Quantum Control and Statistical Information

by

Maison Clouatre

B.S.E. in Electrical Engineering, Mercer University, 2022
B.S. in Mathematics, Mercer University, 2022

Submitted to the Department of Aeronautics and Astronautics
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE IN AERONAUTICS AND ASTRONAUTICS

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

February 2026

© 2026 Maison Clouatre. All rights reserved.

The author hereby grants to MIT a nonexclusive, worldwide, irrevocable, royalty-free license to exercise any and all rights under copyright, including to reproduce, preserve, distribute and publicly display copies of the thesis, or release the thesis under an open-access license.

Authored by: Maison Clouatre
Department of Aeronautics and Astronautics
January 16, 2026

Certified by: Moe Z. Win
Robert R. Taylor Professor
Thesis Supervisor

Accepted by: Jonathan How
Ford Professor of Engineering
Chair, Graduate Program Committee

Quantum Control and Statistical Information

by

Maison Clouatre

Submitted to the Department of Aeronautics and Astronautics
on January 16, 2026 in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE IN AERONAUTICS AND ASTRONAUTICS

ABSTRACT

Quantum control plays a crucial role in quantum inference tasks such as parameter estimation and state discrimination. Despite this, the fields of quantum control and quantum inference have remained largely disjoint. We bridge these fields, revealing that (i) quantum control determines the fundamental performance limits of quantum estimation and state discrimination, and (ii) control of statistical information measures enables performance approaching these limits. This thesis begins by establishing an axiomatic formalization of quantum control theory and developing a mathematical framework for analyzing, approximating, and controlling the evolution of quantum systems. A self-contained treatment of optimal quantum control is presented. This treatment exploits the unique structure of quantum dynamics rather than merely borrowing notions from classical control. Quantum control systems are shown to be universally contractive, which plays an important role in the treatment of optimal quantum control and in the derivation of various approximation results. Next, the thesis proceeds by addressing statistical inference problems, with emphasis on parameter estimation and state discrimination. Controlled information measures are introduced, providing more sensible alternatives to existing information measures for quantum inference, e.g., quantum Fisher information. Controlled information measures are shown to exhibit useful properties, e.g., data-processing inequalities. Methods for deriving achievability and converse bounds on controlled information measures are developed, and adaptive control techniques for information maximization are proposed. This thesis leverages a cross-pollination of control theory, statistical inference, and quantum information towards the development of an application-focused theory of quantum inference.

Thesis supervisor: Moe Z. Win
Title: Robert R. Taylor Professor

Acknowledgments

My trajectory has been profoundly influenced by mentors that stretch, challenge, and inspire: Professors Moe Z. Win (MIT), Stefano Marano (University of Salerno), Andrea Conti (University of Ferrara), Peter L. Falb (MIT/Brown), Makhin Thitsa (Mercer University), and Mark Balas (Texas A&M University), among others. I dedicate this thesis to them. The topic of this research—*quantum information control*—was catalyzed by intimate connections between control (the topic that first made me fall in love with mathematics) and statistical inference (an area that my advisor Win is an expert in). The idea for quantum information control was formed when I took Win's course on statistical inference at MIT in Fall 2023. The research took off in collaboration with Marano, Falb, and Conti.

I'd like to especially thank Professor Win for his guidance and insights. He is relentlessly passionate about technical finesse, expositional clarity, and impact. The world needs more of his "LIDS-style thinking". He's also a humanist who has taught me much about life beyond the technical details. I'm also especially grateful for my undergraduate advisor Professor Thitsa. She introduced me to the field of control theory and fought to give me every opportunity possible. I'd like to thank my family and friends for always supporting my dreams; my peers at MIT and the University of Ferrara for challenging me to grow both technically and personally; Professor Marano for his technical precision and for countless hours spent working out the finer details of our manuscripts (including providing a critical review of this thesis); Professor Conti for his charisma and eye for mathematical elegance; Professor Falb for his boldness and originality of thought; and Professor Balas for his insatiable excitement for mathematics as well as for many shared meals at Cooper's BBQ. Lastly, I'd like to thank Professor Daniel Stroock (MIT) for a truly special independent study on measure theory in Spring 2024.

My graduate studies have been supported, in part, by the National Defense Science and Engineering Graduate (NDSEG) Fellowship, the MathWorks Fellowship, and the National Science Foundation (NSF) under grant ECCS-2430953.

Contents

List of Figures 9

Notation 13

Acronyms 17

1 Introduction **19**

2 Quantum control systems **23**

2.1 Quantum states and measurements . . . 23
2.2 Abstract quantum control systems . . . 25
2.3 Contractivity of quantum control systems . . . 27
2.4 Differential quantum control systems . . . 29
2.4.1 Equations of motion . . . 29
2.4.2 Unitary quantum systems . . . 34
2.4.3 Unital quantum systems . . . 35
2.4.4 Approximating quantum dynamics . . . 39
2.4.5 On the "open" and "closed" terminology . . . 41

3 Optimal quantum control **43**

3.1 Differentiation in Banach and Hilbert spaces . . . 43
3.2 Optimal control laws . . . 49
3.2.1 Gradient based optimization . . . 53
3.2.2 Minimum principle for quantum systems . . . 54

4 Controlled Fisher information for quantum estimation **57**

4.1 Classical and quantum Fisher information . . . 57
4.2 Controlled Fisher information . . . 59
4.2.1 Properties of controlled Fisher information . . . 61
4.3 Adaptive information control . . . 63
4.3.1 ϵ-coverings of information controls . . . 65
4.3.2 Adaptation via maximum likelihood estimation . . . 67
4.3.3 Adaptation via composite estimation . . . 72
4.3.4 Interplay between estimation and control in retrospective . . . 75
4.4 Designing and certifying information controls . . . 75
4.4.1 Greedy information control strategy . . . 76
4.4.2 Projected gradient ascent algorithm . . . 76

4.4.3 Certification of representative control strategies 78
4.5 Case studies: Information control for quantum estimation 79
4.5.1 Linear parameter encoding 79
4.5.2 Nonlinear parameter encoding 84

5 Controlled f_γ-divergence for quantum state discrimination **89**
5.1 Classical hypothesis testing and quantum state discrimination 89
5.2 Controlled f_γ-divergence 92
5.2.1 Converse bounds on controlled f_γ-divergence 94
5.3 Universally optimal state discrimination via control 99

6 Conclusion **107**

A Facts from measure theory and analysis **109**

B Supplemental material for Theorem 2.3 **111**
B.1 Supplemental definitions 111
B.2 Existence and uniqueness of solutions 112
B.3 Convergence proofs 115

List of Figures

1.1 **Visualization of an adaptive information controller.** There is a quantum system equipped with a fixed measurement apparatus, which is represented mathematically by a positive operator-valued measure (POVM) $\mathcal{M}$. The goal is to infer a parameter θ encoded in the initial state $\boldsymbol{\Xi}_\theta(t_0)$ of the quantum system. To improve the information about θ extracted by a fixed measurement apparatus, a control law $\boldsymbol{u}(t)$ is applied to the system over a time interval $[t_0, t_\mathrm{f}]$, and then the state of the system is measured at time t_f. The random outcome y of the measurement is used to form an estimator $\hat{\theta}(\mathsf{y})$ of the unknown parameter. The estimator is fed back to the controller for improved information extraction during subsequent experiments. 22

2.1 **Visualization of the sublinear error propagation property (2.15) when $n = 3$.** The goal is to transfer an initial operator $\boldsymbol{X}_0$ to a target operator $\boldsymbol{X}_3$ via an appropriate choice of control law $\boldsymbol{u} \in \Omega$. Rather than directly address the global transfer (which may be a convoluted problem), the problem is broken into simpler sub-transfers from $\boldsymbol{X}_{i-1}$ to $\boldsymbol{X}_i$, $i \in \mathbb{N}_1^3$. The control transfers $\boldsymbol{X}_{i-1}$ to $\boldsymbol{X}_i$ over time intervals $[t_{i-1}, t_i]$ with transfer error $\boldsymbol{E}_i$ for $i \in \mathbb{N}_1^3$. Because quantum systems are universally contracting, the true trajectory (the effect of $\boldsymbol{u}$ on $\boldsymbol{X}_0$ over the full time interval $[t_0, t_3]$) approximately transfers $\boldsymbol{X}_0$ to $\boldsymbol{X}_3$. The error between $\Phi(t_3; t_0, \boldsymbol{X}_0, \boldsymbol{u})$ and $\boldsymbol{X}_3$ is bounded by $\sum_{i=1}^3 \|\boldsymbol{E}_i\|_1$. 28

4.1 **Estimating the depolarization of a superconducting qubit.** For the case $\tau_\mathrm{dep} = 5\,\mu$s, (a) depicts the information control law $u^{(1)}(t)$ for the representative parameter value $\theta_\mathrm{r}^{(1)} = 0.5$ generated by the greedy information control strategy and the gradient ascent algorithm. The optimized measurement time for these two methods are $t_\mathrm{f}^{(1)} = 2\,$ns and $t_\mathrm{f}^{(1)} = 1.4546\,$ns, respectively. For various values of τ_dep, (b) depicts the quantum Fisher information (QFI), the controlled Fisher information (CFI) converse (4.102), and the $(u^{(1)}, t_\mathrm{f}^{(1)})$-CFI achieved through gradient ascent. Tight CFI converse and achievability results reveal the CFI to high accuracy, e.g., in the case $\tau_\mathrm{dep} = 5\,\mu$s the CFI is known within precision 0.2195. In this case, the CFI is roughly 70% (resp. 10%) tighter than the QFI for large (resp. small) parameter values. 81

4.2 **State estimation case study.** (a) Depicts the QFI, the CFI converse (4.102), and the $(u^{(i)}, t_{\mathrm{f}}^{(i)})$-CFI achieved by the two representative control strategies. Note that the CFI converse and achievability bounds are tight; the bound provided by the QFI is loose. (b) Depicts Monte Carlo simulations of the composite estimator $\hat{\theta}_{\mathrm{cp}}(\mathbf{y}_{1:n})$ for two different values of the true parameter θ_0. The dark blue (resp. red) line denotes the sample mean of the Monte Carlo simulation outcomes in the case that $\theta_0 = 0.2$ (resp. $\theta_0 = -0.2$). The light blue and light red shaded regions denote the range of the simulation outcomes. 82

4.3 **State estimation case study.** The average Fisher information extracted by the adaptive information controller $(\mathcal{R}, \hat{\theta}_{\mathrm{cp}}(\cdot), \pi_{\mathrm{n}}(\cdot))$ as a function of number of experiments n in the case that $\theta_0 = 0.35$. The dark blue line denotes the sample mean of the Monte Carlo simulation outcomes; the light blue shaded area denotes the range of the simulation outcomes. The vertical axis of the plot scales linearly between the worst-case Fisher information 6.32 and the best-case Fisher information 6.63. 82

4.4 **Visualization of magnetic field sensing.** The three phases of the sensing process: preparation, encoding, and measurement. The encoding phase occurs over a long exposure interval of duration t_{exp}. The measurement phase occurs over the shorter time interval of maximum duration $t_{\mathrm{f}} - t_0$. Control is leveraged in the preparation and measurement phases. The effect of the weak magnetic field is tangible only over the long exposure interval. 85

4.5 **Magnetic field sensing case study.** (a) Compares the Fisher information achieved by the set of representative controls to the QFI and to the Fisher information achieved by an ideal Ramsey experiment. (b) Depicts the greedy decision rule $\pi_{\mathrm{g}}(\theta)$. By inspection, this decision rule is robust on Θ and, due to the finite number of discontinuities, is strongly robust at all but finitely many $\theta \in \Theta$. 87

4.6 **Magnetic field sensing case study.** (a) Depicts Monte Carlo simulations of the composite estimator $\hat{\theta}_{\mathrm{ml}}(\mathbf{y}_{1:n})$ for two different values of the true parameter θ_0. The dark red (resp. blue) denotes the sample mean of the Monte Carlo simulation outcomes in the case that $\theta_0 = 1.3$ (resp. $\theta_0 = 0.8$). The light red and light blue shaded regions denote the range of simulation outcomes. (b) Depicts the average Fisher information extracted by the adaptive information controller $(\mathcal{R}, \hat{\theta}_{\mathrm{ml}}(\cdot), \pi(\cdot))$ across the simulations in the case $\theta_0 = 1.3$. The vertical axis of the plot scales linearly between the worst-case Fisher information 0.22 and the best-case Fisher information 2.71. 88

5.1 **Visualization of the f_γ-divergence's evolution under control.** There are two instantiations of the system, one initially in state $\boldsymbol{\Xi}_0(t_0)$ and one initially in state $\boldsymbol{\Xi}_1(t_0)$ at time t_0. These instantiations are controlled jointly using the same control law over the time interval $[t_0, t_{\mathrm{f}}]$, resulting in tethered evolution as depicted by the lines connecting the trajectories of each system. The states are poorly distinguishable at the initial time t_0, but control makes them more easily distinguishable at a later time t_{f}. 93

5.2 **Visualization of the concepts involved in the proof of Theorem 5.3.** (a) Visualizes the approximation concept where controls are designed for the idealized system and then applied to the true system. Lemma 5.3 shows that the error between the true and idealized trajectories is bounded by the integrated environment-induced dissipation rates. (b) Visualizes the decomposition of a global transfer into sub-transfers. The sublinear error propagation property of Theorem 2.2, which stems from the fact that the system trajectories are naturally contracting, implies that the true trajectory remains close to the sub-transfer trajectories. (c) Visualizes the population transfer between the system's energy levels while under idealized Lindblad control, as explained by Lemma 5.4. 102

5.3 **Visualization of the state transfers involved in the proof of Theorem 5.3.** Visualization of the transformation from the initial matrix $\boldsymbol{Z}$ to the desired matrix $\boldsymbol{Z}_{\text{des}}$. This is decomposed into three sub-transfers. Hamiltonian control transforms $\boldsymbol{Z}$ to $\boldsymbol{Z}_1$, which diagonalizes $\boldsymbol{Z}$ in the eigenbasis of the free Hamiltonian $\boldsymbol{H}_0$. The dashed line in the figure separates positive and negative eigenvalues of $\boldsymbol{Z}$. Next, Lindblad control maps the positive eigenvalues of $\boldsymbol{Z}$ to the span of $|h_1\rangle$ (the eigenvector corresponding to the largest eigenvalue of $\boldsymbol{H}_0$) and maps the negative eigenvalues of $\boldsymbol{Z}$ to the span of $|h_m\rangle$ (the eigenvector corresponding to the largest negative eigenvalue of $\boldsymbol{Z}$). Finally, Hamiltonian control transforms $\boldsymbol{Z}_2$ to $\boldsymbol{Z}_{\text{des}}$, which diagonalizes $\boldsymbol{Z}_2$ in the eigenbasis of $\boldsymbol{M}$. 104

Notation

Random variables are displayed in sans serif, upright fonts; their realizations in serif, italic fonts. Vectors are denoted in bold lowercase font; matrices and operators are denoted in bold uppercase font. For example, a random variable and its realization are denoted by x and x; a random vector and its realization are denoted by $\mathbf{x}$ and $\boldsymbol{x}$; a random operator and its realization are denoted $\mathbf{X}$ and $\boldsymbol{X}$, respectively. The (i,j)th element of a matrix $\boldsymbol{X}$ is denoted $[\boldsymbol{X}]_{ij}$. The notation $\mathsf{x} \sim P$ means that x is distributed according to distribution P. Given numbers $x_1, x_2, \ldots, x_n$, the notation $\boldsymbol{x}_{1:n} = [x_1\ x_2\ \ldots\ x_n]$ is used. The Dirac notation $|\psi\rangle$ denotes an element of a Hilbert space and $\langle\psi|$ denotes its dual. Accents as in $\breve{x}$ and $\hat{x}$ are used to denote lower and upper bounds on a quantity x. A summation such as $\sum_{k=m}^{n} x_k$ is defined to be zero if $n < m$. Commonly used symbols are defined below.

Commonly used symbols

$\imath$	the imaginary unit $\sqrt{-1}$
$\varnothing$	empty set
$\mathbb{R}, \mathbb{C}, \mathbb{N}$	set of real, complex, and natural numbers
$\mathbb{R}_{\geqslant 0}$	set of non-negative real numbers
$\mathbb{C}_{\mathrm{H}}^{d\times d}$	set of complex d-by-d Hermitian matrices
$\mathbb{N}_k^n$	set of natural numbers between k and n if $k \leqslant n$, empty set otherwise
$\operatorname{Re} z$, $\operatorname{Im} z$	real and imaginary parts of $z \in \mathbb{C}$, respectively
$a \wedge b$	minimum of a and b
$a \vee b$	maximum of a and b
$\triangleq$	equal by definition
$\tilde{\forall}$	for almost every, with respect to Lebesgue measure
$\subseteq, \subset$	subset and strict subset
$(\mathcal{Y}, \mathscr{E})$	measurable space
$(\Omega, \mathscr{F}, \mathbb{P})$	probability space
$L^p(\Omega)$	space of Lebesgue p-integrable functions $f : \Omega \to \mathbb{R}$
$L^p(\Omega; \mathbb{R}^n)$	space of Lebesgue p-integrable functions $f : \Omega \to \mathbb{R}^n$
$\mu(\cdot)$	Lebesgue measure
$\mathbb{E}\{\cdot\}, \mathbb{V}\{\cdot\}$	expectation and variance
$\mathbb{1}_{\mathcal{X}}(\cdot)$	indicator function of a set $\mathcal{X}$

$\boldsymbol{\lambda}(\boldsymbol{A})$	vector of eigenvalues of $\boldsymbol{A} \in \mathbb{C}^{d\times d}$
$\lambda_k^{\downarrow}(\boldsymbol{A})$	kth largest eigenvalue of $\boldsymbol{A} \in \mathbb{C}_{\mathrm{H}}^{d\times d}$
$\lambda_k^{\uparrow}(\boldsymbol{A})$	kth smallest eigenvalue of $\boldsymbol{A} \in \mathbb{C}_{\mathrm{H}}^{d\times d}$
$\lambda_{\max}(\boldsymbol{A})$	largest eigenvalue of $\boldsymbol{A} \in \mathbb{C}_{\mathrm{H}}^{d\times d}$
$\lambda_{\min}(\boldsymbol{A})$	smallest eigenvalue of $\boldsymbol{A} \in \mathbb{C}_{\mathrm{H}}^{d\times d}$
$\|\cdot\|_{\mathcal{X}}$	norm on a Banach space $\mathcal{X}$
$\|\cdot\|_{L^p}$	Lebesgue p-norm
$\|\cdot\|_{\mathrm{op}}$	operator norm
$\|\cdot\|_1$	trace norm
$\|\cdot\|_{\mathrm{HS}}$	Hilbert-Schmidt norm
$\|\cdot\|_p$	Schatten p-norm
$\|\cdot\|$	Euclidean norm
$\langle\cdot,\cdot\rangle_{\mathcal{H}}$	inner product on $\mathcal{H}$
$\langle\cdot,\cdot\rangle_{\mathrm{HS}}$	Hilbert-Schmidt inner product
$\langle\cdot,\cdot\rangle$	Euclidean inner product
$\dagger$	adjoint $\boldsymbol{A}^{\dagger}$ of $\boldsymbol{A} : \mathbb{C}^d \to \mathbb{C}^d$ w.r.t. $\langle\cdot,\cdot\rangle$
$\ddagger$	adjoint $F^{\ddagger}$ of $F : \mathbb{C}^{d\times d} \to \mathbb{C}^{d\times d}$ w.r.t. $\langle\cdot,\cdot\rangle_{\mathrm{HS}}$
$\mathrm{rank}\{\cdot\}$	rank
$\mathrm{krank}\{\cdot\}$	Kraus rank
$\mathrm{tr}\{\cdot\}$	trace
$[\![\boldsymbol{A},\boldsymbol{B}]\!]_-$	commutator of $\boldsymbol{A}$ and $\boldsymbol{B}$, i.e., $[\![\boldsymbol{A},\boldsymbol{B}]\!]_- = \boldsymbol{A}\boldsymbol{B} - \boldsymbol{B}\boldsymbol{A}$
$[\![\boldsymbol{A},\boldsymbol{B}]\!]_+$	anti-commutator of $\boldsymbol{A}$ and $\boldsymbol{B}$, i.e., $[\![\boldsymbol{A},\boldsymbol{B}]\!]_+ = \boldsymbol{A}\boldsymbol{B} + \boldsymbol{B}\boldsymbol{A}$
$\mathcal{H}$	complex Hilbert space
$\dim\mathcal{H}$	dimension of $\mathcal{H}$
$\mathcal{L}(\mathcal{H})$	linear operators on $\mathcal{H}$
$\mathcal{D}(\mathcal{H})$	density operators on $\mathcal{H}$
$\mathcal{C}(\mathcal{H})$	set of quantum channels on $\mathcal{H}$
$\mathcal{C}_{N_{\mathrm{k}}}(\mathcal{H})$	set of quantum channels on $\mathcal{H}$ with Kraus rank no larger than N_{k}
Σ	quantum control system
T	time domain of Σ
$\mathcal{V}$	control inputs of Σ
Ω	control laws of Σ
Φ	state transition map of Σ
$(\Upsilon,\mathscr{E},\boldsymbol{M})$	measurable space endowed with a POVM
ψ	output map of Σ
$(\boldsymbol{u}, t_{\mathrm{f}})$	control strategy
N_{c}	number of control inputs
F	generator of differential system's dynamics

$\hbar$	reduced Planck constant
$\boldsymbol{H}$	Hamiltonian operator
$\boldsymbol{L}$	Lindblad operator
$\mathcal{M}$	finite cardinality POVM
N_{m}	cardinality of $\mathcal{M}$
$\boldsymbol{P}_x, \boldsymbol{P}_y, \boldsymbol{P}_z$	Pauli-x, -y, and -z operators, respectively
$\tau_{\mathrm{rel}}, \tau_{\mathrm{dep}}$	qubit relaxation and depolarization times, respectively
$D_x f(x)[\cdot]$	Fréchet derivative of $f : \mathcal{X} \to \mathbb{R}$ at x
$\nabla_x f$	gradient of $f : \mathcal{H} \to \mathbb{R}$ at x, i.e., $\langle \nabla_x f, \cdot \rangle_{\mathcal{H}} = D_x f(x)[\cdot]$
$S(\cdot)$	von Neumann entropy
$S_{\mathrm{s}}(\cdot)$	Shannon entropy
$J_{\mathrm{y}}(\theta)$	Fisher information
$J^{\mathrm{q}}_{\boldsymbol{\Xi}_\theta}(\theta)$	quantum Fisher information
$J^{\mathrm{c}}_{\boldsymbol{\Xi}_\theta}(\theta; \Sigma)$	controlled Fisher information
$D_\gamma(P_0 \Vert P_1)$	f_γ-divergence
$D^{\mathrm{q}}_\gamma(\boldsymbol{\Xi}_0 \Vert \boldsymbol{\Xi}_1)$	quantum f_γ-divergence
$D^{\mathrm{c}}_\gamma(\boldsymbol{\Xi}_0 \Vert \boldsymbol{\Xi}_1; \Sigma)$	controlled f_γ-divergence
$D(P_0 \Vert P_1)$	Kullback-Leibler divergence
$\boldsymbol{x} \succ_{\mathrm{m}} \boldsymbol{y}$	vector $\boldsymbol{x}$ majorizes vector $\boldsymbol{y}$
$\boldsymbol{A} \succ_{\mathrm{m}} \boldsymbol{B}$	Hermitian operator $\boldsymbol{A}$ majorizes Hermitian operator $\boldsymbol{B}$
$\boldsymbol{A} \succcurlyeq \boldsymbol{B}$	operator $\boldsymbol{A} - \boldsymbol{B}$ is positive semidefinite
$\boldsymbol{A} \succ \boldsymbol{B}$	operator $\boldsymbol{A} - \boldsymbol{B}$ is positive definite
$U(\mathcal{H})$	group of unitary operators on $\mathcal{H}$
$U(d)$	group of d-by-d unitary matrices
$SU(d)$	group of d-by-d special unitary matrices
$su(d)$	algebra of d-by-d traceless, skew Hermitian matrices
Θ	parameter space, compact subset of $\mathbb{R}$
Θ_{r}	set of representative parameters
$\theta^{(i)}_{\mathrm{r}}$	representative parameter
θ_0	true parameter
$\boldsymbol{\Xi}_\theta$	parameterized quantum state
$\mathcal{R}$	set of representative control strategies
$(\theta^{(i)}_{\mathrm{r}}, \boldsymbol{u}^{(i)}, t^{(i)}_{\mathrm{f}})$	representative control strategy
N_{r}	number of representative control strategies
π	decision rule
$\Pi_\epsilon(\theta)$	set of ϵ-information maximizing controls at $\theta \in \Theta$
$\pi_{\mathrm{n}}, \pi_{\mathrm{g}}$	nearest-neighbor and greedy decision rules, respectively

$\hat{\theta}(\cdot)$	estimator of an unknown parameter θ
$\hat{\theta}_{\mathrm{ml}}(\cdot)$	maximum likelihood estimator
$\hat{\theta}_{\mathrm{cp}}(\cdot)$	composite estimator

Acronyms

CFI	controlled Fisher information
DPI	data processing inequality
FPGA	field-programmable gate array
IID	independent and identically distributed
NV	nitrogen vacancy
ODE	ordinary differential equation
POVM	positive operator-valued measure
QFI	quantum Fisher information

Chapter 1

Introduction

Control sits at the heart of the contemporary quantum engineering, and it plays a crucial role in the quantum inference tasks of estimation [1, 2] and state discrimination [3, 4]. While this role has been recognized, the fields of quantum control and quantum inference have remained largely disjoint. This thesis bridges said gap, revealing that (i) quantum control determines the fundamental performance limits of quantum estimation and state discrimination, and (ii) control of statistical information measures enables performance approaching said limits. Two main contributions of this thesis are the axiomatization of quantum control theory and the introduction of controlled information measures.

Quantum control systems

The central object in classical systems theory (see Kalman, Falb, and Arbib [5]; Falb [6]) is a *control system*—a mathematical abstraction formalizing states, inputs, outputs, and the mappings between them. Chapter 2 defines *quantum* control systems and investigates the immediate consequences of this definition. Equations of motion, classes of dynamics (unital, unitary), and design of optimal control laws are discussed. The focus is finite-dimensional systems; however, many of the results naturally extend to infinite-dimensional quantum systems on separable Hilbert space.

The first notable differences between a quantum control system and a classical control system are that (i) a quantum state transition map has the additional requirement of being completely positive and trace-preserving, and (ii) the act of measurement affects a quantum system's state. Chapter 2 illustrates how this additional structure leads to certain important consequences for quantum systems. For instance, for any control input, quantum systems are contracting under the trace norm. This universal contractivity property underlies a key result that we call the *sublinear error propagation*, which is central to the design of quantum control strategies. Chapter 2 defines differential quantum control systems, i.e., quantum systems whose state transition maps arise from (Lindblad) differential equations, which typically occur in practical applications. The classes of unital and unitary systems are studied; the sublinear error propagation property is shown to hold for arbitrary Schatten p-norms in unital systems; necessary and sufficient conditions for unitality of a differential system are derived. The approximation of a quantum system's trajectory by that of a simpler system (e.g., a simpler system that ignores the effects of certain Lindblad operators) is studied, and a bound on the

normed error between the two trajectories is provided. This bound is exponentially tighter than bounds based on Grönwall's inequality (a major workhorse of classical control theory [6, p. 64]).

Chapter 3 addresses optimal control of differential quantum systems. Compared to classical optimal control, we find that the additional structure of differential quantum systems enables an expedited and self-contained treatment of optimality. Necessary conditions for optimality of control laws are given, and a quantum minimum principle (also known as Pontryagin's maximum principle) is derived. The results on quantum control systems in Chapters 2 and 3 are derived using a first-principles approach rather than merely borrowing known results from classical control theory. The sublinear error propagation property, contraction-based approximation techniques, and self-contained treatment of optimal quantum control are examples of the fruit to bear from this first-principles approach.

Controlled information measures

Controlled information measures quantify the amount of statistical information, regarding an estimation or discrimination task, that can be extracted from a physical quantum system. To the best of our knowledge, this thesis and our paper [2] are the first works to introduce notions of statistical information based on quantum control. To define controlled information measures and demonstrate their utility, it is helpful to recall notions of statistical information commonly used in quantum parameter estimation and state discrimination.

The problem of quantum parameter estimation arises in many applications including quantum sensing, computing, and communications. The goal is to infer an unknown parameter that is encoded into the state of a quantum system. When the state is measured by any particular quantum measurement apparatus, the optimal (asymptotic) estimation performance is quantified by the classical Fisher information [7]. This information can be calculated using Born's rule, which depends on both the measurement apparatus and the parameterization of the quantum state. The ultimate performance limit of quantum estimation (the quantum Cramér-Rao bound [8, 9]) is quantified by the quantum Fisher information (QFI), which is defined as the supremum of the classical Fisher information over all possible quantum measurements [10]. QFI analysis has led to theoretical insights and is used, e.g., to predict quantum advantage in sensing [11].

The problem of quantum state discrimination is crucial to many applications including quantum communications and quantum radar. This thesis focuses on binary state discrimination where the goal is to determine which of two possible states a quantum system is in. When the unknown state is measured by any particular quantum measurement apparatus, the optimal discrimination error probability (DEP) is determined by the f_γ-divergence (see Chapter 5 for the definition of this quantity). The ultimate performance limit of quantum state discrimination (the Helstrom bound [12]) is quantified by the quantum f_γ-divergence. The Helstrom bound has led to theoretical insights and is used, e.g., to design quantum states for communication [13]. We show that the quantum f_γ-divergence can be defined as the supremum of the f_γ-divergence over all possible quantum measurements, thus providing a variational characterization of the quantum f_γ-divergence.

Through the parameter estimation and state discrimination problems described above, a trend emerges:

1. the best-obtainable performance for a statistical inference task using any particular measurement apparatus is quantified by a classical information measure; and

2. a quantum information measure is defined as the supremum of the classical measure over all possible quantum measurements.

While this trend has led to many innovations, the resulting quantum information measures are not always indicative of the performance that one can achieve in practical scenarios. In state discrimination, for example, restricted measurements are routinely used; these measurements may not achieve the Helstrom bound [14, 15, 16, 17, 18]. This restriction is often because a quantum system has been designed with a general-purpose measurement apparatus or because a sub-optimal measurement apparatus is easier to implement [19].

These considerations motivate us to consider the practical scenario where a quantum system is equipped with a single fixed measurement apparatus. To increase information extraction, we allow the system to evolve under the influence of carefully-designed control laws prior to employing the fixed measurement apparatus—see Fig. 1.1. This crucial step encodes more information into the output of the measurement than would be achievable without control. We define controlled information measures as the maximal amount of statistical information that can be extracted via control. These measures are system-dependent and lead, by construction, to achievable performance limits. The axiomatization of quantum control established in Chapter 2 provides the mathematical basis for these developments. Chapter 4 focuses on estimation and controlled Fisher information (CFI); Chapter 5 focuses on state discrimination and controlled f_γ-divergence. The ideas developed in Chapters 4 and Chapter 5 lead to new paradigms, one example of which is the adaptive information controller depicted in Fig. 1.1. This controller works in tandem with a parameter estimator to encode the maximal amount of information into the output of the system's fixed measurement apparatus.

Bibliographic remarks

The "modern" functional analytic approach to classical systems theory began around the 1960s with works like [5, 20, 21]; recent treatments of classical systems theory are given in [6, 22]. Quantum control began to receive considerable interest during the quantum information revolution of the 1990s [23, 24]. The books [25, 26, 27, 28] and tutorials [29, 30, 31, 32, 33] provide a comprehensive overview of the subject and recent developments. The theory developed in Chapter 2 has been greatly inspired by the aforementioned literature. Our goal is to provide a reasonably general mathematical framework to accommodate emerging quantum control modalities, such as incoherent control [34, 35, 36, 37]. In particular, control laws are allowed to enter the Lindblad operators in the definition of a differential quantum control system, which is necessary for incoherent control. The contraction phenomena established for abstract quantum systems has been foretold by the work [38] on differential quantum systems.

Optimal quantum control, the subject of Chapter 3, has been widely studied [39, 40, 41]. The goal herein is to provide a self-contained treatment of optimality starting from the

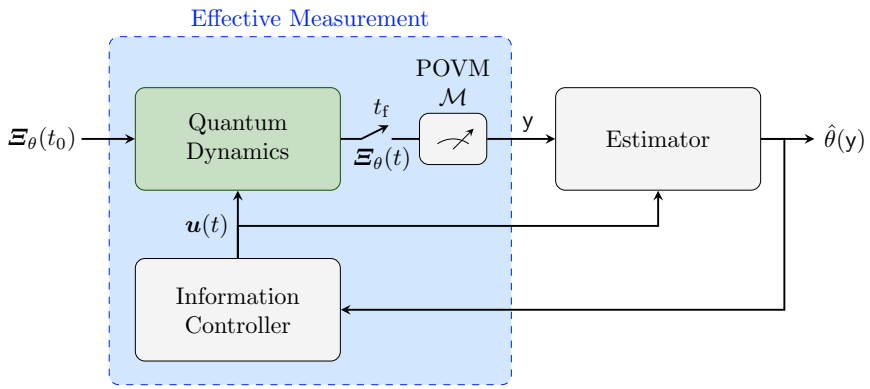


Figure 1.1: Visualization of an adaptive information controller. There is a quantum system equipped with a fixed measurement apparatus, which is represented mathematically by a positive operator-valued measure (POVM) $\mathcal{M}$. The goal is to infer a parameter θ encoded in the initial state $\boldsymbol{\Xi}_\theta(t_0)$ of the quantum system. To improve the information about θ extracted by a fixed measurement apparatus, a control law $\boldsymbol{u}(t)$ is applied to the system over a time interval $[t_0, t_\mathrm{f}]$, and then the state of the system is measured at time t_f. The random outcome y of the measurement is used to form an estimator $\hat{\theta}(\mathsf{y})$ of the unknown parameter. The estimator is fed back to the controller for improved information extraction during subsequent experiments.

differential dynamics defined in Chapter 2. This goal has been inspired by the papers [42, 43] (and talks at MIT given by their first author) which take a first-principles approach to optimal control problems in quantum biology. While our focus is analytical, a variety of numerical techniques for generating optimal controls have been established in the literature [44, 45, 46, 47]. It is widely observed that these numerical techniques produce effective controls, and gradient-based algorithms can converge to global optima [48].

The foundations of quantum estimation and detection theory were laid in the 1970s with works like [12]. The statistical inference problems considered in this thesis are inspired by recent works on quantum communication [49, 50, 51, 52], networking [53, 54, 55, 56, 57, 58], and sensing [59, 60, 61]. Quantum estimation and detection have received significant attention by the control community in recent years [62, 63], several works having studied control of quantum information measures like the QFI [64, 65, 66, 67, 68, 69]. While this is an interesting research direction in its own right, practical limitations of controlling the QFI have been pointed out [65, 68, 70]. In particular, if one uses control to optimize the QFI, subsequent parameter estimation requires that a measurement achieving the resulting QFI is physically realizable. In contrast, this thesis considers systems equipped with a fixed measurement apparatus and leverages control to maximize the information extracted from said apparatus.

Chapter 2

Quantum control systems

2.1 Quantum states and measurements

Quantum theory is built upon a separable complex Hilbert space $\mathcal{H}$. We will focus on finite-dimensional systems, so take (without loss of generality) $\mathcal{H} = \mathbb{C}^d$ for some $d \geqslant 2$. The state of a quantum system is represented mathematically by a vector $|\psi\rangle \in \mathcal{H}$. It is convention to normalize this vector such that $\langle\psi|\psi\rangle = \||\psi\rangle\|^2 = 1$.

Often times, the state of a quantum system is not known deterministically. Rather, the system is known to be described by some ensemble of quantum states—this is where quantum and statistical mechanics merge. More precisely, suppose that the state of a quantum system is known to take value $|\psi_i\rangle \in \mathcal{H}$ with probability $p_i \geqslant 0$ for $i \in \mathbb{N}$. This probabilistic knowledge is represented by a *density matrix*[1]

$$\boldsymbol{\Xi} = \sum_{i \in \mathbb{N}} p_i |\psi_i\rangle\langle\psi_i| \tag{2.1}$$

which is unit-trace and positive semidefinite. The set of all possible density matrices on $\mathbb{C}^d$ is denoted

$$\mathcal{D}(\mathbb{C}^d) \triangleq \left\{ \boldsymbol{\Xi} \in \mathbb{C}^{d \times d} \,:\, \mathrm{tr}\{\boldsymbol{\Xi}\} = 1,\, \boldsymbol{\Xi} \succcurlyeq \mathbf{0} \right\} \tag{2.2}$$

where $\mathbf{0}$ is the zero matrix and $\boldsymbol{\Xi} \succcurlyeq \mathbf{0}$ denotes that $\boldsymbol{\Xi}$ is positive semidefinite. Formally, a density matrix $\boldsymbol{\Xi}$ *describes* the state of a quantum system. However, as is convention in the literature, the terminology *quantum state* and *density matrix describing a quantum state* will be used interchangeably henceforth. By the spectral theorem, each quantum state $\boldsymbol{\Xi} \in \mathcal{D}(\mathbb{C}^d)$ has a spectral decomposition of the form

$$\boldsymbol{\Xi} = \sum_{i=1}^{d} \lambda_i |\xi_i\rangle\langle\xi_i| \tag{2.3}$$

where $\lambda_i \geqslant 0$ are eigenvalues summing to 1 and $\{|\xi_i\rangle \in \mathbb{C}^d \,:\, i \in \mathbb{N}_1^d\}$ are orthonormal eigenvectors. The spectral decomposition (2.3) resembles (2.1), and the λ_i's form a probability distribution. However, in (2.1) the $|\psi_i\rangle$'s are not required to be orthogonal.

[1] At this point, it is not a priori clear why the density matrix $\boldsymbol{\Xi}$ is a sensible mathematical object to describe the state of a quantum system. However, it can be shown that the density matrix contains all of the information about the quantum system that can be obtained by measurement. See [71] for details.

There are intimate connections between quantum mechanics and statistical information measures.[2] Investigating such connections forms the backbone of this thesis. Let us begin with the most well-known information measure in quantum theory.

Definition 2.1 (von Neumann entropy)**.** The *von Neumann entropy* of a quantum state $\boldsymbol{\Xi} \in \mathcal{D}(\mathbb{C}^d)$ is

$$S(\boldsymbol{\Xi}) = -\operatorname{tr}\{\boldsymbol{\Xi} \log \boldsymbol{\Xi}\} \tag{2.4}$$

where $\log \boldsymbol{\Xi}$ is the matrix logarithm. □

Equivalently, the von Neumann entropy of $\boldsymbol{\Xi}$ is the Shannon entropy [72] of the probability vector $\boldsymbol{\lambda}(\boldsymbol{\Xi}) \in \mathbb{R}^d$ containing $\boldsymbol{\Xi}$'s eigenvalues. The von Neumann entropy $S(\boldsymbol{\Xi})$ quantifies how much uncertainty an observer has about the true quantum state $|\psi\rangle$. Various quantum control tasks are related to the ability to remove von Neumann entropy from a system [73].

To extract information from a quantum system, one must perform a measurement. A quantum measurement is described by a matrix-valued function $\boldsymbol{M}(\cdot) : \mathscr{E} \to \mathbb{C}^{d\times d}$ on a measurable space $(\Upsilon, \mathscr{E})$ satisfying

1. $\boldsymbol{M}(\emptyset) = \mathbf{0}$ and $\boldsymbol{M}(\Upsilon) = \boldsymbol{I}$,
2. $\boldsymbol{M}(\mathcal{B}) \succcurlyeq \mathbf{0}$ for all $\mathcal{B} \in \mathscr{E}$, and
3. if $\{\mathcal{B}_i : i \in \mathbb{N}\}$ is a disjoint sequence of elements from the sigma algebra $\mathscr{E}$ then

$$\boldsymbol{M}\Big(\bigcup_{i\in\mathbb{N}} \mathcal{B}_i\Big) = \sum_{i\in\mathbb{N}} \boldsymbol{M}(\mathcal{B}_i) \tag{2.5}$$

where the summation (2.5) converges weakly.

A matrix-valued function satisfying 1-3 is said to be a (non-orthogonal) *resolution of the identity* or a *positive operator-valued measure (POVM)* [74, pp. 43-45]. The elements of the sigma algebra $\mathscr{E}$ are the possible outcomes of the quantum experiment. The probability of observing outcome $\mathcal{B} \in \mathscr{E}$ is given by the state-dependent probability measure

$$\mathbb{P}_{\boldsymbol{\Xi}}\{\mathcal{B}\} = \operatorname{tr}\{\boldsymbol{M}(\mathcal{B})\boldsymbol{\Xi}\}\,. \tag{2.6}$$

One can verify using properties 1-3 above that this is indeed a probability measure [75]. The tuple $(\Upsilon, \mathscr{E}, \mathbb{P}_{\boldsymbol{\Xi}})$ is a probability space. Equation (2.6) is known as Born's rule.

Remark 2.1 (Statistical inference problems)**.** In quantum theory, inference problems arise when we wish to infer an unknown parameter $\theta \in \Theta \subseteq \mathbb{R}$ (or vector of parameters) that is encoded in a quantum state $\boldsymbol{\Xi}_\theta$. This gives rise to a parameterized family of probability spaces $(\Upsilon, \mathscr{E}, (\mathbb{P}_{\boldsymbol{\Xi}_\theta})_{\theta\in\Theta})$. By performing experiments and observing the frequencies of outcomes $\mathcal{B} \in \mathscr{E}$, we can infer which probability distribution $\mathbb{P}_{\boldsymbol{\Xi}_\theta}$ the observations are coming from. This in turn provides us with an estimate of the unknown parameter. If the set Θ is a continuum, then the inference problem is known as *parameter estimation*, which is the subject of Chapter 4. If the set Θ is discrete (e.g., $\Theta = \{0, 1\}$), then the inference problem is known as *state discrimination* or *hypothesis testing*, which is the subject of Chapter 5. □

[2] This is expected since the state of a quantum system is represented by a density matrix—a statistical mechanical object.

In finite-dimensional systems (i.e., systems for which this thesis is concerned), the measure space $(\Upsilon, \mathscr{E})$ often takes the form of $(\mathbb{N}_1^{N_\mathrm{m}}, 2^{\mathbb{N}_1^{N_\mathrm{m}}})$. That is, the outcomes of a quantum measurement can be numbered $1, 2, \ldots, N_\mathrm{m}$ for some $N_\mathrm{m} \in \mathbb{N}$. In such cases, we will refer to a POVM as the set

$$\mathcal{M} = \{\boldsymbol{M}_1, \boldsymbol{M}_2, \ldots, \boldsymbol{M}_{N_\mathrm{m}}\} \tag{2.7}$$

where $\boldsymbol{M}_i \triangleq \boldsymbol{M}(\{i\}) \succcurlyeq \boldsymbol{0}$ for $i \in \mathbb{N}_1^{N_\mathrm{m}}$. This set satisfies $\sum_{i=1}^{N_\mathrm{m}} \boldsymbol{M}_i = \boldsymbol{I}$ and the probability of observing outcome i given that the system is in state $\boldsymbol{\Xi} \in \mathcal{D}(\mathbb{C}^d)$ is equal to $\mathrm{tr}\{\boldsymbol{M}_i \boldsymbol{\Xi}\}$. In systems with a finite-cardinality POVM, the measure-theoretic formalism of quantum measurement simplifies and we simply speak in terms of the set $\mathcal{M}$ instead of the tuple $(\Upsilon, \mathscr{E}, \boldsymbol{M})$.

2.2 Abstract quantum control systems

An axiomatic approach to quantum control theory is now introduced. The definition of a quantum channel—the most general transformation of a quantum state allowed by the axioms of quantum mechanics—will be needed.

Definition 2.2 (Completely positive map [76, p. 106])**.** A linear map $K : \mathbb{C}^{d\times d} \to \mathbb{C}^{d\times d}$ is said to be *completely positive* if, for any finite collections of vectors $\{|\phi_n\rangle\}$, $\{|\psi_m\rangle\} \subset \mathbb{C}^d$, it follows that

$$\sum_n \sum_m \langle\phi_n| K\big(|\psi_n\rangle\langle\psi_m|\big) |\phi_m\rangle \geqslant 0\,. \tag{2.8}$$

□

Definition 2.3 (Quantum channel [71, p. 133])**.** A linear map $K : \mathbb{C}^{d\times d} \to \mathbb{C}^{d\times d}$ is said to be a *quantum channel* if K is completely positive and trace-preserving, i.e.,

$$\mathrm{tr}\{K(\boldsymbol{X})\} = \mathrm{tr}\{\boldsymbol{X}\}\,, \qquad \forall \boldsymbol{X} \in \mathbb{C}^{d\times d}\,. \tag{2.9}$$

The set of all quantum channels on $\mathcal{H}$ is denoted $\mathcal{C}(\mathcal{H})$. □

Since a quantum channel should map $\mathcal{D}(\mathbb{C}^d)$ into itself, it is clear that a quantum channel should be positive (i.e., it should map positive semidefinite operators to positive semidefinite operators) and trace-preserving. At a first glance, the complete positivity requirement can be surprising; however, it has physical significance [71].

One can show that the concatenation of two quantum channels is yet another quantum channel. A key result in quantum information is the following, which gives a canonical representation of any quantum channel. Theorem 2.1 will be used in Chapter 5.

Theorem 2.1 (Kraus representation [76])**.** Let $K : \mathbb{C}^{d\times d} \to \mathbb{C}^{d\times d}$ be a quantum channel. Then, there exists operators $\{\boldsymbol{V}_k \in \mathbb{C}^{d\times d} : k \in \mathbb{N}_1^r\}$, $1 \leqslant r \leqslant d^2$, satisfying the completeness relation $\sum_{k=1}^r \boldsymbol{V}_k^\dagger \boldsymbol{V}_k = \boldsymbol{I}_\mathcal{H}$ such that

$$K(\boldsymbol{X}) = \sum_{k=1}^{r} \boldsymbol{V}_k \boldsymbol{X} \boldsymbol{V}_k^\dagger\,. \tag{2.10}$$

Conversely, if K is a map of the form 2.10 then K is a quantum channel. □

Definition 2.4 (Kraus rank)**.** The minimum $r \in \mathbb{N}$ needed to represent a quantum channel $K : \mathbb{C}^{d\times d} \to \mathbb{C}^{d\times d}$ in the form (2.10) is said to be the *Kraus rank* of the channel, which is denoted $\operatorname{krank}\{K(\cdot)\}$. □

We are now ready to define a quantum control system. This is a mathematical description of how quantum states evolve under the effect of control laws and how these effects are measured.

Definition 2.5 (Quantum control system)**.** A *quantum control system* is a tuple $\Sigma = (T, \mathcal{H}, \mathcal{V}, \Omega, \Phi, (\Upsilon, \mathscr{E}, \boldsymbol{M}))$ of the following form:

1. T is the *time* over which the system may evolve, which is a totally ordered set, and $[t_0, t_1] \triangleq \{t \in T \,:\, t_0 \leqslant t \leqslant t_1\}$ is a subset of time for $t_0, t_1 \in T$ with $t_0 \leqslant t_1$;
2. $\mathcal{H} = \mathbb{C}^d$ is the Hilbert space over which the system is defined;
3. $\mathcal{V}$ is the set of *control inputs* and $\Omega \subseteq \{\boldsymbol{u} \,|\, \boldsymbol{u} : T \to \mathcal{V}\}$ is the non-empty set of *control laws* which satisfies the concatenation property that, if $\boldsymbol{u}_1, \boldsymbol{u}_2 \in \Omega$ and $t_a \in T$, then $\boldsymbol{u}_3$ given by $\boldsymbol{u}_3(t) \triangleq \boldsymbol{u}_1(t)$ for $t \leqslant t_a$ and $\boldsymbol{u}_3(t) \triangleq \boldsymbol{u}_2(t)$ for $t > t_a$ is an element of Ω;
4. Φ is the *state transition map* having the properties that
   (a) (Domain) Φ is defined on the set
   $$\mathcal{D}_\Phi \triangleq \left\{(t_b; t_a, \boldsymbol{X}, \boldsymbol{u}) \,:\, t_b, t_a \in T,\, t_b \geqslant t_a,\, \boldsymbol{X} \in \mathbb{C}^{d\times d},\, \boldsymbol{u} \in \Omega\right\}, \tag{2.11}$$
   (b) (Channel property) $\Phi(t_b; t_a, \cdot, \boldsymbol{u}) : \mathbb{C}^{d\times d} \to \mathbb{C}^{d\times d}$ is a quantum channel (i.e., a completely positive and trace-preserving map) for any $t_b \geqslant t_a$ and $\boldsymbol{u}$,
   (c) (Identity) $\Phi(t; t, \boldsymbol{X}, u) = \boldsymbol{X}$ for all $t \in T$, $\boldsymbol{X} \in \mathbb{C}^{d\times d}$, and $\boldsymbol{u} \in \Omega$, and
   (d) (Divisibility) $\Phi(t_c; t_a, \boldsymbol{X}, \boldsymbol{u}) = \Phi(t_c; t_b, \Phi(t_b; t_a, \boldsymbol{X}, \boldsymbol{u}), \boldsymbol{u})$ for all $\boldsymbol{u}$, $\boldsymbol{X}$, and $t_a \leqslant t_b \leqslant t_c$; and
5. the tuple $(\Upsilon, \mathscr{E}, \boldsymbol{M})$ describes the system's measurement apparatus; $(\Upsilon, \mathscr{E})$ is a measurable space and $\boldsymbol{M} : \mathscr{E} \to \mathbb{C}^{d\times d}$ is a resolution of the identity.

The possible *states* of a quantum system Σ are the density matrices $\mathcal{D}(\mathcal{H})$ on $\mathcal{H}$. A control law $\boldsymbol{u} \in \Omega$ is said to *drive* or *transfer* $\boldsymbol{X}_a$ to $\boldsymbol{X}_b$ on an interval $[t_a, t_b]$ if $\boldsymbol{X}_b = \Phi(t_b; t_a, \boldsymbol{X}_a, \boldsymbol{u})$. A system *trajectory* is a function of time $\boldsymbol{\Xi}(t) = \Phi(t; t_a, \boldsymbol{\Xi}_0, \boldsymbol{u})$, $t \geqslant t_a$, which describes how an initial state $\boldsymbol{\Xi}_0 \in \mathcal{D}(\mathcal{H})$ at time $t_a \in T$ evolves under the effect of a control law $\boldsymbol{u} \in \Omega$. The system's state may be measured once along any trajectory.[3] A measurement at time $t_\mathrm{f} \geqslant t_a$ results in a random outcome $\psi(\boldsymbol{\Xi}(t_\mathrm{f})) \in \mathscr{E}$ distributed according to Born's rule:

$$\mathbb{P}\{\psi(\boldsymbol{\Xi}(t_\mathrm{f})) = \mathcal{B}\} = \operatorname{tr}\{\boldsymbol{M}(\mathcal{B})\boldsymbol{\Xi}(t_\mathrm{f})\}. \tag{2.12}$$

When the system Σ's measurement apparatus is described by a finite-cardinality POVM $\mathcal{M}$, we speak in simply in terms of the set $\mathcal{M}$ rather than the tuple $(\Upsilon, \mathscr{E}, \boldsymbol{M})$. In such situations, Σ is simply written $\Sigma = (T, \mathcal{H}, \mathcal{V}, \Omega, \Phi, \mathcal{M})$. □

[3] Once measured, the system's state "collapses" and the subsequent evolution is not necessarily defined. In some scenarios, it is possible to determine the post-measurement state of a quantum system. Such scenarios are not considered here.

Chapters 2 and 3 are primarily concerned with dynamics of a quantum system, i.e., the state transition map. Chapters 4 and 5 study inference problems, and measurement will play a central role.

Remark 2.2 (Fixed measurement apparatuses)**.** In many laboratory experiments and commercial devices, there is only a single *fixed* physical measurement apparatus available. This measurement apparatus might not reveal (with desired accuracy or efficiency) a property of particular quantum state. Chapters 4 and 5 propose a solution to overcome this: the state is allowed to evolve over some period of time under the effect of a properly designed control law, then the resulting state is measured using the fixed apparatus. This has the effect of realizing an effectively different measurement. The fixed measurement scenario is a central pillar of this thesis, and quantum control systems (Definition 2.5) provide a formal mathematical description of such a scenario. □

2.3 Contractivity of quantum control systems

An important feature of quantum control systems is that they are *universally* contracting. The prefix "universally" is used to emphasize that this contraction holds regardless of the input $\boldsymbol{u} \in \Omega$ and the time interval $[t_a, t_b]$ over which the system evolves.[4] This contractivity property follows immediately from Definition 2.5 and the contractivity of quantum channels under the trace norm [78, eq. 3.3].

Lemma 2.1 (Universal contractivity of state transition map)**.** Let Σ be a quantum control system. Then, for any $\boldsymbol{X} \in \mathbb{C}^{d\times d}$, it follows that

$$\|\Phi(t_1; t_0, \boldsymbol{X}, \boldsymbol{u})\|_1 \leqslant \|\boldsymbol{X}\|_1 \,, \qquad \forall \boldsymbol{u} \in \Omega,\, t_1 \geqslant t_0 \tag{2.13}$$

where $\|\cdot\|_1$ is the trace norm. □

An immediate consequence of the fact that quantum systems universally contract is a property that we refer to as "sublinear error propagation". This result is visualized in Figure 2.1. Suppose one wishes to design a control law $\boldsymbol{u} \in \Omega$ that transfers an initial operator $\boldsymbol{X}_0 \in \mathbb{C}^{d\times d}$ to a desired final operator $\boldsymbol{X}_{\text{des}} \in \mathbb{C}^{d\times d}$. While directly designing such a control law can be difficult, it is often possible to break the transfer into simpler "sub-transfers". For instance, suppose one can design a control $\boldsymbol{u} \in \Omega$ that (approximately) drives $\boldsymbol{X}_{i-1} \in \mathbb{C}^{d\times d}$ to $\boldsymbol{X}_i \in \mathbb{C}^{d\times d}$ over a time interval $[t_{i-1}, t_i]$ for each $i \in \mathbb{N}_1^n$, where $n \in \mathbb{N}$ and $\boldsymbol{X}_n \triangleq \boldsymbol{X}_{\text{des}}$. Then, the sublinear error propagation property shows that the control (approximately) transfers $\boldsymbol{X}_0$ to $\boldsymbol{X}_{\text{des}}$ over the time interval $[t_0, t_n]$.

Theorem 2.2 (Sublinear error propagation)**.** Let Σ be a quantum control system, $\boldsymbol{X}_0 \in \mathbb{C}^{d\times d}$, and $n \in \mathbb{N}$. Suppose that there exists a sequence of times $t_1 \leqslant t_2 \leqslant \ldots \leqslant t_n$ and a sequence of operators $\boldsymbol{X}_i, \boldsymbol{E}_i \in \mathbb{C}^{d\times d}$ for $i \in \mathbb{N}_1^n$ such that

$$\Phi(t_i; t_{i-1}, \boldsymbol{X}_{i-1}, \boldsymbol{u}) = \boldsymbol{X}_i + \boldsymbol{E}_i \,. \tag{2.14}$$

[4] Most classical systems are not universally contracting. In fact, practitioners often search for particular control strategies that would make a classical system contractive [77].

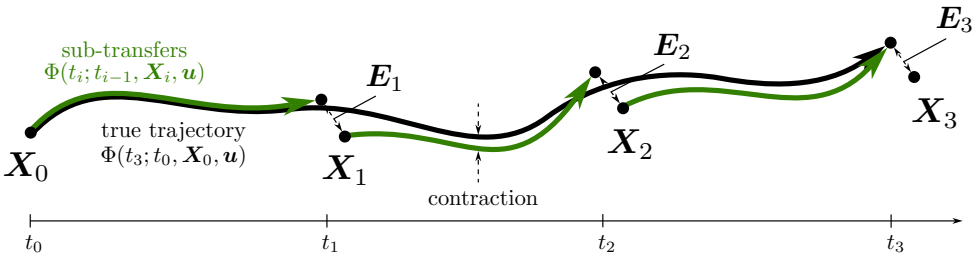


Figure 2.1: Visualization of the sublinear error propagation property (2.15) when $n = 3$. The goal is to transfer an initial operator $\boldsymbol{X}_0$ to a target operator $\boldsymbol{X}_3$ via an appropriate choice of control law $\boldsymbol{u} \in \Omega$. Rather than directly address the global transfer (which may be a convoluted problem), the problem is broken into simpler sub-transfers from $\boldsymbol{X}_{i-1}$ to $\boldsymbol{X}_i$, $i \in \mathbb{N}_1^3$. The control transfers $\boldsymbol{X}_{i-1}$ to $\boldsymbol{X}_i$ over time intervals $[t_{i-1}, t_i]$ with transfer error $\boldsymbol{E}_i$ for $i \in \mathbb{N}_1^3$. Because quantum systems are universally contracting, the true trajectory (the effect of $\boldsymbol{u}$ on $\boldsymbol{X}_0$ over the full time interval $[t_0, t_3]$) approximately transfers $\boldsymbol{X}_0$ to $\boldsymbol{X}_3$. The error between $\Phi(t_3; t_0, \boldsymbol{X}_0, \boldsymbol{u})$ and $\boldsymbol{X}_3$ is bounded by $\sum_{i=1}^3 \|\boldsymbol{E}_i\|_1$.

Then, it follows that

$$\|\Phi(t_n; t_0, \boldsymbol{X}_0, \boldsymbol{u}) - \boldsymbol{X}_n\|_1 \leqslant \sum_{i=1}^{n} \|\boldsymbol{E}_i\|_1 \,. \tag{2.15}$$

□

Proof. The proof proceeds by induction on n.

(Base case): By assumption, $\|\boldsymbol{X}_1 - \Phi(t_1; t_0, \boldsymbol{X}_0, \boldsymbol{u})\|_1 = \|\boldsymbol{E}_1\|_1$. Using the divisibility and linearity properties of Φ, it follows that

$$\begin{aligned}
\|\boldsymbol{X}_2 - \Phi(t_2; t_1, \boldsymbol{X}_0, \boldsymbol{u})\|_1 &= \|\boldsymbol{X}_2 - \Phi(t_2; t_1, \Phi(t_1; t_0, \boldsymbol{X}_0, \boldsymbol{u}), \boldsymbol{u})\|_1 \\
&= \|\boldsymbol{X}_2 - \Phi(t_2; t_1, \boldsymbol{X}_1 + \boldsymbol{E}_1, \boldsymbol{u})\|_1 \\
&\leqslant \|\boldsymbol{X}_2 - \Phi(t_2; t_1, \boldsymbol{X}_1, \boldsymbol{u})\|_1 + \|\Phi(t_2; t_1, \boldsymbol{E}_1, \boldsymbol{u})\|_1 \\
&= \|\boldsymbol{E}_2\|_1 + \|\Phi(t_2; t_1, \boldsymbol{E}_1, \boldsymbol{u})\|_1 \\
&\leqslant \sum_{i=1}^{2} \|\boldsymbol{E}_i\|_1
\end{aligned} \tag{2.16}$$

by the universal contractivity result of quantum systems. So, (2.15) holds for $n = 2$.

(Induction step): Suppose that (2.15) holds for $n \geqslant 2$. Letting $\boldsymbol{F} \triangleq \Phi(t_n; t_0, \boldsymbol{X}_0, \boldsymbol{u}) - \boldsymbol{X}_n$, $\|\boldsymbol{F}\|_1 \leqslant \sum_{i=1}^{n} \|\boldsymbol{E}_i\|_1$. It follows that

$$\begin{aligned}
\|\boldsymbol{X}_{n+1} - \Phi(t_{n+1}; t_0, \boldsymbol{X}_0, u)\|_1 &= \|\boldsymbol{X}_{n+1} - \Phi(t_{n+1}; t_n, \Phi(t_n; t_0, \boldsymbol{X}_0, \boldsymbol{u}), \boldsymbol{u})\|_1 \\
&= \|\boldsymbol{X}_{n+1} - \Phi(t_{n+1}; t_n, \boldsymbol{X}_n + \boldsymbol{F}, \boldsymbol{u})\|_1 \\
&\leqslant \|\boldsymbol{X}_{n+1} - \Phi(t_{n+1}; t_n, \boldsymbol{X}_n, \boldsymbol{u})\|_1 + \|\Phi(t_{n+1}; t_n, \boldsymbol{F}, \boldsymbol{u})\|_1 \\
&\leqslant \|\boldsymbol{E}_{n+1}\|_1 + \|\boldsymbol{F}\|_1 \\
&\leqslant \sum_{i=1}^{n+1} \|\boldsymbol{E}_i\|_1 \,.
\end{aligned} \tag{2.17}$$

⊠

Proving the sublinear error propagation property (2.15) merely relies on (i) the universal contractivity of the underlying system and (ii) the triangle inequality. Hence, the formula would hold if the trace norm is replaced by any pseudometric under which quantum systems are universally contracting. It is natural to ask whether or not the formula holds, e.g., for Schatten norms other than the trace norm. Without imposing any additional structure on the dynamics, the answer is negative: the trace norm is the only Schatten norm that is contracting for general quantum channels [79]. For the special class of *unital* quantum systems, the sublinear error propagation property holds for arbitrary Schatten norms—see Section 2.4.3.

