed compatible in the sense of (IV.5). We have

$$(AQ)_{ij} = \sum_{k=1}^{n} A_{ik}Q_{kj} = \begin{cases} Q_{i+1,j} & \text{if } i < n \\ -\sum_{k=1}^{n} p_{k-1}Q_{kj} & \text{if } i = n \end{cases} \tag{V.13}$$

From (IV.20) and (IV.4) it follows that $(AQ)_{ij} = -(AQ)_{ji}$ unless $i = j = n$. This gives

$$Q_{12} = Q_{21} = (AQ)_{11} = 0$$

$$Q_{23} = Q_{32} = (AQ)_{22} = 0$$

$$Q_{14} = Q_{41} = (AQ)_{31} = -(AQ)_{13} = -Q_{23} = 0$$

and so on. Proceeding in this fashion, we can show that in fact

$$Q_{ij} = 0 \quad \text{whenever} \quad i + j \ \text{ is odd.} \tag{V.14}$$

Now,

$$(AQc)_i = \sum_{j=1}^{n} (AQ)_{ij}c_{j-1} = (AQ)_{i1}p_1 + (AQ)_{i3}p_3 + \cdots + (AQ)_{ir}p_r \tag{V.15}$$

where $r = 2\lfloor (n-1)/2 \rfloor + 1$. We took $a = 1$ in (IV.24) for simplicity.

Let us prove that $(AQc)_i = 0$ for any $i < n$, which would establish compatibility. We need to consider several possible cases.

1. i is odd. Using (V.13), we can rewrite (V.15) as $(AQc)_i = Q_{i+1,1}p_1 + Q_{i+1,3}p_3 + \cdots + Q_{i+1,r}p_r$, and each term is zero by (V.14).

2. i is even, n is even. We can rewrite (V.15) as

$$\begin{aligned}
(AQc)_i &= -(AQ)_{1i}p_1 - (AQ)_{3i}p_3 - \cdots - (AQ)_{n-1,i}p_{n-1} \\
&= -Q_{2i}p_1 - Q_{4i}p_3 - \cdots - Q_{ni}p_{n-1} \\
&= -Q_{1i}p_0 - Q_{2i}p_1 - Q_{3i}p_2 - Q_{4i}p_3 - \cdots - Q_{ni}p_{n-1} \quad \text{(see (V.14))} \\
&= (AQ)_{ni} = -(AQ)_{in} = -Q_{i+1,n} = 0.
\end{aligned}$$

3. i is even, n is odd. In this case the last term of (V.15) that corresponds to $c_{n-1} = 1$ should be treated separately:

$$-Q_{1i}p_0 - Q_{2i}p_1 - \cdots + Q_{i+1,n} = -Q_{i+1,n} + Q_{i+1,n} = 0.$$

This completes the proof.

References

1. G. Birkhoff, G.-C. Rota, *Ordinary Differential equations*, Ginn and Co., Boston, Mass., 1962.
2. A. M. Bloch, R. W. Brockett, T. S. Ratiu, *Completely integrable gradient flows*, Comm. Math. Phys. **147** (1992), 57–74.
3. R. W. Brockett, *On the stability of nonlinear feedback systems*, IEEE Trans. Applications and Industry **83** (1964), 443–449.
4. R. W. Brockett, *Finite Dimensional Linear Systems*, John Wiley, New York, 1970.
5. R. W. Brockett, *Lie algebras and Lie groups in control theory*, Geometric Methods in System Theory (D. Q. Mayne and R. W. Brockett, Eds.), Reidel Publishing, Dordrecht, Netherlands, 1973, 43–82.
6. R. W. Brockett, *Parametrically stochastic differential equations*, Math. Programming Stud. **5** (1976), 8–21.
7. R. W. Brockett, J. C. Willems, *Stochastic control and the second law of thermodynamics*, Proc. 1978 IEEE Conf. Decision and Control, New York: IEEE, 1978, pp. 1007–1011.
8. R. W. Brockett, *Nonlinear systems and nonlinear estimation theory*, Stochastic Systems: the Mathematics of Filtering and Identification and Applications (Les Arcs, 1980), NATO Adv. Study Inst. Ser. C: Math. Phys. Sci., 78, Reidel, Dordrecht–Boston, Mass., 1981, pp. 441–477.
9. R. W. Brockett, C. I. Byrnes, *Multivariable Nyquist criteria, root loci, and pole placement: a geometric viewpoint*, IEEE Trans. Automat. Control **AC-26** (1981), 271–283.
10. R. W. Brockett, *Nonlinear control theory and differential geometry*, Proc. International Congress of Mathematicians, Warszawa, August 16–24, 1983, pp. 1357–1368.
11. R. W. Brockett, *Dynamical systems that sort lists, diagonalize matrices, and solve linear programming problems*, Linear Algebra Appl. **146** (1991), 79–91.
12. R. W. Brockett, *Hybrid models for motion control systems*, Essays on control: perspectives in the theory and its applications (Groningen, 1993), 29–53, Progr. Systems Control Theory, 14, Birkhäuser Boston, Boston, MA, 1993.
13. R. W. Brockett, *Lecture Notes on Stochastic Control*, Harvard University, Cambridge, MA, 1995.
14. R. W. Brockett, *Oscillatory descent for function minimization*, Current and future directions in applied mathematics (M. Alber et al., Eds.), 65–82, Birkhäuser Boston, Boston, MA, 1997.
15. R. W. Brockett, *Notes on stochastic processes on manifolds*, Systems and control in the twenty-first century (St. Louis, MO, 1996), 75–100, Progr. Systems Control Theory, 22, Birkhäuser Boston, Boston, MA, 1997.
16. R. W. Brockett, *Minimum attention control*, Proc. 36th Conf. Decision and Control, San Diego, CA, 1997, pp. 2628–2632.
17. R. Y. Chang, M. L. Wang, *The properties of modified Hermite polynomials and their applications to functional differential equations*, J. Chinese Inst. Engrs **9** (1986), 75–81.
18. J.-H. Chou, S.-H. Chen, I.-R. Horng, *Robust stability bound on linear time-varying uncertainties for linear digital control systems under finite wordlength effects*, JSME Internat. J. Series C **39** (1996), 767–771.
19. H. Cramér, M. R. Leadbetter, *Stationary and Related Stochastic Processes*, John Wiley, New York, 1967.
20. D. F. Delchamps, *Stabilizing a linear system with quantized state feedback*, IEEE Trans. Automat. Control **35** (1990), 916–924.

