ABSTRACT

Title of dissertation: NONLINEAR DETECTION, ESTIMATION, AND CONTROL FOR FREE-SPACE OPTICAL COMMUNICATION

Arash Komaee, Doctor of Philosophy, 2008

Dissertation directed by: Professor P. S. Krishnaprasad
Professor Prakash Narayan
Department of Electrical and Computer Engineering

In free-space optical communication, the intensity of a laser beam is modulated by a message, the beam propagates through free-space or atmosphere, and eventually strikes the receiver. At the receiver, an optical sensor converts the optical energy into an electrical signal, which is processed to reconstruct the original message. The promising features of this communication scheme such as high-bandwidth, power efficiency, and security, render it a viable means for high data rate point-to-point communication.

In this dissertation, we adopt a stochastic approach to address two major issues associated with free-space optics: digital communication over an atmospheric channel and maintaining optical alignment between the transmitter and the receiver, in spite of their relative motion. Associated with these issues, we consider several detection, estimation, and optimal control problems with point process observations. Although these problems are motivated by applications in free-space optics, they are also of direct relevance to the general field of estimation theory and stochastic

control.

We study the detection aspect of digital communication over an atmospheric channel. This problem is formulated as an M-ary hypothesis testing problem involving a doubly stochastic marked and filtered Poisson process in white Gaussian noise. The formal solutions we obtain for this problem are hard to express in an explicit form, thus we approximate them by appropriate closed form expressions. These approximations can be implemented using finite-dimensional, nonlinear, causal filters.

Regarding the optical alignment issue, we consider two problems: active pointing and cooperative optical beam tracking. In the active pointing scheme that we develop for short range applications, the receiving station estimates the center of its incident optical beam based on the output of a position-sensitive photodetector. The transmitter receives this estimate via an independent communication link and incorporates it to accurately aim at the receiving station.

A cooperative optical beam tracking system consists of two stations in such a manner that each station points its optical beam toward the other one. The stations employ the arrival direction of the incident optical beams as a guide to precisely point their own beam toward the other station. We develop a detailed stochastic model for this system and employ it to determine a control law which maximizes the flow of optical energy between the stations. In so doing, we consider the effect of light propagation delay, which requires a point-ahead mechanism to compensate for the displacement of the receiving station during propagation time.

NONLINEAR DETECTION, ESTIMATION, AND CONTROL FOR FREE-SPACE OPTICAL COMMUNICATION

by

Arash Komaee

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2008

Advisory Committee:

Professor P. S. Krishnaprasad, Co-Chair/Advisor
Professor Prakash Narayan, Co-Chair/Co-Advisor
Professor Thomas E. Murphy
Professor Steve Marcus
Professor Rajarshi Roy

DEDICATION

To my parents
for their endless love, encouragement, and support

ACKNOWLEDGMENTS

I would like to express my sincere gratitude to my advisors Professor P. S. Krishnaprasad and Professor Prakash Narayan for their thoughtful guidance, criticism of my work, and every means of support. Their liberal attitude and unquestionable ethics and integrity made working with them a pleasure. It has been my privilege to have them as my mentors, and I endeavor to follow their admirable professional standards.

I am grateful to my dissertation committee members Professor Thomas Murphy, Professor Steve Marcus, and Professor Rajarshi Roy for their time and instructive comments during the defense. I am particularly in debt to Professor Marcus for his suggestions that improved this dissertation and Professor Murphy for the invaluable discussions we had during the course of my research.

I would like to express my gratitude and affection to my parents who have been my primary teachers. Without their continued support and encouragement, I doubt I could have completed my higher education.

I acknowledge the financial support of my research by the Army Research Office under ODDR&E MURI01 Program Grant No. DAAD19-01-1-0465 to the Center for Communicating Networked Control Systems (through Boston University), by the National Science Foundation under Grant No. ECS 0636613, by the Army Research Office under ARO Grant No. W911NF0610325, and by the Office of Naval Research under the ODDR&E MURI2007 Program Grant No. N000140710734.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

Free-space optics is regarded as a high-bandwidth and power efficient means for point-to-point communication with a wide range of applications including fixed-location terrestrial communication [1], communication between mobile robots [2], airborne communication [3], and intersatellite communication [4]. In this mode of communication, (digital) transmitting data modulates the instantaneous power of a laser beam, which propagates through free-space or atmosphere, and eventually strikes the receiver. At the receiver, an optical sensor (photodetector) converts the optical energy into an electrical signal, which is processed to reconstruct the original data.

Two major issues are associated with this communication scheme: optical fade caused by the atmosphere and misalignment of the stations (transmitter and receiver) due to their relative motion. This research investigates a stochastic approach in finding solutions to these problems. Although the detection, estimation, and control problems considered in this work are motivated by applications in free-space optical communication, they are of direct relevance to the general field of

estimation theory and stochastic control.

Throughout this chapter, we briefly explain these issues, the models adopted or developed to describe them, and our solutions to the associated problems. In the last section, we fix the notation that will be used in the following chapters.

1.1 Atmospheric Turbulent Channels

The atmosphere, as an optical medium, introduces random fluctuations in the power of the propagating optical field. The atmospheric turbulence caused by differential heating of the air is characterized in terms of a (slowly-varying lognormal) random process which modulates the optical power at the receiver. These random fluctuations (fade) are a characteristic feature of atmospheric channels in contrast with conventional fiber optic channels.

In general, the output of an optical sensor can be modeled by a marked and filtered Poisson process, which is a stream of randomly weighted narrow pulses arriving at the jump times of a Poisson process [5]. Also, the electronic circuit which follows the sensor, corrupts this signal by thermal noise which is modeled by an additive white Gaussian process. Thus, the problem of detecting digital signals over an atmospheric optical channel can be formulated as a M-ary hypothesis testing problem with observations which consists of a doubly stochastic marked and filtered Poisson process in additive white Gaussian noise.

While most prior studies of the detection problem above are based on some simplified version of the previous model [6, 7, 8, 9, 10] or involve a linearity constraint on the detector structure [11, 12, 13, 14, 15], a few tackle the problem in its

general form [5, 16, 17, 18]. A state-space approach developed in [5, 16] succeeds to formulate a solution in terms of a stochastic partial differential-difference-integral equation; however, the solution of this equation must be approximated using a finite-dimensional filter with an unclear approximation error.

We adopt another approach following that in [17, 18][1]. This approach is based on the fact that conditioned on the number and arrival times of the pulses, the problem reduces to one of M-ary detection of a deterministic signal in white Gaussian noise, which has a known solution. Then, averaging over the number and arrival times of the Poisson process, the solution to the original problem can be obtained. This leads to an expression involving an infinite sum of multiple integrals, which is hard to reduce to an explicit (closed form) expression; however, under certain assumptions, this infinite sum can be approximated by an explicit formula or a mathematically tractable equation.

In Chapter 2, after presenting the model of an atmospheric optical channel and stating its associated detection problem, we discuss our approach in more detail. Based on the infinite sum mentioned above, we establish mathematically tractable upper and lower bounds on the exact solution and study the behavior of these bounds for some important limiting cases. We show that in these limiting cases, the lower bound tends to the upper bound. This motivates us to approximate the exact solution with the upper (or lower) bound under conditions which closely approximate the limiting cases.

[1] In [17, 18], the avalanche gain of the optical sensor and the turbulent fade are not considered. The later is essential for an atmospheric channel.

Furthermore, in Chapter 2, we introduce a novel technique for expressing the infinite sum in terms of an expectation taken with respect to a stochastic process. This new expression is then used in order to develop several approximate solutions. We remark here that the results of Chapter 2 are directly applicable to the fiber optic channels as a special case of the atmospheric turbulent channels.

1.2 Optical Alignment

A major challenge in free-space optical communication is to maintain optical alignment between the stations despite their relative motion. This relative motion, caused by the mobility of the stations or mechanical vibration, can be comparable in magnitude to the size of the narrow laser beams employed by the optical link. Therefore, a closed-loop fine alignment mechanism is required to maintain the alignment after the link is established through a coarse open-loop alignment operation referred to as spatial acquisition.

The closed-loop fine alignment can be decomposed into two operations: active pointing and optical beam tracking. The goal of the first operation is to aim the transmitted beam toward the receiver within an acceptable accuracy, while the second operation is intended to maintain the transmitter within the field of view of the receiver. For the purpose of alignment (pointing and tracking), the receiving and transmitting optical devices are installed on electromechanical pointing assemblies[1], which adjust the direction of the devices according to control signals generated by

[1]Alternatively, the stations can be equipped with steerable flat mirrors to control the light direction.

appropriate closed-loop controllers.

The closed-loop controller employed by an optical alignment system is usually fed by the output of a position-sensitive photodetector (e.g. quadrant detector). The output of this device is normally modeled by a vector-valued point process or ideally by a space-time point process. Thus, the control aspect of an optical alignment system can be formulated in terms of a stochastic optimal control problem with point process observations. In the remainder of this section, we briefly explain the operations of optical beam tracking and active pointing, and our stochastic approach for a resolution of the associated problems.

1.2.1 Optical Beam Tracking

In free-space optics, optical beam tracking is an active operation with the goal of keeping the transmitter in the field of view of the receiver. Figure 1.1 illustrates a simple optical receiver employed in a free-space optical link. The receiver

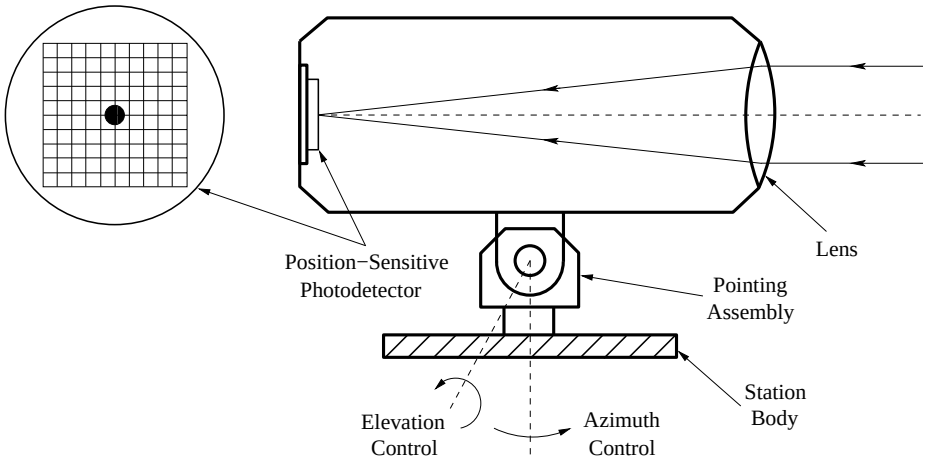

Figure 1.1: Schematic diagram of a simple optical receiver.

is equipped with a lens (or curved mirror) to focus the incident optical field on a position-sensitive photodetector. The position-sensitive photodetector is a photoelectron converter whose surface is partitioned into small regions (pixels) with independent outputs. The image of the incident optical field is a spot of light which randomly moves on the surface of the photodetector due to the relative motion of the transmitter and the receiver. In the absence of an adjusting mechanism, the effect of the relative motion might be large enough to move the spot of light beyond the surface of the photodetector. An active mechanism is needed to maintain the spot of light at the center of the photodetector by consistently adjusting the direction of the receiver. For this purpose, the receiver is installed on an electromechanical pointing assembly which controls the direction of the receiver (in azimuth and elevation), based on the control signals generated by a closed-loop controller. The closed-loop controller estimates the location of the spot of light from the output of the position-sensitive photodetector and determines a proper control in order to direct the spot of light toward the center of the photodetector.

The system explained above has been modeled by a linear state-space equation which is driven by a control vector and a vector-valued Wiener process [19]. Under the assumption that the photodetector has an infinite spatial resolution, the output of the photodetector has been described by a space-time point process whose rate is modulated by the state of the system [19]. Also, the goal of closed-loop control can be formulated in terms of minimizing a quadratic cost functional [19]. Assuming that the observation of the space-time point process is provided on $\mathbb{R}^2$ (which is practically justified) and that the rate of this process has a Gaussian profile, the

solution to this optimal control problem (and its associated estimation problem) is finite-dimensional and has been obtained in [20].

In Chapter 3, we discuss this problem in more detail and attempt to relax the assumption of a "Gaussian profile". This leads us to reformulate the state estimation problem in terms of estimating the state of a discrete-time linear model with additive white non-Gaussian measurement noise.

For a practical system with a finite resolution photodetector, the estimation and control problems above are infinite-dimensional, thus some sort of approximation is required to solve them. A possible approach to this approximation problem, specially for a high resolution photodetector, is to modify the results of [20] for a finite resolution photodetector[1]. This approach is motivated by the fact that the results of [20] are exact and are expressed in an explicit form.

1.2.2 Active Pointing

For short range applications, in which the size of the receiving aperture is comparable to the size of the optical beam, we develop an active pointing scheme in Chapter 4. In this scheme, the receiver is equipped with a position-sensitive photodetector in order to measure the intensity profile of its incident optical beam. The output of the photodetector is used to estimate the center of the received optical beam, whereupon the estimate is conveyed to the transmitter through an optical link or a low-bandwidth RF channel. Based on this estimate, a pointing assembly adjusts

[1]The alternative approach is to start from the stochastic partial differential equation which describes the temporal evolution of the posterior density of the state vector and try to approximate its solution using a finite-dimensional filter.

the transmitter direction with the goal of maintaining the center of the optical beam close to the center of the receiving aperture. Note that the pointing direction must be adjusted consistently in order to compensate for the relative motion between the transmitter and the receiver. The block diagram in Figure 1.2 illustrates this active pointing scheme.

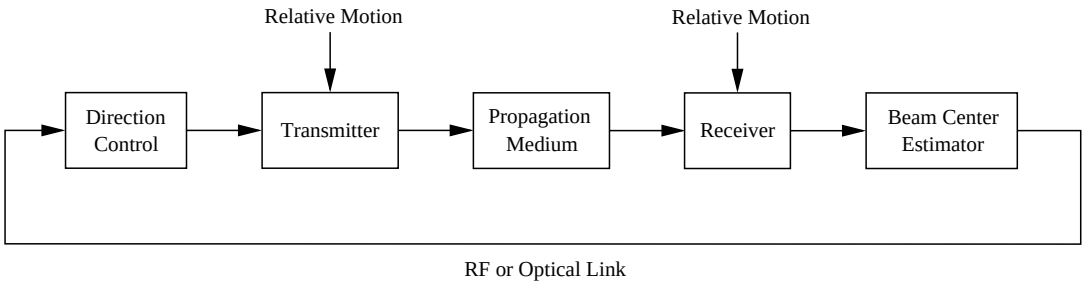

Figure 1.2: Active pointing scheme for a short range free-space optical channel.

We model the dynamics of the pointing assembly and the relative motion by a stochastic linear state-space equation. The observation of the optical intensity on the receiving aperture (photodetector output) will be modeled by a space-time point process with a rate depending on the state of the dynamical model. Also, we formulate the pointing control problem in terms of seeking a control law that minimizes a quadratic cost functional of the state and the control vectors.

We note that for active pointing, the observations are provided only over a subset of $\mathbb{R}^2$, in contrast to the optical beam tracking in which the observation is given over $\mathbb{R}^2$. This causes significant technical difficulties, since the solution to the estimation and control problems associated with this model are infinite-dimensional. For this reason, instead of an exact solution, we obtain an approximate estimator

and controller which are asymptotically optimal in the sense that they tend to the optimal estimator and controller, as the aperture tends to $\mathbb{R}^2$.

1.2.3 Cooperative Optical Beam Tracking

Cooperative optical beam tracking is a viable solution to the alignment problem, especially for long range free-space optical communication. An optical link which incorporates this alignment scheme consists of two stations in such a manner that each station points its optical beam toward the other one. The stations employ the arrival direction of the incident optical beams as a guide to precisely point their own beam toward the other station. In short range applications in which the light propagation delay is negligible, the transmitter points its optical beam along the arrival direction, while in long range applications with significant propagation delay (e. g., intersatellite communication), a point-ahead mechanism compensates for the displacement of the receiving station during propagation time. The concept of "cooperative optical beam tracking" will be explained in detail in Chapter 5, with reference to the architecture of the optical transceivers employed in this alignment method.

The model we develop in Chapter 5 for this alignment scheme consists of two dynamically coupled subsystems, such that each subsystem is modeled similar to an optical beam tracking system. In developing the model, we incorporate nonlinear effects, major sources of disturbance, light propagation delay, and the fluctuations of the optical intensity due to the modulation of data and optical fade. We believe that including these details in the modeling procedure leads to a fairly accurate

model. For a special and important case in which the relative motion can be decomposed into a predetermined large component and an unknown small component (e.g. intersatellite applications), the nonlinear dynamics of the system will be linearized around a nominal state trajectory. This linearized model[1] will be used later in Chapter 6.

While cooperative optical beam tracking has been already analyzed using simple deterministic models [21, 22], we shall use a stochastic approach in Chapter 6 in order to study this alignment scheme. The design goal is to obtain two controllers (one for each station) to maximize an objective functional which is defined as the expected flow of optical energy between the stations. Note that this control problem is decentralized in the sense that each station has access only to its own local observation. For a negligible propagation delay, we directly maximize the objective functional, while for the case of a significant propagation delay, we maximize a lower bound on the objective functional.

1.3 Notations

In the following chapters, dependence on time will be displayed by subscript t, e.g., a stochastic process or a deterministic signal will be denoted as $(\cdot)_t$. This convention will be occasionally violated in Chapter 2 by using $(\cdot)(t)$ to show the dependence on time or using subscripts i through n in order to index a variable or a time-signal over the integer set. In the later case, a time-signal will be denoted by $(\cdot)_i(t)$. All matrices will be denoted by capital letters and we shall occasionally use capital

[1]Note that only the dynamical equations are linearized, not the observation model.

letters for vectors or scalars. We shall not use distinct notation to differentiate between deterministic versus stochastic or vector versus scalar quantities; thus, the nature of an entity should be understood from its context.

Chapter 2

Nonlinear Detection for Digital Communication Over Optical Channels

2.1 Introduction

The receiving end of any optical communication link (fiber or free-space) is equipped with one of the several types of photodetectors (photoemissive, photovoltaic, or photoconductive) to convert the received optical power into an electrical signal. The output of a photodetector, regardless of its type, is a stream of narrow pulses which occur with a rate depending on the instantaneous optical power striking the surface of the device [5]. Each pulse of this stream corresponds to an electron generated through a photo-electron conversion. Avalanche photodiodes and photomultiplier tubes are designed in such a manner that each photo-generated charge carrier releases additional charge carriers [23]. This mechanism introduces an internal gain modeled by i.i.d. random variables which multiply the amplitude of the pulses [5]. In accordance with the description above, in the most general case, the output of a photodetector is modeled by a marked and filtered Poisson process [24] whose rate is modulated by the incident optical power.

Normally, the output of the photodetector is degraded by the thermal noise generated by the internal photodetector resistance, the amplifier circuit, and the load resistor [5]. The thermal noise is well-modeled by an additive, zero-mean, white Gaussian noise.

In free-space optical channels, atmospheric turbulence significantly perturbs the optical power at the receiving end of the link. Mathematically, this phenomenon is characterized in terms of a (lognormal) random process which modulates the optical power at the receiver [25, 26]. In this case, the channel output must be modeled as a doubly stochastic marked and filtered Poisson process [24] in additive white Gaussian noise.

In order to transmit a single symbol through a digital optical communication link, a waveform associated with the symbol is picked from a set of predetermined waveforms to modulate the power of the transmitting optical source during a symbol transmission time. At the receiving end, based on the channel output during the symbol transmission time, a detector decides which symbol was transmitted, in such a manner that the probability of erroneous decision is minimized. The structure and design of such a detector is the subject of this chapter.

During the last four decades, the detection problem above has been studied with different levels of model complexity. An idealized model which assumes infinite bandwidth, constant avalanche gain, and zero thermal noise for a photodetector, is provided by a (doubly stochastic) Poisson process. Using this simplified model, binary and M-ary hypothesis testing problems have been addressed in [6, 7, 8, 9, 10]. As a suboptimal solution to this hypothesis testing problem, some authors have

proposed a linear detector [11, 12, 13, 14, 15]. Gardner [11] introduced an equivalent linear model for a marked and filtered Poisson process and employed this model to develop a linear detector. In [12, 13, 14, 15], the linear detectors are designed based on minimizing the Chernoff bound for the probability of error.

Hoversten *et al.* [5], Kurimoto [16], and Bhanji [17] tackled the problem by exploiting a general formula due to Duncan [27] and Kailath [28] for the likelihood ratio function of a stochastic process in white Gaussian noise. They used a state-space approach to develop an approximate estimator which contributes to the Itô integral based estimator-correlator structure of the likelihood ratio function.

We approach the problem using the fact that conditioned on the Poisson point process associated with the photodetector output, the problem reduces to one of M-ary detection of a deterministic signal in white Gaussian noise, which has a known solution. Then, averaging over this point process, we obtain the solution to the original problem. This procedure leads to an expression involving an infinite sum of multiple integrals, which seems impossible to solve for an explicit expression. The method explained above has been already applied to a special case of the problem by Bhanji [17] and Hero [18], albeit with limited results.

In order to derive useful results from the mentioned infinite sum, we follow two different directions. In the first direction, we establish upper and lower bounds on the infinite sum in terms of two integral equations. Then, we show that the lower bound approaches the upper bound, as the pulse duration tends to zero. Based on this fact and for a small pulse duration, we approximate the infinite sum by the solution of an integral equation.

In the second direction, we introduce a new technique to rewrite the infinite sum in terms of an expectation taken with respect to a stochastic process. This formula which is suitable for further approximations, will be the point of departure for developing several approximate detectors, obtained under different set of assumptions.

2.2 The Model and Problem Statement

We consider an optical channel in which a nonnegative input signal s_t modulates the power of an optical source at the transmitter. The optical signal strikes a photodetector at the receiver, after propagating through an optical medium, which in addition to attenuation, introduces random fluctuations in the optical power (when the atmosphere is the propagation medium). For the purpose of amplification, the photodetector is followed by an electronic circuit, which gives rise to corrupting thermal noise. The output of this circuit is regarded as the channel output y_t.

The model we use for an optical channel (with some modification) is adopted from [5, 29]. In order to keep the description self-contained, we reproduce the model here. We first summarize the model and then discuss in more detail the physical significance of the model parameters.

2.2.1 Stochastic Model of an Optical Link

Let $\{N_t,\ t \geqslant 0\}$ be a doubly stochastic Poisson process with jump times $\{t_n\}_{n=1}^{N_t}$ and a stochastic rate

$$\lambda_t = \alpha s_t + \mu \tag{2.1}$$

where α is a nonnegative random variable with a known probability density function, $\{s_t\}$ is a nonnegative stochastic process regarded as the channel input, and μ is a known nonnegative constant. Let $\{q_n\}_{n=1}^{\infty}$ be an i.i.d. sequence of random variables with a known cumulative distribution function. Denote by $\{w_t\}$ a standard Wiener process and let σ to be a known constant. It is assumed that $\{t_n\}_{n=1}^{N_t}$, $\{q_n\}_{n=1}^{\infty}$, α, and $\{w_t\}$ are statistically independent. Moreover, $\{s_t\}$ is statistically independent of $\{q_n\}_{n=1}^{\infty}$, α, and $\{w_t\}$. Suppose that $\pi(\cdot)$ is a unit area[1] deterministic function such that $\pi(t) = 0$ for $t < 0$. Then, the channel output y_t can be modeled [5] by the stochastic differential equation[2]

$$dy_t = \sum_{n=1}^{N_t} q_n \pi(t - t_n)\, dt + \sigma dw_t. \tag{2.2}$$

We note that a doubly stochastic Poisson point process is fully characterized [24] by

$$\Pr\{N_t = 0 | \lambda_\tau, \tau \in [0,t]\} = \exp\left(-\int_0^t \lambda_\tau d\tau\right) \tag{2.3}$$

and for $n = 1, 2, 3, \ldots$,

$$\begin{aligned}\Pr\{N_t = n, t_i \in [\tau_i, \tau_i + d\tau_i); i = 1, 2, \ldots, n | \lambda_\tau, \tau \in [0,t]\} \\ = \left(\prod_{i=1}^{n} \lambda_{\tau_i} d\tau_i\right) \exp\left(-\int_0^t \lambda_\tau d\tau\right)\end{aligned} \tag{2.4}$$

when $0 \leqslant \tau_1 \leqslant \tau_2 \leqslant \cdots \leqslant \tau_n \leqslant t$, and 0, otherwise.

[1] This means that $\int_0^{\infty} \pi(t)\, dt = 1$.

[2] We take $\sum_{n=1}^{0}(\cdot)$ equal 0.

2.2.2 Significance of the Model Parameters

The nonnegative constant μ in (2.1) represents the sum of the photodetector dark current rate and the rate associated with the background radiation in the case of free-space channels [29]. Except for the atmospheric channel, the parameter α in (2.1) is a known constant which characterizes the multiplicative combination of the antenna gain, the path attenuation, and the photodetector sensitivity. For atmospheric channels, α is modeled by a nonnegative random variable to reflect the random fluctuations of the optical power caused by atmospheric turbulence. When the receiving aperture is smaller than the turbulence coherence length[1], α is a lognormal random variable [25] defined as $\alpha = \bar{\alpha}\mathrm{e}^{2\chi}$, where $\bar{\alpha} = \mathrm{E}[\alpha]$ and χ is a normal random variable with mean $-\sigma_\chi^2$ and variance σ_χ^2 [26]. Here, σ_χ^2 is a known constant depending on the wavelength of the light, the propagation distance, the refractive-index structure constant, and the shape of the optical field [30]. Note that the turbulent fade is a time-varying phenomenon; however, since its coherence time is much longer than the transmission interval [25], it can be accurately modeled as a random variable during the transmission of a single message. We remark that when the receiver possesses the perfect information of the channel fade, α can be modeled as a constant. Also, when imperfect information of the channel fade is provided to the receiver as an estimate for α, the distribution of α must be modified accordingly.

The integer-valued i.i.d. random variables $\{q_n\}_{n=1}^{\infty}$, stand for the random avalanche gains, i.e., the number of released charge carriers due to a single photo-

[1]For definition of the turbulence coherence length see [25, 30].

generated charge carrier. The probability distribution of q_n has been derived (theoretically) by McIntyre [31] and verified experimentally by Conradi [32].

The area under $\pi(\cdot)$ is equal to the charge of an electron multiplied by the gain of the amplifier which follows the photodetector. For sake of simplicity, we normalize this quantity to 1. The shape of pulses at the output of a practical photodetector vary from one pulse to another, i.e., the pulse shape is a random function [33]; however, since a reasonable stochastic model for the pulse shape is not available, $\pi(\cdot)$ is characterized by the average of the random pulses. For an avalanche photodiode, this averaged pulse shape is derived in [33, 34, 35].

The standard Wiener process $\{w_t\}$ in (2.2) represents the thermal noise generated by the amplifier which follows the photodetector. The strength of the thermal noise is characterized by the known constant σ^2.

2.2.3 Problem Statement

Suppose that a random message m taken from $\{1, 2, \ldots, M\}$ is to be transmitted through an optical channel during $[0, T]$. Here, the transmission time $T > 0$ is a known constant. Denote by p_m, $m = 1, 2, \ldots, M$, the prior probability of message m. Assume that a deterministic, nonnegative, bounded waveform $s_m(t)$ is assigned to each message $m = 1, 2, \ldots, M$. Then, in order to transmit m, we let $s_t = s_m(t)$ during $t \in [0, T]$.

Let $(\Omega, \mathscr{F}, P)$ be the underlying probability space for the stochastic model of Section 2.2.1. On this probability space, we define $\mathscr{Y}_T$ as the σ-algebra generated by $\{y_t\}$ during $[0, T]$. The goal is to obtain a $\mathscr{Y}_T$-measurable detection

rule $\hat{m}\left(\mathscr{Y}_T\right) \in \{1, 2, \ldots, M\}$, which minimizes the probability of error defined as $P_e = \Pr\left\{\hat{m}\left(\mathscr{Y}_T\right) \neq m\right\}$.

2.3 The Optimal Detection Law

In this section, we consider the hypothesis testing problem of Section 2.2.3 and determine a formal solution for it via Lemma 2.3.1 below. Then, through Theorems 2.3.1, 2.3.2, and 2.3.3, three different expressions will be presented for this solution.

Lemma 2.3.1. For $i = 1, 2, \ldots, M$, consider the binary hypothesis testing problem

$$\begin{aligned} H_i: &\quad dy_t = \sum_{n=1}^{N_t} q_n \pi\left(t - t_n\right) dt + \sigma dw_t, \quad \lambda_t = \alpha s_i\left(t\right) + \mu, \quad t \in [0, T] \\ H_0: &\quad dy_t = \sigma dw_t, \quad t \in [0, T] \end{aligned} \tag{2.5}$$

and let $L_i\left(T\right)$ denote its associated likelihood ratio function given $\mathscr{Y}_T$. Then, the optimal detection rule which minimizes the probability of error $P_e = \Pr\left\{\hat{m}\left(\mathscr{Y}_T\right) \neq m\right\}$ is given by the maximum a posteriori estimator

$$\hat{m}\left(\mathscr{Y}_T\right) = \arg \max_{i=1,2,\ldots,M} p_i L_i\left(T\right). \tag{2.6}$$

Proof. Define $\delta\left(i, j\right)$ such that $\delta\left(i, i\right) = 1$ and $\delta\left(i, j\right) = 0$ for $i \neq j$. Then, the

probability of error P_e can be expressed as

$$\begin{aligned} P_e &= \mathrm{E}\left[1 - \delta\left(m, \hat{m}\right)\right] \\ &= 1 - \mathrm{E}\left[\mathrm{E}\left[\delta\left(m, \hat{m}\left(\mathscr{Y}_T\right)\right)|\mathscr{Y}_T\right]\right] \\ &= 1 - \mathrm{E}\left[\sum_{i=1}^{M} \delta\left(i, \hat{m}\left(\mathscr{Y}_T\right)\right) \Pr\left\{m = i|\mathscr{Y}_T\right\}\right]. \end{aligned}$$

By the chain rule for likelihood ratios [36], we can write

$$\frac{\Pr\left\{m = i|\mathscr{Y}_T\right\}/p_i}{\Pr\left\{m = j|\mathscr{Y}_T\right\}/p_j} = \frac{L_i\left(T\right)}{L_j\left(T\right)}$$

which leads to

$$P_e = 1 - \mathrm{E}\left[\frac{\sum_{i=1}^{M} \delta\left(i, \hat{m}\left(\mathscr{Y}_T\right)\right) p_i L_i\left(T\right)}{\sum_{i=1}^{M} p_i L_i\left(T\right)}\right].$$

In order to minimize P_e, the sum in the numerator of the expression above must be maximized, which results in the detection rule (2.6). □

Corollary 2.3.1. The solution for the binary case ($M = 2$) of the hypothesis testing problem in Section 2.2.3 is the threshold test

$$\hat{m} = \begin{cases} 2 & \text{if} \quad L^b\left(T\right) \geqslant \frac{p_1}{p_2} \\ 1 & \text{if} \quad L^b\left(T\right) < \frac{p_1}{p_2} \end{cases} \tag{2.7}$$

where the likelihood ratio function $L^b\left(T\right)$ is given by

$$L^b\left(T\right) = \frac{L_2\left(T\right)}{L_1\left(T\right)}.$$

Proof. The proof follows directly from (2.6). □

Our goal for the remainder of this section is to obtain proper expressions for $L_i(T)$, $i = 1, 2, \ldots, M$. Theorem 2.3.1 below represents $L_i(T)$ in terms of an infinite sum of multiple integrals.

Theorem 2.3.1. Fix α and define $\lambda_i(t) = \alpha s_i(t) + \mu$. Then, the likelihood ratio function $L_i(T)$ can be expressed as

$$L_i(T) = \Lambda\left(\mathscr{Y}_T; \{\lambda_i(t)\}, T\right) \tag{2.8}$$

where, for any deterministic function λ_t, the functional $\Lambda(\cdot)$ is given by

$$\Lambda\left(\mathscr{Y}_T; \{\lambda_t\}, T\right) = \exp\left(-\int_0^T \lambda_t dt\right) \cdot \left(1 + \sum_{n=1}^{\infty} \frac{1}{n!} \int_0^T \cdots \int_0^T \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} \tilde{\Lambda}_T^{(n)}(\tau_1, \ldots, \tau_n, \theta_1, \ldots, \theta_n) \prod_{k=1}^{n} dP_q(\theta_k) \prod_{k=1}^{n} \lambda_{\tau_k} d\tau_k\right). \tag{2.9}$$

Here, $\tilde{\Lambda}_T^{(n)}(\tau_1, \ldots, \tau_n, \theta_1, \ldots, \theta_n)$ is defined as

$$\tilde{\Lambda}_T^{(n)}(\tau_1, \ldots, \tau_n, \theta_1, \ldots, \theta_n) = \exp\left\{\frac{1}{\sigma^2} \int_0^T \sum_{k=1}^{n} \theta_k \pi(t - \tau_k)\, dy_t - \frac{1}{2\sigma^2} \int_0^T \left(\sum_{k=1}^{n} \theta_k \pi(t - \tau_k)\right)^2 dt\right\} \tag{2.10}$$

and $P_q(\cdot)$ is the common cumulative distribution function of $\{q_n\}_{n=1}^{\infty}$.

Proof. Consider the random vector $V_T = (t_1, \ldots, t_{N_T}, q_1, \ldots, q_{N_T})$. Conditioned on V_T, the hypothesis testing problem (2.5) is one of detecting a deterministic

signal in white Gaussian noise. For the realization $(\tau_1, \ldots, \tau_n, \theta_1, \ldots, \theta_n)$ of V_T, the likelihood ratio function associated with this hypothesis testing problem is given by (2.10). The likelihood ratio function for the original problem can be obtained [36] by averaging (2.10) over all realizations of V_T, where the rate associated with V_T is $\lambda_i(t) = \alpha s_i(t) + \mu$. Thus, we can write

$$L_i(T) = \mathrm{E}\left[\tilde{\Lambda}_T^{(N_T)}(t_1, \ldots, t_{N_T}, q_1, \ldots, q_{N_T}) \,\middle|\, \mathscr{Y}_T\right]. \tag{2.11}$$

Using (2.3) and (2.4), it is easy to verify that this expectation can be written as (2.8) with

$$\Lambda(\mathscr{Y}_T; \{\lambda_t\}, T) = \exp\left(-\int_0^T \lambda_t dt\right) \cdot \left(1 + \sum_{n=1}^{\infty} \int_0^T \int_0^{\tau_n} \cdots \int_0^{\tau_2} \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} \tilde{\Lambda}_T^{(n)}(\tau_1, \ldots, \tau_n, \theta_1, \ldots, \theta_n) \prod_{k=1}^{n} dP_q(\theta_k) \prod_{k=1}^{n} \lambda_{\tau_k} d\tau_k\right). \tag{2.12}$$

Since the integrand of each multiple integral is invariant under any change in the order of τ_k's, (2.12) can be rewritten as (2.9). □

Corollary 2.3.2. For the case of a random α, the likelihood ratio function $L_i(T)$ can be expressed as

$$L_i(T) = \Lambda_\alpha(\mathscr{Y}_T; \{s_i(t)\}, T) \tag{2.13}$$

where, for any deterministic function s_t, the functional $\Lambda_\alpha(\cdot)$ is given by

$$\Lambda_\alpha(\mathscr{Y}_T; \{s_t\}, T) = \int_0^{\infty} \Lambda(\mathscr{Y}_T; \{as_t + \mu\}, T) \, dP_\alpha(a). \tag{2.14}$$

Here, $P_\alpha(\cdot)$ denotes the cumulative distribution function of α.

Proof. Applying the law of total expectation to (2.11) and using the fact that α and $\mathscr{Y}_T$ are statistically independent, we get

$$\begin{aligned} L_i(T) &= \mathrm{E}\left[\mathrm{E}\left[\tilde{\Lambda}_T^{(N_T)}(t_1,\ldots,t_{N_T},q_1,\ldots,q_{N_T})\,\middle|\,\mathscr{Y}_T,\alpha\right]\middle|\,\mathscr{Y}_T\right] \\ &= \mathrm{E}\big[\Lambda(\mathscr{Y}_T;\{\alpha s_i(t)+\mu\},T)\,\big|\,\mathscr{Y}_T\big] \\ &= \int_0^\infty \Lambda(\mathscr{Y}_T;\{a s_i(t)+\mu\},T)\,dP_\alpha(a). \end{aligned}$$

This leads to (2.13) with $\Lambda_\alpha(\cdot)$ defined by (2.14). □

Remark 2.3.1. We observe from (2.9) and (2.10) that the dependence of the functional $\Lambda(\mathscr{Y}_T;\{\lambda_t\},T)$ on $\{y_t\}$ is through the stochastic process $\{z_t\}$ defined as

$$z_t = \frac{1}{\sigma^2}\int_0^T \pi(\tau - t)\,dy_\tau,\ t\in[0,T]. \tag{2.15}$$

This implies that for implementing $\Lambda(\mathscr{Y}_T;\{\lambda_t\},T)$, the first stage is a linear filter characterized by (2.15). We remark that (2.15) is a matched filter for $\pi(\cdot)$.

The following theorem introduces a new technique to rewrite (2.9) in a simpler form, which is suitable for the purpose of approximation.

Theorem 2.3.2. Let $\{\xi_t\}$ be a standard Wiener process independent of $\{t_n\}_{n=1}^{N_t}$, $\{q_n\}_{n=1}^{\infty}$, and $\{w_t\}$, which is defined on the probability space $(\Omega,\mathscr{F},P)$. Fix α and

define the stochastic process

$$\bar{\xi}_t = \frac{1}{\sigma} \int_0^T \pi\left(\tau - t\right) d\xi_\tau,\ t \in [0, T]. \tag{2.16}$$

Then, with probability 1, the functional $\Lambda\left(\cdot\right)$ can be expressed as

$$\Lambda\left(\mathscr{Y}_T; \left\{\lambda_t\right\}, T\right) = \mathrm{E}\left[\exp\left(\int_0^T \lambda_t \Phi_q\left(z_t + j\bar{\xi}_t\right) dt\right) \middle| \mathscr{Y}_T\right] \tag{2.17}$$

where z_t is given by (2.15), $j = \sqrt{-1}$, and $\Phi_q\left(\cdot\right)$ is defined as

$$\Phi_q\left(z\right) = \int_{-\infty}^{+\infty} \exp\left(\theta z\right) dP_q\left(\theta\right) - 1 \tag{2.18}$$

when it exists.

Proof. We can verify that

$$\exp\left\{-\frac{1}{2\sigma^2} \int_0^T \left(\sum_{k=1}^n \theta_k \pi\left(t - \tau_k\right)\right)^2 dt\right\} = \mathrm{E}\left[\exp\left\{\frac{j}{\sigma} \int_0^T \sum_{k=1}^n \theta_k \pi\left(t - \tau_k\right) d\xi_t\right\}\right] \tag{2.19}$$

by noting that the integral on the right side is a Gaussian random variable. Substituting (2.15), (2.16), and (2.19) into (2.10), and noting that $\mathscr{Y}_T$ and $\{\xi_t\}$ are

statistically independent, we get

$$\tilde{\Lambda}_T^{(n)}(\tau_1,\ldots,\tau_n,\theta_1,\ldots,\theta_n) = \prod_{k=1}^{n}\exp(\theta_k z_{\tau_k})\,\mathrm{E}\left[\prod_{k=1}^{n}\exp\left(j\theta_k\bar{\xi}_{\tau_k}\right)\Big|\mathscr{Y}_T\right]$$
$$= \mathrm{E}\left[\prod_{k=1}^{n}\exp\left\{\theta_k\left(z_{\tau_k}+j\bar{\xi}_{\tau_k}\right)\right\}\Big|\mathscr{Y}_T\right]. \qquad (2.20)$$

Since with probability 1, the sample paths of $\{z_t\}$ are bounded, the rest of this proof will be presented for bounded sample paths of $\{z_t\}$. Thus, any result which is obtained based on the boundedness of $\{z_t\}$ will be stated with probability 1.

Let $\Re\{\cdot\}$ and $\Im\{\cdot\}$ denote the "real part" and the "imaginary part", respectively. We can write

$$\left|\Re\left\{\prod_{k=1}^{n}\exp\left\{\theta_k\left(z_{\tau_k}+j\bar{\xi}_{\tau_k}\right)\right\}\right\}\right| \leqslant \left|\prod_{k=1}^{n}\exp\left\{\theta_k\left(z_{\tau_k}+j\bar{\xi}_{\tau_k}\right)\right\}\right| = \prod_{k=1}^{n}\exp(\theta_k z_{\tau_k}).$$

Assuming that $\Phi_q(\cdot)$ exist, this leads to

$$\int_0^T\cdots\int_0^T\int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty}\mathrm{E}\left[\left|\Re\left\{\prod_{k=1}^{n}\exp\left\{\theta_k\left(z_{\tau_k}+j\bar{\xi}_{\tau_k}\right)\right\}\right\}\right|\Big|\mathscr{Y}_T\right]$$
$$\cdot\prod_{k=1}^{n}dP_q(\theta_k)\prod_{k=1}^{n}\lambda_{\tau_k}d\tau_k \leqslant \prod_{k=1}^{n}\int_0^T\left(\int_{-\infty}^{+\infty}\exp(\theta_k z_{\tau_k})\,P_q(\theta_k)\right)\lambda_{\tau_k}d\tau_k$$
$$= \left(\int_0^T\lambda_t\big(\Phi_q(z_t)+1\big)dt\right)^n.$$

For every bounded sample paths of $\{z_t\}$, the right side of the inequality above is

bounded, thus we conclude from Fubini's theorem [37] that

$$\int_0^T \cdots \int_0^T \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} \mathrm{E}\left[\Re\left\{ \prod_{k=1}^{n} \exp\left\{ \theta_k \left(z_{\tau_k} + j\bar{\xi}_{\tau_k} \right) \right\} \right\} \middle| \mathscr{Y}_T \right] \cdot \prod_{k=1}^{n} dP_q\left(\theta_k\right) \prod_{k=1}^{n} \lambda_{\tau_k} d\tau_k$$
$$= \mathrm{E}\Bigg[\int_0^T \cdots \int_0^T \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} \Re\left\{ \prod_{k=1}^{n} \exp\left\{ \theta_k \left(z_{\tau_k} + j\bar{\xi}_{\tau_k} \right) \right\} \right\} \cdot \prod_{k=1}^{n} dP_q\left(\theta_k\right) \prod_{k=1}^{n} \lambda_{\tau_k} d\tau_k \Big| \mathscr{Y}_T \Bigg].$$

A similar argument can be applied to the imaginary part of

$$\prod_{k=1}^{n} \exp\left\{ \theta_k \left(z_{\tau_k} + j\bar{\xi}_{\tau_k} \right) \right\}$$

in order to show that

$$\int_0^T \cdots \int_0^T \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} \mathrm{E}\left[\prod_{k=1}^{n} \exp\left\{ \theta_k \left(z_{\tau_k} + j\bar{\xi}_{\tau_k} \right) \right\} \middle| \mathscr{Y}_T \right] \prod_{k=1}^{n} dP_q\left(\theta_k\right) \prod_{k=1}^{n} \lambda_{\tau_k} d\tau_k$$
$$= \mathrm{E}\Bigg[\int_0^T \cdots \int_0^T \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} \prod_{k=1}^{n} \exp\left\{ \theta_k \left(z_{\tau_k} + j\bar{\xi}_{\tau_k} \right) \right\} \cdot \prod_{k=1}^{n} dP_q\left(\theta_k\right) \prod_{k=1}^{n} \lambda_{\tau_k} d\tau_k \Big| \mathscr{Y}_T \Bigg]$$
$$= \mathrm{E}\left[\left(\int_0^T \lambda_t \big(\Phi_q\left(z_t + j\bar{\xi}_t \right) + 1 \big) dt \right)^n \middle| \mathscr{Y}_T \right].$$

Substituting (2.20) into (2.9) and using the result above, we obtain

$$\Lambda\left(\mathscr{Y}_T; \{\lambda_t\}, T\right) = \exp\left(- \int_0^T \lambda_t dt \right) \sum_{n=0}^{\infty} \frac{1}{n!} \mathrm{E}\left[Z^n | \mathscr{Y}_T \right]$$

where the random variable Z is defined as

$$Z = \int_0^T \lambda_t \big(\Phi_q \left(z_t + j\bar{\xi}_t \right) + 1 \big) dt.$$

We note that $|Z|$ satisfies the inequality

$$\begin{aligned} |Z| &\leqslant \int_0^T \lambda_t \left| \Phi_q \left(z_t + j\bar{\xi}_t \right) + 1 \right| dt \\ &= \int_0^T \lambda_t \left| \int_{-\infty}^{+\infty} \exp \left\{ \theta \left(z_t + j\bar{\xi}_t \right) \right\} \right| dP_q \left(\theta \right) dt \\ &\leqslant \int_0^T \lambda_t \int_{-\infty}^{+\infty} \exp \left(\theta z_t \right) dP_q \left(\theta \right) dt \\ &= \int_0^T \lambda_t \big(\Phi_q \left(z_t \right) + 1 \big) dt \triangleq |Z|_u. \end{aligned}$$

From this inequality and the fact that $|Z|_u < \infty$ for every bounded sample path of $\{z_t\}$, we conclude that $|Z| < \infty$. Let N be an integer and define the sequence of complex-valued random variables X_N as

$$X_N = \sum_{n=0}^{N} \frac{1}{n!} Z^n.$$

From $|Z| < \infty$, we conclude that $X_N \to \exp \left(Z \right)$, and as consequence, $\Re \left\{ X_N \right\} \to \Re \left\{ \exp \left(Z \right) \right\}$ and $\Im \left\{ X_N \right\} \to \Im \left\{ \exp \left(Z \right) \right\}$. On the other hand we have

$$|X_N| \leqslant \sum_{n=0}^{N} \frac{1}{n!} |Z|^n \leqslant \sum_{n=0}^{\infty} \frac{1}{n!} |Z|^n \leqslant \sum_{n=0}^{\infty} \frac{1}{n!} |Z|_u^n = \exp \left(|Z|_u \right).$$

This leads to $|\Re \left\{ X_N \right\}| \leqslant |X_N| \leqslant \exp \left(|Z|_u \right)$ and $|\Im \left\{ X_N \right\}| \leqslant |X_N| \leqslant \exp \left(|Z|_u \right)$.