Remark 2.3 (A non-classical result)**.** Since classical systems are not (in general) universally contracting in any metric, the sublinear error propagation property is a uniquely quantum result. □

2.4 Differential quantum control systems

Until this section, quantum control systems have been treated in a general manner. This treatment allowed certain results to be deduced without introducing any unnecessary mathematical structure. However, most practical quantum systems have additional differential structure. Differential quantum control systems are defined below. Then, the differential dynamics of unitary and unital quantum systems are formalized.

2.4.1 Equations of motion

A differential quantum control system is a system whose state transition map Φ satisfies the requirements of Definition 2.5 *and* represents the solution to an appropriate linear[5] differential equation. The most general form of time-independent differential dynamics resulting in a quantum dynamical semigroup (a semigroup that is completely positive and trace preserving) is given by the Lindblad equation [80]:

$$\dot{\boldsymbol{X}}(t) = -\frac{\imath}{\hbar}[\![\boldsymbol{H}, \boldsymbol{X}(t)]\!]_- + \sum_{j\in\mathcal{J}} 2\boldsymbol{L}_j\boldsymbol{X}(t)\boldsymbol{L}_j^\dagger - [\![\boldsymbol{L}_j^\dagger\boldsymbol{L}_j, \boldsymbol{X}(t)]\!]_+ \tag{2.18}$$

where $\boldsymbol{H} = \boldsymbol{H}^\dagger \in \mathbb{C}_{\mathrm{H}}^{d\times d}$; $\boldsymbol{L}_j \in \mathbb{C}^{d\times d}$ for all $j \in \mathcal{J}$ and $|\mathcal{J}| < \infty$; $\hbar$ is the reduced Planck constant. Here, $\boldsymbol{H}$ is known as the system's *Hamiltonian* and the $\boldsymbol{L}_j$'s are known as *Lindblad* operators. When controlling a quantum dynamical system, one often considers time-dependent dynamics of the form

$$\begin{aligned}\dot{\boldsymbol{X}}(t) = &-\frac{\imath}{\hbar}[\![\boldsymbol{H}(\boldsymbol{u}(t),t), \boldsymbol{X}(t)]\!]_- \\ &+\sum_{j\in\mathcal{J}} 2\boldsymbol{L}_j(\boldsymbol{u}(t),t)\boldsymbol{X}(t)\boldsymbol{L}_j^\dagger(\boldsymbol{u}(t),t) - [\![\boldsymbol{L}_j^\dagger(\boldsymbol{u}(t),t)\boldsymbol{L}_j(\boldsymbol{u}(t),t), \boldsymbol{X}(t)]\!]_+ \\ &\triangleq F(\boldsymbol{X}(t), \boldsymbol{u}(t), t)\end{aligned} \tag{2.19}$$

[5]Such a differential equation is necessarily linear in the system's state because the state transition map is required to be a quantum channel. However, these differential equations are generally nonlinear in the control.

where $\boldsymbol{u}(\cdot) \in \Omega$ is the control input, $\boldsymbol{H}(\boldsymbol{u}(t), t) = \boldsymbol{H}^\dagger(\boldsymbol{u}(t), t) \in \mathbb{C}_{\mathrm{H}}^{d\times d}$, and $\boldsymbol{L}_j(\boldsymbol{u}(t), t) \in \mathbb{C}^{d\times d}$. However, one must be careful to ensure that the time dependent dynamics (2.19) lead to valid (completely positive and trace-preserving) state transition maps when the generator F has general control- and time-dependence. This is established in Theorem 2.3.

Definition 2.6 (Differential quantum control system)**.** Let $\mathcal{H} = \mathbb{C}^d$ be a finite-dimensional space. A tuple $\Sigma = (T, \mathcal{H}, \mathcal{V}, \Omega, \Phi, (\Upsilon, \mathscr{E}, \boldsymbol{M}))$ is said to be a *differential quantum control system* if

1. the time set T is a compact interval $[t_0, t_1] \subset \mathbb{R}$ or a ray $[t_0, \infty) \subset \mathbb{R}$, where t_0 is referred to as the initial time;

2. the set of control inputs $\mathcal{V}$ is a Borel subset of $\mathbb{R}^{N_c}$, the set of control laws Ω is a subset of measurable functions from T into $\mathcal{V}$, and the set Ω satisfies the concatenation property of Definition 2.5;

3. for any $\boldsymbol{X}_0 \in \mathbb{C}^{d\times d}$, $\boldsymbol{u} \in \Omega$, and $[t_a, t_b] \subseteq T$, the state transition map $\Phi(t_b; t_a, \boldsymbol{X}_0, \boldsymbol{u})$ is the unique Carathéodory solution[6] $\boldsymbol{X}(t)$ to the Lindblad equation (2.19) starting from the initial condition $\boldsymbol{X}(t_a) = \boldsymbol{X}_0$;

4. the Hamiltonian $\boldsymbol{H}(\cdot, \cdot)$ and Lindblad operators $\boldsymbol{L}_j(\cdot, \cdot)$ from (2.19) are Carathéodory[7] with the property that, for any control law $\boldsymbol{u} \in \Omega$, the functions $\|\boldsymbol{H}(\boldsymbol{u}(t), t)\|_{\mathrm{HS}}$ and $\|\boldsymbol{L}_j(\boldsymbol{u}(t), t)\|_{\mathrm{HS}}^2$ are locally integrable;[8] and

5. the tuple $(\Upsilon, \mathscr{E}, \boldsymbol{M})$ consists of a measurable space $(\Upsilon, \mathscr{E})$ and a resolution of the identity $\boldsymbol{M} : \mathscr{E} \to \mathbb{C}^{d\times d}$. □

The following theorem shows that a differential quantum control system is indeed a quantum control system in the sense of Definition 2.5. In particular, the theorem shows that the state transition map arising from the equation (2.19) is completely positive and trace-preserving even if the Hamiltonian and Lindblad operators have very general control- and time-dependence.

Theorem 2.3 (Well-posedness of differential quantum control systems)**.** Let Σ be a differential quantum control system, in the sense of Definition 2.6. Then,

1. for any $\boldsymbol{u} \in \Omega$, $[t_a, t_b] \subseteq T$, and $\boldsymbol{X}_0 \in \mathbb{C}^{d\times d}$, there exists a unique Carathéodory solution $\boldsymbol{X}(t)$ to (2.19) on the interval $[t_a, t_b]$ such that $\boldsymbol{X}(t_a) = \boldsymbol{X}_0$; and

2. Σ is a quantum control system in the sense of Definition 2.5.

In particular, the system Σ's state transition map $\Phi(t_b; t_a, \cdot, \boldsymbol{u})$ is completely positive and trace-preserving. □

[6] There indeed exists a unique solution. See Theorem 2.3.

[7] See Definition B.1 in the appendix.

[8] A function $f : T \to \mathbb{R}$ is said to be locally integrable if, for any compact $K \subseteq T$, it follows that $\int_K |f(t)|\,\mathrm{d}t < \infty$. Integrability of f implies local integrability.

Proof. The proof of existence and uniqueness is deferred to the appendix (Theorem B.1).

Let $\boldsymbol{u} \in \Omega$. To show that Σ is a quantum control system in the sense of Definition 2.5, it must be verified that the state transition map $\Phi(t_b; t_a, \cdot, \boldsymbol{u})$ associated with the dynamics (2.19) satisfies the channel, identity, and divisibility properties. The last two are trivial. Let us now show that $\Phi(t_b; t_a, \cdot, \boldsymbol{u})$ is completely positive and trace-preserving, i.e., $\Phi(t_b; t_a, \cdot, \boldsymbol{u})$ is a quantum channel. For any $\boldsymbol{X}, \boldsymbol{H}, \boldsymbol{L} \in \mathbb{C}^{d\times d}$, the cyclic property of the trace proves that

$$\operatorname{tr}\{[\![\boldsymbol{H}, \boldsymbol{X}]\!]_-\} = \operatorname{tr}\{\boldsymbol{H}\boldsymbol{X} - \boldsymbol{H}\boldsymbol{X}\} = 0 \tag{2.20a}$$

$$\operatorname{tr}\{2\boldsymbol{L}\boldsymbol{X}\boldsymbol{L}^\dagger - [\![\boldsymbol{L}^\dagger\boldsymbol{L}, \boldsymbol{X}]\!]_+\} = \operatorname{tr}\{2\boldsymbol{L}^\dagger\boldsymbol{L}\boldsymbol{X} - 2\boldsymbol{L}^\dagger\boldsymbol{L}\boldsymbol{X}\} = 0\,. \tag{2.20b}$$

Therefore, the state transition map Φ is trace preserving: any solution $\boldsymbol{X}(t) = \Phi(t; t_a, \boldsymbol{X}_0, \boldsymbol{u})$ to (2.19) satisfies

$$\operatorname{tr}\{\boldsymbol{X}(t)\} = \operatorname{tr}\{\boldsymbol{X}_0\} + \int_{t_a}^{t} \underbrace{\operatorname{tr}\{F(\boldsymbol{X}(s), \boldsymbol{u}(s), s)\}}_{=0}\, \mathrm{d}s = \operatorname{tr}\{\boldsymbol{X}_0\}\,. \tag{2.21}$$

Next, it will be shown that the state transition map $\Phi(t_b; t_a, \cdot, \boldsymbol{u})$ is a completely positive map. The idea of the proof is to approximate the differential equation (2.19) by a sequence of equations corresponding to completely positive state transition maps. Complete positivity of $\Phi(t_b; t_a, \cdot, \boldsymbol{u})$ is then obtained by a limiting procedure. For notational simplicity, it will be assumed that $[t_a, t_b] = [0, 1]$. Define the operator $A(t) \in \mathcal{L}(\mathbb{C}^{d\times d})$ as

$$A(t)\boldsymbol{X} \triangleq F(\boldsymbol{X}, \boldsymbol{u}(t), t)\,, \qquad \forall (t, \boldsymbol{X}) \in [0, 1] \times \mathbb{C}^{d\times d}\,. \tag{2.22}$$

Let $\{|\phi_n\rangle\}$, $\{|\psi_m\rangle\} \subset \mathbb{C}^d$ be any finite sets of vectors and define $\boldsymbol{X}_{nm}(t) \triangleq \Phi(t; 0, |\psi_n\rangle\langle\psi_m|, \boldsymbol{u})$. Establishing complete positivity of $\Phi(t; 0, \cdot, \boldsymbol{u})$ is equivalent to showing that, for all $t \in [0, 1]$,

$$\sum_n \sum_m \langle\phi_n|\, \Phi(t; 0, |\psi_n\rangle\langle\psi_m|, \boldsymbol{u})\, |\phi_m\rangle = \sum_n \sum_m \langle\phi_n|\, \boldsymbol{X}_{nm}(t)\, |\phi_m\rangle \geqslant 0\,. \tag{2.23}$$

Step (i): Suppose that the matrices $\boldsymbol{H}(\boldsymbol{u}(t), t)$ and $\boldsymbol{L}_j(\boldsymbol{u}(t), t)$, $j \in \mathcal{J}$ are continuous functions of time (this will be relaxed in Step (ii)). Using the submultiplicative property of the trace norm, it can be shown that

$$\|A(t)\boldsymbol{X}\|_{\mathrm{HS}} \leqslant \Big[\tfrac{2}{\hbar}\|\boldsymbol{H}(\boldsymbol{u}(t), t)\|_{\mathrm{HS}} + \sum_{j\in\mathcal{J}} 4\|\boldsymbol{L}_j(\boldsymbol{u}(t), t)\|_{\mathrm{HS}}^2\Big]\|\boldsymbol{X}\|_{\mathrm{HS}}\,, \qquad \forall \boldsymbol{X} \in \mathbb{C}^{d\times d}\,. \tag{2.24}$$

Let $\|\cdot\|_{\mathrm{op}}$ be the operator norm on $\mathcal{L}(\mathbb{C}^{d\times d})$ induced by the Hilbert-Schmidt norm, i.e.,

$$\|B\|_{\mathrm{op}} \triangleq \sup\left\{\frac{\|B(\boldsymbol{X})\|_{\mathrm{HS}}}{\|\boldsymbol{X}\|_{\mathrm{HS}}} \;:\; \boldsymbol{X} \in \mathbb{C}^{d\times d},\, \boldsymbol{X} \neq \boldsymbol{0}\right\}, \qquad B \in \mathcal{L}(\mathbb{C}^{d\times d})\,. \tag{2.25}$$

By (2.24), the operator norm of $A(t)$ is uniformly bounded by a finite constant on the compact interval $[0, 1]$:

$$\|A(t)\|_{\mathrm{op}} \leqslant \max_{t\in[0,1]} \Big[\tfrac{2}{\hbar}\|\boldsymbol{H}(\boldsymbol{u}(t), t)\|_{\mathrm{HS}} + \sum_{j\in\mathcal{J}} 4\|\boldsymbol{L}_j(\boldsymbol{u}(t), t)\|_{\mathrm{HS}}^2\Big] \triangleq C < \infty\,. \tag{2.26}$$

The maximum in (2.26) is well-defined because $\boldsymbol{H}(\boldsymbol{u}(t),t)$ and $\boldsymbol{L}_j(\boldsymbol{u}(t),t)$ were assumed to be continuous at the start of Step (i).

Define the sequence of step functions

$$A^{(k)}(t) \triangleq \sum_{i=1}^{k} \mathbb{1}_{I_{i,k}}(t)\, A(\tfrac{i-1}{k})\,, \qquad k \in \mathbb{N} \tag{2.27}$$

where $I_{i,k} \triangleq [\frac{i-1}{k}, \frac{i}{k})$. By continuity, $\|A^{(k)}(t) - A(t)\|_{\mathrm{op}} \to 0$ pointwise on $[0,1]$ as $k \to \infty$. For a proof of this fact, see Appendix B.3. For each $k \in \mathbb{N}$, let $\boldsymbol{X}^{(k)}_{nm}(\cdot) : [0,1] \to \mathbb{C}^{d\times d}$ be the unique solution to the differential equation

$$\boldsymbol{X}^{(k)}_{nm}(t) = |\psi_n\rangle\langle\psi_m| + \int_0^t A^{(k)}(s)\boldsymbol{X}^{(k)}_{nm}(s)\,\mathrm{d}s\,. \tag{2.28}$$

By construction, these solutions are obtained by stitching together solutions to time-independent Lindblad equations. The solution maps to time-independent Lindblad equations are quantum channels, and the composition of quantum channels is another quantum channel. Therefore, each $A^{(k)}$ generates a completely positive and trace-preserving state transition map. The solutions of (2.28) are hence contracting (Lemma 2.1):

$$\|\boldsymbol{X}^{(k)}_{nm}(t)\|_{\mathrm{HS}} \leqslant \|\boldsymbol{X}^{(k)}_{nm}(t)\|_1 \leqslant \||\psi_n\rangle\langle\psi_m|\|_1 \leqslant \max_{n,m} \||\psi_n\rangle\langle\psi_m|\|_1 \triangleq c\,. \tag{2.29}$$

for all $(k,t) \in \mathbb{N}\times[0,1]$. Without loss of generality, suppose $c = 1$.

Now, it will be shown that $\boldsymbol{X}^{(k)}_{nm}(t) \to \boldsymbol{X}_{nm}(t)$ uniformly in t, n, and m. Invoking contractivity,

$$\begin{aligned}
\|\boldsymbol{X}_{nm}(t) - \boldsymbol{X}^{(k)}_{nm}(t)\|_{\mathrm{HS}} &\leqslant \int_0^t \|A(s)\boldsymbol{X}_{nm}(s) - A^{(k)}(s)\boldsymbol{X}^{(k)}_{nm}(s)\|_{\mathrm{HS}}\,\mathrm{d}s \\
&\leqslant \int_0^t \|A(s)\|_{\mathrm{op}}\,\|\boldsymbol{X}_{nm}(s) - \boldsymbol{X}^{(k)}_{nm}(s)\|_{\mathrm{HS}}\,\mathrm{d}s \\
&\qquad + \int_0^t \|A(s) - A^{(k)}(s)\|_{\mathrm{op}}\,\|\boldsymbol{X}^{(k)}_{nm}(s)\|_{\mathrm{HS}}\,\mathrm{d}s \\
&\leqslant \int_0^t \|A(s)\|_{\mathrm{op}}\,\|\boldsymbol{X}_{nm}(s) - \boldsymbol{X}^{(k)}_{nm}(s)\|_{\mathrm{HS}}\,\mathrm{d}s \\
&\qquad + \int_0^1 \|A(s) - A^{(k)}(s)\|_{\mathrm{op}}\,\mathrm{d}s\,.
\end{aligned} \tag{2.30}$$

Grönwall's inequality (Lemma A.2) then provides the bound

$$\max_{n,m}\max_{t\in[0,1]} \|\boldsymbol{X}_{nm}(t) - \boldsymbol{X}^{(k)}_{nm}(t)\|_{\mathrm{HS}} \leqslant \exp\Big\{\int_0^1 \|A(s)\|_{\mathrm{op}}\,\mathrm{d}s\Big\}\int_0^1 \|A(s) - A^{(k)}(s)\|_{\mathrm{op}}\,\mathrm{d}s\,. \tag{2.31}$$

Note that $\|A(s) - A^{(k)}(s)\|_{\mathrm{op}} \leqslant \|A(s)\|_{\mathrm{op}} + \|A^{(k)}(s)\|_{\mathrm{op}} \leqslant 2C$. Given this bound and the fact that $\|A^{(k)}(t) - A(t)\|_{\mathrm{op}} \to 0$ pointwise on $[0,1]$ as $k \to \infty$, the bounded convergence theorem (Theorem A.1) implies that

$$\lim_{k\to\infty} \int_0^1 \|A(s) - A^{(k)}(s)\|_{\mathrm{op}}\,\mathrm{d}s = 0\,. \tag{2.32}$$

By (2.31), uniform convergence is proved:

$$\lim_{k\to\infty} \max_{n,m} \max_{t\in[0,1]} \|\boldsymbol{X}_{nm}(t) - \boldsymbol{X}_{nm}^{(k)}(t)\|_{\mathrm{HS}} \leqslant \lim_{k\to\infty} e^{C} \int_0^1 \|A(s) - A^{(k)}(s)\|_{\mathrm{op}}\, \mathrm{d}s = 0\,. \tag{2.33}$$

Since each $A^{(k)}$ generates a completely positive state transition map, it follows that

$$\sum_n \sum_m \langle \phi_n |\, \boldsymbol{X}_{nm}(t)\, | \phi_m \rangle = \lim_{k\to\infty} \sum_n \sum_m \langle \phi_n |\, \boldsymbol{X}_{nm}^{(k)}(t)\, | \phi_m \rangle \geqslant 0\,. \tag{2.34}$$

Hence, the state transition map $\Phi(t;0,\cdot,\boldsymbol{u})$ is completely positive positive when $\boldsymbol{H}(\boldsymbol{u}(t),t)$ and $\boldsymbol{L}_j(\boldsymbol{u}(t),t)$ are continuous functions of time.

Step (ii): Now, consider the general case where $\boldsymbol{H}(\boldsymbol{u}(t),t)$ and $\boldsymbol{L}_j(\boldsymbol{u}(t),t)$ are measurable functions of time. Recall from Definition 2.6 that $\|\boldsymbol{H}(\boldsymbol{u}(t),t)\|_{\mathrm{HS}}$ and $\|\boldsymbol{L}(\boldsymbol{u}(t),t)\|_{\mathrm{HS}}^2$ are locally integrable. Therefore, the measurable function $m : [0,1] \to \mathbb{R}$ defined by

$$m(t) \triangleq \tfrac{2}{\hbar}\|\boldsymbol{H}(\boldsymbol{u}(t),t)\|_{\mathrm{HS}} + \sum_{j\in\mathcal{J}} 4\|\boldsymbol{L}_j(\boldsymbol{u}(t),t)\|_{\mathrm{HS}}^2\,, \qquad \forall t \in [0,1] \tag{2.35}$$

is integrable. Let $h_k > 0$ be a sequence of real numbers converging to zero as $k \to \infty$. Define the averaged operators

$$\boldsymbol{H}^{(k)}(t) \triangleq \frac{1}{2h_k} \int_{t-h_k}^{t+h_k} \boldsymbol{H}(\boldsymbol{u}(s),s)\, \mathbb{1}_{[0,1]}(s)\, \mathrm{d}s \tag{2.36a}$$

$$\boldsymbol{L}_j^{(k)}(t) \triangleq \frac{1}{2h_k} \int_{t-h_k}^{t+h_k} \boldsymbol{L}_j(\boldsymbol{u}(s),s)\, \mathbb{1}_{[0,1]}(s)\, \mathrm{d}s \tag{2.36b}$$

for $t \in [0,1]$, $j \in \mathcal{J}$, and $k \in \mathbb{N}$. By construction, these operators are continuous functions of time. Define the associated generators $A^{(k)}(t) \in \mathcal{L}(\mathbb{C}^{d\times d})$, $k \in \mathbb{N}$, according to

$$A^{(k)}(t)\boldsymbol{X} = -\frac{\imath}{\hbar}[\![\boldsymbol{H}^{(k)}(t), \boldsymbol{X}]\!]_- + \sum_{j\in\mathcal{J}} 2\boldsymbol{L}_j^{(k)}(t)\boldsymbol{X}\boldsymbol{L}_j^{(k)\dagger}(t) - [\![\boldsymbol{L}_j^{(k)\dagger}(t)\boldsymbol{L}_j^{(k)}(t), \boldsymbol{X}]\!]_+ \tag{2.37}$$

for $\boldsymbol{X} \in \mathbb{C}^{d\times d}$. For each n and m, let $\boldsymbol{X}_{nm}^{(k)}(\cdot) : [0,1] \to \mathbb{C}^{d\times d}$ be the unique solution to the differential equation (2.28) with $A^{(k)}(\cdot)$ from (2.37). Since Step (i) proves that continuous Hamiltonians and Lindblad operators result in completely positive and trace-preserving evolution, Lemma 2.1 can be invoked to conclude that, as in (2.29), $\|\boldsymbol{X}_{nm}^{(k)}(t)\|_{\mathrm{HS}} \leqslant 1$ for all $(k,t) \in \mathbb{N}\times[0,1]$. Using the same argumentation as Step (i), Grönwall's inequality implies that

$$\max_{n,m} \max_{t\in[0,1]} \|\boldsymbol{X}_{nm}(t) - \boldsymbol{X}_{nm}^{(k)}(t)\|_{\mathrm{HS}} \leqslant e^{\|m\|_{L^1}} \int_0^1 \|A(s) - A^{(k)}(s)\|_{\mathrm{op}}\, \mathrm{d}s\,. \tag{2.38}$$

It can be verified that $\int_0^1 \|A(s) - A^{(k)}(s)\|_{\mathrm{op}}\, \mathrm{d}s \to 0$ as $k \to \infty$ (see Appendix B.3), therefore

$$\lim_{k\to\infty} \max_{n,m} \max_{t\in[0,1]} \|\boldsymbol{X}_{nm}(t) - \boldsymbol{X}_{nm}^{(k)}(t)\|_{\mathrm{HS}} = 0\,. \tag{2.39}$$

Using the same limiting argumentation as in case (i), complete positivity of $\Phi(t;0,\cdot,\boldsymbol{u})$ is proved. ⊠

There are two special cases of the dynamics (2.19) that deserve special attention.

1. Most often in practice, the generator F decomposes as

$$F(\boldsymbol{X}, \boldsymbol{u}, t) = F_0(\boldsymbol{X}, t) + F_{\mathrm{c}}(\boldsymbol{X}, \boldsymbol{u}, t) \tag{2.40}$$

where F_0 is *linear* in the state variable for any fixed t and F_{c} *bilinear* in the state and input variables for any fixed t. That is,

$$F_{\mathrm{c}}(\alpha\boldsymbol{X} + \beta\boldsymbol{Y}, \boldsymbol{u}, t) = \alpha F_{\mathrm{c}}(\boldsymbol{X}, \boldsymbol{u}, t) + \beta F_{\mathrm{c}}(\boldsymbol{Y}, \boldsymbol{u}, t) \tag{2.41a}$$

$$F_{\mathrm{c}}(\boldsymbol{X}, \alpha\boldsymbol{u} + \beta\boldsymbol{v}, t) = \alpha F_{\mathrm{c}}(\boldsymbol{X}, \boldsymbol{u}, t) + \beta F_{\mathrm{c}}(\boldsymbol{X}, \boldsymbol{v}, t) \tag{2.41b}$$

for $\alpha, \beta \in \mathbb{C}$. For instance, (2.40) holds when the Lindblad operators $\boldsymbol{L}_j(\boldsymbol{u}(t), t)$ do not depend on the control input $\boldsymbol{u}(t)$ and the Hamiltonian takes the form

$$\boldsymbol{H}(\boldsymbol{u}(t), t) = \boldsymbol{H}_0(t) + \sum_{\ell=1}^{N_{\mathrm{c}}} u_\ell(t)\, \boldsymbol{H}_\ell(t) \tag{2.42}$$

for $\boldsymbol{u}(t) \in \mathbb{R}^{N_{\mathrm{c}}}$, $\boldsymbol{H}_\ell(t) \in \mathbb{C}^{d\times d}$, $\ell \in \mathbb{N}_0^{N_{\mathrm{c}}}$, and $N_{\mathrm{c}} \in \mathbb{N}$. The qubit system in Example 2.2 provides an example; see also [26].

2. The Lindblad operators $\boldsymbol{L}_j(\boldsymbol{u}(t), t)$ often take the form $\boldsymbol{L}_j(\boldsymbol{u}(t), t) = \sqrt{\gamma_j(\boldsymbol{u}(t), t)}\, \bar{\boldsymbol{L}}_j$ where $\gamma_j(\boldsymbol{u}(t), t) \in \mathbb{R}_{\geqslant 0}$ is known as a *dissipation rate*. The dynamics are then of the *rate* form

$$\dot{\boldsymbol{X}}(t) = -\frac{\imath}{\hbar}[\![\boldsymbol{H}(\boldsymbol{u}(t), t), \boldsymbol{X}(t)]\!]_- + \sum_{j\in\mathcal{J}} \gamma_j(\boldsymbol{u}(t), t)\Big[2\bar{\boldsymbol{L}}_j \boldsymbol{X}(t) \bar{\boldsymbol{L}}_j^\dagger - [\![\bar{\boldsymbol{L}}_j^\dagger \bar{\boldsymbol{L}}_j, \boldsymbol{X}(t)]\!]_+\Big]\,. \tag{2.43}$$

The dephasing qubit system in Example 2.1 provides an example.

These special cases provided extra structure which, when exploited appropriately, enables richer mathematical results.

2.4.2 Unitary quantum systems

Unitary systems were among the first quantum systems studied around the birth of quantum mechanics. Differential systems with dynamics given by the well-known von Neumann equation are unitary systems.

Definition 2.7 (Unitary channel)**.** A quantum channel $K(\cdot) \in \mathcal{C}(\mathbb{C}^d)$ is said to be *unitary* if there exists a unitary operator $\boldsymbol{U} \in U(d)$ such that

$$K(\boldsymbol{X}) = \boldsymbol{U}\boldsymbol{X}\boldsymbol{U}^\dagger\,, \quad \forall \boldsymbol{X} \in \mathbb{C}^{d\times d}\,. \tag{2.44}$$

□

Definition 2.8 (Unitary system)**.** A quantum control system Σ on $\mathcal{H} = \mathbb{C}^d$ is said to be *unitary* if, for every $\boldsymbol{u} \in \Omega$ and $t_{\mathrm{f}} \geqslant t_0$, the quantum channel $\Phi(t_{\mathrm{f}}; t_0, \cdot, \boldsymbol{u}) \in \mathcal{C}(\mathbb{C}^d)$ is unitary. □

The key feature of a unitary system is that the spectrum of an initial state $\boldsymbol{\Xi} \in \mathcal{D}(\mathbb{C}^d)$ is invariant along the trajectories of the system, i.e., the eigenvalues of $\Phi(t_{\mathrm{f}}; t_0, \boldsymbol{\Xi}, \boldsymbol{u})$ are the same as $\boldsymbol{\Xi}$. This implies that the von Neumann entropy (Definition 2.1) is constant along system trajectories, i.e., $S(\Phi(t_{\mathrm{f}}; t_0, \boldsymbol{\Xi}, \boldsymbol{u})) = S(\boldsymbol{\Xi})$.

If the Lindblad operators $\boldsymbol{L}_j(\boldsymbol{u}(t), t)$ in (2.19) are zero for all t and $\boldsymbol{u}$, then one obtains the *von Neumann equation*

$$\dot{\boldsymbol{X}}(t) = -\frac{\imath}{\hbar}[\![\boldsymbol{H}(\boldsymbol{u}(t), t), \boldsymbol{X}(t)]\!]_- \,. \tag{2.45}$$

The solution to (2.45) starting from initial condition $\boldsymbol{X}(t_0) \in \mathbb{C}^{d\times d}$ is given by

$$\boldsymbol{X}(t) = \boldsymbol{U}(t)\boldsymbol{X}(t_0)\boldsymbol{U}^\dagger(t) \tag{2.46}$$

where $\boldsymbol{U}(t)$ is the solution to $\dot{\boldsymbol{U}}(t) = -\frac{\imath}{\hbar}\boldsymbol{H}(\boldsymbol{u}(t), t)\boldsymbol{U}(t)$ with initial condition $\boldsymbol{U}(t_0) = \boldsymbol{I}$. Since $-\frac{\imath}{\hbar}\boldsymbol{H}(\boldsymbol{u}(t), t)$ is in the tangent space of the unitary group $U(d)$, it follows that $\boldsymbol{U}(t)$ is a unitary operator. Therefore, a differential system with dynamics of the form (2.45) is unitary in the sense of Definition 2.8.

Remark 2.4. The theory of unitary quantum systems is deep and well-established—see [26] and the references therein. □

2.4.3 Unital quantum systems

The class of unital systems is a strict superset of the class of unitary systems and is a proper subset of the class of general quantum systems. This section formalizes (abstract and differential) unital quantum control systems.

Definition 2.9 (Unital and mixed unitary quantum channels [81])**.** Let $K(\cdot) \in \mathcal{C}(\mathbb{C}^d)$ be a quantum channel.

1. The channel $K(\cdot)$ is said to be *unital* if it preserves the identity, i.e., $K(\boldsymbol{I}_d) = \boldsymbol{I}_d$.
2. The channel $K(\cdot)$ is said to be *mixed unitary* if there exists a probability mass function $p_j \geqslant 0$ (on a finite alphabet [82]) and unitary operators $\boldsymbol{U}_j \in U(d)$ such that

$$K(\boldsymbol{X}) = \sum_j p_j \boldsymbol{U}_j \boldsymbol{X} \boldsymbol{U}_j^\dagger \,, \quad \forall \boldsymbol{X} \in \mathbb{C}^{d\times d} \,. \tag{2.47}$$

□

Clearly, every mixed unitary channel is a unital channel; the converse is not true.

Definition 2.10 (Unital quantum control system)**.** A quantum control system Σ on $\mathcal{H} = \mathbb{C}^d$ is said to be *unital* if, for every $\boldsymbol{u} \in \Omega$ and $t_{\mathrm{f}} \geqslant t_0$, the quantum channel $\Phi(t_{\mathrm{f}}; t_0, \cdot, \boldsymbol{u}) \in \mathcal{C}(\mathbb{C}^d)$ is unital. □

Recall the sublinear error propagation property established in Theorem 2.2 for general quantum control systems. In unital quantum systems, this property holds even if the trace norm is replaced by an arbitrary Schatten p-norm.

Theorem 2.4 (Sublinear error propagation for unital systems)**.** Let Σ be a unital quantum control system. Let $p \in [1, \infty]$, $\boldsymbol{X}_0 \in \mathbb{C}^{d\times d}$, and $n \in \mathbb{N}$. Suppose that there exists a sequence of times $t_1 \leqslant t_2 \leqslant \ldots \leqslant t_n$ and a sequence of operators $\boldsymbol{X}_i, \boldsymbol{E}_i \in \mathbb{C}^{d\times d}$ for $i \in \mathbb{N}_1^n$ such that

$$\Phi(t_i; t_{i-1}, \boldsymbol{X}_{i-1}, \boldsymbol{u}) = \boldsymbol{X}_i + \boldsymbol{E}_i \,. \tag{2.48}$$

Then, it follows that

$$\|\Phi(t_n; t_0, \boldsymbol{X}_0, \boldsymbol{u}) - \boldsymbol{X}_n\|_p \leqslant \sum_{i=1}^{n} \|\boldsymbol{E}_i\|_p \tag{2.49}$$

where $\|\cdot\|_p$ is the Schatten p-norm. □

Proof. By [79, Thm. 2.4], unital quantum channels are contractive under any Schatten p-norm. The proof of the theorem then follows from the definition of a unital quantum control system and the proof of Theorem 2.2. ⊠

Theorem 2.5 (Unital differential systens, Clouâtré et al. [1])**.** Let Σ be a differential quantum control system on $\mathcal{H} = \mathbb{C}^d$ whose dynamics are given by (2.19). The system Σ is unital if and only if, for each $\boldsymbol{u} \in \Omega$,

$$\sum_{j\in\mathcal{J}} \boldsymbol{L}_j(\boldsymbol{u}(t), t)\boldsymbol{L}_j^\dagger(\boldsymbol{u}(t), t) = \sum_{j\in\mathcal{J}} \boldsymbol{L}_j^\dagger(\boldsymbol{u}(t), t)\boldsymbol{L}_j(\boldsymbol{u}(t), t) \tag{2.50}$$

for almost all $t \in T$ (with respect to Lebesgue measure). □

Proof. ($\Longrightarrow$) Suppose that the system Σ is unital. It must be shown that (2.50) holds for almost all $t \in T$. We will provide a proof for the case $T = [t_0, t_1]$; the case when $T = [t_0, \infty)$ follows by countable additivity of Lebesgue measure. The differential equation (2.19) can be written equivalently as

$$\begin{aligned}\boldsymbol{X}(t) = \boldsymbol{X}(t_0) + \int_{t_0}^{t} \Big[&-\frac{\imath}{\hbar}[\![\boldsymbol{H}(\boldsymbol{u}(s), s), \boldsymbol{X}(s)]\!]_- \\ &+ \sum_{j\in\mathcal{J}} 2\boldsymbol{L}_j(\boldsymbol{u}(s), s)\boldsymbol{X}(s)\boldsymbol{L}_j^\dagger(\boldsymbol{u}(s), s) - [\![\boldsymbol{L}_j^\dagger(\boldsymbol{u}(s), s)\boldsymbol{L}_j(\boldsymbol{u}(s), s), \boldsymbol{X}(s)]\!]_+ \Big] \mathrm{d}s\end{aligned} \tag{2.51}$$

for $t \in T$. Since Σ is unital, $\boldsymbol{I}/d$ is a steady state of the dynamics. The integral solution starting from the initial condition $\boldsymbol{X}(t_0) = \boldsymbol{I}/d$ yields

$$\boldsymbol{0} = \int_{t_0}^{t} \sum_{j\in\mathcal{J}} \Big[\boldsymbol{L}_j(\boldsymbol{u}(s), s)\boldsymbol{L}_j^\dagger(\boldsymbol{u}(s), s) - \boldsymbol{L}_j^\dagger(\boldsymbol{u}(s), s)\boldsymbol{L}_j(\boldsymbol{u}(s), s)\Big] \,\mathrm{d}s \tag{2.52}$$

for all $t \in T$. By the Lebesgue differentiation theorem (Theorem A.3),

$$\begin{aligned}\boldsymbol{0} &= \lim_{h\searrow 0} \frac{1}{h} \int_{t_0}^{t+h} \sum_{j\in\mathcal{J}} \Big[\boldsymbol{L}_j(\boldsymbol{u}(s), s)\boldsymbol{L}_j^\dagger(\boldsymbol{u}(s), s) - \boldsymbol{L}_j^\dagger(\boldsymbol{u}(s), s)\boldsymbol{L}_j(\boldsymbol{u}(s), s)\Big] \,\mathrm{d}s \\ &= \lim_{h\searrow 0} \frac{1}{h} \Big[\int_{t_0}^{t} \sum_{j\in\mathcal{J}} \Big[\boldsymbol{L}_j(\boldsymbol{u}(s), s)\boldsymbol{L}_j^\dagger(\boldsymbol{u}(s), s) - \boldsymbol{L}_j^\dagger(\boldsymbol{u}(s), s)\boldsymbol{L}_j(\boldsymbol{u}(s), s)\Big] \,\mathrm{d}s \\ &\qquad + \int_{t}^{t+h} \sum_{j\in\mathcal{J}} \Big[\boldsymbol{L}_j(\boldsymbol{u}(s), s)\boldsymbol{L}_j^\dagger(\boldsymbol{u}(s), s) - \boldsymbol{L}_j^\dagger(\boldsymbol{u}(s), s)\boldsymbol{L}_j(\boldsymbol{u}(s), s)\Big] \,\mathrm{d}s \Big] \\ &= \sum_{j\in\mathcal{J}} \boldsymbol{L}_j(\boldsymbol{u}(t), t)\boldsymbol{L}_j^\dagger(\boldsymbol{u}(t), t) - \boldsymbol{L}_j^\dagger(\boldsymbol{u}(t), t)\boldsymbol{L}_j(\boldsymbol{u}(t), t)\end{aligned} \tag{2.53}$$

for almost all $t \in (t_0, t_1)$, and therefore (2.50) holds for almost all $t \in T$.

($\Longleftarrow$) Suppose that (2.50) holds for almost all $t \in T$. If $\boldsymbol{X}(t_0) = \boldsymbol{I}/d$, then $\boldsymbol{X}(t) = \boldsymbol{I}/d$ is a solution to (2.51) (equivalently, a solution to (2.19)). By Theorem 2.3, this solution is unique. Therefore, the system Σ is unital. ⊠

Theorem 2.5 motivates the following definition.

Definition 2.11 (Differential unital quantum control system)**.** A quantum control system Σ is said to be *differential unital quantum control system* if

1. it is a differential system in the sense of Definition 2.6, and
2. (2.50) holds for almost all $t \in T$. □

Example 2.1 (Dephasing qubit)**.** A qubit lives in $\mathcal{H} = \mathbb{C}^2$, which is spanned by the standard basis vectors $|0\rangle$ and $|1\rangle$. The dynamics of a dephasing qubit are of the form

$$\begin{aligned}\dot{\boldsymbol{X}}(t) &= -\frac{\imath}{\hbar}[\![\boldsymbol{H}(\boldsymbol{u}(t), t), \boldsymbol{X}(t)]\!]_- - \gamma(t)\big[2\boldsymbol{P}_z\boldsymbol{X}(t)\boldsymbol{P}_z^\dagger - [\![\boldsymbol{P}_z^\dagger\boldsymbol{P}_z, \boldsymbol{X}(t)]\!]_+\big] \\ &= -\frac{\imath}{\hbar}[\![\boldsymbol{H}(\boldsymbol{u}(t), t), \boldsymbol{X}(t)]\!]_- - 2\gamma(t)\big[\boldsymbol{P}_z\boldsymbol{X}(t)\boldsymbol{P}_z - \boldsymbol{X}(t)\big] \qquad (2.54)\end{aligned}$$

where $\gamma(t) \geqslant 0$ is a time-dependent dephasing rate and $\boldsymbol{P}_z \triangleq |0\rangle\langle 0| - |1\rangle\langle 1|$ is the Pauli-z operator. The dynamics (2.54) are of the special Lindblad form (2.43), and can be identified with the general Lindblad form (2.19) using $\boldsymbol{L}(t) = \sqrt{\gamma(t)}\boldsymbol{P}_z$. This Lindblad operator is Hermitian, so (2.50) is satisfied and the system is unital. □

Unital quantum channels are known to be closely related to operator majorization [81, ch. 4]. There have been efforts to study when differential quantum control systems are majorization monotone. We give a quick proof that directly lifts the results from the study of quantum channels.[9] The following definitions are well-known, e.g., see [83].

Definition 2.12 (Vector majorization)**.** Let $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$ and let $x_i^\downarrow$ denote the ith largest coordinate of $\boldsymbol{x}$. The vector $\boldsymbol{x}$ *majorizes* $\boldsymbol{y}$, denoted $\boldsymbol{y} \prec_{\mathrm{m}} \boldsymbol{x}$, if the following conditions hold:

1. $\sum_{i=1}^k y_i^\downarrow \leqslant \sum_{i=1}^k x_i^\downarrow$ for all $k \in \mathbb{N}_1^{n-1}$, and
2. $\sum_{i=1}^n y_i^\downarrow = \sum_{i=1}^n x_i^\downarrow$. □

Vector majorization is a partial order, implying that $\boldsymbol{y} = \boldsymbol{x}$ if both $\boldsymbol{y} \prec_{\mathrm{m}} \boldsymbol{x}$ and $\boldsymbol{x} \prec_{\mathrm{m}} \boldsymbol{y}$. Vector majorization naturally lifts to majorization of Hermitian operators by considering the vector $\boldsymbol{\lambda}(\boldsymbol{A})$ of a Hermitian operator $\boldsymbol{A}$'s eigenvalues.

Definition 2.13 (Operator majorization)**.** Let $\boldsymbol{A}, \boldsymbol{B} \in \mathbb{C}_{\mathrm{H}}^{d\times d}$ be Hermitian. The operator $\boldsymbol{B}$ *majorizes* $\boldsymbol{A}$ if $\boldsymbol{\lambda}(\boldsymbol{A}) \prec_{\mathrm{m}} \boldsymbol{\lambda}(\boldsymbol{B})$. This is denoted $\boldsymbol{A} \prec_{\mathrm{op}} \boldsymbol{B}$. □

The theory of operator majorization is well-developed, and it provides means for deriving results about unital quantum systems throughout this work. We quote the following result from quantum information theory.

[9] This situation is an example of when it is helpful to think about a quantum control system in terms of its state transition map rather than its differential dynamics.

Theorem 2.6 (Uhlmann [81, p. 242])**.** Let $\boldsymbol{A}, \boldsymbol{B} \in \mathbb{C}_{\mathrm{H}}^{d\times d}$ be Hermitian. The following are equivalent:

1. (operator majorization) $\boldsymbol{B} \prec_{\mathrm{op}} \boldsymbol{A}$;
2. (image under mixed unitary channel) there exists a mixed unitary channel $K(\cdot) \in \mathcal{C}(\mathbb{C}^d)$ such that $\boldsymbol{B} = K(\boldsymbol{A})$;
3. (image under unital channel) there exists a unital channel $K(\cdot) \in \mathcal{C}(\mathbb{C}^d)$ such that $\boldsymbol{B} = K(\boldsymbol{A})$; and
4. (image under unital map) there exists a positive, trace preserving, and unital map $K(\cdot) : \mathbb{C}^{d\times d} \to \mathbb{C}^{d\times d}$ such that $\boldsymbol{B} = K(\boldsymbol{A})$. □

Corollary 2.1. Let Σ be a unital quantum control system. For any $\boldsymbol{\Xi} \in \mathcal{D}(\mathcal{H})$, $\boldsymbol{u} \in \Omega$, and $t_{\mathrm{f}} \geqslant t_0$, the following hold:

1. (entropy monotone) $S(\Phi(t_{\mathrm{f}}; t_0, \boldsymbol{\Xi}, \boldsymbol{u})) \geqslant S(\boldsymbol{\Xi})$;
2. (rank monotone) $\operatorname{rank}\{\Phi(t_{\mathrm{f}}; t_0, \boldsymbol{\Xi}, \boldsymbol{u})\} \geqslant \operatorname{rank}\{\boldsymbol{\Xi}\}$. □

Proof. Let $\boldsymbol{\Xi} \in \mathcal{D}(\mathcal{H})$, $\boldsymbol{u} \in \Omega$, and $t_{\mathrm{f}} \geqslant t_0$.

1. By the relation between the von Neumann and Shannon entropies:

$$S(\Phi(t_{\mathrm{f}}; t_0, \boldsymbol{\Xi}, \boldsymbol{u})) = S_{\mathrm{s}}(\boldsymbol{\lambda}(\Phi(t_{\mathrm{f}}; t_0, \boldsymbol{\Xi}, \boldsymbol{u}))) \tag{2.55a}$$
$$S(\boldsymbol{\Xi}) = S_{\mathrm{s}}(\boldsymbol{\lambda}(\boldsymbol{\Xi}))\,. \tag{2.55b}$$

   By Definition 2.10 and Theorem 2.6, $\boldsymbol{\lambda}(\Phi(t_{\mathrm{f}}; t_0, \boldsymbol{\Xi}, \boldsymbol{u})) \prec_{\mathrm{m}} \boldsymbol{\lambda}(\boldsymbol{\Xi})$. Since Shannon entropy is Schur concave[10] [83, p. 556], $S_{\mathrm{s}}(\boldsymbol{\lambda}(\boldsymbol{\Xi})) \leqslant S_{\mathrm{s}}(\boldsymbol{\lambda}(\Phi(t_{\mathrm{f}}; t_0, \boldsymbol{\Xi}, \boldsymbol{u})))$ and the result follows.

2. By Definition 2.10 and Theorem 2.6, $\Phi(t_{\mathrm{f}}; t_0, \boldsymbol{\Xi}, \boldsymbol{u})$ can be written in mixed unitary form

$$\Phi(t_{\mathrm{f}}; t_0, \boldsymbol{\Xi}, \boldsymbol{u}) = \sum_j p_j\, \boldsymbol{U}_j \boldsymbol{\Xi} \boldsymbol{U}_j^\dagger \tag{2.56}$$

   where $p_j\, \boldsymbol{U}_j \boldsymbol{\Xi} \boldsymbol{U}_j^\dagger \succcurlyeq \boldsymbol{0}$ for each j. Thus,

$$\operatorname{rank}\{\Phi(t_{\mathrm{f}}; t_0, \boldsymbol{\Xi}, \boldsymbol{u})\} \;\geqslant\; \min_{j\,:\,p_j\neq 0} \operatorname{rank}\{\, p_j\, \boldsymbol{U}_j \boldsymbol{\Xi} \boldsymbol{U}_j^\dagger\} = \operatorname{rank}\{\boldsymbol{\Xi}\} \tag{2.57}$$

   and the result is proved.

⊠

Definition 2.14 (Majorization monotone system)**.** A quantum control system Σ is said to be (operator) *majorization monotone* if or any $\boldsymbol{\Xi} \in \mathcal{D}(\mathcal{H})$, $\boldsymbol{u} \in \Omega$, and $t_{\mathrm{f}} \geqslant t_0$, it follows that $\Phi(t_{\mathrm{f}}; t_0, \boldsymbol{\Xi}, \boldsymbol{u}) \prec_{\mathrm{op}} \boldsymbol{\Xi}$. □

[10] A function $f : \mathbb{R}^n \to \mathbb{R}$ is said to be *Schur concave* if $\boldsymbol{y} \prec_{\mathrm{m}} \boldsymbol{x}$ implies that $f(\boldsymbol{x}) \leqslant f(\boldsymbol{y})$.

It turns out that unitality and majorization monotonicity are equivalent.

Corollary 2.2. Let Σ be a quantum control system on $\mathcal{H} = \mathbb{C}^d$. The system Σ is unital if and only if it is majorization monotone. □

Proof. ($\implies$) If the system Σ is unital, then $\Phi(t_\mathrm{f}; t_0, \cdot, \boldsymbol{u})$ is a unital quantum channel. Equivalence between statements 1 and 3 of Theorem 2.6 implies that $\Phi(t_\mathrm{f}; t_0, \boldsymbol{\Xi}, \boldsymbol{u}) \prec_{\mathrm{op}} \boldsymbol{\Xi}$, and therefore the system Σ is majorization monotone.

($\impliedby$) If the system Σ is majorization monotone then $\Phi(t_\mathrm{f}; t_0, \frac{1}{d}\boldsymbol{I}, \boldsymbol{u}) \prec_{\mathrm{op}} \frac{1}{d}\boldsymbol{I}$. Since the maximally mixed state is majorized by all quantum states, the reverse direction $\frac{1}{d}\boldsymbol{I} \prec_{\mathrm{op}} \Phi(t_\mathrm{f}; t_0, \frac{1}{d}\boldsymbol{I}, \boldsymbol{u})$ holds as well. Since vector majorization is a partial order, the eigenvalues of $\Phi(t_\mathrm{f}; t_0, \frac{1}{d}\boldsymbol{I}, \boldsymbol{u})$ and $\frac{1}{d}\boldsymbol{I}$ are the same and the two states are unitarily equivalent. But this means that $\Phi(t_\mathrm{f}; t_0, \frac{1}{d}\boldsymbol{I}, \boldsymbol{u}) = \frac{1}{d}\boldsymbol{I}$ and therefore system Σ is unital. ⊠

The following result was studied in [84] for differential quantum systems with constant Lindblad operators $\boldsymbol{L}_j(\boldsymbol{u}(t), t) = \boldsymbol{L}_j$. With the preliminaries established herein, the generalization is immediate.

Corollary 2.3. Let Σ be a differential quantum control system with dynamics (2.19). The system Σ is majorization monotone if and only if, for each $\boldsymbol{u} \in \Omega$,

$$\sum_{j\in\mathcal{J}} \boldsymbol{L}_j(\boldsymbol{u}(t), t)\boldsymbol{L}_j^\dagger(\boldsymbol{u}(t), t) = \sum_{j\in\mathcal{J}} \boldsymbol{L}_j^\dagger(\boldsymbol{u}(t), t)\boldsymbol{L}_j(\boldsymbol{u}(t), t) \tag{2.58}$$

for almost all $t \in T$ (with respect to Lebesgue measure). □

Proof. By Theorem 2.5, the differential quantum control system Σ is unital if and only if (2.58) holds. The result follows by Corollary 2.2. ⊠

2.4.4 Approximating quantum dynamics

Consider two differential quantum systems with dynamics

$$\dot{\boldsymbol{X}}(t) = F(\boldsymbol{X}(t), \boldsymbol{u}(t), t) + \sum_{i\in\mathcal{I}} \gamma_i(t)\Big[2\boldsymbol{L}_i\boldsymbol{X}(t)\boldsymbol{L}_i^\dagger - [\![\boldsymbol{L}_i^\dagger\boldsymbol{L}_i, \boldsymbol{X}(t)]\!]_+\Big] \tag{2.59a}$$

$$\dot{\boldsymbol{X}}^{\mathrm{i}}(t) = F(\boldsymbol{X}^{\mathrm{i}}(t), \boldsymbol{u}(t), t) \tag{2.59b}$$

where $\mathcal{I}$ is a set of finite cardinality, $\boldsymbol{L}_i \in \mathbb{C}^{d\times d}$ for each $i \in \mathcal{I}$, and $\gamma_i(\cdot)$ is non-negative and locally integrable for each $i \in \mathcal{I}$. Here, the "i" in $\boldsymbol{X}^{\mathrm{i}}(t)$ stands for "idealized". The question of interest is: how far apart are the true and idealized trajectories at some time $t_\mathrm{f} \geqslant t_0$ if both start from the same initial condition $\boldsymbol{X}(t_0) = \boldsymbol{X}^{\mathrm{i}}(t_0) \in \mathbb{C}^{d\times d}$. The following theorem provides a bound, and the key to the proof is the contraction property (Lemma 2.1) of quantum systems.

Theorem 2.7 (Approximating quantum dynamics)**.** For any $t_\mathrm{f} \geqslant t_0$, the trajectories of (2.59) satisfy the inequality

$$\|\boldsymbol{X}(t_\mathrm{f}) - \boldsymbol{X}^{\mathrm{i}}(t_\mathrm{f})\|_1 \leqslant 4\|\boldsymbol{X}(t_0)\|_1 \int_{t_0}^{t_\mathrm{f}} \sum_{i\in\mathcal{I}} \gamma_i(t)\|\boldsymbol{L}_i\|_1^2 \,\mathrm{d}t\,. \tag{2.60}$$

If the system described by state $\boldsymbol{X}^{\mathrm{i}}(t)$ is unital (see Theorem 2.5) then, for any $p \in [1, \infty]$,

$$\|\boldsymbol{X}(t_{\mathrm{f}}) - \boldsymbol{X}^{\mathrm{i}}(t_{\mathrm{f}})\|_p \leqslant 4\|\boldsymbol{X}(t_0)\|_1 \int_{t_0}^{t_{\mathrm{f}}} \sum_{i\in\mathcal{I}} \gamma_i(t)\|\boldsymbol{L}_i\|_p^2 \,\mathrm{d}t \tag{2.61}$$

where $\|\cdot\|_p$ is the Schatten p-norm. If the system described by state $\boldsymbol{X}(t)$ is also unital, then $\|\boldsymbol{X}(t_0)\|_1$ in (2.61) can be replaced by $\|\boldsymbol{X}(t_0)\|_p$. □

Proof. Define an error system with state $\boldsymbol{E}(t) \triangleq \boldsymbol{X}(t) - \boldsymbol{X}^{\mathrm{i}}(t)$ and dynamics

$$\dot{\boldsymbol{E}}(t) = F(\boldsymbol{E}(t), \boldsymbol{u}(t), t) + \sum_{i\in\mathcal{I}} \gamma_j(t)\Big[2\boldsymbol{L}_i\boldsymbol{X}(t)\boldsymbol{L}_i^\dagger - [\![\boldsymbol{L}_i^\dagger\boldsymbol{L}_i, \boldsymbol{X}(t)]\!]_+\Big]. \tag{2.62}$$

Let Φ denote the state transition map associated with $\boldsymbol{X}^{\mathrm{i}}(\cdot)$'s dynamics. By Lemma 2.1, $\Phi(t; t_0, \cdot, \boldsymbol{u})$ is a contraction (with respect to the trace norm) for all $t \in [t_0, t_{\mathrm{f}}]$. By the same lemma, $\|\boldsymbol{X}(t)\|_1 \leqslant \|\boldsymbol{X}(t_0)\|_1$ for all $t \in [t_0, t_{\mathrm{f}}]$. It follows from the variation of constants formula [21, p. 40] that

$$\boldsymbol{E}(t_{\mathrm{f}}) = \Phi(t_{\mathrm{f}}; t_0, \boldsymbol{0}, \boldsymbol{u}) + \int_{t_0}^{t_{\mathrm{f}}} \Phi\Big(t_{\mathrm{f}}; t, \sum_{i\in\mathcal{I}} \gamma_i(t)\Big[2\boldsymbol{L}_i\boldsymbol{X}(t)\boldsymbol{L}_i^\dagger - [\![\boldsymbol{L}_i^\dagger\boldsymbol{L}_i, \boldsymbol{X}(t)]\!]_+\Big], \boldsymbol{u}\Big)\,\mathrm{d}t\,. \tag{2.63}$$

Since $\Phi(t_{\mathrm{f}}; t_0, \cdot, \boldsymbol{u})$ is linear, $\Phi(t_{\mathrm{f}}; t_0, \boldsymbol{0}, \boldsymbol{u}) = \boldsymbol{0}$. It follows that

$$\begin{aligned}
\|\boldsymbol{E}(t_{\mathrm{f}})\|_1 &\leqslant \int_{t_0}^{t_{\mathrm{f}}} \Big\|\Phi\Big(t_{\mathrm{f}}; t, \sum_{i\in\mathcal{I}} \gamma_i(t)\Big[2\boldsymbol{L}_i\boldsymbol{X}(t)\boldsymbol{L}_i^\dagger - [\![\boldsymbol{L}_i^\dagger\boldsymbol{L}_i, \boldsymbol{X}(t)]\!]_+\Big], \boldsymbol{u}\Big)\Big\|_1 \,\mathrm{d}t \\
&\leqslant \int_{t_0}^{t_{\mathrm{f}}} \Big\|\sum_{i\in\mathcal{I}} \gamma_i(t)\Big[2\boldsymbol{L}_i\boldsymbol{X}(t)\boldsymbol{L}_i^\dagger - [\![\boldsymbol{L}_i^\dagger\boldsymbol{L}_i, \boldsymbol{X}(t)]\!]_+\Big]\Big\|_1 \,\mathrm{d}t \\
&\leqslant 4\int_{t_0}^{t_{\mathrm{f}}} \sum_{i\in\mathcal{I}} \gamma_i(t)\|\boldsymbol{X}(t)\|_1\|\boldsymbol{L}_i\|_1^2 \,\mathrm{d}t \\
&\leqslant 4\|\boldsymbol{X}(t_0)\|_1 \int_{t_0}^{t_{\mathrm{f}}} \sum_{i\in\mathcal{I}} \gamma_i(t)\|\boldsymbol{L}_i\|_1^2 \,\mathrm{d}t
\end{aligned} \tag{2.64}$$

by the triangle inequality and the submultiplicativity of the trace norm [85, Thm. 7.4.10.1]. This proves (2.60).

The proof of (2.61) proceeds as before. Using the fact that the state transition map of a unital quantum system is contractive under any Schatten p-norm, one arrives at the conclusion

$$\|\boldsymbol{E}(t_{\mathrm{f}})\|_p \leqslant 4\int_{t_0}^{t_{\mathrm{f}}} \sum_{i\in\mathcal{I}} \gamma_i(t)\|\boldsymbol{X}(t)\|_p\|\boldsymbol{L}_i\|_p^2 \,\mathrm{d}t\,. \tag{2.65}$$

The monotonicity of Schatten norms gives $\|\boldsymbol{X}(t)\|_p \leqslant \|\boldsymbol{X}(t)\|_1 \leqslant \|\boldsymbol{X}(t_0)\|_1$ for all $t \in [t_0, t_{\mathrm{f}}]$. This proves (2.61).

If the system described by state $\boldsymbol{X}(t)$ is also unital, then one has $\|\boldsymbol{X}(t)\|_p \leqslant \|\boldsymbol{X}(t_0)\|_p$. Combined with (2.65), the final claim of the theorem is proved. ⊠

Remark 2.5 (Alternative to Grönwall's inequality)**.** In the theory of classical control, Grönwall's inequality [6, p. 64] is used to bound the distance between trajectories of two systems starting from the same initial condition. This bound is exponential in the perturbation dynamics. However, using the contractivity of quantum systems, the bound in Theorem 2.7 circumvents exponentiation. In this thesis, Lemma 2.1 and Theorem 2.7 fill the roles typically filled by Grönwall. □

Example 2.2 (Unitary gates in quantum computing)**.** A model for qubit dynamics ($\mathcal{H} = \mathbb{C}^2$) is

$$\dot{\boldsymbol{\Xi}}(t) = -\imath[\![\tfrac{\omega_0}{2}\boldsymbol{P}_z + u(t)\boldsymbol{P}_y, \boldsymbol{\Xi}(t)]\!]_- + \frac{1}{2\tau_{\text{rel}}}\Big[2\boldsymbol{P}_-\boldsymbol{\Xi}(t)\boldsymbol{P}_-^\dagger - [\![\boldsymbol{P}_-^\dagger\boldsymbol{P}_-, \boldsymbol{\Xi}(t)]\!]_+\Big] + \frac{1}{4\tau_{\text{dep}}}\Big[2\boldsymbol{P}_z\boldsymbol{\Xi}(t)\boldsymbol{P}_z^\dagger - [\![\boldsymbol{P}_z^\dagger\boldsymbol{P}_z, \boldsymbol{\Xi}(t)]\!]_+\Big] \tag{2.66a}$$

$$= -\imath[\![\tfrac{\omega_0}{2}\boldsymbol{P}_z + u(t)\boldsymbol{P}_y, \boldsymbol{\Xi}(t)]\!]_- + \frac{1}{2\tau_{\text{rel}}}\Big[2\boldsymbol{P}_-\boldsymbol{\Xi}(t)\boldsymbol{P}_+ - [\![\boldsymbol{P}_+\boldsymbol{P}_-, \boldsymbol{\Xi}(t)]\!]_+\Big] + \frac{1}{2\tau_{\text{dep}}}\Big[\boldsymbol{P}_z\boldsymbol{\Xi}(t)\boldsymbol{P}_z^\dagger - \boldsymbol{\Xi}(t)\Big] \tag{2.66b}$$

where $u(t) \in \mathbb{R}$; $\omega_0 \in \mathbb{R}$ is the qubit frequency; $\boldsymbol{P}_y$ and $\boldsymbol{P}_z$ are the Pauli-y and -z operators; $\boldsymbol{P}_- \triangleq |0\rangle\langle 1|$ is the lowering operator; and $\boldsymbol{P}_+ \triangleq |1\rangle\langle 0|$ is the raising operator. The constant $\tau_{\text{rel}} > 0$ is the "relaxation time" (commonly denoted by T_1) and $\tau_{\text{dep}} > 0$ is the "pure dephasing time" (commonly denoted by T_ϕ) [86, p. 11].