21. R. Felipe, *Algebraic aspects of Brockett type equations*, preprint, March 1997.
22. X. Feng, K. A. Loparo, *Active probing for information in control systems with quantized state measurements: a minimum entropy approach*, IEEE Trans. Automat. Control **42** (1997), 216–238.
23. A. F. Filippov, *Differential Equations with Discontinuous Righthand Sides*, Mathematics and its Applications (Soviet Series), 18, Kluwer Academic Publishers, Dordrecht, 1988.
24. I. I. Gikhman, A. V. Skorokhod, *Stochastic Differential Equations*, Springer-Verlag, New York, 1972.
25. H. W. Gould, A. T. Hopper, *Operational formulas connected with two generalizations of Hermite polynomials*, Duke Math. J. **29** (1962), 51–63.
26. J. K. Hale, M. A. Cruz, *Existence, uniqueness and continuous dependence for hereditary systems*, Ann. Mat. Pura Appl (4) **85** (1970), 63–81.
27. R. Z. Has'minskii, *Stochastic Stability of Differential Equations*, Sijthoff & Noordhoff, Alphen aan den Rijn, Netherlands, 1980.
28. R. A. Holley, S. Kusuoka, D. W. Stroock, *Asymptotics of the spectral gap with applications to the theory of simulated annealing*, J. Funct. Anal. **83** (1989), 333–347.
29. L. V. Kantorovich, G. P. Akilov, *Functional Analysis*, Pergamon Press, Oxford, England, 1982.
30. J. P. LaSalle, *Some extensions of Liapunov's second method*, IRE Trans. Circuit Theory **CT-7** (1960), 520–527.
31. A. S. Morse, *Dwell-time switching*, Proc. 2nd European Control Conf., 1993, pp. 176–181.
32. Y. Nakamura, *Completely integrable gradient flows on the manifolds of Gaussian and multinomial distributions*, Japan J. Indust. Appl. Math. **10** (1993), 179–189.
33. E. D. Rainville, *Infinite Series*, Macmillan Co., New York, 1967.
34. J. Raisch, *Control of continuous plants by symbolic output feedback*, Hybrid Systems II, Springer-Verlag, Berlin, 1995, pp. 370–390.
35. C. E. Rohrs, J. L. Melsa, D. G. Schultz, *Linear Control Systems*, McGraw-Hill, New York, 1993.
36. A. D. Rueda, *A class of two dimensional vector fields which are Fokker-Planck integrable*, Ann. Univ. Sci. Budapest Sect. Math. **36** (1993), 133–137.
37. R. L. Stratonovich, *A new representation for stochastic integrals and equations*, SIAM J. Control **4** (1966), 362–371.
38. W. A. Wolovich, *Linear Multivariable Systems*, Springer-Verlag, New York, 1974.
39. W. S. Wong, R. W. Brockett, *Systems with finite communication bandwidth constraints I: State estimation problems*, IEEE Trans. Automat. Control **42** (1997), 1294–1299.
40. W. S. Wong, R. W. Brockett, *Systems with finite communication bandwidth constraints II: Stabilization with limited information feedback*, IEEE Trans. Automat. Control, to appear.
41. M. Zakai, *A Liapunov criterion for the existence of stationary probability distributions for systems perturbed by noise*, SIAM J. Control **7** (1969), 390–397.