Also, for every bounded sample path of $\{z_t\}$, we have

$$\mathrm{E}\left[\exp\left(|Z|_u\right)|\mathscr{Y}_T\right] = \exp\left(|Z|_u\right) < \infty.$$

Thus, we can apply the dominated convergence theorem [37] separately to $\Re\{X_N\}$ and $\Im\{X_N\}$ in order to show that

$$\sum_{n=0}^{\infty} \frac{1}{n!} \mathrm{E}\left[Z^n|\mathscr{Y}_T\right] = \lim_{N\to\infty} \mathrm{E}\left[X_N|\mathscr{Y}_T\right] = \mathrm{E}\left[\exp\left(Z\right)|\mathscr{Y}_T\right].$$

This completes the proof. □

Corollary 2.3.3. For the case of a random α, under the assumptions of Theorem 2.3.2 and assuming that $\{\xi_t\}$ is independent of α, with probability 1, we have

$$\Lambda_\alpha\left(\mathscr{Y}_T; \{s_t\}, T\right) = \mathrm{E}\left[\Phi_\alpha\left(\int_0^T s_t \Phi_q\left(z_t + j\bar{\xi}_t\right) dt\right) \exp\left(\int_0^T \mu \Phi_q\left(z_t + j\bar{\xi}_t\right) dt\right) \Big| \mathscr{Y}_T\right] \tag{2.21}$$

where $\Phi_\alpha(\cdot)$ is defined as

$$\Phi_\alpha(z) = \int_0^\infty \exp(az)\, dP_\alpha(a) \tag{2.22}$$

when it exists.

Proof. Let $\mathscr{X}_T$ denote the σ-algebra generated by $\{\xi_t\}$ during $[0, T]$. Using the law

of total expectation, we write

$$\begin{aligned}\Lambda_{\alpha}\left(\mathscr{Y}_T;\{s_t\},T\right) &= \mathrm{E}\left[\mathrm{E}\left[\exp\left(\int_0^T\left(\alpha s_t+\mu\right)\Phi_q\left(z_t+j\bar{\xi}_t\right)dt\right)\Big|\mathscr{Y}_T,\alpha\right]\Big|\mathscr{Y}_T\right]\\ &= \mathrm{E}\left[\mathrm{E}\left[\exp\left(\int_0^T\left(\alpha s_t+\mu\right)\Phi_q\left(z_t+j\bar{\xi}_t\right)dt\right)\Big|\mathscr{Y}_T,\mathscr{X}_T\right]\Big|\mathscr{Y}_T\right]\\ &= \mathrm{E}\left[\int_0^{\infty}\exp\left(\int_0^T\left(as_t+\mu\right)\Phi_q\left(z_t+j\bar{\xi}_t\right)dt\right)dP_{\alpha}\left(a\right)\Big|\mathscr{Y}_T\right]\end{aligned}$$

where the last equality is concluded from the fact that α is independent of $\mathscr{Y}_T$ and $\mathscr{X}_T$. Since the sample paths of $\{z_t\}$ and $\{\xi_t\}$ are almost surely bounded, assuming that $\Phi_{\alpha}(z)$ exists for every bounded z, this leads to (2.21) with probability 1. □

The following theorem provides an alternative expression for (2.9), which later will be used to establish a lower bound on (2.9).

Theorem 2.3.3. Fix α and let $q_k=\bar{q},\ k=1,2,3,\ldots$ be a constant. Then, with probability 1, (2.9) can be expressed as

$$\begin{aligned}&\Lambda\left(\mathscr{Y}_T;\{\lambda_t\},T\right)\\ &\quad=\exp\left(\int_0^T\lambda_t\Phi_{\bar{q}}\left(z_t\right)dt\right)\mathrm{E}\left[\exp\left\{-\frac{1}{2\sigma^2}\int_0^T\left(\sum_{k=1}^{N_T^*}\bar{q}\pi\left(t-t_n^*\right)\right)^2dt\right\}\Big|\mathscr{Y}_T\right]\end{aligned}\tag{2.23}$$

where $\Phi_{\bar{q}}(\cdot)$ is defined as

$$\Phi_{\bar{q}}\left(z\right)=\mathrm{e}^{\bar{q}z}-1\tag{2.24}$$

and $\{t_n^*\}_{n=1}^{N_t^*}$, is a doubly stochastic Poisson point process with the rate

$$\lambda_t^* = \lambda_t e^{\bar{q} z_t}. \tag{2.25}$$

Proof. Noting that $q_k = \bar{q}$ is a constant, we substitute (2.10), (2.24), and (2.25) into (2.12) and rewrite it as

$$\Lambda\left(\mathscr{Y}_T; \{\lambda_t\}, T\right) = \exp\left(\int_0^T \lambda_t \Phi_{\bar{q}}\left(z_t\right) dt\right) \exp\left(-\int_0^T \lambda_t^* dt\right)$$
$$\cdot \left(1 + \sum_{n=1}^{\infty} \int_0^T \int_0^{\tau_n} \cdots \int_0^{\tau_2} \exp\left\{-\frac{1}{2\sigma^2} \int_0^T \left(\sum_{k=1}^{n} \bar{q}\pi\left(t - \tau_k\right)\right)^2 dt\right\} \prod_{k=1}^{n} \lambda_{\tau_k}^* d\tau_k\right).$$

For bounded sample paths of $\{\lambda_t^*\}$, it is easy to verify that this expression is equivalent to (2.23). Since with probability 1, the sample paths of $\{z_t\}$, and as a consequence the sample paths of $\{\lambda_t^*\}$, are bounded, (2.23) holds with probability 1. □

2.4 Upper and Lower Bounds on $\Lambda(\cdot)$

In this section, we determine simple upper and lower bounds for $\Lambda\left(\mathscr{Y}_T; \{\lambda_t\}, T\right)$. Since the expressions obtained for $\Lambda\left(\mathscr{Y}_T; \{\lambda_t\}, T\right)$ in Section 2.3 are complicated, these bounds are useful in studying the behavior of $\Lambda\left(\mathscr{Y}_T; \{\lambda_t\}, T\right)$ for some important limiting cases. In addition, under conditions close to these limiting cases, the bounds can be used to approximate $\Lambda\left(\mathscr{Y}_T; \{\lambda_t\}, T\right)$. Throughout this section, we shall assume that λ_t is a deterministic function.

Theorem 2.4.1. Assume that $q_k \geqslant 0,\ k = 1, 2, 3, \ldots$ and $\pi(t) \geqslant 0$ for every $t \geqslant 0$.

Define the functional

$$B_1^u\left(\mathscr{Y}_T;\{\lambda_t\},T\right)=\exp\left(\int_0^T\lambda_t\Phi_q^*\left(z_t,b_t\right)dt\right) \tag{2.26}$$

where

$$b_t=\frac{1}{\sigma^2}\int_0^T\pi^2\left(\tau-t\right)d\tau \tag{2.27}$$

and assuming that $\Phi_q^*\left(z,b\right)$ exists for every bounded z and b, it is defined as

$$\Phi_q^*\left(z,b\right)=\int_{-\infty}^{+\infty}\exp\left(\theta z-\frac{1}{2}\theta^2 b\right)dP_q\left(\theta\right)-1. \tag{2.28}$$

Then, with probability 1, we have

$$\Lambda\left(\mathscr{Y}_T;\{\lambda_t\},T\right)\leqslant B_1^u\left(\mathscr{Y}_T;\{\lambda_t\},T\right). \tag{2.29}$$

Proof. Under the assumptions of the theorem, the second term on the right side of

$$\left(\sum_{k=1}^n\theta_k\pi\left(t-\tau_k\right)\right)^2=\sum_{k=1}^n\theta_k^2\pi^2\left(t-\tau_k\right)+\sum_{k=1}^n\sum_{\substack{l=1\\l\neq k}}^n\theta_k\theta_l\pi\left(t-\tau_k\right)\pi\left(t-\tau_l\right)$$

is nonnegative. This indicates that the left side is not greater than the first term on the right side. Applying this result to (2.10), we have

$$\tilde{\Lambda}_T^{(n)}\left(\tau_1,\ldots,\tau_n,\theta_1,\ldots,\theta_n\right)\leqslant\prod_{k=1}^n\exp\left(\theta_k z_{\tau_k}-\frac{1}{2}\theta_k^2 b_{\tau_k}\right).$$

Substituting this inequality into (2.9), we get

$$\Lambda\left(\mathscr{Y}_T;\{\lambda_t\},T\right)\leqslant\exp\left(-\int_0^T\lambda_t dt\right)\sum_{n=1}^{\infty}\frac{1}{n!}\left(\int_0^T\lambda_t\big(\Phi_q^*\left(z_t,b_t\right)+1\big)dt\right)^n.$$

For every bounded sample path of $\{z_t\}$, the right side of this inequality converges to (2.26). Since with probability 1, the sample paths of $\{z_t\}$ are bounded, this indicates that (2.29) holds with probability 1. □

Remark 2.4.1. The more conservative upper bound

$$\Lambda\left(\mathscr{Y}_T;\{\lambda_t\},T\right)\leqslant B_1^u\left(\mathscr{Y}_T;\{\lambda_t\},T\right)\leqslant\exp\left(\int_0^T\lambda_t\Phi_q\left(z_t\right)dt\right)\tag{2.30}$$

does not require the assumptions of Theorem 2.4.1. The proof follows from

$$\exp\left\{-\frac{1}{2\sigma^2}\int_0^T\left(\sum_{k=1}^{n}\theta_k\pi\left(t-\tau_k\right)\right)^2dt\right\}\leqslant 1.$$

Theorem 2.3.3 and Lemma 2.4.1 stated below can be used to establish a lower bound on $\Lambda\left(\cdot\right)$. This lower bound will be presented in Theorem 2.4.2.

Lemma 2.4.1. Let $\{t_n\}_{n=1}^{N_t}$ be a Poisson point process with rate λ_t. Then, we have

$$\begin{aligned}&\mathrm{E}\left[\int_0^T\left(\sum_{k=1}^{N_T}\pi\left(t-t_n\right)\right)^2dt\right]\\&\qquad=\int_0^T\int_0^T\lambda_\tau\pi^2\left(t-\tau\right)d\tau dt+\int_0^T\left(\int_0^T\lambda_\tau\pi\left(t-\tau\right)d\tau\right)^2dt.\end{aligned}\tag{2.31}$$

Proof. The left side of (2.31) can be written as

$$\begin{aligned}
&\exp\left(-\int_0^T \lambda_t dt\right) \sum_{n=1}^{\infty} \frac{1}{n!} \int_0^T \int_0^T \cdots \int_0^T \left[\int_0^T \left(\sum_{k=1}^{n} \pi\left(t-\tau_k\right)\right)^2 dt\right] \prod_{k=1}^{n} \lambda_{\tau_k} d\tau_k \\
&\quad = \exp\left(-\int_0^T \lambda_t dt\right) \sum_{n=1}^{\infty} \frac{1}{n!} \Bigg[n \left(\int_0^T \lambda_t dt\right)^{n-1} \int_0^T \int_0^T \lambda_\tau \pi^2\left(t-\tau\right) d\tau dt \\
&\qquad + n\left(n-1\right) \left(\int_0^T \lambda_t dt\right)^{n-2} \int_0^T \int_0^T \int_0^T \lambda_{\tau_1} \lambda_{\tau_2} \pi\left(t-\tau_1\right) \pi\left(t-\tau_2\right) d\tau_1 d\tau_2 dt \Bigg] \\
&\quad = \int_0^T \int_0^T \lambda_\tau \pi^2\left(t-\tau\right) d\tau dt \\
&\qquad + \int_0^T \left(\int_0^T \lambda_{\tau_1} \pi\left(t-\tau_1\right) d\tau_1\right) \left(\int_0^T \lambda_{\tau_2} \pi\left(t-\tau_2\right)\right) dt
\end{aligned}$$

which is equal to its right side. □

Theorem 2.4.2. Under the assumptions of Theorem 2.3.3, with probability 1, the functional

$$\begin{aligned}
B_1^{\ell}\left(\mathscr{Y}_T; \{\lambda_t\}, T\right) = \exp\Bigg\{ &\int_0^T \lambda_t \Phi_{\bar{q}}\left(z_t\right) dt - \frac{\bar{q}}{2\sigma^2} \int_0^T \int_0^T \lambda_\tau \Phi'_{\bar{q}}\left(z_\tau\right) \pi^2\left(t-\tau\right) d\tau dt \\
&- \frac{1}{2\sigma^2} \int_0^T \left(\int_0^T \lambda_\tau \Phi'_{\bar{q}}\left(z_\tau\right) \pi\left(t-\tau\right) d\tau\right)^2 dt \Bigg\} \qquad (2.32)
\end{aligned}$$

is a lower bound for $\Lambda\left(\mathscr{Y}_T; \{\lambda_t\}, T\right)$. Here $\Phi'_{\bar{q}}\left(\cdot\right)$ is the derivative of $\Phi_{\bar{q}}\left(\cdot\right)$ which is defined by (2.24).

Proof. Noting that exponential is a convex function, we apply Jensen's inequality

to (2.23) to get

$$\Lambda\left(\mathscr{Y}_T;\{\lambda_t\},T\right) \geqslant \exp\left(\int_0^T \lambda_t \Phi_{\bar{q}}\left(z_t\right) dt\right) \cdot \exp\left\{-\frac{1}{2\sigma^2}\,\mathrm{E}\left[\int_0^T \left(\sum_{k=1}^{N_T^*} \bar{q}\pi\left(t-t_n^*\right)\right)^2 dt \,\middle|\, \mathscr{Y}_T\right]\right\}.$$

Next, we apply Lemma 2.4.1 to the right side of this inequality to obtain (2.32). □

Corollary 2.4.1. Under the assumptions of Theorem 2.3.3, with probability 1, we have

$$\lim_{\sigma\to\infty} \frac{\Lambda\left(\mathscr{Y}_T;\{\lambda_t\},T\right)}{B_1^{\ell}\left(\mathscr{Y}_T;\{\lambda_t\},T\right)} = \lim_{\sigma\to\infty} \frac{\Lambda\left(\mathscr{Y}_T;\{\lambda_t\},T\right)}{B_1^{u}\left(\mathscr{Y}_T;\{\lambda_t\},T\right)} = 1\,. \tag{2.33}$$

Proof. From (2.26), (2.30), and (2.32), with probability 1, we have

$$\exp\left\{-\frac{\bar{q}}{2\sigma^2}\int_0^T\int_0^T \lambda_\tau \Phi'_{\bar{q}}\left(z_\tau\right)\pi^2\left(t-\tau\right)d\tau dt - \frac{1}{2\sigma^2}\int_0^T\left(\int_0^T \lambda_\tau \Phi'_{\bar{q}}\left(z_\tau\right)\pi\left(t-\tau\right)d\tau\right)^2 dt\right\} \leqslant \frac{B_1^{\ell}\left(\mathscr{Y}_T;\{\lambda_t\},T\right)}{B_1^{u}\left(\mathscr{Y}_T;\{\lambda_t\},T\right)} \leqslant \frac{\Lambda\left(\mathscr{Y}_T;\{\lambda_t\},T\right)}{B_1^{u}\left(\mathscr{Y}_T;\{\lambda_t\},T\right)} \leqslant 1\,.$$

As $\sigma \to \infty$, for every bounded sample path of $\{z_t\}$, the expression on the left side of this inequality tends to 1. This proves that (2.33) holds with probability 1. □

Remark 2.4.2. Using Theorems 2.4.1 and 2.4.2 and (2.14), we can find upper and

lower bounds on $\Lambda_\alpha\left(\mathscr{Y}_T;\{s_t\},T\right)$ as

$$\Lambda_\alpha\left(\mathscr{Y}_T;\{s_t\},T\right)\geqslant\int_0^\infty B_1^\ell\left(\mathscr{Y}_T;\{as_t+\mu\},T\right)dP_\alpha\left(a\right)$$
$$\Lambda_\alpha\left(\mathscr{Y}_T;\{s_t\},T\right)\leqslant\int_0^\infty B_1^u\left(\mathscr{Y}_T;\{as_t+\mu\},T\right)dP_\alpha\left(a\right).$$

In the remainder of this section, we adopt a different approach to establish upper and lower bounds on $\Lambda\left(\mathscr{Y}_T;\{\lambda_t\},T\right)$. Theorems 2.4.3 and 2.4.4 below explain this approach.

Theorem 2.4.3. Assume that $q_k=\bar{q},\ k=1,2,3,\ldots$ is a constant and $\pi\left(t\right)\geqslant 0$ for every $t\geqslant 0$. Let X_t^u be the solution of the integral equation

$$X_t^u=1+\int_0^t\gamma^u\left(t,\tau\right)\lambda_\tau c_\tau f\left(z_\tau\right)X_\tau^u d\tau \tag{2.34}$$

where $f\left(z\right)=\mathrm{e}^{\bar{q}z}$, $c_t=\exp\left(-\bar{q}^2b_t/2\right)$, and

$$\gamma^u\left(\tau_1,\tau_2\right)=\exp\left(-\frac{\bar{q}^2}{\sigma^2}\int_0^T\pi\left(t-\tau_1\right)\pi\left(t-\tau_2\right)dt\right). \tag{2.35}$$

Then, with probability 1, we have

$$\Lambda\left(\mathscr{Y}_T;\{\lambda_t\},T\right)\leqslant B_2^u\left(\mathscr{Y}_T;\{\lambda_t\},T\right)\leqslant B_1^u\left(\mathscr{Y}_T;\{\lambda_t\},T\right) \tag{2.36}$$

where $B_2^u\left(\cdot\right)$ is defined as

$$B_2^u\left(\mathscr{Y}_T;\{\lambda_t\},T\right)=X_T^u\exp\left(-\int_0^T\lambda_t dt\right). \tag{2.37}$$

Proof. We first define

$$B_2^u\left(\mathscr{Y}_T;\{\lambda_t\},T\right)=\exp\left(-\int_0^T\lambda_t dt\right)\cdot\left(1+\int_0^T\lambda_{\tau_1}c_{\tau_1}f\left(z_{\tau_1}\right)d\tau_1\right.$$
$$\left.+\sum_{n=2}^{\infty}\int_0^T\int_0^{\tau_n}\cdots\int_0^{\tau_2}\lambda_{\tau_n}c_{\tau_n}f\left(z_{\tau_n}\right)d\tau_n\prod_{k=1}^{n-1}\gamma^u\left(\tau_{k+1},\tau_k\right)\lambda_{\tau_k}c_{\tau_k}f\left(z_{\tau_k}\right)d\tau_k\right) \quad (2.38)$$

and show that this expression can be written as

$$B_2^u\left(\mathscr{Y}_T;\{\lambda_t\},T\right)=Y_T^u\exp\left(-\int_0^T\lambda_t dt\right) \quad (2.39)$$

where

$$Y_T^u=1+\int_0^T\lambda_t c_t f\left(z_t\right)X_t^u dt.$$

To achieve this goal, for every bounded sample path of $\{z_t\}$, every $t>0$, and every integer $N\geqslant 2$, we define

$$Z_t^N=1+\sum_{n=2}^{N}\left[\int_0^{\tau_n}\cdots\int_0^{\tau_2}\prod_{k=1}^{n-1}\gamma^u\left(\tau_{k+1},\tau_k\right)\lambda_{\tau_k}c_{\tau_k}f\left(z_{\tau_k}\right)d\tau_k\right]_{\tau_n=t}.$$

We note that for every fixed $t>0$, Z_t^N is increasing in N and satisfies

$$Z_t^N\leqslant\exp\left(\int_0^t\lambda_\tau c_\tau f\left(z_\tau\right)d\tau\right).$$

This shows that for every bounded sample path of $\{z_t\}$, $Z_t^\infty\triangleq\lim_{N\to\infty}Z_t^N$ exists.

For every $N \geqslant 3$, we can write

$$Z_t^N = 1 + \int_0^t \gamma^u(t,\tau)\,\lambda_\tau c_\tau f(z_\tau)\, Z_\tau^{N-1} d\tau$$

which leads to

$$\begin{aligned} Z_t^\infty &= 1 + \lim_{N\to\infty} \int_0^t \gamma^u(t,\tau)\,\lambda_\tau c_\tau f(z_\tau)\, Z_\tau^{N-1} d\tau \\ &= 1 + \int_0^t \gamma^u(t,\tau)\,\lambda_\tau c_\tau f(z_\tau)\, Z_\tau^{\infty} d\tau. \end{aligned}$$

Here, the second equality is concluded from the monotone convergence theorem [38]. This result indicates that $X_t^u = Z_t^\infty$ for $t > 0$. Using the monotone convergence theorem, we express (2.38) as

$$\begin{aligned} B_2^u(\mathscr{Y}_T; \{\lambda_t\}, T) &= \exp\left(-\int_0^T \lambda_t dt\right) \lim_{N\to\infty} \left(1 + \int_0^T \lambda_t c_t f(z_t)\, Z_t^N dt\right) \\ &= \exp\left(-\int_0^T \lambda_t dt\right) \left(1 + \int_0^T \lambda_t c_t f(z_t)\, Z_t^\infty dt\right) \end{aligned}$$

which proves that (2.39) holds. Expressing Y_T^u as

$$Y_T^u = 1 + \int_0^T \gamma^u(T,t)\,\lambda_t c_t f(z_t)\, X_t^u dt + \int_0^T \left(1 - \gamma^u(T,t)\right) \lambda_t c_t f(z_t)\, X_t^u dt$$

and noting that $\gamma^u(T,t) = 1$ for every $0 \leqslant t \leqslant T$, we find that $Y_T^u = X_T^u$. This verifies that with probability 1, (2.38) is equal to (2.37).

To prove the first inequality (from left) of (2.36), we use (2.38) and (2.12) to

construct

$$B_2^u\left(\mathscr{Y}_T;\{\lambda_t\},T\right)-\Lambda\left(\mathscr{Y}_T;\{\lambda_t\},T\right)$$
$$=\exp\left(-\int_0^T\lambda_t dt\right)\sum_{n=2}^{\infty}\int_0^T\int_0^{\tau_n}\cdots\int_0^{\tau_2}\chi_n^u\left(\tau_1,\tau_2,\ldots,\tau_n\right)\prod_{k=1}^{n}\lambda_{\tau_k}f\left(z_{\tau_k}\right)d\tau_k \tag{2.40}$$

where $\chi_n^u(\cdot)$ is given by

$$\chi_n^u\left(\tau_1,\tau_2,\ldots,\tau_n\right)=\prod_{k=1}^{n}c_{\tau_k}\prod_{k=1}^{n-1}\gamma^u\left(\tau_{k+1},\tau_k\right)-\exp\left\{-\frac{\bar{q}^2}{2\sigma^2}\int_0^T\left(\sum_{k=1}^{n}\pi\left(t-\tau_k\right)\right)^2dt\right\}.$$

It is easy to show that $\chi_n^u\left(\tau_1,\tau_2,\ldots,\tau_n\right)\geqslant 0$, by rearranging the expression above as

$$\chi_n^u\left(\tau_1,\tau_2,\ldots,\tau_n\right)$$
$$=\prod_{k=1}^{n}c_{\tau_k}\prod_{k=1}^{n-1}\gamma^u\left(\tau_{k+1},\tau_k\right)\left(1-\exp\left\{-\frac{\bar{q}^2}{2\sigma^2}\int_0^T\sum_{\mathscr{S}(i,j)}\pi\left(t-\tau_i\right)\pi\left(t-\tau_j\right)dt\right\}\right)$$

where $\mathscr{S}(i,j)=\{(i,j):i\neq j,|i-j|\neq 1\}$. This verifies that (2.40) is nonnegative and completes the proof. The second inequality of (2.36) follows from $\gamma^u(\cdot,\cdot)\leqslant 1$. □

Theorem 2.4.4. Assume that $q_k=\bar{q},\ k=1,2,3,\ldots$ is a constant, $\pi(t)\geqslant 0$ for every $t\geqslant 0$, and $\pi(t)=0$ for $t>\epsilon$, where $\epsilon>0$ is a known constant. Let X_t^ℓ and Y_t^ℓ be the solutions of

$$X_t^\ell=1+\int_0^t\gamma^\ell\left(t,\tau\right)\lambda_\tau c_\tau f\left(z_\tau\right)X_\tau^\ell d\tau \tag{2.41}$$

$$Y_t^\ell=1+\int_0^t\lambda_\tau c_\tau f\left(z_\tau\right)X_\tau^\ell d\tau \tag{2.42}$$

where $\gamma^{\ell}(\cdot,\cdot)$ is defined such that

$$\gamma^{\ell}(\tau_1,\tau_2)=\gamma^{\ell}(\tau_2,\tau_1)=\begin{cases}1 & 0\leqslant\tau_1\leqslant\max(0,\tau_2-\epsilon)\\ 0 & \max(0,\tau_2-\epsilon)<\tau_1\leqslant\tau_2.\end{cases}$$

Then, with probability 1, we have

$$B_2^{\ell}(\mathscr{Y}_T;\{\lambda_t\},T)\leqslant B_3^{\ell}(\mathscr{Y}_T;\{\lambda_t\},T)\leqslant\Lambda(\mathscr{Y}_T;\{\lambda_t\},T) \tag{2.43}$$

where

$$B_2^{\ell}(\mathscr{Y}_T;\{\lambda_t\},T)=X_T^{\ell}\exp\left(-\int_0^T\lambda_t dt\right) \tag{2.44}$$

$$B_3^{\ell}(\mathscr{Y}_T;\{\lambda_t\},T)=Y_T^{\ell}\exp\left(-\int_0^T\lambda_t dt\right). \tag{2.45}$$

Proof. The proof is similar to the proof of Theorem 2.4.3, by replacing superscript u with ℓ. For this case, $\chi_n^{\ell}(\cdot)$ is given by

$$\chi_n^{\ell}(\tau_1,\tau_2,\ldots,\tau_n)=\prod_{k=1}^{n}c_{\tau_k}\left(\prod_{j=2}^{n}\gamma^{\ell}(\tau_j,\tau_{j-1})-\prod_{j=2}^{n}\prod_{i=1}^{j-1}\gamma^{u}(\tau_j,\tau_i)\right).$$

From the definitions of $\gamma^{u}(\cdot,\cdot)$ and $\gamma^{\ell}(\cdot,\cdot)$ and the assumptions of the theorem, it is easy to show that for $0\leqslant\tau_1\leqslant\tau_2\leqslant\ldots\leqslant\tau_j$, we have

$$\gamma^{\ell}(\tau_j,\tau_{j-1})\leqslant\prod_{i=1}^{j-1}\gamma^{u}(\tau_j,\tau_i)$$

which leads to $\chi_n^{\ell}(\tau_1,\tau_2,\ldots,\tau_n)\leqslant 0,\ n\geqslant 2$. This proves the inequality on the right

side of (2.43). The inequality on the left side follows from $X_T^\ell \leqslant {Y_T}^\ell$. □

The following theorem provides a closed form expression for a lower bound on (2.45).

Theorem 2.4.5. Let the assumptions of Theorem 2.4.4 hold and define

$$B_4^\ell\left(\mathscr{Y}_T;\{\lambda_t\},T\right)=\exp\left\{-\int_0^T\lambda_t dt + \int_0^T\lambda_t c_t f\left(z_t\right)\left(1+\int_0^t\left(1-\gamma^\ell\left(t,\tau\right)\right)\lambda_\tau c_\tau f\left(z_\tau\right)d\tau\right)^{-1}dt\right\}. \quad (2.46)$$

Then, with probability 1, we have

$$B_4^\ell\left(\mathscr{Y}_T;\{\lambda_t\},T\right)\leqslant B_3^\ell\left(\mathscr{Y}_T;\{\lambda_t\},T\right). \quad (2.47)$$

Proof. Using (2.41), we can write

$$X_t^\ell=1+\int_0^t\lambda_\tau c_\tau f\left(z_\tau\right)X_\tau^\ell d\tau-\int_0^t\left(1-\gamma^\ell\left(t,\tau\right)\right)\lambda_\tau c_\tau f\left(z_\tau\right)X_\tau^\ell d\tau.$$

From (2.42) and the fact that X_t^ℓ is increasing in t, we find that

$$X_t^\ell\geqslant Y_t^\ell-X_t^\ell\int_0^t\left(1-\gamma^\ell\left(t,\tau\right)\right)\lambda_\tau c_\tau f\left(z_\tau\right)d\tau.$$

We know from (2.42) that $X_t^\ell=\dot{Y}_t^\ell/\lambda_t c_t f\left(z_t\right)$, thus substituting this result into the

inequality above and rearranging it, we get

$$\frac{\dot{Y}_t^\ell}{Y_t^\ell} \geqslant \frac{\lambda_t c_t f\left(z_t\right)}{1+\int_0^t \left(1-\gamma^\ell\left(t,\tau\right)\right)\lambda_\tau c_\tau f\left(z_\tau\right)d\tau}.$$

Upon integrating both sides of this inequality over $[0,T]$, we obtain (2.47). □

2.5 Behavior of $\Lambda(\cdot)$ for a Small Pulse Duration

In this section, we study the behavior of $\Lambda\left(\mathscr{Y}_T;\{\lambda_t\},T\right)$ when the pulse duration ϵ tends to zero. Before addressing the main topic, we discuss a technical difficulty arising from $\epsilon \to 0$, namely

$$\lim_{\epsilon\to 0} \mathrm{E}\left[\Lambda\left(\mathscr{Y}_T;\{\lambda_t\},T\right)\right] = \infty. \tag{2.48}$$

As mentioned in Section 2.2.1, $\pi(\cdot)$ has a unit area, thus it can be expressed as $\pi(t) = \epsilon^{-1}\tilde{\pi}(t/\epsilon)$, where $\tilde{\pi}(\cdot)$ has a unit area and $|\tilde{\pi}(\cdot)|$ is bounded for every $t \geqslant 0$. In addition, the condition $\pi(t) = 0$, $t > \epsilon$ implies that $\tilde{\pi}(t) = 0$, $t > 1$. We define $\nu(\tau_1,\tau_2) = 1/\gamma^u(\tau_1,\tau_2)$, where $\gamma^u(\cdot,\cdot)$ is given by (2.35). This function can be expressed as

$$\nu(\tau_1,\tau_2) = \exp\left(\frac{\bar{q}^2}{\sigma^2\epsilon}\Pi^\epsilon(\tau_1,\tau_2)\right) \tag{2.49}$$

where $\Pi^\epsilon(\cdot,\cdot)$ is defined as

$$\Pi^\epsilon(\tau_1,\tau_2) = \int_0^{T/\epsilon} \tilde{\pi}\left(t-\tau_1/\epsilon\right)\tilde{\pi}\left(t-\tau_2/\epsilon\right)dt.$$

Let $\tilde{\pi}(\cdot)$ be nonnegative over $[0,1]$. Then, for some subset of $|\tau_1 - \tau_2| \leqslant \epsilon$ with nonzero Lebesgue measure, we have $\Pi^{\epsilon}(\tau_1, \tau_2) > 0$. Referring to (2.49), this implies that $\lim_{\epsilon \to 0} \epsilon \nu(\tau_1, \tau_2) = \infty$ over a subset of $|\tau_1 - \tau_2| \leqslant \epsilon$. Thus, for every function $\lambda_t > 0$, $t \in [0,T]$ and every $t^* \in [0,T]$, we have

$$\lim_{\epsilon \to 0} \int_0^T \lambda_t \nu(t, t^*)\, dt = \infty. \tag{2.50}$$

We note that the stochastic process $\{c_t f(z_t)\}_{t=0}^T$ can be expressed as[1]

$$c_t f(z_t) = \rho_t \prod_{n=1}^{N_T} \nu(t, t_n) \tag{2.51}$$

where $\{\rho_t\}$ is defined as

$$\rho_t = \exp\left(\frac{\bar{q}}{\sigma} \int_0^T \pi(\tau - t)\, dw_\tau - \frac{\bar{q}^2}{2\sigma^2} \int_0^T \pi^2(\tau - t)\, d\tau \right).$$

It is easy to verify that for every fixed $t \geqslant 0$, the lognormal random variable ρ_t has a unit mean. When $\epsilon \to 0$, since the pulses in $\{z_t\}$ do not overlap each other, (2.51) can be written as

$$c_t f(z_t) = \rho_t + \rho_t \sum_{n=1}^{N_T} (\nu(t, t_n) - 1). \tag{2.52}$$

We remind from Theorem 2.4.4 that $X_t^\ell \geqslant 1$, $t \in [0,T]$. Thus, using (2.42),

[1]We take $\prod_{n=1}^{0}(\cdot)$ equal 1.

(2.52), and $\mathrm{E}[\rho_t] = 1$, we can write

$$\begin{aligned}\mathrm{E}\left[Y_T^{\ell}\right] &\geqslant \mathrm{E}\left[1 + \int_0^T \lambda_t c_t f(z_t)\, dt\right] \\ &= 1 + \int_0^T \lambda_t dt + \int_0^T \lambda_t \mathrm{E}\left[\sum_{n=1}^{N_T} (\nu(t, t_n) - 1)\right] dt.\end{aligned}$$

Applying (2.50) to this inequality, we get $\lim_{\epsilon \to 0} \mathrm{E}\left[Y_T^{\ell}\right] = \infty$, which together with (2.43) and (2.45) prove (2.48).

In spite of the difficulty mentioned above, we present useful results below for the case of $\epsilon \to 0$. These results provide appropriate means for approximating $\Lambda(\cdot)$ when ϵ is small.

Theorem 2.5.1. Assume that the unit area function $\pi(\cdot)$ has the property that $\pi(t) \geqslant 0$ for $t \in [0, \epsilon]$ and $\pi(t) = 0$ beyond this interval. Then, with probability 1, we have

$$\lim_{\epsilon \to 0} \frac{\Lambda(\mathscr{Y}_T; \{\lambda_t\}, T)}{B_2^{\ell}(\mathscr{Y}_T; \{\lambda_t\}, T)} = \lim_{\epsilon \to 0} \frac{\Lambda(\mathscr{Y}_T; \{\lambda_t\}, T)}{B_2^{u}(\mathscr{Y}_T; \{\lambda_t\}, T)} = 1\,.$$

Proof. From (2.34), (2.41), and the fact that $\gamma^{\ell}(\tau_1, \tau_2) = \gamma^{\ell}(\tau_1, \tau_2)\, \gamma^{u}(\tau_1, \tau_2)$, we can write

$$\begin{aligned}\beta_t \triangleq \frac{X_t^u - X_t^{\ell}}{X_t^u} = \frac{1}{X_t^u}\Bigg[&\int_0^t \gamma^u(t, \tau)\, \lambda_\tau c_\tau f(z_\tau) \left(X_\tau^u - X_\tau^{\ell}\right) d\tau \\ &+ \int_0^t \left(1 - \gamma^{\ell}(t, \tau)\right) \gamma^u(t, \tau)\, \lambda_\tau c_\tau f(z_\tau)\, X_\tau^{\ell} d\tau \Bigg].\end{aligned}$$

The increasing property of X_t^u and the facts that $X_t^u \geqslant X_t^\ell$ and $\gamma^u(\cdot,\cdot) \leqslant 1$, result in

$$\beta_t \leqslant \int_0^t \lambda_\tau c_\tau f(z_\tau) \beta_\tau d\tau + \delta_t \tag{2.53}$$

where δ_t is defined as

$$\delta_t = \int_0^t \left(1 - \gamma^\ell(t,\tau)\right) \gamma^u(t,\tau) \lambda_\tau c_\tau f(z_\tau) d\tau.$$

Using (2.51) and $\mathrm{E}[\rho_t] = 1$, we find the mean of δ_t as

$$\mathrm{E}[\delta_t] = \mathrm{E}\left[\int_0^t \left(1 - \gamma^\ell(t,\tau)\right) \gamma^u(t,\tau) \lambda_\tau \prod_{n=1}^{N_T} \nu(\tau, t_n) d\tau\right]. \tag{2.54}$$

For a fixed sample path of $\{t_n\}_{n=1}^{N_T}$, let t' and t'' be, respectively, the closest and the next closest t_n to t. Assume that $|t-t'| > 0$. Then, for $\epsilon < |t-t'|/2$, over the interval $\tau \in [\max(0, t-\epsilon), t]$ in which $1-\gamma^\ell(t,\tau) \neq 0$, we have $\prod_{n=1}^{N_T} \nu(\tau, t_n) = 1$. Now, we assume that $t' = t$. Then, noting that $\Pr\{|t'-t''| > 0\} = 1$ for $\epsilon < |t'-t''|/2$, we have $\gamma^u(t,\tau) \prod_{n=1}^{N_T} \nu(\tau, t_n) = \gamma^u(t,\tau)\nu(\tau,t') = 1$ for $\tau \in [\max(0, t-\epsilon), t]$. These results indicate that the integrand in (2.54) is almost surely bounded, and as a consequence, the integral tends to 0 as $\epsilon \to 0$. This proves that $\lim_{\epsilon\to 0} \mathrm{E}[\delta_t] = 0$ for every $t \geqslant 0$, which in combination with $\delta_t \geqslant 0$ lead to $\lim_{\epsilon\to 0} \delta_t = 0$, with probability 1.

By letting $\delta_t = 0$ on the right side of (2.53) and noting that $\beta_0 = 0$, we find that $\beta_t \leqslant 0$ for $t \geqslant 0$. On the other hand, we know from definition that $\beta_t \geqslant 0$, thus we conclude that $\lim_{\epsilon\to 0} \beta_t = 0$, almost surely for every $t \geqslant 0$. Finally, we complete

the proof by applying $\lim_{\epsilon\to 0} X_T^{\ell}/X_T^{u} = 1$ to the inequality

$$B_2^{\ell}\left(\mathscr{Y}_T;\{\lambda_t\},T\right) \leqslant \Lambda\left(\mathscr{Y}_T;\{\lambda_t\},T\right) \leqslant B_2^{u}\left(\mathscr{Y}_T;\{\lambda_t\},T\right).$$

□

Theorem 2.5.2. Assume that for $0 \leqslant \tau \leqslant t \leqslant T$, the mapping $\gamma(\cdot,\cdot)$ satisfies the inequality

$$\gamma^{\ell}(t,\tau) \leqslant \gamma(t,\tau) \leqslant \gamma^{u}(t,\tau). \tag{2.55}$$

Let X_t be the solution of the integral equation

$$X_t = 1 + \int_0^t \gamma(t,\tau)\,\lambda_\tau c_\tau f(z_\tau)\,X_\tau d\tau \tag{2.56}$$

and define

$$B\left(\mathscr{Y}_T;\{\lambda_t\},T\right) = X_T \exp\left(-\int_0^T \lambda_t dt\right). \tag{2.57}$$

Then, under the assumptions of Theorem 2.5.1, with probability 1, we have

$$\lim_{\epsilon\to 0} \frac{\Lambda\left(\mathscr{Y}_T;\{\lambda_t\},T\right)}{B\left(\mathscr{Y}_T;\{\lambda_t\},T\right)} = 1.$$

Proof. The proof follows from the proof of Theorem 2.5.1 and $X_T^{\ell} \leqslant X_T \leqslant X_T^{u}$. □

Remark 2.5.1. The proof of Theorem 2.5.1 suggests that $B\left(\mathscr{Y}_T;\{\lambda_t\},T\right)$ in (2.57) is an appropriate approximation for $\Lambda\left(\mathscr{Y}_T;\{\lambda_t\},T\right)$, if ϵ is small enough to ensure that most sample paths of $\{y_t\}$ are free from overlapping pulses.

Theorem 2.5.3. Let $\lambda_1(t)$ and $\lambda_2(t)$ be nonnegative functions defined over $[0,T]$. Then, under the assumptions of Theorem 2.5.1, with probability 1, we have

$$\lim_{\epsilon\to 0}\frac{\Lambda\left(\mathscr{Y}_T;\{\lambda_1(t)\},T\right)}{\Lambda\left(\mathscr{Y}_T;\{\lambda_2(t)\},T\right)}=\exp\left(\int_0^T\left(\lambda_2(t)-\lambda_1(t)\right)dt\right)\prod_{n=1}^{N_T}\frac{\lambda_1(t_n)}{\lambda_2(t_n)}. \tag{2.58}$$

Proof. Consider a sample path of the stochastic process $\{c_t f(z_t)\}$ defined by (2.51) and assume that ϵ is much smaller than the minimum distance between two successive occurrence times t_n and t_{n+1}. This ensures that the pulses in (2.51) do not overlap each other. Under this condition, the goal is to solve the integral equation (2.41), whose simplified form is given by

$$X_t^\ell=1+\int_0^{t-\epsilon}\lambda_\tau c_\tau f(z_\tau)X_\tau^\ell d\tau. \tag{2.59}$$

The solution of this equation encounters a "big jump" during $[t_n,t_n+2\epsilon]$, $n=1,2,\ldots,N_T$. Therefore, we have to solve (2.59) separately for two cases: inside the intervals $[t_n,t_n+2\epsilon]$ and outside these intervals.

For $t\in(t_n,t_n+\epsilon]$, we rewrite the integral equation (2.59) as

$$X_t^\ell=X_{t_n}^\ell+\int_{t_n-\epsilon}^{t-\epsilon}\lambda_\tau\rho_\tau\nu(\tau,t_n)X_\tau^\ell d\tau$$

and solve it as follows. Since X_τ, ρ_τ, and λ_τ are (almost surely) continuous over $\tau\in(t_n-\epsilon,t_n]$, the solution of this equation can be approximated by

$$X_t^\ell=X_{t_n}^\ell+\lambda_{t_n}\rho_{t_n}X_{t_n}^\ell\int_{t_n-\epsilon}^{t-\epsilon}\nu(\tau,t_n)\,d\tau,\ t\in(t_n,t_n+\epsilon]. \tag{2.60}$$

Note that as $\epsilon \to 0$, this approximation tends to the exact solution. Also, the solution of (2.59) for $X^{\ell}_{t_n+2\epsilon}$ can be approximated as

$$\begin{aligned} X^{\ell}_{t_n+2\epsilon} &= X^{\ell}_{t_n+\epsilon} + \int_{t_n}^{t_n+\epsilon} \lambda_t \rho_t \nu\left(t, t_n\right) X^{\ell}_t dt \\ &= X^{\ell}_{t_n+\epsilon} + \lambda_{t_n} \rho_{t_n} \int_{t_n}^{t_n+\epsilon} \nu\left(t, t_n\right) X^{\ell}_t dt. \end{aligned}$$

Upon substituting (2.60) into this expression, we get

$$\begin{aligned} X^{\ell}_{t_n+2\epsilon} = X^{\ell}_{t_n} \Bigg(1 + \lambda_{t_n} \rho_{t_n} \int_{t_n-\epsilon}^{t_n+\epsilon} \nu\left(t, t_n\right) dt \\ + \lambda^2_{t_n} \rho^2_{t_n} \int_{t_n}^{t_n+\epsilon} \int_{t_n-\epsilon}^{t-\epsilon} \nu\left(t, t_n\right) \nu\left(\tau, t_n\right) d\tau dt \Bigg). \end{aligned}$$

For sake of simplicity, we express this result as

$$X^{\ell}_{t_n+2\epsilon} = X^{\ell}_{t_n} \left(1 + \lambda_{t_n} \rho_{t_n} J\left(\epsilon\right) + \lambda^2_{t_n} \rho^2_{t_n} \tilde{J}\left(\epsilon\right) \right) \tag{2.61}$$

where $J\left(\epsilon\right)$ and $\tilde{J}\left(\epsilon\right)$ are defined as

$$\begin{aligned} J\left(\epsilon\right) &= \int_0^{2\epsilon} \nu\left(t, \epsilon\right) dt \\ \tilde{J}\left(\epsilon\right) &= \int_{\epsilon}^{2\epsilon} \int_0^{t-\epsilon} \nu\left(t, \epsilon\right) \nu\left(\tau, \epsilon\right) d\tau dt. \end{aligned}$$

In order to obtain X^{ℓ}_t for $t \in (t_n + 2\epsilon, t_{n+1}]$, we need to solve the integral equation

$$X^{\ell}_t = X^{\ell}_{t_n+2\epsilon} + \int_{t_n+\epsilon}^{t-\epsilon} \lambda_\tau \rho_\tau X^{\ell}_\tau d\tau.$$

As $\epsilon \to 0$, the solution of this equation at $t = t_{n+1}$ is given by

$$X^{\ell}_{t_{n+1}} = X^{\ell}_{t_n+2\epsilon} \exp\left(\int_{t_n}^{t_{n+1}} \lambda_\tau \rho_\tau d\tau\right). \tag{2.62}$$

Starting from

$$X^{\ell}_{t_1} = \exp\left(\int_0^{t_1} \lambda_\tau \rho_\tau d\tau\right)$$

and using (2.61) and (2.62) in a recursive procedure, it is straightforward to show that

$$X^{\ell}_{T} = \exp\left(\int_0^{T} \lambda_\tau \rho_\tau d\tau\right) \prod_{n=1}^{N_T} \left(1 + \lambda_{t_n} \rho_{t_n} J(\epsilon) + \lambda^2_{t_n} \rho^2_{t_n} \tilde{J}(\epsilon)\right).$$

This result leads to

$$\frac{\Lambda\left(\mathscr{Y}_T; \{\lambda_1(t)\}, T\right)}{\Lambda\left(\mathscr{Y}_T; \{\lambda_2(t)\}, T\right)} = U(\epsilon) \exp\left(\int_0^{T} \big(\lambda_2(t) - \lambda_1(t)\big) dt\right) \prod_{n=1}^{N_T} \frac{\lambda_1(t_n)}{\lambda_2(t_n)} \tag{2.63}$$

where $U(\epsilon)$ is defined as

$$U(\epsilon) = \exp\left(\int_0^{T} \big(\lambda_1(t) - \lambda_2(t)\big) \rho_t dt\right) \cdot \prod_{n=1}^{N_T} \frac{1 + \big(\lambda_1(t_n)\rho_{t_n} J(\epsilon)\big)^{-1} + \lambda_1(t_n)\rho_{t_n} \tilde{J}(\epsilon)/J(\epsilon)}{1 + \big(\lambda_2(t_n)\rho_{t_n} J(\epsilon)\big)^{-1} + \lambda_2(t_n)\rho_{t_n} \tilde{J}(\epsilon)/J(\epsilon)}.$$

We can show that $\lim_{\epsilon\to 0} U(\epsilon) = 1$, by the fact that for every $t \in [0, T]$, with probability 1, we have $\lim_{\epsilon\to 0} \rho_t = 0$, $\lim_{\epsilon\to 0} \big(\rho_t J(\epsilon)\big)^{-1} = 0$, and $\lim_{\epsilon\to 0} \rho_t \tilde{J}(\epsilon)/J(\epsilon) = 0$. This proves (2.58), upon being applied to (2.63). □

Remark 2.5.2. As $\epsilon \to 0$, the stochastic differential equation (2.2) tends to

$$dy_t = \bar{q} dN_t + \sigma dw_t. \tag{2.64}$$

On the other hand, (2.58) is the likelihood ratio function associated with a channel whose output $\tilde{y}_t$ is described by

$$d\tilde{y}_t = \bar{q} dN_t. \tag{2.65}$$

We conclude from these facts that subject to the detection problem of Section 2.2.3, the channels described by (2.64) and (2.65) are equivalent, in the sense that they have equal probability of error. In addition, since this argument is valid for every T, every integer M, and every set of waveforms $\{\lambda_1(t), \lambda_2(t), \ldots, \lambda_M(t)\}$, we argue that (2.64) and (2.65) have identical channel capacity.