In quantum computing, one is interested in using control to achieve various unitary transformations of the initial qubit state $\boldsymbol{\Xi}(t_0) \in \mathcal{D}(\mathbb{C}^2)$. These transformations form the logical gates of a quantum computer. The true dynamics (2.66) are approximated, on short time intervals, by the idealized dynamics

$$\dot{\boldsymbol{\Xi}}^{\text{i}}(t) = -\imath[\![\tfrac{\omega_0}{2}\boldsymbol{P}_z + u(t)\boldsymbol{P}_y, \boldsymbol{\Xi}^{\text{i}}(t)]\!]_- \,. \tag{2.67}$$

The time τ_{g} required for control to implement a unitary transformation of the state $\boldsymbol{\Xi}^{\text{i}}(t_0)$ is called the *gate time*. Given a gate time τ_{g}, Theorem 2.7 provides a bound on the error between the true and idealized trajectories after $n \in \mathbb{N}$ gates have been implemented:

$$\|\boldsymbol{\Xi}(n\tau_{\text{g}}) - \boldsymbol{\Xi}^{\text{i}}(n\tau_{\text{g}})\|_1 \leqslant n\tau_{\text{g}}\Big[\frac{2}{\tau_{\text{rel}}} + \frac{4}{\tau_{\text{dep}}}\Big] \tag{2.68}$$

where the fact that $\|\boldsymbol{\Xi}(t_0)\|_1 = 1$ (since $\boldsymbol{\Xi}(t_0) \in \mathcal{D}(\mathbb{C}^2)$) has been used. □

2.4.5 On the "open" and "closed" terminology

The expressions "open quantum systems" and "closed quantum systems" are interpreted in a myriad of different ways in today's literature. Consistent with the physical interpretation of such expressions, the following conventions are adopted:

1. an *open* quantum system is one that interacts with its environment, and
2. a *closed* quantum system is one that is perfectly isolated from its environment.

Hence, open and closed are mutually exclusive—a quantum system cannot be both open and closed. The Lindblad equation (2.19) can model either type of system.

Some existing literature refers to a (differential) quantum system whose dynamics take the form (2.19) as an "open quantum system". This is inconsistent with the terminology used herein. To avoid confusion, this thesis uses the terminology *differential quantum control system* to describe a generic quantum control system whose dynamics arise from (2.19). Such a system may be open or closed. The terminology *unital* is used to describe a system whose state transition map is unital. The prefix *differential* is added when appropriate. A unital system may be open or closed. The terminology *unitary* is used to describe a system whose state transition map is unitary. A unitary system is almost certainly closed.

Chapter 3

Optimal quantum control

Optimal control theory is a well-developed area of classical control, e.g., see [5, 6, 20, 87]. There is significant literature devoted to applying concepts from optimal control theory to quantum systems, e.g., see [40, 88] and the references therein. The classic works [5] and [6] give a formal treatment of optimal control in infinite dimensional Banach space. From this perspective, optimal quantum control[1] problems are a special case of existing classical theory. The vision that motivated the present thesis, however, is that the additional structure provided by differential quantum control systems allows for a rather expedient treatment of optimality. The goal of this chapter is to establish a self-contained treatment of optimal quantum control leveraging quantum structure.

This chapter begins by recalling notions of differentiation in Banach and Hilbert spaces. This provides the formal language for optimizing functionals on said spaces. The necessary conditions for optimal control are then obtained by viewing an optimal control objective as a functional on Banach space defined implicitly through the dynamics of a differential quantum control system. The necessary conditions give rise to the well-known gradient-based optimization schemes for optimal quantum control. A gradient algorithm for free terminal time optimal quantum control problems is provided. Lastly, a concise proof of the minimum principle is given.[2]

3.1 Differentiation in Banach and Hilbert spaces

Banach space provides the structure needed to define a derivative in infinite dimensional vector space. An authoritative reference on the subject is [89]. Let $\mathcal{X}$ be a Banach space with norm $\|\cdot\|_{\mathcal{X}}$, let $\mathcal{A} \subseteq \mathcal{X}$ be a non-empty open set, and let $a \in \mathcal{A}$. A functional $g : \mathcal{A} \to \mathbb{R}$ is said to be "little-o of $\|x-a\|_{\mathcal{X}}$ as $x \to a$", written $g(x) = o(\|x-a\|_{\mathcal{X}})$, if

$$\lim_{x \to a} \frac{|g(x)|}{\|x-a\|_{\mathcal{X}}} = 0\,. \tag{3.1}$$

[1] Most of the existing literature uses the terminology "quantum optimal control". This terminology could be misleading as it may refer to quantum methods for solving classical optimal control problems.

[2] Also known as "Pontryagin's maximum principle."

Since the open set $\mathcal{A}$ contains the vector a, this limit is well-defined. The condition (3.1) can be equivalently written in ϵ-δ form: for every $\epsilon > 0$, there exists $\delta > 0$ such that

$$x \in \mathcal{A}, \;\; 0 < \|x - a\|_{\mathcal{X}} \leqslant \delta \implies |g(x)| \leqslant \epsilon \|x - a\|_{\mathcal{X}} \,. \tag{3.2}$$

Definition 3.1 (Fréchet derivative [6, p. 79])**.** Let $\mathcal{X}$ be a Banach space with norm $\|\cdot\|_{\mathcal{X}}$ and let $\mathcal{O} \subseteq \mathcal{X}$ be a non-empty open set. A functional $f : \mathcal{O} \to \mathbb{R}$ is said to be *Fréchet differentiable* at a point $x \in \mathcal{O}$ if there exists a bounded (i.e., continuous) linear functional $D_x f(x)[\cdot] : \mathcal{O} \to \mathbb{R}$ such that

$$f(x') = f(x) + D_x f(x)[x' - x] + o(\|x' - x\|_{\mathcal{X}}) \,. \tag{3.3}$$

The operator $D_x f(x)[\cdot]$ is called the *Fréchet derivative* of f at x. □

In the sequel, the prefix "Fréchet" is omitted unless there is need for emphasis. From (3.3), it is clear that the derivative of f at x is the best linear approximation of the function at that point. This perspective will prove useful in the sequel.

A Hilbert space is a special case of a Banach space, thus Definition 3.1 applies to differentiation in Hilbert space as well. However, the additional inner product structure that accompanies a Hilbert space allows for an alternative (but equivalent) definition of differentiation and gives rise to the notion of a gradient.

Definition 3.2. Let $\mathcal{H}$ be a Hilbert space with inner product $\langle \cdot, \cdot \rangle_{\mathcal{H}}$ and induced norm $\|\cdot\|_{\mathcal{H}}$; let $\mathcal{O} \subseteq \mathcal{H}$ be a non-empty open set. A functional $f : \mathcal{O} \to \mathbb{R}$ is said to be *differentiable* at a point $x \in \mathcal{O}$ if there exists $g \in \mathcal{H}$ such that

$$f(x') = f(x) + \langle g, x' - x \rangle_{\mathcal{H}} + o(\|x' - x\|_{\mathcal{H}}) \tag{3.4}$$

The operator $D_x f(x)[\cdot] = \langle g, \cdot \rangle_{\mathcal{H}}$ is said to be the *derivative* of f at x, and g is said to be the *gradient* of f at x. The notation $\nabla_x f(x) = g$ is used to denote the gradient of f at x. □

The Riesz representation theorem [90, p. 345] ensures that Definitions 3.1 and 3.2 are in harmony. The Riesz representation theorem is one motivating factor for insisting that the "derivative" in Definition 3.1 be bounded. The following examples demonstrate how Fréchet derivatives are calculated. These examples will prove useful in the subsequent derivation of optimal quantum control results.

Example 3.1 (Derivatives of a linear functionals)**.** In ordinary calculus, the derivative of a linear function $x \mapsto ax$ (a and x real numbers) is simply the number a itself (the number a is a mapping, via multiplication, from $\mathbb{R}$ to $\mathbb{R}$). This behavior extends to functionals in Hilbert (or Banach) space. Consider the case of $\mathcal{H} = L^2(\mathbb{R})$ (the Hilbert space of Lebesgue square integrable functions on the real line), $g \in \mathcal{H}$, and the linear functional $f(h) = \int_{\mathbb{R}} g(x)h(x)\mathrm{d}x \triangleq \langle g, h \rangle_{\mathcal{H}}$. The derivative of f is simply f itself, and the gradient of f is g. □

The following examples demonstrate how to calculate derivatives and gradients of functionals taking the form $f(h) = \int_\Omega g(h(x))\,\mathrm{d}x$ or $f(h) = \int_\Omega g(h(x), x)\,\mathrm{d}x$. Two different techniques for calculating said gradients are given—the technical assumptions for each technique are different.

Example 3.2 (Derivatives and gradients in $L^2(\Omega)$)**.** Let $\Omega \subset \mathbb{R}$ be a set of finite Lebesgue measure and let $L^2(\Omega)$ denote the Hilbert space of square Lebesgue integrable functions from Ω to $\mathbb{R}$. Let $g : \mathbb{R} \to \mathbb{R}$ be bounded with bounded, continuous first and second derivatives; let $f : L^2(\Omega) \to \mathbb{R}$ be the functional

$$f(h) = \int_\Omega g(h(x))\,\mathrm{d}x\,, \qquad h \in L^2(\Omega)\,. \tag{3.5}$$

For $h, k \in L^2(\Omega)$ and $x \in \Omega$, Taylor's theorem [91, pp. 110-111] applied to the function $g(\cdot)$ at the point $h(x)$ gives

$$g(h(x)+k(x)) = g(h(x)) + g'(h(x))k(x) + \underbrace{\frac{g''(c)}{2}k^2(x)}_{\triangleq R(x)} \tag{3.6}$$

for some $c \in \mathbb{R}$ between $h(x)$ and $k(x)$. By assumption, there exists $C > 0$ such that $|g''(c)| \leqslant 2C$ for all $c \in \mathbb{R}$. Therefore,

$$|R(x)| = \big|g(h(x)+k(x)) - g(h(x)) - g'(h(x))k(x)\big| \leqslant C\,|k(x)|^2\,, \qquad \forall x \in \Omega \tag{3.7}$$

and consequently $\int_\Omega R(x)\,dx = o(\|k(\cdot)\|_{L^2})$ since

$$\Big|\int_\Omega R(x)\,dx\Big| \leqslant \int_\Omega |R(x)|\,dx \leqslant C\,\|k(\cdot)\|_{L^2}^2 \implies \lim_{k(\cdot)\to 0} \frac{|\int_\Omega R(x)\,dx|}{\|k(\cdot)\|_{L^2}} = 0\,. \tag{3.8}$$

Integrating (3.6) gives

$$f(h+k) = f(h) + \int_\Omega g'(h(x))k(x)\,\mathrm{d}x + o(\|k(\cdot)\|_{L^2})\,. \tag{3.9}$$

Note that $k(\cdot) \mapsto \int_\Omega g'(h(x))k(x)\,\mathrm{d}x$ is a bounded operator: Cauchy-Schwarz and boundedness of $g'(\cdot)$ give

$$\Big|\int_\Omega g'(h(x))k(x)\,\mathrm{d}x\Big| \leqslant \|g'(h(\cdot))\|_{L^2}\,\|k(\cdot)\|_{L^2} \leqslant M\|k(\cdot)\|_{L^2} \tag{3.10}$$

for some $M > 0$. Therefore, the derivative of the functional $f(\cdot)$ at $h(\cdot)$ is

$$D_h f(h)[\cdot] = \int_\Omega g'(h(x))[\cdot]\,\mathrm{d}x = \langle g' \circ h, \cdot\rangle_{L^2} \tag{3.11}$$

and its gradient is $g' \circ h \in L^2(\Omega)$. □

Example 3.3 (Derivatives and gradients in $L^2(\Omega;\mathbb{R}^n)$)**.** Let $\Omega \subseteq \mathbb{R}$ be a set of finite Lebesgue measure and let $L^2(\Omega;\mathbb{R}^n)$ denote the Hilbert space of square Lebesgue integrable functions from Ω to $\mathbb{R}$. Let $g : \mathbb{R}^n \times \Omega \to \mathbb{R}$ be Carathéodory (see Definition B.1) and continuously differentiable in its first argument. Suppose there exists $C > 0$, $\ell_1(\cdot) \in L^1(\Omega)$, and $\ell_2(\cdot) \in L^2(\Omega)$ such that $|g(u,x)| \leqslant C\|u\|^2 + \ell_1(x)$ and $\|g'(u,x)\| \leqslant C\|u\| + \ell_2(x)$ for all $x \in \mathbb{R}$. Let $f : L^2(\Omega;\mathbb{R}^n) \to \mathbb{R}$ be the functional

$$f(h) = \int_\Omega g(h(x),x)\,\mathrm{d}x\,, \qquad h \in L^2(\Omega;\mathbb{R}^n)\,, \tag{3.12}$$

which is well-defined by the $|g(u,x)| \leqslant C\|u\|^2 + \ell_1(x)$ hypothesis. We will show that the derivative of f at h is

$$D_h f(h)[\cdot] = \int_\Omega \langle g'(h(x),x), \cdot \rangle \,\mathrm{d}x \tag{3.13}$$

where $g'(h(x),x) = \nabla_u g(u,x)\big|_{u=h(x)}$, and the gradient of f at h is therefore $g'(h(\cdot),\cdot) \in L^2(\Omega;\mathbb{R}^n)$. One can readily verify that the linear functional (3.13) is bounded. Hence, we need only verify that it satisfies the definition (3.3) of the derivative. For $h, k \in L^2(\Omega;\mathbb{R}^n)$,

$$\begin{aligned} g(h(x)+k(x),x) - g(h(x),x) - \langle g'(h(x),x), k(x)\rangle & \\ &= \Big[\int_0^1 \frac{\mathrm{d}}{\mathrm{d}t} g(h(x)+tk(x))\,\mathrm{d}t\Big] - \langle g'(h(x),x), k(x)\rangle \\ &= \int_0^1 \langle g'(h(x)+tk(x)) - g'(h(x)), k(x)\rangle \,\mathrm{d}t \end{aligned} \tag{3.14}$$

and therefore, invoking Tonelli's theorem [92, p. 111] and the Cauchy-Schwarz inequality,

$$\begin{aligned} \Big| f(h+k) - f(h) - \int_\Omega \langle g'(h(x),x), k(x)\rangle \,\mathrm{d}x\Big| & \\ &\leqslant \int_\Omega \int_0^1 \big|\langle g'(h(x)+tk(x)) - g'(h(x)), k(x)\rangle\big| \,\mathrm{d}t\,\mathrm{d}x \\ &= \int_0^1 \int_\Omega \big|\langle g'(h(x)+tk(x)) - g'(h(x)), k(x)\rangle\big| \,\mathrm{d}x\,\mathrm{d}t \\ &\leqslant \int_0^1 \int_\Omega \|g'(h(x)+tk(x)) - g'(h(x))\|\,\|k(x)\| \,\mathrm{d}x\,\mathrm{d}t \\ &\leqslant \|k(\cdot)\|_{L^2} \int_0^1 \|g'(h(\cdot)+tk(\cdot)) - g'(h(\cdot))\|_{L^2} \,\mathrm{d}t\,. \end{aligned} \tag{3.15}$$

We want to show that $|f(h+k) - f(h) - \int_\Omega \langle g'(h(x),x), k(x)\rangle \,\mathrm{d}x| = o(\|k(\cdot)\|_{L^2})$. Let $\{k_n(\cdot)\}_{n\in\mathbb{N}} \subset L^2(\Omega;\mathbb{R}^n)$ be a sequence tending to zero in $L^2(\Omega;\mathbb{R}^n)$. Given (3.15), it need only be proved that $\int_0^1 \|g'(h(\cdot)+tk_n(\cdot)) - g'(h(\cdot))\|_{L^2}\,\mathrm{d}t \to 0$ as $n\to\infty$. To do this, we will invoke the bounded convergence theorem (see Theorem A.1), which requires that the family of functions

$$\Big\{\zeta_n(t) \triangleq \|g'(h(\cdot)+tk_n(\cdot),\cdot) - g'(h(\cdot))\|_{L^2} \;:\; n\in\mathbb{N}\Big\} \tag{3.16}$$

is uniformly bounded and converges to zero for all $t\in[0,1]$. Using the inequality $(a+b)^2 \leqslant 2(a^2+b^2)$ for $a,b\in\mathbb{R}$:

$$\begin{aligned} \zeta_n^2(t) &= \|g'(h(\cdot)+tk_n(\cdot),\cdot) - g'(h(\cdot))\|_{L^2}^2 \\ &= \int_\Omega \|g'(h(x)+tk_n(x),x) - g'(h(x))\|^2 \,\mathrm{d}x \\ &\leqslant 2\int_\Omega \|g'(h(x)+tk_n(x),x)\|^2 + \|g'(h(x))\|^2 \,\mathrm{d}x \\ &\leqslant 2\int_\Omega (\ell_2(x) + C\|h(x)+tk_n(x)\|)^2 + (\ell_2(x)+C\|h(x)\|)^2 \,\mathrm{d}x \end{aligned}$$

$$
\begin{aligned}
&\leqslant 4\int_\Omega 2\ell_2^2(x) + C^2\|h(x)+tk_n(x)\|^2 + C^2\|h(x)\|^2\,\mathrm{d}x\\
&\leqslant 4\int_\Omega 2\ell_2^2(x) + 2C^2(\|h(x)\|^2 + t^2\|k_n(x)\|^2) + C^2\|h(x)\|^2\,\mathrm{d}x\\
&= 4\int_\Omega 2\ell_2^2(x) + 3C^2\|h(x)\|^2 + 2t^2\|k_n(x)\|^2\,\mathrm{d}x\\
&\leqslant 8\|\ell_2(\cdot)\|_{L^2}^2 + 12C^2\|h(\cdot)\|_{L^2}^2 + 8\max_{n\in\mathbb{N}}\|k_n(\cdot)\|_{L^2}^2 < \infty\,. \qquad (3.17)
\end{aligned}
$$

Hence, the family of $\zeta_n(\cdot)$'s is uniformly bounded.

To show that the functions $\zeta_n(t)$ converge to zero for each $t\in[0,1]$, Vitali's convergence theorem (see Theorem A.2) will be used. Fix $t\in[0,1]$. Based on the earlier deductions,

$$
\|g'(h(x)+tk_n(x),x) - g'(h(x))\|_2^2 \leqslant 8\ell_2^2(x) + 12C^2\|h(x)\|^2 + 8\|k_n(x)\|^2\,. \qquad (3.18)
$$

The family of functions $\{\|k_n(\cdot)\|^2\in L^1(\Omega) : n\in\mathbb{N}\}$ on the finite Lebesgue measure set Ω is uniformly integrable since $\|k_n(\cdot)\|_{L^2}\to 0$ (this follows from Theorem A.2). By (3.18), it thus follows that the family $\{\|g'(h(x)+tk_n(x),x)-g'(h(x))\|_2^2 : n\in\mathbb{N}\}$ is uniformly integrable. On the other hand, we know that $k_n(\cdot)\to 0$ in measure. Since g' is continuous and Ω is of finite measure, $\|g'(h(\cdot)+tk_n(\cdot),\cdot)-g'(h(\cdot))\|^2\to 0$ in measure. By Vitali's convergence theorem, it follows that

$$
\begin{aligned}
\lim_{n\to\infty}\zeta_n(t) &= \lim_{n\to\infty}\int_\Omega \|g'(h(x)+tk_n(x),x)-g'(h(x))\|^2\,\mathrm{d}x\\
&= \int_\Omega \lim_{n\to\infty}\|g'(h(x)+tk_n(x),x)-g'(h(x))\|^2\,\mathrm{d}x\\
&= \int_\Omega 0\,\mathrm{d}x = 0\,. \qquad (3.19)
\end{aligned}
$$

Since the $\zeta_n(t)$'s are bounded and converge to zero pointwise, the bounded convergence theorem (Theorem A.1) implies that $\int_0^1\|g'(h(\cdot)+tk_n(\cdot))-g'(h(\cdot))\|_{L^2}\,\mathrm{d}t\to 0$ as $n\to\infty$. □

Example 3.4 (Differentiating functionals of complex matrices)**.** Consider a functional $K:\mathbb{C}^{d\times d}\to\mathbb{R}$. We view $\mathbb{C}^{d\times d}$ as a *real* vector space endowed with the inner product $\langle\cdot,\cdot\rangle:\mathbb{C}^{d\times d}\times\mathbb{C}^{d\times d}\to\mathbb{R}$ given by

$$
\langle \boldsymbol{A},\boldsymbol{B}\rangle = \mathrm{Re}\,\mathrm{tr}\{\boldsymbol{A}^\dagger\boldsymbol{B}\}\,,\qquad \forall \boldsymbol{A},\boldsymbol{B}\in\mathbb{C}^{d\times d}\,. \qquad (3.20)
$$

Equipped with this inner product, $\mathbb{C}^{d\times d}$ is a real Hilbert space with the Hilbert-Schmidt norm. For K to be Fréchet differentiable at $\boldsymbol{A}\in\mathbb{C}^{d\times d}$ means there exists a linear functional $L:\mathbb{C}^{d\times d}\to\mathbb{R}$ such that[3]

$$
K(\boldsymbol{B}) = K(\boldsymbol{A}) + L(\boldsymbol{B}-\boldsymbol{A}) + o(\|\boldsymbol{B}-\boldsymbol{A}\|_{\mathrm{HS}})\,,\qquad \forall \boldsymbol{B}\in\mathbb{C}^{d\times d}\,. \qquad (3.21)
$$

[3]All norms on a finite dimensional space are topologically equivalent, and equivalent norms lead to the same derivative [89, p. 27]. Hence, one could in principle replace $\|\cdot\|_{\mathrm{HS}}$ in (3.21) with any norm. Since $\|\cdot\|_{\mathrm{HS}}$ is generated by an inner product, working with this norm means that one has the additional notion of a gradient.

By the Riesz representation theorem, there exists $\boldsymbol{G} \in \mathbb{C}^{d\times d}$ such that

$$K(\boldsymbol{B}) = K(\boldsymbol{A}) + \operatorname{Re}\operatorname{tr}\{\boldsymbol{G}^\dagger(\boldsymbol{B}-\boldsymbol{A})\} + o(\|\boldsymbol{B}-\boldsymbol{A}\|_{\mathrm{HS}})\,, \qquad \forall \boldsymbol{B} \in \mathbb{C}^{d\times d}\,. \tag{3.22}$$

□

Example 3.5 (Differentiating functionals of Hermitian matrices)**.** Consider the *real* Hilbert space $\mathbb{C}_{\mathrm{H}}^{d\times d}$ of Hermitian matrices on $\mathbb{C}^{d\times d}$. The Hilbert-Schmidt inner product $\langle\cdot,\cdot\rangle_{\mathrm{HS}} : \mathbb{C}_{\mathrm{H}}^{d\times d} \times \mathbb{C}_{\mathrm{H}}^{d\times d} \to \mathbb{R}$ given by[4]

$$\langle \boldsymbol{A}, \boldsymbol{B}\rangle_{\mathrm{HS}} = \operatorname{tr}\{\boldsymbol{A}^\dagger\boldsymbol{B}\} = \operatorname{tr}\{\boldsymbol{A}\boldsymbol{B}\}\,, \qquad \forall \boldsymbol{A},\boldsymbol{B} \in \mathbb{C}_{\mathrm{H}}^{d\times d} \tag{3.23}$$

generates the Hilbert-Schmidt norm $\|\cdot\|_{\mathrm{HS}}$ on $\mathbb{C}_{\mathrm{H}}^{d\times d}$; equipped with this inner product $\mathbb{C}_{\mathrm{H}}^{d\times d}$ is a Hilbert space. For $K : \mathbb{C}_{\mathrm{H}}^{d\times d} \to \mathbb{R}$ to be Fréchet differentiable at $\boldsymbol{A} \in \mathbb{C}_{\mathrm{H}}^{d\times d}$ means that there exists a linear functional $M : \mathbb{C}_{\mathrm{H}}^{d\times d} \to \mathbb{R}$ such that

$$K(\boldsymbol{B}) = K(\boldsymbol{A}) + M(\boldsymbol{B}-\boldsymbol{A}) + o(\|\boldsymbol{B}-\boldsymbol{A}\|_{\mathrm{HS}})\,, \qquad \forall \boldsymbol{B} \in \mathbb{C}_{\mathrm{H}}^{d\times d}\,. \tag{3.24}$$

By the Riesz representation theorem, there exists $\boldsymbol{J} \in \mathbb{C}_{\mathrm{H}}^{d\times d}$ such that

$$\begin{aligned} K(\boldsymbol{B}) &= K(\boldsymbol{A}) + \langle \boldsymbol{J}, \boldsymbol{B}-\boldsymbol{A}\rangle_{\mathrm{HS}} + o(\|\boldsymbol{B}-\boldsymbol{A}\|_{\mathrm{HS}}) \\ &= K(\boldsymbol{A}) + \operatorname{tr}\{\boldsymbol{J}(\boldsymbol{B}-\boldsymbol{A})\} + o(\|\boldsymbol{B}-\boldsymbol{A}\|_{\mathrm{HS}})\,, \qquad \forall \boldsymbol{B} \in \mathbb{C}_{\mathrm{H}}^{d\times d}\,. \end{aligned} \tag{3.25}$$

Suppose that K is the same functional as in Example 3.4, restricted to the subset $\mathbb{C}_{\mathrm{H}}^{d\times d}$ of $\mathbb{C}^{d\times d}$. When $\boldsymbol{A} \in \mathbb{C}_{\mathrm{H}}^{d\times d}$, the gradients $\boldsymbol{J}$ and $\boldsymbol{G}$ are related by the equation

$$\boldsymbol{J} = \frac{\boldsymbol{G}+\boldsymbol{G}^\dagger}{2}\,. \tag{3.26}$$

To prove this, let $\boldsymbol{B} \in \mathbb{C}_{\mathrm{H}}^{d\times d}$ and observe that

$$\operatorname{Re}\operatorname{tr}\{\boldsymbol{G}^\dagger(\boldsymbol{B}-\boldsymbol{A})\} = \operatorname{Re}\overline{\operatorname{tr}\{\boldsymbol{G}^\dagger(\boldsymbol{B}-\boldsymbol{A})\}} = \operatorname{Re}\operatorname{tr}\{(\boldsymbol{B}-\boldsymbol{A})\boldsymbol{G}\} = \operatorname{Re}\operatorname{tr}\{\boldsymbol{G}(\boldsymbol{B}-\boldsymbol{A})\}\,. \tag{3.27}$$

Since the trace of the product of Hermitian matrices is real,

$$\operatorname{tr}\Big\{\frac{\boldsymbol{G}+\boldsymbol{G}^\dagger}{2}(\boldsymbol{B}-\boldsymbol{A})\Big\} = \operatorname{Re}\operatorname{tr}\Big\{\frac{\boldsymbol{G}+\boldsymbol{G}^\dagger}{2}(\boldsymbol{B}-\boldsymbol{A})\Big\} = \operatorname{Re}\operatorname{tr}\{\boldsymbol{G}^\dagger(\boldsymbol{B}-\boldsymbol{A})\}\,. \tag{3.28}$$

Combining this expression with (3.22) reveals that

$$K(\boldsymbol{B}) = K(\boldsymbol{A}) + \langle \tfrac{1}{2}(\boldsymbol{G}+\boldsymbol{G}^\dagger), \boldsymbol{B}-\boldsymbol{A}\rangle_{\mathrm{HS}} + o(\|\boldsymbol{B}-\boldsymbol{A}\|_{\mathrm{HS}})\,, \qquad \forall \boldsymbol{B} \in \mathbb{C}_{\mathrm{H}}^{d\times d}\,. \tag{3.29}$$

Since the Fréchet derivative/gradient of a function is unique [6, p. 79], it follows from (3.25) that $\boldsymbol{J} = \frac{1}{2}(\boldsymbol{G}+\boldsymbol{G}^\dagger)$. Thus, if the gradient of K on $\mathbb{C}^{d\times d}$ is known, the gradient of K on $\mathbb{C}_{\mathrm{H}}^{d\times d}$ can be computed. □

[4]Note that the trace of the product of Hermitian matrices is real.

3.2 Optimal control laws

In this section we will look at a differential quantum control system Σ. The generator of the system's dynamics is assumed to take the usual form introduced in (2.40):

$$F(\boldsymbol{X}, \boldsymbol{u}, t) = F_0(\boldsymbol{X}, t) + F_c(\boldsymbol{X}, \boldsymbol{u}, t) \tag{3.30}$$

where F_0 is linear (for fixed t) in the state and F_c is bilinear in the state and input variables (for fixed t). A typical optimal quantum control problem takes the form

$$\begin{aligned} \mathscr{P}: \quad \underset{\boldsymbol{u} \in \Omega}{\text{minimize}} \quad & K(\boldsymbol{X}(t_f)) + \int_{t_0}^{t_f} R(\boldsymbol{u}(t), t)\, \mathrm{d}t && \text{(3.31a)} \\ \text{subject to} \quad & \dot{\boldsymbol{X}}(t) = F(\boldsymbol{X}(t), \boldsymbol{u}(t), t)\,, && \text{(3.31b)} \\ & \boldsymbol{X}(t_0) = \boldsymbol{X}_0 \in \mathbb{C}^{d\times d}\,, \quad t \in [t_0, t_f] && \text{(3.31c)} \end{aligned}$$

where $K : \mathbb{C}^{d\times d} \to \mathbb{R}$ and $R : \mathcal{V} \times [t_0, t_f] \to \mathbb{R}$ describe the cost of a control strategy $\boldsymbol{u}$ over the time interval $[t_0, t_f]$. The term $K(\boldsymbol{X}(t_f))$ is known as a "terminal cost" and the integral of $R(\boldsymbol{u}(t), t)$ is known as a "running cost". Letting Φ denote the system's state transition map, the optimal control problem can be equivalently seen as finding a control law $\boldsymbol{u} \in \Omega$ that minimizes the functional

$$Q(\boldsymbol{u}) \triangleq K(\Phi(t_f; t_0, \boldsymbol{X}_0, \boldsymbol{u})) + \int_{t_0}^{t_f} R(\boldsymbol{u}(t), t)\, \mathrm{d}t\,. \tag{3.32}$$

For our purposes, we assume that such minimum exists; see [6] for a treatment of existence theory.

We will show that Q is Fréchet differentiable. Then, the necessary condition for optimality of a control law $\boldsymbol{u} \in \Omega$ is that the derivative of Q vanishes at $\boldsymbol{u}$. We summarize various hypotheses that will be needed in the sequel. Each theorem/lemma statement will identify which of these hypotheses are needed.

Hypotheses:

H1. the set of control inputs is $\mathcal{V} = \mathbb{R}^{N_c}$ for some $N_c \in \mathbb{N}$ and the set of allowed control laws is the Hilbert space $\Omega = L^2([t_0, t_f]; \mathbb{R}^{N_c})$ of square Lebesgue integrable functions from $[t_0, t_f]$ into $\mathbb{R}^{N_c}$;

H2. the generator F of the state transition map Φ takes the form (3.30);

H3. the operators F_0 and F_c are measurable and uniformly bounded with respect to time, i.e., there exists $B_0, B_c > 0$ such that[5]

$$\|F_0(\boldsymbol{X}, t)\|_1 \leqslant B_0 \|\boldsymbol{X}\|_1 \quad \text{and} \quad \|F_c(\boldsymbol{X}, \boldsymbol{u}, t)\|_1 \leqslant B_c \|\boldsymbol{X}\|_1 \|\boldsymbol{u}\| \tag{3.33}$$

for all $t \in [t_0, t_f]$, $\boldsymbol{X} \in \mathbb{C}^{d\times d}$, and $\boldsymbol{u} \in \mathcal{V}$;

H4. the functional $K : \mathbb{C}^{d\times d} \to \mathbb{R}$ is Fréchet differentiable (see Example 3.4);

[5]Here, $\|\boldsymbol{u}\|$ denotes the Euclidean norm on $\mathbb{R}^{N_c}$.

H5a. the functional R is Carathéodory and continuously differentiable with respect to its first argument; moreover, there exists $\ell_1 \in L^1([t_0, t_\mathrm{f}])$, $\ell_2 \in L^2([t_0, t_\mathrm{f}])$, and $C > 0$ such that

$$|R(\boldsymbol{u}, t)| \leqslant C\|\boldsymbol{u}\|^2 + \ell_1(t) \quad \text{and} \quad \|\nabla_{\boldsymbol{v}} R(\boldsymbol{v}, t)\big|_{\boldsymbol{v}=\boldsymbol{u}}\| \leqslant C\|\boldsymbol{u}\| + \ell_2(t) \tag{3.34}$$

for all $(\boldsymbol{u}, t) \in \mathcal{V} \times [t_0, t_\mathrm{f}]$;

H5b. the functional R is Carathéodory; moreover, there exists $\ell \in L^1([t_0, t_\mathrm{f}])$ and $C > 0$ such that $|R(\boldsymbol{v}, t)| \leqslant C\|\boldsymbol{v}\|^2 + \ell(t)$ for all $(\boldsymbol{v}, t) \in \mathcal{V} \times [t_0, t_\mathrm{f}]$.

Theorem 3.1 (Fréchet derivative of optimal quantum control functional)**.** Suppose that hypotheses H1-H4 and H5a are satisfied. Let $\boldsymbol{u} \in \Omega$ and let $\boldsymbol{X}(t) = \Phi(t; t_0, \boldsymbol{X}_0, \boldsymbol{u})$ for $t \in [t_0, t_\mathrm{f}]$. Then, the functional $Q(\cdot)$ defined in (3.32) is Fréchet differentiable with

$$D_{\boldsymbol{u}} Q(\boldsymbol{u})[\cdot] = \int_{t_0}^{t_\mathrm{f}} \mathrm{Re}\langle \boldsymbol{\Lambda}(t), F_\mathrm{c}(\boldsymbol{X}(t), \cdot, t)\rangle_\mathrm{HS} + \langle \nabla_{\boldsymbol{v}} R(\boldsymbol{v}, t)\big|_{\boldsymbol{v}=\boldsymbol{u}(t)}, \cdot\rangle \, \mathrm{d}t \tag{3.35}$$

where $\boldsymbol{\Lambda}(t) \in \mathbb{C}^{d\times d}$ for $t \in [t_0, t_\mathrm{f}]$ is the solution to the final value problem

$$\dot{\boldsymbol{\Lambda}}(t) = -F^\ddagger(\boldsymbol{\Lambda}(t), \boldsymbol{u}(t), t) \tag{3.36}$$

$$\boldsymbol{\Lambda}(t_\mathrm{f}) = \nabla_{\boldsymbol{X}} K(\boldsymbol{X})\big|_{\boldsymbol{X}=\Phi(t_\mathrm{f}; t_0, \boldsymbol{X}_0, \boldsymbol{u})} \tag{3.37}$$

and $F^\ddagger$ denotes the adjoint of F with respect to the Hilbert-Schmidt inner product, i.e., $\langle \boldsymbol{A}, F(\boldsymbol{B})\rangle_\mathrm{HS} = \langle F^\ddagger(\boldsymbol{A}), \boldsymbol{B}\rangle_\mathrm{HS}$ for $\boldsymbol{A}, \boldsymbol{B} \in \mathbb{C}^{d\times d}$. □

The proof of Theorem 3.1 is preceeded by preliminary lemmata. These lemmata also establish results that will be needed in the derivation of the minimum principle in Section 3.2.2. In the control literature $\boldsymbol{\Lambda}(t)$ is commonly referred to as a "co-state" [5, p. 115].

Lemma 3.1. Suppose hypotheses H1-H3 hold. Let $\boldsymbol{u}, \delta\boldsymbol{u} \in \Omega$ be two control laws and define the corresponding trajectories

$$\boldsymbol{X}_1(t) \triangleq \boldsymbol{X}_0 + \int_{t_0}^{t} F(\boldsymbol{X}_1(t), \boldsymbol{u}(t), t) \, \mathrm{d}t \tag{3.38a}$$

$$\boldsymbol{X}_2(t) \triangleq \boldsymbol{X}_0 + \int_{t_0}^{t} F(\boldsymbol{X}_2(t), \boldsymbol{u}(t) + \delta\boldsymbol{u}(t), t) \tag{3.38b}$$

for $t \in [t_0, t_\mathrm{f}]$. Define the state variation $\delta\boldsymbol{X}(t) \triangleq \boldsymbol{X}_2(t) - \boldsymbol{X}_1(t)$. Then, the following statements hold:

1. there exists $L > 0$ such that[6]

$$\max_{t\in[t_0,t_\mathrm{f}]} \|\delta\boldsymbol{X}(t)\|_\mathrm{HS} \leqslant L\|\delta\boldsymbol{u}\|_{L^1} ; \tag{3.39}$$

2. if a functional g satisfies $g(\boldsymbol{X}) = o(\|\boldsymbol{X}\|_\mathrm{HS})$ then $g(\delta\boldsymbol{X}(t)) = o(\|\delta\boldsymbol{u}\|_{L^1})$ and $g(\delta\boldsymbol{X}(t)) = o(\|\delta\boldsymbol{u}\|_{L^2})$ uniformly for $t \in [t_0, t_\mathrm{f}]$. □

[6]This maximum exists because $\delta\boldsymbol{X}(t)$ is continuous and $[t_0, t_\mathrm{f}]$ is compact.

Proof. The differential dynamics of the state variation $\delta\boldsymbol{X}(t)$ are

$$\begin{aligned}\delta\dot{\boldsymbol{X}}(t) &= F_0(\boldsymbol{X}_2(t),t) + F_{\mathrm{c}}(\boldsymbol{X}_2(t),\boldsymbol{u}(t)+\delta\boldsymbol{u}(t),t) - F_0(\boldsymbol{X}_1(t),t) - F_{\mathrm{c}}(\boldsymbol{X}_2(t),\boldsymbol{u}(t),t)\\ &= F_0(\delta\boldsymbol{X}(t),t) + F_{\mathrm{c}}(\delta\boldsymbol{X}(t),\boldsymbol{u}(t),t) + F_{\mathrm{c}}(\boldsymbol{X}_2(t),\delta\boldsymbol{u}(t),t)\\ &= F(\delta\boldsymbol{X}(t),\boldsymbol{u}(t),t) + F_{\mathrm{c}}(\boldsymbol{X}_2(t),\delta\boldsymbol{u}(t),t) \qquad (3.40)\end{aligned}$$

with initial condition $\delta\boldsymbol{X}(t_0)=\boldsymbol{0}$. Using the variation of constants formula [21, p. 40],

$$\delta\boldsymbol{X}(t) = \Phi(t;t_0,\boldsymbol{0},\boldsymbol{u}) + \int_{t_0}^{t}\Phi(t;s,F_{\mathrm{c}}(\boldsymbol{X}_2(s),\delta\boldsymbol{u}(s),s),\boldsymbol{u})\,\mathrm{d}s\,. \qquad (3.41)$$

By the universal contraction property of Lemma 2.1, for any $t\in[t_0,t_{\mathrm{f}}]$,[7]

$$\begin{aligned}\|\delta\boldsymbol{X}(t)\|_1 &\leqslant \int_{t_0}^{t}\|\Phi(t;s,F_{\mathrm{c}}(\boldsymbol{X}_2(s),\delta\boldsymbol{u}(s),s),\boldsymbol{u})\|_1\,\mathrm{d}s\\ &\leqslant \int_{t_0}^{t}\|F_{\mathrm{c}}(\boldsymbol{X}_2(s),\delta\boldsymbol{u}(s),s)\|_1\,\mathrm{d}s\\ &\leqslant \int_{t_0}^{t}B_{\mathrm{c}}\|\boldsymbol{X}_2(s)\|_1\,\|\delta\boldsymbol{u}(s)\|\,\mathrm{d}s\\ &\leqslant B_{\mathrm{c}}\|\boldsymbol{X}_2(t_0)\|_1\int_{t_0}^{t}\|\delta\boldsymbol{u}(s)\|\,\mathrm{d}s\\ &\leqslant B_{\mathrm{c}}\|\boldsymbol{X}_0\|_1\|\delta\boldsymbol{u}\|_{L^1}\,. \qquad (3.42)\end{aligned}$$

The first claim of the lemma follows by monotonicity of Schatten norms: $\|\delta\boldsymbol{X}(t)\|_{\mathrm{HS}}\leqslant\|\delta\boldsymbol{X}(t)\|_1\leqslant B_{\mathrm{c}}\|\boldsymbol{X}_0\|_1\|\delta\boldsymbol{u}\|_{L^1}$.

Now, consider the lemma's second claim and let $\epsilon>0$. Since $g(\boldsymbol{X})=o(\|\boldsymbol{X}\|_{\mathrm{HS}})$, there exists $\beta>0$ such that

$$\frac{|g(\boldsymbol{X})|}{\|\boldsymbol{X}\|_{\mathrm{HS}}}\leqslant\frac{\epsilon}{L}$$

if $0<\|\boldsymbol{X}\|_{\mathrm{HS}}\leqslant\beta$. So, taking $\|\delta\boldsymbol{u}\|_{L^1}\leqslant\beta/L$ implies via (3.39) that

$$\|\delta\boldsymbol{X}(t)\|_{\mathrm{HS}}\leqslant L\,\|\delta\boldsymbol{u}\|_{L^1}\leqslant\beta\,,\qquad\forall t\in[t_0,t_{\mathrm{f}}]$$

and

$$\frac{|g(\delta\boldsymbol{X}(t))|}{\|\delta\boldsymbol{u}\|_{L^1}}\leqslant\frac{L\,|g(\delta\boldsymbol{X}(t))|}{\|\delta\boldsymbol{X}(t)\|_{\mathrm{HS}}}\leqslant\epsilon$$

for $\|\delta\boldsymbol{X}(t)\|_{\mathrm{HS}}\neq0$. Since $\epsilon>0$ was arbitrary, this proves that $g(\delta\boldsymbol{X}(t))=o(\|\delta\boldsymbol{u}\|_{L^1})$. The proof that $g(\delta\boldsymbol{X}(t))=o(\|\delta\boldsymbol{u}\|_{L^2})$ follows the exact same procedure, given the Cauchy-Schwarz inequality $\|\delta\boldsymbol{u}\|_{L^1}\leqslant\sqrt{t_{\mathrm{f}}-t_0}\,\|\delta\boldsymbol{u}\|_{L^2}$. ⊠

Lemma 3.2. Suppose that hypotheses H1-H4 hold; let $\boldsymbol{u}$, $\boldsymbol{X}(t)$, and $\boldsymbol{\Lambda}(t)$ be as in the statement of Theorem 3.1. Then,

$$K(\Phi(t_{\mathrm{f}};t_0,\boldsymbol{X}_0,\boldsymbol{u}+\delta\boldsymbol{u}))-K(\Phi(t_{\mathrm{f}};t_0,\boldsymbol{X}_0,\boldsymbol{u})) = \int_{t_0}^{t_{\mathrm{f}}}\mathrm{Re}\langle\boldsymbol{\Lambda}(t),F_{\mathrm{c}}(\boldsymbol{X}(t),\cdot,t)\rangle_{\mathrm{HS}}\,\mathrm{d}t + E(\delta\boldsymbol{u}) \qquad (3.43)$$

for all $\delta\boldsymbol{u}\in\Omega$, where $E(\delta\boldsymbol{u})=o(\|\delta\boldsymbol{u}\|_{L^1})$ and $E(\delta\boldsymbol{u})=o(\|\delta\boldsymbol{u}\|_{L^2})$. □

[7] Note that this is the same bounding technique that we developed in the proof of Theorem 2.7.

Proof. Let the control laws and trajectories $\boldsymbol{u}$, $\delta\boldsymbol{u}$, $\boldsymbol{X}_1$, $\boldsymbol{X}_2$, and $\delta\boldsymbol{X}$ be as in Lemma 3.1. By assumption, $K : \mathbb{C}^{d\times d} \to \mathbb{R}$ is Fréchet differentiable. Since $\mathbb{C}^{d\times d}$ endowed with the inner product $\mathrm{Re}\langle\cdot,\cdot\rangle_{\mathrm{HS}}$ is a Hilbert space equipped with the Hilbert-Schmidt norm, this implies that (see Example 3.4)

$$K(\boldsymbol{B}) = K(\boldsymbol{A}) + \mathrm{Re}\langle\nabla_{\boldsymbol{A}} K(\boldsymbol{A}), \boldsymbol{B}-\boldsymbol{A}\rangle_{\mathrm{HS}} + o(\|\boldsymbol{B}-\boldsymbol{A}\|_{\mathrm{HS}}) \tag{3.44}$$

for $\boldsymbol{A}, \boldsymbol{B} \in \mathbb{C}^{d\times d}$. By Lemma 3.1,

$$\begin{aligned} K(\boldsymbol{X}_2(t_{\mathrm{f}})) - K(\boldsymbol{X}_1(t_{\mathrm{f}})) &= \mathrm{Re}\langle\nabla_{\boldsymbol{X}} K(\boldsymbol{X})\big|_{\boldsymbol{X}=\boldsymbol{X}_1(t_{\mathrm{f}})}, \delta\boldsymbol{X}(t_{\mathrm{f}})\rangle_{\mathrm{HS}} + o(\|\delta\boldsymbol{X}(t_{\mathrm{f}})\|_{\mathrm{HS}}) \\ &\triangleq \mathrm{Re}\langle\nabla_{\boldsymbol{X}} K(\boldsymbol{X})\big|_{\boldsymbol{X}=\boldsymbol{X}_1(t_{\mathrm{f}})}, \delta\boldsymbol{X}(t_{\mathrm{f}})\rangle_{\mathrm{HS}} + E_1(\delta\boldsymbol{u}) \end{aligned} \tag{3.45}$$

where $E_1(\delta\boldsymbol{u}) = o(\|\delta\boldsymbol{u}\|_{L^1})$ and $E_1(\delta\boldsymbol{u}) = o(\|\delta\boldsymbol{u}\|_{L^2})$.

Invoking the definitions of $\delta\boldsymbol{X}(t)$ and $\boldsymbol{\Lambda}(t)$ then integrating by parts gives

$$\begin{aligned} \langle \tfrac{\mathrm{d}}{\mathrm{d}\boldsymbol{X}} K(\boldsymbol{X})\big|_{\boldsymbol{X}=\boldsymbol{X}_1(t_{\mathrm{f}})}, \delta\boldsymbol{X}(t_{\mathrm{f}})\rangle_{\mathrm{HS}} &= \langle\boldsymbol{\Lambda}(t_{\mathrm{f}}), \delta\boldsymbol{X}(t_{\mathrm{f}})\rangle_{\mathrm{HS}} - \langle\boldsymbol{\Lambda}(t_0), \delta\boldsymbol{X}(t_0)\rangle_{\mathrm{HS}} \\ &= \int_{t_0}^{t_{\mathrm{f}}} \frac{\mathrm{d}}{\mathrm{d}t}\langle\boldsymbol{\Lambda}(t), \delta\boldsymbol{X}(t)\rangle_{\mathrm{HS}}\,\mathrm{d}t \\ &= \int_{t_0}^{t_{\mathrm{f}}} \langle\dot{\boldsymbol{\Lambda}}(t), \delta\boldsymbol{X}(t)\rangle_{\mathrm{HS}} + \langle\boldsymbol{\Lambda}(t), \delta\dot{\boldsymbol{X}}(t)\rangle_{\mathrm{HS}}\,\mathrm{d}t\,. \end{aligned} \tag{3.46}$$

The dynamics (3.40) of the variation $\delta\boldsymbol{X}(t)$ are equivalently written

$$\delta\dot{\boldsymbol{X}}(t) = F(\delta\boldsymbol{X}(t), \boldsymbol{u}(t), t) + F_{\mathrm{c}}(\delta\boldsymbol{X}(t), \delta\boldsymbol{u}(t), t) + F_{\mathrm{c}}(\boldsymbol{X}_1(t), \delta\boldsymbol{u}(t), t)\,. \tag{3.47}$$

Shortly, this expression will be substituted into (3.46). However, first notice the following. By Lemma 3.1, there exists constants $C_1, C_2, C_3 > 0$ such that

$$\begin{aligned} |E_2(\delta\boldsymbol{u})| &\triangleq \Big| \int_{t_0}^{t_{\mathrm{f}}} \mathrm{Re}\langle\boldsymbol{\Lambda}(t), F_{\mathrm{c}}(\delta\boldsymbol{X}(t), \delta\boldsymbol{u}(t), t)\rangle_{\mathrm{HS}}\,\mathrm{d}t \Big| \\ &\leqslant \int_{t_0}^{t_{\mathrm{f}}} \|\boldsymbol{\Lambda}(t)\|_{\mathrm{HS}} \|F_{\mathrm{c}}(\delta\boldsymbol{X}(t), \delta\boldsymbol{u}(t), t)\|_{\mathrm{HS}}\,\mathrm{d}t \\ &\leqslant C_1 \int_{t_0}^{t_{\mathrm{f}}} \|F_{\mathrm{c}}(\delta\boldsymbol{X}(t), \delta\boldsymbol{u}(t), t)\|_{\mathrm{HS}}\,\mathrm{d}t \\ &\leqslant C_2 \int_{t_0}^{t_{\mathrm{f}}} \|\delta\boldsymbol{X}(t)\|_{\mathrm{HS}}\, \|\delta\boldsymbol{u}(t)\|\,\mathrm{d}t \\ &\leqslant C_2 \Big[\max_{t\in[t_0,t_{\mathrm{f}}]} \|\delta\boldsymbol{X}(t)\|_{\mathrm{HS}}\Big] \int_{t_0}^{t_{\mathrm{f}}} \|\delta\boldsymbol{u}(t)\|\,\mathrm{d}t \\ &\leqslant C_3 \|\delta\boldsymbol{u}\|_{L^1}^2\,. \end{aligned} \tag{3.48}$$

This proves $E_2(\delta\boldsymbol{u}) = o(\|\delta\boldsymbol{u}\|_{L^1})$ and $E_2(\delta\boldsymbol{u}) = o(\|\delta\boldsymbol{u}\|_{L^2})$.

Let $E(\delta \boldsymbol{u}) \triangleq E_1(\delta \boldsymbol{u}) + E_2(\delta \boldsymbol{u})$. Using (3.36) and (3.45)-(3.47), we conclude the proof:

$$\begin{aligned} K(\boldsymbol{X}_2(t_\mathrm{f})) - K(\boldsymbol{X}_1(t_\mathrm{f})) &= \int_{t_0}^{t_\mathrm{f}} \Big[\mathrm{Re}\langle \dot{\boldsymbol{\Lambda}}(t), \delta \boldsymbol{X}(t) \rangle_{\mathrm{HS}} + \mathrm{Re}\langle \boldsymbol{\Lambda}(t), F(\delta \boldsymbol{X}(t), \boldsymbol{u}(t), t) \rangle_{\mathrm{HS}} \\ &\qquad + \mathrm{Re}\langle \boldsymbol{\Lambda}(t), F_\mathrm{c}(\boldsymbol{X}_1(t), \delta \boldsymbol{u}(t), t) \rangle_{\mathrm{HS}} \Big] \, \mathrm{d}t + E(\delta \boldsymbol{u}) \\ &= \int_{t_0}^{t_\mathrm{f}} \Big[\mathrm{Re}\langle \dot{\boldsymbol{\Lambda}}(t) + F^\ddagger(\boldsymbol{\Lambda}(t), \boldsymbol{u}(t), t), \delta \boldsymbol{X}(t) \rangle_{\mathrm{HS}} \\ &\qquad + \mathrm{Re}\langle \boldsymbol{\Lambda}(t), F_\mathrm{c}(\boldsymbol{X}_1(t), \delta \boldsymbol{u}(t), t) \rangle_{\mathrm{HS}} \Big] \, \mathrm{d}t + E(\delta \boldsymbol{u}) \\ &= \int_{t_0}^{t_\mathrm{f}} \mathrm{Re}\langle \boldsymbol{\Lambda}(t), F_\mathrm{c}(\boldsymbol{X}_1(t), \delta \boldsymbol{u}(t), t) \rangle_{\mathrm{HS}} \; + \; E(\delta \boldsymbol{u}) \, . \end{aligned} \tag{3.49}$$

⊠

Proof of Theorem 3.1. The Fréchet derivative is linear [89]. Hence, the theorem follows from Lemma 3.2 and Example 3.3. ⊠

Remark 3.1. It is insightful to revisit the proofs of Lemmas 3.1 and 3.2 to observe where universal contractivity and bilinearity played a role. □

3.2.1 Gradient based optimization

The "gradient method" [5, pp. 144-145] is the most common approach to numerically determining local optima of the optimal control problem in (3.31). Since the gradient method will be used in Chapter 4, we provide a brief summary of it here. Starting from an initial control $\boldsymbol{u}_0 \in \Omega = L^2([t_0, t_\mathrm{f}]; \mathbb{R}^{N_\mathrm{c}})$, the gradient method produces a sequence of iteratively refined controls according to the following procedure: for $n = 1, 2, \ldots$

1. compute $\boldsymbol{X}_n(t)$ by numerically solving the initial value problem

$$\dot{\boldsymbol{X}}_n(t) = F(\boldsymbol{X}_n(t), \boldsymbol{u}_n(t), t) \tag{3.50a}$$
$$\boldsymbol{X}_n(t_0) = \boldsymbol{X}_0 \tag{3.50b}$$

2. compute $\boldsymbol{\Lambda}_n(t)$ by numerically solving the differential equation

$$\dot{\boldsymbol{\Lambda}}_n(t) = -F^\ddagger(\boldsymbol{\Lambda}_n(t), \boldsymbol{u}(t), t) \tag{3.51a}$$
$$\boldsymbol{\Lambda}_n(t_\mathrm{f}) = \nabla_{\boldsymbol{X}(t_\mathrm{f})} K(\boldsymbol{X}(t_\mathrm{f})) \tag{3.51b}$$

3. update the current control law according to

$$\boldsymbol{u}_{n+1} = \boldsymbol{u}_n - \sigma_n \nabla_{\boldsymbol{u}_n} Q(\boldsymbol{u}_n) \tag{3.52}$$

   where $\sigma_n > 0$ and $\nabla_{\boldsymbol{u}_n} Q(\boldsymbol{u}_n)$ is the Fréchet derivative in Theorem 3.1, which is completely determined by $\boldsymbol{X}_n(\cdot)$ and $\boldsymbol{\Lambda}_n(\cdot)$ that were computed in steps 1 and 2.

In practice, the stepsize σ_n is chosen adaptively (e.g., according to the Armijo criterion [93, p. 29]) to ensure good convergence properties.

3.2.2 Minimum principle for quantum systems

The Hamiltonian functional[8] $h_{\mathrm{c}} : \mathbb{C}^{d\times d} \times \mathbb{C}^{d\times d} \times \mathcal{V} \times [t_0, t_{\mathrm{f}}] \to \mathbb{R}$ associated with the optimal control problem is defined to be

$$h_{\mathrm{c}}(\boldsymbol{X}, \boldsymbol{\Lambda}, \boldsymbol{u}, t) \triangleq \mathrm{Re}\langle \boldsymbol{\Lambda}, F_{\mathrm{c}}(\boldsymbol{X}, \boldsymbol{u}, t)\rangle_{\mathrm{HS}} + R(\boldsymbol{u}, t)\,. \tag{3.53}$$

The following theorem provides a minimum principle for quantum systems.

Theorem 3.2 (Quantum minimum principle)**.** Suppose that the set $\mathcal{V}$ of control inputs is a compact subset of $\mathbb{R}^{N_{\mathrm{c}}}$ and that the set of control laws Ω is the set of measurable functions from T into $\mathcal{V}$. Further, suppose that hypotheses H2-H4 and H5b hold. If $\boldsymbol{u} \in \Omega$ is an optimal control, then

$$\boldsymbol{u}(t) \;\in\; \arg\min_{\boldsymbol{v}\in\mathcal{V}} \; h_{\mathrm{c}}(\boldsymbol{X}(t), \boldsymbol{\Lambda}(t), \boldsymbol{v}, t)\,, \qquad \tilde{\forall} t \in [t_0, t_{\mathrm{f}}]\,. \tag{3.54}$$

□

Proof. We will first show that, for any $\boldsymbol{v} \in \mathcal{V}$,

$$h_{\mathrm{c}}(\boldsymbol{X}(t), \boldsymbol{\Lambda}(t), \boldsymbol{u}(t), t) \leqslant h_{\mathrm{c}}(\boldsymbol{X}(t), \boldsymbol{\Lambda}(t), \boldsymbol{v}, t)\,, \qquad \tilde{\forall} t \in [t_0, t_{\mathrm{f}}] \tag{3.55}$$

and then use a separability argument to obtain (3.54). Take $\boldsymbol{v} \in \mathcal{V}$ to be an arbitrary input. The function of time

$$\begin{aligned} &h_{\mathrm{c}}(\boldsymbol{X}(t), \boldsymbol{\Lambda}(t), \boldsymbol{v}, t) - h_{\mathrm{c}}(\boldsymbol{X}(t), \boldsymbol{\Lambda}(t), \boldsymbol{u}(t), t) \\ &= \mathrm{Re}\langle \boldsymbol{\Lambda}(t), F_{\mathrm{c}}(\boldsymbol{X}(t), \boldsymbol{v}, t)\rangle_{\mathrm{HS}} + R(\boldsymbol{v}, t) - \mathrm{Re}\langle \boldsymbol{\Lambda}(t), F_{\mathrm{c}}(\boldsymbol{X}(t), \boldsymbol{u}(t), t)\rangle_{\mathrm{HS}} - R(\boldsymbol{u}(t), t) \end{aligned} \tag{3.56}$$

is measurable and Lebesgue integrable. Indeed, F_{c} is continuous in the control variable (it is bounded and bilinear); $\boldsymbol{u}(t)$ is measurable with respect to time; $\boldsymbol{X}(t)$ and $\boldsymbol{\Lambda}(t)$ are absolutely continuous functions of time; and R is Carathéodory. Hence, the measurability claim follows. Integrability follows from the fact that[9]

$$\begin{aligned} &\int_{t_0}^{t_{\mathrm{f}}} |h_{\mathrm{c}}(\boldsymbol{X}(t), \boldsymbol{\Lambda}(t), \boldsymbol{v}, t) - h_{\mathrm{c}}(\boldsymbol{X}(t), \boldsymbol{\Lambda}(t), \boldsymbol{u}(t), t)|\,\mathrm{d}t \\ &\quad= \int_{t_0}^{t_{\mathrm{f}}} |\mathrm{Re}\langle \boldsymbol{\Lambda}(t), F_{\mathrm{c}}(\boldsymbol{X}(t), \boldsymbol{v} - \boldsymbol{u}(t), t)\rangle_{\mathrm{HS}} + R(\boldsymbol{v}, t) - R(\boldsymbol{u}(t), t)|\,\mathrm{d}t \\ &\quad\leqslant \int_{t_0}^{t_{\mathrm{f}}} \|\boldsymbol{\Lambda}(t)\|_{\mathrm{HS}}\, \|F_{\mathrm{c}}(\boldsymbol{X}(t), \boldsymbol{v} - \boldsymbol{u}(t), t)\|_{\mathrm{HS}} + 2\ell(t) + C(\|\boldsymbol{v}\|^2 + \|\boldsymbol{u}(t)\|^2)\,\mathrm{d}t \\ &\quad\leqslant \int_{t_0}^{t_{\mathrm{f}}} B_{\mathrm{c}}\, \|\boldsymbol{\Lambda}(t)\|_{\mathrm{HS}}\, \|\boldsymbol{X}(t)\|_1\, \|\boldsymbol{v} - \boldsymbol{u}(t)\| + 2\ell(t) + C(\|\boldsymbol{v}\|^2 + \|\boldsymbol{u}(t)\|^2)\,\mathrm{d}t \\ &\quad< \infty \end{aligned} \tag{3.57}$$

since $\boldsymbol{X}(t), \boldsymbol{\Lambda}(t)$ are bounded on $[t_0, t_{\mathrm{f}}]$ (continuous functions on a compact interval are bounded), $\ell \in L^1([t_0, t_{\mathrm{f}}])$, and $\boldsymbol{u} \in \Omega$.