Remark 2.5.3. We consider the case that $\sigma = \sigma(\epsilon)$ is a decreasing function of ϵ such that $\lim_{\epsilon \to 0} \sigma(\epsilon) = \infty$. We can verify that the results of Theorems 2.5.1 and 2.5.3 still hold, if for every $0 < \epsilon < 1$ and some $\delta > 0$, $\sigma(\epsilon)$ satisfies the inequality

$$\sigma(\epsilon) < \frac{\bar{q}}{1+\delta} \sqrt{-\frac{\Pi^{\epsilon}(0,0)}{\epsilon \ln \epsilon}}.$$

Under this condition, Theorem 2.5.3 implies that as $\epsilon \to 0$, the probability of error of channel (2.2) tends to the probability of error of the ideal channel (2.65). On the other hand, any linear filtering scheme fails to reconstruct the transmitted message,

since $\lim_{\epsilon\to 0}\sigma(\epsilon)=\infty$ implies that the signal-to-noise-ratio at the output of that filter will be 0.

2.6 Approximate Implementation

The formal solutions introduced in Section 2.3 can be implemented only by means of infinite-dimensional systems; however, under certain assumptions, finite-dimensional approximations for $\Lambda(\mathscr{Y}_T;\{\lambda_t\},T)$ can be derived from (2.17) and the results of Sections 2.4 and 2.5. The goal of this section is to determine such approximate implementations and discuss the conditions under which they are useful. We shall keep the assumption of Section 2.5 in which $\pi(\cdot)$ is of finite duration, i.e., $\pi(t)=0$ for $t\notin[0,\epsilon]$. The interpretation of this assumption is that most of the pulse energy is concentrated in $[0,\epsilon]$ such that the energy beyond this interval is negligible.

According to (2.15), the stochastic process $\{z_t\}$ can be only implemented using an anticausal system; however, due to the assumption above, the stochastic process $\tilde{z}_t\triangleq z_{t-\epsilon}$, $t\in[\epsilon,T+\epsilon]$ can be implemented by a causal, linear, time-varying system. This system is characterized by

$$\tilde{z}_t=\frac{1}{\sigma^2}\int_{-\infty}^{+\infty}h(t,\tau)\,dy_\tau=\frac{1}{\sigma^2}\int_0^t h(t,\tau)\,dy_\tau \tag{2.66}$$

where the impulse response $h(t,\tau)$ is given by

$$h(t,\tau)=\pi(\epsilon-(t-\tau))\,u(T-\tau)$$

with $u(\cdot)$ denoting the unit step function. Note that (2.9) and (2.17) remain unchanged if we shift up the limits of integration by ϵ (i.e. replacing 0 with ϵ and T with $T+\epsilon$) and in the same time replace z_t with $\tilde{z}_t$, e.g., (2.17) can be written as

$$\Lambda\left(\mathscr{Y}_T;\{\lambda_t\},T\right)=\mathrm{E}\left[\exp\left(\int_{\epsilon}^{T+\epsilon}\lambda_{t-\epsilon}\Phi_q\left(\tilde{z}_t+j\bar{\xi}_{t-\epsilon}\right)dt\right)\Big|\mathscr{Y}_T\right].$$

This shows that by accepting a delay of ϵ in the decision time, we can implement (2.9) and (2.17) using the causal system (2.66). For sake of simplicity, in the rest of this section, we keep the time frame $[0,T]$, while we know how to replace it with $[\epsilon,T+\epsilon]$ in order to implement z_t by means of the causal filter (2.66).

Throughout this section, we introduce two categories of approximation for $\Lambda(\cdot)$. The first category will be derived from expression (2.17), while the second one is based on Theorem 2.5.2. While the first category can be used for the general case of α and $\{q_k\}$, the second category is only applicable to the case that α and $q_k=\bar{q}$ are deterministic values.

2.6.1 Approximation: Category I

We derive our first approximation for $\Lambda(\cdot)$ from (2.17) by approximating

$$\Phi_q\left(z_t+j\bar{\xi}_t\right)\simeq\Phi_q\left(z_t\right) \tag{2.67}$$

which results in

$$\Lambda\left(\mathscr{Y}_T;\{\lambda_t\},T\right)\simeq\exp\left(\int_0^T\lambda_t\Phi_q\left(z_t\right)dt\right). \tag{2.68}$$

For the case of a random α, using (2.14), (2.22), and (2.68) we can write

$$\Lambda_\alpha\left(\mathscr{Y}_T;\{s_t\},T\right)\simeq\Phi_\alpha\left(\int_0^T s_t\Phi_q\left(z_t\right)dt\right)\exp\left(\int_0^T\mu\Phi_q\left(z_t\right)dt\right). \tag{2.69}$$

The block diagrams in Figure 2.1 illustrate the implementation of (2.68) and (2.69).

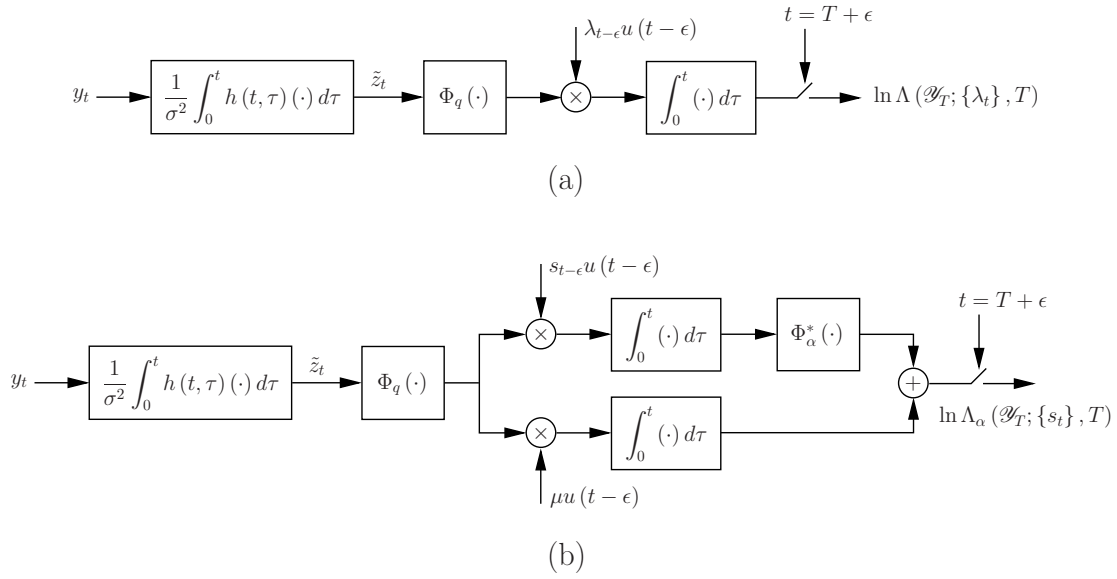

Figure 2.1: Implementation of (a) approximation (2.68) and (b) approximation (2.69). In (b), the nonlinear mapping $\Phi_\alpha^*(\cdot)$ is defined as $\Phi_\alpha^*(\cdot)=\ln\Phi_\alpha(\cdot)$.

In order to obtain a condition under which (2.68) and (2.69) are useful, we

rewrite (2.17) as

$$\Lambda\left(\mathscr{Y}_T;\{\lambda_t\},T\right)=\exp\left(-\int_0^T\lambda_t dt\right)\mathrm{E}\left[\exp\left(\int_0^T\int_{-\infty}^{+\infty}\lambda_t \mathrm{e}^{\theta z_t}\mathrm{e}^{j\theta\bar{\xi}_t}dP_q\left(\theta\right)dt\right)\Big|\mathscr{Y}_T\right].$$

It can be observed from this expression that the approximation (2.67) is equivalent to approximating $\mathrm{e}^{j\theta\bar{\xi}_t}\simeq 1$. Let b_t be given by (2.27) and q_{max} be defined such that $\Pr\left\{|q_k|>q_{max}\right\}\simeq 0$. Then, the approximation $\mathrm{e}^{j\theta\bar{\xi}_t}\simeq 1$ is justified, if for every $t\in[0,T]$ and every $|\theta|<q_{max}$, we have

$$\mathrm{E}\left[\left(1-\cos\theta\bar{\xi}_t\right)^2\right]\simeq\frac{3}{4}\left(\theta^2 b_t\right)^2\ll 1 \tag{2.70}$$

$$\mathrm{E}\left[\left(0-\sin\theta\bar{\xi}_t\right)^2\right]\simeq\theta^2 b_t\ll 1. \tag{2.71}$$

It is straightforward to verify that both conditions (2.70) and (2.71) are satisfied if

$$\frac{q_{max}^2}{\sigma^2}\int_0^\infty\pi^2\left(t\right)dt\ll 1\,. \tag{2.72}$$

Remark 2.6.1. As mentioned earlier, the first stage for implementing $\Lambda\left(\cdot\right)$ is a linear filter which has an impulse response with duration ϵ. Since the bandwidth of this filter is roughly $1/\epsilon$, the effective power of the white noise (thermal noise) is $2\sigma^2/\epsilon$. For simplicity of discussion, assume that $q_k=\bar{q}$ is a deterministic value. Then, (2.72) can be expressed as

$$\frac{1}{\epsilon}\int_0^\epsilon\bar{q}^2\pi^2\left(t\right)dt\ll\frac{1}{2}\left(\frac{2\sigma^2}{\epsilon}\right). \tag{2.73}$$

The interpretation of (2.73) is that the average power of a single pulse must be much smaller than the effective power of the white noise. Note that (2.73) does not necessarily require a small signal-to-noise-ratio (SNR), since with a large rate λ_t, we can maintain a large SNR, while satisfying (2.73).

We can improve (2.68) and (2.69) by approximating

$$\Phi_q\left(z_t + j\bar{\xi}_t\right) \simeq \Phi_q\left(z_t\right) + j\Phi_q'\left(z_t\right)\bar{\xi}_t \tag{2.74}$$

which is equivalent to

$$\mathrm{e}^{j\theta\bar{\xi}_t} \simeq 1 + j\theta\bar{\xi}_t.$$

For this approximation to be valid, in addition to (2.70), the condition

$$\mathrm{E}\left[\left(\theta\bar{\xi}_t - \sin\theta\bar{\xi}_t\right)^2\right] \simeq \frac{5}{12}\left(\theta^2 b_t\right)^3 \ll 1 \tag{2.75}$$

must be satisfied for $t \in [0, T]$ and $|\theta| < q_{max}$. A unified condition that satisfies both (2.70) and (2.75) is obtained as

$$\frac{3}{4}\left(\frac{q_{max}^2}{\sigma^2}\int_0^\infty \pi^2\left(t\right)dt\right)^2 \ll 1$$

which is less restrictive than (2.72).

In order to determine an approximation for $\Lambda\left(\cdot\right)$ based on (2.74), we substi-

tute (2.74) into (2.17) to find

$$\Lambda\left(\mathscr{Y}_T;\{\lambda_t\},T\right)\simeq\exp\left(\int_0^T\lambda_t\Phi_q\left(z_t\right)dt\right)\mathrm{E}\left[\exp\left(j\int_0^T\lambda_t\Phi_q'\left(z_t\right)\bar{\xi}_t dt\right)\Big|\mathscr{Y}_T\right]. \tag{2.76}$$

The second integral in (2.76) can be written as

$$\int_0^T\lambda_t\Phi_q'\left(z_t\right)\bar{\xi}_t dt=\int_0^T x_t d\xi_t \tag{2.77}$$

where x_t is defined as

$$x_t=\frac{1}{\sigma}\int_0^T\pi\left(t-\tau\right)\lambda_\tau\Phi_q'\left(z_\tau\right)d\tau. \tag{2.78}$$

Note that x_t can be implemented using a causal, linear, time-varying system with the impulse response

$$g\left(t,\tau\right)=\pi\left(t-\tau\right)u\left(\tau\right)u\left(T-\tau\right).$$

Let $\check{z}_t$ and $\check{x}_t$ be the sample paths of $\{z_t\}$ and $\{x_t\}$, respectively, noting that $\check{x}_t$ is associated with $\check{z}_t$ through (2.78). Then, for the sample path $\check{z}_t$, the left side of (2.77) is a zero-mean Gaussian random variable with a variance $\int_0^T\check{x}_t^2$. Thus, noting that $\{z_t\}$ and $\{x_t\}$ are independent of $\{\xi_t\}$, we can write

$$\mathrm{E}\left[\exp\left(j\int_0^T\lambda_t\Phi_q'\left(\check{z}_t\right)\bar{\xi}_t dt\right)\right]=\exp\left(-\frac{1}{2}\int_0^T\check{x}_t^2 dt\right). \tag{2.79}$$

Since $\{z_t\}$ and $\{x_t\}$ are smooth stochastic processes with (almost surely) continuous sample paths, in (2.79), we can replace the sample paths $\check{z}_t$ and $\check{x}_t$ with the stochastic processes $\{z_t\}$ and $\{x_t\}$, respectively. Using this fact and substituting (2.79) into (2.76), we obtain

$$\Lambda\left(\mathscr{Y}_T; \{\lambda_t\}, T\right) \simeq \exp\left\{\int_0^T \left(\lambda_t \Phi_q\left(z_t\right) - \frac{1}{2} x_t^2\right) dt\right\}. \tag{2.80}$$

The implementation of this approximation is illustrated in Figure 2.2.

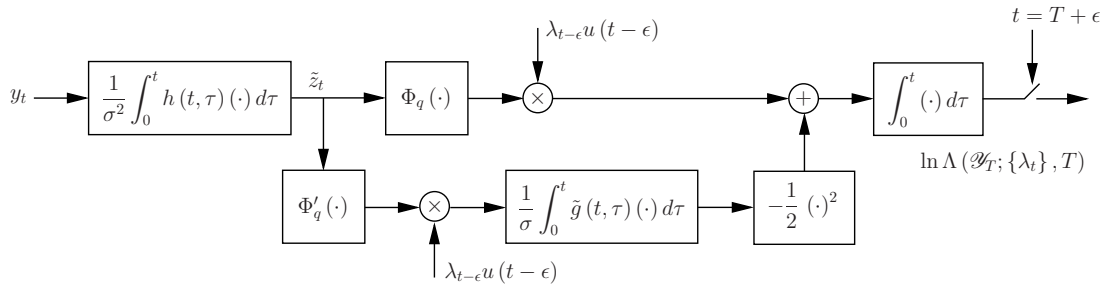

Figure 2.2: Implementation of (2.80). Here, the impulse response $\tilde{g}\left(t,\tau\right)$ is defined as $\tilde{g}\left(t,\tau\right) = g\left(t-\epsilon, \tau-\epsilon\right)$.

In order to extend (2.80) to the case of a random α, we define the stochastic processes

$$\tilde{x}_t = \frac{1}{\sigma}\int_0^T \pi\left(t-\tau\right) s_\tau \Phi_q'\left(z_\tau\right) d\tau$$
$$\bar{x}_t = \frac{1}{\sigma}\int_0^T \pi\left(t-\tau\right) \mu \Phi_q'\left(z_\tau\right) d\tau$$

where s_t is a deterministic function. Next, substituting $x_t = \alpha\tilde{x}_t + \bar{x}_t$ into (2.80)

and using (2.14), we get

$$\Lambda_\alpha\left(\mathscr{Y}_T;\{s_t\},T\right)\simeq\exp\left\{\int_0^T\left(\mu\Phi_q\left(z_t\right)-\frac{1}{2}\,\bar{x}_t^2\right)dt\right\}$$
$$\cdot F_\alpha\left(\int_0^T\left(s_t\Phi_q\left(z_t\right)-\tilde{x}_t\bar{x}_t\right)dt\,,\int_0^T\tilde{x}_t^2dt\right)\qquad(2.81)$$

where $F_\alpha\left(\cdot,\cdot\right)$ is defined as

$$F_\alpha\left(v_1,v_2\right)=\int_0^\infty\exp\left(av_1-\frac{1}{2}a^2v_2\right)dP_\alpha\left(a\right).$$

The implementation of this approximation is illustrated in Figure 2.3.

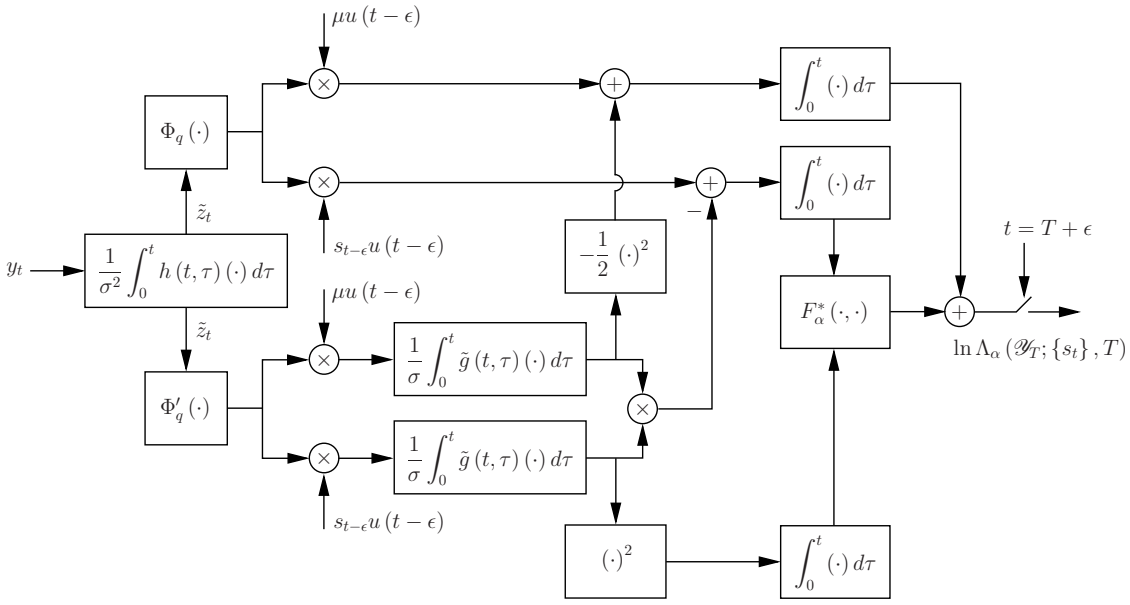

Figure 2.3: Implementation of (2.81). In this block diagram, we have $\tilde{g}\left(t,\tau\right)=g\left(t-\epsilon,\tau-\epsilon\right)$ and $F_\alpha^*\left(v_1,v_2\right)=\ln F_\alpha\left(v_1,v_2\right)$.

2.6.2 Approximation: Category II

The second category of approximation we purpose for $\Lambda(\cdot)$ is based on Theorem 2.5.2. Using the results of this theorem, we argue that for a small ϵ, we can approximate

$$\Lambda\left(\mathscr{Y}_T; \{\lambda_t\}, T\right) \simeq B\left(\mathscr{Y}_T; \{\lambda_t\}, T\right) \tag{2.82}$$

where $B(\cdot)$ is determined by solving the integral equation (2.56) and using (2.57). In order to establish a condition under which (2.82) is a close approximation, we focus on the proof of Theorem 2.5.1. This proof indicates that for a nonzero ϵ, the claim of the theorem is approximately valid, if the sample paths of $\{z_t\}$ are free from overlapping pulses. This is equivalent to having a small probability for occurrence of more than one pulse in an interval with duration ϵ. We know from the properties of Poisson process that this condition is satisfied if we have $\epsilon^2 \lambda_i^2(t) \ll 1$ for $i = 1, 2, \ldots M$ and every $t \in [0, T]$. The structure of a system which determines $B\left(\mathscr{Y}_T; \{\lambda_t\}, T\right)$ is illustrated in Figure 2.4.

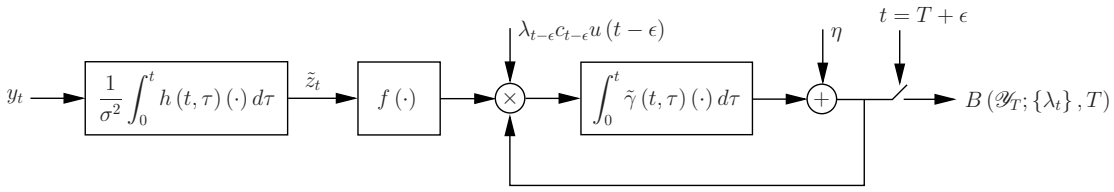

Figure 2.4: Structure of a system which determines $B\left(\mathscr{Y}_T; \{\lambda_t\}, T\right)$ by solving (2.56). In this block diagram, we have $\eta = \exp\left(-\int_0^T \lambda_t dt\right)$ and $\tilde{\gamma}(t, \tau) = \gamma(t - \epsilon, \tau - \epsilon)$.

Chapter 3

Estimation and Control with Space-Time Point Process Observations

3.1 Introduction

The concern of this chapter is to estimate the state of a stochastic dynamical model which modulates the rate of a space-time point process. In addition, an associated optimal control problem will be discussed which has a direct application in the optical beam tracking aspect of free-space optics. Sections 3.2 and 3.3 present the prior work by Snyder and Fishman [39] and Rhodes and Snyder [20], with the intention of providing the necessary background for the next chapters. In Section 3.4, we introduce a new formulation of the problem which is useful in generalizing the results of [39, 20]. An approximation method will be developed in Section 3.5 which incorporates the results of [39, 20] to explore a suboptimal control law for an optical beam tracking system with a finite resolution position-sensitive photodetector.

3.2 The Model and Problem Statement

In this section, we present a stochastic model in which the state of a linear stochastic state-space equation modulates the rate of a space-time point process which is regarded as the system output. In our description of the model, we closely follow [39, 20]. For a rigorous and complete treatment of the space-time point process we refer the reader to [40]. After introducing the model, we state its associated estimation and control problems.

3.2.1 The Model

Consider the stochastic linear dynamical model

$$dx_t = A_t x_t dt + B_t u_t dt + D_t dw_t \tag{3.1}$$

where $x_t \in \mathbb{R}^n$ and $u_t \in \mathbb{R}^k$ are random vectors standing for state and control, $\{w_t\}$ is a p-dimensional standard Wiener process, and A_t, B_t, and D_t are uniformly bounded deterministic matrices with proper dimensions. The initial state x_0 is a Gaussian vector with mean $\bar{x}_0$ and covariance matrix $\bar{\Sigma}_0$ and is independent of $\{w_t\}$.

The observation of the state vector is provided via a space-time point process defined over $[0, \infty) \times \mathcal{A}$, where $\mathcal{A} \subseteq \mathbb{R}^m$, $m \leqslant n$. Each point of this process is characterized in terms of a temporal component $t_i \in [0, \infty)$ and a spatial component $r_i \in \mathcal{A}$. Let the nonnegative scalar map $\lambda_t(r, x_t)$ which is defined over $t \in [0, \infty)$ and $r \in \mathcal{A}$ and is parameterized by the state vector x_t, be the rate associated with

this process. Then, the space-time point process is (statistically) characterized as follows.

Let $N(\mathcal{T} \times \mathcal{S})$ denote the number of points occurring in $\mathcal{T} \times \mathcal{S}$, where $\mathcal{T}$ and $\mathcal{S}$ are Borel sets in $[0, \infty)$ and $\mathcal{A}$, respectively. Associated with $N(\mathcal{T} \times \mathcal{S})$, define the random variable

$$\Lambda(\mathcal{T} \times \mathcal{S}) = \int_{\mathcal{T} \times \mathcal{S}} \lambda_t(r, x_t)\, dtdr.$$

Then, conditioned on $\Lambda(\mathcal{T} \times \mathcal{S})$, the random variable $N(\mathcal{T} \times \mathcal{S})$ is Poisson distributed, i.e.,

$$\Pr\{N(\mathcal{T} \times \mathcal{S}) = n | \Lambda(\mathcal{T} \times \mathcal{S})\} = \frac{e^{-\Lambda(\mathcal{T} \times \mathcal{S})} \Lambda^n(\mathcal{T} \times \mathcal{S})}{n!}.$$

Moreover, for disjoint $\mathcal{T}_1 \times \mathcal{S}_1$ and $\mathcal{T}_2 \times \mathcal{S}_2$, conditioned on $\Lambda(\mathcal{T}_1 \times \mathcal{S}_1)$ and $\Lambda(\mathcal{T}_2 \times \mathcal{S}_2)$, the random variables $N(\mathcal{T}_1 \times \mathcal{S}_1)$ and $N(\mathcal{T}_2 \times \mathcal{S}_2)$ are statistically independent.

We shall assume that the rate of the space-time process has the form

$$\lambda_t(r, x_t) = \mu_t \gamma_t(r - C_t x_t) \tag{3.2}$$

where C_t is a bounded $m \times n$ matrix, the known function μ_t is nonnegative for every $t \geqslant 0$, and the nonnegative map $\gamma_t(\cdot) : \mathbb{R} \times \mathbb{R}^m \to \mathbb{R}$ satisfies

$$\int_{\mathbb{R}^m} \gamma_t(r)\, dr = 1. \tag{3.3}$$

In particular, we are interested in the case that $\gamma_t(\cdot)$ is a Gaussian map, i.e.,

$$\gamma_t(r) = \Phi_m(r; 0, R_t) \tag{3.4}$$

where $R_t = R_t^T$ is a $m \times m$ positive definite matrix and $\Phi_k : \mathbb{R}^k \times \mathbb{R}^k \times \mathbb{R}^{k \times k} \to \mathbb{R},\ k = 1, 2, 3, \ldots$ is defined as

$$\Phi_k(z; \bar{z}, \Theta) = (2\pi)^{-k/2} (\det \Theta)^{-1/2} \exp\left(-\frac{1}{2}(z - \bar{z})^T \Theta^{-1} (z - \bar{z})\right). \tag{3.5}$$

3.2.2 Problem Statement

Let $(\Omega, \mathscr{F}, P)$ be the underlying probability space for the stochastic model of Section 3.2.1. Define $\mathscr{B}_t$ as the σ-algebra generated by the space-time point process over $[0, t)$ and assume that the control vector u_t is $\mathscr{B}_t$-measurable. We say u_t is an admissible control if it is $\mathscr{B}_t$-measurable and the solution to (3.1) is well-defined. Subject to the model of Section 3.2.1, we define the following problems.

Estimation Problem: For every $t > 0$, determine $p_{x_t}(x|\mathscr{B}_t)$, the posterior density of the state vector x_t given $\mathscr{B}_t$. In particular, determine the conditional expectation $\hat{x}_t \triangleq \mathrm{E}[x_t|\mathscr{B}_t]$ and the conditional covariance matrix $\Sigma_t \triangleq \mathrm{E}\left[(x_t - \hat{x}_t)(x_t - \hat{x}_t)^T \middle| \mathscr{B}_t\right]$.

Control Problem: Find an admissible control $\{u_t,\ t \in [0, T]\}$ which minimizes the

cost functional

$$J = \mathrm{E}\left[\int_0^T \left(x_t^T Q_t x_t + u_t^T P_t u_t\right) dt + x_T^T S x_T\right] \tag{3.6}$$

where $P_t = P_t^T > 0$, $Q_t = Q_t^T \geqslant 0$, and $S = S^T \geqslant 0$ are matrices of proper dimensions and $T > 0$ is a fixed time horizon.

3.3 Relevant Prior Work

In this section, we state the results obtained in [39, 20] regarding the estimation and control problems defined in Section 3.2.2. These results provide an adequate framework for our discussion in the next chapters. Theorem 3.3.1 below provides a solution to the estimation problem in the most general case. The rest of theorems address the special case in which the rate of the space-time point process is given by (3.2) and (3.4).

Before stating the theorems, we fix notation. Let $(t_{k-1}, t_k]$ be the interval between two successive occurrence of the space-time point process and r_k be the location of k^{th} occurring point. Assume that $h_t(r, \xi_t)$ is continuous in r and left continuous in t and ξ_t. Then, the stochastic differential equation

$$d\xi_t = \int_{\mathcal{A}} h_t(r, \xi_t)\, N(dt \times dr)$$

is defined such that $d\xi_t = 0$ during (t_{k-1}, t_k) and ξ_t encounters a jump of $h_{t_k}(r_k, \xi_{t_k})$

at $t = t_k$, i.e.,

$$\xi_{t_k^+} = \xi_{t_k^-} + h_{t_k}\left(r_k, \xi_{t_k^-}\right).$$

Theorem 3.3.1 (Rhodes and Snyder 1977 [20])**.** Consider the state-space equation (3.1) and its associated space-time observation defined in Section 3.2.1. Assume that the increasing family of σ-algebras $\mathscr{B}_t$ are given and that u_t is $\mathscr{B}_t$-measurable. Then, the posterior density of the state vector x_t is the solution of the stochastic partial differential-integral equation

$$\begin{aligned} dp_{x_t}\left(x|\mathscr{B}_t\right) = \mathcal{L}\left\{p_{x_t}\left(x|\mathscr{B}_t\right)\right\}dt + p_{x_t}\left(x|\mathscr{B}_t\right)\int_{\mathcal{A}}\left(\lambda_t\left(r,x\right)\hat{\lambda}_t^{-1}\left(r\right) - 1\right)N\left(dt\times dr\right) \\ - p_{x_t}\left(x|\mathscr{B}_t\right)\int_{\mathcal{A}}\left(\lambda_t\left(r,x\right) - \hat{\lambda}_t\left(r\right)\right)drdt \end{aligned} \tag{3.7}$$

where $\hat{\lambda}_t\left(r\right) = \mathrm{E}\left[\lambda_t\left(r,x_t\right)|\mathscr{B}_t\right]$ and $\mathcal{L}\{\cdot\}$ is the forward Kolmogorov operator associated with (3.1) defined as

$$\mathcal{L}\{\cdot\} = -\sum_{i=1}^{n}\partial\left[\left(A_t x + B_t u_t\right)(\cdot)\right]_i/\partial x^i + \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\partial^2\left[D_t D_t^T(\cdot)\right]_{ij}/\partial x^i\partial x^j.$$

Proof. See [39, 20]. □

Corollary 3.3.1 (Rhodes and Snyder 1977 [20])**.** Assume that $\mathcal{A} = \mathbb{R}^m$ and $\lambda_t\left(r,x_t\right)$ is given by (3.2). Let $\{\mu_t\}$ be a nonnegative stochastic process which is statistically independent of x_0 and $\{w_t\}$. Then, under the assumptions of Theorem 3.3.1, the

posterior density $p_{x_t}(x|\mathscr{B}_t)$ is the solution of

$$dp_{x_t}(x|\mathscr{B}_t) = \mathcal{L}\{p_{x_t}(x|\mathscr{B}_t)\}\,dt + p_{x_t}(x|\mathscr{B}_t)\int_{\mathbb{R}^m}\left(\gamma_t(r - C_t x)\,\hat{\gamma}_t^{-1}(r) - 1\right)N(dt\times dr) \tag{3.8}$$

where $\hat{\gamma}_t(\cdot)$ is defined as $\hat{\gamma}_t(r) = \mathrm{E}[\gamma_t(r - C_t x_t)\,|\mathscr{B}_t]$.

Proof. Following [20], we temporarily replace $\mathscr{B}_t$ with $\mathscr{B}'_t = \mathscr{B}_t \cup \mathscr{M}$, where $\mathscr{M}$ is the σ-algebra generated by $\{\mu_t\}$ over $[0,\infty)$. Then, Theorem 3.3.1 indicates that $p_{x_t}(x|\mathscr{B}'_t)$ is the solution of (3.7) with $\lambda_t(r, x_t)$ replaced from (3.2) and $\mathscr{B}_t$ replaced with $\mathscr{B}'_t$. From condition (3.3), it is easy to verify that the second integral on the right side of this equation is identically zero, which leads to (3.8) with $\mathscr{B}_t$ replaced by $\mathscr{B}'_t$. Since this equation does not depend on $\{\mu_t\}$, we can replace $\mathscr{B}'_t$ with $\mathscr{B}_t$ to show that $p_{x_t}(x|\mathscr{B}_t)$ satisfies (3.8). □

Theorem 3.3.2 (Rhodes and Snyder 1977 [20])**.** Let $\gamma_t(\cdot)$ be the Gaussian map (3.4). Then, under the assumptions of Corollary 3.3.1, the posterior density $p_{x_t}(x|\mathscr{B}_t)$ is Gaussian, i.e.,

$$p_{x_t}(x|\mathscr{B}_t) = \Phi_n(x; \hat{x}_t, \Sigma_t).$$

Here, the conditional expectation $\hat{x}_t$ and the conditional covariance matrix Σ_t are the solutions of the stochastic differential equations

$$d\hat{x}_t = A_t\hat{x}_t dt + B_t u_t dt + \int_{\mathbb{R}^m} M_t(r - C_t\hat{x}_t)\,N(dt\times dr) \tag{3.9a}$$

$$d\Sigma_t = A_t\Sigma_t dt + \Sigma_t A_t^T dt + D_t D_t^T dt - M_t C_t \Sigma_t dN_t \tag{3.9b}$$

with the initial states $\hat{x}_0 = \bar{x}_0$ and $\Sigma_0 = \bar{\Sigma}_0$, where $N_t = N([0,t) \times \mathbb{R}^m)$ and M_t is defined as

$$M_t = \Sigma_t C_t^T \left(C_t \Sigma_t C_t^T + R_t\right)^{-1}.$$

Moreover, the conditional covariance matrix Σ_t is almost surely positive definite for $t > 0$, provided that $\bar{\Sigma}_0$ is positive definite.

Proof. We outline the proof and refer the reader to [39] for the details. Let t_1 be the first occurrence time of the space-time point process. During $t \in [0, t_1)$, the integral on the right side of (3.8) is zero and the equation is reduced to

$$\frac{\partial p_{x_t}(x|\mathscr{B}_t)}{\partial t} = \mathcal{L}\left\{p_{x_t}(x|\mathscr{B}_t)\right\}. \tag{3.10}$$

The solution to this equation is a Gaussian function with mean $\hat{x}_t$ and covariance matrix Σ_t. In the transition from t_1^- to t_1^+, the integral on the right side of (3.8) causes $p_{x_t}(x|\mathscr{B}_t)$ to jump from $\Phi_n\left(x; \hat{x}_{t_1^-}, \Sigma_{t_1^-}\right)$ into $\Phi_n\left(x; \hat{x}_{t_1^+}, \Sigma_{t_1^+}\right)$. Continuing this procedure, we find that $p_{x_t}(x|\mathscr{B}_t)$ is Gaussian for every $t \geqslant 0$.

To prove the second statement, we note that Σ_t is positive definite during $t \in [0, t_1)$, since in equation (3.9b), the initial state $\bar{\Sigma}_0$ and $D_t D_t^T$ are positive definite. Also, at $t = t_1$ we have

$$\begin{aligned}
\Sigma_{t_1^+} &= \Sigma_{t_1^-} - \Sigma_{t_1^-} C_{t_1}^T \left(C_{t_1} \Sigma_{t_1^-} C_{t_1}^T + R_{t_1}\right)^{-1} C_{t_1} \Sigma_{t_1^-} \\
&= \left(\Sigma_{t_1^-}^{-1} + C_{t_1}^T R_{t_1}^{-1} C_{t_1}\right)^{-1}
\end{aligned}$$

which indicates that $\Sigma_{t_1^+}$ is positive definite. The proof can be completed by con-

tinuing this procedure. □

Theorem 3.3.3 (Rhodes and Snyder 1977 [20])**.** Under the assumptions of Theorem 3.3.2, the admissible control u_t^* which minimizes the cost functional (3.6) is uniquely given by

$$u_t^* = -P_t^{-1} B_t K_t \hat{x}_t \tag{3.11}$$

where the $n \times n$ nonnegative definite matrix K_t is the backward solution of the Riccati equation

$$\dot{K}_t = -K_t A_t - A_t^T K_t + K_t B_t P_t^{-1} B_t^T K_t - Q_t$$

with the terminal condition $K_T = S$. The minimum of the cost functional J associated with (3.11) is given by

$$J^* = \mathrm{E}\left[x_0^T K_0 x_0\right] + \int_0^T \mathrm{tr}\left(K_t B_t P_t^{-1} B_t^T K_t \mathrm{E}\left[\Sigma_t\right] + K_t D_t D_t^T\right) dt.$$

Proof. According to [20], J can be expressed as

$$\begin{aligned} J = \mathrm{E}\left[\int_0^T \left(u_t + P_t^{-1} B_t K_t \hat{x}_t\right)^T P_t \left(u_t + P_t^{-1} B_t K_t \hat{x}_t\right) dt\right] \\ + \mathrm{E}\left[x_0^T K_0 x_0\right] + \int_0^T \mathrm{tr}\left(K_t B_t P_t^{-1} B_t^T K_t \mathrm{E}\left[\Sigma_t\right] + K_t D_t D_t^T\right) dt. \end{aligned}$$

It is easy to verify that the second and the third terms on the right side of the expression above do not depend on u_t, thus we must minimize the first term. This term is always nonnegative with a minimum of zero which is achieved by choosing u_t

as (3.11). □

3.4 A New Formulation for the Estimation Problem

Throughout this section, we assume that $\lambda_t(r, x_t)$ is given by (3.2) and $\gamma_t(\cdot)$ satisfies (3.3), while it is not necessarily Gaussian. Further, we assume that $\mathcal{A} = \mathbb{R}^m$, the pair of (A_t, D_t) is controllable, and the $\mathscr{B}_t$-measurable u_t is a piecewise continuous and almost surely bounded stochastic process. Under these assumptions, we determine the solution of the estimation problem in terms of the posterior density associated with a discrete-time model. The procedure for obtaining this new representation is explained below.

Lemma 3.4.1. Consider the stochastic differential equation

$$d\tilde{x}_t = A_t \tilde{x}_t dt + D_t dw_t \tag{3.12}$$

and assume that for the fixed but arbitrary $t^* \geqslant 0$, $\tilde{x}_{t^*}$ is independent of $\{w_t,\ t \geqslant t^*\}$ and its probability density function is known. Then, for every $t \geqslant t^*$, the probability density function $p_{\tilde{x}_t}(\tilde{x})$ is given by

$$p_{\tilde{x}_t}(\tilde{x}) = \int_{\mathbb{R}^n} \Phi_n\left(\tilde{x}; \Phi^A(t, t^*)\, x^*, W(t, t^*)\right) p_{\tilde{x}_{t^*}}(x^*)\, dx^*. \tag{3.13}$$

Here, $\Phi^A(t,\tau)$ is the transition matrix associated with A_t which satisfies

$$\begin{aligned} \frac{\partial \Phi^A(t,\tau)}{\partial t} &= A_t \Phi^A(t,\tau) \\ \Phi^A(\tau,\tau) &= I_{n\times n} \end{aligned} \tag{3.14}$$

for every $t \geqslant \tau$ and $W(t,t^*)$ is defined as

$$W(t,t^*) = \int_{t^*}^{t} \Phi^A(t,\tau)\, D_\tau D_\tau^T \left(\Phi^A(t,\tau)\right)^T d\tau.$$

Proof. We note that for every $t \geqslant t^*$, the solution of (3.12) is given by

$$x_t = \Phi^A(t,t^*)\, x_{t^*} + \int_{t^*}^{t} \Phi^A(t,\tau)\, D_\tau dw_\tau.$$

The two random vectors on the right side of this expression are independent and the second one is a zero-mean Gaussian random vector with covariance matrix $W(t,t^*)$. This leads to the convolution described by (3.13). □

Consider the linear discrete-time state-space model

$$\theta_{k+1} = F_k \theta_k + G_k \omega_k \tag{3.15a}$$

$$\rho_k = L_k \theta_k + \nu_k \tag{3.15b}$$

where $\theta_k \in \mathbb{R}^n$ is the state vector and the i.i.d. random vectors $\omega_k \in \mathbb{R}^n$, $k = 1,2,3,\ldots$ are zero-mean and Gaussian with covariance matrix $I_{n\times n}$. The matrices F_k and L_k in (3.15) are defined as $F_k = \Phi^A(t_{k+1},t_k)$ and $L_k = C_{t_k}$, respectively,

where $t_0 = 0$ and $t_k,\ k = 1, 2, 3, \ldots$ is the k^{th} occurrence time of the space-time point process. The $n \times n$ matrix G_k is defined in such a manner that $G_k G_k^T = W(t_{k+1}, t_k)$. The random vector $\nu_k \in \mathbb{R}^m$ is distributed according to the probability density $p_{\nu_k}(\nu) = \gamma_{t_k}(\nu)$ and ν_k and ν_l are independent for $k \neq l$. The initial state θ_0 is a Gaussian random vector with mean $\bar{x}_0$ and covariance matrix $\bar{\Sigma}_0$ and is independent of $\{\omega_k\}$ and $\{\nu_k\}$. We denote the history of the measurement vector ρ_k (up to k) by $\mathscr{R}_k = \{\rho_1, \rho_2, \ldots, \rho_k\},\ k = 1, 2, 3, \ldots$, and $\mathscr{R}_0 = \emptyset$.

Theorem 3.4.1. Suppose that the measurement vector ρ_k in (3.15b) is generated according to $\rho_k = r_k - C_{t_k} v_{t_k}$, where r_k is the location of the event occurred at $t = t_k$ (associated with the space-time point process) and the stochastic process $\{v_t,\ t \geqslant 0\}$ is defined as

$$v_t = \int_0^t \Phi^A(t, \tau) B_\tau u_\tau d\tau.$$

Then, for $t \in [t_k, t_{k+1}),\ k = 0, 1, 2, \ldots$, the posterior density $p_{x_t}(x|\mathscr{B}_t)$ can be expressed as

$$p_{x_t}(x|\mathscr{B}_t) = \int_{\mathbb{R}^n} \Phi_n\left(x; \Phi^A(t, t_k)\theta + v_t, W(t, t_k)\right) p_{\theta_k}(\theta|\mathscr{R}_k)\, d\theta \tag{3.16}$$

where $p_{\theta_k}(\theta|\mathscr{R}_k)$ is the posterior density of the state vector θ_k in (3.15a), conditioned on $\mathscr{R}_k$.

Proof. Let $\tilde{x}_t = x_t - v_t$ be the solution of (3.12) and assume that $p_{\tilde{x}_t}(\tilde{x}|\mathscr{B}_t)$ is known at $t = t_k$. For the time interval $t \in [t_k, t_{k+1})$ in which (3.8) reduces to (3.10), we use

Lemma 3.4.1 to obtain

$$p_{\tilde{x}_t}\left(\tilde{x}|\mathscr{B}_t\right) = \int_{\mathbb{R}^n} \Phi_n\left(\tilde{x}; \Phi^A\left(t,t_k\right)x^*, W\left(t,t_k\right)\right) p_{\tilde{x}_{t_k}}\left(x^*|\mathscr{B}_{t_k}\right) dx^*. \tag{3.17}$$

Also, from $x_t = \tilde{x}_t + v_t$ and the fact that v_t is measurable with respect to $\mathscr{B}_t$, we have

$$p_{x_t}\left(x|\mathscr{B}_t\right) = \int_{\mathbb{R}^n} \Phi_n\left(x; \Phi^A\left(t,t_k\right)x^* + v_t, W\left(t,t_k\right)\right) p_{\tilde{x}_{t_k}}\left(x^*|\mathscr{B}_{t_k}\right) dx^*. \tag{3.18}$$

At $t = t_{k+1}^-$, (3.17) can be expressed as

$$p_{\tilde{x}_{t_{k+1}^-}}\left(\tilde{x}|\mathscr{B}_{t_{k+1}^-}\right) = \int_{\mathbb{R}^n} \Phi_n\left(\tilde{x}; F_k x^*, G_k G_k^T\right) p_{\tilde{x}_{t_k}}\left(x^*|\mathscr{B}_{t_k}\right) dx^*. \tag{3.19}$$

We solve (3.8) between t_{k+1}^- and t_{k+1}, in order to get

$$p_{x_{t_{k+1}}}\left(x|\mathscr{B}_{t_{k+1}}\right) = \frac{p_{x_{t_{k+1}^-}}\left(x|\mathscr{B}_{t_{k+1}^-}\right)\gamma_{t_{k+1}}\left(r_{k+1} - C_{t_{k+1}}x\right)}{\displaystyle\int_{\mathbb{R}^n} p_{x_{t_{k+1}^-}}\left(x^*|\mathscr{B}_{t_{k+1}^-}\right)\gamma_{t_{k+1}}\left(r_{k+1} - C_{t_{k+1}}x^*\right)dx^*}.$$

Using this result, the equality $x_t = \tilde{x}_t + v_t$, the continuity of v_t, and $r_{k+1} - C_{t_{k+1}} v_{t_{k+1}} = \rho_{k+1}$, we obtain

$$p_{\tilde{x}_{t_{k+1}}}\left(\tilde{x}|\mathscr{B}_{t_{k+1}}\right) = \frac{p_{\tilde{x}_{t_{k+1}^-}}\left(\tilde{x}|\mathscr{B}_{t_{k+1}^-}\right)p_{\nu_{k+1}}\left(\rho_{k+1} - L_{k+1}\tilde{x}\right)}{\displaystyle\int_{\mathbb{R}^n} p_{\tilde{x}_{t_{k+1}^-}}\left(x^*|\mathscr{B}_{t_{k+1}^-}\right)p_{\nu_{k+1}}\left(\rho_{k+1} - L_{k+1}x^*\right)dx^*}. \tag{3.20}$$

The recursive formulas (3.19) and (3.20) specify a two-step procedure for determining $p_{\tilde{x}_{t_{k+1}}}\left(\tilde{x}|\mathscr{B}_{t_{k+1}}\right)$ in terms of $p_{\tilde{x}_{t_k}}\left(\tilde{x}|\mathscr{B}_{t_k}\right)$. We shall show that the same procedure

can be used to determine $p_{\theta_{k+1}}(\theta|\mathscr{R}_{k+1})$ in terms of $p_{\theta_k}(\theta|\mathscr{R}_k)$.