[8] In classical control theory, h_{c} is given the name "Hamiltonian". As in [1], we use the name "Hamiltonian functional" to distinguish h_{c} from the quantum system's Hamiltonian $\boldsymbol{H}$.

[9] Monotonicity of Schatten norms is used here, i.e., $\|\boldsymbol{X}\|_{\mathrm{HS}} \leqslant \|\boldsymbol{X}\|_1$ for all $\boldsymbol{X} \in \mathbb{C}^{d\times d}$.

The Lebesgue differentiation theorem (Theorem A.3) provides the relation

$$\begin{aligned} &h_{\mathrm{c}}(\boldsymbol{X}(t), \boldsymbol{\Lambda}(t), \boldsymbol{v}, t) - h_{\mathrm{c}}(\boldsymbol{X}(t), \boldsymbol{\Lambda}(t), \boldsymbol{u}(t), t) \\ &\qquad = \lim_{|B|\to 0} \frac{1}{|B|} \int_B h_{\mathrm{c}}(\boldsymbol{X}(s), \boldsymbol{\Lambda}(s), \boldsymbol{v}, s) - h_{\mathrm{c}}(\boldsymbol{X}(s), \boldsymbol{\Lambda}(s), \boldsymbol{u}(s), s)\, \mathrm{d}s \end{aligned} \tag{3.58}$$

for almost all $t \in (t_0, t_{\mathrm{f}})$, where B is a closed interval containing t. Let $T_{h_{\mathrm{c}}} \subset [t_0, t_{\mathrm{f}}]$ denote the (possibly empty) set of time instants where (3.58) does *not* hold. Another application of the Lebesgue differentiation theorem implies that

$$\lim_{|B|\to 0} \frac{1}{|B|} \int_B \|\boldsymbol{u}(s) - \boldsymbol{u}(t)\|\, \mathrm{d}s = 0 \tag{3.59}$$

for almost all $t \in [t_0, t_{\mathrm{f}}]$. Let $T_{\boldsymbol{u}} \subset [t_0, t_{\mathrm{f}}]$ denote the set of time instants that (3.59) equality does not hold, and let $\mu(\cdot)$ denote the Lebesgue measure on $[t_0, t_{\mathrm{f}}]$. By the union bound $\mu(T_{h_{\mathrm{c}}} \cup T_{\boldsymbol{u}}) \leqslant \mu(T_{h_{\mathrm{c}}}) + \mu(T_{\boldsymbol{u}}) = 0$, so (3.58) and (3.59) hold simultaneously for almost all $t \in [t_0, t_{\mathrm{f}}]$.

Let $t \in [t_0, t_{\mathrm{f}}] \setminus (T_{h_{\mathrm{c}}} \cup T_{\boldsymbol{u}})$. For each closed interval $B \subseteq [t_0, t_{\mathrm{f}}]$ containing t, specify a perturbation $\delta\boldsymbol{u}(s) \triangleq [\boldsymbol{v} - \boldsymbol{u}(s)]\, \mathbb{1}_B(s)$ for $s \in [t_0, t_{\mathrm{f}}]$. First, note that

$$\lim_{|B|\to 0} \frac{1}{|B|} \int_B \|\delta\boldsymbol{u}(s)\|\, \mathrm{d}s = \|\delta\boldsymbol{u}(t)\| = \|\boldsymbol{v} - \boldsymbol{u}(t)\| \tag{3.60}$$

since

$$\begin{aligned} \Big| \lim_{|B|\to 0} \frac{1}{|B|} \int_B \|\delta\boldsymbol{u}(s)\|\, \mathrm{d}s - \|\delta\boldsymbol{u}(t)\| \Big| &\leqslant \lim_{|B|\to 0} \frac{1}{|B|} \int_B \|\delta\boldsymbol{u}(s) - \delta\boldsymbol{u}(t)\|\, \mathrm{d}s \\ &\leqslant \lim_{|B|\to 0} \frac{1}{|B|} \int_B \|\boldsymbol{u}(s) - \boldsymbol{u}(t)\|\, \mathrm{d}s = 0\,. \end{aligned} \tag{3.61}$$

By optimality of $\boldsymbol{u}$ and by Lemma 3.2,

$$\begin{aligned} 0 &\leqslant Q(\boldsymbol{u} + \delta\boldsymbol{u}) - Q(\boldsymbol{u}) \\ &= \int_B \mathrm{Re}\langle \boldsymbol{\Lambda}(t), F_{\mathrm{c}}(\boldsymbol{X}(s), \boldsymbol{v} - \boldsymbol{u}(s), s)\rangle_{\mathrm{HS}}\, \mathrm{d}s + E(\delta\boldsymbol{u}) \\ &\qquad\qquad + \int_{t_0}^{t_{\mathrm{f}}} R(\boldsymbol{u}(s) + \delta\boldsymbol{u}(s), s) - R(\boldsymbol{u}(s), s)\, \mathrm{d}s \\ &= \int_B \mathrm{Re}\langle \boldsymbol{\Lambda}(t), F_{\mathrm{c}}(\boldsymbol{X}(s), \boldsymbol{v} - \boldsymbol{u}(s), s)\rangle_{\mathrm{HS}}\, \mathrm{d}s + E(\delta\boldsymbol{u}) + \int_B R(\boldsymbol{v}, s) - R(\boldsymbol{u}(s), s)\, \mathrm{d}s \\ &= \int_B h_{\mathrm{c}}(\boldsymbol{X}(s), \boldsymbol{\Lambda}(s), \boldsymbol{v}, s) - h_{\mathrm{c}}(\boldsymbol{X}(t), \boldsymbol{\Lambda}(s), \boldsymbol{u}(s), s)\, \mathrm{d}s + E(\delta\boldsymbol{u}) \end{aligned} \tag{3.62}$$

where $E(\delta\boldsymbol{u}) = o(\|\delta\boldsymbol{u}\|_{L^1})$. Suppose $\boldsymbol{v} \neq \boldsymbol{u}(t)$. Because $E(\delta\boldsymbol{u}) = o(\|\delta\boldsymbol{u}\|_{L^1})$, for each $\epsilon > 0$ there exists $\delta > 0$ such that

$$|E(\delta\boldsymbol{u})| \leqslant \frac{\epsilon}{\|\boldsymbol{v} - \boldsymbol{u}(t)\|} \|\delta\boldsymbol{u}\|_{L^1} \tag{3.63}$$

if $\|\delta \boldsymbol{u}\|_{L^1} \leqslant \delta$. Since $\|\delta \boldsymbol{u}\|_{L^1} \to 0$ as $|B| \to 0$, there exists $\delta' > 0$ such that $\|\delta \boldsymbol{u}\|_{L^1} \leqslant \delta$ if $|B| \leqslant \delta'$. By (3.60),

$$\lim_{\substack{|B|\to 0 \\ |B| \in (0,\delta']}} \frac{|E(\delta \boldsymbol{u})|}{|B|} \leqslant \frac{\epsilon}{\|\boldsymbol{v} - \boldsymbol{u}(t)\|} \lim_{\substack{|B|\to 0 \\ |B| \in (0,\delta']}} \frac{1}{|B|} \|\delta \boldsymbol{u}\|_{L^1} = \epsilon. \tag{3.64}$$

If $\boldsymbol{v} = \boldsymbol{u}(t)$, similar argumentation shows that $\lim_{|B|\to 0} |E(\delta \boldsymbol{u})|/|B| \leqslant \epsilon$. Since $\epsilon > 0$ was arbitrary, it follows that $\lim_{|B|\to 0} |E(\delta \boldsymbol{u})|/|B| = 0$. Dividing both sides of (3.62) by $|B|$ and taking the limit as $|B| \to 0$ yields

$$h_{\mathrm{c}}(\boldsymbol{X}(t), \boldsymbol{\Lambda}(t), \boldsymbol{u}(t), t) \leqslant h_{\mathrm{c}}(\boldsymbol{X}(t), \boldsymbol{\Lambda}(t), \boldsymbol{v}, t)\,. \tag{3.65}$$

This proves (3.55).

It will now be shown that (3.55) implies (3.54). The space $\mathcal{V}$ is separable, so there exists a countable dense subset $\{\boldsymbol{v}_n : n \in \mathbb{N}\} \subset \mathcal{V}$. Define the (possibly empty) sets

$$T_n \triangleq \Big\{ t \in [t_0, t_{\mathrm{f}}] : h_{\mathrm{c}}(\boldsymbol{X}(t), \boldsymbol{\Lambda}(t), \boldsymbol{u}(t), t) > h_{\mathrm{c}}(\boldsymbol{X}(t), \boldsymbol{\Lambda}(t), \boldsymbol{v}_n, t) \Big\}, \quad n \in \mathbb{N} \tag{3.66a}$$

$$T \triangleq \Big\{ t \in [t_0, t_{\mathrm{f}}] : \exists n \in \mathbb{N} \text{ s.t. } h_{\mathrm{c}}(\boldsymbol{X}(t), \boldsymbol{\Lambda}(t), \boldsymbol{u}(t), t) > h_{\mathrm{c}}(\boldsymbol{X}(t), \boldsymbol{\Lambda}(t), \boldsymbol{v}_n, t) \Big\}. \tag{3.66b}$$

By (3.55), each of the sets T_n has Lebesgue measure zero. By the union bound, so does T:

$$\mu(T) = \mu\Big(\bigcup_{n\in\mathbb{N}} T_n \Big) \leqslant \sum_{n\in\mathbb{N}} \mu(T_n) = 0\,. \tag{3.67}$$

This proves that

$$h_{\mathrm{c}}(\boldsymbol{X}(t), \boldsymbol{\Lambda}(t), \boldsymbol{u}(t), t) \leqslant \inf_{n\in\mathbb{N}} h_{\mathrm{c}}(\boldsymbol{X}(t), \boldsymbol{\Lambda}(t), \boldsymbol{v}_n, t)\,, \qquad \tilde{\forall} t \in [t_0, t_{\mathrm{f}}]\,. \tag{3.68}$$

Lastly, since R is Carathéodory, $R(\boldsymbol{v}, t)$ is continuous in $\boldsymbol{v}$ for almost all $t \in [t_0, t_{\mathrm{f}}]$. Since $\mathrm{Re}\langle \boldsymbol{\Lambda}(t), F_{\mathrm{c}}(\boldsymbol{X}(t), \boldsymbol{v}, t)\rangle$ is linear in $\boldsymbol{v}$, it follows that the Hamiltonian functional $h_{\mathrm{c}}(\boldsymbol{X}(t), \boldsymbol{\Lambda}(t), \boldsymbol{v}, t)$ is continuous in $\boldsymbol{v}$ for almost all $t \in [t_0, t_{\mathrm{f}}]$. By density of $\{\boldsymbol{v}_n : n \in \mathbb{N}\}$ in $\mathcal{V}$,

$$\inf_{n\in\mathbb{N}} h_{\mathrm{c}}(\boldsymbol{X}(t), \boldsymbol{\Lambda}(t), \boldsymbol{v}_n, t) = \min_{\boldsymbol{v}\in\mathcal{V}} h_{\mathrm{c}}(\boldsymbol{X}(t), \boldsymbol{\Lambda}(t), \boldsymbol{v}, t)\,, \qquad \tilde{\forall} t \in [t_0, t_{\mathrm{f}}]\,. \tag{3.69}$$

Combining (3.68) and (3.69) gives

$$h_{\mathrm{c}}(\boldsymbol{X}(t), \boldsymbol{\Lambda}(t), \boldsymbol{u}(t), t) \leqslant \min_{\boldsymbol{v}\in\mathcal{V}} h_{\mathrm{c}}(\boldsymbol{X}(t), \boldsymbol{\Lambda}(t), \boldsymbol{v}, t)\,, \qquad \tilde{\forall} t \in [t_0, t_{\mathrm{f}}]\,. \tag{3.70}$$

Since $\boldsymbol{u}(t) \in \mathcal{V}$ for all $t \in [t_0, t_{\mathrm{f}}]$, the proof is complete. ⊠

Chapter 4

Controlled Fisher information for quantum estimation

This chapter uncovers how quantum control is central to quantum parameter estimation. A new statistical information measure referred to as *controlled Fisher information (CFI)* is defined, which connects estimation performance to the structure of a particular quantum control system. The chapter first reviews the classical and quantum Fisher information (QFI), then the CFI is introduced. As in prior chapters, let $\mathcal{H} = \mathbb{C}^d$ with $d \geqslant 2$.

4.1 Classical and quantum Fisher information

The goal of parameter estimation is to infer an unknown parameter θ, belonging to a parameter space Θ, based on observations of a random variable x whose distribution depends on θ. The parameter space Θ is taken to be a compact subset of $\mathbb{R}$. The distribution of the random variable x is denoted $p_\theta(\cdot)$, which belongs to a parameterized family

$$\mathcal{P} = \left\{ p_\theta(\cdot) \,:\, \theta \in \Theta \right\} \tag{4.1}$$

that has a common support across the parameter space Θ. Here, $p_\theta(\cdot)$ is taken to be continuously differentiable with respect to θ on some open set containing Θ. This family is also referred to as a *statistical model*. The random variable x contains information about θ, which can be quantified by the Fisher information.

Definition 4.1 (Fisher information)**.** The *Fisher information* about $\theta \in \Theta$ contained in the the random variable $\mathsf{x} \sim p_\theta(\cdot)$ is

$$J_{\mathsf{x}}(\theta) \;=\; \mathbb{E}\{\jmath(\mathsf{x};\theta)\} \tag{4.2}$$

where $\jmath(\boldsymbol{x};\theta) = [\frac{\mathrm{d}}{\mathrm{d}\theta} \ln p_\theta(\boldsymbol{x})]^2$ for each $\boldsymbol{x}$ in the support of $p_\theta(\cdot)$. □

To infer the unknown parameter θ, one can form an estimator $\hat{\theta}(\mathsf{x})$. An estimator is any function of x that is not also a function of the unknown parameter θ [94, pp. 211, 311]. An estimator is said to be *unbiased* if its expected value is the unknown parameter, i.e.,

$\mathbb{E}\{\hat{\theta}(\mathsf{x})\} = \theta$. A central result in classical statistics is that the variance of any unbiased estimator is lower bounded by the inverse of the Fisher information:

$$\mathbb{V}\{\hat{\theta}(\mathsf{x})\} \geqslant \frac{1}{J_{\mathsf{x}}(\theta)}\,. \tag{4.3}$$

This result, known as the information inequality [95, p. 126], foretells a deep connection between estimation error (concerning operational performance) and Fisher information (concerning the statistical structure of the parameterized family $\mathcal{P}$). If equality holds in (4.3), $\hat{\theta}(\mathsf{x})$ is said to be an *efficient* estimator [96, p. 66].

Random vectors of the form $\mathsf{x}_{1:n} \triangleq [\mathsf{x}_1\ \mathsf{x}_2\ \ldots\ \mathsf{x}_n]$ often arise, where $\mathsf{x}_1, \mathsf{x}_2, \ldots, \mathsf{x}_n$ are independent random variables. In such case, the Fisher information is additive: $J_{\mathsf{x}_{1:n}}(\theta) = \sum_{i=1}^{n} J_{\mathsf{x}_i}(\theta)$. If the x_i's are independent and identically distributed (IID) then $J_{\mathsf{x}_{1:n}}(\theta) = nJ_{\mathsf{x}_1}(\theta)$ [95, p. 129]. In the IID setting, the connection between estimation performance and Fisher information is sealed by the fact that, under appropriate regularity conditions [94, p. 516], there exist asymptotically efficient estimators. That is, as n grows large, there exist estimators whose variance scales as $1/(nJ_{\mathsf{x}_1}(\theta))$. Such estimators may include maximum likelihood estimators and one-step estimators [94], [95, prob. 4.4.12].

So far, the discussion has only concerned classical statistics. Suppose now that the unknown parameter θ is encoded into a density matrix $\boldsymbol{\Xi}_\theta \in \mathcal{D}(\mathcal{H})$ describing the state of a quantum system. Here, $\boldsymbol{\Xi}_\theta$ is taken to be *twice* continuously differentiable on some open set containing the parameter space Θ.[1] Suppose further that the quantum state is measured using a positive operator-valued measure (POVM) $\mathcal{M}$, which results in a random measurement outcome y distributed according to Born's rule. This constitutes a POVM-dependent parameterized family of distributions:

$$\mathcal{P}(\mathcal{M}) = \left\{\, p_{\mathsf{y}}(\cdot\,; \theta, \mathcal{M}) \;:\; \theta \in \Theta \,\right\}. \tag{4.4}$$

For a given $\mathcal{M}$, the performance limit of estimating θ is characterized by the POVM-dependent Fisher information $J_{\mathsf{y}}(\theta; \mathcal{M})$. If there is freedom in selecting $\mathcal{M}$, more can accomplished.

Definition 4.2 (Quantum Fisher information [10])**.** The *quantum Fisher information* about $\theta \in \Theta$ contained in the density matrix $\boldsymbol{\Xi}_\theta$ is[2]

$$J^{\mathrm{q}}_{\boldsymbol{\Xi}_\theta}(\theta) = \max_{\mathcal{M}} J_{\mathsf{y}}(\theta; \mathcal{M}) \tag{4.5}$$

where the maximization is performed over all possible POVMs. □

Definition 4.2 implicitly assumes enough regularity on $\boldsymbol{\Xi}_\theta$'s parameter dependence to have a well-defined Fisher information under any POVM system $\mathcal{M}$.[3] A sufficient regularity condition is that $\boldsymbol{\Xi}_\theta$ is positive *definite* and depends on θ smoothly [8]. This condition allows the QFI to be calculated as

$$J^{\mathrm{q}}_{\boldsymbol{\Xi}_\theta}(\theta) = \mathrm{tr}\{\boldsymbol{\Xi}_\theta \boldsymbol{L}_\theta^2\} \tag{4.6}$$

[1] Twice continuous differentiability will ensure that the CFI, to be defined later, is Lipschitz continuous.

[2] Note that the subscript of the QFI is a density matrix while the subscript of the classical Fisher information is a random variable.

[3] Recall that the Fisher information is defined on a parameterized family of probability distributions with common support across the parameter space.

where $\boldsymbol{L}_\theta \in \mathbb{C}^{d\times d}$ is the symmetric logarithmic derivative (SLD), which is defined as the solution to the Lyapunov equation

$$2\frac{\mathrm{d}}{\mathrm{d}\theta}\boldsymbol{\Xi}_\theta = \boldsymbol{\Xi}_\theta \boldsymbol{L}_\theta + \boldsymbol{L}_\theta \boldsymbol{\Xi}_\theta \,. \tag{4.7}$$

A particular measurement achieving the QFI is a projective measurement system onto the eigenstates of $\boldsymbol{L}_\theta$ [97].

The QFI describes the *ultimate* limit of estimating θ. Namely, assuming that all possible quantum measurements are at one's disposal, the QFI is the maximum information about θ that can be extracted from $\boldsymbol{\Xi}_\theta$. However, the *fundamental* limit of a particular quantum technology or hardware setup may differ from the ultimate limit. In particular, if the quantum measurements that one can realize are constrained, then it may be impossible to achieve the QFI. For the remainder of this chapter, we aim to investigate the fundamental limits of quantum estimation with control-enabled measurements.

4.2 Controlled Fisher information

We are interested in a quantum system Σ, which is (by definition) equipped with a fixed measurement apparatus. Take the measurement apparatus to be represented by a POVM $\mathcal{M} = \{\boldsymbol{M}_k \,:\, k \in \mathbb{N}_1^{N_\mathrm{m}}\}$, $N_\mathrm{m} \in \mathbb{N}$. The system is initially in state $\boldsymbol{\Xi}_\theta$ at time t_0, where $\boldsymbol{\Xi}_\theta$ is the parameterized state from the prior section. The state evolves under the effect of a control law $\boldsymbol{u} \in \Omega$, and the resulting state is measured at a time $t_\mathrm{f} \in T$ to infer θ. The random measurement outcome is $\psi(\Phi(t_\mathrm{f}; t_0, \boldsymbol{\Xi}_\theta, \boldsymbol{u}))$. Notice that, even though the system Σ has a fixed measurement apparatus, the distribution of $\psi(\Phi(t_\mathrm{f}; t_0, \boldsymbol{\Xi}_\theta, \boldsymbol{u}))$ depends on the control law $\boldsymbol{u}$ and the measurement time t_f. The goal is thus to select $\boldsymbol{u}$ and t_f so as to maximize the information $J_{\psi(\Phi(t_\mathrm{f};t_0,\boldsymbol{\Xi}_\theta,\boldsymbol{u}))}(\theta)$ about θ contained in $\psi(\Phi(t_\mathrm{f}; t_0, \boldsymbol{\Xi}_\theta, \boldsymbol{u}))$. In order for this information to be well-defined, the following well-posedness condition is assumed.

Assumption 4.1. The random variable $\psi(\Phi(t_\mathrm{f}; t_0, \boldsymbol{\Xi}_\theta, \boldsymbol{u}))$ has common support for all $\theta \in \Theta$, $\boldsymbol{u} \in \Omega$, and $t_\mathrm{f} \in T$. □

In cases of practical interest, all measurement outcomes will have non-zero probability, in which case Assumption 4.1 is satisfied. The following lemma gives sufficient mathematical conditions for Assumption 4.1 to hold.

Lemma 4.1. Either of the following conditions is sufficient for $\psi(\Phi(t_\mathrm{f}; t_0, \boldsymbol{\Xi}_\theta, \boldsymbol{u}))$ to have full support for all $\theta \in \Theta$, $\boldsymbol{u} \in \Omega$, and $t_\mathrm{f} \in T$:

1. the elements of the POVM system $\mathcal{M}$ are positive definite; or
2. the density matrix $\boldsymbol{\Xi}_\theta$ is positive definite for all $\theta \in \Theta$ and the quantum control system Σ is unital (see Definition 2.10). □

Proof. The inner product of a non-zero, positive *semi*definite matrix and a positive *definite* matrix is a strictly positive number. If either $\boldsymbol{M}_k$ or $\Phi(t_\mathrm{f}; t_0, \boldsymbol{\Xi}_\theta, \boldsymbol{u})$ is positive definite, then Born's rule implies that measurement outcome k has non-zero probability. Hence, the first

condition in the Lemma's statement implies that the random variable $\psi(\Phi(t_\mathrm{f}; t_0, \boldsymbol{\Xi}_\theta, \boldsymbol{u}))$ has full support.

Suppose that $\boldsymbol{\Xi}_\theta$ is positive definite. If Σ is unital, then $\mathrm{rank}\{\Phi(t_\mathrm{f}; t_0, \boldsymbol{\Xi}_\theta, \boldsymbol{u})\} \geqslant \mathrm{rank}\{\boldsymbol{\Xi}_\theta\}$ by Corollary 2.1. Hence, $\Phi(t_\mathrm{f}; t_0, \boldsymbol{\Xi}_\theta, \boldsymbol{u})$ is positive definite if (i) $\boldsymbol{\Xi}_\theta$ is positive definite and (ii) Σ is unital. This proves sufficiency of the second condition. ⊠

The CFI quantifies the fundamental limit of information extraction by the quantum control system Σ.

Definition 4.3 (Controlled Fisher information)**.** A pair $(\boldsymbol{u}, t_\mathrm{f}) \in \Omega \times T$ consisting of a control law and a measurement time is referred to as a *control strategy*. The Fisher information $J_{\psi(\Phi(t_\mathrm{f};t_0,\boldsymbol{\Xi}_\theta,\boldsymbol{u}))}(\theta)$ about θ extracted by Σ using the control strategy $(\boldsymbol{u}, t_\mathrm{f})$ is referred to as the $(\boldsymbol{u}, t_\mathrm{f})$-*controlled Fisher information*. The *controlled Fisher information* about θ attainable by the quantum control system Σ is

$$J^{\mathrm{c}}_{\boldsymbol{\Xi}_\theta}(\theta; \Sigma) = \sup_{(\boldsymbol{u}, t_\mathrm{f}) \in \Omega \times T} J_{\psi(\Phi(t_\mathrm{f};t_0,\boldsymbol{\Xi}_\theta,\boldsymbol{u}))}(\theta)\,. \tag{4.8}$$

□

Obtaining an expression for the $(\boldsymbol{u}, t_\mathrm{f})$-CFI is straightforward. For any k in the support of $\psi(\Phi(t_\mathrm{f}; t_0, \boldsymbol{\Xi}_\theta, \boldsymbol{u}))$,

$$\frac{\mathrm{d}}{\mathrm{d}\theta} \ln\langle \boldsymbol{M}_k, \Phi(t_\mathrm{f}; t_0, \boldsymbol{\Xi}_\theta, \boldsymbol{u})\rangle_{\mathrm{HS}} = \frac{\langle \boldsymbol{M}_k, \Phi(t_\mathrm{f}; t_0, \frac{\mathrm{d}}{\mathrm{d}\theta}\boldsymbol{\Xi}_\theta, \boldsymbol{u})\rangle_{\mathrm{HS}}}{\langle \boldsymbol{M}_k, \Phi(t_\mathrm{f}; t_0, \boldsymbol{\Xi}_\theta, \boldsymbol{u})\rangle_{\mathrm{HS}}} \tag{4.9}$$

since $\Phi(t_\mathrm{f}; t_0, \cdot, \boldsymbol{u}) : \mathbb{C}^{d\times d} \to \mathbb{C}^{d\times d}$ is a linear mapping that does not depend on θ. Suppose, without loss of generality, that the support of $\psi(\Phi(t_\mathrm{f}; t_0, \boldsymbol{\Xi}_\theta, \boldsymbol{u}))$ is $\mathbb{N}_1^{N_\mathrm{m}}$. Applying Definition 4.1, the $(\boldsymbol{u}, t_\mathrm{f})$-CFI is given by the equivalent expressions

$$J_{\psi(\Phi(t_\mathrm{f};t_0,\boldsymbol{\Xi}_\theta,\boldsymbol{u}))}(\theta) = \sum_{k=1}^{N_\mathrm{m}} \frac{\langle \boldsymbol{M}_k, \Phi(t_\mathrm{f}; t_0, \frac{\mathrm{d}}{\mathrm{d}\theta}\boldsymbol{\Xi}_\theta, \boldsymbol{u})\rangle^2_{\mathrm{HS}}}{\langle \boldsymbol{M}_k, \Phi(t_\mathrm{f}; t_0, \boldsymbol{\Xi}_\theta, \boldsymbol{u})\rangle_{\mathrm{HS}}} = \sum_{k=1}^{N_\mathrm{m}} \frac{\langle \Phi^\ddagger(t_\mathrm{f}; t_0, \boldsymbol{M}_k, \boldsymbol{u}), \frac{\mathrm{d}}{\mathrm{d}\theta}\boldsymbol{\Xi}_\theta\rangle^2_{\mathrm{HS}}}{\langle \Phi^\ddagger(t_\mathrm{f}; t_0, \boldsymbol{M}_k, \boldsymbol{u}), \boldsymbol{\Xi}_\theta\rangle_{\mathrm{HS}}}\,. \tag{4.10}$$

Determining the value of the CFI generally consists of producing two types of bounds.

1. *Achievability*. An achievability result is of the form, "the CFI $J^{\mathrm{c}}_{\boldsymbol{\Xi}_\theta}(\theta; \Sigma)$ is at least $b_1(\theta)$," for some $b_1(\theta) \geqslant 0$. This type of result can be produced by evaluating the performance of a particular control strategy. Section 4.4 addresses control design, providing means to construct achievability results.

2. *Converse*. A converse bound is of the form, "the CFI $J^{\mathrm{c}}_{\boldsymbol{\Xi}_\theta}(\theta; \Sigma)$ is no larger than $b_2(\theta)$," for some $b_2(\theta) \geqslant 0$. Converse bound, which are generally more difficult to derive than achievability results, are addressed in Section 4.5.

Producing achievability and converse bounds gives bounds $b_1(\theta) \leqslant J^{\mathrm{c}}_{\boldsymbol{\Xi}_\theta}(\theta; \Sigma) \leqslant b_2(\theta)$, and if the gap $b_2(\theta) - b_1(\theta)$ is sufficiently small then $J^{\mathrm{c}}_{\boldsymbol{\Xi}_\theta}(\theta; \Sigma)$ is revealed to high-accuracy.

Several modifications to Definition 4.3 of the CFI are possible. In this regard, the following remark is in order.

Remark 4.1. It may not be necessary to optimize the measurement time t_{f}. In the case of certain unitary quantum systems, the set of states reachable by control is constant after t_{f} crosses a certain threshold [26, p. 95]. It may also be the case that the measurement time t_{f} is specified a priori, in which case there is no need to optimize over t_{f}. In these scenarios, the supremum in (4.8) is taken over $\boldsymbol{u} \in \Omega$ with some fixed $t_{\mathrm{f}} \in T$. □

4.2.1 Properties of controlled Fisher information

The CFI has several properties (e.g., Lipschitzness and the data processing inequality (DPI)) that will be used throughout this chapter. These properties are summarized in the lemmata below.

Lemma 4.2. Let Σ be a quantum control system and $\theta, \theta_1, \theta_2 \in \Theta$. The following properties hold:

1. (Non-negativity) $0 \leqslant J^{\mathrm{c}}_{\boldsymbol{\Xi}_\theta}(\theta; \Sigma)$;
2. (Universal converse) $J^{\mathrm{c}}_{\boldsymbol{\Xi}_\theta}(\theta; \Sigma) \leqslant J^{\mathrm{q}}_{\boldsymbol{\Xi}_\theta}(\theta)$; and
3. (Lipschitzness) if either of the conditions in Lemma 4.1 holds, then there exists $L > 0$ such that
$$\left| J_{\psi(\Phi(t_{\mathrm{f}};t_0,\boldsymbol{\Xi}_{\theta_1},\boldsymbol{u}))}(\theta_1) - J_{\psi(\Phi(t_{\mathrm{f}};t_0,\boldsymbol{\Xi}_{\theta_2},\boldsymbol{u}))}(\theta_2) \right| \leqslant L \left| \theta_1 - \theta_2 \right|. \tag{4.11}$$

□

Proof. 1) Non-negativity of the CFI follows from the fact that the CFI is the supremum over a set of Fisher informations, which are themselves non-negative.

2) For any $(\boldsymbol{u}, t_{\mathrm{f}}) \in \Omega \times T$, the set $\{\Phi^{\ddagger}(t_{\mathrm{f}}; t_0, \boldsymbol{M}_k, \boldsymbol{u}) \,:\, k \in \mathbb{N}_1^{N_{\mathrm{m}}}\}$ is a POVM system since the adjoint state transition map $\Phi^{\ddagger}(t_{\mathrm{f}}; t_0, \cdot, \boldsymbol{u})$ is completely positive and unital [71].[4] Equation (4.10) shows that the $(\boldsymbol{u}, t_{\mathrm{f}})$-CFI equals the Fisher information obtained by measuring $\boldsymbol{\Xi}_\theta$ with said POVM. Given Definition 4.2, each $(\boldsymbol{u}, t_{\mathrm{f}})$-CFI is upper bounded as $J_{\psi(\Phi(t_{\mathrm{f}};t_0\boldsymbol{\Xi}_\theta,\boldsymbol{u}))}(\theta) \leqslant J^{\mathrm{q}}_{\boldsymbol{\Xi}_\theta}(\theta)$. Taking the supremum of the left-hand side of this equality proves the universal converse.

3) Suppose that either of the conditions in Lemma 4.1 holds. It will be shown that the family of functions
$$\left\{ J_{\psi(\Phi(t_{\mathrm{f}};t_0,\boldsymbol{\Xi}_\theta,\boldsymbol{u}))} : \Theta \to \mathbb{R} \;\middle|\; (\boldsymbol{u}, t_{\mathrm{f}}) \in \Omega \times T \right\} \tag{4.12}$$
is uniformly Lipschitz continuous, implying Lipschitz continuity of the CFI via Lemma A.3. Towards this end, note that there exists $c > 0$ such that
$$\langle \boldsymbol{M}_k, \Phi(t_{\mathrm{f}}; t_0, \boldsymbol{\Xi}_\theta, \boldsymbol{u}) \rangle_{\mathrm{HS}} \geqslant c \tag{4.13}$$
for any $(\boldsymbol{u}, t_{\mathrm{f}}, k, \theta) \in \Omega \times T \times \mathbb{N}_1^{N_{\mathrm{m}}} \times \Theta$. To see this, first define the constants
$$\breve{\lambda} \triangleq \min_{\theta \in \Theta} \lambda_{\min}(\boldsymbol{\Xi}_\theta) \geqslant 0 \tag{4.14a}$$
$$\breve{m} \triangleq \min_{k \in \mathbb{N}_1^{N_{\mathrm{m}}}} \operatorname{tr}\{\boldsymbol{M}_k\} > 0 \tag{4.14b}$$

[4] The adjoint state transition map is unital regardless of the unitality of Σ.

where $\lambda_{\min}(\boldsymbol{\Xi}_\theta)$ denotes the smallest eigenvalue of $\boldsymbol{\Xi}_\theta$. The minimum in (4.14a) is justified because $\boldsymbol{\Xi}_\theta$ depends on θ continuously, the eigenvalues of a matrix are continuous [85, pp. 563-564], and Θ is compact. If the elements of the POVM system $\mathcal{M}$ are positive definite, then Lemma A.4 implies that

$$\begin{aligned}\langle \boldsymbol{M}_k, \Phi(t_\mathrm{f}; t_0, \boldsymbol{\Xi}_\theta, \boldsymbol{u})\rangle_{\mathrm{HS}} &\geqslant \lambda_{\min}(\boldsymbol{M}_k)\,\mathrm{tr}\{\Phi(t_\mathrm{f}; t_0, \boldsymbol{\Xi}_\theta, \boldsymbol{u})\} \\ &= \lambda_{\min}(\boldsymbol{M}_k) > 0 \end{aligned} \tag{4.15}$$

since the state transition map is trace-preserving and $\boldsymbol{\Xi}_\theta$ has unit trace. Hence, if the first condition of Lemma 4.1 holds, take $c \triangleq \lambda_{\min}(\boldsymbol{M}_k)$. On the other hand, if the quantum control system Σ is unital then

$$\begin{aligned}\langle \boldsymbol{M}_k, \Phi(t_\mathrm{f}; t_0, \boldsymbol{\Xi}_\theta, \boldsymbol{u})\rangle_{\mathrm{HS}} &= \langle \Phi^\ddagger(t_\mathrm{f}; t_0, \boldsymbol{M}_k, \boldsymbol{u}), \boldsymbol{\Xi}_\theta\rangle_{\mathrm{HS}} \\ &\geqslant \mathrm{tr}\{\Phi^\ddagger(t_\mathrm{f}; t_0, \boldsymbol{M}_k, \boldsymbol{u})\}\,\lambda_{\min}(\boldsymbol{\Xi}_\theta) \\ &= \mathrm{tr}\{\boldsymbol{M}_k\}\,\lambda_{\min}(\boldsymbol{\Xi}_\theta) \\ &\geqslant \breve{m}\,\breve{\lambda}\,. \end{aligned} \tag{4.16}$$

This follows from Lemma A.4 and the fact that the adjoint state transition map of a unital quantum system is trace-preserving. If the density matrix $\boldsymbol{\Xi}_\theta$ is positive definite for all $\theta \in \Theta$, then $\breve{\lambda} > 0$. Hence, if the second condition of Lemma 4.1 holds, take $c \triangleq \breve{m}\,\breve{\lambda}$.

Consider any $(\boldsymbol{u}, t_\mathrm{f}) \in \Omega \times T$. Leveraging (4.10) and defining $\boldsymbol{M}'_k \triangleq \Phi^\ddagger(t_\mathrm{f}; t_0, \boldsymbol{M}_k, \boldsymbol{u})$ for $k \in \mathbb{N}_1^{N_\mathrm{m}}$, the derivative of the $(\boldsymbol{u}, t_\mathrm{f})$-CFI with respect to θ is

$$\begin{aligned}\zeta(\theta) &\triangleq \frac{\partial}{\partial\theta} J_{\psi(\Phi(t_\mathrm{f};t_0,\boldsymbol{\Xi}_\theta,\boldsymbol{u}))}(\theta) \\ &= \sum_{k=1}^{N_\mathrm{m}} \frac{2\langle \boldsymbol{M}'_k, \frac{\mathrm{d}}{\mathrm{d}\theta}\boldsymbol{\Xi}_\theta\rangle_{\mathrm{HS}}\langle \boldsymbol{M}'_k, \frac{\mathrm{d}^2}{\mathrm{d}\theta^2}\boldsymbol{\Xi}_\theta\rangle_{\mathrm{HS}}}{\langle \boldsymbol{M}'_k, \boldsymbol{\Xi}_\theta\rangle_{\mathrm{HS}}} - \frac{\langle \boldsymbol{M}'_k, \frac{\mathrm{d}}{\mathrm{d}\theta}\boldsymbol{\Xi}_\theta\rangle^3_{\mathrm{HS}}}{\langle \boldsymbol{M}'_k, \boldsymbol{\Xi}_\theta\rangle^2_{\mathrm{HS}}}\,. \end{aligned} \tag{4.17}$$

It follows, by the fact that $\boldsymbol{\Xi}_\theta$ is twice continuously differentiable and that Θ is compact,

$$\begin{aligned}\max_{\theta\in\Theta} |\zeta(\theta)| &\leqslant \max_{\theta\in\Theta} \sum_{k=1}^{N_\mathrm{m}} \left[\frac{2}{c}\left|\langle \boldsymbol{M}'_k, \tfrac{\mathrm{d}}{\mathrm{d}\theta}\boldsymbol{\Xi}_\theta\rangle_{\mathrm{HS}}\right|\left|\langle \boldsymbol{M}'_k, \tfrac{\mathrm{d}^2}{\mathrm{d}\theta^2}\boldsymbol{\Xi}_\theta\rangle_{\mathrm{HS}}\right| + \frac{1}{c^2}\left|\langle \boldsymbol{M}'_k, \tfrac{\mathrm{d}}{\mathrm{d}\theta}\boldsymbol{\Xi}_\theta\rangle_{\mathrm{HS}}\right|^3\right] \\ &\triangleq L < \infty\,. \end{aligned} \tag{4.18}$$

By the mean value theorem, given $\theta_1, \theta_2 \in \Theta$, there exists $\theta^\star \in \Theta$ such that

$$\left|J_{\psi(\Phi(t_\mathrm{f};t_0,\boldsymbol{\Xi}_{\theta_1},\boldsymbol{u}))}(\theta_1) - J_{\psi(\Phi(t_\mathrm{f};t_0,\boldsymbol{\Xi}_{\theta_2},\boldsymbol{u}))}(\theta_2)\right| = |\zeta(\theta^\star)|\,|\theta_1 - \theta_2| \leqslant L\,|\theta_1 - \theta_2|\,. \tag{4.19}$$

Since the constant L does not depend on $(\boldsymbol{u}, t_\mathrm{f}) \in \Omega \times T$, this proves that the family (4.12) of $(\boldsymbol{u}, t_\mathrm{f})$-CFIs is uniformly L-Lipschitz. It follows from Lemma A.3 that the CFI (4.8) is also L-Lipschitz. ⊠

DPIs are useful when deriving bounds on an information measure. We now establish a DPI for the CFI. Recall [98, p. 8] that a POVM $\mathcal{M}'$ of cardinality $N'_\mathrm{m} \in \mathbb{N}$ is said to be a

post-processing of $\mathcal{M}$ if there exists a conditional probability distribution $p_{j|k}$ (which is not a function of θ) on $N'_{\mathrm{m}} \times N_{\mathrm{m}}$ such that

$$\boldsymbol{M}'_j = \sum_{k=1}^{N_{\mathrm{m}}} p_{j|k} \boldsymbol{M}_k \qquad \forall j \in \mathbb{N}_1^{N'_{\mathrm{m}}} \tag{4.20}$$

and $\mathcal{M}' = \{\boldsymbol{M}'_j \,:\, j \in \mathbb{N}_1^{N'_{\mathrm{m}}}\}$.

Lemma 4.3 (DPI for CFI)**.** Let Σ be a quantum control system equipped with the POVM $\mathcal{M}$, and let Σ' be the same system equipped instead with a post-processing $\mathcal{M}'$ of $\mathcal{M}$. For all $\theta \in \Theta$,

$$J^{\mathrm{c}}_{\boldsymbol{\Xi}_\theta}(\theta; \Sigma') \leqslant J^{\mathrm{c}}_{\boldsymbol{\Xi}_\theta}(\theta; \Sigma)\,. \tag{4.21}$$

□

Proof. Let $\psi' : \mathcal{D}(\mathcal{H}) \to \mathbb{N}_1^{N'_{\mathrm{m}}}$ denote the output map of the system Σ'. For any control strategy $(\boldsymbol{u}, t_{\mathrm{f}}) \in \Omega \times T$, it follows that

$$\begin{aligned}
\mathbb{P}\{\psi'(\Phi(t_{\mathrm{f}}; t_0, \boldsymbol{\Xi}_\theta, \boldsymbol{u})) = j\} &= \langle \boldsymbol{M}'_j, \Phi(t_{\mathrm{f}}; t_0, \boldsymbol{\Xi}_\theta, \boldsymbol{u})\rangle_{\mathrm{HS}} \\
&= \sum_{k=1}^{N_{\mathrm{m}}} p_{j|k} \langle \boldsymbol{M}_k, \Phi(t_{\mathrm{f}}; t_0, \boldsymbol{\Xi}_\theta, \boldsymbol{u})\rangle_{\mathrm{HS}} \\
&= \sum_{k=1}^{N_{\mathrm{m}}} p_{j|k}\, \mathbb{P}\{\psi(\Phi(t_{\mathrm{f}}; t_0, \boldsymbol{\Xi}_\theta, \boldsymbol{u})) = k\}
\end{aligned} \tag{4.22}$$

for each $j \in \mathbb{N}_1^{N'_{\mathrm{m}}}$. Since $p_{j|k}$ is independent of the parameter θ, the DPI for classical Fisher information [72, pp. 186-187] implies that

$$J_{\psi'(\Phi(t_{\mathrm{f}}; t_0, \boldsymbol{\Xi}_\theta, \boldsymbol{u}))}(\theta) \leqslant J_{\psi(\Phi(t_{\mathrm{f}}; t_0, \boldsymbol{\Xi}_\theta, \boldsymbol{u}))}(\theta)\,. \tag{4.23}$$

Supremizing both sides of (4.23) over all $(\boldsymbol{u}, t_{\mathrm{f}}) \in \Omega \times T$ proves the claim. ⊠

4.3 Adaptive information control

Determining control strategies achieving the CFI is tantamount to solving[5] the optimal control problem

$$\mathscr{P}(\theta): \quad \underset{(\boldsymbol{u}, t_{\mathrm{f}}) \in \Omega \times T}{\text{supremize}} \quad J_{\psi(\Phi(t_{\mathrm{f}}; t_0, \boldsymbol{\Xi}, \boldsymbol{u}))}(\theta) \tag{4.24}$$

Since problem $\mathscr{P}(\theta)$ is parameter-dependent, solving it generally requires knowledge of θ [1]. This would make a priori implementation of control strategies that solve $\mathscr{P}(\theta)$ infeasible, motivating adaptive information control. Using n sequential instances of a parameterized quantum state, the goal of adaptive information control is to:

[5] Because of the supremum, it may be the case that an exact solution does not exist. Therefore, we will use the terminology "solution" to mean a control-measurement time pair that leads arbitrarily close to the supremum.

1. apply a suitable sequence of information control strategies to each of the states, and
2. estimate the unknown parameter θ with optimal asymptotic variance $1/(nJ^{\mathrm{c}}_{\boldsymbol{\Xi}_\theta}(\theta;\Sigma))$ determined by the CFI.

This section establishes adaptive information controllers that achieve this goal by leveraging iterative experiments.

Suppose that multiple instances of the state $\boldsymbol{\Xi}_\theta(t_0) = \boldsymbol{\Xi}_\theta$ are available and consider, for creative inspiration, the following experimentation protocol.

1. A baseline control strategy $(\boldsymbol{u}, t_{\mathrm{f}})$ is applied to the first instance, resulting in a random outcome y_1. Based on this outcome, an estimator $\hat{\theta}(\mathsf{y}_1)$ for θ is established.
2. A new control strategy is generated by solving the optimal control problem $\mathscr{P}(\hat{\theta}(\mathsf{y}_1))$. This strategy is applied to the next instance of the state, resulting in a new random measurement outcome y_2 and parameter estimator $\hat{\theta}(\mathbf{y}_{1:2})$.
3. Step two is repeated iteratively. Letting $\mathbf{y}_{1:j}$ denote the random outcomes of the first j experiments, the control strategy used in the $(j+1)$st experiment is a solution to $\mathscr{P}(\hat{\theta}(\mathbf{y}_{1:j}))$.

At first glance, this protocol seems reasonable. However, quantum experiments generally occur on a faster timescale than that at which quantum optimal control problems can be solved. The prior protocol would require pauses between quantum experiments in order to reoptimize control strategies. On the other hand, many suboptimal quantum experiments could have been performed during these pauses, potentially leading to (cumulatively) more information extraction than a reoptimized control strategy could supply from a single experiment. Nonetheless, suboptimal experimentation is an inefficient use of resources as it precludes convergence of the estimation error with optimal scaling $1/(nJ^{\mathrm{c}}_{\boldsymbol{\Xi}_\theta}(\theta;\Sigma))$.

Accordingly, consider a different approach. Prior to experimentation, a finite set of $N_{\mathrm{r}} \in \mathbb{N}$ representative parameter values

$$\Theta_{\mathrm{r}} = \left\{\theta_{\mathrm{r}}^{(i)} \,:\, i \in \mathbb{N}_1^{N_{\mathrm{r}}}\right\} \subset \Theta \tag{4.25}$$

is selected. A set of representative information control strategies is constructed

$$\mathcal{R} = \left\{(\theta_{\mathrm{r}}^{(i)}, \boldsymbol{u}^{(i)}, t_{\mathrm{f}}^{(i)}) \,:\, i \in \mathbb{N}_1^{N_{\mathrm{r}}}\right\} \subset \Theta \times \Omega \times T \tag{4.26}$$

where $(\boldsymbol{u}^{(i)}, t_{\mathrm{f}}^{(i)})$ is an approximate solution to $\mathscr{P}(\theta_{\mathrm{r}}^{(i)})$. Exact solutions are not needed; this is made precise by Definition 4.4 in the next section. During experimentation, control strategies from the representative set are applied adaptively based on the current parameter estimate. Specifically, at the $(j+1)$st instance, a control strategy is selected from the pre-established representative set $\mathcal{R}$ according to $\pi(\hat{\theta}(\mathbf{y}_{1:j}))$, where $\pi : \Theta \to \mathbb{N}_1^{N_{\mathrm{r}}}$ is a decision rule. Example decision rules include:

$$\pi_{\mathrm{g}}(\theta) \in \arg\max_{i \in \mathbb{N}_1^{N_{\mathrm{r}}}} J_{\psi(\Phi(t_{\mathrm{f}}^{(i)}; t_0, \boldsymbol{\Xi}_\theta, \boldsymbol{u}^{(i)}))}(\theta) \tag{4.27a}$$

$$\pi_{\mathrm{n}}(\theta) \in \arg\min_{i \in \mathbb{N}_1^{|\mathcal{R}|}} |\theta_{\mathrm{r}}^{(i)} - \theta| \tag{4.27b}$$

which are known as the "greedy" and "nearest neighbor" decision rules, respectively. The experimentation protocol described in this paragraph is defined entirely by the tuple $(\mathcal{R}, \hat{\theta}(\cdot), \pi(\cdot))$, which we refer to as an *adaptive information controller*.

Computation of $\pi(\hat{\theta}(\mathbf{y}_{1:n}))$ is significantly lighter than reoptimizing a control strategy. Also, in light of advances in quantum feedback control via field-programmable gate arrays (FPGAs) [99, 100], adaptive information control could likely be implemented in hardware. In what follows, it is shown that adaptive information control enables optimal information extraction and estimation error converging to zero at a rate arbitrarily close to $1/(nJ^{\mathrm{c}}_{\boldsymbol{\Xi}_\theta}(\theta; \Sigma))$. Representative control strategies and decision rules are addressed first, followed by estimator design. The remainder of this section (i.e., Section 4.3) will suppose the following.

Assumption 4.2. At least one of the conditions of Lemma 4.1 is satisfied. □

4.3.1 ϵ-coverings of information controls

Sets of representative control strategies have been motivated by experimental considerations. Now, it will be shown that finite sets of representative control strategies can be constructed without sacrificing performance.

Definition 4.4 (ϵ-covering)**.** Fix a particular $\epsilon \geqslant 0$. A finite set $\mathcal{R} \subset \Theta \times \Omega \times T$ is said to be an *ϵ-covering of information control strategies* if, for each $\theta \in \Theta$, there exists $(\theta_{\mathrm{r}}, \boldsymbol{u}, t_{\mathrm{f}}) \in \mathcal{R}$ such that

$$J_{\psi(\Phi(t_{\mathrm{f}}; t_0, \boldsymbol{\Xi}_\theta, \boldsymbol{u}))}(\theta) \;\geqslant\; J^{\mathrm{c}}_{\boldsymbol{\Xi}_\theta}(\theta; \Sigma) - \epsilon\,. \tag{4.28}$$

Such control strategy is said to be an *ϵ-information maximizing control strategy* for the parameter $\theta \in \Theta$. Denoting the elements of $\mathcal{R}$ as $(\theta_{\mathrm{r}}^{(i)}, \boldsymbol{u}^{(i)}, t_{\mathrm{f}}^{(i)})$, $i \in \mathbb{N}_1^{|\mathcal{R}|}$, the set of indices corresponding to ϵ-information maximizing control strategies for θ is denoted $\Pi_\epsilon(\theta) \subseteq \mathbb{N}_1^{|\mathcal{R}|}$. □

Given an ϵ-covering $\mathcal{R}$, the following notation will be useful:

$$\check{J}(\theta) \triangleq \min_{i \in \mathbb{N}_1^{|\mathcal{R}|}} J_{\psi(\Phi(t_{\mathrm{f}}^{(i)}; t_0, \boldsymbol{\Xi}_\theta, \boldsymbol{u}^{(i)}))}(\theta) \tag{4.29a}$$

$$\hat{J}(\theta) \triangleq \max_{i \in \mathbb{N}_1^{|\mathcal{R}|}} J_{\psi(\Phi(t_{\mathrm{f}}^{(i)}; t_0, \boldsymbol{\Xi}_\theta, \boldsymbol{u}^{(i)}))}(\theta)\,. \tag{4.29b}$$

Definition 4.5 (Robust decision rule)**.** Given an ϵ-covering of information control strategies $\mathcal{R}$, a decision rule $\pi : \Theta \to \mathbb{N}_1^{|\mathcal{R}|}$ is said to be *robust on* Θ if for each $\theta \in \Theta$ there exists $r(\theta) > 0$ such that $\pi(\theta') \in \Pi_\epsilon(\theta)$ if $|\theta' - \theta| < r(\theta)$, $\theta' \in \Theta$. Such a rule is said to be *strongly robust at* $\theta \in \Theta$ if there exists $r(\theta) > 0$ such that $\pi(\theta') = \pi(\theta) \in \Pi_\epsilon(\theta)$ if $|\theta' - \theta| < r(\theta)$, $\theta' \in \Theta$. □

Remark 4.2. Suppose π is robust on Θ. A (trivial) sufficient condition for strong robustness of π at θ is that $|\Pi_\epsilon(\theta)| = 1$. □

An ϵ-covering of information control strategies enables information extraction within ϵ of the CFI uniformly across the entire parameter space. Robustness allows one to determine an ϵ-information maximizing control strategy using an estimate of the true parameter. The next theorem shows that ϵ-coverings and robust decision rules exist for all $\epsilon > 0$.

Theorem 4.1 (Existence of ϵ-coverings and robust decision rules)**.** For any $\epsilon > 0$, there exists an ϵ-covering of information control strategies $\mathcal{R}$ with the property that the nearest neighbor decision rule $\pi_{\mathrm{n}}(\cdot)$ defined in (4.27b) is (i) robust on Θ and (ii) strongly robust at all but finitely many points in Θ. □

Proof. Let $\epsilon > 0$ be arbitrary. Under the standing hypothesis of Assumption 4.2, Lemma 4.2 proves that the CFI is L-Lipschitz continuous for some $L > 0$. Denoting the maximum and minimum elements of Θ by $\hat{\theta}$ and $\check{\theta}$, let $C \triangleq 2L(\hat{\theta} - \check{\theta})$. An ϵ-covering $\mathcal{R}$ can be constructed as follows.

1. Place a finite number of equally spaced representative parameter values on Θ such that each element of the parameter space Θ is no further than $\epsilon/2L$ from a representative parameter value.

2. For each representative parameter value $\theta_{\mathrm{r}}^{(i)}$, take a control strategy $(\boldsymbol{u}^{(i)}, t_{\mathrm{f}}^{(i)}) \in \Omega \times T$ such that
$$J^{\mathrm{c}}_{\boldsymbol{\Xi}_{\theta_{\mathrm{r}}^{(i)}}}(\theta_{\mathrm{r}}^{(i)}; \Sigma) - \frac{\epsilon}{2} \leqslant J_{\psi(\Phi(t_{\mathrm{f}}^{(i)}; t_0, \boldsymbol{\Xi}_{\theta_{\mathrm{r}}^{(i)}}, \boldsymbol{u}^{(i)}))}(\theta_{\mathrm{r}}^{(i)}) \,. \tag{4.30}$$
Such strategies exist by definition of the CFI.

To verify that $\mathcal{R}$ is an ϵ-covering, let $\theta \in \Theta$ and observe that $\pi_{\mathrm{n}}(\theta)$ selects a representative control strategy, say $(\boldsymbol{u}^{(1)}, t_{\mathrm{f}}^{(1)})$, with the property that
$$|\theta_{\mathrm{r}}^{(1)} - \theta| \leqslant \frac{\epsilon}{2L} \,. \tag{4.31}$$
By Lipschitzness of the CFI and construction of the representative control strategies,
$$\begin{aligned} J^{\mathrm{c}}_{\boldsymbol{\Xi}_{\theta}}(\theta; \Sigma) &\leqslant J^{\mathrm{c}}_{\boldsymbol{\Xi}_{\theta_{\mathrm{r}}^{(1)}}}(\theta_{\mathrm{r}}^{(1)}; \Sigma) + L|\theta_{\mathrm{r}}^{(1)} - \theta| \\ &\leqslant J^{\mathrm{c}}_{\boldsymbol{\Xi}_{\theta_{\mathrm{r}}^{(1)}}}(\theta_{\mathrm{r}}^{(1)}; \Sigma) + \frac{\epsilon}{2} \\ &\leqslant J_{\psi(\Phi(t_{\mathrm{f}}^{(1)}; t_0, \boldsymbol{\Xi}_{\theta_{\mathrm{r}}^{(1)}}, \boldsymbol{u}^{(1)}))}(\theta_{\mathrm{r}}^{(1)}) + \epsilon \,. \end{aligned} \tag{4.32}$$
This proves that $\mathcal{R}$ is an ϵ-covering and that the nearest-neighbor decision rule is robust on Θ. The strong robustness claim follows from Remark 4.2 and the fact that the representative parameter values are equally spaced. ⊠

Coverings of information controls generalize notions established in our prior papers [1, 2].

1. If it exists, a 0-covering of cardinality 1 has the property that a single control strategy $(\boldsymbol{u}, t_{\mathrm{f}})$ is universally optimal, maximizing the Fisher information across the entire parameter space. This type of control, known as an "admissible control", was introduced in [1].

2. For the quantum system studied in [2], it was shown that there exists a set of two control strategies with the property that one of them achieves the CFI at any point in the parameter space. These two control strategies form a 0-covering of cardinality 2.

While 0-coverings do not always exist, Theorem 4.1 proves that ($\epsilon > 0$)-coverings do. Hence, ϵ-coverings are a strict generalization of these existing notions.

Once an ϵ-covering of information control strategies has been determined, the next step in designing an adaptive information controller is to construct an estimator. In Sections 4.3.2 and 4.3.3, two different types of estimators are investigated. The main results of these sections show that: (i) with high probability, adaptive information control selects ϵ-information maximizing control strategies after a sufficient number of experiments; and (ii) the estimator's variance has optimal asymptotic scaling. That is, the variance scales according to the inverse of the CFI (times the number n of experiments performed).

4.3.2 Adaptation via maximum likelihood estimation

The first type of estimator is the maximum likelihood estimator. Given an ϵ-covering $\mathcal{R}$, it is useful to define the probability distributions $p_\theta^{(i)}(y) \triangleq \langle \boldsymbol{M}_y^{(i)}, \boldsymbol{\Xi}_\theta \rangle_{\mathrm{HS}}$, where

$$\boldsymbol{M}_y^{(i)} \triangleq \Phi^\ddagger(t_{\mathrm{f}}^{(i)}; t_0, \boldsymbol{M}_y, \boldsymbol{u}^{(i)}) \tag{4.33}$$

for each $(i, y) \in \mathbb{N}_1^{|\mathcal{R}|} \times \mathbb{N}_1^{N_{\mathrm{m}}}$. An estimator $\hat{\theta}_{\mathrm{ml}}(\mathbf{y}_{1:n})$ is said to be a maximum likelihood estimator if

$$\hat{\theta}_{\mathrm{ml}}(\mathsf{y}_1) \in \arg\max_{\theta \in \Theta} \; p_\theta^{(1)}(\mathsf{y}_1) \tag{4.34}$$

and, for $n \geqslant 2$,

$$\hat{\theta}_{\mathrm{ml}}(\mathbf{y}_{1:n}) \in \arg\max_{\theta \in \Theta} \; p_\theta^{(1)}(\mathsf{y}_1) \prod_{j=2}^{n} p_\theta^{(\pi(\hat{\theta}_{\mathrm{ml}}(\mathbf{y}_{1:(j-1)})))}(\mathsf{y}_j)\,. \tag{4.35}$$

Definition 4.6 (cf. [94, p. 523])**.** A statistical model $\{\, p_\theta \,:\, \theta \in \Theta \,\}$ is said to be *identifiable* if for every pair of distinct parameters $\theta_1, \theta_2 \in \Theta$ the Kullback-Leibler divergence $D(p_{\theta_1} \| p_{\theta_2})$ between p_{θ_1} and p_{θ_2} is positive. □

A statistical model must be identifiable in order to distinguish the true value of the parameter from other elements of the parameter space. The following theorem establishes performance guarantees for the adaptive information controller $(\mathcal{R}, \hat{\theta}_{\mathrm{ml}}(\cdot), \pi(\cdot))$.