Using the law of total probability, we can write

$$\begin{aligned} p_{\theta_{k+1}}(\theta|\mathscr{R}_k) &= \int_{\mathbb{R}^n} p_{\theta_{k+1}}(\theta|\theta_k = \theta^*, \mathscr{R}_k)\, p_{\theta_k}(\theta^*|\mathscr{R}_k)\, d\theta^* \\ &= \int_{\mathbb{R}^n} \Phi_n\left(\theta; F_k\theta^*, G_k G_k^T\right) p_{\theta_k}(\theta^*|\mathscr{R}_k)\, d\theta^* \end{aligned} \tag{3.21}$$

where the second equality is obtained from (3.15a), noting that

$$p_{\theta_{k+1}}(\theta|\theta_k = \theta^*, \mathscr{R}_k) = p_{\theta_{k+1}}(\theta|\theta_k = \theta^*)\,.$$

Also, from Bayes' rule we obtain

$$\begin{aligned} p_{\theta_{k+1}}(\theta|\mathscr{R}_{k+1}) &= p_{\theta_{k+1}}(\theta|\mathscr{R}_k, \rho_{k+1}) \\ &= \frac{p_{\theta_{k+1}}(\theta|\mathscr{R}_k)\, p_{\rho_{k+1}}(\rho_{k+1}|\theta_{k+1} = \theta, \mathscr{R}_k)}{\displaystyle\int_{\mathbb{R}^n} p_{\theta_{k+1}}(\theta^*|\mathscr{R}_k)\, p_{\rho_{k+1}}(\rho_{k+1}|\theta_{k+1} = \theta^*, \mathscr{R}_k)\, d\theta^*} \\ &= \frac{p_{\theta_{k+1}}(\theta|\mathscr{R}_k)\, p_{\nu_{k+1}}(\rho_{k+1} - L_{k+1}\theta)}{\displaystyle\int_{\mathbb{R}^n} p_{\theta_{k+1}}(\theta^*|\mathscr{R}_k)\, p_{\nu_{k+1}}(\rho_{k+1} - L_{k+1}\theta^*)\, d\theta^*} \end{aligned} \tag{3.22}$$

where the last equality is derived from (3.15b). Comparing the pair of formulas (3.21) and (3.22) with (3.19) and (3.20), we conclude that $p_{\tilde{x}_{t_k}}(\theta|\mathscr{B}_{t_k}) = p_{\theta_k}(\theta|\mathscr{R}_k)$, $k = 0, 1, 2, \ldots$, which leads to (3.16) upon substituting into (3.18). □

For the special case that $\nu_k,\ k = 1, 2, 3, \ldots$ is a Gaussian random vector, the estimation problem associated with the discrete-time model (3.15) has an exact Gaussian solution, which is consistent with the results of [39, 20]. For the gen-

eral case, this estimation problem is difficult to solve, i.e., the problem is infinite-dimensional. While it seems that Theorem 3.4.1 converts a hard-to-solve problem into another hard-to-solve problem, the new formulation might be easier to approach, due to the discrete-time and linear nature of the model.

3.5 Optical Beam Tracking

In Section 1.2.1, we briefly discussed the operation of optical beam tracking. This operation has been studied by Snyder [19] in a stochastic framework using an idealized model for the photodetector. In that work, the dynamics of the pointing assembly and the relative motion is modeled by (3.1), where u_t is the (control) input of the pointing assembly. Also, the location of the center of spot of light is modeled by $C_t x_t$ and its optical intensity is described by (3.2). In addition, [19] considers the Gaussian model (3.4) for the intensity pattern of the spot of light[1]. Regarding the position-sensitive photodetector, [19] makes two ideal assumptions: the photodetector has an infinite spatial resolution and an infinite area ($\mathcal{A} = \mathbb{R}^2$). The first assumption allows us to model the output of the photodetector by a space-time point process with rate[2] $\lambda_t(r, x_t)$ in (3.2) and the second one makes it possible to use the results of Theorems 3.3.2 and 3.3.3. Finally, the problem of optical beam tracking can be formulated in terms of minimizing (3.6) with $Q_t = C_t^T C_t$. For a detailed derivation of this model see Chapter 5.

In a more realistic model, while keeping (3.1), (3.2), and (3.4), we describe

[1] To evaluate the validity of this assumption see Chapter 5

[2] Here, the background radiation and the dark current noise are ignored.

the output of the position-sensitive photodetector by a point process vector. For this purpose, let $\mathcal{A}^i,\ i = 1, 2, \ldots, q$ denote the i^{th} partition on the surface of the photodetector such that $\bigcup_{i=1}^{q} \mathcal{A}^i = \mathcal{A}$. The output of the region $\mathcal{A}^i$ will be modeled by a doubly stochastic Poisson process Y_t^i with rate $\Lambda_t^i(x_t)$, where $\Lambda_t^i(\cdot) : \mathbb{R}\times\mathbb{R}^n \to \mathbb{R}$ is defined as

$$\Lambda_t^i(x) = \int_{\mathcal{A}^i} \lambda_t(r, x)\, dr.$$

To have a compact notation, we put $Y_t^i,\ i = 1, 2, \ldots, q$ in a vector Y_t and express the output of the position-sensitive photodetector by $Y_t = (Y_t^1, Y_t^2, \ldots, Y_t^q)$.

In order to solve the optimal control problem associated with this new model, we need to obtain an equation which describes the temporal evolution of the posterior density (similar to (3.7)). The filtering problem associated with this equation is infinite-dimensional, which requires some sort of approximation to reduce it into a finite-dimensional problem. An alternative to this approach is to start from the idealized model explained above and derive an appropriate approximation from the "exact" results of Theorem 3.3.2 and Theorem 3.3.3. To justify this approach, we note that the "infinite resolution" assumption provides a reasonable approximation for a high spatial resolution photodetector. Also, the "infinite area" assumption is appropriate when the photodetector area is significantly larger than the size and the displacement of the spot of light.

Following this approach, we approximate the optimal control associated with

the "finite resolution" model as

$$\tilde{u}_t = -P_t^{-1} B_t K_t \tilde{x}_t$$

where $\tilde{x}_t$ is the solution of the stochastic differential equations

$$d\tilde{x}_t = A_t \tilde{x}_t dt + B_t \tilde{u}_t dt + \sum_{i=1}^{q} \tilde{M}_t \left(r_t^i - C_t \tilde{x}_t \right) dY_t^i \tag{3.23a}$$

$$d\tilde{\Sigma}_t = A_t \tilde{\Sigma}_t dt + \tilde{\Sigma}_t A_t^T dt + D_t D_t^T dt - \sum_{i=1}^{q} \tilde{M}_t C_t \tilde{\Sigma}_t dY_t^i \tag{3.23b}$$

with the initial state $\tilde{x}_0 = \bar{x}_0$ and $\tilde{\Sigma}_0 = \bar{\Sigma}_0$. Here, $r_t^i \in \mathcal{A}^i$ is a representative point of the region $\mathcal{A}^i$ and $\tilde{M}_t$ is defined as

$$\tilde{M}_t = \tilde{\Sigma}_t C_t^T \left(C_t \tilde{\Sigma}_t C_t^T + R_t \right)^{-1}.$$

We note that Y_t^i is the integral over $\mathcal{A}^i$ of the space-time point process, i.e., each point (event) of this process which occurs on $\mathcal{A}^i$ increases the value of Y_t^i by one unit. The information we lose by replacing the space-time point process with Y_t is the knowledge of the exact occurrence location of the points on $\mathcal{A}^i$. In fact, we derived (3.23a) from (3.9a) by replacing the exact occurrence location of the points with r_t^i as a representative point of $\mathcal{A}^i$. This suggests that to achieve the best performance of the estimator (3.23), r_t^i must be chosen as a "good" estimate of the occurrence locations, based on the past observation of Y_t.

Based on the explanation above, an appropriate choice for r_t^i is the minimum

mean squared error (MMSE) estimator

$$r_t^i = \frac{\displaystyle\int_{\mathcal{A}^i} r\mathrm{E}\left[\Phi_m\left(r; C_t x_t, R_t\right)|\mathscr{Y}_t\right] dr}{\displaystyle\int_{\mathcal{A}^i} \mathrm{E}\left[\Phi_m\left(r; C_t x_t, R_t\right)|\mathscr{Y}_t\right] dr}$$

where $\mathscr{Y}_t$ is the σ-algebra generated by Y_t over $[0,t)$. Approximating $p_{x_t}\left(x|\mathscr{Y}_t\right)$ with $\Phi_n\big(x; \tilde{x}_t, \tilde{\Sigma}_t\big)$, the expression above can be written as[1]

$$r_t^i = \frac{\displaystyle\int_{\mathcal{A}^i} r\Phi_m\left(r; C_t\tilde{x}_t, R_t + C_t\tilde{\Sigma}_t C_t^T\right) dr}{\displaystyle\int_{\mathcal{A}^i} \Phi_m\left(r; C_t\tilde{x}_t, R_t + C_t\tilde{\Sigma}_t C_t^T\right) dr}.$$

Another suitable choice for r_t^i is the maximum a posteriori (MAP) estimator

$$r_t^i = \arg\max_{r\in\mathcal{A}^i} \Phi_m\left(r; C_t\tilde{x}_t, R_t + C_t\tilde{\Sigma}_t C_t^T\right).$$

When the partition $\mathcal{A}^1, \mathcal{A}^2, \ldots, \mathcal{A}^q$ is fine enough, for sake of simplicity, r_t^i can be a predetermined point of $\mathcal{A}^i$ such as

$$r_t^i = \frac{\displaystyle\int_{\mathcal{A}^i} r\Phi_m\left(r; 0, R_t\right) dr}{\displaystyle\int_{\mathcal{A}^i} \Phi_m\left(r; 0, R_t\right) dr}.$$

[1]See Lemma 4.5.1.

Chapter 4

Active Pointing

4.1 Introduction

In this chapter we study the estimation and control problems defined in Section 3.2.2 for the case of $\mathcal{A} \neq \mathbb{R}^m$. Since under this assumption, the associated filtering problem is infinite-dimensional, we focus our attention on an approximate solution for the problem, which leads to a suboptimal estimator and controller.

The motivation for this study is its application in an active fine pointing scheme for short range free-space optical communication. The one-way optical link under consideration comprises an optical transmitter and an optical receiver which are subject to relative motion. The optical transmitter is equipped with an electromechanical pointing assembly which can control the azimuth and elevation of a transmitting laser source. The optical beam emitted by the laser source has a nonuniform intensity profile which is assumed to be Gaussian [41]. Normally, the aperture of the receiver is smaller than the received optical beam, so that the receiver can collect only a fraction of the optical power. In order to enlarge this captured fraction, the goal of active pointing is to hold the center of the optical beam at the center

of the receiving aperture. The receiver employs a position-sensitive photodetector to measure the intensity profile of the optical beam that strikes its aperture. The output of the photodetector is used to estimate the center of the received optical beam, which is then conveyed to the transmitter through an optical link or a low-bandwidth RF channel. The pointing assembly then adjusts the orientation of the transmitter based on this estimate. The concept of this active pointing method is illustrated in the block diagram of Figure 1.2.

The performance of the proposed active pointing scheme depends significantly on the accuracy of the estimate of the beam center. In order to achieve a good estimate of the beam center, it is necessary that the size of the receiving aperture be comparable with the size of the beam. This requirement limits the application of the method to short distance links.

The remainder of this chapter is organized as follows. In the next section, we show that the overall active pointing scheme can be described in terms of the model of Section 3.2.1 and its associated estimation and control problems in Section 3.2.2. Sections 4.3 and 4.4 consider the estimation and control problems, respectively. Since the proof of the theorems stated in these sections are long, for sake of continuity of discussion, we present the proofs in Section 4.5.

4.2 The Model

Let the two-dimensional vector θ_t denote the azimuth and elevation angles of the transmitter axis with respect to some fixed coordinate system. Similarly, α_t denotes the azimuth and elevation angles of the line-of-sight of the stations (passing through

the center of the receiving aperture) with respect to the same coordinate system. We assume that the receiving aperture is held perpendicular to the line-of-sight by means of an optical beam tracking system. Then, for a small pointing error $\theta_t - \alpha_t$, the displacement of the center of the optical beam with respect to the center of the receiving aperture is given by $y_t = l\,(\theta_t - \alpha_t)$, where the known constant l is the distance between the stations. Figure 4.1 illustrates the optical beam in the plane of the receiving aperture and the displacement vector y_t.

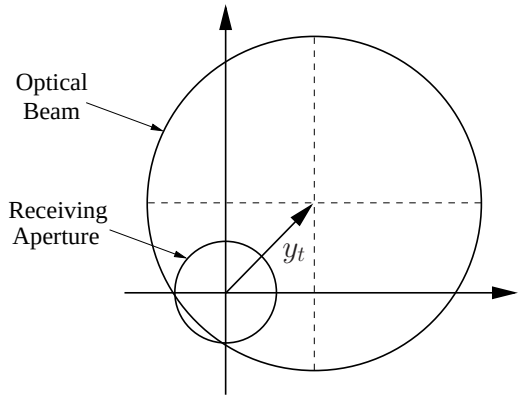

Figure 4.1: Receiving aperture, optical beam, and the displacement vector y_t.

The pointing assembly is an electromechanical system with the input vector $u_t \in \mathbb{R}^2$ and the output vector $\theta_t \in \mathbb{R}^2$. We describe this system by the linear stochastic state-space equations

$$\begin{aligned} dx_t^p &= A_t^p x_t^p dt + B_t^p u_t dt + D_t^p dw_t^p \\ \theta_t &= C_t^p x_t^p \end{aligned} \tag{4.1}$$

where $x_t^p \in \mathbb{R}^{n^p}$ is the state vector and $\{w_t^p\}$ is a m^p-dimensional standard Wiener

process. In this equation, A_t^p, B_t^p, D_t^p, and C_t^p are known uniformly bounded matrices of appropriate dimensions. Using a linear model for the pointing assembly is justified by the fact that the system operates over small angles during the active pointing regime[1].

We model α_t by a Gauss-Markov stochastic process described by the state-space equations

$$\begin{aligned} dx_t^d &= A_t^d x_t^d dt + D_t^d dw_t^d \\ \alpha_t &= C_t^d x_t^d \end{aligned} \tag{4.2}$$

with the state vector $x_t^d \in \mathbb{R}^{n^d}$, m^d-dimensional standard Wiener process $\left\{w_t^d\right\}$, and known uniformly bounded matrices A_t^d, D_t^d, and C_t^d of proper dimensions.

The displacement vector $y_t = l\left(\theta_t - \alpha_t\right)$ is a linear function of x_t^p and x_t^d, so that (4.1) and (4.2) can be combined in a compact form

$$\begin{aligned} dx_t &= A_t x_t dt + B_t u_t dt + D_t dw_t \\ y_t &= C_t x_t \end{aligned} \tag{4.3}$$

with the state vector $x_t \in \mathbb{R}^n$ and m-dimensional standard Wiener process $\{w_t\}$, where $n = n^p + n^d$ and $m = m^p + m^d$. The initial state x_0 is assumed to be Gaussian with mean $\bar{x}_0$ and covariance matrix $\bar{\Sigma}_0$, and independent of $\{w_t\}$.

Let r denote the position vector of an arbitrary point on the plane of the receiving aperture with respect to a coordinate system centered at the center of the

[1]For a detailed discussion of this issue see Chapter 5.

aperture. Then, for a Gaussian beam centered at $y_t = C_t x_t$, the optical intensity over the plane of the aperture is proportional to $\Phi_2(r - y_t; 0, R_t) = \Phi_2(r; C_t x_t, R_t)$, where $R_t = R_t^T$ is a 2×2 positive definite matrix describing the shape of the beam. For a circular symmetric beam with a constant radius $\varrho > 0$, R_t can be expressed as $R_t = \varrho^2 I_{2\times 2}$.

Let $\mathcal{A}$ denote the set of points on the receiving aperture. In a practical system, the optical field over the receiving aperture is focused on a position-sensitive photodetector of small surface area by means of a focusing lens. The photodetector measures the intensity profile of the imaged optical field, which is a scaled-down version of the optical intensity over the receiving aperture. Therefore, we consider the combination of the lens and the photodetector as a virtual photodetector of area $\mathcal{A}$, i.e., we assume that the virtual photodetector provides the observation of the optical intensity over $\mathcal{A}$.

Following [19] and our discussion in Section 3.5, we use an "infinite resolution" model for a high resolution photodetector employed by the receiver. According to this model, we describe the output of the photodetector by a space-time point process defined over $\mathcal{A}$ with a rate given by

$$\lambda_t(r, x_t) = \mu_t \Phi_2(r; C_t x_t, R_t).$$

where μ_t is a known nonnegative function. We remind from Section 3.2.2 that $\mathscr{B}_t$ denotes the σ-algebra generated by the space-time point process over $[0, t)$.

The central objective of an active pointing system is to maintain the centroid of

the optical beam as close as possible to the center of the photodetector. This control task can be interpreted as one of minimizing y_t with respect to some appropriate norm. For this purpose, we adopt the quadratic cost functional

$$J = \mathrm{E}\left[\int_0^T \left(x_t^T Q_t x_t + u_t^T P_t u_t\right) dt + x_T^T S x_T\right] \tag{4.4}$$

with $Q_t = C_t^T C_t$, $P_t = \rho I_{2\times 2}$, and $S = 0$, where $\rho > 0$ is a known constant.

Our discussion up to this point indicates that the controller design for an active pointing system can be formulated in terms of the control problem of Section 3.2.2 subject to the model of Section 3.2.1. An intermediate step for solving the control problem is to obtain the posterior density $p_{x_t}(x|\mathscr{B}_t)$. In the next section, we discuss this problem and develop an approximation for the posterior density. In Section 4.4, we employ this approximation in order to determine an approximate solution for the optimal control problem. Although for the specific application of active pointing, the space-time point process is defined over $\mathcal{A} \subset \mathbb{R}^2$, our results can be applied to the general case of $\mathcal{A} \subset \mathbb{R}^m$. Hence, for sake of generality, we present these results for an arbitrary integer m.

4.3 Estimation Problem

We remind from Chapter 3 that the posterior density $p_{x_t}(x|\mathscr{B}_t)$ is the solution of the stochastic partial differential-integral equation (3.7). For the case that $\mathcal{A} \neq \mathbb{R}^m$, the filtering problem associated with this equation is infinite-dimensional; however, when $\mathcal{A}$ is large compared with the size of the optical beam, a finite-dimensional

approximation is reasonable. The fact that for $\mathcal{A} = \mathbb{R}^m$, the posterior density $p_{x_t}(x|\mathscr{B}_t)$ is Gaussian[1], motivates us to consider a Gaussian approximation for $p_{x_t}(x|\mathscr{B}_t)$ when $\mathcal{A} \neq \mathbb{R}^m$. In the remainder of this section, we develop a method to determine the mean and covariance matrix of such a Gaussian approximation. The cumulant generating function associated with $p_{x_t}(x|\mathscr{B}_t)$ plays a central role in this development.

The conditional cumulant generating function of x_t given $\mathscr{B}_t$ is defined as

$$\psi_t(s) = \ln \mathrm{E}\left[\exp\left(j\omega^T x_t\right)|\mathscr{B}_t\right]\Big|_{j\omega=s}.$$

This function can be expanded in terms of the conditional cumulants $\kappa_t^{i_1 i_2 \cdots i_j}$ [42] as

$$\psi_t(s) = \sum_{j=1}^{\infty} \sum_{\mathscr{I}_j^n} \frac{1}{j!} \kappa_t^{i_1 i_2 \cdots i_j} s_{i_1} s_{i_2} \cdots s_{i_j} \tag{4.5}$$

where $\mathscr{I}_j^n = \{1, 2, \ldots, n\}^j$ and $s = (s_1, s_2, \ldots, s_n)$. Note that $\hat{x}_t$ and Σ_t are represented in terms of the first and the second order cumulants as $\hat{x}_t = (\kappa_t^1, \kappa_t^2, \ldots, \kappa_t^n)$ and $\Sigma_t = [\kappa_t^{ij}]$. The temporal evolution of $\psi_t(\cdot)$ is described by a partial differential-integral equation derived from (3.7) and is stated next.

Theorem 4.3.1. Let $\psi_t(\cdot)$ be the conditional cumulant generating function of x_t given $\mathscr{B}_t$ where x_t is the solution of (4.3). Then, the temporal evolution of $\psi_t(\cdot)$ is

[1]It is also assumed that $\gamma_t(\cdot)$ is a Gaussian map.

described by

$$d\psi_t(s) = s^T\left(A_t\frac{\partial\psi_t(s)}{\partial s} + B_t u_t\right)dt + \frac{1}{2}\,s^T D_t D_t^T s\,dt + \int_{\mathcal{A}}\big(\ln\beta_t(r,s) - \ln\beta_t(r,0)\big)N(dt\times dr) - \int_{\mathcal{A}}\big(\beta_t(r,s) - \beta_t(r,0)\big)drdt \tag{4.6}$$

where $\beta_t(\cdot,\cdot)$ is defined as

$$\beta_t(r,s) = \exp\{-\psi_t(j\omega)\}\,\mathrm{E}\left[\exp\left(j\omega^T x_t\right)\lambda_t(r,x_t)|\mathscr{B}_t\right]\big|_{j\omega=s}. \tag{4.7}$$

Moreover, if the Fourier transform of $\lambda_t(r,\cdot)$,

$$\Lambda_t(r,j\omega) = \int_{\mathbb{R}^n}\lambda_t(r,x)\exp\left(-j\omega^T x\right)dx$$

exists, $\beta_t(\cdot,\cdot)$ can be expressed as

$$\beta_t(r,s) = \frac{1}{(2\pi)^n}\int_{\mathbb{R}^n}\Lambda_t(r,j\nu)\exp\big\{\psi_t(j\nu+s) - \psi_t(s)\big\}d\nu. \tag{4.8}$$

Proof. See Section 4.5.1. □

The temporal evolution of the cumulants is described by a (generally infinite) set of nonlinear stochastic differential equations driven by the space-time point process $N(\mathcal{T}\times\mathcal{S})$. This set of equations can be derived from (4.6) by matching the coefficients of corresponding $s_{i_1}s_{i_2}\cdots s_{i_j}$ on the two sides of (4.6). We usually suppose that the first few cumulants approximate $p_{x_t}(x|\mathscr{B}_t)$ within an acceptable

precision. This means that the infinite set of equations can be approximated by a finite-dimensional one.

Regarding this approach, two issues should be addressed. First, we need to compute $\beta_t(\cdot,\cdot)$ in terms of the cumulants via equations (4.7) or (4.8) and expansion (4.5), which is not straightforward for an arbitrary number of cumulants. Second, when we truncate (4.5) to a finite number of terms, the corresponding approximation for $p_{x_t}(x|\mathscr{B}_t)$ might not be a valid probability density function, i.e., it might be negative for some x. When we limit the expansion (4.5) to the first and second order terms (Gaussian approximation), these difficulties are avoided. In this case, $\beta_t(\cdot,\cdot)$ can be easily calculated (when $\gamma_t(\cdot)$ is Gaussian) and the truncated expansion leads to a valid probability density.

We have used the method above to approximate $p_{x_t}(x|\mathscr{B}_t)$ with a Gaussian probability density. It is shown in Section 4.5.2 that the mean $\tilde{x}_t$ and the covariance matrix $\tilde{\Sigma}_t$ of this Gaussian approximation are the solutions of the stochastic differential equations

$$\begin{aligned} d\tilde{x}_t &= A_t\tilde{x}_t dt + B_t u_t dt + \int_{\mathcal{A}} \tilde{M}_t\left(r - C_t\tilde{x}_t\right) N\left(dt \times dr\right) - \mu_t h_t\big(\tilde{x}_t, \tilde{\Sigma}_t\big)dt \\ d\tilde{\Sigma}_t &= A_t\tilde{\Sigma}_t dt + \tilde{\Sigma}_t A_t^T dt + D_t D_t^T dt - \tilde{M}_t C_t \tilde{\Sigma}_t dN_t + \mu_t H_t\big(\tilde{x}_t, \tilde{\Sigma}_t\big)dt \end{aligned} \tag{4.9}$$

with the initial state $\tilde{x}_0 = \bar{x}_0$ and $\tilde{\Sigma}_0 = \bar{\Sigma}_0$. Here, we have $N_t = N\left([0,t)\times\mathcal{A}\right)$ and $\tilde{M}_t = \Gamma_t\big(\tilde{\Sigma}_t\big)$, where $\Gamma_t(\cdot) : \mathbb{R}\times\mathbb{R}^{n\times n} \to \mathbb{R}^{n\times m}$ is defined as

$$\Gamma_t(\Sigma) = \Sigma C_t^T\left(C_t\Sigma C_t^T + R_t\right)^{-1}.$$

Also, $h_t(\cdot,\cdot) : \mathbb{R} \times \mathbb{R}^n \times \mathbb{R}^{n\times n} \to \mathbb{R}^n$ and $H_t(\cdot,\cdot) : \mathbb{R} \times \mathbb{R}^n \times \mathbb{R}^{n\times n} \to \mathbb{R}^{n\times n}$ are given by

$$h_t(x,\Sigma) = \int_{\mathcal{A}} \Gamma_t(\Sigma)(r - C_t x)\,\Phi_m\left(r; C_t x, C_t\Sigma C_t^T + R_t\right) dr$$

$$H_t(x,\Sigma) = \int_{\mathcal{A}} \Gamma_t(\Sigma)\left(C_t\Sigma C_t^T + R_t - (r - C_t x)(r - C_t x)^T\right)\Gamma_t^T(\Sigma) \cdot \Phi_m\left(r; C_t x, C_t\Sigma C_t^T + R_t\right) dr.$$

Note that $\tilde{x}_t$ and $\tilde{\Sigma}_t$ are approximations of $\hat{x}_t$ and Σ_t, not their exact values.

Remark 4.3.1. The definition of $h(\cdot,\cdot)$ and $H(\cdot,\cdot)$ imply that as $\mathcal{A} \to \mathbb{R}^m$, $h(\cdot,\cdot) \to 0$ and $H(\cdot,\cdot) \to 0$, and as a consequence, the approximate estimator (4.9) tends to the exact estimator (3.9). In this sense, we can say that (4.9) is an asymptotically optimal estimator.

4.4 Control Problem

We exploit the results of the previous section in solving the control problem as stated in Theorem 4.4.1 below. Before presenting this result, we fix notation. Let $\Sigma = [\sigma_{ij}]$ denote a symmetric $n \times n$ matrix and $f(\Sigma)$ be a scalar function of Σ. Assume that the partial derivatives of $f(\Sigma)$ with respect to the elements of Σ exist. We denote by $\partial f(\Sigma)/\partial\Sigma$ a $n \times n$ symmetric matrix $F(\Sigma) = [F_{ij}(\Sigma)]$ such that $F_{ii} = \partial f/\partial\sigma_{ii}$ and $F_{ij} = (1/2)\,\partial f/\partial\sigma_{ij}$ for $i \neq j$. Let $g_t(x,\Sigma)$ be a scalar function of $x \in \mathbb{R}^n$ and $n \times n$ symmetric matrix Σ. Assume that the partial derivatives of $g_t(x,\Sigma)$ with

respect to x and Σ exist. Define the linear operator $\mathcal{L}_t\{\cdot\}$ as

$$\begin{aligned}\mathcal{L}_t\{g_t(x,\Sigma)\} = \int_{\mathcal{A}} \Big(g_t\big(x + \Gamma_t(\Sigma)(r - C_t x), \Sigma - \Gamma_t(\Sigma) C_t \Sigma\big) - g_t(x,\Sigma) \Big) \\ \cdot \Phi_m\left(r; C_t x, C_t \Sigma C_t^T + R_t\right) dr \\ - \left(\partial g_t(x,\Sigma)/\partial x\right)^T h_t(x,\Sigma) + \operatorname{tr}\left\{\left(\partial g_t(x,\Sigma)/\partial \Sigma\right) H_t(x,\Sigma)\right\}.\end{aligned} \tag{4.10}$$

Finally, we use $\|\cdot\|_{P_t}^2$ to denote $(\cdot)^T P_t(\cdot)$.

Theorem 4.4.1. Let $x \in \mathbb{R}^n$ and Σ be a $n \times n$ symmetric matrix. Suppose that $g_t(x,\Sigma)$, $t \in [0,T]$ is the solution of the partial differential equation

$$\begin{aligned}-\frac{\partial}{\partial t} g_t(x,\Sigma) = \left(\frac{\partial}{\partial x} g_t(x,\Sigma)\right)^T A_t x - \frac{1}{4}\left(\frac{\partial}{\partial x} g_t(x,\Sigma)\right)^T B_t P_t^{-1} B_t^T \left(\frac{\partial}{\partial x} g_t(x,\Sigma)\right) \\ + \operatorname{tr}\left\{\left(\frac{\partial}{\partial \Sigma} g_t(x,\Sigma)\right)\left(A_t \Sigma + \Sigma A_t^T + D_t D_t^T\right) + Q_t \Sigma\right\} \\ + x^T Q_t x + \mu_t \mathcal{L}_t\{g_t(x,\Sigma)\}\end{aligned} \tag{4.11}$$

with the boundary condition $g_T(x,\Sigma) = x^T S x$. Then, the cost functional (4.4) can be expressed as

$$J = g_0\big(\tilde{x}_0, \tilde{\Sigma}_0\big) + \mathrm{E}\left[\int_0^T \delta_t dt\right] + \mathrm{E}\left[\int_0^T \left\| u_t + \frac{1}{2} P_t^{-1} B_t^T \left(\frac{\partial}{\partial \tilde{x}_t} g_t\big(\tilde{x}_t, \tilde{\Sigma}_t\big)\right)\right\|_{P_t}^2 dt\right] \tag{4.12}$$

where

$$\delta_t = \int_{\mathbb{R}^n} \Big\{ x^T Q_t x + \int_{\mathcal{A}} \Big(g_t\big(\tilde{x}_t + \tilde{M}_t \left(r - C_t \tilde{x}_t\right), \tilde{\Sigma}_t - \tilde{M}_t C_t \tilde{\Sigma}_t\big) - g_t\big(\tilde{x}_t, \tilde{\Sigma}_t\big)\Big) \lambda_t\left(r, x\right) dr \Big\} \Big(p_{x_t}\left(x|\mathscr{B}_t\right) - \tilde{p}_{x_t}\left(x|\mathscr{B}_t\right)\Big) dx \quad (4.13)$$

is the error term resulting from replacing the posterior density $p_{x_t}\left(x|\mathscr{B}_t\right)$ by its Gaussian approximation $\tilde{p}_{x_t}\left(x|\mathscr{B}_t\right)$.

Proof. See Section 4.5.3. □

The first term on the right side of (4.12) does not depend on u_t and so is not involved in the minimization. While the hard-to-compute error term δ_t in (4.12) depends on u_t, it is supposed to be small. Therefore, in minimizing (4.12), we ignore the term involving δ_t and only minimize the third term. We note that the minimum of the third term is zero which is achieved when u_t is given by

$$u_t^* = -\frac{1}{2} P_t^{-1} B_t^T \left(\frac{\partial}{\partial \tilde{x}_t} g_t\big(\tilde{x}_t, \tilde{\Sigma}_t\big) \right). \quad (4.14)$$

Then the cost associated with u_t^* will be

$$J^* = g_0\big(\bar{x}_0, \bar{\Sigma}_0\big) + \int_0^T \mathrm{E}\left[\delta_t|_{u_t = u_t^*}\right] dt.$$

When $\mathcal{A} = \mathbb{R}^m$, a simple solution can be obtained for (4.11). This solution which is stated by the following theorem confirms that the control (4.14) is consistent with that obtained for $\mathcal{A} = \mathbb{R}^m$ by Rhodes and Snyder [20] in Theorem 3.3.3. This

indicates that the suboptimal control (4.14) tends to the optimal control as $\mathcal{A} \to \mathbb{R}^m$.

Theorem 4.4.2. When $\mathcal{A} = \mathbb{R}^m$, the solution of the partial differential equation (4.11) with the boundary condition $g_T(x, \Sigma) = x^T S x$ can be expressed as

$$g_t(x, \Sigma) = x^T K_t x + f_t(\Sigma) \tag{4.15}$$

where K_t is the backward solution to the Riccati equation

$$\dot{K}_t = -K_t A_t - A_t^T K_t + K_t B_t P_t^{-1} B_t^T K_t - Q_t \tag{4.16}$$

with $K_T = S$, and $f_t(\Sigma)$ is the solution of the partial differential equation

$$\begin{aligned} -\frac{\partial}{\partial t} f_t(\Sigma) = \operatorname{tr}\left\{ \left(\frac{\partial}{\partial \Sigma} f_t(\Sigma) \right) \left(A_t \Sigma + \Sigma A_t^T + D_t D_t^T \right) + Q_t \Sigma \right\} \\ + \mu_t \big(f_t(\Sigma - \Gamma_t(\Sigma) C_t \Sigma) - f_t(\Sigma) \big) + \mu_t \operatorname{tr}\left\{ \Gamma_t(\Sigma) C_t \Sigma K_t \right\} \end{aligned} \tag{4.17}$$

with the boundary condition $f_T(\Sigma) = 0$.

Proof. See Section 4.5.4. □

We observe from (4.14) and (4.15) that when $\mathcal{A} = \mathbb{R}^m$, the optimal control is given by

$$u_t^* = -P_t^{-1} B_t^T K_t \hat{x}_t \tag{4.18}$$

with the optimal cost

$$J^* = \bar{x}_0^T K_0 \bar{x}_0 + f_0\left(\bar{\Sigma}_0\right). \tag{4.19}$$

While the optimal control (4.18) has been obtained by Rhodes and Snyder [20], the value of the corresponding optimal cost (4.19) is newly obtained here.

4.5 Proof of the Theorems

4.5.1 Proof of Theorem 4.3.1

The Fourier transform of (3.7) is given by [20] as

$$\begin{aligned} d\phi_t(j\omega) = \mathrm{E}\left[\exp\left(j\omega^T x_t\right)\left(j\omega^T\left(A_t x_t + B_t u_t\right) - \frac{1}{2}\omega^T D_t D_t^T \omega\right)\Big|\mathscr{B}_t\right] dt \\ + \int_{\mathcal{A}} \mathrm{E}\left[\exp\left(j\omega^T x_t\right)\left(\lambda_t(r, x_t)\,\hat{\lambda}_t^{-1}(r) - 1\right)\Big|\mathscr{B}_t\right] N(dt \times dr) \\ - \int_{\mathcal{A}} \mathrm{E}\left[\exp\left(j\omega^T x_t\right)\left(\lambda_t(r, x_t) - \hat{\lambda}_t(r)\right)\Big|\mathscr{B}_t\right] dr dt. \end{aligned} \tag{4.20}$$

Let $t_0 = 0$ and $t_1 < t_2 < t_3 < \cdots$ be the occurrence times of the space-time process $N(\mathcal{T} \times \mathcal{S})$. During the interval $(t_k, t_{k+1}),\ k = 0, 1, 2, \ldots$, the first integral on the right side of (4.20) is zero, thus we can write

$$\begin{aligned} d\exp\{\psi_t(j\omega)\} = \mathrm{E}\left[\exp\left(j\omega^T x_t\right)\left(j\omega^T\left(A_t x_t + B_t u_t\right) - \frac{1}{2}\omega^T D_t D_t^T \omega\right)\Big|\mathscr{B}_t\right] dt \\ - \int_{\mathcal{A}} \mathrm{E}\left[\exp\left(j\omega^T x_t\right)\left(\lambda_t(r, x_t) - \hat{\lambda}_t(r)\right)\Big|\mathscr{B}_t\right] dr dt. \end{aligned}$$

Using the identity

$$\begin{aligned} \mathrm{E}\left[x_t \exp\left(j\omega^T x_t\right)|\mathscr{B}_t\right] &= \frac{\partial}{\partial j\omega}\mathrm{E}\left[\exp\left(j\omega^T x_t\right)|\mathscr{B}_t\right] \\ &= \frac{\partial \psi_t(j\omega)}{\partial j\omega}\exp\{\psi_t(j\omega)\} \end{aligned}$$

we rewrite this equation as

$$\exp\{\psi_t(j\omega)\}\, d\psi_t(j\omega) = \exp\{\psi_t(j\omega)\}\left[j\omega^T\left(A_t\frac{\partial\psi_t(j\omega)}{\partial j\omega} + B_t u_t\right) - \frac{1}{2}\,\omega^T D_t D_t^T \omega\right] dt$$
$$- \int_{\mathcal{A}} \mathrm{E}\left[\exp\left(j\omega^T x_t\right)\left(\lambda_t\left(r, x_t\right) - \hat{\lambda}_t\left(r\right)\right)\middle|\mathscr{B}_t\right] dr dt.$$

Multiplying both sides of this equation by $\exp\{-\psi_t(j\omega)\}$ and substituting $\beta_t\left(r, j\omega\right)$ from (4.7) into the result, we obtain

$$d\psi_t\left(j\omega\right) = j\omega^T\left(A_t\frac{\partial\psi_t(j\omega)}{\partial j\omega} + B_t u_t\right) dt$$
$$- \frac{1}{2}\,\omega^T D_t D_t^T \omega\, dt - \int_{\mathcal{A}}\left(\beta_t\left(r, j\omega\right) - \beta_t\left(r, 0\right)\right) dr dt. \tag{4.21}$$

The discontinuity at $t = t_k$ is treated as follows. Let r_k be the spatial component of the event occurring at t_k. Then, from (4.20) we get

$$\phi_{t_k^+}\left(j\omega\right) - \phi_{t_k^-}\left(j\omega\right) = \mathrm{E}\left[\exp\left(j\omega^T x_{t_k^-}\right)\left(\lambda_{t_k^-}\left(r_k, x_{t_k^-}\right)\hat{\lambda}_{t_k^-}^{-1}\left(r_k\right) - 1\right)\middle|\mathscr{B}_{t_k^-}\right]$$

which can be simplified as

$$\phi_{t_k^+}\left(j\omega\right) = \mathrm{E}\left[\exp\left(j\omega^T x_{t_k^-}\right)\lambda_{t_k^-}\left(r_k, x_{t_k^-}\right)\middle|\mathscr{B}_{t_k^-}\right]\hat{\lambda}_{t_k^-}^{-1}\left(r_k\right).$$

Multiplying both sides of this equality by $\exp\left\{-\psi_{t_k^-}\left(j\omega\right)\right\}$ and taking logarithm, we

obtain

$$\begin{aligned}
&\psi_{t_k^+}(j\omega) - \psi_{t_k^-}(j\omega) \\
&\qquad = \ln\Big(\exp\big\{-\psi_{t_k^-}(j\omega)\big\}\mathrm{E}\left[\exp\big(j\omega^T x_{t_k^-}\big)\lambda_{t_k^-}\big(r_k, x_{t_k^-}\big)\big|\mathscr{B}_{t_k^-}\right]\Big) - \ln\hat{\lambda}_{t_k^-}(r_k) \\
&\qquad = \ln\beta_{t_k^-}(r_k, j\omega) - \ln\beta_{t_k^-}(r_k, 0)\,.
\end{aligned}$$

Combining this result with (4.21) and replacing $j\omega$ by s, we obtain (4.6).

From the definition of $\beta_t(r, j\omega)$ in (4.7), we have

$$\beta_t(r, j\omega) = \exp\{-\psi_t(j\omega)\}\int_{\mathbb{R}^n} p_{x_t}(x|\mathscr{B}_t)\exp\left(j\omega^T x\right)\lambda_t(r, x)\,dx. \tag{4.22}$$

Note that $p_{x_t}(x|\mathscr{B}_t)$ is the Fourier transform of $\exp\{\psi_t(j\omega)\}$, and so we can write

$$p_{x_t}(x|\mathscr{B}_t) = \frac{1}{(2\pi)^n}\int_{\mathbb{R}^n}\exp\{\psi_t(j\nu)\}\exp\left(-j\nu^T x\right)d\nu.$$

Upon substituting this expression into (4.22) and interchanging the order of integration[1], we obtain

$$\beta_t(r, j\omega) = \frac{1}{(2\pi)^n}\int_{\mathbb{R}^n}\exp\big\{\psi_t(j\nu) - \psi_t(j\omega)\big\}\int_{\mathbb{R}^n}\lambda_t(r, x)\exp\Big\{-j\left(\nu-\omega\right)^T x\Big\}\,dx d\nu.$$

Replacing the second integral above by $\Lambda_t(r, j\nu - j\omega)$ and changing the variable of

[1]This interchange is permissible since for any fixed t, the integrand is continuous in x and ν.

integration ν with $\nu + \omega$, we get

$$\beta_t(r, j\omega) = \frac{1}{(2\pi)^n} \int_{\mathbb{R}^n} \Lambda_t(r, j\nu) \exp\{\psi_t(j\nu + j\omega) - \psi_t(j\omega)\} d\nu.$$

Finally, we obtain (4.8) upon replacing $j\omega$ with s.

4.5.2 Derivation of (4.9)

We first state a technical lemma from [43] which will be used later in deriving (4.9). For sake of completeness, we repeat below the proof from [43].

Lemma 4.5.1. Let $z_k, \bar{z}_k \in \mathbb{R}^k$, $z_l \in \mathbb{R}^l$, and Θ_k and Θ_l be respectively $k \times k$ and $l \times l$ positive definite matrices. Assume that G is any $l \times k$ matrix. Then, we have

$$\int_{\mathbb{R}^k} \Phi_k(z_k; \bar{z}_k, \Theta_k) \Phi_l(z_l; Gz_k, \Theta_l) dz_k = \Phi_l\left(z_l; G\bar{z}_k, \Theta_l + G\Theta_k G^T\right). \tag{4.23}$$

Proof. Denoting the Fourier transform of the left side of (4.23) by $F_l(\omega_l)$, we can write

$$\begin{aligned} F_l(\omega_l) &= \int_{\mathbb{R}^k} \Phi_k(z_k; \bar{z}_k, \Theta_k) \exp\left(j\omega_l^T G z_k - \tfrac{1}{2}\omega_l^T \Theta_l \omega_l\right) dz_k \\ &= \exp\left(j\omega_l^T G\bar{z}_k - \tfrac{1}{2}\omega_l^T G\Theta_k G^T \omega_l\right) \exp\left(-\tfrac{1}{2}\omega_l^T \Theta_l \omega_l\right) \\ &= \exp\left(j\omega_l^T G\bar{z}_k - \tfrac{1}{2}\omega_l^T \left(\Theta_l + G\Theta_k G^T\right)\omega_l\right). \end{aligned}$$

Taking inverse Fourier transform of the expression above, we get the right side of (4.23). □

The probability density function associated with the truncated expansion $\tilde{\psi}_t(s) = s^T\tilde{x}_t + \frac{1}{2}\, s^T\tilde{\Sigma}_t s$ is Gaussian with mean $\tilde{x}_t$ and covariance matrix $\tilde{\Sigma}_t$. With this approximate probability density function and with $\lambda_t\left(r, x_t\right) = \mu_t \Phi_m\left(r; C_t x_t, R_t\right)$, the approximation of $\beta_t\left(\cdot,\cdot\right)$ is given by

$$\tilde{\beta}_t\left(r,s\right) = \exp\left\{-\tilde{\psi}_t\left(s\right)\right\} \int_{\mathbb{R}^n} \Phi_n\big(x; \tilde{x}_t, \tilde{\Sigma}_t\big) \exp\left(s^T x\right) \mu_t \Phi_m\left(r; C_t x, R_t\right) dx.$$

A simple calculation yields that

$$\exp\left\{-\tilde{\psi}_t\left(s\right)\right\} \Phi_n\big(x; \tilde{x}_t, \tilde{\Sigma}_t\big) \exp\left(s^T x\right) = \Phi_n\big(x; \tilde{x}_t + \tilde{\Sigma}_t s, \tilde{\Sigma}_t\big).$$

Then, using Lemma 4.5.1, we get

$$\tilde{\beta}_t\left(r,s\right) = \mu_t \Phi_m\left(r; C_t\tilde{x}_t + C_t\tilde{\Sigma}_t s, C_t\tilde{\Sigma}_t C_t^T + R_t\right)$$

which leads to

$$\begin{aligned}\ln\tilde{\beta}_t\left(r,s\right) - \ln\tilde{\beta}_t\left(r,0\right) = {} & s^T\tilde{\Sigma}_t C_t^T \left(C_t\tilde{\Sigma}_t C_t^T + R_t\right)^{-1}\left(r - C_t\tilde{x}_t\right) \\ & - \frac{1}{2}\, s^T\tilde{\Sigma}_t C_t^T \left(C_t\tilde{\Sigma}_t C_t^T + R_t\right)^{-1} C_t\tilde{\Sigma}_t s \end{aligned} \tag{4.24}$$

and

$$\begin{aligned}\tilde{\beta}_t(r,s) - \tilde{\beta}_t(r,0) = \mu_t \Phi_m\left(r; C_t\tilde{x}_t, C_t\tilde{\Sigma}_t C_t^T + R_t\right) \\ \cdot \Big\{ s^T\tilde{\Sigma}_t C_t^T \left(C_t\tilde{\Sigma}_t C_t^T + R_t\right)^{-1}(r - C_t\tilde{x}_t) \\ + \frac{1}{2}\left[s^T\tilde{\Sigma}_t C_t^T \left(C_t\tilde{\Sigma}_t C_t^T + R_t\right)^{-1}(r - C_t\tilde{x}_t)\right]^2 \\ - \frac{1}{2} s^T\tilde{\Sigma}_t C_t^T \left(C_t\tilde{\Sigma}_t C_t^T + R_t\right)^{-1} C_t\tilde{\Sigma}_t s + O\left(\|s\|^3\right)\Big\}. \quad (4.25)\end{aligned}$$

We combine (4.24), (4.25), and (4.6) and match the coefficients of $s^T(\cdot)$ and $s^T(\cdot)s$ from both sides to obtain (4.9).

4.5.3 Proof of Theorem 4.4.1

Our proof consists of the following four steps.

Step I: Using the properties of conditional expectation, it is easy to show that

$$\mathrm{E}\left[x_t^T Q_t x_t\right] = \mathrm{E}\Big[\hat{x}_t^T Q_t \hat{x}_t + \mathrm{tr}\left\{Q_t\Sigma_t\right\}\Big].$$

Then the cost functional (4.4) can be expressed as

$$J = \mathrm{E}\left[\int_0^T \left(\tilde{x}_t^T Q_t \tilde{x}_t + \mathrm{tr}\big\{Q_t\tilde{\Sigma}_t\big\} + u_t^T P_t u_t + \delta_t^1\right) dt + x_T^T S x_T\right] \quad (4.26)$$

where δ_t^1 is defined as

$$\delta_t^1 = \mathrm{tr}\left\{Q_t\left(\hat{x}_t\hat{x}_t^T - \tilde{x}_t\tilde{x}_t^T + \Sigma_t - \tilde{\Sigma}_t\right)\right\}.$$

Step II: For $t \geqslant 0$ and for any positive ϵ, $\Delta N_t \triangleq N_{t+\epsilon} - N_t$ is a conditionally Poisson random variable with the stochastic rate

$$\bar{\lambda}_t^{\epsilon} = \int_t^{t+\epsilon} \int_{\mathcal{A}} \lambda_\tau \left(r, x_\tau\right) dr d\tau.$$

Thus, using the law of total probability, we can write

$$\begin{aligned} \Pr\left\{\Delta N_t = 1 \middle| \mathscr{B}_t\right\} &= \mathrm{E}\left[\Pr\left\{\Delta N_t = 1 | \bar{\lambda}_t^{\epsilon}\right\} \middle| \mathscr{B}_t\right] \\ &= \mathrm{E}\left[\bar{\lambda}_t^{\epsilon} \exp\left(-\bar{\lambda}_t^{\epsilon}\right) \middle| \mathscr{B}_t\right] \\ &= \epsilon q_t + O\left(\epsilon^2\right) \end{aligned} \tag{4.27}$$

where q_t is defined as

$$q_t = \int_{\mathcal{A}} \mathrm{E}\left[\lambda_t \left(r, x_t\right) | \mathscr{B}_t\right] dr. \tag{4.28}$$

In a similar manner, we can show that

$$\begin{aligned} \Pr\left\{\Delta N_t = 0 \middle| \mathscr{B}_t\right\} &= 1 - \epsilon q_t + O\left(\epsilon^2\right) \\ \Pr\left\{\Delta N_t \geqslant 2 \middle| \mathscr{B}_t\right\} &= O\left(\epsilon^2\right). \end{aligned} \tag{4.29}$$

Let the random vector $R \in \mathbb{R}^m$ denote the location of a single event occurring during $[t, t+\epsilon)$. We show that

$$p_{R}\left(r | \Delta N_t = 1, \mathscr{B}_t\right) = \frac{\mathrm{E}\left[\lambda_t \left(r, x_t\right) | \mathscr{B}_t\right]}{q_t} I_{\mathcal{A}}\left(r\right) + O\left(\epsilon\right) \tag{4.30}$$

where $I_{\mathcal{A}}\left(r\right) = 1$ if $r \in \mathcal{A}$ and $I_{\mathcal{A}}\left(r\right) = 0$ otherwise. For this purpose, let $\mathcal{D}\left(r\right) \subset \mathcal{A}$

denote a m-dimensional cube with side length Δr which is centered at $r \in \mathcal{A}$. Defining $\mathscr{T} = [t, t+\epsilon)$ and using Bayes' rule, we can write

$$\begin{aligned} p_R\left(r|\Delta N_t = 1, \mathscr{B}_t\right) &= \lim_{\Delta r \to 0} \Delta r^{-2} \Pr\left\{R \in \mathcal{D}\left(r\right)|\Delta N_t = 1, \mathscr{B}_t\right\} \\ &= \lim_{\Delta r \to 0} \Delta r^{-2} \Pr\left\{N\left(\mathscr{T} \times \mathcal{D}\left(r\right)\right) = 1|\Delta N_t = 1, \mathscr{B}_t\right\} \\ &= \lim_{\Delta r \to 0} \frac{1}{\Delta r^2} \cdot \frac{\Pr\left\{N\left(\mathscr{T} \times \mathcal{D}\left(r\right)\right) = 1, N\left(\mathscr{T} \times \mathcal{A}\right) = 1|\mathscr{B}_t\right\}}{\Pr\left\{\Delta N_t = 1|\mathscr{B}_t\right\}}. \end{aligned} \tag{4.31}$$

Note that the event of $N\left(\mathscr{T} \times \mathcal{D}\left(r\right)\right) = 1$ and $N\left(\mathscr{T} \times \mathcal{A}\right) = 1$ is equivalent to the event of $N\left(\mathscr{T} \times \mathcal{D}\left(r\right)\right) = 1$ and $N\left(\mathscr{T} \times \left(\mathcal{A} - \mathcal{D}\left(r\right)\right)\right) = 0$. Therefore, defining $\mathscr{X}_t = \left\{x_\tau | \, \tau \in \mathscr{T}\right\}$ and using the law of total probability and properties of a space-time point process, we get

$$\begin{aligned} &\Pr\left\{N\left(\mathscr{T} \times \mathcal{D}\left(r\right)\right) = 1, N\left(\mathscr{T} \times \mathcal{A}\right) = 1|\mathscr{B}_t\right\} \\ &\quad = \mathrm{E}\Big[\Pr\left\{N\left(\mathscr{T} \times \mathcal{D}\left(r\right)\right) = 1, N\left(\mathscr{T} \times \left(\mathcal{A} - \mathcal{D}\left(r\right)\right)\right) = 0 \,|\, \mathscr{X}_t, \mathscr{B}_t\right\} \Big| \mathscr{B}_t\Big] \\ &\quad = \mathrm{E}\Big[\Pr\left\{N\left(\mathscr{T} \times \mathcal{D}\left(r\right)\right) = 1|\mathscr{X}_t\right\} \Pr\left\{N\left(\mathscr{T} \times \left(\mathcal{A} - \mathcal{D}\left(r\right)\right)\right) = 0 \,|\, \mathscr{X}_t\right\} \Big| \mathscr{B}_t\Big] \\ &\quad = \mathrm{E}\left[\int_{\mathscr{T} \times \mathcal{D}(r)} \lambda_\tau\left(s, x_\tau\right) d\tau ds \left(1 - O\left(\epsilon \Delta r^2\right)\right) \Big| \, \mathscr{B}_t\right] \\ &\quad = \epsilon \Delta r^2 \mathrm{E}\left[\lambda_t\left(r, x_t\right)|\mathscr{B}_t\right] + O\left(\epsilon \Delta r^3\right) + O\left(\epsilon^2 \Delta r^2\right). \end{aligned} \tag{4.32}$$

Substituting (4.27) and (4.32) into (4.31), we obtain (4.30).

Assume that $\tilde{p}_{x_t}\left(x|\mathscr{B}_t\right)$ is the Gaussian approximation of $p_{x_t}\left(x|\mathscr{B}_t\right)$. Then,

using Lemma 4.5.1, we can write

$$\begin{aligned}
\mathrm{E}\left[\lambda_t\left(r, x_t\right)|\mathscr{B}_t\right] &= \int_{\mathbb{R}^n} \tilde{p}_{x_t}\left(x|\mathscr{B}_t\right)\lambda_t\left(r,x\right)dx \\
&\quad + \int_{\mathbb{R}^n}\Big(p_{x_t}\left(x|\mathscr{B}_t\right) - \tilde{p}_{x_t}\left(x|\mathscr{B}_t\right)\Big)\lambda_t\left(r,x\right)dx \\
&= \mu_t \Phi_m\big(r; C_t\tilde{x}_t, C_t\tilde{\Sigma}_t C_t^T + R_t\big) \\
&\quad + \int_{\mathbb{R}^n}\Big(p_{x_t}\left(x|\mathscr{B}_t\right) - \tilde{p}_{x_t}\left(x|\mathscr{B}_t\right)\Big)\lambda_t\left(r,x\right)dx. \qquad (4.33)
\end{aligned}$$

Step III: Let $g_t\left(x,\Sigma\right)$ be a scalar function of $x \in \mathbb{R}^n$ and $n \times n$ symmetric matrix Σ. Assume that the partial derivatives of $g_t\left(x,\Sigma\right)$ with respect to t, x, and Σ exist. Using the law of total probability we can write

$$\mathrm{E}\left[g_{t+\epsilon}\big(\tilde{x}_{t+\epsilon}, \tilde{\Sigma}_{t+\epsilon}\big)\big|\mathscr{B}_t\right] = \sum_{k=0}^{\infty}\mathrm{E}\left[g_{t+\epsilon}\big(\tilde{x}_{t+\epsilon}, \tilde{\Sigma}_{t+\epsilon}\big)\big|\mathscr{B}_t, \Delta N_t = k\right]\Pr\left\{\Delta N_t = k|\mathscr{B}_t\right\}.$$

Replacing $\Pr\left\{\Delta N_t = k|\mathscr{B}_t\right\}$ from (4.27) and (4.29) into this expression, and using the law of total probability again, we find

$$\begin{aligned}
\mathrm{E}\left[g_{t+\epsilon}\big(\tilde{x}_{t+\epsilon}, \tilde{\Sigma}_{t+\epsilon}\big)\big|\mathscr{B}_t\right] &= \mathrm{E}\left[g_{t+\epsilon}\big(\tilde{x}_{t+\epsilon}, \tilde{\Sigma}_{t+\epsilon}\big)\big|\mathscr{B}_t, \Delta N_t = 0\right](1-\epsilon q_t) \\
&\quad + \mathrm{E}\left[\mathrm{E}\left[g_{t+\epsilon}\big(\tilde{x}_{t+\epsilon}, \tilde{\Sigma}_{t+\epsilon}\big)\big|\mathscr{B}_t, \Delta N_t = 1, R = r\right]\big|\mathscr{B}_t, \Delta N_t = 1\right]\epsilon q_t + O\left(\epsilon^2\right).
\end{aligned}$$

Substituting (4.28) and (4.30) into this equality and rearranging terms, we obtain

$$\begin{aligned}\mathrm{E}\left[g_{t+\epsilon}\big(\tilde{x}_{t+\epsilon},\tilde{\Sigma}_{t+\epsilon}\big)\big|\mathscr{B}_t\right] &= \mathrm{E}\left[g_{t+\epsilon}\big(\tilde{x}_{t+\epsilon},\tilde{\Sigma}_{t+\epsilon}\big)\big|\mathscr{B}_t,\Delta N_t=0\right]\\ &+\epsilon\int_{\mathcal{A}}\Big(\mathrm{E}\left[g_{t+\epsilon}\big(\tilde{x}_{t+\epsilon},\tilde{\Sigma}_{t+\epsilon}\big)\big|\mathscr{B}_t,\Delta N_t=1,R=r\right]\\ &-\mathrm{E}\left[g_{t+\epsilon}\big(\tilde{x}_{t+\epsilon},\tilde{\Sigma}_{t+\epsilon}\big)\big|\mathscr{B}_t,\Delta N_t=0\right]\Big)\mathrm{E}\left[\lambda_t\left(r,x_t\right)|\mathscr{B}_t\right]dr+O\left(\epsilon^2\right). \quad (4.34)\end{aligned}$$

Conditioned on $\mathscr{B}_t$ and $\Delta N_t = 0$, (4.9) can be solved during $[t, t+\epsilon)$ to obtain

$$\begin{aligned}\tilde{x}_{t+\epsilon} &= \tilde{x}_t+\epsilon A_t\tilde{x}_t+\epsilon B_t u_t-\epsilon\mu_t h_t\big(\tilde{x}_t,\tilde{\Sigma}_t\big)+O\left(\epsilon^2\right)\\ \tilde{\Sigma}_{t+\epsilon} &= \tilde{\Sigma}_t+\epsilon A_t\tilde{\Sigma}_t+\epsilon\tilde{\Sigma}_t A_t^T+\epsilon D_t D_t^T+\epsilon\mu_t H_t\big(\tilde{x}_t,\tilde{\Sigma}_t\big)+O\left(\epsilon^2\right).\end{aligned} \quad (4.35)$$

Also, conditioned on $\mathscr{B}_t$, $\Delta N_t = 1$, and $R = r$, we can write

$$\begin{aligned}\tilde{x}_{t+\epsilon} &= \tilde{x}_t+\tilde{M}_t\left(r-C_t\tilde{x}_t\right)+O\left(\epsilon\right)\\ \tilde{\Sigma}_{t+\epsilon} &= \tilde{\Sigma}_t-\tilde{M}_t C_t\tilde{\Sigma}_t+O\left(\epsilon\right).\end{aligned} \quad (4.36)$$

Substituting (4.35) and (4.36) into (4.34), and linearizing with respect to ϵ, we get

$$\begin{aligned}\mathrm{E}\left[g_t\big(\tilde{x}_{t+\epsilon},\tilde{\Sigma}_{t+\epsilon}\big)\big|\mathscr{B}_t\right] &= g_t\big(\tilde{x}_t,\tilde{\Sigma}_t\big)\\ &+\epsilon\frac{\partial}{\partial t}g_t\big(\tilde{x}_t,\tilde{\Sigma}_t\big)+\epsilon\left(\frac{\partial}{\partial\tilde{x}_t}g_t\big(\tilde{x}_t,\tilde{\Sigma}_t\big)\right)^T\left(A_t\tilde{x}_t+B_t u_t-\mu_t h_t\big(\tilde{x}_t,\tilde{\Sigma}_t\big)\right)\\ &+\epsilon\,\mathrm{tr}\left\{\left(\frac{\partial}{\partial\tilde{\Sigma}_t}g_t\big(\tilde{x}_t,\tilde{\Sigma}_t\big)\right)\left(A_t\tilde{\Sigma}_t+\tilde{\Sigma}_t A_t^T+D_t D_t^T+\mu_t H_t\big(\tilde{x}_t,\tilde{\Sigma}_t\big)\tilde{M}_t^T\right)\right\}\\ &+\epsilon\int_{\mathcal{A}}\left[g_t\left(\tilde{x}_t+\tilde{M}_t\left(r-C_t\tilde{x}_t\right),\tilde{\Sigma}_t-\tilde{M}_t C_t\tilde{\Sigma}_t\right)-g_t\big(\tilde{x}_t,\tilde{\Sigma}_t\big)\right]\mathrm{E}\left[\lambda_t\left(r,x_t\right)|\mathscr{B}_t\right]dr\\ &+O\left(\epsilon^2\right).\end{aligned}$$

We replace $\mathrm{E}\left[\lambda_t\left(r, x_t\right)|\mathscr{B}_t\right]$ from (4.33) into the expression above and use the linear operator $\mathcal{L}_t\{\cdot\}$ defined by (4.10) to obtain the simplified form

$$\begin{aligned}\mathrm{E}\left[g_t\big(\tilde{x}_{t+\epsilon}, \tilde{\Sigma}_{t+\epsilon}\big)\big|\mathscr{B}_t\right] &= g_t\big(\tilde{x}_t, \tilde{\Sigma}_t\big) + \epsilon\frac{\partial}{\partial t} g_t\big(\tilde{x}_t, \tilde{\Sigma}_t\big) \\ &\quad + \epsilon\left(\frac{\partial}{\partial \tilde{x}_t} g_t\big(\tilde{x}_t, \tilde{\Sigma}_t\big)\right)^T \left(A_t\tilde{x}_t + B_t u_t\right) \\ &\quad + \epsilon\,\mathrm{tr}\left\{\left(\frac{\partial}{\partial \tilde{\Sigma}_t} g_t\big(\tilde{x}_t, \tilde{\Sigma}_t\big)\right)\left(A_t\tilde{\Sigma}_t + \tilde{\Sigma}_t A_t^T + D_t D_t^T\right)\right\} \\ &\quad + \epsilon\mu_t\mathcal{L}_t\left\{g_t\big(\tilde{x}_t, \tilde{\Sigma}_t\big)\right\} + \epsilon\delta_t^2 + O\left(\epsilon^2\right) \qquad (4.37)\end{aligned}$$

where the error term δ_t^2 is defined as

$$\begin{aligned}\delta_t^2 = \int_{\mathbb{R}^n}\int_{\mathcal{A}} &\left[g_t\left(\tilde{x}_t + \tilde{M}_t\left(r - C_t\tilde{x}_t\right), \tilde{\Sigma}_t - \tilde{M}_t C_t\tilde{\Sigma}_t\right) - g_t\big(\tilde{x}_t, \tilde{\Sigma}_t\big)\right] \\ &\cdot\left(p_{x_t}\left(x|\mathscr{B}_t\right) - \tilde{p}_{x_t}\left(x|\mathscr{B}_t\right)\right)\lambda_t\left(r, x\right) dr dx.\end{aligned}$$

Define the nonlinear operator $\mathcal{K}_t\{\cdot\}$ as

$$\begin{aligned}\mathcal{K}_t\left\{g_t\left(x, \Sigma\right)\right\} &= \frac{\partial}{\partial t} g_t\left(x, \Sigma\right) + \left(\frac{\partial}{\partial x} g_t\left(x, \Sigma\right)\right)^T A_t x \\ &\quad - \frac{1}{4}\left(\frac{\partial}{\partial x} g_t\left(x, \Sigma\right)\right)^T B_t P_t^{-1} B_t^T\left(\frac{\partial}{\partial x} g_t\left(x, \Sigma\right)\right) \\ &\quad + \mathrm{tr}\left\{\left(\frac{\partial}{\partial \Sigma} g_t\left(x, \Sigma\right)\right)\left(A_t\Sigma + \Sigma A_t^T + D_t D_t^T\right) + Q_t\Sigma\right\} \\ &\quad + x^T Q_t x + \mu_t\mathcal{L}_t\left\{g_t\left(x, \Sigma\right)\right\}.\end{aligned}$$

Then, (4.37) can be written as

$$\begin{aligned}\mathrm{E}\left[g_{t+\epsilon}\big(\tilde{x}_{t+\epsilon},\tilde{\Sigma}_{t+\epsilon}\big)\big|\mathscr{B}_t\right] = g_t\big(\tilde{x}_t,\tilde{\Sigma}_t\big) + \epsilon\left\| u_t + \frac{1}{2}\,P_t^{-1}B_t^T\,\left(\frac{\partial}{\partial\tilde{x}_t}\,g_t\big(\tilde{x}_t,\tilde{\Sigma}_t\big)\right)\right\|_{P_t}^2 + \epsilon\delta_t^2 \\ - \epsilon\left(\tilde{x}_t^TQ_t\tilde{x}_t + \mathrm{tr}\left\{Q_t\tilde{\Sigma}_t\right\} + u_t^TP_tu_t\right) \\ + \epsilon\mathcal{K}_t\left\{g_t\big(\tilde{x}_t,\tilde{\Sigma}_t\big)\right\} + O\left(\epsilon^2\right). \qquad (4.38)\end{aligned}$$

Let $g_t\left(\cdot,\cdot\right)$ be the solution of the partial differential equation (4.11) with the boundary condition $g_T\left(x,\Sigma\right) = x^TSx$. This implies that $\mathcal{K}_t\big\{g_t\big(\tilde{x}_t,\tilde{\Sigma}_t\big)\big\} = 0$. Under this condition, we take expectation from (4.38) to get

$$\begin{aligned}\mathrm{E}\left[g_{t+\epsilon}\big(\tilde{x}_{t+\epsilon},\tilde{\Sigma}_{t+\epsilon}\big)\right] = \mathrm{E}\left[g_t\big(\tilde{x}_t,\tilde{\Sigma}_t\big)\right] + \epsilon\mathrm{E}\left[\left\| u_t + \frac{1}{2}\,P_t^{-1}B_t^T\,\left(\frac{\partial}{\partial\tilde{x}_t}\,g_t\big(\tilde{x}_t,\tilde{\Sigma}_t\big)\right)\right\|_{P_t}^2\right] \\ + \epsilon\mathrm{E}\left[\delta_t^2\right] - \epsilon\mathrm{E}\left[\tilde{x}_t^TQ_t\tilde{x}_t + \mathrm{tr}\left\{Q_t\tilde{\Sigma}_t\right\} + u_t^TP_tu_t\right] + O\left(\epsilon^2\right). \qquad (4.39)\end{aligned}$$

Step IV: We partition the interval $[0,T)$ into K subintervals $[t_k,t_{k+1}),\ k=0,1,\ldots,K-1$, where $t_0=0$, $t_K=T$, and $t_{k+1}-t_k \triangleq \epsilon_k > 0$. Recalling that $x_T^TSx_T = g_{t_K}\big(\tilde{x}_{t_K},\tilde{\Sigma}_{t_K}\big)$, we approximate the cost functional (4.26) by the finite sum

$$J \simeq J_K = \sum_{k=0}^{K-1}\epsilon_k\mathrm{E}\left[\tilde{x}_{t_k}^TQ_{t_k}\tilde{x}_{t_k} + \mathrm{tr}\left\{Q_{t_k}\tilde{\Sigma}_{t_k}\right\} + u_{t_k}^TP_{t_k}u_{t_k} + \delta_{t_k}^1\right] + \mathrm{E}\left[g_{t_K}\big(\tilde{x}_{t_K},\tilde{\Sigma}_{t_K}\big)\right]$$

and rearrange it as

$$J_K = \sum_{k=0}^{K-2} \epsilon_k \mathrm{E}\left[\tilde{x}_{t_k}^T Q_{t_k} \tilde{x}_{t_k} + \mathrm{tr}\left\{Q_{t_k}\tilde{\Sigma}_{t_k}\right\} + u_{t_k}^T P_{t_k} u_{t_k} + \delta_{t_k}^1\right] + \mathrm{E}\left[\delta_{t_{K-1}}^1\right]$$
$$+ \epsilon_{K-1}\mathrm{E}\left[\tilde{x}_{t_{K-1}}^T Q_{t_{K-1}} \tilde{x}_{t_{K-1}} + \mathrm{tr}\left\{Q_{t_{K-1}}\tilde{\Sigma}_{t_{K-1}}\right\} + u_{t_{K-1}}^T P_{t_{K-1}} u_{t_{K-1}}\right]$$
$$+ \mathrm{E}\left[g_{t_K}\big(\tilde{x}_{t_K}, \tilde{\Sigma}_{t_K}\big)\right].$$

In this expression, we replace $\mathrm{E}\left[g_{t_K}\big(\tilde{x}_{t_K}, \tilde{\Sigma}_{t_K}\big)\right]$ by the right side of (4.39). With minor manipulations, and upon defining $\delta_t = \delta_t^1 + \delta_t^2$ according to (4.13), we find that

$$J_K = \sum_{k=0}^{K-2} \epsilon_k \mathrm{E}\left[\tilde{x}_{t_k}^T Q_{t_k} \tilde{x}_{t_k} + \mathrm{tr}\left\{Q_{t_k}\tilde{\Sigma}_{t_k}\right\} + u_{t_k}^T P_{t_k} u_{t_k} + \delta_{t_k}^1\right] + \mathrm{E}\left[g_{t_{K-1}}\big(\tilde{x}_{t_{K-1}}, \tilde{\Sigma}_{t_{K-1}}\big)\right]$$
$$+ \epsilon_{K-1}\mathrm{E}\left[\left\| u_{t_{K-1}} + \frac{1}{2} P_{t_{K-1}}^{-1} B_{t_{K-1}}^T \left(\frac{\partial}{\partial \tilde{x}_{t_{K-1}}} g_{t_{K-1}}\big(\tilde{x}_{t_{K-1}}, \tilde{\Sigma}_{t_{K-1}}\big)\right)\right\|_{P_{t_{K-1}}}^2\right]$$
$$+ \epsilon_{K-1}\mathrm{E}\left[\,\delta_{t_{K-1}}\right] + O\left(\epsilon_{K-1}^2\right).$$

Repeating this procedure for $k = K-2, K-3, \ldots, 1, 0$, we obtain

$$J_K = \mathrm{E}\left[g_{t_0}\big(\tilde{x}_{t_0}, \tilde{\Sigma}_{t_0}\big)\right] + \sum_{k=0}^{K-1} \epsilon_k \,\mathrm{E}\left[\,\delta_{t_k}\right]$$
$$+ \sum_{k=0}^{K-1} \epsilon_k \,\mathrm{E}\left[\left\| u_{t_k} + \frac{1}{2} P_{t_k}^{-1} B_{t_k}^T \left(\frac{\partial}{\partial \tilde{x}_{t_k}} g_{t_k}\big(\tilde{x}_{t_k}, \tilde{\Sigma}_{t_k}\big)\right)\right\|_{P_{t_k}}^2\right] + \sum_{k=0}^{K-1} O\left(\epsilon_k^2\right).$$

Finally, we take the limit of J_K as $K \to \infty$ and $\max \epsilon_k \to 0$ to obtain (4.12).

4.5.4 Proof of Theorem 4.4.2

For $\mathcal{A} = \mathbb{R}^m$ and $g_t(x, \Sigma)$ given by (4.15), we can show that

$$\mathcal{L}_t \{g_t(x, \Sigma)\} = f_t\big(\Sigma - \Gamma_t(\Sigma) C_t \Sigma\big) - f_t(\Sigma) + \operatorname{tr}\{\Gamma_t(\Sigma) C_t \Sigma K_t\}$$

which clearly does not depend on x. Therefore, (4.11) can be decomposed into two decoupled equations: the partial differential equation (4.17) with the boundary condition $f_T(\Sigma) = 0$ and the equation

$$\begin{aligned} -\frac{\partial\left(x^T K_t x\right)}{\partial t} = & \left(\frac{\partial\left(x^T K_t x\right)}{\partial x}\right)^T A_t x \\ & - \frac{1}{4}\left(\frac{\partial\left(x^T K_t x\right)}{\partial x}\right)^T B_t P_t^{-1} B_t^T \left(\frac{\partial\left(x^T K_t x\right)}{\partial x}\right) + x^T Q_t x \end{aligned}$$

with the boundary condition $x^T K_T x = x^T S x$. This equation holds for any arbitrary x if and only if K_t satisfies (4.16) with the terminal condition $K_T = S$.

Chapter 5

Cooperative Optical Beam Tracking: Concept and Model

5.1 Introduction

In free-space optical communication using narrow laser beams, it is required to maintain the optical alignment between the stations in spite of their relative motion. This relative motion is caused by the mobile nature of the stations, mechanical vibration, or accidental shocks. In order to establish and maintain a free-space optical link, a two-phase optical alignment mechanism is required. In the first phase, a coarse alignment is achieved through the open-loop operation of spatial acquisition [41, 44]. Following the coarse alignment phase, data transmission is established and simultaneously a closed-loop fine alignment operation is performed to precisely compensate for the persistent relative motion of the stations. A possible scheme to achieve this fine alignment is cooperative (reciprocal) optical beam tracking.

A cooperative optical beam tracking system consists of two stations in such a manner that each station points its optical beam toward the other one. The receiving station continuously measures the arrival direction of its incident optical beam in order to employ it as a guide to precisely point its own beam toward the other

station. In short range applications with negligible light propagation delay, this direction is approximately along the line-of-sight of the stations, thus the stations transmit their optical beams along this measured direction. In applications with a large propagation delay, the optical beams must be transmitted within a certain angle with respect to the arrival direction in order to compensate for the variation of the line-of-sight during the travel time of the transmitted beams. This requires the transmitter to predict the future location of the receiver and point its optical beam toward the predicted location.

To implement the alignment scheme above, the stations are equipped with a position-sensitive photodetector (e.g., quadrant detector) and a focusing lens (or an arrangement of curved mirrors) to measure the azimuth and elevation components of the beam arrival direction. In addition, each station employs an electromechanical pointing assembly to adjust the direction of its optical devices according to the control signals provided by a closed-loop controller. The controller incorporates the output of the position-sensitive photodetector and generates proper azimuth and elevation control signals. As an alternative (or complement) to adjusting the transceiver direction, the incoming and outgoing optical fields can be directed using an arrangement of steerable flat mirrors.

The goal of this chapter is to develop a mathematical model for a cooperative optical beam tracking system, which includes the nonlinear effects, major disturbance sources, and light propagation delay. For analyzing the optical alignment between two fast maneuvering stations (e.g. aircrafts), the nonlinearity of the dynamical equations is essential; however, in applications such as intersatellite commu-

nication in which the relative motion consists of a predetermined large component and an unknown small component, we can linearize the nonlinear dynamics around a nominal state trajectory.

In the last section, we shall describe the relative motion of the stations by means of a set of stochastic differential equations. This stochastic model will be used later in Chapter 6 for a stochastic analysis of the system, as an alternative to the deterministic approach of [21, 22].

5.2 System Architecture

In this section, we first consider the structure and components of an optical transceiver and then describe the operation of a cooperative optical beam tracking system which employs two transceivers of this type.

5.2.1 Transceiver Structure

A schematic diagram of a simple transceiver used in short range free-space optical links is illustrated in Figure 5.1 (see also [1]). This transceiver comprises a lens, a position-sensitive photodetector, and a narrow laser source, all installed on a rigid platform. The photodetector surface is perpendicular to the lens axis and its center is placed at the focus of the lens. The axes of the lens and the laser source are parallel to transceiver axis. The azimuth and elevation of the transceiver axis can be controlled by means of an electromechanical pointing assembly, which is mounted on the station body.

The optical beam generated by the laser source is used for two purposes: as a

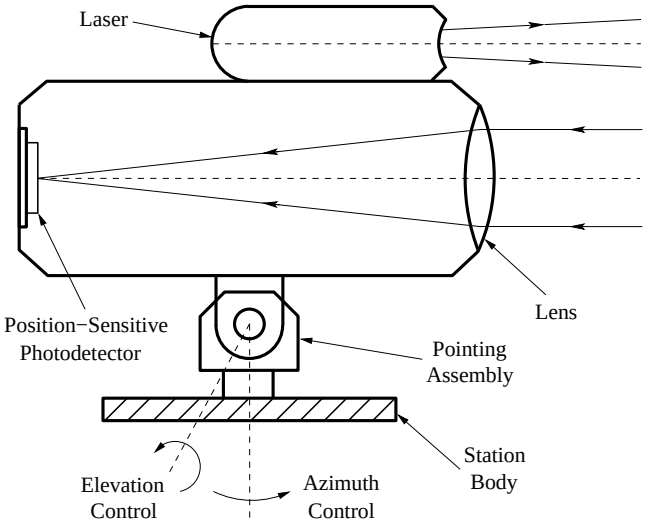

Figure 5.1: Schematic diagram of an optical transceiver for short range applications.

carrier of information and as a beacon assisting the opposite station in its tracking and pointing operations. For the purpose of communication, the instantaneous laser power is modulated by the information-bearing signal, usually with a digital form of on-off-keying.

The position-sensitive photodetector is a photoelectron converter whose surface is partitioned into small regions. The output of each region counts the number of converted electrons regardless of their location on the region. The photoelectron conversion rate depends on the instantaneous optical power absorbed by the region. The image of the received optical field on the surface of the photodetector is a spot of light with a bell-shaped intensity pattern whose location depends on the angle of arrival of the optical field with respect to the transceiver axis. Hence, using the position-sensitive photodetector, this angle can be tracked by measuring the location of the spot of light. Many practical optical beam tracking systems employ a

quadrant detector[1] as their optical sensing device, while the low spatial resolution of a quadrant detector can be improved using a finer partition. For instance, the authors of [45] describe a beam tracking system which employs a photodetector with 512×512 pixels.

The pointing assembly is usually a two-axes gimballed system with two independent motor which control the azimuth and elevation of the transceiver. Gimballed pointing systems generally suffer from low bandwidth (in order of 10 Hz) and low slew rate, while being able to cover a large solid angle. Also, they have the disadvantage of being singular at certain points, which limits their coverage region [46]. To resolve this difficulty, Omni-Wrist III is an alternative antenna pointer with double universal joints and linear actuators, which has 2π steradian range of motion without singularity [46].

A more sophisticated transceiver design, used for intersatellite communication, is illustrated in Figure 5.2 (for detailed discussion see [47, 48]). Similar to Figure 5.1, this design employs a position-sensitive photodetector, a pointing assembly, and a laser source; however, instead of a lens, it employs a reflecting telescope.

The telescope which is shared between the receiving and the transmitting optics, consists of a primary and a secondary curved mirror with one of the several common designs. The most popular [47] design, Cassegrainian telescope, employs a parabolic primary mirror and a hyperbolic secondary mirror which share the same focus. In addition to the telescope, an arrangement of lenses (not shown in Figure 5.2) might be used for extra magnification [47]. In design of the transceiver,

[1]Quadrant detector is a position-sensitive photodetector with a four-region partition.

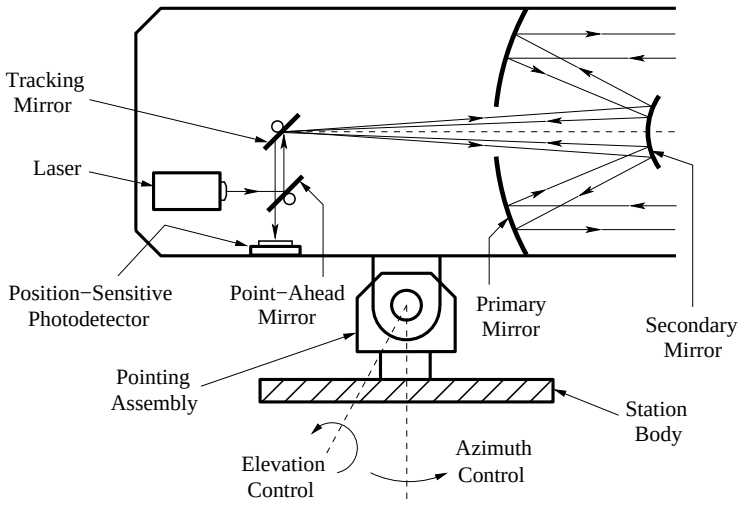

Figure 5.2: Optical transceiver for intersatellite communication (based on [48]).

the incoming and outgoing optical fields must be isolated as much as possible, since the backscattered photons caused by the outgoing light emerge as a source of noise for the photodetector. This can be achieved by a combination of spectral isolation, spatial separation, and polarization isolation [47]. In the situations that these techniques cannot provide enough isolation, two separated telescopes are required for the incoming and the outgoing optical beams [47], while this dual telescope approach leaves the tracking function of the transceiver unchanged.

The tracking mirror in Figure 5.2 is intended to control the direction of the incoming light toward the position-sensitive photodetector and the outgoing light toward the target. This steerable flat mirror, which is equipped with miniature actuators, provides a complementary (or alternative) means for the pointing assembly. The steering machinery consists of a support plate with a single pivot and three or four piezoelectric linear actuators (fast steering mirror). Although, the scanning region of a steerable mirror is small (less than 5 degrees in each direction), its small

mass and fast actuators result in a high bandwidth (up to 1 kHz) and high slew rate. This provides considerable assistance to the pointing assembly in suppressing the high bandwidth disturbance.

The point-ahead mirror is another steerable flat mirror with the purpose of compensating for the displacement of the receiver during light propagation time. This mirror provides an additional degree of freedom in controlling the pointing direction of the outgoing light.

5.2.2 The Concept of Cooperative Optical Beam Tracking

We consider a two-way optical link consisted of two transceivers of the type discussed earlier, in such a manner that each station transmits its optical beam toward the other station and receives the optical beam from the other side. We assume that the stations are subjected to relative motion.

For a simple description of the alignment scheme, consider the transceiver of Figure 5.1 and suppose that a uniform optical field strikes the transceiver aperture (i.e., the lens). When the striking optical field propagates along the axis of the lens, its image is a spot of light at the center of the position-sensitive photodetector, while any deviation from this direction shifts the spot of light from the center. This shift can be detected by the position-sensitive photodetector. The output of the photodetector is fed to a closed-loop controller, which adjusts the transceiver direction by applying proper signals to the pointing assembly. For short range applications (negligible propagation delay), the goal of the controller is to eliminate the angle between the lens axis and the arrival direction of the incident optical beam.

Since the axes of the lens and the laser source are parallel, this operation aligns the propagation direction of the transmitted optical beam with the arrival direction of the received optical beam. Assuming that both stations actively perform this operation, the propagation direction of the optical beams stay close to the line-of-sight, in spite of the relative motion between the stations. The block diagram in Figure 5.3 illustrates the interconnection between the components of a cooperative optical beam tracking system.

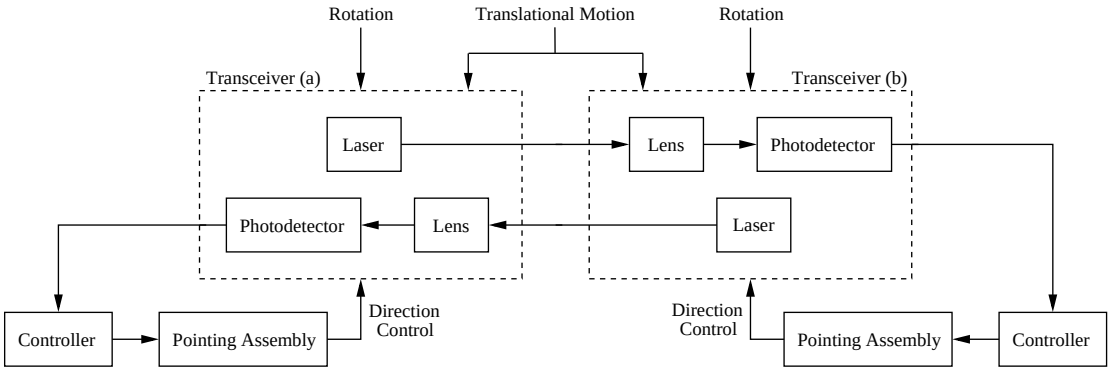

Figure 5.3: Interconnection between the components of a cooperative optical beam tracking system.

In the optical transceiver of Figure 5.2, the task of adjusting the light direction is distributed between the pointing assembly and the tracking mirror. In applications such as intersatellite communication, the relative motion consists of a large, slowly varying component and a small, high bandwidth term. Accordingly, the control law consists of an open-loop, coarse control and a closed-loop, fine control. In this case, a reasonable design employs the pointing assembly for the purpose of coarse (open-loop) control and the tracking mirror for closed-loop fine control.

In the transceiver of Figure 5.1, a successful tracking operation which keeps

the spot of light close to the center of the photodetector, requires the transmitted optical beam to propagate along the arrival direction of the received optical beam; however, in applications with a large propagation delay, the optical beam must be pointed ahead with respect to this arrival direction. The required point-ahead angle can be accommodated by means of the point-ahead mirror.

5.2.3 Assisting Equipments

Application of inertial sensors (gyro and accelerometer) in the alignment aspect of intersatellite optical communication is considered in [49]. Through measuring the angular velocity and acceleration, these sensors provide information regarding the position of the stations. This information can be combined with the output of the photodetectors to improve the overall performance of the system.

Another possibility for improving the performance of the system is to exchange information between the stations. Sharing the output of the photodetectors and the inertial sensors enables each individual station to produce a more accurate estimate of the line-of-sight, which in turn, increases the capability of the stations to compensate for their relative motion. The means for exchange of information can be provided by the optical channel itself or an independent low bandwidth RF channel.

In the optical transceiver of Figure 5.1, an additional pointing error can be introduced by a misalignment between the axes of the lens and the laser source. This misalignment might occur due to the imperfect manufacturing process. The same type of pointing error arises in the optical transceiver of Figure 5.2, because of the imperfect positioning of the laser source and the point-ahead mirror. Note

that this type of pointing error can be measured only by the receiving station, thus, in order to eliminate it, the stations need to share their outputs. This is another reason which indicates that information exchange between the stations improves the performance of the system.

5.3 The Model

In this section, we develop a mathematical model for a cooperative optical beam tracking system. In order to avoid unnecessary complications in notation, we shall assume that the optical link under consideration consists of two identical transceivers (Figure 5.1 or Figure 5.2). We begin by fixing notation and defining necessary coordinate systems. Then we determine the optical field on the aperture of each transceiver, which will be used later to derive a formula for the optical intensity on the photodetector surface. This will be followed by discussing the effect of atmospheric turbulence on the optical intensity. Next, we shall present a statistical description for the photodetector output in terms of the optical intensity. Finally, we introduce the dynamical equations which describe the temporal evolution of the system.

5.3.1 Notation and Coordinate Systems

In what follows, we distinguish the stations by superscripts a and b or $i = a, b$ when referring to both stations. Let $\mathcal{I}$ be an inertial coordinate system. Consider the position vector of station i with respect to the origin of $\mathcal{I}$ and denote by r_t^i its representation in coordinate system $\mathcal{I}$ at time t. The light travel time between the

stations (with abuse of notation) will be denoted by t_d and can be determined from

$$t_d = c^{-1} \left\| r_t^a - r_t^b \right\|$$

where c is the light velocity in the propagation medium (vacuum). Note that in general, t_d depends on time which is not emphasized by this notation.

For station i and at time t, we define the coordinate system $\mathcal{R}_t^i$ which is attached to the center of the transceiver aperture in such a manner that its z axis extends outward perpendicular to the aperture plane, its x axis is parallel to the elevation axis (see Figure 5.1 and Figure 5.2), and its y axis is the cross product of z and x axes. We denote by Ω_t^i the rotation matrix from coordinate system $\mathcal{I}$ to coordinate system $\mathcal{R}_t^i$. Let $\mathcal{B}_t^i$ be a coordinate system fixed to the body of station i and denote by ω_t^i the angular velocity of $\mathcal{I}$ with respect to $\mathcal{B}_t^i$ represented in $\mathcal{I}$. The angular velocity of $\mathcal{B}_t^i$ with respect to $\mathcal{R}_t^i$ represented in $\mathcal{R}_t^i$ will be denoted by υ_t^i.

We shall assume that the optical field transmitted by station i propagates along the z axis of the coordinate system $\mathcal{T}_t^i$. This coordinate system is obtained by two successive rotations of $\mathcal{R}_t^i$ in the following manner. First, rotate $\mathcal{R}_t^i$ around its x axis by an angle $-\delta_t^{x,i}$ to get the coordinate system $\tilde{\mathcal{R}}_t^i$, and then rotate $\tilde{\mathcal{R}}_t^i$ around its y axis by an angle $-\delta_t^{y,i}$ to obtain $\mathcal{T}_t^i$. The rotation matrix from $\mathcal{R}_t^i$ to $\mathcal{T}_t^i$ is given by

$$\Delta_t^i = \begin{bmatrix} \cos\delta_t^{y,i} & \sin\delta_t^{x,i}\sin\delta_t^{y,i} & \cos\delta_t^{x,i}\sin\delta_t^{y,i} \\ 0 & \cos\delta_t^{x,i} & -\sin\delta_t^{x,i} \\ -\sin\delta_t^{y,i} & \sin\delta_t^{x,i}\cos\delta_t^{y,i} & \cos\delta_t^{x,i}\cos\delta_t^{y,i} \end{bmatrix}.$$

For small $\delta_t^{x,i}$ and $\delta_t^{y,i}$, this matrix can be approximated by

$$\Delta_t^i \simeq \begin{bmatrix} 1 & 0 & \delta_t^{y,i} \\ 0 & 1 & -\delta_t^{x,i} \\ -\delta_t^{y,i} & \delta_t^{x,i} & 1 \end{bmatrix}. \tag{5.1}$$

Let $a = [a_1 \quad a_2 \quad a_3]^T$ and define the operator $[\cdot]_\times$ such that

$$[a]_\times = \begin{bmatrix} 0 & -a_3 & a_2 \\ a_3 & 0 & -a_1 \\ -a_2 & a_1 & 0 \end{bmatrix}.$$

Then, (5.1) can be expressed in the compact form[1]

$$\Delta_t^i = I + \left[I_*^T \delta_t^i\right]_\times \tag{5.2}$$

where I is the 3×3 identity matrix, $\delta_t^i = \begin{bmatrix}\delta_t^{x,i} & \delta_t^{y,i}\end{bmatrix}^T$, and the 2×3 matrix I_* is defined by

$$I_* = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}.$$

In the optical transceiver of Figure 5.1, for the ideal case that the axes of the lens and the laser source are perfectly aligned, we have $\delta_t^i = 0$; however, as mentioned in Section 5.2.3, due to an imperfect manufacturing process, there might be a small angle between the lens axis and the laser axis. This can be modeled by letting $\delta_t^i = \varepsilon_t^i$ in (5.2), where the elements of ε_t^i are the misalignment errors in x

[1]Afterward, we consider (5.1) as an exact formula.

and y directions (elevation and azimuth).

For the optical transceiver of Figure 5.2, δ_t^i depends on the orientations of the tracking and point-ahead mirrors which are described by the 2-dimensional vectors α_t^i and β_t^i, respectively. These vectors are defined as follows. Let ζ_t^i be a vector normal to the tracking mirror i and define $\bar{\zeta}^i$ such that $\bar{\zeta}^i = \zeta_t^i$ when the actuators of the mirror are in the "rest" condition. Then, the elements of α_t^i are the deviation (angles) of ζ_t^i from $\bar{\zeta}^i$ in x and y directions. Since the elements of α_t^i and β_t^i are small, their contributions to δ_t^i appear linearly, i.e., we can write

$$\delta_t^i = K\alpha_t^i + L\beta_t^i + \varepsilon_t^i$$

where K and L are 2×2 known matrices. Replacing this result into (5.2), we express the rotation matrix from $\mathcal{R}_t^i$ to $\mathcal{T}_t^i$ as

$$\Delta_t^i = I + \left[I_*^T \left(K\alpha_t^i + L\beta_t^i + \varepsilon_t^i\right)\right]_\times . \tag{5.3}$$

Note that this expression can be used for the transceiver of Figure 5.1 as well, by setting $\alpha_t^i = \beta_t^i = 0$.

Regarding the optical transceiver of Figure 5.1, we define the 2-dimensional coordinate system $\mathcal{P}_t^i$ in the focal plane of the lens (also the photodetector plane) such that its center is located at the focus of the lens and its x and y axes are parallel to x and y axes of $\mathcal{R}_t^i$, respectively. For the optical transceiver of Figure 5.2, the coordinate system $\mathcal{P}_t^i$ is defined in the plane of the photodetector surface such

that its origin is located at the projection of the focus of the secondary mirror on the photodetector surface, its x axis is parallel to the x axis of $\mathcal{R}_t^i$, and its y axis is perpendicular to its x axis.

5.3.2 Optical Field on the Transceiver Aperture

Consider the coordinate system $\mathcal{O}$ with axes (a_x, a_y, a_z) and assume that a laser beam with a wavelength λ and a divergence angle $\bar{\psi}$ propagates along a_z. Since the major fraction of the laser power is concentrated in its fundamental transverse mode (TEM_{00}), we shall use a Gaussian beam model for the optical field generated by the laser. Based on this model, the complex amplitude of the optical field at a point $r = (x, y, z)\,,\ z > 0$ is given [23] by

$$U(r) = \sqrt{\frac{2P}{\pi}} \cdot \frac{1}{w(z)} \exp\left\{-jkz - \left(\frac{1}{w^2(z)} + \frac{jk}{2R(z)}\right)\left(x^2 + y^2\right)\right\} \tag{5.4}$$

where $P > 0$ is the laser power, $k = 2\pi/\lambda$ is the wave number, and $w(z)$ and $R(z)$ are defined as

$$w(z) = w_0\left(1 + (z/z_0)^2\right)^{1/2}$$

$$R(z) = z\left(1 + (z_0/z)^2\right)$$

with[1] $w_0 = \lambda/\pi\bar{\psi}$ and $z_0 = \lambda/\pi\bar{\psi}^2$.

Suppose that r can be decomposed as $r = \bar{r} + \delta r$ such that $\bar{r} = (\bar{x}, \bar{y}, \bar{z})$ and $\delta r = (\delta x, \delta y, \delta z)$ satisfy the conditions $\|\delta r\|/\|\bar{r}\| \ll \bar{\psi}$ and $|\delta z| \ll z_0 \ll \|\bar{r}\|$.

[1] In spite of our convention, here, the subscript 0 is not used as time.