Theorem 4.2 (Convergence of MLE-based adaptive information control)**.** Let $\theta_0 \in \Theta$ be the true parameter, $\mathcal{R}$ be an ϵ-covering, $\hat{\theta}_{\mathrm{ml}}(\cdot)$ be a maximum likelihood estimator, and $\pi(\cdot)$ be robust on Θ. Suppose that the statistical model $\{\, p_\theta^{(i)} \,:\, \theta \in \Theta \,\}$ induced by each representative control strategy $i \in \mathbb{N}_1^{|\mathcal{R}|}$ is identifiable. Then, there exist positive constants c_1 and c_2 such that the probability of the adaptive information controller $(\mathcal{R}, \hat{\theta}_{\mathrm{ml}}(\cdot), \pi(\cdot))$ selecting an ϵ-information maximizing control strategy satisfies

$$\mathbb{P}\big\{\pi(\hat{\theta}_{\mathrm{ml}}(\mathbf{y}_{1:n})) \in \mathit{\Pi}_\epsilon(\theta_0)\big\} \geqslant 1 - c_1\, e^{-n c_2} \tag{4.36}$$

for all $n \in \mathbb{N}$. If $\pi(\cdot)$ is strongly robust at θ_0 then there exists a maximum likelihood estimator such that

$$\sqrt{n}\big(\hat{\theta}_{\mathrm{ml}}(\mathbf{y}_{1:n}) - \theta_0\big) \xrightarrow{\mathrm{d}} \mathcal{N}\Big(0, \frac{1}{J^{\mathrm{c}}_{\boldsymbol{\Xi}_\theta}(\theta_0; \Sigma) - \epsilon'}\Big) \tag{4.37}$$

for some $\epsilon' \in [0, \epsilon]$, where $\xrightarrow{\mathrm{d}}$ denotes convergence in distribution. □

The idea of the proof is as follows. First, a concentration inequality is invoked to show that, with high probability, the adaptive controller selects an ϵ-information maximizing control strategy. Then, this high probability bound is used to quantify the asymptotic behavior of the estimation error. To reduce the notational burden of the proof, given a sequence of n experiments, denote the index of the control strategy used at the jth experiment by

$$\nu_j \triangleq \begin{cases} 1 & , \quad \text{if } j = 1 \\ \pi(\hat{\theta}_{\mathrm{ml}}(\mathbf{y}_{1:j-1})) \, , & \text{if } j \in \mathbb{N}_2^n \end{cases} \tag{4.38}$$

and let $\boldsymbol{\nu} \triangleq [\nu_1 \, \nu_2 \, \ldots \, \nu_n]$.

Proof of Theorem 4.2. Define an open ball $B \triangleq \{\theta \in \Theta \, : \, |\theta - \theta_0| < r(\theta_0)\}$ of radius $r(\theta_0)$ centered at the true parameter θ_0 such that, if the parameter estimate $\hat{\theta}$ falls within this ball, then $\pi(\hat{\theta}) \in \Pi_\epsilon(\theta_0)$. Such a ball exists since $\pi(\cdot)$ is robust on Θ. Monotonicity of probability implies that

$$\begin{aligned} \mathbb{P}\big\{\pi(\hat{\theta}_{\mathrm{ml}}(\mathbf{y}_{1:n})) \in \Pi_\epsilon(\theta_0)\big\} &= 1 - \mathbb{P}\big\{\pi(\hat{\theta}_{\mathrm{ml}}(\mathbf{y}_{1:n})) \notin \Pi_\epsilon(\theta_0)\big\} \\ &\geqslant 1 - \mathbb{P}\big\{\hat{\theta}_{\mathrm{ml}}(\mathbf{y}_{1:n}) \notin B\big\} \, . \end{aligned} \tag{4.39}$$

Consider any instantiation $\boldsymbol{\nu}$ of $\boldsymbol{\nu}$. The monotonicity of probability as well as the logarithm gives

$$\begin{aligned} \mathbb{P}\big\{\hat{\theta}_{\mathrm{ml}}(\mathbf{y}_{1:n}) \notin B \,|\, \boldsymbol{\nu} = \boldsymbol{\nu}\big\} &\leqslant \mathbb{P}\left\{ \frac{\sup\limits_{\theta \in \Theta \setminus B} \prod\limits_{j=1}^{n} p_\theta^{(\nu_j)}(\mathsf{y}_j)}{\sup\limits_{\theta \in B} \prod\limits_{j=1}^{n} p_\theta^{(\nu_j)}(\mathsf{y}_j)} \geqslant 1 \right\} \\ &\leqslant \mathbb{P}\left\{ \sup_{\theta \in \Theta \setminus B} \prod_{j=1}^{n} \frac{p_\theta^{(\nu_j)}(\mathsf{y}_j)}{p_{\theta_0}^{(\nu_j)}(\mathsf{y}_j)} \geqslant 1 \right\} \\ &= \mathbb{P}\left\{ \sup_{\theta \in \Theta \setminus B} \sum_{j=1}^{n} \log\left(\frac{p_\theta^{(\nu_j)}(\mathsf{y}_j)}{p_{\theta_0}^{(\nu_j)}(\mathsf{y}_j)} \right) \geqslant 0 \right\} \qquad (4.40a) \\ &\leqslant \mathbb{P}\left\{ \sum_{j=1}^{n} \sup_{\theta \in \Theta \setminus B} \log\left(\frac{p_\theta^{(\nu_j)}(\mathsf{y}_j)}{p_{\theta_0}^{(\nu_j)}(\mathsf{y}_j)} \right) \geqslant 0 \right\} . \qquad (4.40b) \end{aligned}$$

At this point, if the random variables in the summation of (4.40b) were bounded and had negative mean, one could invoke a concentration inequality to constrain $\mathbb{P}\big\{\hat{\theta}_{\mathrm{ml}}(\mathbf{y}_{1:n}) \notin B\big\}$. However, the random variables in question do not necessarily have negative mean. This hurdle is addressed next using a covering argument.

Let $i \in \mathbb{N}_1^{|\mathcal{R}|}$ and $\theta \in \Theta \setminus B$. Given the identifiability of the statistical model $\{ \, p_\theta^{(i)} \, : \, \theta \in \Theta \, \}$ and the fact that[6]

$$D(p_{\theta_0}^{(i)} \| p_\theta^{(i)}) = - \lim_{\rho \to 0} \mathbb{E}_{\mathsf{y} \sim p_{\theta_0}^{(i)}} \left\{ \sup_{\theta' \, : \, |\theta' - \theta| \leqslant \rho} \log\left(\frac{p_{\theta'}^{(i)}(\mathsf{y})}{p_{\theta_0}^{(i)}(\mathsf{y})} \right) \right\} \tag{4.41}$$

[6]Equation (4.41) was proven by Wald in [101] prior to the introduction of the Kullback-Leibler divergence in [102].

there exists an open set $\mathcal{O}_\theta \subseteq \Theta \setminus B$ containing θ such that

$$\mathbb{E}_{\mathsf{y}\sim p^{(i)}_{\theta_0}}\left\{ \sup_{\theta'\in\mathcal{O}_\theta} \log\left(\frac{p^{(i)}_{\theta'}(\mathsf{y})}{p^{(i)}_{\theta_0}(\mathsf{y})}\right)\right\} < 0\,. \tag{4.42}$$

Since Θ is compact and B is open, $\Theta \setminus B$ is compact and there exists a *finite* open cover $\mathcal{C}^{(i)} = \{\mathcal{O}^{(i)}_1, \mathcal{O}^{(i)}_2, \ldots, \mathcal{O}^{(i)}_{N^{(i)}}\}$ of $\Theta \setminus B$ such that each element of the cover has property (4.42). Construct a new open cover of $\Theta \setminus B$ as

$$\mathcal{C} \triangleq \left\{\mathcal{O} \neq \varnothing \,:\, \mathcal{O} = \bigcap_{i=1}^{|\mathcal{R}|} \mathcal{O}^{(i)}_{\alpha_i},\, \alpha_i \in \mathbb{N}_1^{N^{(i)}}\right\},\quad |\mathcal{C}| < \infty\,. \tag{4.43}$$

By construction, this new open cover has the property that

$$\xi \triangleq \max_{(i,\mathcal{O})\in\mathbb{N}_1^{|\mathcal{R}|}\times\mathcal{C}} \mathbb{E}_{\mathsf{y}\sim p^{(i)}_{\theta_0}}\left\{ \sup_{\theta'\in\mathcal{O}} \log\left(\frac{p^{(i)}_{\theta'}(\mathsf{y})}{p^{(i)}_{\theta_0}(\mathsf{y})}\right)\right\} < 0\,. \tag{4.44}$$

By the proof of Lemma 4.2, there exists $c > 0$ such that $p^{(i)}_\theta(y) = \langle \boldsymbol{M}^{(i)}_y, \boldsymbol{\Xi}_\theta\rangle_{\mathrm{HS}} \geqslant c$ for all $\theta \in \Theta$. Hence,

$$\log(c) \leqslant \log\left(\frac{p^{(i)}_\theta(y)}{p^{(i)}_{\theta_0}(y)}\right) \leqslant -\log(c) \tag{4.45}$$

for all $(i, y, \theta) \in \mathbb{N}_1^{|\mathcal{R}|} \times \mathbb{N}_1^{N_{\mathrm{m}}} \times \Theta$ and therefore

$$\log(c) \leqslant \sup_{\theta'\in\mathcal{O}} \log\left(\frac{p^{(i)}_{\theta'}(y)}{p^{(i)}_{\theta_0}(y)}\right) \leqslant -\log(c) \tag{4.46}$$

for any $(i, y, \mathcal{O}) \in \mathbb{N}_1^{|\mathcal{R}|} \times \mathbb{N}_1^{N_{\mathrm{m}}} \times \mathcal{C}$.

The cover $\mathcal{C}$ enables the application of a concentration inequality. Considering any instantiation $\boldsymbol{\nu}$ of $\boldsymbol{\nu}$,

$$\begin{aligned}
\mathbb{P}\left\{ \sup_{\theta\in\Theta\setminus B} \sum_{j=1}^{n} \log\left(\frac{p^{(\nu_j)}_\theta(\mathsf{y}_j)}{p^{(\nu_j)}_{\theta_0}(\mathsf{y}_j)}\right) \geqslant 0\right\} &= \mathbb{P}\left\{ \bigvee_{\mathcal{O}\in\mathcal{C}} \left[\sup_{\theta\in\mathcal{O}} \sum_{j=1}^{n} \log\left(\frac{p^{(\nu_j)}_\theta(\mathsf{y}_j)}{p^{(\nu_j)}_{\theta_0}(\mathsf{y}_j)}\right)\right] \geqslant 0\right\} \\
&\leqslant \sum_{\mathcal{O}\in\mathcal{C}} \mathbb{P}\left\{ \sup_{\theta\in\mathcal{O}} \sum_{j=1}^{n} \log\left(\frac{p^{(\nu_j)}_\theta(\mathsf{y}_j)}{p^{(\nu_j)}_{\theta_0}(\mathsf{y}_j)}\right) \geqslant 0\right\} \\
&\leqslant \sum_{\mathcal{O}\in\mathcal{C}} \mathbb{P}\left\{ \sum_{j=1}^{n} \sup_{\theta\in\mathcal{O}} \log\left(\frac{p^{(\nu_j)}_\theta(\mathsf{y}_j)}{p^{(\nu_j)}_{\theta_0}(\mathsf{y}_j)}\right) \geqslant 0\right\} \\
&\leqslant |\mathcal{C}| \exp\left\{-\frac{2n\xi^2}{\log^2(c)}\right\}
\end{aligned} \tag{4.47}$$

where (4.47) is due to Hoeffding's inequality [103, p. 25]. It follows from (4.40a) and (4.47)

that

$$\begin{aligned}\mathbb{P}\big\{\hat{\theta}_{\mathrm{ml}}(\mathbf{y}_{1:n}) \notin B\big\} &= \sum_{\boldsymbol{\nu}} \mathbb{P}\big\{\hat{\theta}_{\mathrm{ml}}(\mathbf{y}_{1:n}) \notin B \,|\, \boldsymbol{\nu} = \boldsymbol{\nu}\big\}\mathbb{P}\{\boldsymbol{\nu} = \boldsymbol{\nu}\} \\ &\leqslant \sum_{\boldsymbol{\nu}} \mathbb{P}\left\{ \sup_{\theta \in \Theta \setminus B} \sum_{j=1}^{n} \log\left(\frac{p_{\theta}^{(\nu_j)}(\mathsf{y}_j)}{p_{\theta_0}^{(\nu_j)}(\mathsf{y}_j)}\right) \geqslant 0 \right\} \mathbb{P}\{\boldsymbol{\nu} = \boldsymbol{\nu}\} \\ &\leqslant |\mathcal{C}| \exp\left\{ -\frac{2n\xi^2}{\log^2(c)} \right\}. \end{aligned} \tag{4.48}$$

By (4.39), the claim (4.36) is proven with $c_1 \triangleq |\mathcal{C}|$ and $c_2 \triangleq 2\xi^2/\log^2(c)$.

We now prove (4.37). Suppose that the decision rule $\pi(\cdot)$ is strongly robust at the true parameter θ_0. Without loss of generality, suppose that the first representative control strategy is information maximizing, i.e., $\pi(\theta_0) = 1$. By definition,

$$J^{(1)}(\theta_0) = J_{\psi(\Phi(t_{\mathrm{f}}^{(1)}; t_0, \boldsymbol{\Xi}_{\theta_0}, \boldsymbol{u}^{(1)}))}(\theta_0) = J^{\mathrm{c}}_{\boldsymbol{\Xi}_{\theta_0}}(\theta_0; \Sigma) - \epsilon' \tag{4.49}$$

where $\epsilon' \in [0, \epsilon]$. Let $\mathsf{J}_{1:n}(\theta_0)$ denote the Fisher information extracted by the first $n \in \mathbb{N}$ experiments, i.e.,

$$\mathsf{J}_{1:n}(\theta_0) \triangleq \sum_{j=1}^{n} J_{\psi(\Phi(t_{\mathrm{f}}^{(\nu_j)}; t_0, \boldsymbol{\Xi}_{\theta_0}, \boldsymbol{u}^{(\nu_j)}))}(\theta_0) \tag{4.50}$$

which is bounded as

$$\mathsf{n}^{(1)} J^{(1)}(\theta_0) + (n - \mathsf{n}^{(1)}) \breve{J}(\theta_0) \leqslant \mathsf{J}_{1:n}(\theta_0) \tag{4.51a}$$

$$\mathsf{n}^{(1)} J^{(1)}(\theta_0) + (n - \mathsf{n}^{(1)}) \hat{J}(\theta_0) \geqslant \mathsf{J}_{1:n}(\theta_0) \tag{4.51b}$$

where $\mathsf{n}^{(1)}$ is the number of times the first representative control strategy is applied. The result (4.37) will follow (e.g., from [104, p. 427]) if it can be shown that

$$\frac{1}{n}\mathsf{J}_{1:n}(\theta_0) \xrightarrow{\mathrm{p}} J^{(1)}(\theta_0) \overset{(4.49)}{=} J^{\mathrm{c}}_{\boldsymbol{\Xi}_{\theta_0}}(\theta_0; \Sigma) - \epsilon'. \tag{4.52}$$

This is tantamount to showing that, for any $\delta > 0$ and $\sigma \in (0, 1]$, the inequalities

$$\eta_1 \triangleq \mathbb{P}\left\{ \frac{1}{n}\mathsf{J}_{1:n}(\theta_0) - J^{(1)}(\theta_0) \geqslant -\delta \right\} \geqslant 1 - \frac{\sigma}{2} \tag{4.53a}$$

$$\eta_2 \triangleq \mathbb{P}\left\{ \frac{1}{n}\mathsf{J}_{1:n}(\theta_0) - J^{(1)}(\theta_0) \leqslant +\delta \right\} \geqslant 1 - \frac{\sigma}{2} \tag{4.53b}$$

hold for sufficiently large $n \in \mathbb{N}$ since

$$\mathbb{P}\left\{ \left|\frac{1}{n}\mathsf{J}_{1:n}(\theta_0) - J^{(1)}(\theta_0)\right| \leqslant \delta \right\} \geqslant \eta_1 + \eta_2 - 1. \tag{4.54}$$

We will construct $N_{\delta,\sigma} \in \mathbb{N}$ such that (4.53a) holds if $n \geqslant N_{\delta,\sigma}$. Similar argumentation shows there exists $M_{\delta,\sigma} \in \mathbb{N}$ such that (4.53b) holds if $n \geqslant M_{\delta,\sigma}$. From there, (4.37) follows since

$$\mathbb{P}\left\{ \left|\frac{1}{n}\mathsf{J}_{1:n}(\theta_0) - J^{(1)}(\theta_0)\right| \leqslant \delta \right\} \geqslant 1 - \sigma \tag{4.55}$$

holds for all $n \geqslant \max\{N_{\delta,\sigma},\, M_{\delta,\sigma}\}$.

Let $\delta > 0$ and $\sigma \in (0,1]$; let c_1 and c_2 be the constants defined on the prior page. Because the series $\sum_{j=1}^{\infty} c_1 e^{-jc_2}$ converges, there exists $K \in \mathbb{N}$ such that $\sum_{j=K+1}^{\infty} c_1 e^{-jc_2} \leqslant \sigma/2$. Choose $N_{\delta,\sigma} \geqslant K$ large enough so that

$$\frac{K}{N_{\delta,\sigma}}\big(J^{(1)}(\theta_0) - \breve{J}(\theta_0)\big) \leqslant \delta\,. \tag{4.56}$$

Let $n \geqslant N_{\delta,\sigma}$. Conditioned on the event that $\mathsf{n}^{(1)} \geqslant (n-K)$, the random variable $\mathsf{J}_{1:n}(\theta_0)/n$ satisfies the bound

$$\frac{1}{n}\mathsf{J}_{1:n}(\theta_0) \geqslant \frac{n-K}{n} J^{(1)}(\theta_0) + \frac{K}{n}\breve{J}(\theta_0) \geqslant J^{(1)}(\theta_0) - \delta\,. \tag{4.57}$$

In other words,

$$\mathbb{P}\Big\{ \frac{1}{n}\mathsf{J}_{1:n}(\theta_0) - J^{(1)}(\theta_0) \geqslant -\delta \,\Big|\, \mathsf{n}^{(1)} \geqslant n-K \Big\} = 1\,. \tag{4.58}$$

On the other hand, monotonicity of probability and the union bound give

$$\begin{aligned}
\mathbb{P}\{\mathsf{n}^{(1)} \geqslant n-K\} &\geqslant \mathbb{P}\big\{\hat{\theta}_{\mathrm{ml}}(\mathbf{y}_{1:j}) \in B \;\; \forall j \in \mathbb{N}_{K+1}^{n}\big\} \\
&= 1 - \mathbb{P}\big\{\exists j \in \mathbb{N}_{K+1}^{n} \,:\, \hat{\theta}_{\mathrm{ml}}(\mathbf{y}_{1:j}) \notin B\big\} \\
&\geqslant 1 - \sum_{j=K+1}^{n} \mathbb{P}\big\{\hat{\theta}_{\mathrm{ml}}(\mathbf{y}_{1:j}) \notin B\big\}\,.
\end{aligned} \tag{4.59}$$

Combining (4.58), (4.59), and (4.48) with our choice of K,

$$\eta_1 = \mathbb{P}\Big\{ \frac{1}{n}\mathsf{J}_{1:n}(\theta_0) - J^{(1)}(\theta_0) \geqslant -\delta \Big\} \geqslant \mathbb{P}\{\mathsf{n}^{(1)} \geqslant n-K\} \geqslant 1 - \frac{\sigma}{2} \tag{4.60}$$

proving that (4.53a) holds for $n \geqslant N_{\delta,\sigma}$. ⊠

Theorem 4.2 shows that, with overwhelming probability, the adaptive controller will eventually select an ϵ-information maximizing control strategy. Moreover, the variance of the estimator scales with near-optimal asymptotic rate.

Let $(\mathcal{R}, \hat{\theta}(\cdot), \pi(\cdot))$ be an adaptive information controller. The mathematical structure essential to proving results like Theorem 4.2 is the following.

1. *Approximability:* The set $\mathcal{R}$ of representative controls should be an ϵ-covering so that one can extract information at a level within ϵ of the CFI.

2. *Consistency:* The estimator $\hat{\theta}(\mathbf{y}_{1:n})$ should converge to the true parameter as the number of experiments n grows.

3. *Robustness:* The decision rule $\pi(\cdot)$ should be robust such that, if $\hat{\theta}$ is near the true parameter, then $\hat{\theta}$ can be used to select an ϵ-information maximizing control.

One can leverage this structure to prove convergence results in a variety of settings. In the next section, this structure is exploited to prove a result similar to Theorem 4.2 when the maximum likelihood estimator is replaced by a so-called composite estimator.

4.3.3 Adaptation via composite estimation

Consider an adaptive information controller based on an ϵ-covering $\mathcal{R}$. Let $\theta_0 \in \Theta$ be the true parameter and let the minimal Fisher information (4.29a) be positive $\breve{J}(\theta_0) > 0$. In a sequence of n experiments, let $\mathsf{n}^{(i)}$ denote the number of times the control strategy $(\boldsymbol{u}^{(i)}, t_{\mathrm{f}}^{(i)})$ is used and let $\mathbf{y}^{(i)}$ denote the subset of measurement results obtained when the control strategy $(\boldsymbol{u}^{(i)}, t_{\mathrm{f}}^{(i)})$ is used, $i \in \mathbb{N}_1^{|\mathcal{R}|}$. The elements of $\mathbf{y}^{(i)}$ are IID $\psi(\Phi(t_{\mathrm{f}}^{(i)}; t_0, \boldsymbol{\Xi}_{\theta_0}, \boldsymbol{u}^{(i)}))$ random variables. Each $\mathsf{n}^{(i)}$ is random and $n = \mathsf{n}^{(1)} + \mathsf{n}^{(2)} + \ldots + \mathsf{n}^{(|\mathcal{R}|)}$. Define $\mathbf{n} \triangleq [\mathsf{n}^{(1)}\, \mathsf{n}^{(2)} \ldots \mathsf{n}^{(|\mathcal{R}|)}]$.

Suppose that, for each representative control strategy $i \in \mathbb{N}_1^{|\mathcal{R}|}$, there exists an unbiased estimator $\hat{\theta}^{(i)}(\mathbf{y}^{(i)}) \in \Theta$ such that

$$\mathbb{V}\{\hat{\theta}^{(i)}(\mathbf{y}^{(i)}) \,|\, \mathsf{n}^{(i)} = n^{(i)}\} = \frac{1}{n^{(i)} J_{\psi(\Phi(t_{\mathrm{f}}^{(i)}; t_0, \boldsymbol{\Xi}_{\theta_0}, \boldsymbol{u}^{(i)}))}(\theta_0)} \tag{4.61}$$

holds for any realization $n^{(i)}$ of $\mathsf{n}^{(i)}$. That is, $\hat{\theta}^{(i)}$ is an efficient estimator given random variables generated by the ith control strategy. The *composite estimator* induced by the efficient estimators $\{\, \hat{\theta}^{(i)}(\cdot) \,:\, i \in \mathbb{N}_1^{|\mathcal{R}|} \,\}$ is

$$\hat{\theta}_{\mathrm{cp}}(\mathbf{y}_{1:n}) \triangleq \sum_{i=1}^{|\mathcal{R}|} \frac{\mathbf{n}^{(i)}}{n} \hat{\theta}^{(i)}(\mathbf{y}^{(i)}) \tag{4.62}$$

which is itself unbiased. To state the next theorem, let $\hat{\theta}$ and $\breve{\theta}$ denote the maximum and minimum values of the compact set $\Theta \subset \mathbb{R}$, respectively.

Theorem 4.3 (Convergence of composite estimation-based adaptive information control)**.** Let $\theta_0 \in \Theta$ be the true parameter, $\mathcal{R}$ be an ϵ-covering, $\hat{\theta}_{\mathrm{cp}}(\cdot)$ be a composite estimator, and $\pi(\cdot)$ be robust on Θ. Then, there exists a positive constant $c_3 > 0$ such that the probability of the adaptive information controller $(\mathcal{R}, \hat{\theta}_{\mathrm{cp}}(\cdot), \pi(\cdot))$ selecting an ϵ-information maximizing control strategy satisfies

$$\mathbb{P}\big\{\pi(\hat{\theta}_{\mathrm{cp}}(\mathbf{y}_{1:n})) \in \mathit{\Pi}_\epsilon(\theta_0)\big\} \geqslant 1 - 2\exp\left\{-\frac{1}{2}\frac{c_3^2}{\frac{1}{n\breve{J}(\theta_0)} + (\hat{\theta} - \breve{\theta})c_3}\right\} \tag{4.63}$$

for all $n \in \mathbb{N}$. If $\pi(\cdot)$ is strongly robust at θ_0 then

$$n\,\mathbb{V}\big\{\hat{\theta}_{\mathrm{cp}}(\mathbf{y}_{1:n}) \,|\, \mathbf{n}\big\} \xrightarrow{\mathrm{p}} \frac{1}{J^{\mathrm{c}}_{\boldsymbol{\Xi}_{\theta_0}}(\theta_0; \mathit{\Sigma}) - \epsilon'} \tag{4.64}$$

for some $\epsilon' \in [0, \epsilon]$, where $\xrightarrow{\mathrm{p}}$ denotes convergence in probability. □

Proof. We begin by proving (4.63). Since the individual estimators are unbiased, it follows that the composite estimator is unbiased. Moreover, $\hat{\theta}_{\mathrm{cp}}(\mathbf{y}_{1:n})$ has finite variance: for any realization $\boldsymbol{n}$ of $\mathbf{n}$,

$$\mathbb{V}\big\{\hat{\theta}_{\mathrm{cp}}(\mathbf{y}_{1:n}) \,|\, \mathbf{n} = \boldsymbol{n}\big\} = \sum_{i=1}^{|\mathcal{R}|} \left(\frac{n^{(i)}}{n}\right)^2 \mathbb{V}\big\{\hat{\theta}^{(i)}(\mathbf{y}^{(i)})\big\} \tag{4.65}$$

and it follows from the additivity property of Fisher information that

$$\frac{1}{n\hat{J}(\theta_0)} \leqslant \mathbb{V}\big\{\hat{\theta}_{\mathrm{cp}}(\mathbf{y}_{1:n}) \,|\, \mathbf{n}\big\} \leqslant \frac{1}{n\breve{J}(\theta_0)}\,. \tag{4.66}$$

Lastly, note that $\hat{\theta}_{\mathrm{cp}}(\mathbf{y}_{1:n})$ is bounded since Θ is. Now, any standard concentration inequality states that $\hat{\theta}_{\mathrm{cp}}(\mathbf{y}_{1:n})$ concentrates around its mean θ with high probability. Define an open ball $B \triangleq \{\theta \in \Theta : |\theta - \theta_0| < r(\theta_0)\}$ of radius $r(\theta_0)$ centered at the true parameter θ_0 such that, if the parameter estimate $\hat{\theta}_{\mathrm{cp}}$ falls within this ball, then $\pi(\hat{\theta}_{\mathrm{cp}}) \in \mathit{\Pi}_\epsilon(\theta_0)$. Such a ball exists since $\pi(\cdot)$ is robust on Θ. Bernstein's inequality shows [103, p. 28]

$$\mathbb{P}\big\{\hat{\theta}_{\mathrm{cp}}(\mathbf{y}_{1:n}) \notin B\big\} \leqslant 2\exp\left\{-\frac{1}{2}\frac{r^2(\theta_0)}{\frac{1}{n\breve{J}(\theta_0)} + (\hat{\theta} - \breve{\theta})r(\theta_0)}\right\}. \tag{4.67}$$

Equation (4.63) follows from (4.67) and monotonicity of probability:

$$\begin{aligned}\mathbb{P}\big\{\pi(\hat{\theta}_{\mathrm{cp}}(\mathbf{y}_{1:n})) \in \mathit{\Pi}_\epsilon(\theta_0)\big\} &= 1 - \mathbb{P}\big\{\pi(\hat{\theta}_{\mathrm{cp}}(\mathbf{y}_{1:n})) \notin \mathit{\Pi}_\epsilon(\theta_0)\big\} \\ &\geqslant 1 - \mathbb{P}\big\{\hat{\theta}_{\mathrm{cp}}(\mathbf{y}_{1:n}) \notin B\big\}\,. \end{aligned} \tag{4.68}$$

To prove (4.64), let the decision rule $\pi(\cdot)$ be strongly robust at the true parameter θ_0. Without loss of generality, suppose that that the first control strategy is ϵ-information maximizing, i.e., $\pi(\theta_0) = 1$. By definition,

$$J^{(1)}(\theta_0) = J_{\psi(\Phi(t_{\mathrm{f}}^{(1)};t_0,\boldsymbol{\Xi}_{\theta_0},\boldsymbol{u}^{(1)}))}(\theta_0) = J^{\mathrm{c}}_{\boldsymbol{\Xi}_{\theta_0}}(\theta_0;\Sigma) - \epsilon' \tag{4.69}$$

Thus, proving (4.64) is tantamount to showing that, for any $\delta > 0$ and $\sigma \in (0,1]$, the inequalities

$$\eta_1 \triangleq \mathbb{P}\bigg\{\, n\mathbb{V}\big\{\hat{\theta}_{\mathrm{cp}}(\mathbf{y}_{1:n}) \,|\, \mathbf{n}\big\} - \frac{1}{J^{(1)}(\theta_0)} \leqslant +\delta \,\bigg\} \geqslant 1 - \frac{\sigma}{2} \tag{4.70a}$$

$$\eta_2 \triangleq \mathbb{P}\bigg\{\, n\mathbb{V}\big\{\hat{\theta}_{\mathrm{cp}}(\mathbf{y}_{1:n}) \,|\, \mathbf{n}\big\} - \frac{1}{J^{(1)}(\theta_0)} \geqslant -\delta \,\bigg\} \geqslant 1 - \frac{\sigma}{2} \tag{4.70b}$$

hold for sufficiently large $n \in \mathbb{N}$. We will construct $N_{\delta,\sigma} \in \mathbb{N}$ such that (4.70a) holds if $n \geqslant N_{\delta,\sigma}$. Similar argumentation shows there exists $M_{\delta,\sigma} \in \mathbb{N}$ such that (4.70b) holds if $n \geqslant M_{\delta,\sigma}$. From there, (4.64) holds since

$$\mathbb{P}\bigg\{\, \bigg| n\mathbb{V}\big\{\hat{\theta}_{\mathrm{cp}}(\mathbf{y}_{1:n}) \,|\, \mathbf{n}\big\} - \frac{1}{J^{(1)}(\theta_0)}\bigg| \leqslant \delta \,\bigg\} \geqslant 1 - \sigma \tag{4.71}$$

for $n \geqslant \max\{N_{\delta,\sigma},\, M_{\delta,\sigma}\}$.

Let $\delta > 0$ and $\sigma \in (0,1]$ be arbitrary. Observe that for any $k \in \mathbb{N} \cup \{0\}$ the series

$$S_k \triangleq \sum_{j=k+1}^{\infty} 2\exp\left\{-\frac{1}{2}\frac{r^2(\theta_0)}{\frac{1}{j\breve{J}(\theta_0)} + (\hat{\theta} - \breve{\theta})r(\theta_0)}\right\} \tag{4.72}$$

converges (this can be proven, e.g., using a ratio test). Hence, there exists $K \in \mathbb{N} \cup \{0\}$ such that $S_K \leqslant \sigma/2$. Now, take $N_{\delta,\sigma} > K$ large enough such that

$$\frac{K}{N_{\delta,\sigma}}\Big(\frac{1}{\breve{J}(\theta_0)} - \frac{1}{J^{(1)}(\theta_0)}\Big) \leqslant \delta \tag{4.73}$$

and consider $n \geqslant N_{\delta,\sigma}$. Equations (4.65) and (4.73) give

$$\begin{aligned}\mathbb{V}\big\{\,\hat{\theta}_{\mathrm{cp}}(\mathbf{y}_{1:n})\,|\,\mathbf{n},\,\mathsf{n}^{(1)} \geqslant (n-K)\,\big\} &\leqslant \frac{\mathsf{n}^{(1)}}{n^2}\frac{1}{J^{(1)}(\theta_0)} + \frac{n-\mathsf{n}^{(1)}}{n^2}\frac{1}{\breve{J}(\theta_0)} \\ &\leqslant \frac{1}{nJ^{(1)}(\theta_0)} + \frac{\delta}{n}\,. \end{aligned} \tag{4.74}$$

This proves that

$$\mathbb{P}\Big\{\, n\mathbb{V}\big\{\hat{\theta}_{\mathrm{cp}}(\mathbf{y}_{1:n})\,|\,\mathbf{n}\big\} - \frac{1}{J^{(1)}(\theta_0)} \leqslant \delta \,\Big|\, \mathsf{n}^{(1)} \geqslant n-K \,\Big\} = 1\,. \tag{4.75}$$

Following the derivation of (4.59),

$$\mathbb{P}\{\mathbf{n}^{(1)} \geqslant n-K\} \geqslant 1 - \sum_{j=K+1}^{n} \mathbb{P}\big\{\hat{\theta}_{\mathrm{cp}}(\mathbf{y}_{1:j}) \notin B\big\}\,. \tag{4.76}$$

Combining (4.75), (4.76), and (4.67) with the fact that $S_K \leqslant \sigma/2$,

$$\begin{aligned}\eta_1 &= \mathbb{P}\Big\{\, n\mathbb{V}\big\{\hat{\theta}_{\mathrm{cp}}(\mathbf{y}_{1:n})\,|\,\mathbf{n}\big\} - \frac{1}{J^{(1)}(\theta_0)} \leqslant \delta \,\Big\} \\ &\geqslant \mathbb{P}\{\mathsf{n}^{(1)} \geqslant n-K\} \\ &\geqslant 1 - \frac{\sigma}{2} \end{aligned} \tag{4.77}$$

proving that (4.70a) holds for $n \geqslant N_{\delta,\sigma}$. ⊠

The worst-case information extracted by a representative strategy in $\mathcal{R}$ is $\breve{J}(\theta_0)$. Equation (4.63) shows that the controller converges to an ϵ-information maximizing strategy quickly if $\breve{J}(\theta_0)$ is large. The following is an immediate corollary of Theorem 4.3.

Corollary 4.1 (Convergence of composite estimation-based adaptive information control)**.** Suppose that the conditions of Theorem 4.3 hold. Regardless of strong robustness of $\pi(\cdot)$ at θ_0, the following weaker version of (4.64) holds: for every $\delta > 0$,

$$\lim_{n\to\infty} \mathbb{P}\Big\{\, n\,\mathbb{V}\big\{\hat{\theta}_{\mathrm{cp}}(\mathbf{y}_{1:n})\,|\,\mathbf{n}\big\} \leqslant \frac{1}{J^{\mathrm{c}}_{\boldsymbol{\Xi}_{\theta_0}}(\theta_0;\Sigma) - \epsilon} + \delta \,\Big\} = 1\,. \tag{4.78}$$

□

Proof. The proof follows along the lines of the proof of Theorem 4.3. Suppose, without loss of generality, that the first control strategy is ϵ-information maximizing, i.e., $1 \in \mathit{\Pi}_\epsilon(\theta_0)$. Note that, by definition,

$$J_{\psi(\Phi(t_{\mathrm{f}}^{(i)};t_0,\boldsymbol{\Xi}_{\theta_0},\boldsymbol{u}^{(i)}))}(\theta_0) \geqslant J^{\mathrm{c}}_{\boldsymbol{\Xi}_{\theta_0}}(\theta_0;\Sigma) - \epsilon \quad \forall i \in \mathit{\Pi}_\epsilon(\theta_0)\,. \tag{4.79}$$

It must be shown that, for any $\delta > 0$ and $\sigma \in (0,1]$, there exists $N_{\delta,\sigma} \in \mathbb{N}$ with the property

$$\mathbb{P}\Big\{\, n\mathbb{V}\big\{\hat{\theta}_{\mathrm{cp}}(\mathbf{y}_{1:n})\,|\,\mathbf{n}\big\} - \frac{1}{J^{\mathrm{c}}_{\boldsymbol{\Xi}_{\theta_0}}(\theta_0;\Sigma) - \epsilon} \leqslant \delta \,\Big\} \geqslant 1-\sigma \tag{4.80}$$

for all $n \geqslant N_{\delta,\sigma}$. The proof follows that of Theorem 4.3 using the fact that all control strategies corresponding to the decision region $\mathit{\Pi}_\epsilon(\theta_0)$ extract information at a rate within ϵ of the CFI. ⊠

4.3.4 Interplay between estimation and control in retrospective

Cumulatively, Section 4.3 establishes a mathematical framework to estimate an unknown parameter at the fundamental limit quantified by the CFI. Motivated by the need for rapid, real-time adaptation on the timescale of quantum experiments, ϵ-coverings of information control strategies were introduced. It was shown that such coverings exist for every $\epsilon > 0$. Adaptive information control was formalized and its performance quantified. Theorems 4.2 and 4.3 demonstrate the symbiotic relationship between estimation and control: estimation enables selecting an ϵ-information maximizing control strategy with high probability, and control quenches the estimation error to within ϵ of the optimal scaling $1/(nJ^{\mathrm{c}}_{\boldsymbol{\Xi}_\theta}(\theta;\Sigma))$.

4.4 Designing and certifying information controls

Now that a foundation for adaptive information control based on ϵ-coverings has been established, we turn to the problem of designing ϵ-coverings. The approach has already been outlined in Section 4.3. Namely, a set of $N_{\mathrm{r}} \in \mathbb{N}$ representative parameter values $\Theta_{\mathrm{r}} = \{\theta_{\mathrm{r}}^{(1)}, \theta_{\mathrm{r}}^{(2)}, \ldots, \theta_{\mathrm{r}}^{(N_{\mathrm{r}})}\}$ is placed on the parameter space. The number N_{r} is chosen by the practitioner, with higher N_{r} typically leading to coverings with smaller ϵ. An approximate solution $(\boldsymbol{u}^{(i)}, t_{\mathrm{f}}^{(i)})$ is generated for each optimal control problem $\mathscr{P}(\theta_{\mathrm{r}}^{(i)})$, $i \in \mathbb{N}_1^{N_{\mathrm{r}}}$. This produces a set $\mathcal{R} = \{(\theta_{\mathrm{r}}^{(i)}, \boldsymbol{u}^{(i)}, t_{\mathrm{f}}^{(i)}) : i \in \mathbb{N}_1^{N_{\mathrm{r}}}\}$ of representative control strategies. Generating approximate solutions to the infinite-dimensional non-convex optimization problem $\mathscr{P}(\theta_{\mathrm{r}}^{(i)})$ generally requires numerical techniques. We suggest two approaches: (i) Section 4.4.1 provides a greedy information control strategy that only requires a single call to an ordinary differential equation (ODE) solver, and (ii) Section 4.4.2 provides a projected gradient ascent algorithm leveraging ideas from optimal quantum control. The ascent algorithm utilizes multiple calls to an ODE solver to iteratively refine an information control law. In both sections, we will suppose that Σ is a differential quantum control system. The set of allowed control inputs $\mathcal{V}$ is taken to be a hypercube

$$\mathcal{V} = [u_1^{\min}, u_1^{\max}] \times [u_2^{\min}, u_2^{\max}] \times \ldots \times [u_{N_{\mathrm{c}}}^{\min}, u_{N_{\mathrm{c}}}^{\max}] \subset \mathbb{R}^{N_{\mathrm{c}}} \tag{4.81}$$

and the allowed control laws Ω are measurable functions from a time interval $T \triangleq [t_0, t_1]$ into $\mathcal{V}$. The system's dynamics are taken to be of the form

$$\begin{aligned}\dot{\boldsymbol{X}}(t) &= -\tfrac{\imath}{\hbar}[\![\boldsymbol{H}(\boldsymbol{u}(t),t), \boldsymbol{X}(t)]\!]_- + \sum_{j\in\mathcal{J}} \boldsymbol{L}_j(t)\boldsymbol{X}(t)\boldsymbol{L}_j^\dagger(t) - \tfrac{1}{2}[\![\boldsymbol{L}_j^\dagger(t)\boldsymbol{L}_j(t), \boldsymbol{X}(t)]\!]_+ \\ &\triangleq F(\boldsymbol{X}(t), \boldsymbol{u}(t), t)\end{aligned} \tag{4.82}$$

where the Hamiltonian takes the form

$$\boldsymbol{H}(\boldsymbol{u}(t),t) = \boldsymbol{H}_0(t) + \sum_{\ell=1}^{N_{\mathrm{c}}} u_\ell(t)\, \boldsymbol{H}_\ell(t)\,. \tag{4.83}$$

Once the set $\mathcal{R}$ of representative control strategies has been determined, converse bounds are used to certify that $\mathcal{R}$ is an ϵ-covering. This is detailed in Section 4.4.3. The case studies in Section 4.5 demonstrate coverings with small ϵ.

4.4.1 Greedy information control strategy

The greedy information controller $(\boldsymbol{u}^{\mathrm{gr}}, t_{\mathrm{f}}^{\mathrm{gr}}) \in \Omega \times T$ is defined as follows. At any given time $t \in T$, the control law $\boldsymbol{u}^{\mathrm{gr}}$ maximizes the time derivative of the Fisher information (along the trajectory $\boldsymbol{\Xi}_\theta(t)$ of the quantum control system Σ):

$$\boldsymbol{u}^{\mathrm{gr}}(t) \in \arg\max_{\boldsymbol{u}(t)\in\mathcal{V}} \frac{\mathrm{d}}{\mathrm{d}t} J_{\psi(\boldsymbol{\Xi}_\theta(t))}(\theta) \tag{4.84}$$

for $t \in T$. Using elementary calculus and the dynamics (4.82),

$$\frac{\mathrm{d}}{\mathrm{d}t} J_{\psi(\boldsymbol{\Xi}_\theta(t))}(\theta) = C(t) - \sum_{\ell=1}^{N_{\mathrm{c}}} u_\ell(t)\gamma_\ell(t) \tag{4.85}$$

where $C(t)$ is a term that does not depend on the control $\boldsymbol{u}(t)$ and

$$\begin{aligned}\gamma_\ell(t) \triangleq \sum_{k=1}^{N_{\mathrm{m}}} & 2\left[\frac{\langle \boldsymbol{M}_k, \frac{\mathrm{d}}{\mathrm{d}\theta}\boldsymbol{\Xi}_\theta(t)\rangle_{\mathrm{HS}}}{\langle \boldsymbol{M}_k, \boldsymbol{\Xi}_\theta(t)\rangle_{\mathrm{HS}}}\right]\langle \boldsymbol{M}_k, \tfrac{\imath}{\hbar}[\![\boldsymbol{H}_\ell(t), \tfrac{\mathrm{d}}{\mathrm{d}\theta}\boldsymbol{\Xi}_\theta(t)]\!]_-\rangle_{\mathrm{HS}} \\ & -\left[\frac{\langle \boldsymbol{M}_k, \frac{\mathrm{d}}{\mathrm{d}\theta}\boldsymbol{\Xi}_\theta(t)\rangle_{\mathrm{HS}}}{\langle \boldsymbol{M}_k, \boldsymbol{\Xi}_\theta(t)\rangle_{\mathrm{HS}}}\right]^2\langle \boldsymbol{M}_k, \tfrac{\imath}{\hbar}[\![\boldsymbol{H}_\ell(t), \boldsymbol{\Xi}_\theta(t)]\!]_-\rangle_{\mathrm{HS}}\,.\end{aligned} \tag{4.86}$$

Thus, if $\gamma_\ell(t) \neq 0$, the ℓth component of the greedy control law is

$$u_\ell^{\mathrm{gr}}(t) = \begin{cases} u_\ell^{\max}, & \text{if } \gamma_\ell(t) < 0 \\ u_\ell^{\min}, & \text{if } \gamma_\ell(t) > 0 \end{cases} \tag{4.87}$$

for $t \in T$, $\ell \in \mathbb{N}_1^{N_{\mathrm{c}}}$. If $\gamma_\ell(t) = 0$ then $u_\ell^{\mathrm{gr}}(t)$ can be chosen arbitrarily from the interval $[u_\ell^{\min}, u_\ell^{\max}]$. The greedy measurement time $t_{\mathrm{f}}^{\mathrm{gr}}$ is defined as

$$t_{\mathrm{f}}^{\mathrm{gr}} \in \arg\max_{t\in T} J_{\psi(\Phi(t;t_0,\boldsymbol{\Xi}_\theta,\boldsymbol{u}^{\mathrm{gr}}))}(\theta)\,. \tag{4.88}$$

Given the explicit formula (4.87), the control law $\boldsymbol{u}^{\mathrm{gr}}$ can be calculated using an ODE solver to compute $\frac{\mathrm{d}}{\mathrm{d}\theta}\boldsymbol{\Xi}_\theta(t) = \Phi(t;t_0,\frac{\mathrm{d}}{\mathrm{d}\theta}\boldsymbol{\Xi}_\theta,\boldsymbol{u}^{\mathrm{gr}})$ and compute $\boldsymbol{\Xi}_\theta(t) = \Phi(t;t_0,\boldsymbol{\Xi}_\theta,\boldsymbol{u}^{\mathrm{gr}})$ for $t \in T$. The measurement time $t_{\mathrm{f}}^{\mathrm{gr}}$ can be determined by selecting the time $t \in T$ at which the Fisher information $J_{\psi(\Phi(t;t_0,\boldsymbol{\Xi}_\theta,\boldsymbol{u}^{\mathrm{gr}}))}(\theta)$ is maximized.

4.4.2 Projected gradient ascent algorithm

The goal of the gradient ascent algorithm is to find a control strategy that is a stationary point of the infinite-dimensional optimization problem $\mathscr{P}(\theta)$. The essence of the approach is to determine the Fréchet gradient (Definition 3.2) of $J_{\psi(\Phi(t_{\mathrm{f}};t_0,\boldsymbol{\Xi}_\theta,\boldsymbol{u}))}(\theta)$ with respect to the control strategy $(\boldsymbol{u}, t_{\mathrm{f}})$, perform a gradient update of the control strategy, and then project the update onto the set of feasible control strategies $\Omega \times T$. This process is iterated until convergence.

For the numerical implementation of the algorithm, it is useful to first perform a transformation of variables. Given $\tau \in [0,1]$, let

$$\Delta \triangleq t_{\mathrm{f}} - t_0\,, \quad t \triangleq \Delta\tau + t_0\,, \quad \boldsymbol{c}(\tau) \triangleq \boldsymbol{u}(t) \tag{4.89a}$$

$$\boldsymbol{X}_1(\tau) \triangleq \tfrac{\mathrm{d}}{\mathrm{d}\theta}\boldsymbol{\Xi}_\theta(t)\,, \quad \boldsymbol{X}_2(\tau) \triangleq \boldsymbol{\Xi}_\theta(t)\,. \tag{4.89b}$$

Note that, e.g., $\boldsymbol{X}_2(0) = \boldsymbol{\Xi}_\theta(t_0)$ and $\boldsymbol{X}_2(1) = \boldsymbol{\Xi}_\theta(t_{\mathrm{f}})$. Using (4.82) and (4.89) $\boldsymbol{X}_i(\tau)$, $i \in \{1,2\}$, evolves as

$$\dot{\boldsymbol{X}}_i(\tau) = \Delta F(\boldsymbol{X}_i(\tau), \boldsymbol{c}(\tau), \Delta\tau + t_0) \tag{4.90}$$

where the shorthand $\dot{\boldsymbol{X}}_i(\tau) \triangleq \mathrm{d}\boldsymbol{X}_i(\tau)/\mathrm{d}\tau$ is used. The advantage of this change of variables is that the transformed control law $\boldsymbol{c}(\cdot)$ and the transformed states $\boldsymbol{X}_i(\cdot)$ are defined on the unit interval $[0,1]$ regardless of the scaling parameter $\Delta \in [0, t_1 - t_0] \triangleq \bar{T}$. Without this change of variables, the domain of the control and state vary as the measurement time t_{f} is optimized. Now, the optimal control problem $\mathscr{P}(\theta)$ can be rewritten in the form

$$\begin{aligned}
&\underset{(\boldsymbol{c},\Delta)\,\in\,\bar{\Omega}\times\bar{T}}{\text{supremize}} && J_{\psi(\boldsymbol{X}_2(1))}(\theta) \\
&\text{subject to} && \dot{\boldsymbol{X}}_i(\tau) = \Delta F(\boldsymbol{X}_i(\tau), \boldsymbol{c}(\tau), \Delta\tau + t_0)\,, \\
&&& \boldsymbol{X}_1(0) = \tfrac{\mathrm{d}}{\mathrm{d}\theta}\boldsymbol{\Xi}_\theta\,, \quad \boldsymbol{X}_2(0) = \boldsymbol{\Xi}_\theta\,, \\
&&& i \in \{1,2\}\,, \quad \tau \in [0,1]
\end{aligned}$$

where $\bar{\Omega}$ is the set of time-normalized control laws: measurable functions from $[0,1]$ into $\mathcal{V}$.

To implement a gradient ascent algorithm, gradients of $J_{\psi(\boldsymbol{X}_2(1))}(\theta)$ at the control-measurement time pair $(\boldsymbol{c}, \Delta)$ must be calculated. Towards this end, given a pair of trajectories $\boldsymbol{X}_1(\tau)$ and $\boldsymbol{X}_2(\tau)$ of (4.90) generated by $(\boldsymbol{c}, \Delta)$, define the costates $\boldsymbol{\Lambda}_i(\tau) \in \mathbb{C}^{d\times d}$, $i \in \{1,2\}$, as the solution to the final value problem

$$\dot{\boldsymbol{\Lambda}}_i(\tau) = -\Delta F^{\ddagger}(\boldsymbol{\Lambda}_i(\tau), \boldsymbol{c}(\tau), \Delta\tau + t_0)\,, \tag{4.91a}$$

$$\boldsymbol{\Lambda}_i(1) = \nabla_{\boldsymbol{X}_i(1)} \sum_{k=1}^{N_{\mathrm{m}}} \frac{\langle \boldsymbol{M}_k, \boldsymbol{X}_1(1)\rangle_{\mathrm{HS}}^2}{\langle \boldsymbol{M}_k, \boldsymbol{X}_2(1)\rangle_{\mathrm{HS}}}\,. \tag{4.91b}$$

The costates enable the calculation of the aforementioned gradient using the ideas developed in Chapter 3

Theorem 4.4 ((Informal) Fréchet gradients of Fisher information functional)**.** Let F be continuously differentiable with respect to the time variable. The gradients of $J_{\psi(\boldsymbol{X}_2(1))}(\theta)$ at $(\boldsymbol{c}, \Delta) \in L^2([0,1]; \mathbb{R}^{N_{\mathrm{c}}}) \times \mathbb{R}_{>0}$ with respect to $\boldsymbol{c}$ and Δ are given, respectively, by

$$g_\ell(\tau) = -\tfrac{\imath}{\hbar}\Delta \sum_{i=1}^{2} \mathrm{Re}\langle \boldsymbol{\Lambda}_i(\tau), [\![\boldsymbol{H}_\ell(\Delta\tau + t_0), \boldsymbol{X}_i(\tau)]\!]_-\rangle_{\mathrm{HS}} \qquad \forall (\tau, \ell) \in [0,1] \times \mathbb{N}_1^{N_{\mathrm{c}}} \tag{4.92a}$$

$$h = \int_0^1 \sum_{i=1}^{2} \mathrm{Re}\Big\langle \boldsymbol{\Lambda}_i(\tau), F(\boldsymbol{X}_i(\tau), \boldsymbol{c}(\tau), \Delta\tau + t_0) + \Delta \frac{\partial}{\partial t} F(\boldsymbol{X}_i(\tau), \boldsymbol{c}(\tau), t)\Big|_{t=\Delta\tau+t_0} \Big\rangle_{\mathrm{HS}} \mathrm{d}\tau. \tag{4.92b}$$

□

Now, using projected gradient ascent, control laws can be iteratively refined to maximize $J_{\psi(\boldsymbol{X}_2(1))}(\theta)$. The projection $\mathrm{Proj}_{\bar{\Omega}}(\boldsymbol{c})$ of a measurable function $\boldsymbol{c} : [0,1] \to \mathbb{R}^{N_\mathrm{c}}$ onto the set $\bar{\Omega}$ is defined according to the rule

$$c_\ell(\tau) \mapsto (u_\ell^{\max} \wedge c_\ell(\tau)) \vee u_\ell^{\min} \tag{4.93}$$

for each $\tau \in [0,1]$ and for each component $c_\ell(\tau)$ of $\boldsymbol{c}(\tau)$, $\ell \in \mathbb{N}_1^{N_\mathrm{c}}$. Similarly, the projection $\mathrm{Proj}_{\bar{T}}(\Delta)$ of $\Delta \in \mathbb{R}$ onto the interval $\bar{T}$ is defined according to the rule

$$\Delta \mapsto ((t_1 - t_0) \wedge \Delta) \vee 0\,. \tag{4.94}$$

Projected gradient ascent updates a control strategy $(\boldsymbol{c}, \Delta) \in \Omega \times [0, t_1 - t_0]$ as

$$\boldsymbol{c} \mapsto \operatorname*{Proj}_{\bar{\Omega}}(\boldsymbol{c} + \gamma\, \boldsymbol{g}) \tag{4.95a}$$

$$\Delta \mapsto \operatorname*{Proj}_{\bar{T}}(\Delta + \gamma\, h) \tag{4.95b}$$

where $\gamma > 0$ is the stepsize. The projection step ensures that the values of the updated control law lie inside the set $\mathcal{V}$, which could represent physical limitations on a control signal generator, and that the values of Δ lie within the physical range $\bar{T}$. Implementing the update rule (4.95) requires four parallelizable calls to an ODE solver, one to solve for each of $\boldsymbol{X}_1(\cdot)$, $\boldsymbol{X}_2(\cdot)$, $\boldsymbol{\Lambda}_1(\cdot)$, and $\boldsymbol{\Lambda}_2(\cdot)$.

4.4.3 Certification of representative control strategies

Let $\mathcal{R} = \{(\theta_\mathrm{r}^{(i)}, \boldsymbol{u}^{(i)}, t_\mathrm{f}^{(i)}) : i \in \mathbb{N}_1^{N_\mathrm{r}}\}$ be a set of representative control strategies. To certify that $\mathcal{R}$ is an ϵ-covering, a converse bound is needed, i.e., a function $b : \Theta \to \mathbb{R}$ such that

$$J^{\mathrm{c}}_{\boldsymbol{\Xi}_\theta}(\theta; \Sigma) \leqslant b(\theta) \qquad \forall \theta \in \Theta\,. \tag{4.96}$$

Given such a converse bound, the set $\mathcal{R}$ is an ϵ-covering if

$$b(\theta) - \max_{i \in \mathbb{N}_1^{N_\mathrm{r}}} J_{\psi(\Phi(t_\mathrm{f}^{(i)}; t_0, \boldsymbol{\Xi}_\theta, \boldsymbol{u}^{(i)}))}(\theta) \leqslant \epsilon \qquad \forall \theta \in \Theta\,. \tag{4.97}$$

Once it has been established that $\mathcal{R}$ is an ϵ-covering, a decision rule $\pi(\cdot)$ (e.g., the nearest-neighbor decision rule (4.27b)) is robust on Θ if for each $\theta \in \Theta$ there exists $r(\theta) > 0$ with the property

$$b(\theta) - J_{\psi(\Phi(t_\mathrm{f}^{(\pi(\theta'))}; t_0, \boldsymbol{\Xi}_\theta, \boldsymbol{u}^{(\pi(\theta'))}))}(\theta) \leqslant \epsilon \tag{4.98}$$

for all $\theta' \in \Theta$ such that $|\theta' - \theta| < r(\theta)$.

Converse bounds generally need to be developed on a case-by-case basis leveraging the specific structure of a given quantum system. Lemma 4.2 proved that QFI is a universal converse bound on the CFI; however, this bound can be loose since it is agnostic to Σ. This is demonstrated in Section 4.5 by producing tight converse bounds $b(\cdot)$ for particular systems such that

$$J^{\mathrm{c}}_{\boldsymbol{\Xi}_\theta}(\theta; \Sigma) \leqslant b(\theta) < J^{\mathrm{q}}_{\boldsymbol{\Xi}_\theta}(\theta) \qquad \forall \theta \in \Theta\,. \tag{4.99}$$

Relations like (4.99) demonstrate that the CFI is, in general, a more accurate metric on the amount of information that can be extracted from a quantum system equipped with a fixed measurement apparatus.

4.5 Case studies: Information control for quantum estimation

Three case studies will be presented. The goal of these case studies is to reveal the breadth of effects that parameter encoding has on the CFI, ϵ-coverings, and adaptive information control.

4.5.1 Linear parameter encoding

Let us begin by considering a simple scenario. We have a quantum control system Σ on a finite-dimensional space $\mathcal{H} = \mathbb{C}^d$. The parameter θ is linearly encoded into the state $\boldsymbol{\Xi}_\theta \in \mathcal{D}(\mathcal{H})$ as $\boldsymbol{\Xi}_\theta = \frac{1}{d}\boldsymbol{I} + \theta \boldsymbol{Z}$, where $\boldsymbol{Z} \in \mathbb{C}^{d\times d}$ is traceless. This scenario has applications to depolarization estimation and state estimation, which are studied subsequently. A useful converse bound has been developed in [2]. To present the result, let $\boldsymbol{M} \in \mathbb{C}^{d\times d}$ such that $\boldsymbol{0} \preccurlyeq \boldsymbol{M} \preccurlyeq \boldsymbol{I}$, $\boldsymbol{M} \neq \boldsymbol{0}$, and $\boldsymbol{M} \neq \boldsymbol{I}$. Define

$$z_{\min} \triangleq \sum_{k=1}^{d} \lambda_k^{\uparrow}(\boldsymbol{M})\lambda_k^{\downarrow}(\boldsymbol{Z}) \tag{4.100a}$$

$$z_{\max} \triangleq \sum_{k=1}^{d} \lambda_k^{\downarrow}(\boldsymbol{M})\lambda_k^{\downarrow}(\boldsymbol{Z}) \tag{4.100b}$$

where $\lambda_k^{\uparrow}(\boldsymbol{M})$ and $\lambda_k^{\downarrow}(\boldsymbol{M})$ denote the kth smallest and largest eigenvalues of $\boldsymbol{M}$, respectively. Moreover, define the scalar function $g : \Theta \times \mathbb{R} \to \mathbb{R}$ according to

$$g(\theta; z) \triangleq \frac{z^2}{\eta(1-\eta) + \theta\,(1-2\eta)z - \theta^2 z^2} \tag{4.101}$$

where $\eta \triangleq \operatorname{tr}\{\boldsymbol{M}\}/d \in (0, 1)$.

Theorem 4.5 (Linear parameter encoding converse bound, Clouâtré et al. [2])**.** Consider a unital quantum control system Σ endowed with the POVM $\mathcal{M} = \{\boldsymbol{M}, \boldsymbol{I} - \boldsymbol{M}\}$ and the parameterized density matrix $\boldsymbol{\Xi}_\theta = \frac{1}{d}\boldsymbol{I} + \theta\boldsymbol{Z}$. The following converse bound on the CFI holds:

$$J^{\mathrm{c}}_{\boldsymbol{\Xi}_\theta}(\theta; \Sigma) \leqslant \max\big\{ g(\theta; z_{\min}),\, g(\theta; z_{\max}) \big\}\,. \tag{4.102}$$

□

Depolarization estimation

Consider a qubit system ($\mathcal{H} = \mathbb{C}^2$) subject to dephasing noise. This is mathematically described by a differential quantum control system Σ with dynamics (4.82) where

$$\boldsymbol{H}_0 = -\tfrac{\hbar\omega_0}{2}\,\boldsymbol{P}_z\,, \quad \boldsymbol{H}_1 = \hbar\boldsymbol{P}_y\,, \text{ and } \quad \boldsymbol{L} = \sqrt{\gamma}\,\boldsymbol{P}_z\,. \tag{4.103}$$

Take the qubit frequency to be $\omega_0/2\pi = 5\,\mathrm{GHz}$ and the strength of the dephasing noise to be $\gamma = 1/(2\tau_{\mathrm{dep}})$. Here, τ_{dep} is the pure dephasing time, which is typically on the order of a few microseconds (worst-case) or hundreds of microseconds (best-case). Take the set

of control inputs to be $\mathcal{V} = [u_1^{\min}, u_1^{\max}]$ and the time interval to be $T = [t_0, t_1]$ where $u_1^{\max}/(2\pi) = -u_1^{\min}/(2\pi) = 1\,\text{GHz}$, $t_0 = 0$ s, and $t_1 = 2\,\text{ns}$. This model and parameters are typical of, e.g., a microwave-driven superconducting qubit system [105]. The relaxation time (commonly referred to as the "T_1 time") of such systems can be on the order of hundreds of microseconds, so relaxation is inconsequential on the scale of T. The POVM $\mathcal{M} = \{\boldsymbol{M}, \boldsymbol{I} - \boldsymbol{M}\}$ is taken to be an imperfect measurement in the computational basis:

$$\boldsymbol{M} = \begin{bmatrix} 1 - p_{10} & 0 \\ 0 & p_{01} \end{bmatrix}, \quad \boldsymbol{I} - \boldsymbol{M} = \begin{bmatrix} p_{10} & 0 \\ 0 & 1 - p_{01} \end{bmatrix} \tag{4.104}$$

where p_{10} (resp. p_{01}) is the probability of inferring $|1\rangle$ given $|0\rangle$ (resp. inferring $|0\rangle$ given $|1\rangle$).[7] These errors are due, in part, to the qubit's interaction with its environment during the measurement process. The typical operating temperatures of superconducting qubits make the qubits more likely to lose energy to their environment than to gain [105], so generally $p_{01} > p_{10}$. For concreteness, consider $p_{01} = 0.03$ and $p_{10} = 0.02$. Note that, by Theorem 2.5, the differential system Σ described by (4.103) is unital.