Denote by $\langle \bar{r}, a_z \rangle$ the angle between the vectors $\bar{r}$ and a_z and assume that the magnitude of $\langle \bar{r}, a_z \rangle$ is in order of $\bar{\psi}$ or smaller, i.e. $\langle \bar{r}, a_z \rangle < \bar{\psi}$. Then, for a small $\bar{\psi}$ (e.g., smaller that 1 mrad), the optical field (5.4) can be approximated by

$$U\left(\bar{r}+\delta r\right) \simeq \sqrt{\frac{2P}{\pi}} \cdot \frac{1}{\bar{\psi}\|\bar{r}\|} \exp\left(-\frac{\langle \bar{r}, a_z \rangle^2}{\bar{\psi}^2}\right) \cdot \exp\left(-jk\frac{\bar{r} \cdot \delta r}{\|\bar{r}\|} - jk\frac{\delta x^2 + \delta y^2}{2\|\bar{r}\|}\right) \exp\left(-j\bar{\phi}\right) \tag{5.5}$$

where $\cdot$ denotes the dot product operator and $\bar{\phi}$ is defined as

$$\bar{\phi} = k\bar{z} + k\frac{\bar{x}^2 + \bar{y}^2}{2R\left(\bar{z}\right)} \cdot$$

In the context of free-space optics, $\bar{r}$ represents the line-of-sight of the receiver with respect to the transmitter and δr is the position vector of a point on the aperture plane of the receiver with respect to the center of the aperture. With this assignment, $|\delta z|/\|\delta r\|$ will be in the order of the pointing error $\langle \bar{r}, a_z \rangle$. Table 5.1 presents some typical values of the parameters of a free-space optical link which are relevant to approximation (5.5). Note that in Table 5.1, it is assumed that under the closed-loop regime, the pointing error is in order of 1/10 of the angular spread of the beam. According to this table, the conditions for approximating (5.4) with (5.5) are satisfied for all scenarios, i.e., short, medium, and long range applications.

Note that $\bar{\phi}$ on the right side of (5.5) introduces a constant phase over the aperture of the receiver, which can be dropped without affecting our discussion. In

Parameter	Symbol	Short Range (terrestrial)	Medium Range (airborne)	Long Range (intersatellite)
Wavelength	λ	$1\ \mu$m	$1\ \mu$m	$1\ \mu$m
Angular Spread	$2\bar{\psi}$	$500-3000\ \mu$rad	$100\ \mu$rad	$1-50\ \mu$rad
Range	$\|\bar{r}\|$	$0.1-5$ km	$10-100$ km	$1000-80000$ km
Aperture Diameter	$2\|\delta r\|$	$5-20$ cm	20 cm	$20-100$ cm
Pointing Error	$\langle \bar{r}, a_z \rangle$	$50-300\ \mu$rad	$10\ \mu$rad	$0.1-5\ \mu$rad
—	z_0	$0.15-5$ m	125 m	500 m $-$ 1250 km
—	$\|\delta z\|$	order of μm	order of μm	order of μm

Table 5.1: Typical values of the parameters of a free-space optical link. The data is gathered from [47, 3, 50].

addition, Table 5.1 indicates that for long range applications we have

$$k\frac{\delta x^2+\delta y^2}{2\|\bar{r}\|} \ll 1 \tag{5.6}$$

which allows us to simplify (5.5) as

$$U\left(\bar{r}+\delta r\right) \simeq \sqrt{\frac{2P}{\pi}}\cdot\frac{1}{\bar{\psi}\|\bar{r}\|}\exp\left(-\frac{\langle \bar{r}, a_z\rangle^2}{\bar{\psi}^2}\right)\exp\left(-jk\frac{\bar{r}\cdot\delta r}{\|\bar{r}\|}\right). \tag{5.7}$$

Although (5.7) is not a good approximation for short and medium range applications, still it can be used for these cases, since the phase (5.6) can be compensated by proper adjustment of the distance between the lens and the photodetector surface [51].

We use approximation (5.7) to determine the optical field on the aperture of the stations. For this purpose, consider the position vector of a point on the aperture plane of transceiver i with respect to the center of the aperture and let $I_*^T s^i,\ s^i \in \mathbb{R}^2$ be its representation in the coordinate system $\mathcal{R}_t^i$. Then, the representation of this vector in $\mathcal{I}$ is given by $\left(I_*\Omega_t^i\right)^T s^i$. We define the 2-dimensional (tracking error)

vectors θ_t^a and θ_t^b as

$$\begin{aligned}\theta_t^a &= \frac{I_* \Omega_t^a \left(r_t^b - r_t^a\right)}{\left\|r_t^b - r_t^a\right\|} \\ \theta_t^b &= \frac{I_* \Omega_t^b \left(r_t^a - r_t^b\right)}{\left\|r_t^a - r_t^b\right\|}\,.\end{aligned} \tag{5.8}$$

In order to find the optical field on the aperture of station b, we replace $\bar{r}$ with $r_t^b - r_t^a$ and δr with $\left(I_* \Omega_t^b\right)^T s^b$ in (5.7). This is equivalent to replacing $\bar{r} \cdot \delta r / \|\bar{r}\|$ with $-\theta_t^b \cdot s^b$. To obtain a proper replacement for $\langle \bar{r}, a_z \rangle$, we must take into account the travel time of light between the stations. For this purpose, we define the 2-dimensional (pointing error) vectors ψ_t^a and ψ_t^b as

$$\begin{aligned}\psi_t^a &= \frac{I_* \Delta_t^a \Omega_t^a \left(r_{t+t_d}^b - r_{t+t_d}^a\right)}{\left\|r_{t+t_d}^b - r_{t+t_d}^a\right\|} \\ \psi_t^b &= \frac{I_* \Delta_t^b \Omega_t^b \left(r_{t+t_d}^a - r_{t+t_d}^b\right)}{\left\|r_{t+t_d}^a - r_{t+t_d}^b\right\|}\,.\end{aligned} \tag{5.9}$$

Then, it is easy to show that $\langle \bar{r}, a_z \rangle^2$ must be replaced by $\left\|\psi_{t-t_d}^a\right\|^2$. In a similar manner, we can find proper replacements for determining the optical field on the aperture of station a.

Let $U_t^i(s)$ denote the complex amplitude of the optical field at a point s on the aperture of station i. Then, using the replacements mentioned above, we can write

$$\begin{aligned}U_t^a(s) &= \left(\frac{2P_{t-t_d}^b}{\pi \bar{\psi}^2 \left\|r_t^a - r_t^b\right\|^2}\right)^{1/2} \exp\left(-\left(\left\|\psi_{t-t_d}^b\right\| / \bar{\psi}\right)^2\right) \exp\left(jk\,\theta_t^a \cdot s\right) \\ U_t^b(s) &= \left(\frac{2P_{t-t_d}^a}{\pi \bar{\psi}^2 \left\|r_t^b - r_t^a\right\|^2}\right)^{1/2} \exp\left(-\left(\left\|\psi_{t-t_d}^a\right\| / \bar{\psi}\right)^2\right) \exp\left(jk\,\theta_t^b \cdot s\right)\end{aligned} \tag{5.10}$$

where $P_t^i \geqslant 0$ is the instantaneous power transmitted by station i. It can be observed from (5.10) that the optical intensity at station b depends on the pointing error $\psi_{t-t_d}^a$ of station a, while the phase depends only on the tracking error θ_t^b of station b.

Note that in (5.10), we allow P_t^i to be time-dependent in order to describe the information-bearing signal which modulates the optical power of station i. The nature of the information-bearing signals suggests that a nonnegative stochastic process is an appropriate means for modeling P_t^i. The statistical properties of this stochastic process depends on the type of modulation scheme and channel coding; however, for many applications, the detailed characterization is not necessary and only a few parameters (e.g., expected value and coherence time) and the knowledge of general properties (e.g., stationarity) of the process are enough to use the model.

5.3.3 Optical Intensity on the Photodetector Surface

The optical field on the focal plane of a thin lens can be determined from Fraunhofer diffraction [41, 51]. Consider the optical transceiver of Figure 5.1 and assume that the optical field $U_t^i(s)$ on the receiving lens is given by (5.10). According to Fraunhofer diffraction, the optical field on the $x-y$ plane of $\mathcal{P}_t^i$ is given [41, 51] by

$$U_t^{d,i}(s) = \frac{1}{j\lambda f_c} \exp\left(\frac{jk\|s\|^2}{2f_c}\right) \int_{|\tilde{s}|\leqslant\varrho} U_t^i(\tilde{s}) \exp\left(-j\frac{k}{f_c} s \cdot \tilde{s}\right) d\tilde{s} \tag{5.11}$$

where f_c and ϱ are the focal length and the radius of the (circular) lens, respectively. In terms of (5.11), the optical intensity on the photodetector surface is given by

$$I_t^i(s) = \left| U_t^{d,i}(s) \right|^2. \tag{5.12}$$

Let $J_1(\cdot)$ be a Bessel function of the first kind and define the intensity pattern $\gamma(\cdot) : \mathbb{R}^2 \to \mathbb{R}$ as

$$\gamma(s) = \frac{J_1^2(k\varrho\|s\|/f_c)}{\pi\|s\|^2}. \tag{5.13}$$

Then, using (5.10), (5.11), and (5.12) and defining $y_t^i = f_c\theta_t^i$, we determine the optical intensities $I_t^a(s)$ and $I_t^b(s)$ as

$$\begin{aligned} I_t^a(s) &= \frac{2\varrho^2 P_{t-t_d}^b}{\bar{\psi}^2 \left\| r_t^a - r_t^b \right\|^2} \exp\left(-2\left(\left\| \psi_{t-t_d}^b \right\| / \bar{\psi} \right)^2 \right) \gamma\left(s - y_t^a \right) \\ I_t^b(s) &= \frac{2\varrho^2 P_{t-t_d}^a}{\bar{\psi}^2 \left\| r_t^b - r_t^a \right\|^2} \exp\left(-2\left(\left\| \psi_{t-t_d}^a \right\| / \bar{\psi} \right)^2 \right) \gamma\left(s - y_t^b \right). \end{aligned} \tag{5.14}$$

For the optical transceiver of Figure 5.2, the entire telescope is considered as a "big" lens with a focal length f_c. Also, the effect of the tracking mirror on the optical intensity $I_t^i(s)$ will be included by modifying y_t^i as

$$y_t^i = f_c\theta_t^i + H\alpha_t^i \tag{5.15}$$

where H is a known 2×2 matrix. The linearity of (5.15) is justified by the assumption that $\|\alpha_t^i\|$ is small.

In order to simplify our analysis of Chapter 6, it is desirable to approximate

the intensity pattern $\gamma(s)$ with a Gaussian function, i.e.,

$$\gamma(s) \simeq \Phi_2(r; 0, R) \tag{5.16}$$

where $R = 2(f_c/k\varrho)^2 I_{2\times 2}$. The comparison between (5.13) and (5.16), illustrated in Figure 5.4, indicates that (5.16) is a reasonably close approximation for (5.13). Note that both $\gamma(s)$ and its Gaussian approximation approach the unit impulse $\delta(\|s\|)$, as $k\varrho/f_c \to \infty$.

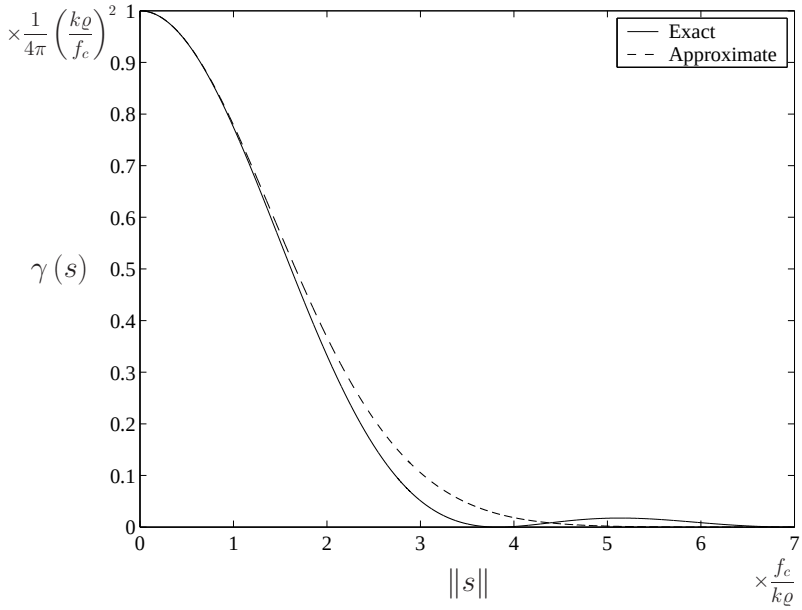

Figure 5.4: Comparison of $\gamma(s)$ with its Gaussian approximation.

5.3.4 Atmospheric Turbulence

The optical intensity (5.14) was determined based on the assumption that the laser beam propagates through free-space (vacuum). While this assumption is justified

for intersatellite applications, for communication through atmosphere, (5.14) must be modified to include the effect of atmospheric turbulence.

Atmospheric turbulence which is caused by differential heating of the air, results in random variations in the refractive index of air. This, in turn, causes random fluctuations in the intensity and phase of the received optical field. The refractive index of air as a function of the position vector r and time t can be modeled as $n_t(r) = \bar{n} + \delta n_t(r)$, where $\bar{n}$ is a constant and $\{\delta n_t(r)\}$ is a stochastic field. The statistical properties of $\{\delta n_t(r)\}$ can be derived from the Kolmogorov theory [52, 30, 26].

The Rytov method[1] is frequently used to analyze and model the propagation of an optical field in the turbulent atmosphere [30, 26]. In this method, the complex amplitude of the optical field is expressed as

$$U_t(r) = T_t(r)\,\bar{U}_t(r) \tag{5.17}$$

where $\bar{U}_t(r)$ is the optical field under the condition $\delta n_t(r) = 0$ and $\{T_t(r)\}$ is a stochastic field which can be determined in terms of $\{\delta n_t(r)\}$ and $\bar{U}_t(r)$.

In general, using (5.17) in obtaining an expression for $I_t^i(r)$ leads to a complicated calculation which is beyond the scope of this study. For short range applications (in order of 1 km for weak to moderate turbulence) in which the diameter of the receiving aperture is smaller than the turbulence coherence length[2], the sto-

[1]This method provides an approximate solution to the Maxwell wave equation.

[2]Roughly speaking, the turbulence coherence length (Fried parameter) is the maximum distance between two points r_1 and r_2 in which $T_t(r_1)$ and $T_t(r_2)$ are highly correlated. For a detailed characterization of this parameter see [52, 30].

chastic field $\{T_t(r)\}$ is approximately uniform over the aperture. Therefore, $\{T_t(r)\}$ can be approximated by a stochastic process. By means of this approximation, we modify (5.14) as

$$\begin{aligned} I_t^a(s) &= \frac{2\varrho^2 \kappa_t^b P_{t-t_d}^b}{\bar{\psi}^2 \left\| r_t^a - r_t^b \right\|^2} \exp\left(-2\left(\left\|\psi_{t-t_d}^b\right\| / \bar{\psi}\right)^2\right) \gamma\left(s - y_t^a\right) \\ I_t^b(s) &= \frac{2\varrho^2 \kappa_t^a P_{t-t_d}^a}{\bar{\psi}^2 \left\| r_t^b - r_t^a \right\|^2} \exp\left(-2\left(\left\|\psi_{t-t_d}^a\right\| / \bar{\psi}\right)^2\right) \gamma\left(s - y_t^b\right) \end{aligned} \tag{5.18}$$

where $\left\{\kappa_t^a\right\}$ and $\left\{\kappa_t^b\right\}$ are nonnegative stochastic processes.

Under the condition of approximation, $\{\kappa_t^i\}$ is a unit mean lognormal stochastic process [30, 26]. The variance of κ_t^i (for a fixed t) depends on the wavelength of the light, the propagation distance, the refractive-index structure constant, and the shape of the optical field [30]. Although, a complete description for the temporal evolution of $\{\kappa_t^i\}$ does not exist, its autocorrelation function can be approximated using the Taylor's frozen-flow hypothesis [52].

5.3.5 The Photodetector Output

In order to model the position-sensitive photodetectors, we first introduce the space-time rate $\lambda_t^i(r)$ which is an affine function of the optical intensity $I_t^i(r)$, i.e.,

$$\lambda_t^i(r) = \eta I_t^i(r) + \bar{\lambda}. \tag{5.19}$$

Here, the known constant $\eta > 0$ is the photodetector sensitivity[1] and $\bar{\lambda} \geqslant 0$ is a known constant which characterizes the combination of the dark current noise and the back ground radiation [29]. Substituting $I_t^i(r)$ from (5.18) into (5.19), we express $\lambda_t^a(r)$ and $\lambda_t^b(r)$ as

$$\begin{aligned} \lambda_t^a(r) &= \nu_{t-t_d}^a \exp\left(-2\left(\left\|\psi_{t-t_d}^b\right\|/\bar{\psi}\right)^2\right)\gamma\left(r-y_t^a\right)+\bar{\lambda} \\ \lambda_t^b(r) &= \nu_{t-t_d}^b \exp\left(-2\left(\left\|\psi_{t-t_d}^a\right\|/\bar{\psi}\right)^2\right)\gamma\left(r-y_t^b\right)+\bar{\lambda} \end{aligned} \tag{5.20}$$

where ν_t^a and ν_t^b are defined as

$$\begin{aligned} \nu_t^a &= \frac{2\eta\varrho^2\kappa_{t+t_d}^b P_t^b}{\bar{\psi}^2\left\|r_{t+t_d}^a - r_{t+t_d}^b\right\|^2} \\ \nu_t^b &= \frac{2\eta\varrho^2\kappa_{t+t_d}^a P_t^a}{\bar{\psi}^2\left\|r_{t+t_d}^b - r_{t+t_d}^a\right\|^2}. \end{aligned}$$

Note that ν_t^i/η is the instantaneous optical power received by station i in the absence of the pointing error.

We denote the set of points on the surface of the photodetector by $\mathcal{A}$. In a position-sensitive photodetector, $\mathcal{A}$ is partitioned into q subsets $\mathcal{A}^k,\ k = 1, 2, \cdots, q$ such that $\cup_{k=1}^q \mathcal{A}^k = \mathcal{A}$. The output of photodetector i is a q- dimensional vector Y_t^i such that its k^{th} element $Y_t^{k,i}$ is the output of the region $\mathcal{A}^k$. We model $Y_t^{k,i}$ as a doubly stochastic Poisson process with the rate process $\left\{\Lambda_t^{k,i}\right\}$, where $\Lambda_t^{k,i}$ is given by

$$\Lambda_t^{k,i} = \int_{\mathcal{A}^k} \lambda_t^i(r)\, dr. \tag{5.21}$$

[1]The photodetector sensitivity is given by $\eta = \zeta/\hbar\bar{f}$, where $\hbar$ is the Planck's constant, $\bar{f}$ is the mean frequency of light, and $0 < \zeta \leqslant 1$ is the quantum efficiency of the photodetector.

Moreover, for $l \neq k$, conditioned on $\left\{\Lambda_t^{k,i}\right\}$ and $\left\{\Lambda_t^{l,i}\right\}$, the stochastic processes $\left\{Y_t^{k,i}\right\}$ and $\left\{Y_t^{l,i}\right\}$ are mutually independent.

Remark 5.3.1. Let $\left|\mathcal{A}^k\right|$ denote the area of region k. Then, under the condition

$$q \to \infty \quad \text{and} \quad \max_{k\in\{1,2,\cdots,q\}} \left|\mathcal{A}^k\right| \to 0,$$

the vector-valued stochastic process Y_t^i tends to a space-time point process with rate (5.20). This is a motivation for approximating the output of a high spatial resolution photodetector by a space-time point process.

Remark 5.3.2. A first order approximation for a Poisson process is its expected value. For the q-dimensional vector Y_t^i, this approximation can be expressed as $Y_t^i \simeq \Lambda_t^i$, where Λ_t^i is a q-dimensional vector with elements $\Lambda_t^{k,i}$. This approximation can be improved as

$$Y_t^i \simeq \Lambda_t^i + n_t^i \tag{5.22}$$

where the noise vector n_t^i is independent of Λ_t^i. This simple approximation is useful when a deterministic approach is adopted to study the system [41, 21].

5.3.6 Dynamical Equations

Let $u_t^{p,i}$, $u_t^{\alpha,i}$, and $u_t^{\beta,i}$ denote the input (control) vectors of the pointing assembly, the tracking mirror, and the point-ahead mirror, respectively. Define the disturbance vector ρ_t^i as

$$\rho_t^a = -\rho_t^b = \frac{r_t^b - r_t^a}{\left\|r_t^b - r_t^a\right\|}\,.$$

The goal of this section is to develop a state-space model which determines the output vector (y_t^i, ψ_t^i) in terms of the control vector $u_t^i \triangleq \left(u_t^{p,i}, u_t^{\alpha,i}, u_t^{\beta,i}\right)$ and the disturbance vector $(\omega_t^i, \rho_t^i, \varepsilon_t^i)$. We note that (5.8) and (5.15) express y_t^i in terms of ρ_t^i, Ω_t^i, and α_t^i. Also, ψ_t^i can be determined in terms of ρ_t^i, ε_t^i, Ω_t^i, α_t^i, and β_t^i using (5.3) and (5.9). Thus, we need to obtain dynamical equations which describe the temporal evolution of Ω_t^i, α_t^i, and β_t^i.

Referring to the definition of υ_t^i and ω_t^i in Section 5.3.1, we can show that Ω_t^i is the solution of the matrix differential equation

$$\dot{\Omega}_t^i = \left[\upsilon_t^i\right]_\times \Omega_t^i + \Omega_t^i \left[\omega_t^i\right]_\times .$$

Here, the angular velocity vector $\upsilon_t^i \in \mathbb{R}^3$ is controlled by the pointing assembly. In the most general case, the relationship between υ_t^i and $u_t^{p,i}$ can be described by the nonlinear state-space equations

$$\begin{aligned} \dot{x}_t^{p,i} &= f\left(x_t^{p,i}, u_t^{p,i}\right) \\ \upsilon_t^i &= g\left(x_t^{p,i}\right) \end{aligned} \tag{5.23}$$

where $x_t^{p,i} \in \mathbb{R}^{n^p}$ is the state vector and $f(\cdot)$ and $g(\cdot)$ are smooth vector fields with proper dimensions. The explicit forms of $f(\cdot)$ and $g(\cdot)$ depend on the structure of the pointing assembly and will not be discussed here. For a two-axes gimballed pointing assembly, the control vector $u_t^{p,i}$ is 2-dimensional.

We model the dynamics of the steerable flat mirrors by the linear state-space

equations

$$\dot{x}_t^{\alpha,i} = A^\alpha x_t^{\alpha,i} + B^\alpha u_t^{\alpha,i}$$
$$\alpha_t^i = C^\alpha x_t^{\alpha,i}$$

and

$$\dot{x}_t^{\beta,i} = A^\beta x_t^{\beta,i} + B^\beta u_t^{\beta,i}$$
$$\beta_t^i = C^\beta x_t^{\beta,i} \tag{5.24}$$

where the superscripts α and β refer to the tracking and point-ahead mirrors, respectively. In these equations, $x_t^{\alpha,i} \in \mathbb{R}^{n^\alpha}$ and $x_t^{\beta,i} \in \mathbb{R}^{n^\beta}$ are the state vectors, $u_t^{\alpha,i} \in \mathbb{R}^2$ and $u_t^{\beta,i} \in \mathbb{R}^2$ are the control vectors, and the matrices have appropriate dimensions. The linearity of the equations is justified by the fact that the flat mirrors operate over small angles. It is worth remarking here that the actuators of the flat mirrors are fast dynamical systems, so that for an approximate analysis, we can ignore their dynamics and approximate $\alpha_t^i \simeq u_t^{\alpha,i}$ and $\beta_t^i \simeq u_t^{\beta,i}$.

As a summary, we characterize each station as a dynamical system with the state vector $\left(x_t^{p,i}, \Omega_t^i, x_t^{\alpha,i}, x_t^{\beta,i}\right)$, the control vector $\left(u_t^{p,i}, u_t^{\alpha,i}, u_t^{\beta,i}\right)$, and the distur-

bance vector $(\omega_t^i, \rho_t^i, \varepsilon_t^i)$, where the state vector evolves in time according to

$$\begin{aligned}
\dot{x}_t^{p,i} &= f\left(x_t^{p,i}, u_t^{p,i}\right) \\
\dot{\Omega}_t^i &= \left[g\left(x_t^{p,i}\right)\right]_\times \Omega_t^i + \Omega_t^i \left[\omega_t^i\right]_\times \\
\dot{x}_t^{\alpha,i} &= A^\alpha x_t^{\alpha,i} + B^\alpha u_t^{\alpha,i} \\
\dot{x}_t^{\beta,i} &= A^\beta x_t^{\beta,i} + B^\beta u_t^{\beta,i}.
\end{aligned} \tag{5.25}$$

Also, the output vector (y_t^i, ψ_t^i) is expressed in terms of the state and disturbance vectors as

$$\begin{aligned}
y_t^i &= f_c I_* \Omega_t^i \rho_t^i + HC^\alpha x_t^{\alpha,i} \\
\psi_t^i &= I_* \left(I + \left[I_*^T \left(KC^\alpha x_t^{\alpha,i} + LC^\beta x_t^{\beta,i} + \varepsilon_t^i \right) \right]_\times \right) \Omega_t^i \rho_{t+t_d}^i.
\end{aligned} \tag{5.26}$$

5.3.7 Model Summary and Discussion

The mathematical model developed in this section is summarized in the block diagram of Figure 5.5. This block diagram describes a dynamical system with the input vector $\left(u_t^a, u_t^b\right)$, the disturbance vector $\left(\omega_t^a, \omega_t^b, \rho_t^a, \rho_t^b, \varepsilon_t^a, \varepsilon_t^b, \nu_{t-t_d}^a, \nu_{t-t_d}^b\right)$, and the output vector $\left(Y_t^a, Y_t^b\right)$. Here, the output vector $\left(Y_t^a, Y_t^b\right)$ "statistically" depends on the state of the system and the disturbance vector. According to Figure 5.5, except for ω_t^a and ω_t^b which appear in the state-space equations, other elements of the disturbance vector only appear in the output equations. Therefore, ρ_t^i, ε_t^i, and $\nu_{t-t_d}^i$, $i = a, b$ affect the state vector only after establishing closed-loop paths from Y_t^i to u_t^i. Also, we observe from Figure 5.5 that when the feedback loops exist, the

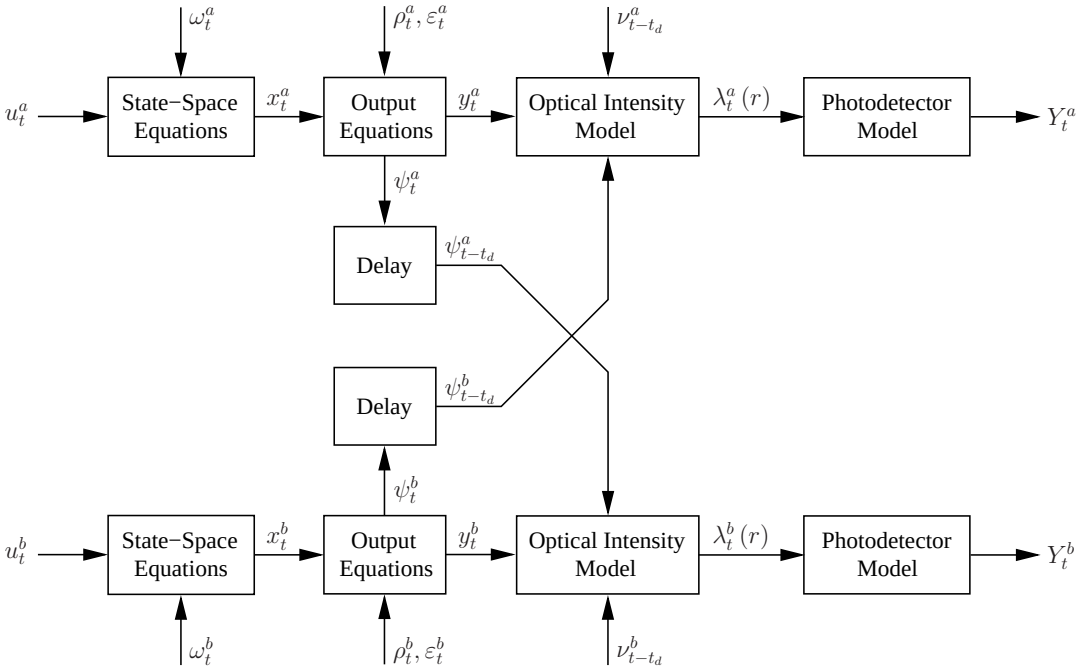

Figure 5.5: Block diagram of a cooperative optical beam tracking system. In this figure, the blocks marked by "State-Space Equations", "Output Equations", and "Optical Intensity Model" refer to (5.25), (5.26), and (5.20), respectively. Also, "Photodetector Model" refers to (5.21) and the vector-valued doubly stochastic Poisson process Y_t^i defined in Section 5.3.5.

subsystems a and b are coupled through (5.20) ("Optical Intensity Model" block in Figure 5.5) and the linear constraint $\rho_t^a + \rho_t^b = 0$.

In a cooperative optical beam tracking system, the goal of the closed-loop control is to maintain some appropriate norm of $\{y_t^i\}$ and $\{\psi_t^i\}$ close to zero. Here, the condition $y_t^i \simeq 0$ is the objective of the optical beam tracking operation, while $\psi_t^i \simeq 0$ is associated with the active pointing. In short range applications with $t_d = 0$, assuming that $\varepsilon_t^i = 0$, the two conditions are equivalent, i.e., the tracking and pointing operations are combined. In long range applications where $t_d \neq 0$, the conditions can be achieved independently, by means of the additional degree of freedom provided by the point-ahead mirrors.

The solution to the control problem above, depends on the structure of the information which is provided to the controllers. In one scenario, the controller of each station has access only to the output of the same station, i.e., the control problem is decentralized. In another possible scheme, the stations share their outputs through the optical link or and independent RF channel. This scheme improves the performance of the system by increasing the information for estimating the common disturbance $\rho_t^a = -\rho_t^b$. On the other hand, sharing the outputs is the only possibility for compensating ε_t^i, since this disturbance vector can be observed only by the receiving station. As an assisting equipment, an arrangement of three (perpendicular) rate gyros and accelerometers which is carried by each station provides relevant information for estimating ω_t^i and ρ_t^i.

Up to this point, we did not offer any model for the disturbance vectors, while such a model is essential for the further analysis. A disturbance model might be deterministic or stochastic, depending on the preferred analysis tools and methods. For a deterministic analysis, the disturbance vectors are modeled using appropriate deterministic functions which are rich enough to represent the family of all possible instances. In this type of analysis, we can also approximate the output vector Y_t^i by (5.22).

A stochastic disturbance model can be characterized by an appropriate stochastic state-space equation driven by a vector-valued Wiener process. Then, a complete stochastic model is constructed by combining this equation with (5.25). This stochastic modeling approach will be considered in Section 5.5.

5.4 Linearizing the Dynamical Equations

In the applications such as intersatellite communication, the relative motion (disturbance) consists of a large, predetermined component and a small, unknown term. Accordingly, the control law consists of an open-loop, coarse control and a small, closed-loop, fine control. In this case, the nonlinear state-space equations (5.25) and the output equations (5.26) can be linearized around a predetermined nominal trajectory, which results in a linear time-varying model for the fine control regime.

In order to obtain the linearized model, every vector contributing in (5.25) and (5.26) will be expressed as the sum of a nominal vector and a (small) deviation vector. For a vector x, the nominal and deviation vectors will be denoted by $\bar{x}$ and δx, respectively. Thus, we express x as $x = \bar{x} + \delta x$. Because the tracking mirrors are not involved in the coarse control regime, we set $\bar{\alpha}_t^i = 0$. In this procedure, ρ_t^i needs a special attention, since it must satisfy the condition $\|\rho_t^i\| = 1$. This norm condition is obviously satisfied for the nominal vector $\bar{\rho}_t^i$ defined as

$$\bar{\rho}_t^a = -\bar{\rho}_t^b = \frac{\bar{r}_t^b - \bar{r}_t^a}{\left\|\bar{r}_t^b - \bar{r}_t^a\right\|} .$$

We define the deviation vectors $\delta\xi_t^a$ and $\delta\xi_t^b$ as

$$\delta\xi_t^a = -\delta\xi_t^b = \frac{\delta r_t^b - \delta r_t^a}{\left\|\bar{r}_t^b - \bar{r}_t^a\right\|} .$$

Then, assuming that $\|\delta\xi_t^i\|^2 \ll 1$, we can show that the condition $\|\bar{\rho}_t^i + \delta\rho_t^i\| = 1$ is

approximately preserved, if the deviation vector $\delta\rho_t^i$ is given by

$$\delta\rho_t^i = \left(I - \bar{\rho}_t^i \left(\bar{\rho}_t^i\right)^T\right) \delta\xi_t^i. \tag{5.27}$$

The rotation matrix Ω_t^i will be expressed as $\Omega_t^i \simeq \left(I + [\delta\phi_t^i]_\times\right)\bar{\Omega}_t^i$, where $\delta\phi_t^i$ is a 3-dimensional vector and $\bar{\Omega}_t^i$ is the nominal rotation matrix satisfying

$$\dot{\bar{\Omega}}_t^i = \left[\bar{\upsilon}_t^i\right]_\times \bar{\Omega}_t^i + \bar{\Omega}_t^i \left[\bar{\omega}_t^i\right]_\times .$$

We can show that $\delta\phi_t^i$ is the solution of the linear differential equation

$$\delta\dot{\phi}_t^i = \left[\bar{\upsilon}_t^i\right]_\times \delta\phi_t^i + \delta\upsilon_t^i + \bar{\Omega}_t^i \delta\omega_t^i.$$

The goal of the open-loop control $\left(\bar{u}_t^{p,i}, \bar{u}_t^{\beta,i}\right)$ is to maintain $\bar{y}_t^i = 0$ and $\bar{\psi}_t^i = 0$. These conditions are respectively equivalent to $\bar{\Omega}_t^i \bar{\rho}_t^i = [0 \quad 0 \quad 1]^T$ and $\bar{\Delta}_t^i \bar{\Omega}_t^i = \bar{\Omega}_{t+t_d}^i$, where $\bar{\Delta}_t^i$ is defined as

$$\bar{\Delta}_t^i = I + \left[I_*^T \left(L\bar{\beta}_t^i + \bar{\varepsilon}_t^i\right)\right]_\times .$$

We can show that $\bar{y}_t^i = 0$ holds, if and only if $\bar{\upsilon}_t^i$ satisfies

$$I_* \bar{\upsilon}_t^i = -I_* \bar{\Omega}_t^i \bar{\omega}_t^i + J I_* \bar{\Omega}_t^i \dot{\bar{\rho}}_t^i \tag{5.28}$$

where the 2×2 matrix J is defined as

$$J = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}.$$

Also, solving the equation $\bar{\Delta}_t^i \bar{\Omega}_t^i = \bar{\Omega}_{t+t_d}^i$ for $\bar{\beta}_t^i$, we find

$$\bar{\beta}_t^i = (JL)^{-1} I_* \bar{\Omega}_{t+t_d}^i \bar{\rho}_t^i - L^{-1} \bar{\varepsilon}_t^i \tag{5.29}$$

which is the condition on $\bar{\beta}_t^i$ that leads to $\bar{\psi}_t^i = 0$. We shall assume that the state-space equations (5.23) and (5.24) allow $\bar{u}_t^{p,i}$ and $\bar{u}_t^{\beta,i}$ to achieve the conditions (5.28) and (5.29).

Let $\bar{x}_t^{p,i}$ denote the state of (5.23) under the nominal control $\bar{u}_t^{p,i}$. We linearize the state-space model (5.23) around $\bar{x}_t^{p,i}$ and $\bar{u}_t^{p,i}$ to obtain

$$\delta \dot{x}_t^{p,i} = A_t^{p,i} \delta x_t^{p,i} + B_t^{p,i} \delta u_t^{p,i}$$

$$\delta \upsilon_t^i = C_t^{p,i} \delta x_t^{p,i}$$

where $A_t^{p,i}$, $B_t^{p,i}$, and $C_t^{p,i}$ are defined as

$$A_t^{p,i} = \frac{\partial f(x,u)}{\partial x}\Big|_{x=\bar{x}_t^{p,i},\, u=\bar{u}_t^{p,i}}$$

$$B_t^{p,i} = \frac{\partial f(x,u)}{\partial u}\Big|_{x=\bar{x}_t^{p,i},\, u=\bar{u}_t^{p,i}}$$

$$C_t^{p,i} = \frac{\partial g(x)}{\partial x}\Big|_{x=\bar{x}_t^{p,i}}.$$

As a summary, we present the linearized version of (5.25) through the following

state-space equations

$$\begin{aligned}
\delta\dot{x}_t^{p,i} &= A_t^{p,i}\delta x_t^{p,i} + B_t^{p,i}\delta u_t^{p,i} \\
\delta\dot{\phi}_t^i &= \left[\bar{v}_t^i\right]_\times \delta\phi_t^i + C_t^{p,i}\delta x_t^{p,i} + \bar{\Omega}_t^i\delta\omega_t^i \\
\delta\dot{x}_t^{\alpha,i} &= A^\alpha\delta x_t^{\alpha,i} + B^\alpha\delta u_t^{\alpha,i} \\
\delta\dot{x}_t^{\beta,i} &= A^\beta\delta x_t^{\beta,i} + B^\beta\delta u_t^{\beta,i}.
\end{aligned} \tag{5.30}$$

Assuming that (5.28) and (5.29) hold, we linearize (5.26) around the nominal trajectory to get

$$\begin{aligned}
y_t^i &= f_c\left(JI_*\delta\phi_t^i + I_*\bar{\Omega}_t^i\delta\xi_t^i\right) + HC^\alpha\delta x_t^{\alpha,i} \\
\psi_t^i &= JI_*\bar{\Delta}_t^i\delta\phi_t^i + I_*\bar{\Omega}_{t+t_d}^i\delta\xi_{t+t_d}^i - \left(I_*\left[\bar{\Omega}_t^i\bar{\rho}_{t+t_d}^i\right]_\times I_*^T\right)\left(KC^\alpha\delta x_t^{\alpha,i} + LC^\beta\delta x_t^{\beta,i} + \delta\varepsilon_t^i\right).
\end{aligned} \tag{5.31}$$

We note that the linearized model is not identical for $i = a, b$, as it depends on the nominal trajectory which is different for stations a and b. When (5.30) is used to describe the transceiver of Figure 5.1, the control vectors $u_t^{\alpha,i}$ and $u_t^{\beta,i}$ are identically zero. For the transceiver of Figure 5.2, since the pointing assembly is used only for the open-loop control, we set $\delta u_t^{p,i} = 0$ in the first equation of (5.30).

5.5 Stochastic Model

As mentioned in Section 5.3.7, the disturbance vector $(\omega_t^i, \rho_t^i, \varepsilon_t^i)$ can be adequately described in a stochastic framework. Since our analysis in Chapter 6 focuses on the linearized model of Section 5.4, the goal of this section is to develop a stochastic

description for the linearized disturbance vectors $\delta\omega_t^i$, $\delta\xi_t^i$, and $\delta\varepsilon_t^i$.

We use a set of stochastic differential equations to model the disturbance vectors $\delta\omega_t^i$, $\delta\xi_t^i$, and $\delta\varepsilon_t^i$. Justified by the assumption that these "deviation" vectors have small norms, we consider a linear structure for the equations, i.e., we describe $\delta\omega_t^i$, $\delta\xi_t^i$, and $\delta\varepsilon_t^i$ by[1]

$$\begin{aligned} dx_t^{\omega,i} &= A_t^{\omega,i} x_t^{\omega,i} dt + D_t^{\omega,i} dw_t^{\omega,i} \\ \delta\omega_t^i &= C_t^{\omega,i} x_t^{\omega,i}, \end{aligned} \tag{5.32a}$$

$$\begin{aligned} dx_t^{\xi,i} &= A_t^{\xi} x_t^{\xi,i} dt + D_t^{\xi} dw_t^{\xi,i} \\ \delta\xi_t^i &= C_t^{\xi} x_t^{\xi,i}, \end{aligned} \tag{5.32b}$$

$$\begin{aligned} dx_t^{\varepsilon,i} &= A_t^{\varepsilon,i} x_t^{\varepsilon,i} dt + D_t^{\varepsilon,i} dw_t^{\varepsilon,i} \\ \delta\varepsilon_t^i &= C_t^{\varepsilon,i} x_t^{\varepsilon,i} \end{aligned} \tag{5.32c}$$

where $x_t^{\omega,i} \in \mathbb{R}^{n^\omega}$, $x_t^{\xi,i} \in \mathbb{R}^{n^\xi}$, and $x_t^{\varepsilon,i} \in \mathbb{R}^{n^\varepsilon}$ are the state vectors and $w_t^{\omega,i} \in \mathbb{R}^{p^\omega}$, $w_t^{\xi,i} \in \mathbb{R}^{p^\xi}$, and $w_t^{\varepsilon,i} \in \mathbb{R}^{p^\varepsilon}$ are vector-valued standard Wiener processes. Here, the matrices are uniformly bounded and have appropriate dimensions. The initial states $x_0^{\omega,i}$, $x_0^{\xi,i}$, and $x_0^{\varepsilon,i}$ are Gaussian random vectors with known mean and covariance matrix. Since $\delta\omega_t^i$, $\delta\xi_t^i$, and $\delta\varepsilon_t^i$ have independent physical origins, the initial states $x_0^{\omega,i}$, $x_0^{\xi,i}$, and $x_0^{\varepsilon,i}$ and the Wiener processes $\left\{w_t^{\omega,i}\right\}$, $\left\{w_t^{\xi,i}\right\}$, and $\left\{w_t^{\varepsilon,i}\right\}$ are assumed to be statistically independent. Also, the initial states and Wiener processes associated with stations a and b are statistically independent,

[1]The equations are defined in the Itô sense.

except for the pairs $\left(x_0^{\xi,a}, x_0^{\xi,b}\right)$ and $\left(w_t^{\xi,a}, w_t^{\xi,b}\right)$ that must satisfy $x_0^{\xi,a} + x_0^{\xi,b} = 0$ and $w_t^{\xi,a} + w_t^{\xi,b} = 0,\ t \geqslant 0$, in order to hold $\delta\xi_t^a + \delta\xi_t^b = 0$.

The state-space equations (5.30) and (5.32) and the output equations (5.31) can be combined in a compact form[1]

$$dx_t^i = A_t^i x_t^i + B_t^i u_t^i dt + D_t^i dw_t^i \tag{5.33a}$$

$$y_t^i = C_t^i x_t^i \tag{5.33b}$$

$$\psi_t^i = \left(\bar{\psi}/2\right) L_t^i z_t^i \tag{5.33c}$$

where

$$\begin{aligned}
x_t^i &= \begin{bmatrix} \delta x_t^{p,i} & \delta\phi_t^i & \delta x_t^{\alpha,i} & \delta x_t^{\beta,i} & x_t^{\omega,i} & x_t^{\xi,i} & x_t^{\varepsilon,i} \end{bmatrix}^T \\
z_t^i &= \begin{bmatrix} \delta x_t^{p,i} & \delta\phi_t^i & \delta x_t^{\alpha,i} & \delta x_t^{\beta,i} & x_t^{\omega,i} & x_{t+t_d}^{\xi,i} & x_t^{\varepsilon,i} \end{bmatrix}^T \\
u_t^i &= \begin{bmatrix} \delta u_t^{p,i} & \delta u_t^{\alpha,i} & \delta u_t^{\beta,i} \end{bmatrix}^T \\
w_t^i &= \begin{bmatrix} w_t^{\omega,i} & w_t^{\xi,i} & w_t^{\varepsilon,i} \end{bmatrix}^T
\end{aligned} \tag{5.34}$$

and the block matrices $\{A_t^i, B_t^i, D_t^i, C_t^i, L_t^i\}$ can be obtained in terms of the matrices appearing in (5.30), (5.31), and (5.32). Here, the initial state x_0^i is a Gaussian random vector with mean $\bar{x}_0^i$ and covariance matrix $\bar{\Sigma}_0^i$. Considering (5.34), we can determine matrices Π and Π^d such that

$$z_t^i = \Pi x_t^i + \Pi^d x_{t+t_d}^i. \tag{5.35}$$

[1]The notation u_t^i which is used here for the linearized model should not be confused with the same notation used in Section 5.3.6 for the nonlinear case.

Also, we note that the linear constraints $x_0^{\xi,a} + x_0^{\xi,b} = 0$ and $w_t^{\xi,a} + w_t^{\xi,b} = 0$ can be expressed as

$$\begin{aligned} \Upsilon^x\big(x_0^a + x_0^b\big) &= 0 \\ \Upsilon^w\big(w_t^a + w_t^b\big) &= 0 \end{aligned} \tag{5.36}$$

with properly defined matrices Υ^x and Υ^w.

In the following, we discuss some properties of matrices B_t^i, C_t^i, and L_t^i which will be used later in Chapter 6. Assume that $\delta\varepsilon_t^i$ is identically zero. Then, regarding the transceiver of Figure 5.1 for which $\delta x_t^{\alpha,i} = 0$, $\delta x_t^{\beta,i} = 0$, and $t_d = 0$, we observe from (5.31) that $y_t^i = f_c\psi_t^i$. Comparing this result with (5.33b) and (5.33c) we find that

$$L_t^i = 2\left(f_c\bar{\psi}\right)^{-1} C_t^i. \tag{5.37}$$

This result also holds for the transceiver of Figure 5.2, if in addition to $t_d = 0$ and $\delta\varepsilon_t^i \equiv 0$, we have $H + f_cK = 0$. Regarding B_t^i, we can verify that this matrix satisfies $\Pi^d B_t^i = 0$ for every $t \geqslant 0$. This follows from the fact that $x_t^{\xi,i}$ does not depend on the control vector u_t^i.

The space-time rates (5.20) can be expressed in terms of x_t^i and z_t^i as

$$\begin{aligned} \lambda_t^a(r) &= \nu_{t-t_d}^a \exp\left(-\tfrac{1}{2}\left\|L_{t-t_d}^b z_{t-t_d}^b\right\|^2\right)\gamma\big(r - C_t^a x_t^a\big) + \bar{\lambda} \\ \lambda_t^b(r) &= \nu_{t-t_d}^b \exp\left(-\tfrac{1}{2}\left\|L_{t-t_d}^a z_{t-t_d}^a\right\|^2\right)\gamma\big(r - C_t^b x_t^b\big) + \bar{\lambda}. \end{aligned} \tag{5.38}$$

We allow $\left\{\nu_t^a\right\}$ and $\left\{\nu_t^b\right\}$ to be nonnegative stochastic processes with piecewise

continuous and bounded sample paths and nonzero mean[1]. Further, we assume that $\left\{\nu_t^a\right\}$ and $\left\{\nu_t^b\right\}$ are mutually independent and are independent of x_0^i and $\left\{w_t^i\right\},\ i = a, b.$

[1]This characterizes the effect of random optical fade and data modulation.