Suppose that a nominal system state $\boldsymbol{\Xi} \in \mathcal{D}(\mathcal{H})$ undergoes a depolarization process—perhaps due to crosstalk from neighboring qubit gates or random gate errors accumulated over a large computation. The state $\boldsymbol{\Xi}$ after the depolarization process is

$$\boldsymbol{\Xi}_\theta = \tfrac{1}{d}\boldsymbol{I} + \theta \underbrace{(\boldsymbol{\Xi} - \tfrac{1}{d}\boldsymbol{I})}_{\triangleq \boldsymbol{Z}} \tag{4.105}$$

where $(1 - \theta)$ is the depolarization rate, $\theta \in [0, 1]$. For concreteness, consider $\boldsymbol{\Xi} = |+\rangle\langle+|$ where $|+\rangle = (|0\rangle + |1\rangle)/\sqrt{2}$. The goal is to infer θ.

A control strategy $(u^{(1)}, t_{\text{f}}^{(1)})$ is generated using the greedy information controller (4.87) and using projected gradient ascent (see Section 4.4.2) based on a representative value of the parameter $\theta_{\text{r}}^{(1)} = 0.5$. Fig. 4.1a depicts the corresponding representative control strategy for the case $\tau_{\text{dep}} = 5\,\mu\text{s}$. The optimized measurement time for the greedy information controller is $t_{\text{f}}^{(1)} = 2\,\text{ns}$ and for the optimized controller is $t_{\text{f}}^{(1)} = 1.4546\,\text{ns}$. The greedy controller achieves a $(u^{(1)}, t_{\text{f}}^{(1)})$-CFI of approximately 1.0021, while the optimized controller achieves a $(u^{(1)}, t_{\text{f}}^{(1)})$-CFI of approximately 1.1793. Fig. 4.1b depicts the optimized $(u^{(1)}, t_{\text{f}}^{(1)})$-CFI in comparison to the converse bound (4.102) for various depolarizing times τ_{dep}. For the most physically-realistic depolarizing time, $\tau_{\text{dep}} = 5\,\mu\text{s}$, the maximum difference between the $(u^{(1)}, t_{\text{f}}^{(1)})$-CFI and the converse bound is less than 0.2195, which reveals the CFI to high-accuracy. Moreover, this shows that $\mathcal{R} = \{(\theta_{\text{r}}^{(1)}, u^{(1)}, t_{\text{f}}^{(1)})\}$ is an ϵ-covering of cardinality 1, with $\epsilon = 0.2195$. For large (resp. small) values of θ, there is roughly a 70% (resp. 10%) difference between the QFI and CFI in the case $\tau_{\text{dep}} = 5\,\mu\text{s}$, indicating the looseness of the QFI.

Consider now estimator design. Suppose that n instances of the state $\boldsymbol{\Xi}_\theta$ are measured using the control strategy $(u^{(1)}, t_{\text{f}}^{(1)})$, resulting in random measurement outcomes $\mathbf{y}_{1:n}$. For convenience, let $\mathsf{y} = +1$ if the outcome associated with POVM element $\boldsymbol{M}$ is recorded,

[7] Here, $|0\rangle$ and $|1\rangle$ are the standard basis vectors in $\mathbb{C}^2$.

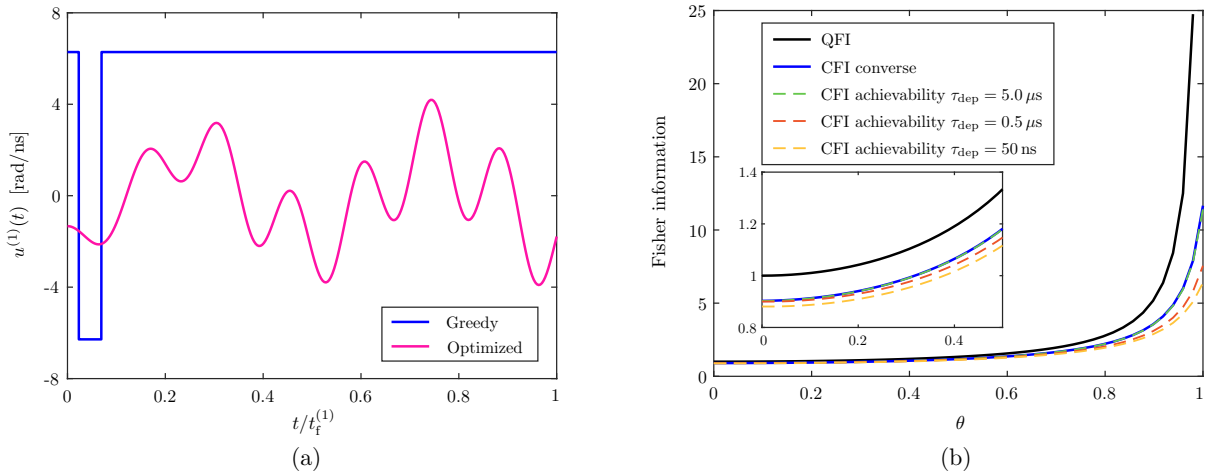


Figure 4.1: Estimating the depolarization of a superconducting qubit. For the case $\tau_{\rm dep} = 5\,\mu$s, (a) depicts the information control law $u^{(1)}(t)$ for the representative parameter value $\theta_{\rm r}^{(1)} = 0.5$ generated by the greedy information control strategy and the gradient ascent algorithm. The optimized measurement time for these two methods are $t_{\rm f}^{(1)} = 2\,$ns and $t_{\rm f}^{(1)} = 1.4546\,$ns, respectively. For various values of $\tau_{\rm dep}$, (b) depicts the QFI, the CFI converse (4.102), and the $(u^{(1)}, t_{\rm f}^{(1)})$-CFI achieved through gradient ascent. Tight CFI converse and achievability results reveal the CFI to high accuracy, e.g., in the case $\tau_{\rm dep} = 5\,\mu$s the CFI is known within precision 0.2195. In this case, the CFI is roughly 70% (resp. 10%) tighter than the QFI for large (resp. small) parameter values.

and let $\mathsf{y} = -1$ if the outcome associated with POVM element $\boldsymbol{I} - \boldsymbol{M}$ is recorded. It is straightforward to show that the estimator

$$\hat{\theta}(\mathsf{y}_{1:n}) = \frac{b + \frac{1}{n}\sum_{i=1}^{n} \mathsf{y}_i}{2\langle \boldsymbol{M}, \Phi(t_{\rm f}^{(1)}; t_0, \boldsymbol{Z}, \boldsymbol{u}^{(1)})\rangle_{\rm HS}} \tag{4.106}$$

has variance

$$\mathbb{V}\{\hat{\theta}(\mathsf{y}_{1:n})\} = \frac{1}{n\, J_{\psi(\Phi(t_{\rm f}^{(1)}; t_0, \boldsymbol{\Xi}_\theta, \boldsymbol{u}^{(1)}))}(\theta)} \tag{4.107a}$$

$$\leqslant \frac{1}{n\,(J^{\rm c}_{\boldsymbol{\Xi}_\theta}(\theta; \Sigma) - \epsilon)}\,. \tag{4.107b}$$

Equality (4.107a) shows that the estimator's variance achieves the information inequality (4.3) corresponding to the $(u^{(1)}, t_{\rm f}^{(1)})$-CFI. Inequality (4.107b) shows that (in the case $\tau_{\rm dep} = 5\,\mu$s) the estimator's variance is at most a factor of $\epsilon = 0.2195$ away from the fundamental limit, which is determined by the information inequality corresponding to the CFI. Since the ϵ-covering $\mathcal{R}$ is of cardinality 1, adaptive information control is not needed in this example.

State estimation

Consider the same dephasing qubit system as before with linear parameter encoding. However, instead of $\theta \in [0, 1]$ as in the prior example, θ can take on positive or negative values. For example, a density matrix of the form $\boldsymbol{\Xi}_\theta = \frac{1}{d}\boldsymbol{I} + \theta\, \boldsymbol{Z}$ could arise in quantum state estimation

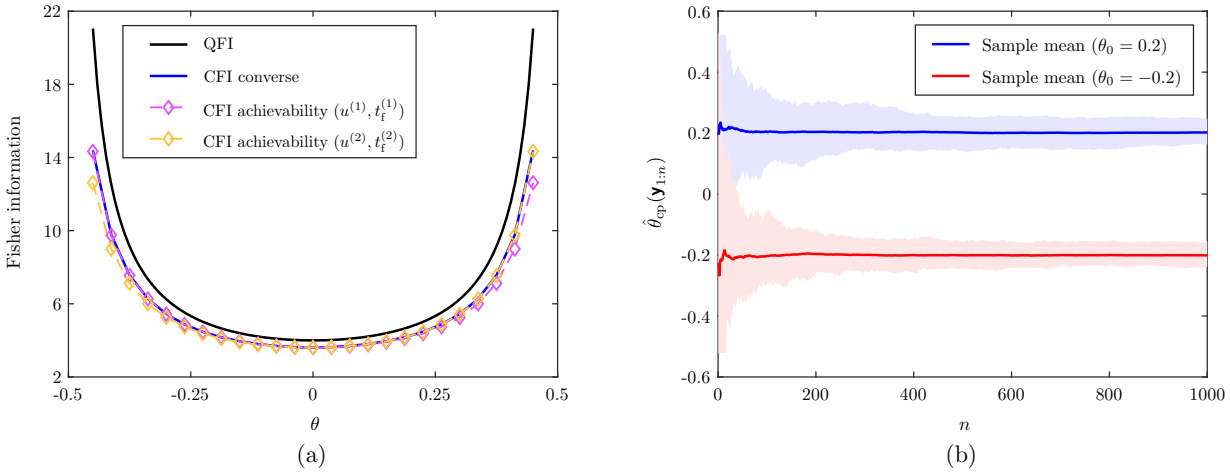


Figure 4.2: State estimation case study. (a) Depicts the QFI, the CFI converse (4.102), and the $(u^{(i)}, t_{\mathrm{f}}^{(i)})$-CFI achieved by the two representative control strategies. Note that the CFI converse and achievability bounds are tight; the bound provided by the QFI is loose. (b) Depicts Monte Carlo simulations of the composite estimator $\hat{\theta}_{\mathrm{cp}}(\mathbf{y}_{1:n})$ for two different values of the true parameter θ_0. The dark blue (resp. red) line denotes the sample mean of the Monte Carlo simulation outcomes in the case that $\theta_0 = 0.2$ (resp. $\theta_0 = -0.2$). The light blue and light red shaded regions denote the range of the simulation outcomes.

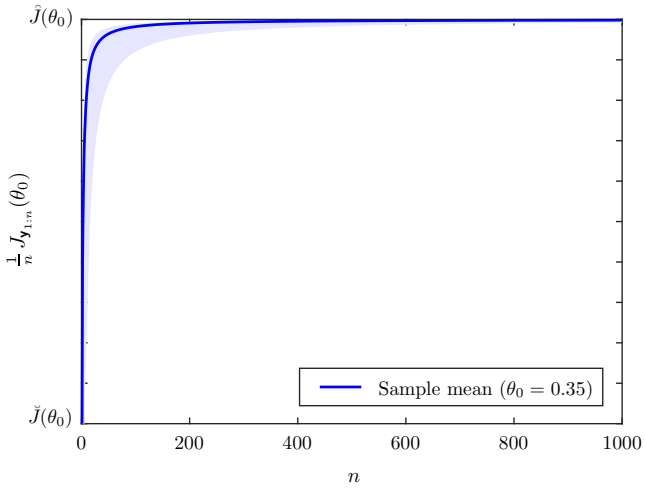


Figure 4.3: State estimation case study. The average Fisher information extracted by the adaptive information controller $(\mathcal{R}, \hat{\theta}_{\mathrm{cp}}(\cdot), \pi_{\mathrm{n}}(\cdot))$ as a function of number of experiments n in the case that $\theta_0 = 0.35$. The dark blue line denotes the sample mean of the Monte Carlo simulation outcomes; the light blue shaded area denotes the range of the simulation outcomes. The vertical axis of the plot scales linearly between the worst-case Fisher information 6.32 and the best-case Fisher information 6.63.

where $\boldsymbol{Z} = \boldsymbol{P}_x$, $\boldsymbol{P}_y$, or $\boldsymbol{P}_z$. Suppose $\boldsymbol{Z} = \boldsymbol{P}_y$ and $\theta \in [-0.45, 0.45]$. Using projected gradient ascent, two control strategies were generated for the representative parameter values $\theta_{\mathrm{r}}^{(i)} = \mp 0.45$, $i \in \{1, 2\}$. This forms a set $\mathcal{R}$ of representative information control strategies. The $(u^{(i)}, t_{\mathrm{f}}^{(i)})$-CFIs as a function of θ are plotted against the converse bound (4.102) in Fig. 4.2a. Observe that for $\theta < 0$, the first representative control strategy extracts more information, and vice-versa for $\theta > 0$. As the true parameter is learned through adaptation, an adaptive information controller will select the information maximizing control strategy. At any point θ, the maximum Fisher information $\hat{J}(\theta)$ of the representative strategies is within 0.06 of the converse bound, showing that $\mathcal{R}$ is an ϵ-covering of cardinality 2, with $\epsilon = 0.06$. Moreover, by inspection, the nearest neighbor decision rule $\pi_{\mathrm{n}}(\cdot)$ is strongly robust across the parameter space. This is clear from Fig. 4.2a because the first representative control strategy (designed for $\theta_{\mathrm{r}}^{(1)} = -0.45$) is optimal for $\theta \leqslant 0$ and the second representative control strategy (designed for $\theta_{\mathrm{r}}^{(2)} = 0.45$) is optimal for $\theta \geqslant 0$.

The composite estimator (4.62) will be used based on the individual estimators

$$\hat{\theta}^{(1)}(\mathbf{y}^{(1)}) \triangleq \frac{b + \frac{1}{\mathsf{n}^{(1)}} \sum_{i=1}^{\mathsf{n}^{(1)}} \mathsf{y}_i^{(1)}}{2\langle \boldsymbol{M}, \Phi(t_{\mathrm{f}}^{(1)}; t_0, \boldsymbol{Z}, u^{(1)})\rangle_{\mathrm{HS}}} \tag{4.108a}$$

$$\hat{\theta}^{(2)}(\mathbf{y}^{(2)}) \triangleq \frac{b + \frac{1}{\mathsf{n}^{(2)}} \sum_{i=1}^{\mathsf{n}^{(2)}} \mathsf{y}_i^{(2)}}{2\langle \boldsymbol{M}, \Phi(t_{\mathrm{f}}^{(2)}; t_0, \boldsymbol{Z}, u^{(2)})\rangle_{\mathrm{HS}}} \tag{4.108b}$$

which are efficient for experiments performed with the first and second control strategies, respectively. By Theorem 4.3, for any $-0.45 \leqslant \theta \leqslant 0.45$, the adaptive information controller $(\mathcal{R}, \hat{\theta}_{\mathrm{cp}}(\cdot), \pi_{\mathrm{n}}(\cdot))$ will select an information maximizing control strategy with high probability as $n \to \infty$ and the variance of $\hat{\theta}_{\mathrm{cp}}(\mathbf{y}_{1:n})$ approaches (within ϵ) the fundamental limit determined by the CFI. Fig. 4.2b demonstrates one-hundred Monte Carlo simulations of the adaptive information controller $(\mathcal{R}, \hat{\theta}_{\mathrm{cp}}(\cdot), \pi_{\mathrm{n}}(\cdot))$ in the case that the true parameter is $\theta_0 = 0.2$ or $\theta_0 = -0.2$. The solid lines depict the sample mean of the Monte Carlo simulation outcomes, whereas the shaded areas depict the range of parameter estimates across the outcomes. As the number of experiments n increases, the uncertainty about the true parameter vanishes as expected. Fig. 4.3 demonstrates the average Fisher information extracted by the adaptive information controller $(\mathcal{R}, \hat{\theta}_{\mathrm{cp}}(\cdot), \pi_{\mathrm{n}}(\cdot))$ in the case that the true parameter is $\theta_0 = 0.35$. The average Fisher information lies between the lower bound $\check{J}(\theta) = 6.32$ and upper bound $\hat{J}(\theta) = 6.63$.

It should be noted that $\langle \boldsymbol{M}, \Phi(t_{\mathrm{f}}^{(1)}; t_0, \boldsymbol{Z}, u^{(1)})\rangle_{\mathrm{HS}}$ and $\langle \boldsymbol{M}, \Phi(t_{\mathrm{f}}^{(2)}; t_0, \boldsymbol{Z}, u^{(2)})\rangle_{\mathrm{HS}}$ can be computed prior to experimentation. Hence, calculating (4.108a), (4.108b), and (4.62) only requires basic arithmetic operations. Such operations can be quickly executed on a FPGA like those used for quantum feedback control, meaning that the adaptive information controller $(\mathcal{R}, \hat{\theta}_{\mathrm{cp}}(\cdot), \pi_{\mathrm{n}}(\cdot))$ is able to operate at the timescale required for quantum experiments.

4.5.2 Nonlinear parameter encoding

This section considers magnetic field sensing with negatively charged nitrogen vacancy (NV) centers in diamond.[8] For introduction to NV-based sensing or its various applications, see [106, 107, 108, 109, 110]. Before proceeding, a useful converse bound will be derived.

Let $\mathcal{H} = \mathbb{C}^d$ and let $\boldsymbol{P} \in \mathbb{C}^{d\times d}$ be a rank-1 projection. Then, consider the POVM system $\mathcal{M} = \{\boldsymbol{M}, \boldsymbol{I} - \boldsymbol{M}\}$ where

$$\boldsymbol{M} = (1 - p_{\mathrm{e}})\boldsymbol{P} + \frac{p_{\mathrm{e}}}{d-1}(\boldsymbol{I} - \boldsymbol{P}) \tag{4.109}$$

and $p_{\mathrm{e}} \in (0, 1 - \frac{1}{d})$. This POVM system models the situation where a quantum state is measured with the projector system $\{\boldsymbol{P}, \boldsymbol{I} - \boldsymbol{P}\}$ and the measurement outcome is subject to random errors. Namely, with probability p_{e}, the outcome associated with $\boldsymbol{P}$ is mistaken for that associated with $\boldsymbol{I} - \boldsymbol{P}$. This type of noisy measurement arises in optical detection where fluorescence is used to distinguish between a ground state and an excited manifold.

Theorem 4.6 (Nonlinear parameter encoding converse bound)**.** Let Σ be a unital quantum control system endowed with the POVM $\mathcal{M} = \{\boldsymbol{M}, \boldsymbol{I} - \boldsymbol{M}\}$, where $\boldsymbol{M}$ is given in (4.109). The following converse bound on the CFI holds:

$$J^{\mathrm{c}}_{\boldsymbol{\Xi}_\theta}(\theta; \Sigma) \leqslant \operatorname{tr}\{\boldsymbol{\Upsilon}_\theta \boldsymbol{R}_\theta^2\} \tag{4.110}$$

where $\boldsymbol{\Upsilon}_\theta \triangleq \frac{dp_{\mathrm{e}}}{d-1}\frac{1}{d}\boldsymbol{I} + (1 - \frac{dp_{\mathrm{e}}}{d-1})\boldsymbol{\Xi}_\theta$ and $\boldsymbol{R}_\theta \in \mathbb{C}^{d\times d}$ is the solution to the Lyapunov equation

$$2\frac{\mathrm{d}}{\mathrm{d}\theta}\boldsymbol{\Upsilon}_\theta = \boldsymbol{\Upsilon}_\theta \boldsymbol{R}_\theta + \boldsymbol{R}_\theta \boldsymbol{\Upsilon}_\theta\,. \tag{4.111}$$

□

Proof. Take Σ and $\mathcal{M}$ as in the statement of the theorem. Take $\mathcal{P} = \{\boldsymbol{P}_k \,:\, k \in \mathbb{N}_1^d\}$ to be any POVM system consisting of rank-1 projections such that $\boldsymbol{P}_1 = \boldsymbol{P}$. Define another POVM system $\mathcal{S} = \{\boldsymbol{S}_k \,:\, k \in \mathbb{N}_1^d\}$ such that

$$\boldsymbol{S}_k \triangleq (1 - p_{\mathrm{e}})\boldsymbol{P}_k + \frac{p_{\mathrm{e}}}{d-1}(\boldsymbol{I} - \boldsymbol{P}_k) \qquad \forall k \in \mathbb{N}_1^d\,. \tag{4.112}$$

Moreover, define quantum control systems Σ_{p} and Σ_{s} as the system Σ with $\mathcal{M}$ replaced with $\mathcal{P}$ and $\mathcal{S}$, respectively. Denote the associated output maps by ψ_{p} and ψ_{s}.

First, note that $\mathcal{M}$ is a post-processing of $\mathcal{S}$. Hence, by Lemma 4.3, it follows that $J^{\mathrm{c}}_{\boldsymbol{\Xi}_\theta}(\theta; \Sigma) \leqslant J^{\mathrm{c}}_{\boldsymbol{\Xi}_\theta}(\theta; \Sigma_{\mathrm{s}})$.

Next, an inequality between $J^{\mathrm{c}}_{\boldsymbol{\Xi}_\theta}(\theta; \Sigma_{\mathrm{s}})$ and $J^{\mathrm{q}}_{\boldsymbol{\Upsilon}_\theta}(\theta)$ is established. Since Σ is unital, for any $(\boldsymbol{u}, t_{\mathrm{f}}) \in \Omega \times T$, it follows that $\Phi^{\ddagger}(t_{\mathrm{f}}; t_0, \cdot, \boldsymbol{u}) : \mathbb{C}^{d\times d} \to \mathbb{C}^{d\times d}$ is a unital and trace-preserving map. Hence, for any $k \in \mathbb{N}_1^d$,

$$\begin{aligned}
&\langle \Phi^{\ddagger}(t_{\mathrm{f}}; t_0, \boldsymbol{S}_k, \boldsymbol{u}), \boldsymbol{\Xi}_\theta \rangle_{\mathrm{HS}} \\
&= (1 - p_{\mathrm{e}}) \operatorname{tr}\big\{\Phi^{\ddagger}(t_{\mathrm{f}}; t_0, \boldsymbol{P}_k, \boldsymbol{u})\boldsymbol{\Xi}_\theta\big\} + \tfrac{p_{\mathrm{e}}}{d-1} \operatorname{tr}\big\{\Phi^{\ddagger}(t_{\mathrm{f}}; t_0, \boldsymbol{I}, \boldsymbol{u})\boldsymbol{\Xi}_\theta\big\} - \tfrac{p_{\mathrm{e}}}{d-1} \operatorname{tr}\big\{\Phi^{\ddagger}(t_{\mathrm{f}}; t_0, \boldsymbol{P}_k, \boldsymbol{u})\boldsymbol{\Xi}_\theta\big\}
\end{aligned}$$

[8] As is standard in the literature, we abbreviate NV without adding a superscript to indicate negative charge. Throughout this chapter, it is understood that NV refers to NV^-.

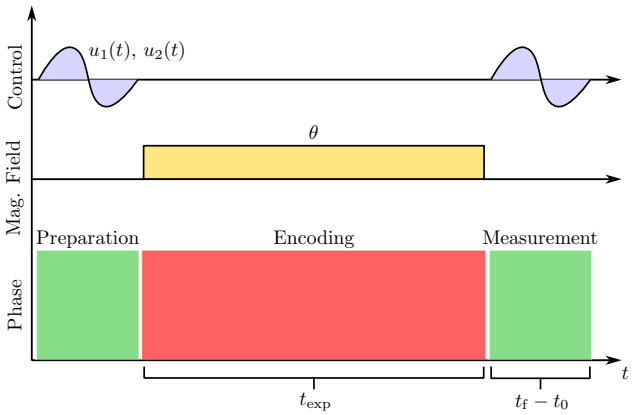


Figure 4.4: Visualization of magnetic field sensing. The three phases of the sensing process: preparation, encoding, and measurement. The encoding phase occurs over a long exposure interval of duration t_{exp}. The measurement phase occurs over the shorter time interval of maximum duration $t_{\text{f}} - t_0$. Control is leveraged in the preparation and measurement phases. The effect of the weak magnetic field is tangible only over the long exposure interval.

$$
\begin{aligned}
&= (1-p_{\text{e}})\operatorname{tr}\big\{\Phi^{\ddagger}(t_{\text{f}};t_0,\boldsymbol{P}_k,\boldsymbol{u})\boldsymbol{\Xi}_\theta\big\} + \tfrac{p_{\text{e}}}{d-1} - \tfrac{p_{\text{e}}}{d-1}\operatorname{tr}\big\{\Phi^{\ddagger}(t_{\text{f}};t_0,\boldsymbol{P}_k,\boldsymbol{u})\boldsymbol{\Xi}_\theta\big\} \\
&= (1-p_{\text{e}})\operatorname{tr}\big\{\Phi^{\ddagger}(t_{\text{f}};t_0,\boldsymbol{P}_k,\boldsymbol{u})\boldsymbol{\Xi}_\theta\big\} + \tfrac{p_{\text{e}}}{d-1}\operatorname{tr}\big\{\Phi^{\ddagger}(t_{\text{f}};t_0,\boldsymbol{P}_k,\boldsymbol{u})\big\} - \tfrac{p_{\text{e}}}{d-1}\operatorname{tr}\big\{\Phi^{\ddagger}(t_{\text{f}};t_0,\boldsymbol{P}_k,\boldsymbol{u})\boldsymbol{\Xi}_\theta\big\} \\
&= \operatorname{tr}\big\{\Phi^{\ddagger}(t_{\text{f}};t_0,\boldsymbol{P}_k,\boldsymbol{u})\big(\tfrac{dp_{\text{e}}}{d-1}\tfrac{1}{d}\boldsymbol{I} + (1-\tfrac{dp_{\text{e}}}{d-1})\boldsymbol{\Xi}_\theta\big)\big\} \\
&= \langle\Phi^{\ddagger}(t_{\text{f}};t_0,\boldsymbol{P}_k,\boldsymbol{u}),\boldsymbol{\Upsilon}_\theta\rangle_{\text{HS}}
\end{aligned}
\tag{4.113}
$$

and $\langle\Phi^{\ddagger}(t_{\text{f}};t_0,\boldsymbol{S}_k,\boldsymbol{u}),\frac{\mathrm{d}}{\mathrm{d}\theta}\boldsymbol{\Xi}_\theta\rangle_{\text{HS}} = \langle\Phi^{\ddagger}(t_{\text{f}};t_0,\boldsymbol{S}_k,\boldsymbol{u}),\frac{\mathrm{d}}{\mathrm{d}\theta}\boldsymbol{\Upsilon}_\theta\rangle_{\text{HS}}$. Using expression (4.10) of the $(\boldsymbol{u},t_{\text{f}})$-CFI,

$$
J_{\psi_{\text{s}}(\Phi(t_{\text{f}};t_0,\boldsymbol{\Xi}_\theta,\boldsymbol{u}))}(\theta) = J_{\psi_{\text{p}}(\Phi(t_{\text{f}};t_0,\boldsymbol{\Upsilon}_\theta,\boldsymbol{u}))}(\theta) \tag{4.114}
$$

for all $(\boldsymbol{u},t_{\text{f}}) \in \Omega\times T$ and hence $J^{\text{c}}_{\boldsymbol{\Xi}_\theta}(\theta;\Sigma_{\text{s}}) = J^{\text{c}}_{\boldsymbol{\Upsilon}_\theta}(\theta;\Sigma_{\text{p}})$. Finally, given the universal converse $J^{\text{c}}_{\boldsymbol{\Upsilon}_\theta}(\theta;\Sigma_{\text{p}}) \leqslant J^{\text{q}}_{\boldsymbol{\Upsilon}_\theta}(\theta)$ from Lemma 4.2, it follows that

$$
J^{\text{c}}_{\boldsymbol{\Xi}_\theta}(\theta;\Sigma) \leqslant J^{\text{c}}_{\boldsymbol{\Xi}_\theta}(\theta;\Sigma_{\text{s}}) \leqslant J^{\text{q}}_{\boldsymbol{\Upsilon}_\theta}(\theta)\,. \tag{4.115}
$$

Invoking (4.6) completes the proof. ⊠

Magnetic field sensing with NV centers

NV centers enable sensing of exceptionally weak magnetic fields, such as those produced by individual nuclear spins. This case study will consider such scenarios. In particular, the goal is to estimate the strength of a weak ($\sim 250\text{nT}$) static magnetic field polarized along the NV center's axis.

The NV center can be modeled as a 3-level system. The free Hamiltonian of the system is [111]

$$
\boldsymbol{H}_0 = \hbar\Delta\boldsymbol{S}_z^2 + g\mu_{\text{B}}B_z\boldsymbol{S}_z \tag{4.116}
$$

where $\Delta = 2\pi\times 2.87\,\text{GHz}$ is the crystal field splitting between the $m_s = \pm 1$ Zeeman sublevels and the $m_s = 0$ sublevel, $g \approx 2$ is the Landé g-factor, $\mu_{\text{B}} = 9.274\times 10^{-24}\,\text{J/T}$ is the Bohr

magneton, B_z is the strength of the magnetic field to be estimated, and $\boldsymbol{S}_z$ is the z spin-1 matrix. The sensing protocol is comprised of three phases, which are summarized in Fig. 4.4 and detailed below.

1. *Probe state preparation:* The state of the NV center is reliably initialized in the $|m_s = 0\rangle$ state by the application of green laser light. The subsequent application of magnetic field control pulses enables arbitrary initialization of the state. The goal of this phase is to prepare the NV center in the pure state

$$|\psi\rangle \triangleq \frac{|m_s = +1\rangle + |m_s = -1\rangle}{\sqrt{2}} \,. \tag{4.117}$$

   For purposes of the case study, suppose that this state is prepared with "efficiency" $\eta \in (0, 1)$, say $\eta = 0.99$. The state of the system at the end of the preparation phase is taken to be the mixed state

$$\boldsymbol{A} \triangleq \eta |\psi\rangle\langle\psi| + (1 - \eta)|m_s = 0\rangle\langle m_s = 0| \,. \tag{4.118}$$

2. *Parameter encoding:* The magnetic field strength B_z is encoded into the probe state under the free evolution of the system over a long exposure interval of duration t_{exp}. Estimating B_z is tantamount to estimating the nondimensionalized parameter $\theta \triangleq g\mu_{\text{B}} B_z/\hbar$. The state after free evolution is

$$\boldsymbol{B}_\theta \triangleq e^{-\frac{\imath}{\hbar}\boldsymbol{H}_0 t_{\text{exp}}} \, \boldsymbol{A} \, e^{\frac{\imath}{\hbar}\boldsymbol{H}_0 t_{\text{exp}}} \,. \tag{4.119}$$

   Since the magnetic field strength is weak, the exposure interval is the longest portion of the sensing protocol. Accounting for relaxation during this exposure, the state of the system after parameter encoding is

$$\boldsymbol{\Xi}_\theta = e^{-\left(\frac{t_{\text{exp}}}{\tau_{\text{rel}}}\right)^3} \boldsymbol{B}_\theta + \left(1 - e^{-\left(\frac{t_{\text{exp}}}{\tau_{\text{rel}}}\right)^3}\right) \text{diag}\{\boldsymbol{B}_\theta\} \tag{4.120}$$

   where τ_{rel} is the relaxation time (commonly referred to as the "T_2 time"). For concreteness, the values $t_{\text{exp}} = 150\,\mu$s and $\tau_{\text{rel}} = 300\,\mu$s are considered.

3. *Measurement:* The state of the NV center is typically measured using optical detection with green laser excitation. Upon measurement, if the state collapses to the excited $m_s = \pm 1$ manifold, the center appears dark due to reduced fluorescence. If the state collapses to the ground level $m_s = 0$, the center exhibits stronger fluorescence. This enables a near-projective measurement to distinguish the excited manifold from the ground state—a binary POVM. Letting $\boldsymbol{P}_{\text{g}} \triangleq |m_s = 0\rangle\langle m_s = 0|$ be the projection onto the ground state, the binary POVM can be represented by $\mathcal{M} = \{\boldsymbol{M}, \boldsymbol{I} - \boldsymbol{M}\}$ where

$$\boldsymbol{M} = (1 - p_{\text{eg}})\boldsymbol{P}_{\text{g}} + \frac{p_{\text{eg}}}{2}(\boldsymbol{I} - \boldsymbol{P}_{\text{g}}) \tag{4.121}$$

   where $p_{\text{eg}} \in (0, 1)$ is the probability of inferring that the state is in the excited manifold given that it is in the ground state. Suppose that $p_{\text{eg}} = 0.04$. To realize

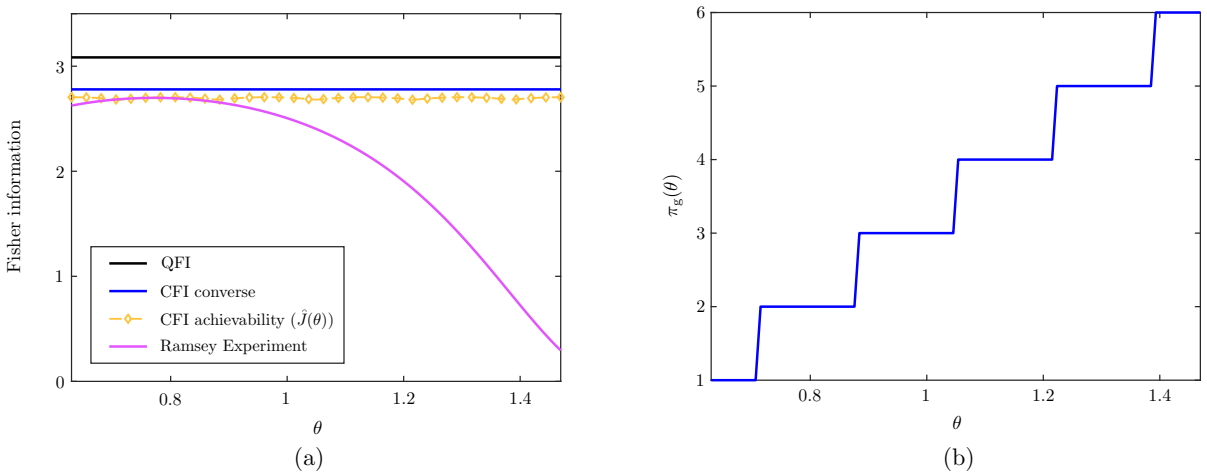


Figure 4.5: Magnetic field sensing case study. (a) Compares the Fisher information achieved by the set of representative controls to the QFI and to the Fisher information achieved by an ideal Ramsey experiment. (b) Depicts the greedy decision rule $\pi_g(\theta)$. By inspection, this decision rule is robust on Θ and, due to the finite number of discontinuities, is strongly robust at all but finitely many $\theta \in \Theta$.

measurements other than $\mathcal{M}$, control is applied prior to optical detection. The control Hamiltonians $\boldsymbol{H}_1 = \hbar \boldsymbol{S}_x$ and $\boldsymbol{H}_2 = \hbar \boldsymbol{S}_z$ are considered, corresponding to magnetic fields polarized along the x and z axes of the NV center, respectively. The control constraints $u_\ell^{\max}/2\pi = -u_\ell^{\min}/2\pi = 1\,\text{GHz}$ are considered, $\ell \in \{1, 2\}$. The effects of the magnetic field being sensed and of decoherence are negligible on the short timescale of the measurement phase. Hence, for the purposes of information control, the system is unitary and its dynamics are parameter-independent. In particular, $\boldsymbol{H}_0 = \hbar \Delta \boldsymbol{S}_z^2$ is used for control design.

A range of magnetic field strengths from 150 nT to 350 nT is considered. A grid of $N_r = 6$ equally spaced representative parameter values were placed over the parameter space and the gradient-based optimization scheme was leveraged to generate N_r representative control strategies. In this scheme, the measurement time is held constant at $t_f = 1\,\text{ns}$, motivated by Remark 4.1. Comparing the Fisher information achieved by the generated control strategies to the converse bound given by Theorem 4.6 certifies that the representative control strategies form an ϵ-covering, with $\epsilon = 0.0992$. This is visualized in Fig. 4.5a. Comparing the QFI to the CFI converse bound, the CFI is at least 10% tighter than the QFI. For context, Fig. 4.5a also plots the Fisher information achieved by a Ramsey experiment. A Ramsey experiment is non-adaptive, so the achievable estimation performance is not optimized over the full range of parameter values. On the other hand, adaptive information control with an ϵ-covering ensures that the estimation error scales asymptotically within ϵ of the fundamental limit $1/(nJ^{\text{c}}_{\boldsymbol{\Xi}_\theta}(\theta; \Sigma))$. We refer to the increase in Fisher information achieved by adaptive information control as the *adaptation gain*. Fig. 4.5b depicts the greedy decision rule π_g, which was defined in (4.27a). By inspection, this decision rule is robust on Θ and is strongly robust at all but finitely many $\theta \in \Theta$.

Fig. 4.6a demonstrates one-hundred Monte Carlo simulations of the adaptive information controller $(\mathcal{R}, \hat{\theta}_{\text{ml}}(\cdot), \pi_g(\cdot))$ in the case that the true parameter is $\theta_0 = 1.3$ or $\theta_0 = 0.8$. The

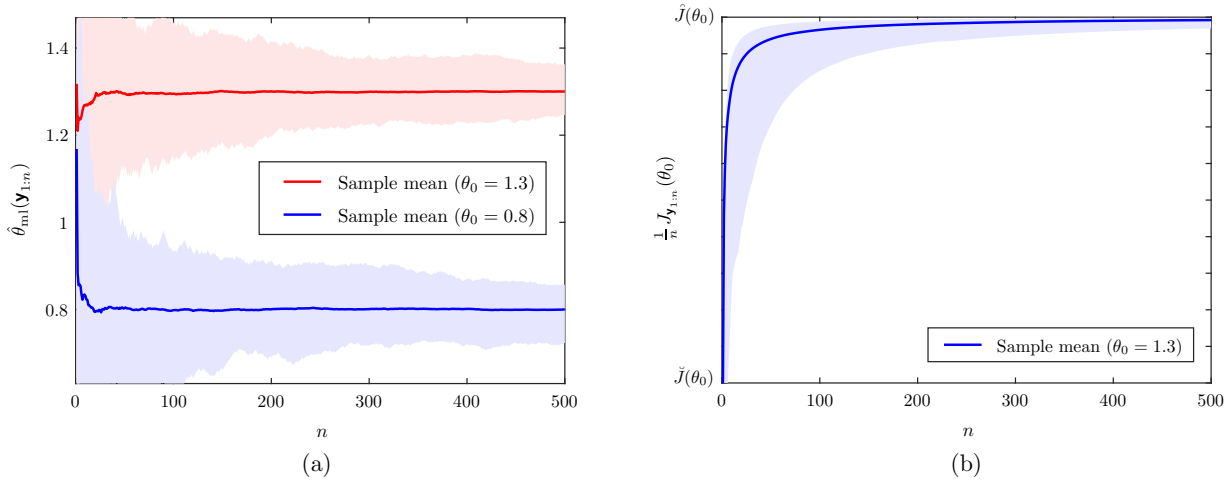


Figure 4.6: Magnetic field sensing case study. (a) Depicts Monte Carlo simulations of the composite estimator $\hat{\theta}_{\mathrm{ml}}(\mathbf{y}_{1:n})$ for two different values of the true parameter θ_0. The dark red (resp. blue) denotes the sample mean of the Monte Carlo simulation outcomes in the case that $\theta_0 = 1.3$ (resp. $\theta_0 = 0.8$). The light red and light blue shaded regions denote the range of simulation outcomes. (b) Depicts the average Fisher information extracted by the adaptive information controller $(\mathcal{R}, \hat{\theta}_{\mathrm{ml}}(\cdot), \pi(\cdot))$ across the simulations in the case $\theta_0 = 1.3$. The vertical axis of the plot scales linearly between the worst-case Fisher information 0.22 and the best-case Fisher information 2.71.

solid lines depict the average estimate whereas the shaded areas depict the range of parameter estimates across the simulations. As the number of experiments n increases, the uncertainty about the true parameters vanishes as expected. Fig. 4.6b demonstrates the average Fisher information extracted by the adaptive information controller $(\mathcal{R}, \hat{\theta}_{\mathrm{ml}}(\cdot), \pi_{\mathrm{g}}(\cdot))$—as a function of the number of experiments performed—in the case of the true parameter $\theta_0 = 1.3$. The average Fisher information lies between the lower bound $\check{J}(\theta) = 0.22$ and upper bound $\hat{J}(\theta) = 2.71$. The average Fisher information rapidly converges to $\hat{J}(\theta)$, which agrees with the non-asymptotic bound (4.36) proved in Theorem 4.2.

Chapter 5

Controlled f_γ-divergence for quantum state discrimination

5.1 Classical hypothesis testing and quantum state discrimination

Consider a classical binary hypothesis testing problem. A random variable $\mathsf{y} \in \mathcal{Y}$, whose distribution is either P_0 or P_1, is observed. The probability that $\mathsf{y} \sim P_0$ is denoted p_0; the probability that $\mathsf{y} \sim P_1$ is denoted p_1. The goal is to infer which of the two distributions generated y, i.e., to distinguish between the two hypotheses

$$\begin{aligned} H_0 &: \quad \mathsf{y} \sim P_0 \\ H_1 &: \quad \mathsf{y} \sim P_1 \,. \end{aligned} \tag{5.1}$$

The distinguishability of the two distributions P_0 and P_1 can be quantified by the f_γ-divergence,[1] an information measure [72, ch. 7] belonging to the class of f-divergences [114, 115].

Definition 5.1 (f_γ-divergence [112, p. 1333], [113, p. 2314])**.** Let $\gamma \geqslant 1$ and define the function $f_\gamma(y) \triangleq \max\{y - \gamma, 0\}$. Let P_0 and P_1 be probability distributions on $\mathcal{Y} \subseteq \mathbb{N}$. The *$f_\gamma$-divergence* between P_0 and P_1 is defined to be

$$D_\gamma(P_0 \| P_1) \triangleq \mathbb{E}_{\mathsf{y} \sim P_1} \left\{ f_\gamma\left(\frac{P_0(\mathsf{y})}{P_1(\mathsf{y})} \right) \right\}. \tag{5.2}$$

If $\mathsf{x} \sim P_0$ and $\mathsf{z} \sim P_1$, then $D_\gamma(\mathsf{x} \| \mathsf{z}) \triangleq D_\gamma(P_0 \| P_1)$. □

The following equivalent expressions can be deduced from the definition:

$$D_\gamma(P_0 \| P_1) = \sum_{y \in \mathcal{Y} \,:\, P_0(y) \geqslant \gamma P_1(y)} P_0(y) - \gamma P_1(y) \tag{5.3a}$$

$$= \frac{1-\gamma}{2} + \frac{1}{2} \sum_{y \in \mathcal{Y}} \left| P_0(y) - \gamma P_1(y) \right| . \tag{5.3b}$$

[1] This divergence has various names in the literature, including the "hockey stick divergence." Inspired by [112, 113], we opt to use the more formal name "f_γ-divergence."

Notice that the f_γ-divergence is exactly the total variation distance when $\gamma = 1$.

A binary hypothesis test is a rule $\pi : \mathcal{Y} \to \{0, 1\}$ that maps a measurement outcome y to an inferred hypothesis 0 or 1. If hypothesis i is inferred when the true hypothesis is H_j, where $i, j \in \{0, 1\}$, then a cost $c_{ij} \triangleq C(i, H_j) \geqslant 0$ is incurred. It will be assumed that $c_{00} < c_{10}$ and $c_{11} < c_{01}$, i.e., inferring the correct hypothesis costs less than inferring the wrong hypothesis. The *risk* of a test $\pi(\cdot)$ given hypothesis H_j is

$$R^\pi_{H_j} \triangleq \mathbb{E}_{\mathsf{y}\sim H_j}\{C(\pi(\mathsf{y}), H_j)\}\,. \tag{5.4}$$

The *Bayes risk* of a test $\pi(\cdot)$ is the expected risk considering the prior probabilities p_0 and p_1:

$$R^\pi \triangleq p_0\, R^\pi_{H_0} + p_1\, R^\pi_{H_1}\,. \tag{5.5}$$

Without loss of generality, suppose that $p_0(c_{10} - c_{00}) \leqslant p_1(c_{01} - c_{11})$.[2] It can be shown that the following test minimizes the Bayes risk:

$$\Lambda(y) \underset{H_0}{\overset{H_1}{\gtrless}} \frac{p_1\,(c_{01} - c_{11})}{p_0\,(c_{10} - c_{00})} \triangleq \gamma \tag{5.6}$$

where $\Lambda(y) \triangleq P_0(y)/P_1(y)$ is the *likelihood ratio* [94, pp. 374-379]. The risk of the optimal test (5.6) depends only on the f_γ-divergence between P_0 and P_1:

$$R^\star = (p_0 c_{10} + p_1 c_{11}) - p_0(c_{10} - c_{00})\, D_\gamma(P_0 \| P_1)\,. \tag{5.7}$$

For completeness, we briefly prove (i) that the likelihood ratio test (5.6) gives the optimal decision rule and (ii) that the optimal risk associated with this test is (5.7). Any decision rule is entirely determined by a decision region $\mathcal{E} \subseteq \mathbb{N}_1^{N_\mathrm{m}}$ such that

$$\pi(y) = \begin{cases} 1\,, & \text{if } y \in \mathcal{E} \\ 0\,, & \text{if } y \in \bar{\mathcal{E}} \end{cases}.$$

Given such region, one may deduce that

$$\begin{aligned} R^\pi_{H_0} &= c_{10} \sum_{y\in\mathcal{Y}} \pi(y) P_0(y) + c_{00} \sum_{y\in\mathcal{Y}} (1 - \pi(y)) P_0(y) \\ &= c_{10} P_0(\mathcal{E}) + c_{00} P_0(\bar{\mathcal{E}}) \\ &= c_{10} - (c_{10} - c_{00}) P_0(\bar{\mathcal{E}}) \end{aligned}$$

where $\bar{\mathcal{E}} \triangleq \mathcal{Y} \setminus \mathcal{E}$. Similarly, $R^\pi_{H_1} = c_{11} + (c_{01} - c_{11}) P_1(\bar{\mathcal{E}})$. This reveals that

$$R^\pi = (p_0 c_{10} + p_1 c_{11}) - p_0(c_{10} - c_{00}) \Big[P_0(\bar{\mathcal{E}}) - \frac{p_1(c_{01} - c_{11})}{p_0(c_{10} - c_{00})} P_1(\bar{\mathcal{E}}) \Big]\,. \tag{5.8}$$

This proves that an optimal decision region $\mathcal{E}^\star$ is such that $y \in \bar{\mathcal{E}}^\star$ if $P_0(y) \geqslant \gamma\, P_1(y)$. This is exactly the region specified by (5.6). Invoking the characterization (5.3a) of D_γ, the optimal risk is (5.7).

[2] If this is not the case, swap the order of the hypotheses.

In the quantum domain, hypothesis testing manifests in the problem of quantum state discrimination. Let $\mathcal{H} = \mathbb{C}^d$ with $d \geqslant 2$ as usual. Consider a random quantum state $\boldsymbol{\Xi}$ taking the value $\boldsymbol{\Xi}_0 \in \mathcal{D}(\mathcal{H})$ or $\boldsymbol{\Xi}_1 \in \mathcal{D}(\mathcal{H})$ with prior probabilities p_0 and p_1, respectively. Suppose that the random variable y is the outcome of measuring $\boldsymbol{\Xi}$ with a positive operator-valued measure (POVM) system $\mathcal{M} = \{\boldsymbol{M}_k : k \in \mathbb{N}_1^{N_\mathrm{m}}\}$. By Born's rule, $P_0(k) = \operatorname{tr}\{\boldsymbol{M}_k \boldsymbol{\Xi}_0\}$ and $P_1(k) = \operatorname{tr}\{\boldsymbol{M}_k \boldsymbol{\Xi}_1\}$ for each $k \in \mathbb{N}_1^{N_\mathrm{m}}$. To emphasize the dependence of P_0 (resp. P_1) on $\boldsymbol{\Xi}_0$ (resp. $\boldsymbol{\Xi}_1$) and $\mathcal{M}$, the abuse of notation

$$D_\gamma(\boldsymbol{\Xi}_0 \| \boldsymbol{\Xi}_1; \mathcal{M}) \triangleq D_\gamma(P_0 \| P_1) \tag{5.9}$$

will be used. The symbol $D_\gamma(\boldsymbol{\Xi}_0 \| \boldsymbol{\Xi}_1; \mathcal{M})$ should be read as "the divergence between $\boldsymbol{\Xi}_0$ and $\boldsymbol{\Xi}_1$ induced by the $\mathcal{M}$." If the POVM $\mathcal{M}$ is fixed, this problem is a classical hypothesis testing problem and $D_\gamma(\boldsymbol{\Xi}_0 \| \boldsymbol{\Xi}_1; \mathcal{M})$, through (5.7), provides the best attainable performance. However, if it is possible to choose the POVM, one would choose a POVM that maximizes the f_γ-divergence.

Definition 5.2 (Quantum f_γ-divergence, cf. [116])**.**[3] The *quantum f_γ-divergence* between $\boldsymbol{\Xi}_0$ and $\boldsymbol{\Xi}_1$ is

$$D_\gamma^{\mathrm{q}}(\boldsymbol{\Xi}_0 \| \boldsymbol{\Xi}_1) = \max_{\mathcal{M}} D_\gamma(\boldsymbol{\Xi}_0 \| \boldsymbol{\Xi}_1; \mathcal{M}) \tag{5.10}$$

where the maximization is performed over all possible POVM systems. □

Notice that the quantum f_γ-divergence is between two quantum states $\boldsymbol{\Xi}_0$ and $\boldsymbol{\Xi}_1$, whereas the classical f_γ-divergence is between two probability distributions. Definition 5.2 is equivalent to the usual definition [116] of quantum f_γ-divergence

$$D_\gamma^{\mathrm{q}}(\boldsymbol{\Xi}_0 \| \boldsymbol{\Xi}_1) = \frac{1-\gamma}{2} + \frac{1}{2} \| \boldsymbol{\Xi}_0 - \gamma \boldsymbol{\Xi}_1 \|_1 \tag{5.11}$$

as shown by the following lemma.

Lemma 5.1. Definitions (5.10) and (5.11) of the quantum f_γ-divergence are equivalent. Moreover, a particular measurement achieving the quantum f_γ-divergence is the binary POVM $\mathcal{M} = \{\boldsymbol{M}, \boldsymbol{I} - \boldsymbol{M}\}$, where $\boldsymbol{M}$ is the projector onto the range of $[\boldsymbol{\Xi}_0 - \gamma \boldsymbol{\Xi}_1]_+$. □

Proof. Let $\boldsymbol{Z} \triangleq \boldsymbol{\Xi}_0 - \gamma \boldsymbol{\Xi}_1$. Using Born's rule and (5.3b) it follows that, for any POVM $\{\boldsymbol{M}_k : k \in \mathbb{N}_1^{N_\mathrm{m}}\}$,

$$D_\gamma(\boldsymbol{\Xi}_0 \| \boldsymbol{\Xi}_1; \mathcal{M}) = \frac{1-\gamma}{2} + \frac{1}{2} \sum_{k=1}^{N_\mathrm{m}} |\operatorname{tr}\{\boldsymbol{M}_k \boldsymbol{Z}\}| \,. \tag{5.12}$$

For any matrix $\boldsymbol{A}$, the inequality $\sum_{k=1}^{N_\mathrm{m}} |\operatorname{tr}\{\boldsymbol{M}_k \boldsymbol{A}\}| \leqslant \|\boldsymbol{A}\|_1$ holds [81, p. 130]. Thus, $D_\gamma^{\mathrm{q}}(\boldsymbol{\Xi}_0 \| \boldsymbol{\Xi}_1) \leqslant \frac{1-\gamma}{2} + \frac{1}{2} \| \boldsymbol{\Xi}_0 - \gamma \boldsymbol{\Xi}_1 \|_1$.

[3] As with the classical case, the literature often refers to this divergence by the name of "quantum hockey stick divergence."

See the reference [117] for additional commentary on this divergence.

On the other hand, let $\mathcal{M}$ be the binary POVM $\{\boldsymbol{M}, \boldsymbol{I}-\boldsymbol{M}\}$ where $\boldsymbol{M}$ is a projector onto the range of $[\boldsymbol{Z}]_+$. If $\nu \in \mathbb{N}_1^d$ is the number of positive eigenvalues of $\boldsymbol{Z}$, then $|\operatorname{tr}\{\boldsymbol{M}\boldsymbol{A}\}| = \sum_{i=1}^{\nu} \lambda_i^{\downarrow}(\boldsymbol{Z})$ and $|\operatorname{tr}\{(\boldsymbol{I}-\boldsymbol{M})\boldsymbol{Z}\}| = |\sum_{i=\nu+1}^{d} \lambda_i^{\downarrow}(\boldsymbol{Z})|$. So,

$$|\operatorname{tr}\{\boldsymbol{M}\boldsymbol{Z}\}| + |\operatorname{tr}\{(\boldsymbol{I}-\boldsymbol{M})\boldsymbol{Z}\}| = \sum_{i=1}^{d} |\lambda_i^{\downarrow}(\boldsymbol{Z})| = \|\boldsymbol{Z}\|_1$$

and $D_\gamma^{\mathrm{q}}(\boldsymbol{\Xi}_0\|\boldsymbol{\Xi}_1) \geqslant D_\gamma(\boldsymbol{\Xi}_0\|\boldsymbol{\Xi}_1;\mathcal{M}) = \frac{1-\gamma}{2} + \frac{1}{2}\|\boldsymbol{\Xi}_0 - \gamma\boldsymbol{\Xi}_1\|_1$ by (5.12). ⊠

Remark 5.1. The variational characterization (5.10) of the quantum f_γ-divergence parallels the variational characterization of quantum Fisher information (QFI) (see Definition 4.2) in terms of maximizing the classical Fisher information over all possible POVMs. □

Remark 5.2. Combining (5.7) and (5.11) gives the same lower bound on the Bayes risk of discriminating two quantum states that was calculated by Helstrom in [12, pp. 106-108]. In the case that $c_{00} = c_{11} = 0$ and $c_{01} = c_{10} = 1$, the risk R^π is exactly the discrimination error probability, and the corresponding lower bound is what's widely-known as the "Helstrom bound" [59, 118, 119, 120]. □

While the quantum f_γ-divergence describes the *ultimate limit* on information maximization, it can be unachievable in quantum systems where only a fixed POVM is available and control is used to effectively realize different measurements. The goal of this thesis is to uncover the *fundamental limits* of information maximization in said quantum systems.

5.2 Controlled f_γ-divergence

The goal of this section is to define the controlled f_γ-divergence, which describes the fundamental limit on information extraction in systems equipped with a fixed measurement apparatus.

Consider a quantum control system Σ whose (fixed) measurement apparatus represented by a POVM $\mathcal{M}$. The goal is still to determine which value the random quantum state $\boldsymbol{\Xi}$ takes (i.e., the problem of discriminating between the quantum states $\boldsymbol{\Xi}_0$ and $\boldsymbol{\Xi}_1$). If the unknown state $\boldsymbol{\Xi}$ is measured by fixed apparatus at an initial time $t_0 \in T$, before evolving under the effect of control, the resulting information is simply $D_\gamma(\boldsymbol{\Xi}_0\|\boldsymbol{\Xi}_1;\mathcal{M})$ as before. However, to increase information extraction, we allow the unknown state to evolve under the influence of a carefully-designed control law prior to employing the fixed measurement apparatus. This crucial step encodes more information into the output of the measurement than would be achievable without control. Mathematically, if the unknown state first evolves under the effect of a control law $\boldsymbol{u} \in \Omega$ over a time interval $[t_0, t_{\mathrm{f}}]$ and is then measured at time $t_{\mathrm{f}} \in T$, the resulting information is $D_\gamma(\psi(\Phi(t_{\mathrm{f}};t_0,\boldsymbol{\Xi}_0,\boldsymbol{u}))\|\psi(\Phi(t_{\mathrm{f}};t_0,\boldsymbol{\Xi}_1,\boldsymbol{u}))$. This is visualized in Fig. 5.1. Optimizing over all possible $(\boldsymbol{u}, t_{\mathrm{f}}) \in \Omega \times T$ gives the fundamental limit on information extraction by a quantum control system Σ.

Definition 5.3 (Controlled f_γ-divergence)**.** A control law-measurement time pair $(\boldsymbol{u}, t_{\mathrm{f}}) \in \Omega \times T$ is said to be a *control strategy*. The $(\boldsymbol{u}, t_{\mathrm{f}})$-*controlled* f_γ-*divergence* is

$$D_\gamma^{\mathrm{c}}(\boldsymbol{\Xi}_0\|\boldsymbol{\Xi}_1;\Sigma,\boldsymbol{u},t_{\mathrm{f}}) = D_\gamma(\psi(\Phi(t_{\mathrm{f}};t_0,\boldsymbol{\Xi}_0,\boldsymbol{u}))\|\psi(\Phi(t_{\mathrm{f}};t_0,\boldsymbol{\Xi}_1,\boldsymbol{u}))\,. \tag{5.13}$$

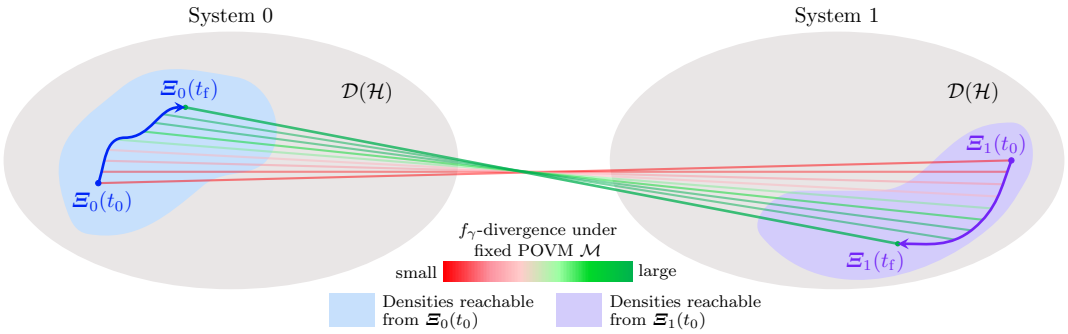


Figure 5.1: Visualization of the f_γ-divergence's evolution under control. There are two instantiations of the system, one initially in state $\boldsymbol{\Xi}_0(t_0)$ and one initially in state $\boldsymbol{\Xi}_1(t_0)$ at time t_0. These instantiations are controlled jointly using the same control law over the time interval $[t_0, t_f]$, resulting in tethered evolution as depicted by the lines connecting the trajectories of each system. The states are poorly distinguishable at the initial time t_0, but control makes them more easily distinguishable at a later time t_f.

The *controlled f_γ-divergence* attainable by the quantum control system Σ is

$$D_\gamma^{\mathrm{c}}(\boldsymbol{\Xi}_0 \| \boldsymbol{\Xi}_1; \Sigma) = \sup_{(\boldsymbol{u}, t_f) \in \Omega \times T} D_\gamma^{\mathrm{c}}(\boldsymbol{\Xi}_0 \| \boldsymbol{\Xi}_1; \Sigma, \boldsymbol{u}, t_f)\,. \tag{5.14}$$

□

The following lemma collects useful properties of the controlled f_γ-divergence (e.g., the data processing inequality (DPI)). These properties are notably similar to those of the controlled Fisher information (CFI) introduced in Chapter 4.