Chapter 6

Cooperative Optical Beam Tracking: Optimal Control

6.1 Introduction

In Chapter 5, we discussed the concept of cooperative optical beam tracking and developed a mathematical model for this alignment scheme. In the next step, the dynamical equations associated with this model were linearized around a nominal trajectory and a stochastic description for the disturbance vectors was introduced. In the present chapter, we use this linearized stochastic model in order to study the problem of controller design for a cooperative optical beam tracking system. The design goal is to maximize the flow of optical energy between the stations. We shall study the problem (separately) for two scenarios: short range applications with $t_d = 0$ and long range applications with $t_d > 0$.

The organization of this chapter is as follows. In Section 6.2, we first introduce some additional assumptions on the model of Chapter 5 and modify it accordingly, and then define its associated control problem. Section 6.3 considers an associated estimation problem and presents an approximate solution for it. This solution will be used in Sections 6.4 and 6.5 in order to develop two different methods for solving

the control problem for $t_d = 0$ and $t_d > 0$, respectively.

6.2 Model and Problem Statement

In order to use the results of Theorem 3.3.2 in our analysis, we shall assume that $\gamma(r)$ in (5.38) has a Gaussian profile (see Section 5.3.3) and ignore the effect of the dark current and the background noises by letting $\bar{\lambda} = 0$ in (5.38). In addition, we use the "infinite resolution" and "infinite area" model for the photodetectors as explained in Section 3.5. These assumptions enable us to describe the output of a position-sensitive photodetector by a space-time point process over $\mathbb{R}^2$. After solving the control problem using this idealized model, we can apply the approximation method of Section 3.5 to modify the solution for a practical finite resolution photodetector.

6.2.1 The Model

Following the model of Section 5.5, we describe the dynamics of station $i = a, b$ by the state-space equation

$$dx_t^i = A_t^i x_t^i dt + B_t^i u_t^i dt + D_t^i dw_t^i \tag{6.1}$$

where $x_t^i \in \mathbb{R}^n$ is the state vector and $\{w_t^i\}$ is a p-dimensional standard Wiener process. The initial state x_0^i is independent of $\{w_t^a\}$ and $\{w_t^b\}$ and is assumed to be a Gaussian vector with mean $\bar{x}_0^i = 0$ and covariance matrix P_0^i. We remind from Section 5.5 that $x_0^i,\ i = a, b$ and $\{w_t^i\},\ i = a, b$ satisfy the linear constraints (5.36). The control vector u_t^i will be discussed separately for $t_d = 0$ and $t_d > 0$ in Section 6.2.2.

The observation at station $i = a, b$ which is provided over $\mathbb{R}^2$ is the space-time point process $N^i(\mathcal{T} \times \mathcal{S})$ with rate

$$\lambda^i\left(r, x_t^i, \mu_t^i\right) = \mu_t^i \Phi_2\left(r; C_t^i x_t^i, R\right) \tag{6.2}$$

where $R = 2\left(f_c/k\varrho\right)^2 I_{2\times 2}$ is defined in Section 5.3.3 and the Gaussian map $\Phi_2(\cdot)$ is given by (3.5). The stochastic processes $\left\{\mu_t^a\right\}$ and $\left\{\mu_t^b\right\}$ are defined as

$$\begin{aligned} \mu_t^a &= \nu_{t-t_d}^a \exp\left(-\tfrac{1}{2}\left\|L_{t-t_d}^b z_{t-t_d}^b\right\|^2\right) \\ \mu_t^b &= \nu_{t-t_d}^b \exp\left(-\tfrac{1}{2}\left\|L_{t-t_d}^a z_{t-t_d}^a\right\|^2\right) \end{aligned} \tag{6.3}$$

where the nonnegative stochastic process $\{\nu_t^i\}$ is statistically independent of x_0^i and $\{w_t^i\}$, $i = a, b$ and has piecewise continuous and bounded sample paths and nonzero mean. The stochastic vector z_t^i is determined in terms of the state vector x_t^i according to[1]

$$z_t^i = \Pi x_t^i + \Pi^d x_{t+t_d}^i. \tag{6.4}$$

We shall assume that prior to $t = 0$, the stations are under the open-loop control $u_t^a = u_t^b = 0$. This implies that for $t \in [-t_d, 0]$, x_t^i is a zero-mean Gaussian random vector with covariance matrix P_t^i which satisfies the matrix differential equation

$$\dot{P}_t^i = A_t^i P_t^i + P_t^i A_t^{iT} + D_t^i D_t^{iT}.$$

Note that due to the propagation delay t_d, the initial state of the system must be

[1] See (5.35) in Section 5.5.

given over the interval $t \in [-t_d, 0]$, instead of a single point $t = 0$.

Over the underlying probability space $(\Omega, \mathscr{F}, P)$ of the above stochastic model, we define $\mathscr{B}_t^i,\ i = a, b$ as the σ-algebra generated by the space-time point process i over $[0, t)$. We say u_t^a and u_t^b are admissible controls if u_t^i is $\mathscr{B}_t^i$-measurable and the solution to (6.1) is well defined for $i = a, b$.

6.2.2 Problem Statement

For long range applications using the transceiver of Figure 5.2, the control u_t^i is a 4-dimensional vector comprised of $u_t^{\alpha,i} \in \mathbb{R}^2$ and $u_t^{\beta,i} \in \mathbb{R}^2$ which are the control vectors associated with the tracking mirror and the point-ahead mirror, respectively. In terms of $u_t^{\alpha,i}$ and $u_t^{\beta,i}$, the state-space equation (6.1) can be expressed as

$$dx_t^i = A_t^i x_t^i dt + B_t^{\alpha,i} u_t^{\alpha,i} dt + B_t^{\beta,i} u_t^{\beta,i} dt + D_t^i dw_t^i \tag{6.5}$$

where $B_t^{\alpha,i}$ and $B_t^{\beta,i}$ are defined such that $B_t^i = \begin{bmatrix} B_t^{\alpha,i} & B_t^{\beta,i} \end{bmatrix}$. Note that the control $u_t^{\alpha,i}$ is employed for the purpose of tracking, i.e., maintaining $\|C_t^i x_t^i\|$ close to 0, while $u_t^{\beta,i}$ is used for precise pointing, i.e., keeping $\|L_t^i z_t^i\|$ as small as possible. We observe from (5.30) and (5.31) that $C_t^i x_t^i$ does not depend on $u_t^{\beta,i}$, thus the tracking control $u_t^{\alpha,i}$ can be designed subject to (6.5) with $u_t^{\beta,i} = 0$. The design problem can be formulated in terms of minimizing the cost functional (3.6) with $Q_t = C_t^{iT} C_t^i$, $P_t = \eta I_{2\times 2}$, $\eta > 0$, and $S = 0$ as explained in Section 3.5. A simpler scheme is to obtain $u_t^{\alpha,i}$ such that $\mathrm{E}\left[C_t^i x_t^i | \mathscr{B}_t^i\right] = 0$. This scheme needs some additional assumptions which will be explained later through Lemma 6.4.1.

After determining an admissible $u_t^{\alpha,i}$, we solve the pointing control problem subject to (6.5) with the already obtained $u_t^{\alpha,i}$. For the nonnegative constants k^a and k^b and the fixed time horizon $T > t_d \geqslant 0$, we define the objective functional

$$J = \mathrm{E}\left[\int_0^T \left(k^a \nu_t^a \exp\left(-\tfrac{1}{2}\left\|L_t^b z_t^b\right\|^2\right) + k^b \nu_t^b \exp\left(-\tfrac{1}{2}\left\|L_t^a z_t^a\right\|^2\right)\right) dt\right] \tag{6.6}$$

as a linear combination of the expected optical energy received by the stations during $[t_d, T + t_d]$. Then, the pointing control problem can be defined as follows. Subject to (6.5), determine the admissible controls $\left\{u_t^{\beta,a},\ t \in [0,T]\right\}$ and $\left\{u_t^{\beta,b},\ t \in [0,T]\right\}$ that maximize the objective functional (6.6).

For short range applications with $t_d = 0$, the control vector $u_t^i \in \mathbb{R}^2$ is associated with either the pointing assembly or the tracking mirror. In this case, the control problem is to obtain the admissible controls $\left\{u_t^a,\ t \in [0,T]\right\}$ and $\left\{u_t^b,\ t \in [0,T]\right\}$ that maximize the objective functional

$$J = \mathrm{E}\left[\int_0^T \left(k^a \nu_t^a \exp\left(-\tfrac{1}{2}\left\|L_t^b x_t^b\right\|^2\right) + k^b \nu_t^b \exp\left(-\tfrac{1}{2}\left\|L_t^a x_t^a\right\|^2\right)\right)\right]. \tag{6.7}$$

We note that for the case of $t_d = 0$, (6.3) can be simplified as

$$\begin{aligned} \mu_t^a &= \nu_t^a \exp\left(-\tfrac{1}{2}\left\|L_t^b x_t^b\right\|^2\right) \\ \mu_t^b &= \nu_t^b \exp\left(-\tfrac{1}{2}\left\|L_t^a x_t^a\right\|^2\right). \end{aligned} \tag{6.8}$$

6.3 Estimation Problem

The first step in solving the control problem is to determine the posterior density $p_{x_t^i}(x|\mathscr{B}_t^i)$. While the structure of the model for a single station[1] is similar to the model of Chapter 3, this model does not completely satisfy the assumptions of Theorem 3.3.2, due to the statistical dependence between $\{\mu_t^i\}$ and $\{w_t^i\}$. This dependence is manifested by (6.3) which indicates that $\{\mu_t^i\}$ depends on the state of the other station, and as a consequence, the statistical dependence between $\left\{x_t^a\right\}$ and $\left\{x_t^b\right\}$ leads to statistical dependence between $\{\mu_t^i\}$ and $\{w_t^i\}$. The dependence between $\left\{x_t^a\right\}$ and $\left\{x_t^b\right\}$ has two origins: the linear constraints (5.36), and the dependence of $\{x_t^i\}$ on $\mathscr{B}_t^i$ through u_t^i, noting that $\mathscr{B}_t^i$ depends on the state of the other station.

The discussion above suggests that the coupling between the stations must be involved in an exact solution for the estimation problem. Such an estimation problem is difficult to solve, not only due to the complexity of the model, but also because of the requirement of determining the optimal controls u_t^a and u_t^b, prior to solving the estimation problem. A suboptimal solution for the estimation problem which avoids these difficulties can be obtained by applying Theorem 3.3.2 to each individual station, ignoring the statistical dependence between $\{\mu_t^i\}$ and $\{w_t^i\}$. To justify this approximation, we note that for a well-designed system, the expected

[1]This model is consisted of the state-space equation (6.1) and the observation of the space-time point process with rate (6.2).

value of the stochastic process

$$\rho_t^i \triangleq \exp\left(-\tfrac{1}{2}\left\|L_t^i z_t^i\right\|^2\right)$$

must stay close to 1 for every $t \in [0,T]$. This implies that the standard deviation of ρ_t^i is small compared to its expected value, i.e., we can approximate

$$\rho_t^i \simeq \mathrm{E}\left[\rho_t^i\right]. \tag{6.9}$$

Based on (6.9), we approximate the stochastic processes $\left\{\mu_t^a\right\}$ and $\left\{\mu_t^b\right\}$ by $\left\{\nu_{t-t_d}^a \mathrm{E}\left[\rho_t^b\right]\right\}$ and $\left\{\nu_{t-t_d}^b \mathrm{E}\left[\rho_t^a\right]\right\}$, respectively. Since $\{\nu_t^i\}$ is statistically independent of x_0^i and $\{w_t^i\}$, these approximations imply that $\{\mu_t^i\}$ is statistically independent of x_0^i and $\{w_t^i\}$. Thus, using Theorem 3.3.2, we approximate

$$p_{x_t^i}\left(x|\mathscr{B}_t^i\right) \simeq \tilde{p}_{x_t^i}\left(x|\mathscr{B}_t^i\right) = \Phi_n\left(x; \hat{x}_t^i, \Sigma_t^i\right) \tag{6.10}$$

where $\hat{x}_t^i$ and Σ_t^i are the solutions of the stochastic differential equations

$$d\hat{x}_t^i = A_t^i \hat{x}_t^i dt + B_t^i u_t^i dt + \int_{\mathbb{R}^2} M_t^i\left(r - C_t^i \hat{x}_t^i\right) N^i\left(dt \times dr\right) \tag{6.11a}$$

$$d\Sigma_t^i = A_t^i \Sigma_t^i dt + \Sigma_t^i A_t^{iT} dt + D_t^i D_t^{iT} dt - M_t^i C_t^i \Sigma_t^i dN_t^i \tag{6.11b}$$

with the initial states $\hat{x}_0^i = \bar{x}_0^i = 0$ and $\Sigma_0^i = P_0^i$. Here, the stochastic process $\{N_t^i\}$

and the stochastic matrix M_t^i are defined as

$$N_t^i = N^i\left([0,t)\times\mathbb{R}^2\right)$$
$$M_t^i = \Sigma_t^i C_t^{iT}\left(C_t^i\Sigma_t^i C_t^{iT}+R\right)^{-1}.$$

From the definition of $N^i\left(\mathcal{T}\times\mathcal{S}\right)$, we know that $\{N_t^i\}$ is a doubly stochastic Poisson process with rate $\{\mu_t^i\}$ which is defined by (6.3) (or by (6.8) for the short range applications).

Note that the estimator (6.11) does not explicitly depend on $\{\mu_t^i, t\geqslant 0\}$; however, the estimates $\hat{x}_t^i$ and Σ_t^i depend on $\{\mu_t^i, t\geqslant 0\}$ through the space-time point process. This dependence can be explained by observing from (6.11b) that the occurrence of each event in the space-time point process decreases Σ_t^i by the positive definite matrix $M_t^i C_t^i \Sigma_t^i$. Thus, a larger μ_t^i leads to a smaller estimation error by increasing the occurrence rate of the events. According to (6.3), a smaller pointing error $\|L_{t-t_d}^b z_{t-t_d}^b\|$ at station b results in a larger μ_t^a and, as a consequence, a closer estimate for x_t^a, which in turn, leads to a smaller pointing error at station a. This explains the mechanism which couples the dynamics of the stations.

The posterior density $p_{z_t^i}\left(z|\mathscr{B}_t^i\right)$ can be obtained in terms of $p_{x_t^i}\left(x|\mathscr{B}_t^i\right)$ as stated in the following theorem.

Theorem 6.3.1. Consider the state-space equation (6.1) and its associated space-time observation with the rate process (6.2). Assume that the increasing family of σ-algebras $\mathscr{B}_t^i$ are given and that u_t^i is $\mathscr{B}_t^i$-measurable. Then, for z_t^i defined by (6.4),

the posterior density $p_{z_t^i}\left(z|\mathscr{B}_t^i\right)$ is given by

$$p_{z_t^i}\left(z|\mathscr{B}_t^i\right) = \int_{\mathbb{R}^n} p_{x_t^i}\left(x|\mathscr{B}_t^i\right) \Phi_n\left(z; H_t x, V_t\right) dx \tag{6.12}$$

where H_t and V_t are defined as[1]

$$H_t = \Pi + \Pi^d \Phi^{A^i}\left(t + t_d, t\right)$$

and

$$V_t = \int_t^{t+t_d} \left(\Pi^d \Phi^{A^i}\left(t + t_d, \tau\right) D_\tau^i\right) \left(\Pi^d \Phi^{A^i}\left(t + t_d, \tau\right) D_\tau^i\right)^T d\tau.$$

Here, $\Phi^{A^i}\left(t, \tau\right)$ is the transition matrix associated with A_t^i.

Proof. We know from Section 5.5 that $\Pi^d B_t^i = 0$ for every $t \geqslant 0$. Thus, solving the linear stochastic differential equation (6.1) from t to $t + t_d$, we get

$$\Pi^d x_{t+t_d}^i = \Pi^d \Phi^{A^i}\left(t + t_d, t\right) x_t^i + \Pi^d v_t^i$$

where

$$v_t^i = \int_t^{t+t_d} \Phi^{A^i}\left(t + t_d, \tau\right) D_\tau^i dw_\tau^i.$$

From the definition of z_t^i and H_t we can write

$$z_t^i = H_t x_t^i + \Pi^d v_t^i.$$

[1]We know from Section 5.5 that H_t and V_t are identical for stations a and b, thus the superscript i has been dropped.

The conditional characteristic function of z_t^i given $\mathscr{B}_t^i$ is given by

$$\begin{aligned}
\mathrm{E}\left[\exp\left(j\omega^T z_t^i\right)|\mathscr{B}_t^i\right] &= \mathrm{E}\left[\exp\left(j\omega^T H_t x_t^i\right)\exp\left(j\omega^T \Pi^d v_t^i\right)\big|\mathscr{B}_t^i\right] \\
&= \mathrm{E}\left[\exp\left(j\omega^T H_t x_t^i\right)|\mathscr{B}_t^i\right]\mathrm{E}\left[\exp\left(j\omega^T \Pi^d v_t^i\right)\right] \\
&= \mathrm{E}\left[\exp\left(j\omega^T H_t x_t^i\right)|\mathscr{B}_t^i\right]\exp\left(-\tfrac{1}{2}\,\omega^T V_t \omega\right).
\end{aligned}$$

Taking the inverse Fourier transform of the last expression above, we obtain (6.12).

□

Using Theorem 6.3.1 and Lemma 4.5.1, we determine the approximation of $p_{z_t^i}\left(z|\mathscr{B}_t^i\right)$ associated with (6.10) as

$$\tilde{p}_{z_t^i}\left(z|\mathscr{B}_t^i\right) = \Phi_n\left(z; \hat{z}_t^i, H_t\Sigma_t^i H_t^T + V_t\right) \tag{6.13}$$

where $\hat{z}_t^i$ is defined as $\hat{z}_t^i = H_t\hat{x}_t^i$. For future reference, we use (6.13) to obtain

$$\int_{\mathbb{R}^n} \tilde{p}_{z_t^i}\left(z|\mathscr{B}_t^i\right)\exp\left(-\tfrac{1}{2}\,\left\|L_t^i z\right\|^2\right)dz = f_t^i\left(\Sigma_t^i\right)\exp\left(-\tfrac{1}{2}\,\left\|Q_t^i L_t^i \hat{z}_t^i\right\|^2\right) \tag{6.14}$$

where $f_t^i(\cdot) : \mathbb{R}^{n\times n} \to \mathbb{R}$ is defined as

$$f_t^i(X) = \left[\det\left(I_{2\times 2} + L_t^i\left(H_t X H_t^T + V_t\right)L_t^{iT}\right)\right]^{-1/2} \tag{6.15}$$

and the positive definite stochastic matrix Q_t^i is given by[1]

$$Q_t^i = \left(I_{2\times 2} + L_t^i \left(H_t \Sigma_t^i H_t^T + V_t\right) L_t^{iT}\right)^{-1/2}. \tag{6.16}$$

For the case of $t_d = 0$, in which $H_t = I_{n\times n}$ and $V_t = 0$, (6.15) and (6.16) are simplified as

$$f_t^i(X) = \left[\det\left(I_{2\times 2} + L_t^i X L_t^{iT}\right)\right]^{-1/2} \tag{6.17}$$

and

$$Q_t^i = \left(I_{2\times 2} + L_t^i \Sigma_t^i L_t^{iT}\right)^{-1/2}. \tag{6.18}$$

6.4 Control Problem: Short Range Applications

We exploit the estimator (6.11) and the approximate posterior density (6.10) in order to prove Theorem 6.4.1 below which is the basis for developing a suboptimal control for the case of $t_d = 0$. Before stating the theorem, we fix notation. Let $g_t\left(\Sigma^a, \Sigma^b\right)$ be a scalar function of $n \times n$ symmetric matrices Σ^a and Σ^b. Assume that the partial derivatives of $g_t\left(\Sigma^a, \Sigma^b\right)$ with respect to the elements of Σ^a and Σ^b exist. We denote by $\partial g_t\left(\Sigma^a, \Sigma^b\right)/\partial\Sigma^i,\ i = a, b$ a $n \times n$ symmetric matrix with diagonal elements $\partial g_t/\partial\sigma_{kk}^i$ and off-diagonal elements $(1/2)\,\partial g_t/\partial\sigma_{kl}^i$, where σ_{kl}^i is the element of Σ^i at k^{th} row and l^{th} column. We define the linear operators $\mathcal{L}_t^a\{\cdot\}$ and $\mathcal{L}_t^b\{\cdot\}$ as

$$\begin{aligned}
\mathcal{L}_t^a\left\{g_t\left(\Sigma^a, \Sigma^b\right)\right\} &= g_t\left(S_t^a\left(\Sigma^a\right), \Sigma^b\right) - g_t\left(\Sigma^a, \Sigma^b\right) \\
\mathcal{L}_t^b\left\{g_t\left(\Sigma^a, \Sigma^b\right)\right\} &= g_t\left(\Sigma^a, S_t^b\left(\Sigma^b\right)\right) - g_t\left(\Sigma^a, \Sigma^b\right)
\end{aligned} \tag{6.19}$$

[1]Here, by $X = Y^{-1/2}$, we mean a matrix X that satisfies $X^T X = Y^{-1}$.

where $S_t^i(\cdot)$ is given by

$$S_t^i(\Sigma) = \Sigma - \Sigma C_t^{iT}\left(C_t^i \Sigma C_t^{iT} + R\right)^{-1} C_t^i \Sigma. \tag{6.20}$$

Theorem 6.4.1. Fix sample paths ν_t^a and ν_t^b, $t \in [0,T]$. Let Σ^a and Σ^b be $n \times n$ symmetric matrices and assume that $g_t\left(\Sigma^a, \Sigma^b\right)$, $t \in [0,T]$ is the solution of the partial differential equation

$$\begin{aligned}
&\frac{\partial g_t\left(\Sigma^a, \Sigma^b\right)}{\partial t} + \nu_t^a f_t^b\left(\Sigma^b\right)\left(k^a + \mathcal{L}_t^a\left\{g_t\left(\Sigma^a, \Sigma^b\right)\right\}\right) \\
&\qquad + \nu_t^b f_t^a\left(\Sigma^a\right)\left(k^b + \mathcal{L}_t^b\left\{g_t\left(\Sigma^a, \Sigma^b\right)\right\}\right) \\
&\qquad + \operatorname{tr}\left\{\frac{\partial g_t\left(\Sigma^a, \Sigma^b\right)}{\partial \Sigma^a}\left(A_t^a \Sigma^a + \Sigma^a A_t^{aT} + D_t^a D_t^{aT}\right)\right\} \\
&\qquad + \operatorname{tr}\left\{\frac{\partial g_t\left(\Sigma^a, \Sigma^b\right)}{\partial \Sigma^b}\left(A_t^b \Sigma^b + \Sigma^b A_t^{bT} + D_t^b D_t^{bT}\right)\right\} = 0
\end{aligned} \tag{6.21}$$

with the boundary condition $g_T(\cdot,\cdot) = 0$, where $f_t^i(\cdot)$ is defined by (6.17). Let $\{\Sigma_t^i,\ t \in [0,T]\}$, $i = a, b$ be the solution of the stochastic differential equation (6.11b) with the initial state P_0^i. Then, for the fixed sample paths ν_t^a and ν_t^b, the objective

functional (6.7) can be expressed as

$$
\begin{aligned}
J = & \, g_0\left(P_0^a, P_0^b\right) \\
& + \int_0^T \mathrm{E}\left[\nu_t^a \delta_t^b \Big(k^a + \mathcal{L}_t^a\left\{g_t\left(\Sigma_t^a, \Sigma_t^b\right)\right\}\Big) + \nu_t^b \delta_t^a \Big(k^b + \mathcal{L}_t^b\left\{g_t\left(\Sigma_t^a, \Sigma_t^b\right)\right\}\Big)\right] dt \\
& - \int_0^T \mathrm{E}\left[\nu_t^a f_t^b\left(\Sigma_t^b\right)\Big(k^a + \mathcal{L}_t^a\left\{g_t\left(\Sigma_t^a, \Sigma_t^b\right)\right\}\Big)\left\{1 - \exp\left(-\tfrac{1}{2}\left\|Q_t^b L_t^b \hat{x}_t^b\right\|^2\right)\right\}\right] dt \\
& - \int_0^T \mathrm{E}\left[\nu_t^b f_t^a\left(\Sigma_t^a\right)\Big(k^b + \mathcal{L}_t^b\left\{g_t\left(\Sigma_t^a, \Sigma_t^b\right)\right\}\Big)\left\{1 - \exp\left(-\tfrac{1}{2}\left\|Q_t^a L_t^a \hat{x}_t^a\right\|^2\right)\right\}\right] dt
\end{aligned}
\tag{6.22}
$$

where $Q_t^i,\ i = a, b$ is given by (6.18) and the error term $\delta_t^i,\ i = a, b$ is defined as

$$
\delta_t^i = \exp\left(-\tfrac{1}{2}\ \left\|L_t^i x_t^i\right\|^2\right) - f_t^i\left(\Sigma_t^i\right) \exp\left(-\tfrac{1}{2}\ \left\|Q_t^i L_t^i \hat{x}_t^i\right\|^2\right). \tag{6.23}
$$

Moreover, if Σ^a and Σ^b are positive definite, $\mathcal{L}_t^a\left\{g_t\left(\Sigma^a, \Sigma^b\right)\right\}$ and $\mathcal{L}_t^b\left\{g_t\left(\Sigma^a, \Sigma^b\right)\right\}$ are nonnegative for every $t \in [0, T]$.

Proof. See Section 6.6. □

In the following, we use the results of Theorem 6.4.1 to develop a suboptimal solution for the control problem. Clearly, the first term on the right side of (6.22) is not involved in the optimization, since it does not depend on u_t^a and u_t^b. Even though the hard-to-compute error terms δ_t^a and δ_t^b do depend on u_t^a and u_t^b, they are small, at least under the suboptimal control that will be obtained. Therefore, in maximizing (6.22), we ignore the second term on the right side and maximize $\tilde{J}$ which is defined as the sum of the third and the fourth terms.

Assuming that $k^i > 0,\ i = a, b$, for the fixed sample paths ν_t^a and ν_t^b, we

have $\tilde{J} \leqslant 0$, with equality holds if and only if $L_t^a \hat{x}_t^a = 0$ for almost every $t \in \{\tau \,|\, \nu_\tau^b \neq 0,\, \tau \in [0,T]\}$ and $L_t^b \hat{x}_t^b = 0$ for almost every $t \in \{\tau \,|\, \nu_\tau^a \neq 0,\, \tau \in [0,T]\}$. Assuming that the stochastic processes $\{\nu_t^a\}$ and $\{\nu_t^b\}$ have nonzero sample paths during $t \in [0,T]$, the condition $\tilde{J} = 0$ holds for all sample paths of $\{\nu_t^a\}$ and $\{\nu_t^b\}$, if and only if for almost every $t \in [0,T]$ we have

$$L_t^a \hat{x}_t^a = L_t^b \hat{x}_t^b = 0. \tag{6.24}$$

Note that this condition is sufficient for $\tilde{J} = 0$, even if the assumption above does not hold, i.e., with a nonzero probability, some of the sample paths of $\{\nu_t^a\}$ and $\{\nu_t^b\}$ are identically zero over an interval $\mathscr{I} \subseteq [0,T]$.

Assuming that under the condition (6.24), the error term (second term) on the right side of (6.22) is ignorable, for fixed sample paths ν_t^a and ν_t^b, the maximum of J can be approximated by $J^* \simeq g_0\big(P_0^a, P_0^b\big)$. In order to generalize this result to the stochastic processes $\{\nu_t^a\}$ and $\{\nu_t^b\}$, we need to average $g_0\big(P_0^a, P_0^b\big)$ over all sample paths of $\{\nu_t^a\}$ and $\{\nu_t^b\}$. For this purpose, consider (6.21) with the sample path ν_t^i replaced with the stochastic process $\{\nu_t^i\},\ i = a, b$ and let the random variable $g_0\big(P_0^a, P_0^b\big)$ be the solution of this stochastic differential equation at $t = 0$. Then, the maximum of J can be approximated as

$$J^* = \mathrm{E}\left[g_0\big(P_0^a, P_0^b\big)\right].$$

Remark 6.4.1. The condition (6.24) which leads to the maximum of $\tilde{J}$ does not

depend on k^a and k^b. In particular, the sufficient condition is the same for $(k^a, k^b) = (1, 0)$ and $(k^a, k^b) = (0, 1)$, which means that under (6.24), both stations approximately receive the maximum possible optical energy. In other words, the stations are not required to pay any cost in order to increase the payoff of the other station, which indicates a cooperative relationship between them.

Remark 6.4.2. According to (6.11), $\hat{x}_t^a$ and $\hat{x}_t^b$ do not explicitly depend on $\left\{\nu_t^a\right\}$ and $\left\{\nu_t^b\right\}$, thus the control law which leads to (6.24) does not explicitly depend on $\left\{\nu_t^a\right\}$ and $\left\{\nu_t^b\right\}$. This is important in particular when a reliable model for $\{\nu_t^i\}$ does not exist.

The following lemma determines a control law u_t^i which achieves (6.24).

Lemma 6.4.1. Consider the stochastic differential equation (6.11a) and assume that $L_0^i \hat{x}_0^i = 0$. Let $L_t^i B_t^i$ be nonsingular and L_t^i be differentiable for $t \geqslant 0$. Then, under the control

$$u_t^i dt = -\left(L_t^i B_t^i\right)^{-1} \left\{ \left(L_t^i A_t^i + \dot{L}_t^i\right) \hat{x}_t^i dt + \int_{\mathbb{R}^2} L_t^i M_t^i \left(r - C_t^i \hat{x}_t^i\right) N^i \left(dt \times dr\right) \right\} \quad (6.25)$$

we have $L_t^i \hat{x}_t^i = 0$ for every $t \geqslant 0$.

Proof. We verify the validity of the lemma by substituting (6.25) into (6.11a) and left multiplying both sides by L_t^i. The resulting equation will be $L_t^i d\hat{x}_t^i = -\dot{L}_t^i \hat{x}_t^i dt$, which yields $d\left(L_t^i \hat{x}_t^i\right) = 0$. Then we argue that $L_t^i \hat{x}_t^i = L_0^i \hat{x}_0^i = 0$ for $t \geqslant 0$. □

Remark 6.4.3. The state-space equation (5.32c) indicates that the elements of x_t^i which are associated with ε_t^i constitute an isolated block in (6.1). Also, according

to (5.31), this block of x_t^i does not appear in $C_t^i x_t^i$, i.e., its corresponding block in C_t^i is identically zero. Considering these facts, we find from (6.11) that the block of $\hat{x}_t^i$ which is associated with ε_t^i is identically zero. This result together with (5.37) indicate[1] that $L_t^i \hat{x}_t^i = 0$ implies $C_t^i \hat{x}_t^i = 0$, which allows us to simplify (6.25) as

$$u_t^i dt = -\left(L_t^i B_t^i\right)^{-1} \left\{ \left(L_t^i A_t^i + \dot{L}_t^i\right) \hat{x}_t^i dt + L_t^i M_t^i \int_{\mathbb{R}^2} r N^i \left(dt \times dr\right) \right\}. \tag{6.26}$$

Remark 6.4.4. As mentioned in Remark 6.4.3, the estimate of ε_t^i generated by (6.11) is identically zero. As a consequence, under control (6.26), the effect of ε_t^i remains uncompensated. Since ε_t^a can be observed only by station b, the only possibility to compensate it is to generate its estimate by station b and send this estimate to station a through the optical channel or an independent RF link.

6.5 Control Problem: Long Range Applications

In this section, we develop a method for solving the control problem when $t_d > 0$. In this method, we first establish an analytically tractable lower bound on the objective functional (6.6) and then maximize that lower bound. We shall continue using the approximation (6.9) which leads to $\mu_t^i \simeq \tilde{\mu}_t^i$, where $\tilde{\mu}_t^i$ is defined as

$$\begin{aligned} \tilde{\mu}_t^a &= \nu_{t-t_d}^a \mathrm{E}\left[\exp\left(-\tfrac{1}{2}\left\|L_{t-t_d}^b z_{t-t_d}^b\right\|^2\right)\right] \\ \tilde{\mu}_t^b &= \nu_{t-t_d}^b \mathrm{E}\left[\exp\left(-\tfrac{1}{2}\left\|L_{t-t_d}^a z_{t-t_d}^a\right\|^2\right)\right]. \end{aligned} \tag{6.27}$$

[1]For the transceiver of Figure 5.2, also the condition $H + f_c K = 0$ is required.

In the remainder of this section, we consider an "approximate model" which is similar to the model of Section 6.2.1, except for μ_t^i being replaced with $\tilde{\mu}_t^i,\ i = a, b$. This approximate model allows us to present our results in exact statements, since its associated estimation problem has an exact solution. Then, these "exact results" are interpreted as "approximate results" for the original model of Section 6.2.1. Noting that $\left\{\tilde{\mu}_t^i\right\}$ is statistically independent of $\{w_t^i\}$ and x_0^i, the exact solution of the estimation problem associated with the approximate model is given by the posterior densities (6.10) and (6.13).

Our solution to the control problem is based on Theorem 6.5.1 which establishes a lower bound on the objective functional J, and Corollary 6.5.1 which determines a sufficient condition for the lower bound to achieve its maximum. In order to prove Theorem 6.5.1 and Corollary 6.5.1, we need the results of Lemmas 6.5.1 and 6.5.2 below.

Lemma 6.5.1. The scalar function $f_t^i(X)$ defined by (6.15) is decreasing in X, i.e., $0 \leqslant X^1 \leqslant X^2$ (in the sense of positive semidefiniteness ordering) implies that $f_t^i(X^1) \geqslant f_t^i(X^2) \geqslant 0$. Furthermore, for any positive definite random matrix X we have

$$\mathrm{E}\left[f_t^i(X)\right] \geqslant f_t^i(\mathrm{E}[X]). \tag{6.28}$$

Proof. The decreasing property of $f_t^i(\cdot)$ follows from the increasing property of $\det(\cdot)$ over the set of positive semidefinite matrices, which is shown in [53]. To prove (6.28), let us denote the argument of the determinant in (6.15) by Y. Since $(\det(\cdot))^{-1}$ is convex over the set of positive definite matrices [53], we use Jensen's inequality to

show

$$\mathrm{E}\left[f_t^i(X)\right] = \mathrm{E}\left[\left(\det Y^{1/2}\right)^{-1}\right] \geqslant \left(\det \mathrm{E}\left[Y^{1/2}\right]\right)^{-1}.$$

Then we exploit the matrix inequality $\mathrm{E}[Y] \geqslant \left(\mathrm{E}\left[Y^{1/2}\right]\right)^2$ and the increasing property of $\det(\cdot)$ to prove (6.28). □

The following lemma compares the solutions of two generalized matrix Riccati differential equations [54].

Lemma 6.5.2. Consider the symmetric matrix differential equations

$$\dot{Y}_t^1 = A_t Y_t^1 + Y_t^1 A_t^T + D_t D_t^T - \sigma_t^1 Y_t^1 C_t^T \left(C_t Y_t^1 C_t^T + R\right)^{-1} C_t Y_t^1$$

$$\dot{Y}_t^2 = A_t Y_t^2 + Y_t^2 A_t^T + D_t D_t^T - \sigma_t^2 Y_t^2 C_t^T \left(C_t Y_t^2 C_t^T + R\right)^{-1} C_t Y_t^2$$

and assume that the scalar functions σ_t^1 and σ_t^2 satisfy $0 \leqslant \sigma_t^2 \leqslant \sigma_t^1$ for $t \in [0, T]$. Then, $0 \leqslant Y_\tau^1 \leqslant Y_\tau^2$ implies that $0 \leqslant Y_t^1 \leqslant Y_t^2$ for $0 \leqslant \tau \leqslant t \leqslant T$.

Proof. The proof follows from [54, Theorem 4.5] by replacing the independent variable t with $-t$ and noting that

$$\begin{bmatrix} D_t D_t^T & 0 \\ 0 & R/\sigma_t^2 \end{bmatrix} \geqslant \begin{bmatrix} D_t D_t^T & 0 \\ 0 & R/\sigma_t^1 \end{bmatrix}.$$

□

Theorem 6.5.1. Fix sample paths ν_t^a and ν_t^b, $t \in [-t_d, T]$. Let $f_t^i(\cdot)$ be given

by (6.15) and β_t^i be defined such that $\beta_t^i = 0$ for $t < 0$ and

$$\beta_t^i = f_t^i\left(0_{n\times n}\right)\left\{1 - \mathrm{E}\left[\exp\left(-\tfrac{1}{2}\left\|Q_t^i L_t^i \hat{z}_t^i\right\|^2\right)\right]\right\}$$

for $t \geqslant 0$, where Q_t^i is defined by (6.16). Assume that Γ_t^a and Γ_t^b, $t \in [0, T]$ are the solutions of the matrix delay differential equations

$$\begin{aligned}
\dot{\Gamma}_t^a &= A_t^a \Gamma_t^a + \Gamma_t^a A_t^{aT} + D_t^a D_t^{aT} \\
&\quad - \nu_{t-t_d}^a \Big(f_{t-t_d}^b\left(\Gamma_{t-t_d}^b\right) - \beta_{t-t_d}^b\Big)\Gamma_t^a C_t^{aT}\left(C_t^a \Gamma_t^a C_t^{aT} + R\right)^{-1} C_t^a \Gamma_t^a \qquad (6.29a) \\
\dot{\Gamma}_t^b &= A_t^b \Gamma_t^b + \Gamma_t^b A_t^{bT} + D_t^b D_t^{bT} \\
&\quad - \nu_{t-t_d}^b \Big(f_{t-t_d}^a\left(\Gamma_{t-t_d}^a\right) - \beta_{t-t_d}^a\Big)\Gamma_t^b C_t^{bT}\left(C_t^b \Gamma_t^b C_t^{bT} + R\right)^{-1} C_t^b \Gamma_t^b \qquad (6.29b)
\end{aligned}$$

with the initial state $\Gamma_t^i = P_t^i$, $i = a, b$ for $t \in [-t_d, 0]$. Then, Γ_t^a and Γ_t^b upper bound $\mathrm{E}\left[\Sigma_t^a\right]$ and $\mathrm{E}\left[\Sigma_t^b\right]$, respectively, i.e. $\Gamma_t^a \geqslant \mathrm{E}\left[\Sigma_t^a\right]$ and $\Gamma_t^b \geqslant \mathrm{E}\left[\Sigma_t^b\right]$ for $t \in [0, T]$. Moreover, J_L defined as

$$J_L = \int_0^T \left(k^a \nu_t^a f_t^b\left(\Gamma_t^b\right) + k^b \nu_t^b f_t^a\left(\Gamma_t^a\right)\right) dt - \int_0^T \left(k^a \nu_t^a \beta_t^b + k^b \nu_t^b \beta_t^a\right) dt \qquad (6.30)$$

is a lower bound for J, i.e. $J_L \leqslant J$.

Proof. It is shown in [20, Theorem 4] that the solution of the matrix differential equation

$$\dot{X}_t^i = A_t^i X_t^i + X_t^i A_t^{iT} + D_t^i D_t^{iT} - \tilde{\mu}_t^i X_t^i C_t^{iT}\left(C_t^i X_t^i C_t^{iT} + R\right)^{-1} C_t^i X_t^i \qquad (6.31)$$

with the initial state $X_0^i = P_0^i$ is an upper bound for $\mathrm{E}\left[\Sigma_t^i\right]$, i.e. $\mathrm{E}\left[\Sigma_t^i\right] \leqslant X_t^i$. During $t \in [0, t_d]$, $x_{t-t_d}^b$ is a zero-mean Gaussian random vector with covariance matrix $P_{t-t_d}^b$. It is easy to show that $z_{t-t_d}^b$ is a zero-mean Gaussian random vector with the covariance matrix $H_{t-t_d} P_{t-t_d}^b H_{t-t_d}^T + V_{t-t_d}$. Then, recalling that $\Gamma_{t-t_d}^b = P_{t-t_d}^b$ during $t \in [0, t_d]$, we find from (6.27) that

$$\tilde{\mu}_t^a = \nu_{t-t_d}^a f_{t-t_d}^b \left(\Gamma_{t-t_d}^b\right), \ t \in [0, t_d]. \tag{6.32}$$

Substituting $\tilde{\mu}_t^a$ from (6.32) into (6.31) with $i = a$ and noting that $\beta_{t-t_d}^b = 0$, we find that (6.31) and (6.29a) are identical during $t \in [0, t_d]$. Hence, we can write $\Gamma_t^a = X_t^a \geqslant \mathrm{E}\left[\Sigma_t^a\right]$ for $t \in [0, t_d]$. With a similar argument, we can show that $\Gamma_t^b = X_t^b \geqslant \mathrm{E}\left[\Sigma_t^b\right]$, $t \in [0, t_d]$.

For $t \in (t_d, T]$, we use the smoothing property of conditional expectation and (6.14) to express (6.27) as

$$\begin{aligned}
\tilde{\mu}_t^a &= \nu_s^a \mathrm{E}\left[\mathrm{E}\left[\exp\left(-\tfrac{1}{2}\left\|L_s^b z_s^b\right\|^2\right) | \mathscr{B}_s^b\right]\right] \\
&= \nu_s^a \mathrm{E}\left[f_s^b\left(\Sigma_s^b\right) \exp\left(-\tfrac{1}{2}\left\|Q_s^b L_s^b \hat{z}_s^b\right\|^2\right)\right] \\
&= \nu_s^a \mathrm{E}\left[f_s^b\left(\Sigma_s^b\right)\right] - \nu_s^a \mathrm{E}\left[f_s^b\left(\Sigma_s^b\right)\left\{1 - \exp\left(-\tfrac{1}{2}\left\|Q_s^b L_s^b \hat{z}_s^b\right\|^2\right)\right\}\right]
\end{aligned}$$

where, for convenience of notation, we have replaced $t - t_d$ with s. From Lemma 6.5.1, we have $f_s^b\left(0_{n\times n}\right) \geqslant f_s^b\left(\Sigma_s^b\right)$ and $\mathrm{E}\left[f_s^b\left(\Sigma_s^b\right)\right] \geqslant f_s^b\left(\mathrm{E}\left[\Sigma_s^b\right]\right)$. These two inequalities

together with the definition of β_t^b result in

$$\tilde{\mu}_t^a \geqslant \nu_{t-t_d}^a \left(f_{t-t_d}^b \left(\mathrm{E} \left[\Sigma_{t-t_d}^b \right] \right) - \beta_{t-t_d}^b \right).$$

We partition the interval $[0, T]$ as

$$\mathscr{I}_1 = [0, t_d], \ \mathscr{I}_2 = (t_d, 2t_d], \ldots, \mathscr{I}_n = (nt_d - t_d, T]. \tag{6.33}$$

For $t \in \mathscr{I}_1$ we already proved that $\Gamma_t^i \geqslant X_t^i \geqslant \mathrm{E}\left[\Sigma_t^i\right]$. Assume that $\Gamma_t^i \geqslant X_t^i \geqslant \mathrm{E}\left[\Sigma_t^i\right]$ holds for $t \in \mathscr{I}_k$. Then, for $t \in \mathscr{I}_{k+1}$ we have $f_s^b \left(\mathrm{E} \left[\Sigma_s^b \right] \right) \geqslant f_s^b \left(\Gamma_s^b \right)$, which results in

$$\tilde{\mu}_t^a \geqslant \nu_{t-t_d}^a \left(f_{t-t_d}^b \left(\Gamma_{t-t_d}^b \right) - \beta_{t-t_d}^b \right). \tag{6.34}$$

In Lemma 6.5.2, let X_t^a, Γ_t^a, and kt_d play the role of Y_t^1, Y_t^2, and τ, respectively. Then inequality (6.34) and $\Gamma_{kt_d}^a \geqslant X_{kt_d}^a$ imply that $\Gamma_t^a \geqslant X_t^a \geqslant \mathrm{E}\left[\Sigma_t^a\right]$ for $t \in \mathscr{I}_{k+1}$. A similar argument shows that $\Gamma_t^b \geqslant X_t^b \geqslant \mathrm{E}\left[\Sigma_t^b\right]$ for $t \in \mathscr{I}_{k+1}$. Repeating this process, we show that $\Gamma_t^i \geqslant X_t^i \geqslant \mathrm{E}\left[\Sigma_t^i\right]$ holds for $t \in \mathscr{I}_k$, $k = 1, 2, \ldots n$, which means that $\Gamma_t^i \geqslant \mathrm{E}\left[\Sigma_t^i\right]$ holds for every $t \in [0, T]$. The second statement of the Theorem follows from (6.34). □

Corollary 6.5.1. Under the assumptions of Theorem 6.5.1, let Γ_t^{*a} and Γ_t^{*b}, $t \in [0, T]$ be the solutions of (6.29) with $\beta_t^a = \beta_t^b = 0$, $t \in [0, T - t_d]$. Then, J_L^* defined as

$$J_L^* = \int_0^T \left(k^a \nu_t^a f_t^b \left(\Gamma_t^{*b} \right) + k^b \nu_t^b f_t^a \left(\Gamma_t^{*a} \right) \right) dt$$

is an upper bound for J_L, i.e. $J_L^* \geqslant J_L$, and equality holds if for almost every $t \in [0, T]$ we have

$$L_t^a \hat{z}_t^a = L_t^b \hat{z}_t^b = 0. \tag{6.35}$$

Proof. Referring to the partition (6.33), since $\beta_t^a = \beta_t^b = 0$ for $t < 0$, we have $\Gamma_t^{*i} = \Gamma_t^i$ for $t \in \mathscr{I}_1$. Assume that $\Gamma_t^{*i} \leqslant \Gamma_t^i$ for $t \in \mathscr{I}_k$. Then, from the decreasing property of $f_t^i(\cdot)$ and the fact that $\beta_t^i \geqslant 0$, we conclude that

$$\nu_{t-t_d}^a f_{t-t_d}^b \left(\Gamma_{t-t_d}^{*b}\right) \geqslant \nu_{t-t_d}^a \left(f_{t-t_d}^b \left(\Gamma_{t-t_d}^b\right) - \beta_{t-t_d}^b\right) \tag{6.36}$$

holds for $t \in \mathscr{I}_{k+1}$. Similar to the proof of Theorem 6.5.1, we employ Lemma 6.5.2 and inequality (6.36) to show that $\Gamma_t^{*i} \leqslant \Gamma_t^i$ for $t \in \mathscr{I}_{k+1}$. As before, repeating this process, we show for $k = 1, 2, \ldots n$ that $\Gamma_t^{*i} \leqslant \Gamma_t^i$, $t \in \mathscr{I}_k$. This means that $\Gamma_t^{*i} \leqslant \Gamma_t^i$ holds for every $t \in [0, T]$. Applying this inequality and $\beta_t^i \geqslant 0$ to (6.30), we find $J_L^* \geqslant J_L$. From the definition of β_t^i, we know that (6.35) results in $\beta_t^a = \beta_t^b = 0$ almost everywhere in $[0, T]$, which leads to $J_L^* = J_L$. □

The following lemma proposes a control law $u_t^{\beta,i}$ which leads to $L_t^i H_t \hat{x}_t^i = 0$, $t \geqslant 0$, or equivalently the condition (6.35).