Lemma 5.2 (Properties of controlled f_γ-divergence)**.** Let Σ be a quantum control system equipped with a POVM $\mathcal{M}$. The following properties hold:

1. (Non-negativity) $0 \leqslant D_\gamma^{\mathrm{c}}(\boldsymbol{\Xi}_0 \| \boldsymbol{\Xi}_1; \Sigma)$;
2. (Universal converse) $D_\gamma^{\mathrm{c}}(\boldsymbol{\Xi}_0 \| \boldsymbol{\Xi}_1; \Sigma) \leqslant D_\gamma^{\mathrm{q}}(\boldsymbol{\Xi}_0 \| \boldsymbol{\Xi}_1)$; and
3. (Data-processing inequality) if Σ' is the same system as Σ equipped instead with a post-processing $\mathcal{M}'$ of $\mathcal{M}$, then

$$D_\gamma^{\mathrm{c}}(\boldsymbol{\Xi}_0 \| \boldsymbol{\Xi}_1; \Sigma') \leqslant D_\gamma^{\mathrm{c}}(\boldsymbol{\Xi}_0 \| \boldsymbol{\Xi}_1; \Sigma)\,. \tag{5.15}$$

□

Proof. Non-negativity follows immediately from the non-negativity of the classical divergence. Let $(\boldsymbol{u}, t_f) \in \Omega \times T$ be any control strategy and $\Phi^\ddagger(t_f; t_0, \cdot, \boldsymbol{u})$ be the adjoint of the state transition map with respect to the Hilbert-Schmidt inner product. Using Born's rule, linearity of the state transition map, and (5.3b), it follows that the $(\boldsymbol{u}, t_f)$-controlled f_γ-divergence is

$$\begin{aligned} D_\gamma^{\mathrm{c}}(\boldsymbol{\Xi}_0 \| \boldsymbol{\Xi}_1; \Sigma, \boldsymbol{u}, t_f) &= \frac{1-\gamma}{2} + \frac{1}{2} \sum_{k=1}^{N_{\mathrm{m}}} |\operatorname{tr}\{\boldsymbol{M}_k \Phi(t_f; t_0, \boldsymbol{\Xi}_0 - \gamma \boldsymbol{\Xi}_1, \boldsymbol{u})\}| \\ &= \frac{1-\gamma}{2} + \frac{1}{2} \sum_{k=1}^{N_{\mathrm{m}}} |\operatorname{tr}\{\Phi^\ddagger(t_f; t_0, \boldsymbol{M}_k, \boldsymbol{u})(\boldsymbol{\Xi}_0 - \gamma \boldsymbol{\Xi}_1)\}|\,. \end{aligned} \tag{5.16}$$

Since the adjoint state transition map is positive and unital, it follows that

$$\mathcal{M}' = \{\Phi^{\ddagger}(t_{\mathrm{f}}; t_0, \boldsymbol{M}_k, \boldsymbol{u}) \,:\, k \in \mathbb{N}_1^{N_{\mathrm{m}}}\} \tag{5.17}$$

is a POVM and therefore $D_\gamma^{\mathrm{c}}(\boldsymbol{\Xi}_0 \| \boldsymbol{\Xi}_1; \Sigma, \boldsymbol{u}, t_{\mathrm{f}}) \leqslant D_\gamma^{\mathrm{q}}(\boldsymbol{\Xi}_0 \| \boldsymbol{\Xi}_1)$. The universal converse follows by taking the supremum of the left hand side of this inequality over all possible control strategies.

Let $\mathcal{M}' = \{\boldsymbol{M}'_j \,:\, j \in \mathbb{N}_1^{N'_{\mathrm{m}}}\}$, $N'_{\mathrm{m}} \in \mathbb{N}$, be a post-processing of $\mathcal{M}$. Then, there exists a conditional probability distribution $p_{j|k}$ on $N'_{\mathrm{m}} \times N_{\mathrm{m}}$, $N'_{\mathrm{m}} = |\mathcal{M}'|$, such that

$$\boldsymbol{M}'_j = \sum_{k=1}^{N_{\mathrm{m}}} p_{j|k} \boldsymbol{M}_k \qquad \forall j \in \mathbb{N}_1^{N'_{\mathrm{m}}} . \tag{5.18}$$

For $i \in \{0, 1\}$ and a control strategy $(\boldsymbol{u}, t_{\mathrm{f}}) \in \Omega \times T$,

$$\begin{aligned} \mathbb{P}\{\psi'(\Phi(t_{\mathrm{f}}; t_0, \boldsymbol{\Xi}_i, \boldsymbol{u})) = j\} &= \langle \boldsymbol{M}'_j, \Phi(t_{\mathrm{f}}; t_0, \boldsymbol{\Xi}_i, \boldsymbol{u})\rangle_{\mathrm{HS}} \\ &= \sum_{k=1}^{N_{\mathrm{m}}} p_{j|k} \langle \boldsymbol{M}_k, \Phi(t_{\mathrm{f}}; t_0, \boldsymbol{\Xi}_i, \boldsymbol{u})\rangle_{\mathrm{HS}} \\ &= \sum_{k=1}^{N_{\mathrm{m}}} p_{j|k} \, \mathbb{P}\{\psi(\Phi(t_{\mathrm{f}}; t_0, \boldsymbol{\Xi}_i, \boldsymbol{u})) = k\} \end{aligned} \tag{5.19}$$

Thus, the distribution of the output $\psi'(\Phi(t_{\mathrm{f}}, t_0, \boldsymbol{\Xi}_i, \boldsymbol{u}))$ of system Σ' is equivalent to composing the conditional distribution $p_{j|k}$ with the distribution of the output $\psi(\Phi(t_{\mathrm{f}}; t_0, \boldsymbol{\Xi}_i, \boldsymbol{u}))$ of system Σ. By the DPI for f-divergences [72, pp. 119-120], it follows that

$$\begin{aligned} D_\gamma(\psi'(\Phi(t_{\mathrm{f}}; t_0, \boldsymbol{\Xi}_0, \boldsymbol{u}))\|\psi'(\Phi(t_{\mathrm{f}}; t_0, \boldsymbol{\Xi}_1, \boldsymbol{u})) &\leqslant D_\gamma(\psi(\Phi(t_{\mathrm{f}}; t_0, \boldsymbol{\Xi}_0, \boldsymbol{u}))\|\psi(\Phi(t_{\mathrm{f}}; t_0, \boldsymbol{\Xi}_1, \boldsymbol{u})) \\ &\leqslant D_\gamma^{\mathrm{c}}(\boldsymbol{\Xi}_0 \| \boldsymbol{\Xi}_1; \Sigma) . \end{aligned} \tag{5.20}$$

The DPI for the controlled f_γ-divergence follows by taking a supremum of the lefthand side of this inequality over all possible control strategies. ⊠

5.2.1 Converse bounds on controlled f_γ-divergence

This section will develop converse bounds on the controlled f_γ-divergence in the case of quantum systems that are equipped with a binary POVM. Recall from Chapter 2.2 that $\mathcal{C}(\mathcal{H})$ denotes the set of quantum channels on $\mathcal{H}$ and $\operatorname{krank}\{K(\cdot)\}$ denotes the Kraus rank of $K(\cdot) \in \mathcal{C}(\mathcal{H})$. The set of Kraus rank constrained channels on $\mathcal{H}$ will be denoted

$$\mathcal{C}_{N_{\mathrm{k}}}(\mathcal{H}) \triangleq \left\{K(\cdot) \in \mathcal{C}(\mathcal{H}) \,:\, \operatorname{krank}\{K(\cdot)\} \leqslant N_{\mathrm{k}}\right\} \tag{5.21}$$

for each $N_{\mathrm{k}} \in \mathbb{N}_1^{d^2}$. The following result—which will be used to establish converse bounds—concerns the maximization and minimization of the quantity $\operatorname{tr}\{\boldsymbol{F} K(\boldsymbol{G})\}$ over Kraus-rank constrained quantum channels $K(\cdot) \in \mathcal{C}_{N_{\mathrm{k}}}(\mathcal{H})$.

Theorem 5.1 (Optimizing linear functionals of Kraus rank-constrained quantum channels)**.** Let $\boldsymbol{F}, \boldsymbol{G} \in \mathbb{C}_{\mathrm{H}}^{d\times d}$ be Hermitian matrices, let $0 \leqslant \nu \leqslant d$ denote the number of positive eigenvalues of $\boldsymbol{G}$, and let $N_{\mathrm{k}} \in \mathbb{N}$. Then, the following hold:

$$\min_{K(\cdot)\in\mathcal{C}_{N_{\mathrm{k}}}(\mathbb{C}^d)} \operatorname{tr}\{\boldsymbol{F}K(\boldsymbol{G})\} = \sum_{j=1}^{\nu} \lambda_j^{\downarrow}(\boldsymbol{G})\lambda_j^{\uparrow}(\boldsymbol{I}_{N_{\mathrm{k}}}\otimes\boldsymbol{F}) + \sum_{j=\nu+(N_{\mathrm{k}}-1)d+1}^{N_{\mathrm{k}}d} \lambda_{j-(N_{\mathrm{k}}-1)d}^{\downarrow}(\boldsymbol{G})\,\lambda_j^{\uparrow}(\boldsymbol{I}_{N_{\mathrm{k}}}\otimes\boldsymbol{F}) \quad (5.22a)$$

$$\max_{K(\cdot)\in\mathcal{C}_{N_{\mathrm{k}}}(\mathbb{C}^d)} \operatorname{tr}\{\boldsymbol{F}K(\boldsymbol{G})\} = \sum_{j=1}^{\nu} \lambda_j^{\downarrow}(\boldsymbol{G})\lambda_j^{\downarrow}(\boldsymbol{I}_{N_{\mathrm{k}}}\otimes\boldsymbol{F}) + \sum_{j=\nu+(N_{\mathrm{k}}-1)d+1}^{N_{\mathrm{k}}d} \lambda_{j-(N_{\mathrm{k}}-1)d}^{\downarrow}(\boldsymbol{G})\,\lambda_j^{\downarrow}(\boldsymbol{I}_{N_{\mathrm{k}}}\otimes\boldsymbol{F})\,. \quad (5.22b)$$

If $N_{\mathrm{k}} \geqslant \max\{\nu, d-\nu\}$, then these expressions reduce to

$$\min_{\substack{K(\cdot)\in\mathcal{C}_{N_{\mathrm{k}}}(\mathcal{H})\\ N_{\mathrm{k}}\geqslant\max\{\nu,d-\nu\}}} \operatorname{tr}\{\boldsymbol{F}K(\boldsymbol{G})\} = \lambda_{\min}(\boldsymbol{F}) \sum_{j=1}^{\nu} \lambda_j^{\downarrow}(\boldsymbol{G}) + \lambda_{\max}(\boldsymbol{F}) \sum_{j=\nu+1}^{d} \lambda_j^{\downarrow}(\boldsymbol{G}) \quad (5.23a)$$

$$\max_{\substack{K(\cdot)\in\mathcal{C}_{N_{\mathrm{k}}}(\mathcal{H})\\ N_{\mathrm{k}}\geqslant\max\{\nu,d-\nu\}}} \operatorname{tr}\{\boldsymbol{F}K(\boldsymbol{G})\} = \lambda_{\max}(\boldsymbol{F}) \sum_{j=1}^{\nu} \lambda_j^{\downarrow}(\boldsymbol{G}) + \lambda_{\min}(\boldsymbol{F}) \sum_{j=\nu+1}^{d} \lambda_j^{\downarrow}(\boldsymbol{G})\,. \quad (5.23b)$$

□

Proof. First consider (5.22b) and (5.22a). A quantum channel $K(\cdot)$ of Kraus rank at most N_{k} may be written as $K(\boldsymbol{G}) = \sum_{i=1}^{N_{\mathrm{k}}} \boldsymbol{V}_i \boldsymbol{G} \boldsymbol{V}_i^{\dagger}$ where the $\boldsymbol{V}_i$'s satisfy the completeness relation $\sum_{i=1}^{N_{\mathrm{k}}} \boldsymbol{V}_i^{\dagger}\boldsymbol{V}_i = \boldsymbol{I}_d$. There exists a set of matrices $\boldsymbol{W}_i \in \mathbb{C}^{d\times(N_{\mathrm{k}}-1)d}$ such that the block matrix

$$\boldsymbol{X} = \left[\begin{array}{cccc} \boldsymbol{V}_1^{\dagger} & \boldsymbol{V}_2^{\dagger} & \cdots & \boldsymbol{V}_{N_{\mathrm{k}}}^{\dagger} \\ \hline \boldsymbol{W}_1^{\dagger} & \boldsymbol{W}_2^{\dagger} & \cdots & \boldsymbol{W}_{N_{\mathrm{k}}}^{\dagger} \end{array}\right] \in \mathbb{C}^{N_{\mathrm{k}}d\times N_{\mathrm{k}}d} \quad (5.24)$$

is unitary. This follows from the completeness relation, which ensures that the first d rows of $\boldsymbol{X}$ are orthonormal. The Gram-Schmidt procedure can be used to construct an orthonormal basis and fill in the remaining $(N_{\mathrm{k}}-1)d$ rows of $\boldsymbol{X}$, resulting in $\boldsymbol{X}^{\dagger}\boldsymbol{X} = \boldsymbol{I}_{N_{\mathrm{k}}d}$. Conversely, any unitary $\boldsymbol{X} \in \mathbb{C}^{N_{\mathrm{k}}d\times N_{\mathrm{k}}d}$ can be partitioned to obtain a Kraus representation of a quantum channel with rank at most N_{k}.

Define $\boldsymbol{A} \triangleq \boldsymbol{I}_{N_{\mathrm{k}}} \otimes \boldsymbol{F}$ and

$$\boldsymbol{B} \triangleq \left[\begin{array}{c|c} \boldsymbol{G} & \boldsymbol{0}_{d\times(N_{\mathrm{k}}-1)d} \\ \hline \boldsymbol{0}_{(N_{\mathrm{k}}-1)d\times d} & \boldsymbol{0}_{(N_{\mathrm{k}}-1)d\times(N_{\mathrm{k}}-1)d} \end{array}\right]. \quad (5.25)$$

Since $\boldsymbol{A}$ and $\boldsymbol{B}$ are Hermitian, there exist eigendecompositions $\boldsymbol{A} = \boldsymbol{Q}_{\boldsymbol{A}}\boldsymbol{\Lambda}_{\boldsymbol{A}}\boldsymbol{Q}_{\boldsymbol{A}}^{\dagger}$ and $\boldsymbol{B} = \boldsymbol{Q}_{\boldsymbol{B}}\boldsymbol{\Lambda}_{\boldsymbol{B}}\boldsymbol{Q}_{\boldsymbol{B}}^{\dagger}$ with $\boldsymbol{Q}_{\boldsymbol{A}}$ and $\boldsymbol{Q}_{\boldsymbol{B}}$ unitary [121, p. 246]. It is straightforward to show, regardless of the $\boldsymbol{W}_i$'s chosen to fill out the matrix (5.24), that

$$\operatorname{tr}\{\boldsymbol{F}K(\boldsymbol{G})\} = \operatorname{tr}\{\boldsymbol{X}\boldsymbol{A}\,\boldsymbol{X}^{\dagger}\boldsymbol{B}\} = \operatorname{tr}\{\boldsymbol{Y}\boldsymbol{\Lambda}_{\boldsymbol{A}}\boldsymbol{Y}^{\dagger}\boldsymbol{\Lambda}_{\boldsymbol{B}}\} \quad (5.26)$$

where $\boldsymbol{Y} \triangleq \boldsymbol{Q}_{\boldsymbol{B}}^{\dagger} \boldsymbol{X} \boldsymbol{Q}_{\boldsymbol{A}}$. Since $\boldsymbol{Y}$ is unitary, $\boldsymbol{Y} \boldsymbol{\Lambda}_{\boldsymbol{A}} \boldsymbol{Y}^{\dagger}$ has the same eigenvalues as $\boldsymbol{A}$. Invoking Lemma A.4 provides the bounds

$$\sum_{j=1}^{N_{\mathrm{k}} d} \lambda_j^{\uparrow}(\boldsymbol{A}) \lambda_j^{\downarrow}(\boldsymbol{B}) \leqslant \operatorname{tr}\{\boldsymbol{F} K(\boldsymbol{G})\} \leqslant \sum_{j=1}^{N_{\mathrm{k}} d} \lambda_j^{\downarrow}(\boldsymbol{A}) \lambda_j^{\downarrow}(\boldsymbol{B}) \tag{5.27}$$

for all Kraus maps of rank no greater than N_{k}. Choosing $\boldsymbol{Y}$ in (5.26) to be an appropriate permutation matrix (which is unitary) allows one to achieve either of the bounds in (5.27). Then, $\boldsymbol{X} = \boldsymbol{Q}_{\boldsymbol{B}} \boldsymbol{Y} \boldsymbol{Q}_{\boldsymbol{A}}^{\dagger}$ may be partitioned to obtain Kraus operators $\boldsymbol{V}_i$ defining a quantum channel $K(\cdot)$ which achieves the upper or lower bound. Noting that the spectrum of $\boldsymbol{B}$ is exactly that of $\boldsymbol{G}$ with the addition of $(N_{\mathrm{k}} - 1)d$ zeros, (5.22b) and (5.22a) are proved.

Equations (5.23b) and (5.23a) follow from the fact that the eigenvalues of $\boldsymbol{I}_{N_{\mathrm{k}}} \otimes \boldsymbol{F}$ are those of $\boldsymbol{F}$ repeated N_{k} times. If $N_{\mathrm{k}} \geqslant \max\{\nu, d - \nu\}$, the ν-largest eigenvalues of $\boldsymbol{I}_{N_{\mathrm{k}}} \otimes \boldsymbol{F}$ equal $\lambda_{\max}(\boldsymbol{F})$ and the $(d - \nu)$-smallest eigenvalues of $\boldsymbol{I}_{N_{\mathrm{k}}} \otimes \boldsymbol{F}$ equal $\lambda_{\min}(\boldsymbol{F})$. These facts complete the proof. ⊠

Now converse bounds are established for various classes of quantum systems. Binary POVMs of the form $\mathcal{M} = \{\boldsymbol{M}, \boldsymbol{I} - \boldsymbol{M}\}$ are of primary interest. Of course $\boldsymbol{0} \preccurlyeq \boldsymbol{M} \preccurlyeq \boldsymbol{I}$ and, to avoid trivialities, $\boldsymbol{M} \neq \boldsymbol{0}$ and $\boldsymbol{M} \neq \boldsymbol{I}$. Recall that:

1. the ultimate limit on binary quantum state discrimination, provided by the quantum f_γ-divergence, is achieved by a binary POVM; and
2. there are many quantum platforms where the fixed measurement apparatus is a binary POVM, e.g., multi-level trapped ion platforms [122].

Consider a quantum control system Σ equipped with the binary POVM $\mathcal{M}$. Define the convex function $g_\gamma(\cdot) : \mathbb{R} \to \mathbb{R}$ as

$$g_\gamma(z) \triangleq \frac{1}{2}\big[(1 - \gamma) + |z| + |(1 - \gamma) - z|\big] , \tag{5.28}$$

let $(\boldsymbol{u}, t_{\mathrm{f}}) \in \Omega \times T$, let $\boldsymbol{Z} \triangleq \boldsymbol{\Xi}_0 - \gamma \boldsymbol{\Xi}_1$, and let $\boldsymbol{Z}(t_{\mathrm{f}}) \triangleq \Phi(t_{\mathrm{f}}; t_0, \boldsymbol{Z}, \boldsymbol{u})$. Since the state transition map $\Phi(t_{\mathrm{f}}; t_0, \cdot, \boldsymbol{u})$ is linear,

$$\boldsymbol{Z}(t_{\mathrm{f}}) = \Phi(t_{\mathrm{f}}; t_0, \boldsymbol{\Xi}_0, \boldsymbol{u}) - \gamma \Phi(t_{\mathrm{f}}; t_0, \boldsymbol{\Xi}_1, \boldsymbol{u}) . \tag{5.29}$$

Using (5.3b), the $(\boldsymbol{u}, t_{\mathrm{f}})$-controlled f_γ-divergence can be expressed as

$$D_\gamma^{\mathrm{c}}(\boldsymbol{\Xi}_0 \| \boldsymbol{\Xi}_1; \Sigma, \boldsymbol{u}, t_{\mathrm{f}}) = g_\gamma\big(\langle \boldsymbol{M}, \boldsymbol{Z}(t_{\mathrm{f}}) \rangle_{\mathrm{HS}}\big) . \tag{5.30}$$

Since $g_\gamma(\cdot)$ is convex, (5.30) reveals that maximizing the f_γ-divergence is tantamount to extremizing $\langle \boldsymbol{M}, \boldsymbol{Z}(t_{\mathrm{f}}) \rangle_{\mathrm{HS}}$. Defining

$$\breve{z} \triangleq \inf_{(\boldsymbol{u}, t_{\mathrm{f}}) \in \Omega \times T} \langle \boldsymbol{M}, \Phi(t_{\mathrm{f}}; t_0, \boldsymbol{Z}, \boldsymbol{u}) \rangle_{\mathrm{HS}} \tag{5.31a}$$

$$\hat{z} \triangleq \sup_{(\boldsymbol{u}, t_{\mathrm{f}}) \in \Omega \times T} \langle \boldsymbol{M}, \Phi(t_{\mathrm{f}}; t_0, \boldsymbol{Z}, \boldsymbol{u}) \rangle_{\mathrm{HS}} \tag{5.31b}$$

the controlled f_γ-divergence is simply

$$D_\gamma^{\mathrm{c}}(\boldsymbol{\Xi}_0\|\boldsymbol{\Xi}_1;\Sigma) = \max\left\{g_\gamma(\breve{z}),\, g_\gamma(\hat{z})\right\}. \tag{5.32}$$

Equation (5.32) is an exact formula, but determining $\breve{z}$ and $\hat{z}$ would require solving the (generally infinite-dimensional and non-convex) optimization problems (5.31a) and (5.31b). However, this formula informs useful converse bounds. If it is possible to derive $\breve{\breve{z}}$ and $\hat{\hat{z}}$ such that $\breve{\breve{z}} \leqslant \breve{z}$ and $\hat{z} \leqslant \hat{\hat{z}}$, then by convexity of $g_\gamma(\cdot)$ it immediately follows that

$$D_\gamma^{\mathrm{c}}(\boldsymbol{\Xi}_0\|\boldsymbol{\Xi}_1;\Sigma) \leqslant \max\left\{g_\gamma(\breve{\breve{z}}),\, g_\gamma(\hat{\hat{z}})\right\}. \tag{5.33}$$

For the next theorem, define the constants

$$\breve{\breve{z}}_{\mathrm{u}} \triangleq \sum_{j=1}^{d} \lambda_j^{\downarrow}(\boldsymbol{M})\lambda_j^{\uparrow}(\boldsymbol{Z}) \tag{5.34a}$$

$$\hat{\hat{z}}_{\mathrm{u}} \triangleq \sum_{j=1}^{d} \lambda_j^{\downarrow}(\boldsymbol{M})\lambda_j^{\downarrow}(\boldsymbol{Z}). \tag{5.34b}$$

Letting $0 \leqslant \nu \leqslant d$ denote the number of positive eigenvalues of $\boldsymbol{Z}$, define the constants

$$\breve{\breve{z}} \triangleq \lambda_{\min}(\boldsymbol{M}) \sum_{j=1}^{\nu} \lambda_j^{\downarrow}(\boldsymbol{Z}) + \lambda_{\max}(\boldsymbol{M}) \sum_{j=\nu+1}^{d} \lambda_j^{\downarrow}(\boldsymbol{Z}) \tag{5.35a}$$

$$\hat{\hat{z}} \triangleq \lambda_{\max}(\boldsymbol{M}) \sum_{j=1}^{\nu} \lambda_j^{\downarrow}(\boldsymbol{Z}) + \lambda_{\min}(\boldsymbol{M}) \sum_{j=\nu+1}^{d} \lambda_j^{\downarrow}(\boldsymbol{Z}). \tag{5.35b}$$

Theorem 5.2 (CFI converse bounds)**.** Let Σ be a quantum control system equipped with a binary POVM $\mathcal{M} = \{\boldsymbol{M}, \boldsymbol{I} - \boldsymbol{M}\}$.

1. If Σ is unital or $d = 2$, the controlled f_γ-divergence is bounded as

$$D_\gamma^{\mathrm{c}}(\boldsymbol{\Xi}_0\|\boldsymbol{\Xi}_1;\Sigma) \leqslant \max\left\{g_\gamma(\breve{\breve{z}}_{\mathrm{u}}),\, g_\gamma(\hat{\hat{z}}_{\mathrm{u}})\right\}. \tag{5.36}$$

2. For any quantum system Σ, the controlled f_γ-divergence is bounded as

$$D_\gamma^{\mathrm{c}}(\boldsymbol{\Xi}_0\|\boldsymbol{\Xi}_1;\Sigma) \leqslant \max\left\{g_\gamma(\breve{\breve{z}}),\, g_\gamma(\hat{\hat{z}})\right\}. \tag{5.37}$$

□

Proof. 1) Begin by considering a *unitary* quantum system Σ. The evolution of the state of a unitary quantum system may be described by a rank-1 Kraus map so that $\boldsymbol{Z}(t_{\mathrm{f}}) = \boldsymbol{U}\boldsymbol{Z}\boldsymbol{U}^\dagger$ for some unitary $\boldsymbol{U} \in U(\mathcal{H})$. The bounds $\breve{\breve{z}}_{\mathrm{u}} \leqslant \breve{z}$ and $\hat{z} \leqslant \hat{\hat{z}}_{\mathrm{u}}$ follow from Theorem 5.1 with $N_{\mathrm{k}} = 1$, and (5.33) implies (5.36). Now, if Σ is unital, for any $\boldsymbol{u} \in \Omega$, Theorem 2.6 implies

that there exists a probability simplex $\{c_k\}$ and set of unitary matrices $\{\boldsymbol{U}_k\} \subset U(d)$ such that

$$\begin{aligned}\langle \boldsymbol{M}, \boldsymbol{Z}(t_{\mathrm{f}})\rangle_{\mathrm{HS}} &= \sum_k c_k \operatorname{tr}\{\, \boldsymbol{M}\boldsymbol{U}_k \boldsymbol{Z}\boldsymbol{U}_k^\dagger \,\} \\ &\leqslant \sum_k \sum_{j=1}^d c_k \max_{\boldsymbol{U}\in U(d)} \operatorname{tr}\{\, \boldsymbol{M}\boldsymbol{U}\boldsymbol{Z}\boldsymbol{U}^\dagger \,\} \\ &= \sum_k \sum_{j=1}^d c_k\, \lambda_j^{\downarrow}(\boldsymbol{M})\lambda_j^{\downarrow}(\boldsymbol{Z}) \\ &= \sum_{j=1}^d \lambda_j^{\downarrow}(\boldsymbol{M})\lambda_j^{\downarrow}(\boldsymbol{Z}) \end{aligned} \tag{5.38}$$

where the inequality is due to (5.22b) with $N_{\mathrm{k}} = 1$. Similarly, one can show $\langle \boldsymbol{M}, \boldsymbol{Z}(t_{\mathrm{f}})\rangle_{\mathrm{HS}} \geqslant \sum_{j=1}^d \lambda_j^{\downarrow}(\boldsymbol{M})\lambda_j^{\uparrow}(\boldsymbol{Z})$. This proves the bound (5.36) for unital quantum systems.

2) Let Σ be a general quantum system. The bounds $\breve{\breve{z}} \leqslant \breve{z}$ and $\hat{z} \leqslant \hat{\hat{z}}$ follow from Theorem 5.1 with $N_{\mathrm{k}} = d^2$, and (5.33) implies (5.37). ⊠

The *spread* of the matrix $\boldsymbol{M}$ is denoted $\operatorname{sp}(\boldsymbol{M}) \triangleq \lambda_{\max}(\boldsymbol{M}) - \lambda_{\min}(\boldsymbol{M}) \in (0, 1]$ [123]. To put the next result in context, a glance at (5.11) is helpful.

Corollary 5.1. Let Σ be a quantum system equipped with the POVM $\mathcal{M} = \{\boldsymbol{M}, \boldsymbol{I} - \boldsymbol{M}\}$. If $\gamma = 1$, the controlled f_γ-divergence is bounded as

$$D_\gamma^{\mathrm{c}}(\boldsymbol{\Xi}_0 \| \boldsymbol{\Xi}_1; \Sigma) \leqslant \frac{1}{2} \operatorname{sp}(\boldsymbol{M})\, \|\boldsymbol{\Xi}_0 - \boldsymbol{\Xi}_1\|\,. \tag{5.39}$$

Consequently, if $\operatorname{sp}(\boldsymbol{M}) < 1$, then the controlled f_γ-divergence is strictly less than the quantum f_γ-divergence. □

Proof. If $\gamma = 1$, the constants $\breve{\breve{z}}$ and $\hat{\hat{z}}$ from (5.35) can be written equivalently as

$$\hat{\hat{z}} = (\lambda_{\max}(\boldsymbol{M}) - \lambda_{\min}(\boldsymbol{M}))\, \frac{\|\boldsymbol{Z}\|_1}{2} \tag{5.40a}$$

$$\breve{\breve{z}} = (\lambda_{\min}(\boldsymbol{M}) - \lambda_{\max}(\boldsymbol{M}))\, \frac{\|\boldsymbol{Z}\|_1}{2}\,. \tag{5.40b}$$

To see this, note that $\boldsymbol{Z} = \boldsymbol{\Xi}_0 - \boldsymbol{\Xi}_1$ has zero trace. Therefore, the eigenvalues of $\boldsymbol{Z}$ sum to zero and $\sum_{j=1}^{\nu} \lambda_j^{\downarrow}(\boldsymbol{Z}) = -\sum_{j=\nu+1}^{d} \lambda_j^{\downarrow}(\boldsymbol{Z})$. The expressions in (5.35) reduce to

$$\hat{\hat{z}} = (\lambda_{\max}(\boldsymbol{M}) - \lambda_{\min}(\boldsymbol{M})) \sum_{j=1}^{\nu} \lambda_j^{\downarrow}(\boldsymbol{Z}) \tag{5.41a}$$

$$\breve{\breve{z}} = (\lambda_{\min}(\boldsymbol{M}) - \lambda_{\max}(\boldsymbol{M})) \sum_{j=1}^{\nu} \lambda_j^{\downarrow}(\boldsymbol{Z})\,. \tag{5.41b}$$

On the other hand, the trace norm of $\boldsymbol{Z}$ is exactly

$$\|\boldsymbol{Z}\|_1 = \sum_{j=1}^{d} |\lambda_j(\boldsymbol{Z})| = 2\sum_{j=1}^{\nu} \lambda_j^{\downarrow}(\boldsymbol{Z})\,. \tag{5.42}$$

When (5.40) holds, $g_\gamma(\hat{\hat{z}}) = g_\gamma(\breve{\breve{z}})$. Invoking Theorem 5.2 gives (5.39). ⊠

5.3 Universally optimal state discrimination via control

Control strategies can be adapted to varying circumstances, so there is the flexibility of choosing a different control strategy depending on which pair of quantum states one wishes to discriminate at any given point in time. In this adaptive scenario, it is natural to ask how well a quantum system can discriminate an arbitrary pair of quantum states. This leads to the following definition.

Definition 5.4 (Universally optimal for quantum state discrimination)**.** A quantum control system Σ on a Hilbert space $\mathcal{H}$ is said to be *universally optimal* for quantum state discrimination if, for all $\boldsymbol{\Xi}_0, \boldsymbol{\Xi}_1 \in \mathcal{D}(\mathcal{H})$ and $\gamma \geqslant 1$,

$$D_\gamma^{\mathrm{c}}(\boldsymbol{\Xi}_0 \| \boldsymbol{\Xi}_1; \Sigma) = D_\gamma^{\mathrm{q}}(\boldsymbol{\Xi}_0 \| \boldsymbol{\Xi}_1)\,. \tag{5.43}$$

□

The prior section derived converse bounds on the controlled f_γ-divergence for a quantum control system Σ equipped with a binary POVM $\mathcal{M} = \{\boldsymbol{M}, \boldsymbol{I} - \boldsymbol{M}\}$. One of the consequences of the bound (5.39) is that the POVM must be of a certain quality (i.e., $\mathrm{sp}(\boldsymbol{M}) = 1$) for the controlled f_γ-divergence to have the possibility of coinciding with the quantum f_γ-divergence when $\gamma = 1$. This necessary condition immediately generalizes to systems equipped with higher-cardinality POVMs.

Corollary 5.2 (Necessary condition for universal optimality)**.** Let Σ be a quantum control system equipped with a POVM $\mathcal{M} = \{\boldsymbol{M}_1, \boldsymbol{M}_2, \ldots, \boldsymbol{M}_{N_{\mathrm{m}}}\}$. If Σ is universally optimal for quantum state discrimination then the following equality holds:

$$\max_{\mathcal{S} \subseteq \mathbb{N}_1^{N_{\mathrm{m}}}} \mathrm{sp}\Big(\sum_{k \in \mathcal{S}} \boldsymbol{M}_k\Big) = 1\,. \tag{5.44}$$

□

We will now inquire about sufficient conditions for universal optimality. Below, two different differential quantum systems are shown to be universally optimal for quantum state discrimination.

d-level system with Hamiltonian and Lindblad controls

Consider a differential quantum system Σ on a finite d-dimensional space $\mathcal{H} = \mathbb{C}^d$. The control input $\boldsymbol{u}(t) = [\boldsymbol{v}(t)\, \boldsymbol{c}(t)]$ for this system is comprised of two pieces: $\boldsymbol{v}(t) \in \mathbb{R}^{N_{\mathrm{v}}}$ (where $N_{\mathrm{v}} \in \mathbb{N}$) represents control entering the system's dynamics through the Hamiltonian, and

$\boldsymbol{c}(t) \in \mathbb{R}^{d(d-1)/2}$ represents control entering the system's dynamics through the Lindblad operators. The system will have a time-invariant Hamiltonian $\boldsymbol{H}(\boldsymbol{v}(t))$ that is continuous in the control input $\boldsymbol{v}(t)$.[4] The *free Hamiltonian* is denoted $\boldsymbol{H}_0 \triangleq \boldsymbol{H}(\boldsymbol{0})$ and it has spectral decomposition $\boldsymbol{H}_0 = \sum_{i=1}^{d} \lambda_i^{\downarrow}(\boldsymbol{H}_0)|h_i\rangle\langle h_i|$. Here, $\lambda_i^{\downarrow}(\boldsymbol{H}_0) \in \mathbb{R}$ are *energy levels* (eigenvalues) and $|h_i\rangle$ are orthonormal eigenvectors of $\boldsymbol{H}_0$. The Lindblad control input $\boldsymbol{c}(t) = \text{vec}\{c_{ij}(t) \geqslant 0 : i, j \in \mathbb{N}_1^d,\, i < j\}$ is comprised of elements $c_{ij}(t)$ that influence the rate of transitions from the jth to ith energy level. The class of allowed controls Ω is taken to *contain* all piecewise constant functions $\boldsymbol{u}(t) = [\boldsymbol{v}(t)\, \boldsymbol{c}(t)]$ with the properties that $\boldsymbol{v}(t) = \boldsymbol{0}$ if $\boldsymbol{c}(t) \neq \boldsymbol{0}$ and vice versa, i.e., only Hamiltonian *or* Lindblad control need be turned on at any given time. The elements of Ω are measurable and essentially bounded. System dynamics of the form

$$\begin{aligned}\dot{\boldsymbol{X}}(t) = -\frac{\imath}{\hbar}[\![\boldsymbol{H}(\boldsymbol{v}(t)), \boldsymbol{X}(t)]\!]_- &+ \sum_{\substack{i,j=1 \\ i<j}}^{d} \gamma_{ij}^{\text{c}}(c_{ij}(t))\Big[2\boldsymbol{L}_{ij}^{\text{c}}\boldsymbol{X}(t)\boldsymbol{L}_{ij}^{\text{c}\dagger} - [\![\boldsymbol{L}_{ij}^{\text{c}\dagger}\boldsymbol{L}_{ij}^{\text{c}}, \boldsymbol{X}(t)]\!]_+\Big] \\ &+ \sum_{i,j=1}^{d} \gamma_{ij}^{\text{e}}(t)\Big[2\boldsymbol{L}_{ij}^{\text{e}}\boldsymbol{X}(t)\boldsymbol{L}_{ij}^{\text{e}\dagger} - [\![\boldsymbol{L}_{ij}^{\text{e}\dagger}\boldsymbol{L}_{ij}^{\text{e}}, \boldsymbol{X}(t)]\!]_+\Big] \qquad (5.45)\end{aligned}$$

are considered. Here, $\gamma_{ij}^{\text{c}}(\cdot)$ is a class $\mathcal{K}_\infty$ function;[5] $\boldsymbol{L}_{ij}^{\text{c}} \triangleq |h_i\rangle\langle h_j|$; $\gamma_{ij}^{\text{e}}(t) \geqslant 0$ is Lebesgue integrable; $\boldsymbol{L}_{ij}^{\text{e}} \in \mathbb{C}^{d\times d}$; and $\|\boldsymbol{L}_{ij}^{\text{e}}\|_1 = 1$. The first set of dissipative terms in (5.45) model control's effect on transitions from the jth to ith energy level; the second set of dissipative terms in (5.45) model uncontrollable environment-induced dissipation. The state transition map associated with the dynamics (5.45) is denoted Φ, as usual.

We will show that, under appropriate hypotheses, the system Σ with dynamics given by (5.45) is universally optimal for state discrimination. Before proceeding on that front, preliminary lemmata are established. The dynamics (5.45) allow for general environment-induced dissipation, which could obfuscate control law design. On the other hand, the idealized dynamics

$$\begin{aligned}\dot{\boldsymbol{X}}^{\text{i}}(t) &= -\frac{\imath}{\hbar}[\![\boldsymbol{H}(\boldsymbol{v}(t)), \boldsymbol{X}^{\text{i}}(t)]\!]_- + \sum_{\substack{i,j=1 \\ i<j}}^{d} \gamma_{ij}^{\text{c}}(c_{ij}(t))\Big[2\boldsymbol{L}_{ij}^{\text{c}}\boldsymbol{X}^{\text{i}}(t)\boldsymbol{L}_{ij}^{\text{c}\dagger} - [\![\boldsymbol{L}_{ij}^{\text{c}\dagger}\boldsymbol{L}_{ij}^{\text{c}}, \boldsymbol{X}^{\text{i}}(t)]\!]_+\Big] \\ &\triangleq F^{\text{i}}(\boldsymbol{X}^{\text{i}}(t), \boldsymbol{u}(t)) \qquad (5.46)\end{aligned}$$

obtained by temporarily disregarding the environment-induced dissipation are more amenable to analysis. Recalling the approximation result established in Chapter 2.4.4, it is possible to bound the error between solutions to the true and idealized dynamics. In particular, the following lemma shows that the error can be mitigated by designing control laws that occur on sufficiently small time intervals.

[4] Continuity ensures that $\boldsymbol{H}(\boldsymbol{v}(\cdot))$ is measurable if $\boldsymbol{v}(\cdot)$ is measurable. This is essential for existence and uniqueness of solutions.

[5] Recall that a function $g : [0, \infty) \to [0, \infty)$ is class $\mathcal{K}_\infty$ if g is continuous, g is strictly increasing, $g(0) = 0$, and $\lim_{x\to\infty} g(x) = \infty$ [124, p. 144].

Lemma 5.3. Let $(\boldsymbol{u}, t_\mathrm{f}) \in \Omega \times T$, let $\boldsymbol{X}(t)$ be the solution to (5.45) originating from $\boldsymbol{X}(t_0) = \boldsymbol{X}_0 \in \mathbb{C}^{d\times d}$, and let $\boldsymbol{X}^\mathrm{i}(t)$ be the solution to (5.46) originating from $\boldsymbol{X}^\mathrm{i}(t_0) = \boldsymbol{X}_0$. Then, the following bound holds:

$$\|\boldsymbol{X}(t_\mathrm{f}) - \boldsymbol{X}^\mathrm{i}(t_\mathrm{f})\|_1 \leqslant 4\|\boldsymbol{X}_0\|_1 \int_{t_0}^{t_\mathrm{f}} \sum_{i,j=1}^{d} \gamma_{ij}^\mathrm{e}(t)\,\mathrm{d}t\,. \tag{5.47}$$

□

Proof. This is a direct corollary of Theorem 2.7. ⊠

The following lemma concerns the eigendecomposition of the generator $F^\mathrm{i}(\cdot, \boldsymbol{c}(t))$ of the idealized dynamics under a time-constant Lindblad control input $\boldsymbol{c}(t)$.

Lemma 5.4. Let $n \in \mathbb{N}_2^{d-1}$, $m = n+1$, and $\zeta_{1i}, \zeta_{mj} > 0$ for $i \in \mathbb{N}_2^n$ and $j \in \mathbb{N}_{m+1}^d$.[6] Then, the linear operator $F_\zeta : \mathbb{C}^{d\times d} \to \mathbb{C}^{d\times d}$, defined by

$$\begin{aligned} F_\zeta(\boldsymbol{X}) \triangleq -\frac{\imath}{\hbar}[\![\boldsymbol{H}_0, \boldsymbol{X}]\!]_- &+ \sum_{i=2}^{n} \zeta_{1i}\left[2\boldsymbol{L}_{1i}^\mathrm{c}\boldsymbol{X}\boldsymbol{L}_{1i}^{\mathrm{c}\dagger} - [\![\boldsymbol{L}_{1i}^{\mathrm{c}\dagger}\boldsymbol{L}_{1i}^\mathrm{c}, \boldsymbol{X}]\!]_+\right] \\ &+ \sum_{j=m+1}^{d} \zeta_{mj}\left[2\boldsymbol{L}_{mj}^\mathrm{c}\boldsymbol{X}\boldsymbol{L}_{mj}^{\mathrm{c}\dagger} - [\![\boldsymbol{L}_{mj}^{\mathrm{c}\dagger}\boldsymbol{L}_{mj}^\mathrm{c}, \boldsymbol{X}]\!]_+\right] \end{aligned} \tag{5.48}$$

for $\boldsymbol{X} \in \mathbb{C}^{d\times d}$, has eigenpairs

1. $(0, |h_1\rangle\langle h_1|)$ and $(0, |h_m\rangle\langle h_m|)$;
2. $(-2\zeta_{1i}\,, |h_i\rangle\langle h_i| - |h_1\rangle\langle h_1|)$ for each $i \in \mathbb{N}_2^n$; and
3. $(-2\zeta_{mj}, |h_j\rangle\langle h_j| - |h_m\rangle\langle h_m|)$ for each $j \in \mathbb{N}_{m+1}^d$.

□

Proof. The proof is purely algebraic and follows from the definitions of the Hamiltonian $\boldsymbol{H}_0$ and the Lindblad operators $\boldsymbol{L}_{1i}^\mathrm{c}$, $\boldsymbol{L}_{mj}^\mathrm{c}$; the details are omitted. ⊠

If the Lindblad control $\boldsymbol{c}(t)$ is set to zero, the idealized dynamics (5.46) reduce to the unitary dynamics

$$\dot{\boldsymbol{X}}^\mathrm{i}(t) = -\frac{\imath}{\hbar}[\![\boldsymbol{H}(\boldsymbol{v}(t)), \boldsymbol{X}^\mathrm{i}(t)]\!]_-\,. \tag{5.49}$$

Controllability of the system (5.49) is closely related to Lie theory [26]. We will say that (5.49) is *fully-actuated* if $\{-\frac{\imath}{\hbar}\boldsymbol{H}(\boldsymbol{x}) : \boldsymbol{x} \in \mathbb{R}^{N_\mathrm{v}}\} \supseteq su(d)$ and that (5.49) is *unitarily controllable* if $\{-\frac{\imath}{\hbar}\boldsymbol{H}(\boldsymbol{x}) : \boldsymbol{x} \in \mathbb{R}^{N_\mathrm{v}}\}_\mathrm{LA} \supseteq su(d)$.[7] In the unitarily controllable case, there exists $t_\mathrm{c} > 0$ such that any $\boldsymbol{X}^\mathrm{i}(t_0) \in \mathbb{C}^{d\times d}$ can be driven any unitarily equivalent matrix at time $t_\mathrm{c} + t_0$ [26, p. 95].

The following result establishes circumstances under which the system Σ (whose dynamics are given by (5.45)) is universally optimal for state discrimination.

[6] If $n = 1$ (resp. $m = d$) then $\mathbb{N}_2^n$ (resp. $\mathbb{N}_{m+1}^d$) are empty, as defined in the notation section.

[7] Here, $\{-\frac{\imath}{\hbar}\boldsymbol{H}(\boldsymbol{x}) : \boldsymbol{x} \in \mathbb{R}^{N_\mathrm{v}}\}_\mathrm{LA}$ denotes the smallest Lie algebra containing $\{-\frac{\imath}{\hbar}\boldsymbol{H}(\boldsymbol{x}) : \boldsymbol{x} \in \mathbb{R}^{N_\mathrm{v}}\}$.

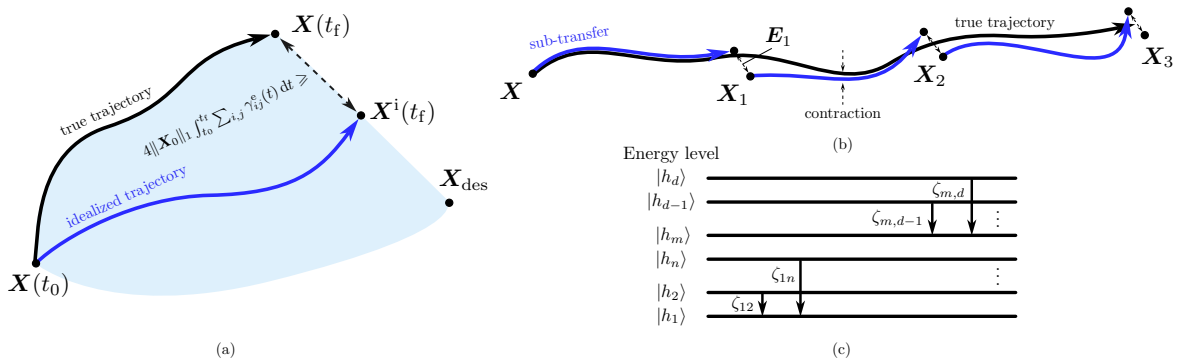


Figure 5.2: Visualization of the concepts involved in the proof of Theorem 5.3. (a) Visualizes the approximation concept where controls are designed for the idealized system and then applied to the true system. Lemma 5.3 shows that the error between the true and idealized trajectories is bounded by the integrated environment-induced dissipation rates. (b) Visualizes the decomposition of a global transfer into sub-transfers. The sublinear error propagation property of Theorem 2.2, which stems from the fact that the system trajectories are naturally contracting, implies that the true trajectory remains close to the sub-transfer trajectories. (c) Visualizes the population transfer between the system's energy levels while under idealized Lindblad control, as explained by Lemma 5.4.

Theorem 5.3 (Sufficient conditions for universal optimality)**.** Let Σ be a differential quantum system with the dynamics (5.45). Suppose the system Σ's measurement is described by a POVM $\mathcal{M} = \{\boldsymbol{M}_1, \boldsymbol{M}_2 \ldots, \boldsymbol{M}_{N_{\mathrm{m}}}\}$ satisfying the necessary condition (5.44). Then, it follows that

1. if the idealized system with dynamics (5.49) is unitarily controllable and the dissipation rates $\gamma_{ij}^{\mathrm{e}}(t)$ are zero for all $i, j \in \mathbb{N}_1^d$ and $t \geqslant t_0$, then Σ is universally optimal for quantum state discrimination; and

2. if the idealized system with dynamics (5.49) is fully-actuated, then Σ is universally optimal for quantum state discrimination. □

Remark 5.3. One can extend Theorem 5.3 to the case when the dissipation rates are non-zero and the idealized dynamics are unitarily controllable. This requires that one weaken the definition of universal optimality to ϵ-optimality. □

Theorem 5.3 will be proved shortly. The proof proceeds by constructing a control strategy $(\boldsymbol{u}, t_{\mathrm{f}})$ that transfers $\boldsymbol{Z} = \boldsymbol{\Xi}_0 - \gamma\boldsymbol{\Xi}_1$ to a desired matrix $\boldsymbol{Z}_{\mathrm{des}}$. Constructing the control strategy relies on the following concepts, which are visualized in Fig. 5.2.

a) Control strategies are designed for the idealized dynamics (5.46). Lemma 5.3 ensures that these controls behave as intended when applied to the true dynamics (5.45) over short time-intervals.

b) The transfer from $\boldsymbol{Z}$ to $\boldsymbol{Z}_{\mathrm{des}}$ is decomposed into sub-transfers, each of which can be accomplished by a relatively simple control strategy. The sublinear error propagation property of Theorem 2.2 is used to stitch together the strategies and obtain a global transfer.

c) An eigendecomposition of the idealized dynamics' generator F^{i} is used to design Lindblad control that is necessary for one of the sub-transfers. This eigendecomposition is the subject of Lemma 5.4.

We are now ready for the proof of Theorem 5.3.

Proof of 1) Let $\gamma \geqslant 1$ and $\boldsymbol{\Xi}_0, \boldsymbol{\Xi}_1 \in \mathcal{D}(\mathcal{H})$ be arbitrary. To prove that Σ is universally optimal for quantum state discrimination, it suffices to show that

$$D_\gamma^{\mathrm{c}}(\boldsymbol{\Xi}_0 \| \boldsymbol{\Xi}_1; \Sigma) \geqslant D_\gamma^{\mathrm{q}}(\boldsymbol{\Xi}_0 \| \boldsymbol{\Xi}_1) - \epsilon\,, \qquad \forall \epsilon > 0\,. \tag{5.50}$$

This follows from the universal converse of Lemma 5.2. If $\boldsymbol{Z} \triangleq \boldsymbol{\Xi}_0 - \gamma \boldsymbol{\Xi}_1$ is positive or negative semidefinite then this inequality is trivial, so consider the case where $\boldsymbol{Z}$ is indefinite.

Fix $\epsilon > 0$. By (5.44), there exists a decision region $\mathcal{S} \subseteq \mathbb{N}_1^{N_{\mathrm{m}}}$ such that $\boldsymbol{M} \triangleq \sum_{k \in \mathcal{S}} \boldsymbol{M}_k$ satisfies $\mathrm{sp}(\boldsymbol{M}) = 1$. Let $\boldsymbol{M} = \sum_{k=1}^{d} \lambda_k^{\downarrow}(\boldsymbol{M}) |e_k\rangle\langle e_k|$ be an eigendecomposition of $\boldsymbol{M}$ in terms of orthonormal eigenvectors $|e_k\rangle$. Because of the spread condition, it follows that $\lambda_1^{\downarrow}(\boldsymbol{M}) = 1$ and $\lambda_d^{\downarrow}(\boldsymbol{M}) = 0$. Let Σ' be the same system as Σ, but let the measurement of Σ' be described by the post-processing $\{\boldsymbol{M}, \boldsymbol{I} - \boldsymbol{M}\}$ of $\mathcal{M}$. By the DPI for controlled f_γ-divergence (Lemma 5.2) it follows that

$$\begin{aligned} D_\gamma^{\mathrm{c}}(\boldsymbol{\Xi}_0 \| \boldsymbol{\Xi}_1; \Sigma) &\geqslant D_\gamma^{\mathrm{c}}(\boldsymbol{\Xi}_0 \| \boldsymbol{\Xi}_1; \Sigma') \\ &\geqslant D_\gamma^{\mathrm{c}}(\boldsymbol{\Xi}_0 \| \boldsymbol{\Xi}_1; \Sigma'; \boldsymbol{u}, t_{\mathrm{f}}) \\ &= g_\gamma(\langle \boldsymbol{M}, \Phi(t_{\mathrm{f}}; t_0, \boldsymbol{Z}, \boldsymbol{u}) \rangle_{\mathrm{HS}})\,, \quad \forall (\boldsymbol{u}, t_{\mathrm{f}}) \in \Omega \times T\,. \end{aligned} \tag{5.51}$$

Let $\nu \in \mathbb{N}_2^{d-1}$ denote the number of positive eigenvalues of $\boldsymbol{Z}$, and let $m = \nu + 1$. Consider the matrix

$$\boldsymbol{Z}_{\mathrm{des}} \triangleq \sum_{k=1}^{\nu} \lambda_k^{\downarrow}(\boldsymbol{Z}) |e_1\rangle\langle e_1| + \sum_{k=\nu+1}^{d} \lambda_k^{\downarrow}(\boldsymbol{Z}) |e_d\rangle\langle e_d|\,. \tag{5.52}$$

Note that $\|\boldsymbol{Z}\|_1 = \|\boldsymbol{Z}_{\mathrm{des}}\|_1$ and that $\boldsymbol{M}$ projects $\boldsymbol{Z}_{\mathrm{des}}$ onto its positive part. It follows that

$$\begin{aligned} D_\gamma^{\mathrm{q}}(\boldsymbol{\Xi}_0 \| \boldsymbol{\Xi}_1) &= \frac{1-\gamma}{2} + \frac{1}{2} \|\boldsymbol{\Xi}_0 - \gamma \boldsymbol{\Xi}_1\|_1 \\ &= \frac{1-\gamma}{2} + \frac{1}{2} \|\boldsymbol{Z}_{\mathrm{des}}\|_1 \\ &= g_\gamma(\langle \boldsymbol{M}, \boldsymbol{Z}_{\mathrm{des}} \rangle_{\mathrm{HS}})\,. \end{aligned} \tag{5.53}$$

By the reverse triangle inequality, for any $a, b \in \mathbb{R}$,

$$\begin{aligned} g_\gamma(a) - g_\gamma(b) \leqslant |g_\gamma(a) - g_\gamma(b)| &\leqslant \tfrac{1}{2} \big| |a| - |b| \big| + \tfrac{1}{2} \big| |(1-\gamma) - a| - |(1-\gamma) - b| \big| \\ &\leqslant |a - b|\,. \end{aligned} \tag{5.54}$$

For any $(\boldsymbol{u}, t_{\mathrm{f}}) \in \Omega \times T$, it follows by the Cauchy-Schwarz inequality and the monotonicity of Schatten norms that

$$\begin{aligned} g_\gamma(\langle \boldsymbol{M}, \Phi(t_{\mathrm{f}}; t_0, \boldsymbol{Z}, \boldsymbol{u}) \rangle_{\mathrm{HS}}) &\geqslant g_\gamma(\langle \boldsymbol{M}, \boldsymbol{Z}_{\mathrm{des}} \rangle_{\mathrm{HS}}) - \big| \langle \boldsymbol{M}, \boldsymbol{Z}_{\mathrm{des}} - \Phi(t_{\mathrm{f}}; t_0, \boldsymbol{Z}, \boldsymbol{u}) \rangle_{\mathrm{HS}} \big| \\ &\geqslant g_\gamma(\langle \boldsymbol{M}, \boldsymbol{Z}_{\mathrm{des}} \rangle_{\mathrm{HS}}) - \|\boldsymbol{M}\|_{\mathrm{HS}} \, \|\boldsymbol{Z}_{\mathrm{des}} - \Phi(t_{\mathrm{f}}; t_0, \boldsymbol{Z}, \boldsymbol{u})\|_{\mathrm{HS}} \\ &\geqslant g_\gamma(\langle \boldsymbol{M}, \boldsymbol{Z}_{\mathrm{des}} \rangle_{\mathrm{HS}}) - \|\boldsymbol{M}\|_{\mathrm{HS}} \, \|\boldsymbol{Z}_{\mathrm{des}} - \Phi(t_{\mathrm{f}}; t_0, \boldsymbol{Z}, \boldsymbol{u})\|_1\,. \end{aligned} \tag{5.55}$$

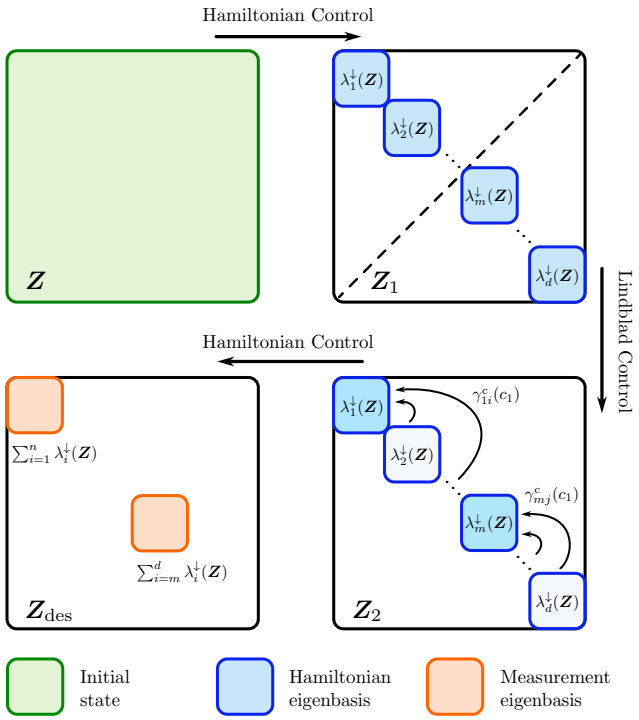


Figure 5.3: Visualization of the state transfers involved in the proof of Theorem 5.3. Visualization of the transformation from the initial matrix $\boldsymbol{Z}$ to the desired matrix $\boldsymbol{Z}_{\mathrm{des}}$. This is decomposed into three sub-transfers. Hamiltonian control transforms $\boldsymbol{Z}$ to $\boldsymbol{Z}_1$, which diagonalizes $\boldsymbol{Z}$ in the eigenbasis of the free Hamiltonian $\boldsymbol{H}_0$. The dashed line in the figure separates positive and negative eigenvalues of $\boldsymbol{Z}$. Next, Lindblad control maps the positive eigenvalues of $\boldsymbol{Z}$ to the span of $|h_1\rangle$ (the eigenvector corresponding to the largest eigenvalue of $\boldsymbol{H}_0$) and maps the negative eigenvalues of $\boldsymbol{Z}$ to the span of $|h_m\rangle$ (the eigenvector corresponding to the largest negative eigenvalue of $\boldsymbol{Z}$). Finally, Hamiltonian control transforms $\boldsymbol{Z}_2$ to $\boldsymbol{Z}_{\mathrm{des}}$, which diagonalizes $\boldsymbol{Z}_2$ in the eigenbasis of $\boldsymbol{M}$.

In light of (5.51), (5.53), and (5.55), we need only construct a control strategy $(\boldsymbol{u}, t_{\mathrm{f}}) \in \Omega \times T$ such that $\|\boldsymbol{Z}_{\mathrm{des}} - \Phi(t_{\mathrm{f}}; t_0, \boldsymbol{Z}, \boldsymbol{u})\|_1 \leqslant \epsilon / \|\boldsymbol{M}\|_{\mathrm{HS}}$ and we will have shown that

$$D_{\gamma}^{\mathrm{c}}(\boldsymbol{\Xi}_0 \| \boldsymbol{\Xi}_1; \Sigma) \geqslant D_{\gamma}^{\mathrm{q}}(\boldsymbol{\Xi}_0 \| \boldsymbol{\Xi}_1) - \epsilon \tag{5.56}$$

as desired.

The control strategy $(\boldsymbol{u}, t_{\mathrm{f}})$ will implement three sequential sub-transfers: one driving $\boldsymbol{Z}$ to $\boldsymbol{Z}_1$; one driving $\boldsymbol{Z}_1$ to within $\epsilon / \|\boldsymbol{M}\|_{\mathrm{HS}}$ of $\boldsymbol{Z}_2$; and one driving $\boldsymbol{Z}_2$ to $\boldsymbol{Z}_3 \triangleq \boldsymbol{Z}_{\mathrm{des}}$, where

$$\boldsymbol{Z}_1 \triangleq \sum_{k=1}^{d} \lambda_k^{\downarrow}(\boldsymbol{Z}) |h_k\rangle\langle h_k| \tag{5.57a}$$

$$\boldsymbol{Z}_2 \triangleq \sum_{k=1}^{\nu} \lambda_k^{\downarrow}(\boldsymbol{Z}) |h_1\rangle\langle h_1| + \sum_{k=n+1}^{d} \lambda_k^{\downarrow}(\boldsymbol{Z}) |h_d\rangle\langle h_d| \,. \tag{5.57b}$$

By the sublinear error propagation property (Theorem 2.2), it will follow that the trace distance between $\boldsymbol{Z}_{\mathrm{des}}$ and $\Phi(t_{\mathrm{f}}; t_0, \boldsymbol{Z}, \boldsymbol{u})$ is no greater than $\epsilon / \|\boldsymbol{M}\|_{\mathrm{HS}}$.