Lemma 6.5.3. Consider the stochastic differential equation (6.11a) with $B_t^i u_t^i = B_t^{\alpha,i} u_t^{\alpha,i} + B_t^{\beta,i} u_t^{\beta,i}$ and assume that $L_0^i H_0 \hat{x}_0^i = 0$. Let $L_t^i H_t B_t^{\beta,i}$ be nonsingular

and $L_t^i H_t$ be differentiable for $t \geqslant 0$. Then, under control

$$u_t^{\beta,i} dt = -\left(L_t^i H_t B_t^{\beta,i}\right)^{-1} \left\{ \left(L_t^i H_t A_t^i + \frac{dL_t^i H_t}{dt} \right) \hat{x}_t^i dt + L_t^i H_t B_t^{\alpha,i} u_t^{\alpha,i} dt \right. \\ \left. + \int_{\mathbb{R}^2} L_t^i H_t M_t^i \left(r - C_t^i \hat{x}_t^i\right) N^i \left(dt \times dr\right) \right\} \tag{6.37}$$

we have $L_t^i H_t \hat{x}_t^i = 0$ for every $t \geqslant 0$.

Proof. The proof is similar to the proof of Lemma 6.4.1. □

As mentioned in Section 6.2.2, the goal of the tracking control is to keep $C_t^i \hat{x}_t^i = 0$, $i = a, b$ for every $t \geqslant 0$. Suppose that $C_0^i \hat{x}_0^i = 0$, $C_t^i B_t^{\alpha,i}$ is nonsingular, and C_t^i is differentiable. Then, according to Lemma 6.4.1, we can achieve $C_t^i \hat{x}_t^i = 0,\ t \geqslant 0$, using the control law

$$u_t^{\alpha,i} dt = -\left(C_t^i B_t^{\alpha,i}\right)^{-1} \left\{ \left(C_t^i A_t^i + \dot{C}_t^i \right) \hat{x}_t^i dt + C_t^i M_t^i \int_{\mathbb{R}^2} r N^i \left(dt \times dr\right) \right\}. \tag{6.38}$$

Upon combining (6.37) and (6.38), we get

$$u_t^i dt = F_t^i \hat{x}_t^i dt + G_t^i M_t^i \int_{\mathbb{R}^2} r N^i \left(dt \times dr\right)$$

where the matrices F_t^i and G_t^i are given by

$$F_t^i = -\begin{bmatrix} I_{2\times 2} & 0_{2\times 2} \\ \left(L_t^i H_t B_t^{\beta,i}\right)^{-1} L_t^i H_t B_t^{\alpha,i} & I_{2\times 2} \end{bmatrix}^{-1} \begin{bmatrix} \left(C_t^i B_t^{\alpha,i}\right)^{-1} \left(C_t^i A_t^i + \dot{C}_t^i\right) \\ \left(L_t^i H_t B_t^{\beta,i}\right)^{-1} \left(L_t^i H_t A_t^i + \dot{L}_t^i H_t + L_t^i \dot{H}_t\right) \end{bmatrix}$$

and

$$G_t^i = -\begin{bmatrix} I_{2\times 2} & 0_{2\times 2} \\ \left(L_t^i H_t B_t^{\beta,i}\right)^{-1} L_t^i H_t B_t^{\alpha,i} & I_{2\times 2} \end{bmatrix}^{-1} \begin{bmatrix} \left(C_t^i B_t^{\alpha,i}\right)^{-1} C_t^i \\ \left(L_t^i H_t B_t^{\beta,i}\right)^{-1} L_t^i H_t \end{bmatrix}.$$

6.6 Proof of Theorem 6.4

Using the differential rule for point processes[1] [24], we can write

$$\begin{aligned} dg_t\left(\Sigma_t^a, \Sigma_t^b\right) = & \left(\partial g_t\left(\Sigma_t^a, \Sigma_t^b\right)/\partial t\right) dt \\ & + \mathrm{tr}\left\{\left(\partial g_t\left(\Sigma_t^a, \Sigma_t^b\right)/\partial \Sigma_t^a\right)\left(A_t^a \Sigma_t^a + \Sigma_t^a A_t^{aT} + D_t^a D_t^{aT}\right)\right\} dt \\ & + \mathrm{tr}\left\{\left(\partial g_t\left(\Sigma_t^a, \Sigma_t^b\right)/\partial \Sigma_t^b\right)\left(A_t^b \Sigma_t^b + \Sigma_t^b A_t^{bT} + D_t^b D_t^{bT}\right)\right\} dt \\ & + \mathcal{L}_t^a\left\{g_t\left(\Sigma_t^a, \Sigma_t^b\right)\right\} dN_t^a + \mathcal{L}_t^b\left\{g_t\left(\Sigma_t^a, \Sigma_t^b\right)\right\} dN_t^b. \end{aligned} \tag{6.39}$$

Noting that $\{\mu_t^i\}$ defined by (6.8) is the rate of $\{N_t^i\}$, we use the smoothing property of conditional expectation to show that

$$\begin{aligned} \mathrm{E}\left[\mathcal{L}_t^i\left\{g_t\left(\Sigma_t^a, \Sigma_t^b\right)\right\} dN_t^i\right] &= \mathrm{E}\left[\mathrm{E}\left[\mathcal{L}_t^i\left\{g_t\left(\Sigma_t^a, \Sigma_t^b\right)\right\} dN_t^i | \mu_t^i\right]\right] \\ &= \mathrm{E}\left[\mathrm{E}\left[\mathcal{L}_t^i\left\{g_t\left(\Sigma_t^a, \Sigma_t^b\right)\right\} \mu_t^i dt | \mu_t^i\right]\right] \\ &= \mathrm{E}\left[\mathcal{L}_t^i\left\{g_t\left(\Sigma_t^a, \Sigma_t^b\right)\right\} \mu_t^i\right] dt, \;\; i = a, b. \end{aligned}$$

[1] This is the counterpart of the Itô differential rule for Wiener process.

Taking expectation from both sides of (6.39), using the last result, and noting that $\mathrm{E}\left[dg_t\big(\Sigma_t^a, \Sigma_t^b\big)\right] = d\,\mathrm{E}\left[g_t\big(\Sigma_t^a, \Sigma_t^b\big)\right]$, we get

$$
\begin{aligned}
d\,\mathrm{E}\left[g_t\big(\Sigma_t^a, \Sigma_t^b\big)\right] = \mathrm{E}\Big[& \partial g_t\big(\Sigma_t^a, \Sigma_t^b\big)/\partial t \\
& + \mathrm{tr}\left\{\big(\partial g_t\big(\Sigma_t^a, \Sigma_t^b\big)/\partial \Sigma_t^a\big)\left(A_t^a \Sigma_t^a + \Sigma_t^a A_t^{aT} + D_t^a D_t^{aT}\right)\right\} \\
& + \mathrm{tr}\left\{\big(\partial g_t\big(\Sigma_t^a, \Sigma_t^b\big)/\partial \Sigma_t^b\big)\left(A_t^b \Sigma_t^b + \Sigma_t^b A_t^{bT} + D_t^b D_t^{bT}\right)\right\} \\
& + \mathcal{L}_t^a\left\{g_t\left(\Sigma_t^a, \Sigma_t^b\right)\right\}\mu_t^a + \mathcal{L}_t^b\left\{g_t\left(\Sigma_t^a, \Sigma_t^b\right)\right\}\mu_t^b\Big] dt.
\end{aligned}
$$

We add the expression $\mathrm{E}\left[k^a\mu_t^a + k^b\mu_t^b\right]dt$ to the both sides of this equation and rearrange the terms in order to obtain

$$
\begin{aligned}
\mathrm{E}\big[k^a\mu_t^a + k^b\mu_t^b\big]dt - \Big\{ & - d\,\mathrm{E}\left[g_t\big(\Sigma_t^a, \Sigma_t^b\big)\right] \\
& + \mathrm{E}\left[\big(\mu_t^a - \nu_t^a f_t^b\big(\Sigma_t^b\big)\big)\left(k^a + \mathcal{L}_t^a\left\{g_t\left(\Sigma_t^a, \Sigma_t^b\right)\right\}\right)\right] dt \\
& + \mathrm{E}\left[\big(\mu_t^b - \nu_t^b f_t^a\big(\Sigma_t^a\big)\big)\left(k^b + \mathcal{L}_t^b\left\{g_t\left(\Sigma_t^a, \Sigma_t^b\right)\right\}\right)\right] dt\Big\} \\
= \mathrm{E}\Bigg[& \frac{\partial g_t\left(\Sigma_t^a, \Sigma_t^b\right)}{\partial t} + \nu_t^a f_t^b\big(\Sigma_t^b\big)\Big(k^a + \mathcal{L}_t^a\left\{g_t\left(\Sigma_t^a, \Sigma_t^b\right)\right\}\Big) \\
& + \nu_t^b f_t^a\big(\Sigma_t^a\big)\Big(k^b + \mathcal{L}_t^b\left\{g_t\left(\Sigma_t^a, \Sigma_t^b\right)\right\}\Big) \\
& + \mathrm{tr}\left\{\frac{\partial g_t\left(\Sigma_t^a, \Sigma_t^b\right)}{\partial \Sigma_t^a}\left(A_t^a \Sigma_t^a + \Sigma_t^a A_t^{aT} + D_t^a D_t^{aT}\right)\right\} \\
& + \mathrm{tr}\left\{\frac{\partial g_t\left(\Sigma_t^a, \Sigma_t^b\right)}{\partial \Sigma_t^b}\left(A_t^b \Sigma_t^b + \Sigma_t^b A_t^{bT} + D_t^b D_t^{bT}\right)\right\}\Bigg] dt.
\end{aligned}
$$

Since $g_t(\cdot,\cdot)$ is the solution of (6.21), the right side of the equation above is identically zero which leads to

$$\begin{aligned}\mathrm{E}\left[k^a\mu_t^a + k^b\mu_t^b\right]dt = &- d\,\mathrm{E}\left[g_t\big(\Sigma_t^a,\Sigma_t^b\big)\right]\\ &+ \mathrm{E}\left[\big(\mu_t^a - \nu_t^a f_t^b\big(\Sigma_t^b\big)\big)\left(k^a + \mathcal{L}_t^a\left\{g_t\left(\Sigma_t^a,\Sigma_t^b\right)\right\}\right)\right]dt\\ &+ \mathrm{E}\left[\big(\mu_t^b - \nu_t^b f_t^a\big(\Sigma_t^a\big)\big)\left(k^b + \mathcal{L}_t^b\left\{g_t\left(\Sigma_t^a,\Sigma_t^b\right)\right\}\right)\right]dt.\end{aligned}$$

Integrating this equation from 0 to T and noting that $\mathrm{E}\left[g_T\left(\Sigma_T^a,\Sigma_T^b\right)\right] = 0$ and $\mathrm{E}\left[g_0\left(\Sigma_0^a,\Sigma_0^b\right)\right] = g_0\left(P_0^a,P_0^b\right)$, we obtain

$$\begin{aligned}J &= \int_0^T \mathrm{E}\left[k^a\mu_t^a + k^b\mu_t^b\right]dt\\ &= g_0\big(P_0^a,P_0^b\big) + \int_0^T \mathrm{E}\left[\big(\mu_t^a - \nu_t^a f_t^b\big(\Sigma_t^b\big)\big)\left(k^a + \mathcal{L}_t^a\left\{g_t\left(\Sigma_t^a,\Sigma_t^b\right)\right\}\right)\right]dt\\ &\qquad + \int_0^T \mathrm{E}\left[\big(\mu_t^b - \nu_t^b f_t^a\big(\Sigma_t^a\big)\big)\left(k^b + \mathcal{L}_t^b\left\{g_t\left(\Sigma_t^a,\Sigma_t^b\right)\right\}\right)\right]dt. \qquad (6.40)\end{aligned}$$

From (6.8) and (6.23), we have

$$\begin{aligned}\mu_t^a - \nu_t^a f_t^b\big(\Sigma_t^b\big) &= \nu_t^a\delta_t^b - \nu_t^a f_t^b\big(\Sigma_t^b\big)\left\{1 - \exp\left(-\tfrac{1}{2}\left\|Q_t^b L_t^b \hat{x}_t^b\right\|^2\right)\right\}\\ \mu_t^b - \nu_t^b f_t^a\big(\Sigma_t^a\big) &= \nu_t^b\delta_t^a - \nu_t^b f_t^a\big(\Sigma_t^a\big)\left\{1 - \exp\left(-\tfrac{1}{2}\left\|Q_t^a L_t^a \hat{x}_t^a\right\|^2\right)\right\}.\end{aligned}$$

Substituting these expressions into (6.40), we obtain (6.22).

In order to prove the second statement of the theorem, we need the following preliminaries.

P-1) In the context of this proof, we say $f(\cdot):\mathbb{R}^{n\times n}\to\mathbb{R}$ is decreasing, if for any

positive definite[1] Σ and any positive semidefinite Δ, we have $f(\Sigma+\Delta) \leqslant f(\Sigma)$. Also, we say $f(\cdot)$ is m-nonnegative, if for any positive definite Σ, we have $f(\Sigma) \geqslant 0$.

P-2) If $f_1(\cdot)$ and $f_2(\cdot)$ are decreasing and m-nonnegative, $f_1(\cdot)+f_2(\cdot)$ and $f_1(\cdot) f_2(\cdot)$ are decreasing and m-nonnegative as well.

P-3) If $f(\cdot)$ is decreasing, for any positive definite Σ, we have $f(\Sigma) \leqslant f(S_t^i(\Sigma))$, where $S_t^i(\cdot)$ is defined by (6.20).

Proof. Applying the matrix inversion lemma to (6.20), it is easy to verify that for any positive definite Σ, $S_t^i(\Sigma)$ is positive definite. Also, we know from (6.20) that $\Delta \triangleq \Sigma - S_t^i(\Sigma)$ is a positive semidefinite matrix. Since $f(\cdot)$ is decreasing, we can write

$$f(\Sigma) = f\left(S_t^i(\Sigma)+\Delta\right) \leqslant f\left(S_t^i(\Sigma)\right).$$

□

P-4) If $f(\cdot)$ is decreasing and m-nonnegative, for any fixed t, $f(S_t^i(\cdot))$ is decreasing and m-nonnegative.

Proof. For any positive definite Σ and any positive semidefinite Δ, we can show

$$\Sigma^{-1} - (\Sigma+\Delta)^{-1} = \Sigma^{-1}\Delta^{1/2}\left(I+\Delta^{1/2}\Sigma^{-1}\Delta^{1/2}\right)^{-1}\Delta^{1/2}\Sigma^{-1} \triangleq \tilde{\Delta} \qquad (6.41)$$

[1]By definition, any positive definite matrix is symmetric.

where I is the identity matrix with proper dimension. This indicates that $\tilde{\Delta}$ is positive semidefinite. Using the matrix inversion lemma and replacing Σ^{-1} with $(\Sigma+\Delta)^{-1}+\tilde{\Delta}$, we can write

$$\begin{aligned} S_t^i(\Sigma+\Delta) - S_t^i(\Sigma) &= \left((\Sigma+\Delta)^{-1} + C_t^{iT} R^{-1} C_t^i\right)^{-1} - \left(\Sigma^{-1} + C_t^{iT} R^{-1} C_t^i\right)^{-1} \\ &= \left((\Sigma+\Delta)^{-1} + C_t^{iT} R^{-1} C_t^i\right)^{-1} \\ &\quad - \left(\left((\Sigma+\Delta)^{-1} + C_t^{iT} R^{-1} C_t^i\right) + \tilde{\Delta}\right)^{-1}. \end{aligned}$$

Applying (6.41) to the last equality, we find that $S_t^i(\Sigma+\Delta) - S_t^i(\Sigma)$ is positive semidefinite. Then, since $S_t^i(\Sigma)$ is positive definite and $f(\cdot)$ is decreasing, we have

$$f\left(S_t^i(\Sigma+\Delta)\right) = f\left(S_t^i(\Sigma) + \left\{S_t^i(\Sigma+\Delta) - S_t^i(\Sigma)\right\}\right) \leqslant f\left(S_t^i(\Sigma)\right)$$

which means that $f(S_t^i(\cdot))$ is decreasing. Moreover, since $f(\cdot)$ is m-nonnegative and $S_t^i(\Sigma)$ is positive definite, $f(S_t^i(\cdot))$ is m-nonnegative. □

P-5) For any fixed t, $f_t^i(\cdot)$ defined by (6.17) is decreasing and m-nonnegative.

Proof. For any positive definite Σ and any positive semidefinite Δ, we can write

$$\begin{aligned} \frac{f_t^i(\Sigma)}{f_t^i(\Sigma+\Delta)} &= \sqrt{\frac{\det\left(I_{2\times 2} + L_t^i \Sigma L_t^{iT} + L_t^i \Delta L_t^{iT}\right)}{\det\left(I_{2\times 2} + L_t^i \Sigma L_t^{iT}\right)}} \\ &= \sqrt{\det\left(I_{2\times 2} + \bar{\Delta}_t^i\right)} \end{aligned}$$

where $\bar{\Delta}_t^i$ is defined as

$$\bar{\Delta}_t^i = \left(\left(I_{2\times 2} + L_t^i \Sigma L_t^{iT}\right)^{-1/2} L_t^i\right) \Delta \left(\left(I_{2\times 2} + L_t^i \Sigma L_t^{iT}\right)^{-1/2} L_t^i\right)^T.$$

This implies that $\det\left(I_{2\times 2} + \bar{\Delta}_t^i\right) \geqslant 1$, which leads to $f_t^i(\Sigma + \Delta) \leqslant f_t^i(\Sigma)$. $\square$

P-6) Let $h_t\left(\Sigma^a, \Sigma^b\right)$ be a scalar function of $n \times n$ matrices Σ^a and Σ^b. Assume that this function is decreasing and m-nonnegative in both Σ^a and Σ^b. For $\epsilon \geqslant 0$ define the linear operator $\mathcal{K}_t^\epsilon$ as

$$\mathcal{K}_t^\epsilon h_t\left(\Sigma^a, \Sigma^b\right) = \left(1 - \epsilon \nu_t^a f_t^b\left(\Sigma^b\right) - \epsilon \nu_t^b f_t^a\left(\Sigma^a\right)\right) h_t\left(X_t^{a,\epsilon}\left(\Sigma^a\right), X_t^{b,\epsilon}\left(\Sigma^b\right)\right)$$

where

$$X_t^{i,\epsilon}(\Sigma) = \Sigma + \epsilon\left(A_t^i \Sigma + \Sigma A_t^{iT} + D_t^i D_t^{iT}\right).$$

Then, for any positive definite Σ^a and Σ^b and any positive semidefinite Δ^a and Δ^b, there exists $\zeta = \zeta\left(\Sigma^a, \Sigma^b, \Delta^a, \Delta^b\right) > 0$ such that for every $0 \leqslant \epsilon < \zeta$, we have

$$\mathcal{K}_t^\epsilon h_t\left(\Sigma^a + \Delta^a, \Sigma^b\right) \leqslant \mathcal{K}_t^\epsilon h_t\left(\Sigma^a, \Sigma^b\right)$$

$$\mathcal{K}_t^\epsilon h_t\left(\Sigma^a, \Sigma^b + \Delta^b\right) \leqslant \mathcal{K}_t^\epsilon h_t\left(\Sigma^a, \Sigma^b\right)$$

$$\mathcal{K}_t^\epsilon h_t\left(\Sigma^a, \Sigma^b\right) \geqslant 0.$$

Therefore, as $\epsilon \to 0^+$, these conditions are satisfied for any choice of Σ^a, Σ^b, Δ^a, and Δ^b. This means that $\mathcal{K}_t^\epsilon h_t\left(\Sigma^a, \Sigma^b\right)$ is decreasing and m-nonnegative

in both Σ^a and Σ^b, as $\epsilon \to 0^+$.

We claim that $g_t\big(\Sigma^a, \Sigma^b\big)$, the solution of equation (6.21) with the boundary condition $g_T\big(\Sigma^a, \Sigma^b\big) = 0$, is decreasing in both Σ^a and Σ^b for every $t \in [0, T]$. Once the claim is proven, we apply (P-3) to (6.19) in order to show that $\mathcal{L}_t^i \left\{ g_t \left(\Sigma^a, \Sigma^b \right) \right\} \geqslant 0$, $i = a, b$ for every positive definite matrices Σ^a and Σ^b and every $t \in [0, T]$.

In order to prove this claim, for any $0 \leqslant t < T$, we partition the interval $[t, T]$ into K subintervals $[t_{k+1}, t_k)$, $k = 0, 1, \ldots, K-1$, where $t_K = t$, $t_0 = T$, and $t_k - t_{k+1} = \epsilon_k > 0$. Using this partition, we discretise the partial differential equation (6.21) to obtain the recursive equation

$$
\begin{aligned}
g_{t_{k+1}}\big(\Sigma^a, \Sigma^b\big) = \epsilon_k \Big(& k^a \nu_{t_k}^a f_{t_k}^b \big(\Sigma^b\big) + k^b \nu_{t_k}^b f_{t_k}^a \big(\Sigma^a\big) \Big) \\
& + \epsilon_k \Big(\nu_{t_k}^a f_{t_k}^b \big(\Sigma^b\big) g_{t_k} \big(S_{t_k}^a \left(\Sigma^a\right), \Sigma^b \big) + \nu_{t_k}^b f_{t_k}^a \big(\Sigma^a\big) g_{t_k} \big(\Sigma^a, S_{t_k}^b \left(\Sigma^b\right) \big) \Big) \\
& + \mathcal{K}_{t_k}^{\epsilon_k} g_{t_k} \big(\Sigma^a, \Sigma^b\big) + O\left(\epsilon_k^2\right). \qquad (6.42)
\end{aligned}
$$

Starting from $g_{t_0}(\cdot, \cdot) = 0$ and using this recursive equation for $k = 0, 1, 2, \ldots, K-1$, we can determine $g_{t_K}(\cdot, \cdot)$. Then, by letting $K \to \infty$ such that $\max \epsilon_k \to 0$, we have $g_{t_K}(\cdot, \cdot) \to g_t(\cdot, \cdot)$.

We prove by induction that as $K \to \infty$ and $\max \epsilon_k \to 0$, for $k = 0, 1, \ldots, K$, $g_{t_k}(\cdot, \cdot)$ is decreasing and m-nonnegative in both Σ^a and Σ^b. Clearly, $g_{t_0}\big(\Sigma^a, \Sigma^b\big) = 0$ is decreasing and m-nonnegative in both Σ^a and Σ^b. Also, we show that if $g_{t_k}\big(\Sigma^a, \Sigma^b\big)$ is decreasing and m-nonnegative, $g_{t_{k+1}}\big(\Sigma^a, \Sigma^b\big)$ is decreasing and m-nonnegative as well. For this purpose, we use (P-2, P-5) and (P-2, P-4, P-5), respectively, to show that the first and the second terms on the right side of (6.42) are decreasing

and m-nonnegative. Also, as $\epsilon_k \to 0^+$, (P-6) implies that the third term on the right side of (6.42) is decreasing and m-nonnegative. Since all three terms on the right side of (6.42) are decreasing and m-nonnegative, we conclude from (P-2) that $g_{t_{k+1}}\left(\Sigma^a, \Sigma^b\right)$ is decreasing and m-nonnegative.

Chapter 7

Conclusion and Directions for Future Work

7.1 Summary of Main Contributions

This dissertation is devoted to finding solutions for two major concerns in free-space optics: digital communication over a free-space optical channel and optical alignment between the transmitter and the receiver. Adopting a stochastic approach, we formulated these concerns in terms of detection, estimation, and optimal control problems with point process observations. This observation model is imposed by the nature of optical sensors which are the essential component of a free-space optical link.

We discussed an M-ary detection problem associated with a marked and filtered Poisson process in additive white Gaussian noise. The motivation for this study comes from the digital communication over the optical channels (fiber or free-space). The stochastic model adopted for the problem is adequate for characterizing the optical sensors (including the avalanche gain), the thermal noise generated by the amplifying circuits, and the atmosphere-induced optical fade. We obtained a solution for the problem in terms of an infinite sum of multiple integrals, which is

hard to express in an explicit form. We established two sets of upper and lower bounds on this infinite sum, where the bounds in the first set are given in explicit forms, while in the second set, two integral equations determine the bounds. In both cases, it was observed that under certain conditions, the lower bound is close to the upper bound, which is a motivation for approximating the solution by one of these bounds. In another effort for simplifying the infinite sum, we expressed it in terms of an expectation taken with respect to a stochastic process. The resulting expression was our point of departure to develop several approximations with different levels of complexity. These approximations can be implemented by means of finite-dimensional, nonlinear, causal filters.

A stochastic dynamical model introduced in [20] plays a central role in our study of optical alignment. This model consists of a linear stochastic state-space equation driven by a control vector and a vector-valued Wiener process, and the observation of a space-time point process with a rate which depends on the state vector. Associated with this model, an optimal control problem is defined in [20] in terms of minimizing a quadratic cost functional. The model and its associated control problem have been used in [19] in order to analyze an optical beam tracking system which employs an infinite resolution position-sensitive photodetector. In that study, it is assumed that the observation is provided over $\mathbb{R}^2$ and that the rate of the space-time point process has a Gaussian profile. Under these assumptions, the control problem and its associated state estimation problem have finite-dimensional exact solutions [20]. We used these solutions to develop a suboptimal control law for an optical beam tracking system with a finite resolution photodetector. Fur-

thermore, in an effort to extend these results to the case of a non-Gaussian rate, we demonstrated that the estimation problem can be formulated in terms of estimating the state of a discrete-time linear model with an observation vector which is corrupted by additive white non-Gaussian noise.

We proposed an active pointing scheme in which the receiving station estimates the center of its incident optical beam by means of a position-sensitive photodetector. The transmitter receives this estimate via an independent communication link and incorporates it to accurately aim at the receiving station. We showed that the stochastic model mentioned above can adequately characterize this alignment scheme, but with the observation which is provided over a subset of $\mathbb{R}^2$ instead of $\mathbb{R}^2$. Regarding this modified model, we determined a suboptimal state estimator and a suboptimal control law. In addition, we demonstrated that our suboptimal results tend toward the optimal results when the observation is provided over the entire $\mathbb{R}^2$.

We studied the concept of cooperative optical beam tracking and developed a detailed nonlinear model for it. Next we linearized the model around a nominal state trajectory and presented a stochastic description for its disturbance vectors. Associated with this stochastic model, we considered an optimal control problem with the goal of maximizing an objective functional defined as the expected optical energy received by the stations of the link. For short range applications with negligible light propagation delay, we proposed a suboptimal control law which approximately maximizes the objective functional. We demonstrated that the proposed control law does not depend on the nature of the optical fade or the information-bearing signals which modulate the optical beams. In addition, we showed that the control

law simultaneously maximizes the received optical energy for both stations. We also addressed the considerations arising from light propagation delay in a cooperative optical beam tracking system. For the case that the propagation delay is significant, a suboptimal control law was developed based on maximizing a lower bound on the objective functional.

7.2 Directions for Future Work

In this section, we sketch some directions for the future research. We first highlight a few problems regarding the topics discussed in this dissertation which can complete or extend our present results. Then, we briefly explain an application of free-space optical communication in space missions and its associated problems which can be viewed in the framework of this study.

7.2.1 Extending the Present Results

In Chapter 2, we developed several detection rules as the solution to our detection problem. In order to evaluate the performance of each detector, it is required to obtain its associated probability of error. In addition, this performance measure can be used to compare the effectiveness of the proposed detection rules. In particular, it is useful to compare the performance of each detection rule with a simple linear detector. This comparison evaluates the improvement of the performance as a result of accepting the complexity of a nonlinear detector. On the other hand, it is important to determine how the probability of error varies with the parameters of the model. Considering the complexity of the detectors and the observation process,

it is difficult to obtain an analytical expression for the probability of error. It seems that the numerical methods such as Monte Carlo simulation are the most convincing way to compute this quantity.

Upon computing the probability of error, we can improve the performance of the detection rules by optimizing their associated threshold. To explain this idea, consider the detection rule (2.7) for the binary case. In this detection rule, the threshold p_1/p_2 is determined in terms of the prior probability of the binary message; however, when we replace the exact likelihood ratio function $L^b(T)$ with its approximation, the optimal threshold is not necessarily p_1/p_2. Thus, to achieve the best performance, we can determine the probability of error as a function of an unknown threshold and then, minimize this function with respect to the threshold. This modification can be easily extended to the general M-ary hypothesis testing problem whose optimal detector is given by (2.6). In this case, the probability of error must be minimized with respect to $p_1, p_2, \ldots, p_M$, subject to the constraints $p_i \geqslant 0,\ i = 1, 2, \ldots, M$ and $p_1 + p_2 + \cdots + p_M = 1$.

In Section 3.5, we proposed a control law for an optical beam tracking system with a finite resolution photodetector. This controller was developed by applying an approximation scheme to the results of [20] for an infinite resolution photodetector. Another approach is to directly solve the estimation and control problems for the original finite resolution photodetector. Since in this case, the estimation problem is infinite-dimensional, we have to approximate its solution using a finite-dimensional filter. The cumulant matching method developed in Section 4.3 is a possible scheme for this approximation. Following this method, we can approximate the posterior

density with a Gaussian density function whose mean an covariance matrix are obtained from a set of stochastic differential equations driven by the photodetector output. Further, a suboptimal control law can be determined by following the procedure in Section 4.4. Only after finding this controller and comparing its performance with the already developed controller, we can decide which one better approximates the solution of the optimal control problem.

In Chapter 5, we mentioned to the possible application of inertial sensors in a cooperative optical beam tracking system. In order to involve these sensors in our analysis, we need to include the output of the sensors in the observation set of the estimators. Then, the estimation and control problems must be solved again for this extend observation set. We note that the output of each inertial sensor is a noisy version of an element of the disturbance vector. If there are convincing indications that the noise is additive and Gaussian, the results of [20] still can be applied to the estimation problem.

Another method for improving the performance of a cooperative optical beam tracking system is to exchange information between the stations. In one scenario, station b estimates the error vector ε_t^a and sends the estimate back to station a. Then, the controller of station a employs this estimate to compensate for ε_t^a by means of the point-ahead mirror. In another scenario, each station transmits its photodetector output to the other station, i.e., the stations estimate their state from the observation of both stations. Clearly, these schemes improve the performance of the system by providing more information to the controllers; however, it is not clear whether this improvement is significant enough to justify the increased complexity.

This must be evaluated through comparing the performance of the schemes by means of computer simulations.

7.2.2 Optical Communication for Space Missions

In recent years, the idea of assigning the task of a large satellite to a cluster of cooperating micro-satellites has drawn attention to certain space missions [55, 56]. A possible application for this idea is a cluster of micro-satellites equipped with small aperture antennas, which cooperatively act as a distributed antenna with an effective aperture size larger than that can be achieved by a single large satellite. In addition to cost reduction [4], a cluster of micro-satellites flying in formation has the advantage of being reconfigurable to meet requirements for different missions [56].

In a multisatellite application, a closed-loop formation-keeping controller maintains the required constellation in spite of the disturbance forces. An essential component to implement this controller is the availability of reliable communication between the micro-satellites forming the constellation [48]. Because of high-bandwidth, power efficiency, and small weight, free-space optics is an attractive means for communication between the micro-satellites [48]. Moreover, a free-space optical link can be used simultaneously for the purpose of range and attitude measurement [57], which is another essential component for formation flying [48].

Deployment of the optical links in a formation flying system raises several questions and design concerns, while at the same time, it opens some windows of opportunity. A fundamental question in this regard is whether the communication subsystem (including the optical alignment components) and the formation-keeping

controller can be designed independently. For two reasons, the answer to this question seems to be negative. First, the limitations of this communication scheme (e. g., communication lost due to sudden lose of alignment) require special considerations in the design of the formation-keeping controller. The second reason is that a certain formation flying mission might impose constraints on the optical alignment system, or it might support a capability for designing a more efficient and less expensive alignment system.

In a formation flying system, the communication between the members can be established through an optical network. This provides with extra flexibility for the optical alignment systems, since for any specific formation state, the network can be reconfigured to achieve the best alignment performance.

A stochastic approach to the problems above is the most convincing one [4]. Finding solutions to these problems is a logical continuation of this research, in light of the well-established models and methods we developed in the present work.

7.3 Possible Applications in the Study of Nervous Systems

The response of an animal's sensory nerve to a physical stimulus is a sequence of electrical pulses called spike or action potential [58, 59]. By means of sensory nerves, the information carried by the stimuli is represented (coded) in terms of the temporal pattern of the sequences of spikes [58]. An important question in the study of nervous systems is how the information is represented through this temporal pattern. This question leads to the associated decoding problem: how to reconstruct the stimulus from the spikes.

The nature of the problem above suggests a stochastic approach to tackle the problem. A complete description of the problem consists of a stochastic model for the "stimulus-spike" relationship, a stochastic model to describe the temporal evolution of the stimulus, and formulating the decoding problem in terms of an estimation problem. For example, the authors of [60] characterized place cells[1] in terms of an array of doubly stochastic Poisson processes (place cell-spike frequency representation). Then, they formulated the decoding problem as estimating the state of a continuous-time state-space model which modulates the rates of the Poisson processes. A discrete-time version of this model has been used in [61], in order to estimate the position of a rat, based on the data collected from its hippocampal place cells. Also, adaptive filtering techniques have been applied to the model in order to analyze the plasticity[2] of neural receptive fields [62].

These examples demonstrate close similarity to the stochastic models we worked with during the present research. Thus, the body of techniques we developed here might be useful in study of nervous systems.

[1]Place cells are sensory nerves which fire when the animal is close to a particular location.
[2]This means that the response of neurons to stimuli change with experience.

Bibliography

[1] H. A. Willebrand and B. S. Ghuman, "Fiber optics without fiber," *Spectrum, IEEE*, vol. 38, pp. 40–45, Aug. 2001.

[2] D. Kerr, K. Bouazza-Marouf, K. Girach, and T. C. West, "Free space laser communication links for short range control of mobile robots using active pointing and tracking techniques," in *IEE Colloquium on Optical Free Space Communication Links*, Feb. 1996, pp. 11/1–11/5.

[3] J. Maynard, "System design considerations," in *Laser Satellite Communication*, M. Katzman, Ed. Prentice-Hall, Inc., 1987, ch. 2, pp. 11–67.

[4] V. W. S. Chan, "Optical satellite networks," *IEEE, Journal of Lightwave Technology*, vol. 21, no. 11, pp. 2811–2827, Nov. 2003.

[5] E. V. Hoversten, D. L. Snyder, R. O. Harger, and K. Kurimoto, "Direct-detection optical communication receivers," *IEEE Transactions on Communications*, vol. COM-22, no. 1, pp. 17–27, Jan. 1974.

[6] B. Reiffen and H. Sherman, "An optimum demodulator for Poisson processes: Photon source detectors," *Proceedings IEEE*, vol. 51, no. 10, pp. 1316–1320, Oct. 1963.

[7] K. Abend, "Optimum photon detection," *IEEE Transactions on Information Theory*, vol. IT-12, no. 1, pp. 64–65, Jan. 1966.

[8] I. Bar-David, "Communication under the Poisson regime," *IEEE Transactions on Information Theory*, vol. IT-15, no. 1, pp. 31–37, Jan. 1969.

[9] D. L. Snyder, "Filtering and detection for doubly stochastic Poisson processes," *IEEE Transactions on Information Theory*, vol. IT-18, no. 1, pp. 91–102, Jan. 1972.

[10] R. M. Gagliardi and S. Karp, "M-ary Poisson detection and optical communications," *IEEE Transactions on Communications*, vol. COM-17, no. 2, pp. 208–216, Apr. 1969.

[11] W. A. Gardner, "An equivalent linear model for marked and filtered doubly stochastic Poisson processes with application to MMSE linear estimation for synchronous m-ary optical data signals," *IEEE Transactions on Communications*, vol. 24, no. 8, pp. 917–921, Aug. 1976.

[12] G. F. Herrmann, "Linear detection in a Poisson regime with random pulse heights," *IEEE Transactions on Communications*, vol. 24, no. 2, pp. 254–259, Feb. 1976.

[13] J. R. F. Da Rocha and J. J. O'Reilly, "Linear direct-detection fiber-optic receiver optimization in the presence of intersymbol interference," *IEEE Transactions on Communications*, vol. COM-34, no. 4, pp. 365–374, Apr. 1986.

[14] J. J. O'Reilly and J. R. F. Da Rocha, "Improved error probability evaluation methods for direct detection optical communication systems," *IEEE Transactions on Information Theory*, vol. IT-33, no. 6, pp. 839–848, Nov. 1987.

[15] K. E. House, "Filters for the detection of binary signaling: Optimization using the Chernoff bound," *IEEE Transactions on Communications*, vol. COM-28, no. 2, pp. 257–259, Feb. 1980.

[16] K. Kurimoto, "Optimum filtering and detection for doubly stochastic shot noise process in white Gaussian noise," Ph.D. dissertation, University of Maryland, 1972.

[17] S. G. Bhanji, "Binary communication with a filtered Poisson process in white Gaussian noise," Ph.D. dissertation, University of Maryland, 1972.

[18] A. O. Hero, "Timing estimation for a filtered Poisson process in Gaussian noise," *IEEE Transactions on Information Theory*, vol. 37, no. 1, pp. 92–106, Jan. 1991.

[19] D. L. Snyder, "Applications of stochastic calculus for point process models arising in optical communication," in *Communication Systems and Random Process Theory*, J. K. Skwirzynski, Ed., Sijthoff and Noordhoff. Alphen aan der Rijn, The Netherlands, 1978, pp. 789–804.

[20] I. B. Rhodes and D. L. Snyder, "Estimation and control performance for space-time point-process observations," *IEEE Transactions on Automatic Control*, vol. AC-22, pp. 338–346, Jun. 1977.

[21] T. S. Wei and R. M. Gagliardi, "Cooprative optical beam tracking performance analysis," in *Acquisition, Tracking, and Pointing II, Proc. SPIE*, vol. 887, 1988, pp. 176–183.

[22] G. Marola, D. Santerini, and G. Prati, "Stability analysis of direct-detection cooperative optical beam tracking," *IEEE Transactions on Aerospace and Electronic System*, vol. 25, no. 3, pp. 325–333, May 1989.

[23] C. C. Davis., *Laser and Electro-Optics: Fundamentals and Engineering*. Cambridge University Press, 2000.

[24] D. L. Snyder, *Random Point Processes*. John Wiley & Sons, Inc., 1975.

[25] J. H. Shapiro, "Imaging and optical communication through atmospheric turbulence," in *Laser Beam Propagation in the Atmosphere*, J. W. Strohbehn, Ed. Berlin Germany: Springer-Verlag, 1978.

[26] J. W. Goodman, *Statistical Optics*. John Wiley & Sons, Inc., 1985.

[27] T. E. Duncan, "Evaluation of likelihood functions," *Information and Control*, vol. 13, no. 1, pp. 62–73, Jul. 1968.

[28] T. Kailath, "A general likelihood-ratio formula for random signals in Gaussian noise," *IEEE Transactions on Information Theory*, vol. IT-15, no. 3, pp. 350–361, May 1969.

[29] E. V. Hoversten, "Optical communication theory," in *Laser Handbook*, F. T. Arecchi and E. O. Schulz-Dubois, Eds., vol. 2. Amsterdam, The Netherlands: North-Holland, 1972, pp. 1805–1862.

[30] S. Karp, R. M. Gagliardi, S. E. Moran, and L. B. Stotts, *Optical Channels: Fibers, Clouds, Water, and the Atmosphere*. Plenum Press, New York, 1988.

[31] R. J. McIntyre, "The distribution of gains in uniformly multiplying avalanche photodiodes: Theory," *IEEE Transactions on Electron Devices*, vol. ED-19, no. 6, pp. 703–713, Jun. 1972.

[32] J. Conradi, "The distribution of gains in uniformly multiplying avalanche photodiodes: Experimental," *IEEE Transactions on Electron Devices*, vol. ED-19, no. 6, pp. 713–718, Jun. 1972.

[33] M. C. Teich, K. Matsuo, and B. E. A. Saleh, "Time and frequency response of the conventional avalanche photodiode," *IEEE Transactions on Electron Devices*, vol. ED-33, no. 10, pp. 1511–1517, Oct. 1986.

[34] K. Matsuo, M. C. Teich, and B. E. A. Saleh, "Noise properties and time response of the staircase avalanche photodiode," *IEEE Transactions on Electron Devices*, vol. ED-32, no. 12, pp. 2615–2623, Dec. 1985.

[35] M. M. Hayat and B. E. A. Saleh, "Statistical properties of the impulse response function of double-carrier multiplication avalanche photodiodes including the effect of dead space," *Journal of Lightwave Technology*, vol. 10, no. 10, pp. 1415–1425, Oct. 1992.

[36] T. Kailath and H. V. Poor, "Detection of stochastic processes," *IEEE Transactions on Information Theory*, vol. 44, no. 6, pp. 2230–2231, Oct. 1998.

[37] A. N. Shiryaev, *Probability*. Springer-Verlag, 1996.

[38] H. L. Royden, *Real analysis*. Macmillan, 1968.

[39] D. L. Snyder and P. M. Fishman, "How to track a swarm of fireflies by observing their flashes," *IEEE Transactions on Information Theory*, vol. IT-21, pp. 692–695, Nov. 1975.

[40] P. M. Fishman and D. L. Snyder, "The statistical analysis of space-time point processes," *IEEE Transactions on Information Theory*, vol. 22, no. 3, pp. 257–274, May 1976.

[41] R. M. Gagliardi and S. Karp, *Optical Communication*, 2nd ed. John Wiley & Sons, Inc., 1995.

[42] J. E. Kolassa, *Series approximation methods in statistics*, ser. Lecture notes in statistics. Springer-Verlag, 1994.

[43] J. A. Gubner, "Filtering results for a joint control-decoding problem in optical communications," Master's thesis, University of Maryland, 1985.

[44] J. M. Lopez and K. Yong, "Laser beam pointing control, acquisition, and tracking," in *Laser Satellite Communication*, M. Katzman, Ed. Prentice-Hall, Inc., 1987, ch. 6, pp. 190–213.

[45] S. Lee, J. W. Alexander, and M. Jeganathan, "Pointing and tracking subsystem design for optical communications link between the international space

station and ground," in *Free-Space Laser Communication Technologies XII, Proc. SPIE*, S. Mecherle, Ed., vol. 3932, 2000, pp. 150–157.

[46] M. E. Rosheim and G. F. Sauter, "Free space optical communication system pointer," in *Free-Space Laser Communication Technologies XV, Proc. SPIE*, G. S. Mecherle, Ed., vol. 4975, Jan. 2003, pp. 126–133.

[47] W. Stoelzner, "Optical configuration and system design," in *Laser Satellite Communication*, M. Katzman, Ed. Prentice-Hall, Inc., 1987, ch. 5, pp. 146–189.

[48] M. Guelman, A. Kogan, A. Kazarian, A. Livne, M. Orenstein, H. Michalik, and S. Arnon, "Acquisition and pointing control for inter-satellite laser communications," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 40, no. 4, pp. 1239–1248, Oct. 2004.

[49] G. G. Ortiz, S. Lee, and J. W. Alexander, "Sub-microradian pointing for deep space optical telecommunications network," in *Proceedings of the* 19^{th} *AIAA International Communications Satellite Systems Conference*, Toulouse, France, 2001.

[50] C. C. Davis, I. I. Smolyaninov, and S. D. Milner, "Flexible optical wireless links and networks," *IEEE Communications Magazine*, vol. 41, no. 3, pp. 51–57, Mar. 2003.

[51] A. Papoulis, *Systems and Transforms with Applications in Optics*. New York: McGraw-Hill, 1968.

[52] M. C. Roggemann, B. M. Welsh, and R. Q. Fugate, "Improving the resolution of ground-based telescopes," *Reviews of Modern Physics*, vol. 69, no. 2, pp. 437–506, Apr. 1997.

[53] D. S. Bernstein, *Matrix Mathematics: Theory, Facts, and Formulas with Application to Linear Systems Theory*. Princeton, N.J.: Princeton University Press, 2005.

[54] G. Freiling and A. Hochhaus, "On a class of rational matrix differential equations arising in stochastic control," *Linear Algebra and its Applications*, vol. 379, pp. 43–68, Mar. 2004.

[55] W. Kang, A. Sparks, and S. Band, "Coordinated control of multisatellite systems," *Journal of Guidance, Control, and Dynamics*, vol. 24, no. 2, pp. 360–368, Apr 2001.

[56] M. B. Milam, N. Petit, and R. M. Murray, "Constrained trajectory generation for micro-satellite formation flying," in *AIAA Guidance, Navigation, and Control Conference and Exhibit*, Montreal, Canada, Aug. 2001.

[57] J. M. Kovalik, W. H. Farr, C. Esproles, and H. Hemmati, "Optical communication system with range and attitude measurement capability," IPN Progress Report 42-161, http://tmo.jpl.nasa.gov/progress_report/42-161/161Q.pdf, May 2005.

[58] P. Dayan and L. Abbott, *Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems*. Cambridge, Mass.: Massachusetts Institute of Technology Press, 2001.

[59] F. Rieke, D. Warland, R. de Ruyter van Steveninck, and W. Bialek, *Spikes: Exploring the Neural Code*. Cambridge, Mass.: MIT Press, 1997.

[60] N. Twum-Danso and R. Brockett, "Trajectory estimation from place cell data," *Neural Networks*, vol. 14, pp. 835–844, 2001.

[61] E. N. Brown, L. M. Frank, D. Tang, M. C. Quirk, and M. A. Wilson, "A statistical paradigm for neural spike train decoding applied to position prediction from ensemble firing patterns of rat hippocampal place cells," *The Journal of Neuroscience*, vol. 18, pp. 7411–7425, Sep. 1998.

[62] E. N. Brown, D. P. Nguyen, L. M. Frank, M. A. Wilson, and V. Solo, "An analysis of neural receptive field plasticity by point process adaptive filtering," *Proceedings of the National Academy of Sciences*, vol. 98, no. 21, pp. 12 261–12 266, Oct. 2001.