Since $\boldsymbol{Z}$ and $\boldsymbol{Z}_1$ are unitarily equivalent, existence of a Hamiltonian control strategy $\boldsymbol{v}(\cdot)$ transferring $\boldsymbol{Z}$ to $\boldsymbol{Z}_1$ over the time interval $[t_0, t_1] \triangleq [t_0, t_0 + t_c]$ is guaranteed by unitary controllability. The same goes for transferring $\boldsymbol{Z}_2$ to $\boldsymbol{Z}_3$.

Now the transfer of $\boldsymbol{Z}_1$ to $\boldsymbol{Z}_2$ is designed. For all $t \in [t_1, t_2]$, set $\boldsymbol{v}(t) = 0$, $c_{1i}(t) = c_1 \; \forall i \in \mathbb{N}_2^{\nu}$, and $c_{mj}(t) = c_1 \; \forall j \in \mathbb{N}_{m+1}^{d}$; where c_1 is a positive constant,

$$t_2 \triangleq t_1 + \max\left\{0, -\frac{1}{2\breve{\gamma}^{\mathrm{c}}(c_1)} \ln\left[\frac{1}{2\|\boldsymbol{Z}\|_1} \frac{\epsilon}{\|\boldsymbol{M}\|_{\mathrm{HS}}}\right]\right\} \tag{5.58}$$

and $\breve{\gamma}^{\mathrm{c}} : \mathbb{R}_{\geqslant 0} \to \mathbb{R}_{\geqslant 0}$ is the function

$$\breve{\gamma}^{\mathrm{c}}(x) \triangleq \Big(\bigwedge_{i=2}^{\nu} \gamma_{1i}^{\mathrm{c}}(x)\Big) \wedge \Big(\bigwedge_{j=m+1}^{d} \gamma_{mj}^{\mathrm{c}}(x)\Big), \quad \forall x \in \mathbb{R}_{\geqslant 0}\,. \tag{5.59}$$

By the eigendecomposition provided in Lemma 5.4,

$$\begin{aligned}\Phi(t_2; t_1, \boldsymbol{Z}_1, \boldsymbol{u}) = \boldsymbol{Z}_2 &+ \sum_{i=2}^{\nu} e^{-2\gamma_{1i}^{\mathrm{c}}(c_1)(t_2-t_1)} \lambda_i^{\downarrow}(\boldsymbol{Z})(|h_i\rangle\langle h_i| - |h_1\rangle\langle h_1|) \\ &+ \sum_{j=m+1}^{d} e^{-2\gamma_{mj}^{\mathrm{c}}(c_1)(t_2-t_1)} \lambda_j^{\downarrow}(\boldsymbol{Z})(|h_j\rangle\langle h_j| - |h_m\rangle\langle h_m|)\end{aligned} \tag{5.60}$$

and, by the orthonormality of the $|h_i\rangle$'s,

$$\begin{aligned}\|\Phi(t_2; t_1, \boldsymbol{Z}_1, \boldsymbol{u}) - \boldsymbol{Z}_2\|_1 &= 2\sum_{i=2}^{\nu} e^{-2\gamma_{1i}^{\mathrm{c}}(c_1)(t_2-t_1)} \, |\lambda_i^{\downarrow}(\boldsymbol{Z})| + 2\sum_{j=m+1}^{d} e^{-2\gamma_{mj}^{\mathrm{c}}(c_1)(t_2-t_1)} |\lambda_j^{\downarrow}(\boldsymbol{Z})| \\ &\leqslant 2e^{-2\breve{\gamma}^{\mathrm{c}}(c_1)(t_2-t_1)} \sum_{\substack{k=2 \\ k\neq m}}^{d} |\lambda_k^{\downarrow}(\boldsymbol{Z})|\,.\end{aligned} \tag{5.61}$$

Given (5.58) and the fact $\|\boldsymbol{Z}\|_1 = \sum_{k=1}^{d} |\lambda_k^{\downarrow}(\boldsymbol{Z})|$, it follows that

$$\|\Phi(t_2; t_1, \boldsymbol{Z}_1, \boldsymbol{u}) - \boldsymbol{Z}_2\|_1 \leqslant 2e^{-2\breve{\gamma}^{\mathrm{c}}(c_1)(t_2-t_1)} \|\boldsymbol{Z}\|_1 \leqslant \frac{\epsilon}{\|\boldsymbol{M}\|_{\mathrm{HS}}} \tag{5.62}$$

and the proof of part 1 is complete. ⊠

Proof of 2) Let ϵ, $\boldsymbol{Z}$, and $\boldsymbol{Z}_{\mathrm{des}}$ be as in the proof of part 1. The proof strategy is the same as before: construct a three-stage control strategy $(\boldsymbol{u}, t_{\mathrm{f}})$ driving $\boldsymbol{Z}$ to $\boldsymbol{Z}_3 = \boldsymbol{Z}_{\mathrm{des}}$. In this case, the three stages must occur on sufficiently small time intervals to overcome the uncontrolled environment-induced dissipation ($\gamma_{ij}^{\mathrm{e}}(t) \neq 0$).

For the first phase of the control, select $t_1 \geqslant t_0$ such that

$$\sigma(t_1; t_0) \triangleq 8 \int_{t_0}^{t_1} \sum_{i,j=1}^{d} \gamma_{ij}^{\mathrm{e}}(t)\, \mathrm{d}t \leqslant \frac{\epsilon}{6\|\boldsymbol{M}\|_{\mathrm{HS}}}\,. \tag{5.63}$$

Continuity of the integral ensures that this can be done. Matrices $\boldsymbol{Z}$ and $\boldsymbol{Z}_1$ are unitarily equivalent, so there exists $\boldsymbol{V} \in SU(d)$ such that $\boldsymbol{Z}_1 = \boldsymbol{V}\boldsymbol{Z}\boldsymbol{V}^{\dagger}$. Let $\imath\boldsymbol{A} \in su(d)$ be such that

$\boldsymbol{V} = e^{\imath \boldsymbol{A}(t_1 - t_0)}$. For $t \in [t_0, t_1]$, set $\boldsymbol{v}(t)$ such that $\boldsymbol{H}(\boldsymbol{v}(t)) = \hbar \boldsymbol{A}$ (this can be done because the system is fully actuated) and set $\boldsymbol{c}(t) = \boldsymbol{0}$. Thus, for the idealized dynamics (5.49), $\boldsymbol{Z}$ is transferred to $\boldsymbol{Z}_1$ over the time interval $[t_0, t_1]$. By Lemma 5.3, the actual trajectory satisfies

$$\|\Phi(t_1; t_0, \boldsymbol{Z}, \boldsymbol{u}) - \boldsymbol{Z}_1\| \leqslant \frac{\epsilon}{3\|\boldsymbol{M}\|_{\mathrm{HS}}}. \tag{5.64}$$

For the second phase of the control, choose $t_2 \geqslant t_1$ such that $\sigma(t_2; t_1) \leqslant \epsilon/(6\|\boldsymbol{M}\|_{\mathrm{HS}})$; choose $c_2 \in \mathbb{R}$ with

$$\breve{\gamma}^{\mathrm{c}}(c_2) \geqslant \max\left\{0, \frac{-1}{2(t_2 - t_1)} \ln\left[\frac{1}{2\|\boldsymbol{Z}\|_1} \frac{\epsilon}{6\|\boldsymbol{M}\|_{\mathrm{HS}}}\right]\right\}. \tag{5.65}$$

such that $2e^{-2\breve{\gamma}^{\mathrm{c}}(c_2)(t_2 - t_1)}\|\boldsymbol{Z}\|_1 \leqslant \epsilon/(6\|\boldsymbol{M}\|_{\mathrm{HS}})$. This can be done because the $\gamma^{\mathrm{c}}_{ij}(\cdot)$'s are class $\mathcal{K}_\infty$ functions. Set $\boldsymbol{v}(t) = \boldsymbol{0}$ and $c_{1i}(t) = c_{mj}(t) = c_2$ for all $t \in [t_1, t_2]$, $i \in \mathbb{N}_2^{\nu}$, and $j \in \mathbb{N}_{m+1}^{d}$. Combining the argument of part 1 and Lemma 5.3,

$$\|\Phi(t_2; t_1, \boldsymbol{Z}_1, \boldsymbol{u}) - \boldsymbol{Z}_2\| \leqslant \frac{\epsilon}{3\|\boldsymbol{M}\|_{\mathrm{HS}}}. \tag{5.66}$$

For the third phase of the control, $t_3 \geqslant t_2$ and $\boldsymbol{u}(t)$ for $t \in [t_2, t_3]$ can be designed by the same procedure as the first phase of the control, resulting in $\|\Phi(t_3; t_2, \boldsymbol{Z}_2) - \boldsymbol{Z}_3\|_1 \leqslant \epsilon/(3\|\boldsymbol{M}\|_{\mathrm{HS}})$. The proof is complete by invoking the sublinear error propagation property, as in the proof of part 1. ⊠

Hamiltonian controlled 2-level systems

In the case of 2-level systems, universally optimal state discrimination can be achieved by Hamiltonian control alone. Let

$$\dot{\boldsymbol{X}}(t) = -\frac{\imath}{\hbar}[\![\boldsymbol{H}(\boldsymbol{v}(t)), \boldsymbol{X}(t)]\!]_- + \sum_{i,j=1}^{2} \gamma^{\mathrm{e}}_{ij}(t)\left[2\boldsymbol{L}^{\mathrm{e}}_{ij}\boldsymbol{X}(t)\boldsymbol{L}^{\mathrm{e}\dagger}_{ij} - [\![\boldsymbol{L}^{\mathrm{e}\dagger}_{ij}\boldsymbol{L}^{\mathrm{e}}_{ij}, \boldsymbol{X}(t)]\!]_+\right] \tag{5.67}$$

denote the dynamics of a 2-level quantum system without Lindblad control.

Theorem 5.4 (Sufficient conditions for universal optimality)**.** Let Σ be a differential quantum system with the dynamics (5.67). Suppose the system Σ's measurement is described by a POVM $\mathcal{M} = \{\boldsymbol{M}_1, \boldsymbol{M}_2 \ldots, \boldsymbol{M}_{N_{\mathrm{m}}}\}$ satisfying the necessary condition (5.44). Then, statements 1 and 2 of Theorem 5.3 hold. □

Proof. Let $\boldsymbol{Z}$ and $\boldsymbol{M}$ be as in the proof of Theorem 5.3. Since $d = 2$, the matrix $\boldsymbol{M}$ is a rank-1 projector $\boldsymbol{M} = |e_1\rangle\langle e_1|$. The goal is then to drive $\boldsymbol{Z}$ to an ϵ neighborhood of

$$\boldsymbol{Z}_{\mathrm{des}} = \lambda_1^{\downarrow}(\boldsymbol{Z})\,|e_1\rangle\langle e_1| + \lambda_2^{\downarrow}(\boldsymbol{Z})\,(\boldsymbol{I} - |e_1\rangle\langle e_1|). \tag{5.68}$$

If condition 1 or 2 hold, a control law $\boldsymbol{u}$ and a $t_{\mathrm{f}} \geqslant t_0$ can be constructed with the property that $\|\Phi(t_{\mathrm{f}}; t_0, \boldsymbol{Z}, \boldsymbol{v}) - \boldsymbol{Z}_{\mathrm{des}}\|_1 \leqslant \epsilon$. The proof technique is the same as in Theorem 5.3, requiring only a single-stage Hamiltonian control strategy to implement an approximate unitary transformation mapping $\boldsymbol{Z}$ to $\boldsymbol{Z}_{\mathrm{des}}$. ⊠

Chapter 6

Conclusion

This thesis established new connections between quantum control and quantum inference, bridging a gap between two traditionally disparate fields of research. This study was motivated, in part, by the recurring practical scenario of a physical quantum system equipped with a fixed measurement apparatus. In such systems, ultimate limits of statistical inference tasks predicted by the quantum Fisher information (QFI) and quantum f_γ-divergence are not always achievable. This lack of achievability is because the QFI and quantum f_γ-divergence are constructed by optimizing the measurement apparatus. In contrast, the vision inspiring this thesis was leveraging quantum control to increase the information extracted by a *fixed* measurement apparatus. Controlled information measures were introduced, quantifying the maximal amount of statistical information that can be extracted via control from a system equipped with a single fixed measurement apparatus. These measures represent achievable fundamental limits for statistical inference tasks and were shown to exhibit key properties, e.g., data-processing inequalities. Methods for deriving achievability and converse bounds on controlled information measures were established, as were adaptive techniques for experimentally realizing information maximizing control strategies.

An axiomatization of quantum control theory was proposed, which provided the mathematical foundation needed to establish controlled information measures. This axiomatization led to certain immediate consequences for all quantum control systems, such as the universal contractivity property and the related sublinear error propagation property. Differential quantum control systems were studied, and approximation techniques were established leveraging the universal contractivity property. An expedited and self-contained treatment of optimal quantum control was presented, exploiting the unique structure and properties of differential quantum systems. This work took a first-principles approach rather than merely importing known results from classical control theory. The sublinear error propagation property, contraction-based approximation techniques, and self-contained treatment of optimal quantum control are fruits of this first-principles approach.

There are several natural ways in which this work could be extended. For instance, this thesis focused on scalar parameter estimation, and multi-parameter estimation is a logical next step. However, the extension of controlled Fisher information (CFI) to the multi-parameter setting is not a direct one. In the multi-parameter setting, the Fisher information is represented by a matrix, and to extend the optimal control techniques developed in this thesis to this setting one must first scalarize [125, pp. 386-287] the Fisher information matrix. There is no

canonical way to do this. Another direction for future research is to extend our axiomatization of quantum control systems to address composite or ensemble systems [126, 127]. These, of course, play an important role in quantum information science. For instance, unique control methodologies arise from allowing a quantum system to interact with an environment (the system and environment form a composite system) [73]. Lastly, while we have focused on achieving the best possible estimation and discrimination performance given a family of quantum states, there is the important related problem of optimal state design [4, 128, 129]. Combining state design with information control is a fertile area for future research.

Appendix A

Facts from measure theory and analysis

This appendix collects a few results from measure theory and mathematical analysis that are used in the main text.

Theorem A.1 (Bounded convergence theorem [91, p. 322])**.** Let $(\Omega, \mathcal{F}, \mu)$ be a *finite* measure space and $\{f_n\}_{n\in\mathbb{N}}$ a sequence of measurable functions on this space. If there exists $M > 0$ such that $|f_n(x)| \leqslant M$ for all $(n, x) \in \mathbb{N} \times \Omega$ and if $f_n \to f$ pointwise for some measurable function f, then

$$\int_\Omega f \,\mathrm{d}\mu = \lim_{n\to\infty} \int_\Omega f_n \,\mathrm{d}\mu\,. \tag{A.1}$$

□

Theorem A.2 (Vitali's convergence theorem [92, p. 93])**.** Fix $p \in [1, \infty)$. Let $(\Omega, \mathcal{F}, \mu)$ be a *finite* measure space, let $f : \Omega \to \mathbb{R}$ be measureable, and let $\{f_n \in L^p(\Omega) : n \in \mathbb{N}\}$ be a sequence of integrable functions. The following are equivalent:

1. $f \in L^p(\Omega)$ and $f_n \to f$ in $L^p(\Omega)$ as $n \to \infty$;
2. f_n converges in measure to f and $\{|f_n|^p : n \in \mathbb{N}\}$ is uniformly integrable.

□

Lemma A.1 (Fatou's lemma [92, p. 81])**.** Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and let $\{f_n : n \in \mathbb{N}\}$ be a sequence of non-negative measurable functions on Ω. Then,

$$\int_\Omega \liminf_n f_n \,\mathrm{d}\mu \leqslant \liminf_n \int_\Omega f_n \,\mathrm{d}\mu\,. \tag{A.2}$$

□

Theorem A.3 (Lebesgue differentiation theorem [130, p. 217])**.** Let E denote a *real* finite-dimensional Euclidean space and let μ denote the Lebesgue measure. If f is a vector-valued Lebesgue integrable function defined on an open set $O \subseteq E$ then

$$\lim_{\mu(B)\to 0} \frac{1}{\mu(B)} \int_B \|f(y) - f(x)\| \,\mathrm{d}\mu(y) = 0 \tag{A.3}$$

for almost all $x \in O$, where B in the limit is a closed cube containing x. Consequently,

$$f(x) = \lim_{\mu(B)\to 0} \frac{1}{\mu(B)} \int_B f(y)\,\mathrm{d}\mu(y) \tag{A.4}$$

for almost all $x \in O$. □

Lemma A.2 (Grönwall inequality [6, p. 64])**.** Let $a > b$, $u \in L^1([a,b])$ with $u \geqslant 0$, and let $\alpha \geqslant 0$. If $\beta : [a,b] \to \mathbb{R}_{\geqslant 0}$ is measurable and essentially bounded and

$$u(t) \leqslant \alpha + \int_a^t \beta(s)u(s)\,\mathrm{d}s\,, \qquad \forall t \in [a,b] \tag{A.5}$$

then it follows that $u(t) \leqslant \alpha e^{\int_a^b \beta(s)\,\mathrm{d}s}$ for all $t \in [a,b]$. □

Lemma A.3 (cf. [131, p. 22])**.** Let $\mathcal{O} \subseteq \mathbb{R}$, $\mathcal{A}$ be an indexing set, and $\{f_\alpha : \mathcal{O} \to \mathbb{R} \,|\, \alpha \in \mathcal{A}\}$ be a family of uniformly L-Lipschitz functions, i.e., for all $\alpha \in \mathcal{A}$:

$$|f_\alpha(x) - f_\alpha(y)| \leqslant L|x-y| \qquad \forall x, y \in \mathbb{R}\,.$$

Then, the function $f(x) \triangleq \sup_{\alpha\in\mathcal{A}} f_\alpha(x)$ is L-Lipschitz. □

Lemma A.4 (*Richter-Mirsky* [132])**.** For any two Hermitian matrices $\boldsymbol{A}, \boldsymbol{B} \in \mathbb{C}_{\mathrm{H}}^{d\times d}$ the following bound holds:

$$\sum_{i=1}^{d} \lambda_i^{\downarrow}(\boldsymbol{A})\lambda_i^{\uparrow}(\boldsymbol{B}) \leqslant \langle \boldsymbol{A}, \boldsymbol{B}\rangle_{\mathrm{HS}} \leqslant \sum_{i=1}^{n} \lambda_i^{\downarrow}(\boldsymbol{A})\lambda_i^{\downarrow}(\boldsymbol{B})\,. \tag{A.6}$$

□

Appendix B

Supplemental material for Theorem 2.3

This appendix collects facts that are useful to the proof of Theorem 2.3. It also provides an existence and uniqueness proof concerning the time-dependent Lindblad dynamics (2.19).

B.1 Supplemental definitions

We will need to integrate complex matrix-valued functions such as $\boldsymbol{X}(\cdot) : [a, b] \to \mathbb{C}^{n\times m}$. Measurability and integrability is defined in the canonical way, e.g., see [91, p. 325]. The function $\boldsymbol{X}(\cdot)$ is made up of $2nm$ real-valued functions, namely the coordinate functions $x^{\mathrm{r}}_{ij}, x^{\mathrm{i}}_{ij} : [a, b] \to \mathbb{R}$ defined by

$$x^{\mathrm{r}}_{ij}(t) \triangleq \mathrm{Re}[\boldsymbol{X}(t)]_{ij} \qquad \text{and} \qquad x^{\mathrm{i}}_{ij}(t) \triangleq \mathrm{Im}[\boldsymbol{X}(t)]_{ij} \tag{B.1}$$

for $i \in \mathbb{N}^n_1$ and $j \in \mathbb{N}^m_1$. The function $\boldsymbol{X}(\cdot)$ is said to be measurable if and only if each of the $2nm$ real-valued functions is measurable; $\boldsymbol{X}(\cdot)$ is said to be integrable if and only if

$$\int_a^b \|\boldsymbol{X}(t)\|_{\mathrm{HS}} < +\infty\,. \tag{B.2}$$

The integral of $\boldsymbol{X}(\cdot)$ on $[a, b]$ is defined to be the matrix whose (i, j)th entry is

$$\int_a^b x^{\mathrm{r}}_{ij}(t)\,\mathrm{d}t + \imath \int_a^b x^{\mathrm{i}}_{ij}(t)\,\mathrm{d}t\,. \tag{B.3}$$

Definition B.1 (Carathéodory function [133, p. 153])**.** Let $n, j, k \in \mathbb{N}$ and $\Omega \subset \mathbb{R}$. A function $f : \mathbb{R}^n \times \Omega \to \mathbb{C}^{j\times k}$ is said to be *Carathéodory* if (i) the mapping $\boldsymbol{x} \mapsto f(\boldsymbol{x}, t)$ is continuous for almost every $t \in \Omega$ and (ii) the mapping $t \mapsto f(\boldsymbol{x}, t)$ is measurable for every $\boldsymbol{x} \in \mathbb{R}^n$. □

A useful property of Carathéodory functions is that if $\boldsymbol{x}(\cdot) : [a, b] \to \mathbb{R}^n$ is measurable and f is Carathéodory, then the composition $f(\boldsymbol{x}(t), t)$ is a measurable function of t [133].

Let $\mathcal{W}$ denote the set of continuous matrix-valued functions $\boldsymbol{X} : [t_a, t_b] \to \mathbb{C}^{d\times d}$ on the compact interval $[t_a, t_b]$, and define the supremum norm

$$\|\boldsymbol{X}\|_{L^\infty} \triangleq \sup_{t\in[t_a,t_b]} \|\boldsymbol{X}(t)\|_{\mathrm{HS}}\,, \qquad \boldsymbol{X} \in \mathcal{W}\,. \tag{B.4}$$

It is well-known that $(\mathcal{W}, \|\cdot\|_{L^\infty})$ is a Banach space [90, p. 148].

B.2 Existence and uniqueness of solutions

Consider a differential quantum control system Σ, defined in Definition 2.6, corresponding to the dynamics (2.19). A "solution" to (2.19) is defined in the following sense.

Definition B.2 (Carathéodory solution [134, p. 4])**.** Let $[t_a, t_b] \subseteq T$, $\boldsymbol{X}_0 \in \mathbb{C}^{d\times d}$, and $\boldsymbol{u} \in \Omega$. An integrable function $\boldsymbol{X}(\cdot) : [t_a, t_b] \to \mathbb{C}^{d\times d}$ is said to be a *Carathéodory solution* to (2.19) on the interval $[t_a, t_b] \subseteq T$, subject to initial condition $\boldsymbol{X}_0$, if

$$\boldsymbol{X}(t) = \boldsymbol{X}_0 + \int_{t_a}^{t} F(\boldsymbol{X}(s), \boldsymbol{u}(s), s)\,\mathrm{d}s \tag{B.5}$$

for all $t \in [t_a, t_b]$. □

One can check that $F(\boldsymbol{X}(t), \boldsymbol{u}(t), t)$ is integrable using Definition 2.6, hence (B.5) is well-defined. It follows (e.g., by the Lebesgue differentiation theorem A.3) that a Carathéodory solution $\boldsymbol{X}(t)$ to the differential equation (2.19) satisfies the differential equation for almost all t. Indeed, for almost all $t \in (t_a, t_b)$,

$$\begin{aligned} F(\boldsymbol{X}(t), \boldsymbol{u}(t), t) &= \lim_{h\searrow 0} \frac{1}{h} \int_{t-h}^{t} F(\boldsymbol{X}(s), \boldsymbol{u}(s), s)\,\mathrm{d}s \\ &= \lim_{h\searrow 0} \frac{1}{h}\Big[\int_{t_a}^{t} F(\boldsymbol{X}(s), \boldsymbol{u}(s), s)\,\mathrm{d}s - \int_{t_a}^{t-h} F(\boldsymbol{X}(s), \boldsymbol{u}(s), s)\,\mathrm{d}s\Big] \\ &= \lim_{h\searrow 0} \frac{\boldsymbol{X}(t) - \boldsymbol{X}(t-h)}{h} = \dot{\boldsymbol{X}}(t)\,. \end{aligned} \tag{B.6}$$

Theorem B.1 (Existence of a unique Carathéodory solution)**.** Let Σ be a differential quantum control system with dynamics (2.19). Let $[t_a, t_b] \subseteq T$, $\boldsymbol{X}_0 \in \mathbb{C}^{d\times d}$, and $\boldsymbol{u} \in \Omega$. Then, there exists a unique Carathéodory solution to (2.19) on the interval $[t_a, t_b]$ with initial condition $\boldsymbol{X}_0$. □

Proof. The proof is essentially the same as that for dynamical systems on $\mathbb{R}^n$, e.g., see [21, 134].

(Existence) Define a sequence $\{\boldsymbol{X}^{(k)}(\cdot) : k = 0, 1, \ldots\}$ of continuous functions with $\boldsymbol{X}^{(0)}(t) \triangleq \boldsymbol{X}_0$ and

$$\boldsymbol{X}^{(k)}(t) \triangleq \boldsymbol{X}_0 + \int_{t_a}^{t} F(\boldsymbol{X}^{(k-1)}(s), \boldsymbol{u}(s), s)\,\mathrm{d}s \tag{B.7}$$

for all $t \in [t_a, t_b]$ and $k \in \mathbb{N}$. It will be shown that the limit of the $\boldsymbol{X}^{(k)}(\cdot)$'s exists and that it is a Carathéodory solution to (2.19).

By Definition 2.6, $\|\boldsymbol{H}(\boldsymbol{u}(t), t)\|_{\mathrm{HS}}$ and $\|\boldsymbol{L}_j(\boldsymbol{u}(t), t)\|_{\mathrm{HS}}^2$ are locally integrable. Hence, the non-negative measurable function $m : [t_a, t_b] \to \mathbb{R}$ defined by

$$m(t) \triangleq \frac{2}{\hbar}\|\boldsymbol{H}(\boldsymbol{u}(t), t)\|_{\mathrm{HS}} + 4\sum_{j\in\mathcal{J}} \|\boldsymbol{L}_j(\boldsymbol{u}(t), t)\|_{\mathrm{HS}}^2 \tag{B.8}$$

is integrable. Submultiplicativity of the Hilbert-Schmidt norm alongside the definition (2.19) of F gives

$$\|F(\boldsymbol{X}, \boldsymbol{u}(t), t)\|_{\mathrm{HS}} \leqslant m(t)\|\boldsymbol{X}\|_{\mathrm{HS}}\,, \qquad \forall (\boldsymbol{X}, t) \in \mathbb{C}^{d\times d} \times [t_a, t_b]\,. \tag{B.9}$$

By definition (B.7) of the $\boldsymbol{X}^{(k)}(\cdot)$'s and linearity of $F(\cdot, \boldsymbol{u}(t), t)$,

$$\begin{aligned}
\|\boldsymbol{X}^{(k)}(t) - \boldsymbol{X}^{(k-1)}(t)\|_{\mathrm{HS}} &= \Big\| \int_{t_a}^{t} F(\boldsymbol{X}^{(k)}(s) - \boldsymbol{X}^{(k-1)}(s), \boldsymbol{u}(s), s)\,\mathrm{d}s \Big\|_{\mathrm{HS}} \\
&\leqslant \int_{t_a}^{t} \|F(\boldsymbol{X}^{(k)}(s) - \boldsymbol{X}^{(k-1)}(s), \boldsymbol{u}(s), s)\|_{\mathrm{HS}}\,\mathrm{d}s \\
&\leqslant \int_{t_a}^{t} m(s)\,\|\boldsymbol{X}^{(k)}(s) - \boldsymbol{X}^{(k-1)}(s)\|_{\mathrm{HS}}\,\mathrm{d}s\,.
\end{aligned} \tag{B.10}$$

Define $\zeta \geqslant 0$ and $\eta : [t_a, t_b] \to [0, \infty)$ according to

$$\zeta \triangleq \max_{s\in[t_a,t_b]} \|\boldsymbol{X}^{(1)}(s) - \boldsymbol{X}_0\|_{\mathrm{HS}} \tag{B.11a}$$

$$\eta(t) \triangleq \int_{t_a}^{t} m(s)\,\mathrm{d}s\,. \tag{B.11b}$$

Observe that, for any $t \in [t_a, t_b]$,

$$\|\boldsymbol{X}^{(2)}(t) - \boldsymbol{X}^{(1)}(t)\|_{\mathrm{HS}} \leqslant \zeta \int_{t_a}^{t} m(s)\,\mathrm{d}s = \zeta\,\eta(t) \tag{B.12a}$$

$$\begin{aligned}
\|\boldsymbol{X}^{(3)}(t) - \boldsymbol{X}^{(2)}(t)\|_{\mathrm{HS}} &\leqslant \int_{t_a}^{t} m(s_2)\,\|\boldsymbol{X}^{(2)}(s_2) - \boldsymbol{X}^{(1)}(s_2)\|_{\mathrm{HS}}\,\mathrm{d}s_2 \\
&\leqslant \zeta \int_{t_a}^{t} m(s_2) \int_{t_a}^{s_2} m(s_1)\,\mathrm{d}s_1\,\mathrm{d}s_2 \\
&= \zeta \int_{t_a}^{t} \frac{\mathrm{d}}{\mathrm{d}s_2}\Big[\frac{\eta^2(s_2)}{2}\Big]\,\mathrm{d}s_2 \\
&= \zeta \frac{\eta^2(t)}{2}\,.
\end{aligned} \tag{B.12b}$$

Iterating this procedure, it follows that

$$\|\boldsymbol{X}^{(k+1)}(t) - \boldsymbol{X}^{(k)}(t)\|_{\mathrm{HS}} \leqslant \zeta\,\frac{\eta^k(t)}{k!} \leqslant \zeta\,\frac{\eta^k(t_b)}{k!}\,, \qquad \forall (k, t) \in \mathbb{N} \times [t_a, t_b]\,. \tag{B.13}$$

For any $k \in \mathbb{N}$, it is possible to write the function $\boldsymbol{X}^{(k)}(t)$ as a sum of increments:

$$\boldsymbol{X}^{(k)}(t) = \boldsymbol{X}_0 + \sum_{i=1}^{k} \boldsymbol{X}^{(i)}(t) - \boldsymbol{X}^{(i-1)}(t)\,. \tag{B.14}$$

Let $\epsilon > 0$ and choose a sufficiently large $K \in \mathbb{N}$ such that $\sum_{i=K+1}^{\infty} \eta^i(t_b)/i! < \epsilon/\zeta$.[1] Then, for any $k, m \geqslant K$ and $t \in [t_a, t_b]$ it follows that

$$\begin{aligned}
\|\boldsymbol{X}^{(k)}(t) - \boldsymbol{X}^{(m)}(t)\|_{\mathrm{HS}} &= \Big\| \sum_{i=(k\wedge m)+1}^{k\vee m} \boldsymbol{X}^{(i)}(t) - \boldsymbol{X}^{(i-1)}(t) \Big\|_{\mathrm{HS}} \\
&\leqslant \sum_{i=(k\wedge m)+1}^{\infty} \|\boldsymbol{X}^{(i)}(t) - \boldsymbol{X}^{(i-1)}(t)\|_{\mathrm{HS}} \\
&\leqslant \zeta \sum_{i=(k\wedge m)+1}^{\infty} \frac{\eta^i(t_b)}{i!} \\
&\leqslant \epsilon
\end{aligned} \tag{B.15}$$

and therefore

$$\|\boldsymbol{X}^{(k)} - \boldsymbol{X}^{(m)}\|_{L^\infty} = \sup_{t\in[t_a,t_b]} \|\boldsymbol{X}^{(k)}(t) - \boldsymbol{X}^{(m)}(t)\|_{\mathrm{HS}} \leqslant \epsilon\,. \tag{B.16}$$

This proves that the sequence of functions $\{\boldsymbol{X}^{(k)}(\cdot) : k = 0, 1, \ldots\}$ is Cauchy. Since the space $(\mathcal{W}, \|\cdot\|_{L^\infty})$ of continuous matrix-valued functions equipped with the supremum norm is a Banach space, this implies that the $\boldsymbol{X}^{(k)}(\cdot)$'s converge uniformly to a continuous function $\boldsymbol{X} : [t_a, t_b] \to \mathbb{C}^{d\times d}$.

Using the sum of increments expression (B.14) it follows that the $\boldsymbol{X}^{(k)}(\cdot)$'s are bounded on the compact interval $[t_a, t_b]$:

$$\|\boldsymbol{X}^{(k)}(t)\|_{\mathrm{HS}} \leqslant \|\boldsymbol{X}_0\|_{\mathrm{HS}} + \zeta e^{\eta(t_b)} < +\infty\,. \tag{B.17}$$

By the bounded convergence theorem (Theorem A.1), the limiting function $\boldsymbol{X}(\cdot)$ is a Carathéodory solution to (2.19):

$$\boldsymbol{X}(t) = \lim_{k\to\infty} \boldsymbol{X}^{(k)}(t) = \boldsymbol{X}_0 + \lim_{k\to\infty} \int_{t_a}^{t} F(\boldsymbol{X}^{(k)}(s), \boldsymbol{u}(s), s)\,\mathrm{d}s = \boldsymbol{X}_0 + \int_{t_a}^{t} F(\boldsymbol{X}(s), \boldsymbol{u}(s), s)\,\mathrm{d}s\,. \tag{B.18}$$

(Uniqueness) Suppose that $\boldsymbol{X}^{(1)}(\cdot)$ and $\boldsymbol{X}^{(2)}(\cdot)$ are two solutions to (2.19) on the interval $[t_a, t_b]$ starting from initial condition $\boldsymbol{X}_0$. Then, for all $t \in [t_a, t_b]$,

$$\begin{aligned}
\|\boldsymbol{X}^{(1)}(t) - \boldsymbol{X}^{(2)}(t)\|_{\mathrm{HS}} &\leqslant \int_{t_a}^{t} \|F(\boldsymbol{X}^{(1)}(s) - \boldsymbol{X}^{(2)}(s), \boldsymbol{u}(s), s)\|_{\mathrm{HS}}\,\mathrm{d}s \\
&\leqslant \int_{t_a}^{t} m(s)\|\boldsymbol{X}^{(1)}(s) - \boldsymbol{X}^{(2)}(s)\|_{\mathrm{HS}}\,\mathrm{d}s\,.
\end{aligned} \tag{B.19}$$

By Grönwall's inequality (Lemma A.2) it follows that $\|\boldsymbol{X}^{(1)}(t) - \boldsymbol{X}^{(2)}(t)\|_{\mathrm{HS}} = 0$ for all $t \in [t_a, t_b]$. Hence, the solution (2.19) is unique. ⊠

[1] This can be done because the series $e^{\eta(t_b)} = \sum_{i=0}^{\infty} \eta^i(t_b)/i!$ converges.

B.3 Convergence proofs

Step (i) of Theorem 2.3's proof relies on the fact that $\|A^{(k)}(t) - A(t)\|_{\text{op}} \to 0$ as $k \to \infty$ for all $t \in [0,1]$. Step (ii) of Theorem 2.3's proof relies on the fact that $\int_0^1 \|A^{(k)}(s) - A(s)\|_{\text{op}}\, \mathrm{d}s \to 0$ as $k \to \infty$. This section serves to prove said facts. Recall that the operator norm on $\mathcal{L}(\mathbb{C}^{d\times d})$ induced by the Hilbert-Schmidt norm is defined as

$$\|B\|_{\text{op}} = \sup\left\{ \frac{\|B(\boldsymbol{X})\|_{\text{HS}}}{\|\boldsymbol{X}\|_{\text{HS}}} \;:\; \boldsymbol{X} \in \mathbb{C}^{d\times d},\, \boldsymbol{X} \neq \boldsymbol{0} \right\}, \qquad B \in \mathcal{L}(\mathbb{C}^{d\times d})\,. \tag{B.20}$$

The following lemma will be needed.

Lemma B.1. Let $i \in \{1,2\}$, $\boldsymbol{A}_i \in \mathbb{C}^{d\times d}$, and $K_i, L_i, M_i \in \mathcal{L}(\mathbb{C}^{d\times d})$ with

$$K_i(\boldsymbol{X}) = -\tfrac{\imath}{\hbar}[\![\boldsymbol{A}_i, \boldsymbol{X}]\!]_-\,, \quad L_i(\boldsymbol{X}) = \boldsymbol{A}_i \boldsymbol{X} \boldsymbol{A}_i^\dagger\,, \quad M_i(\boldsymbol{X}) = \boldsymbol{A}_i^\dagger \boldsymbol{A}_i \boldsymbol{X} \tag{B.21}$$

for all $\boldsymbol{X} \in \mathbb{C}^{d\times d}$. Then,

$$\|K_1 - K_2\|_{\text{op}} \leqslant \tfrac{2}{\hbar}\|\boldsymbol{A}_1 - \boldsymbol{A}_2\|_{\text{HS}} \tag{B.22a}$$

$$\|L_1 - L_2\|_{\text{op}} \leqslant \big[\|\boldsymbol{A}_1\|_{\text{HS}} + \|\boldsymbol{A}_2\|_{\text{HS}}\big]\|\boldsymbol{A}_1 - \boldsymbol{A}_2\|_{\text{HS}} \tag{B.22b}$$

$$\|M_1 - M_2\|_{\text{op}} \leqslant \big[\|\boldsymbol{A}_1\|_{\text{HS}} + \|\boldsymbol{A}_2\|_{\text{HS}}\big]\|\boldsymbol{A}_1 - \boldsymbol{A}_2\|_{\text{HS}}\,. \tag{B.22c}$$

□

Proof. The proofs of (B.22a)-(B.22c) are entirely similar, so only that of (B.22b) is provided. Let $\boldsymbol{X} \in \mathbb{C}^{d\times d}$ with $\boldsymbol{X} \neq \boldsymbol{0}$. The submultiplicative property of the Hilber-Schmidt norm provides the bound

$$\begin{aligned}
\|L_1(\boldsymbol{X}) - L_2(\boldsymbol{X})\|_{\text{HS}} &= \|\boldsymbol{A}_1\boldsymbol{X}\boldsymbol{A}_1^\dagger - \boldsymbol{A}_2\boldsymbol{X}\boldsymbol{A}_2^\dagger\|_{\text{HS}} \\
&= \|\boldsymbol{A}_1\boldsymbol{X}\boldsymbol{A}_1^\dagger - \boldsymbol{A}_2\boldsymbol{X}\boldsymbol{A}_1^\dagger + \boldsymbol{A}_2\boldsymbol{X}\boldsymbol{A}_1^\dagger - \boldsymbol{A}_2\boldsymbol{X}\boldsymbol{A}_2^\dagger\|_{\text{HS}} \\
&\leqslant \|(\boldsymbol{A}_1 - \boldsymbol{A}_2)\boldsymbol{X}\boldsymbol{A}_1^\dagger\|_{\text{HS}} + \|\boldsymbol{A}_2\boldsymbol{X}(\boldsymbol{A}_1^\dagger - \boldsymbol{A}_2^\dagger)\|_{\text{HS}} \\
&\leqslant \|\boldsymbol{A}_1 - \boldsymbol{A}_2\|_{\text{HS}}\,\|\boldsymbol{X}\|_{\text{HS}}\,\|\boldsymbol{A}_1^\dagger\|_{\text{HS}} + \|\boldsymbol{A}_2\|_{\text{HS}}\,\|\boldsymbol{X}\|_{\text{HS}}\,\|\boldsymbol{A}_1^\dagger - \boldsymbol{A}_2^\dagger\|_{\text{HS}}\,.
\end{aligned} \tag{B.23}$$

Dividing both sides of (B.23) by $\|\boldsymbol{X}\|_{\text{HS}}$ proves (B.22b). ⊠

Fix $\boldsymbol{u} \in \Omega$ and $[t_a, t_b] = [0,1]$ as in the proof of Theorem 2.3. Recall the definition of the operator $A(t) \in \mathcal{L}(\mathbb{C}^{d\times d})$:

$$\begin{aligned}
A(t)\boldsymbol{X} &\triangleq F(\boldsymbol{X}, \boldsymbol{u}(t), t) \\
&= -\frac{\imath}{\hbar}[\![\boldsymbol{H}(\boldsymbol{u}(t),t), \boldsymbol{X}(t)]\!]_- \\
&\quad + \sum_{j\in\mathcal{J}} 2\boldsymbol{L}_j(\boldsymbol{u}(t),t)\boldsymbol{X}(t)\boldsymbol{L}_j^\dagger(\boldsymbol{u}(t),t) - [\![\boldsymbol{L}_j^\dagger(\boldsymbol{u}(t),t)\boldsymbol{L}_j(\boldsymbol{u}(t),t), \boldsymbol{X}(t)]\!]_+
\end{aligned} \tag{B.24}$$

for $\boldsymbol{X} \in \mathbb{C}^{d\times d}$. In Step (i) of Theorem 2.3's proof, the Hamiltonian $\boldsymbol{H}(\boldsymbol{u}(t),t)$ and Lindblad operators $\boldsymbol{L}_j(\boldsymbol{u}(t),t)$ are taken to be continuous functions of time. The operator $A(t)$ is approximated by a sequence of operators $A^{(k)}(t) \in \mathcal{L}(\mathbb{C}^{d\times d})$ defined as

$$A^{(k)}(t) \triangleq \sum_{i=1}^{k} \mathbb{1}_{I_{i,k}}(t)\, A(\tfrac{i-1}{k})\,, \qquad k \in \mathbb{N} \tag{B.25}$$

where $I_{i,k} \triangleq [\frac{i-1}{k}, \frac{i}{k})$. Equivalently, $A^{(k)}(t)$ may be written as

$$A^{(k)}(t)\boldsymbol{X} = -\frac{\imath}{\hbar}[\![\boldsymbol{H}^{(k)}(t), \boldsymbol{X}]\!]_- + \sum_{j\in\mathcal{J}} 2\boldsymbol{L}_j^{(k)}(t)\boldsymbol{X}\boldsymbol{L}_j^{(k)\dagger}(t) - [\![\boldsymbol{L}_j^{(k)\dagger}(t)\boldsymbol{L}_j^{(k)}(t), \boldsymbol{X}]\!]_+ \quad \text{(B.26)}$$

for $\boldsymbol{X} \in \mathbb{C}^{d\times d}$ with

$$\boldsymbol{H}^{(k)}(t) \triangleq \sum_{i=1}^{k} \mathbb{1}_{I_{i,k}}(t)\, \boldsymbol{H}(\boldsymbol{u}(\tfrac{i-1}{k}), \tfrac{i-1}{k}) \quad \text{(B.27a)}$$

$$\boldsymbol{L}_j^{(k)}(t) \triangleq \sum_{i=1}^{k} \mathbb{1}_{I_{i,k}}(t)\, \boldsymbol{L}_j(\boldsymbol{u}(\tfrac{i-1}{k}), \tfrac{i-1}{k})\,. \quad \text{(B.27b)}$$

Continuity implies that

$$\sigma \triangleq \max_{j\in\mathcal{J}} \Big[\max_{t\in[0,1]} \|\boldsymbol{L}_j(\boldsymbol{u}(t), t)\|_{\mathrm{HS}} \Big] < \infty\,. \quad \text{(B.28)}$$

Proof that $\|A^{(k)}(t) - A(t)\|_{\mathrm{op}} \to 0$ for each $t \in [0,1]$ as $k \to \infty$ follows from the triangle inequality and Lemma B.1:

$$\begin{aligned}\|A^{(k)}(t) - A(t)\|_{\mathrm{op}} &\leqslant \tfrac{2}{\hbar}\|\boldsymbol{H}^{(k)}(t) - \boldsymbol{H}(\boldsymbol{u}(t), t)\|_{\mathrm{HS}} + \sum_{j\in\mathcal{J}} 4\Big[\|\boldsymbol{L}_j^{(k)}(t)\|_{\mathrm{HS}} + \|\boldsymbol{L}_j(\boldsymbol{u}(t), t)\|_{\mathrm{HS}}\Big] \\ &\qquad \times \|\boldsymbol{L}_j^{(k)}(t) - \boldsymbol{L}_j(\boldsymbol{u}(t), t)\|_{\mathrm{HS}} \quad \text{(B.29a)} \\ &\leqslant \tfrac{2}{\hbar}\|\boldsymbol{H}^{(k)}(t) - \boldsymbol{H}(\boldsymbol{u}(t), t)\|_{\mathrm{HS}} + \sum_{j\in\mathcal{J}} 8\sigma \|\boldsymbol{L}_j^{(k)}(t) - \boldsymbol{L}_j(\boldsymbol{u}(t), t)\|_{\mathrm{HS}}\,. \quad \text{(B.29b)}\end{aligned}$$

Continuity implies that $\|\boldsymbol{H}^{(k)}(t) - \boldsymbol{H}(\boldsymbol{u}(t), t)\|_{\mathrm{HS}} \to 0$ and $\|\boldsymbol{L}_j^{(k)}(t) - \boldsymbol{L}_j(\boldsymbol{u}(t), t)\|_{\mathrm{HS}} \to 0$ as $k \to \infty$ for each $t \in [0,1]$. Combined with (B.29b), it follows that $\|A^{(k)}(t) - A(t)\|_{\mathrm{op}} \to 0$ for each $t \in [0,1]$ as $k \to \infty$.

In Step (ii) of Theorem 2.3's proof $A^{(k)}(t)$ is defined as in (B.26) with $\boldsymbol{H}^{(k)}(t)$ and $\boldsymbol{L}_j^{(k)}(t)$ defined as

$$\boldsymbol{H}^{(k)}(t) \triangleq \frac{1}{2h_k}\int_{t-h_k}^{t+h_k} \boldsymbol{H}(\boldsymbol{u}(s), s)\, \mathbb{1}_{[0,1]}(s)\, \mathrm{d}s \quad \text{(B.30a)}$$

$$\boldsymbol{L}_j^{(k)}(t) \triangleq \frac{1}{2h_k}\int_{t-h_k}^{t+h_k} \boldsymbol{L}_j(\boldsymbol{u}(s), s)\, \mathbb{1}_{[0,1]}(s)\, \mathrm{d}s\,. \quad \text{(B.30b)}$$

The goal is to show that $\int_0^1 \|A(s) - A^{(k)}(s)\|_{\mathrm{op}}\, \mathrm{d}s \to 0$ as $k \to \infty$. As before,

$$\begin{aligned}\|A^{(k)}(t) - A(t)\|_{\mathrm{op}} &\leqslant \tfrac{2}{\hbar}\|\boldsymbol{H}^{(k)}(t) - \boldsymbol{H}(\boldsymbol{u}(t), t)\|_{\mathrm{HS}} + \sum_{j\in\mathcal{J}} 4\Big[\|\boldsymbol{L}_j^{(k)}(t)\|_{\mathrm{HS}} + \|\boldsymbol{L}_j(\boldsymbol{u}(t), t)\|_{\mathrm{HS}}\Big] \\ &\qquad \times \|\boldsymbol{L}_j^{(k)}(t) - \boldsymbol{L}_j(\boldsymbol{u}(t), t)\|_{\mathrm{HS}}\,. \quad \text{(B.31)}\end{aligned}$$

The Cauchy-Schwarz and triangle inequalities imply that

$$\begin{aligned}
&\int_0^1 \|A^{(k)}(t) - A(t)\|_{\mathrm{op}}\,\mathrm{d}t \\
&\quad\leqslant \tfrac{2}{\hbar}\|\boldsymbol{H}^{(k)}(\cdot) - \boldsymbol{H}(\boldsymbol{u}(\cdot),\cdot)\|_{L^1} + \sum_{j\in\mathcal{J}} 4\Big\|\|\boldsymbol{L}_j^{(k)}(\cdot)\|_{\mathrm{HS}} + \|\boldsymbol{L}_j(\boldsymbol{u}(\cdot),\cdot)\|_{\mathrm{HS}}\Big\|_{L^2} \\
&\qquad\qquad \times\|\boldsymbol{L}_j^{(k)}(\cdot) - \boldsymbol{L}_j(\boldsymbol{u}(\cdot),\cdot)\|_{L^2} \qquad \text{(B.32)} \\
&\quad\leqslant \tfrac{2}{\hbar}\|\boldsymbol{H}^{(k)}(\cdot) - \boldsymbol{H}(\boldsymbol{u}(\cdot),\cdot)\|_{L^1} + \sum_{j\in\mathcal{J}} 4\Big[\|\boldsymbol{L}_j^{(k)}(\cdot)\|_{L^2} + \|\boldsymbol{L}_j(\boldsymbol{u}(\cdot),\cdot)\|_{L^2}\Big] \\
&\qquad\qquad \times\|\boldsymbol{L}_j^{(k)}(\cdot) - \boldsymbol{L}_j(\boldsymbol{u}(\cdot),\cdot)\|_{L^2}\,. \qquad \text{(B.33)}
\end{aligned}$$

It can be verified [92, p. 183] that

$$\lim_{k\to\infty} \|\boldsymbol{H}^{(k)}(\cdot) - \boldsymbol{H}(\boldsymbol{u}(\cdot),\cdot)\|_{L^1} = 0 \qquad \text{(B.34a)}$$

$$\lim_{k\to\infty} \|\boldsymbol{L}_j^{(k)}(\cdot) - \boldsymbol{L}_j(\boldsymbol{u}(\cdot),\cdot)\|_{L^2} = 0 \qquad \text{(B.34b)}$$

since $\boldsymbol{H}(\boldsymbol{u}(\cdot),\cdot) \in L^1$ and $\boldsymbol{L}_j(\boldsymbol{u}(\cdot),\cdot) \in L^2$. Moreover, Young's inequality [92, p. 179] provides $\|\boldsymbol{L}_j^{(k)}(\cdot)\|_{L^2} \leqslant \|\boldsymbol{L}_j(\boldsymbol{u}(\cdot),\cdot)\|_{L^2}$. Thus,

$$\lim_{k\to\infty} \int_0^1 \|A^{(k)}(t) - A(t)\|_{\mathrm{op}}\,\mathrm{d}t = 0 \qquad \text{(B.35)}$$

as desired.

Bibliography

[1] M. Clouâtré, S. Marano, P. L. Falb, and M. Z. Win, "Admissible optimal control for parameter estimation in quantum systems," *IEEE Contr. Syst. Lett.*, vol. 8, pp. 2283–2288, Sep. 2024.

[2] M. Clouâtré, S. Marano, and M. Z. Win, "Information control and online estimation for unital quantum systems," in *Proc. IEEE Conf. Decis. and Contr.*, Rio de Janeiro, Brazil, Dec. 2025, pp. 6697–6703.

[3] M. Clouâtré, S. Marano, A. Conti, P. L. Falb, and M. Z. Win, "Optimal quantum state discrimination with fixed measurements," in *Proc. Int. Conf. on Quantum Commun., Netw., and Comput. (QCNC)*, Nara, Japan, Apr. 2025, pp. 247–253.

[4] A. Giani, M. Z. Win, P. L. Falb, and A. Conti, "Unveiling distinguishable non-Gaussian quantum states," *Phys. Rev. A*, pp. 1–16, Jan 2026, to appear.

[5] R. E. Kalman, P. L. Falb, and M. A. Arbib, *Topics in Mathematical System Theory*. New York: McGraw-Hill, 1969.

[6] P. Falb, *Direct Methods in Control Problems*. New York: Springer, 2019.

[7] R. A. Fisher, "Theory of statistical estimation," in *Math. Proc. Cambridge Phil. Soc.*, vol. 22, no. 5. Cambridge University Press, 1925, pp. 700–725.

[8] D. Petz and C. Ghinea, "Introduction to quantum Fisher information," in *Quantum probability and related topics*. World Scientific, 2011, pp. 261–281.

[9] J. Liu, H. Yuan, X.-M. Lu, and X. Wang, "Quantum Fisher information matrix and multiparameter estimation," *J. Phys. A: Math. Theor.*, vol. 53, no. 2, p. 023001, 2020.

[10] S. L. Braunstein and C. M. Caves, "Statistical distance and the geometry of quantum states," *Phys. Rev. Lett.*, vol. 72, no. 22, p. 3439, 1994.

[11] B. Kartal, "*Theoretical Limits of Quantum Ranging*," Master's thesis, Department of Aeronautics and Astronautics, Massachusetts Institute of Technology, Cambridge, MA, May 2025, thesis advisor: Professor Moe Z. Win.

[12] C. Helstrom, *Quantum detection and estimation theory*. New York, USA: Academic Press, 1976.

[13] A. Giani, "Quantum sensing and communication via non-Gaussian states," Ph.D. dissertation, University of Ferrara, Ferrara, Italy, 2024, thesis advisor: Professor Andrea Conti.

[14] H. Ma, D. Dong, I. R. Petersen, C.-J. Huang, and G.-Y. Xiang, "Neural networks for quantum state tomography with constrained measurements," *Quantum Inf. Process.*, vol. 23, no. 9, p. 317, 2024.

[15] F. G. Brandao, A. W. Harrow, J. R. Lee, and Y. Peres, "Adversarial hypothesis testing and a quantum Stein's lemma for restricted measurements," *IEEE Trans. Inf. Theory*, vol. 66, no. 8, pp. 5037–5054, 2020.

[16] C. Lancien and A. Winter, "Distinguishing multi-partite states by local measurements," *Comm. Math. Phys.*, vol. 323, pp. 555–573, 2013.

[17] W. Matthews, M. Piani, and J. Watrous, "Entanglement in channel discrimination with restricted measurements," *Phys. Rev. A*, vol. 82, no. 3, p. 032302, 2010.

[18] W. Matthews, S. Wehner, and A. Winter, "Distinguishability of quantum states under restricted families of measurements with an application to quantum data hiding," *Comm. Math. Phys.*, vol. 291, pp. 813–843, 2009.

[19] R. Malaney, "The quantum car," *IEEE Wireless Commun. Lett.*, vol. 5, no. 6, pp. 624–627, 2016.

[20] M. Athans and P. L. Falb, *Optimal Control: An Introduction to the Theory and Its Applications*. New York: McGraw-Hill, 1966.

[21] R. W. Brockett, *Finite dimensional linear systems*. John Wiley and Sons, 1970.

[22] E. D. Sontag, *Mathematical control theory: deterministic finite dimensional systems*. Springer Science & Business Media, 2013, vol. 6.

[23] S. Lloyd, "Quantum controllers for quantum systems," *arXiv preprint quant-ph/9703042*, 1997.

[24] ——, "Coherent quantum feedback," *Phys. Rev. A*, vol. 62, no. 2, p. 022108, 2000.

[25] D. Dong and I. R. Petersen, *Learning and robust control in quantum technology*. Springer, 2023.

[26] D. D'Alessandro, *Introduction to Quantum Control and Dynamics*. Chapman and Hall/CRC, 2022.

[27] S. Cong, *Control of quantum systems: theory and methods*. John Wiley & Sons, 2014.

[28] H. M. Wiseman and G. J. Milburn, *Quantum measurement and control*. Cambridge university press, 2009.

[29] J. Berberich and D. Fink, "Quantum computing through the lens of control: A tutorial introduction," *IEEE Control Syst. Mag.*, vol. 44, no. 6, pp. 24–49, 2024.

[30] J. Berberich, R. L. Kosut, and T. Schulte-Herbrüggen, "Bringing quantum systems under control: A tutorial invitation to quantum computing and its relation to bilinear control systems," in *Proc. IEEE Conf. Decis. and Contr.* IEEE, 2024, pp. 5231–5247.

[31] M. M. Nicotra, J. Shao, J. Combes, A. C. Theurkauf, P. Axelrad, L.-Y. Chih, M. Holland, A. A. Zozulya, C. K. LeDesma, K. Mehling *et al.*, "Modeling and control of ultracold atoms trapped in an optical lattice: An example-driven tutorial on quantum control," *IEEE Control Syst. Mag.*, vol. 43, no. 1, pp. 28–43, 2023.

[32] C. P. Koch, U. Boscain, T. Calarco, G. Dirr, S. Filipp, S. J. Glaser, R. Kosloff, S. Montangero, T. Schulte-Herbrüggen, D. Sugny *et al.*, "Quantum optimal control in quantum technologies. Strategic report on current status, visions and goals for research in Europe," *EPJ Quant. Tech.*, vol. 9, no. 1, p. 19, 2022.

[33] C. Altafini and F. Ticozzi, "Modeling and control of quantum systems: An introduction," *IEEE Trans. Autom. Control*, vol. 57, no. 8, pp. 1898–1917, 2012.

[34] L. Lokutsievskiy and A. Pechen, "Reachable sets for two-level open quantum systems driven by coherent and incoherent controls," *J. Phys. A: Math. Theor.*, vol. 54, no. 39, p. 395304, 2021.

[35] A. Pechen, "Engineering arbitrary pure and mixed quantum states," *Phys. Rev. A*, vol. 84, no. 4, p. 042106, 2011.

[36] D. Dong, C. Zhang, H. Rabitz, A. Pechen, and T.-J. Tarn, "Incoherent control of locally controllable quantum systems," *J. Chem. Phys.*, vol. 129, no. 15, 2008.

[37] A. Pechen and H. Rabitz, "Teaching the environment to control quantum systems," *Phys. Rev. A*, vol. 73, no. 6, p. 062102, 2006.

[38] P. Rouchon and A. Sarlette, "Contraction and stability analysis of steady-states for open quantum systems described by Lindblad differential equations," in *Proc. IEEE Conf. Decis. and Contr.* IEEE, 2013, pp. 6568–6573.

[39] Q. Ansel, E. Dionis, F. Arrouas, B. Peaudecerf, S. Guérin, D. Guéry-Odelin, and D. Sugny, "Introduction to theoretical and experimental aspects of quantum optimal control," *J. Phys. B*, vol. 57, no. 13, p. 133001, 2024.

[40] U. Boscain, M. Sigalotti, and D. Sugny, "Introduction to the Pontryagin maximum principle for quantum optimal control," *PRX Quantum*, vol. 2, no. 3, p. 030203, 2021.

[41] N. Khaneja, R. Brockett, and S. J. Glaser, "Time optimal control in spin systems," *Phys. Rev. A*, vol. 63, no. 3, p. 032308, 2001.

[42] U. G. Abdulla, J. Rodrigues, P. Jimenez, C. Zhen, and C. Martino, "Bang-bang optimal control in coherent spin dynamics of radical pairs in quantum biology," *Quantum Sci. Tech.*, vol. 9, no. 4, p. 045022, 2024.

[43] U. G. Abdulla, J. H. Rodrigues, and J.-J. Slotine, "Quantum optimal control for coherent spin dynamics of radical pairs via Pontryagin maximum principle," *arXiv preprint arXiv:2508.01806*, 2025.

[44] J. Shao, M. Naris, J. Hauser, and M. M. Nicotra, "Solving quantum optimal control problems using projection-operator-based Newton steps," *Phys. Rev. A*, vol. 109, no. 1, p. 012609, 2024.

[45] M. Clouâtré, M. J. Khojasteh, and M. Z. Win, "Model-predictive quantum control via Hamiltonian learning," *IEEE Trans. Quantum Eng.*, vol. 3, pp. 1–23, Oct. 2022, Art no. 4100623.

[46] J.-S. Li, J. Ruths, T.-Y. Yu, H. Arthanari, and G. Wagner, "Optimal pulse design in quantum control: A unified computational method," *Proceedings of the National Academy of Sciences*, vol. 108, no. 5, pp. 1879–1884, 2011.

[47] N. Khaneja, T. Reiss, C. Kehlet, T. Schulte-Herbrüggen, and S. J. Glaser, "Optimal control of coupled spin dynamics: design of NMR pulse sequences by gradient ascent algorithms," *J. Mag. Res.*, vol. 172, no. 2, pp. 296–305, 2005.

[48] R. Chakrabarti and H. Rabitz, "Quantum control landscapes," *Int. Rev. Phys. Chem.*, vol. 26, no. 4, pp. 671–735, 2007.

[49] L. Hanzo, Z. Babar, Z. Cai, D. Chandra, I. B. Djordjevic, B. Koczor, S. X. Ng, M. Razavi, and O. Simeone, "Quantum information processing, sensing, and communications: Their myths, realities, and futures," *Proc. IEEE*, 2025, early access.

[50] M. S. Winnel, Z. Wang, R. Malaney, R. Aguinaldo, J. Green, and T. C. Ralph, "Classical-quantum dual encoding for laser communications in space," *New J. Phys.*, vol. 26, no. 3, p. 033012, 2024.

[51] A. S. Cacciapuoti, M. Caleffi, R. Van Meter, and L. Hanzo, "When entanglement meets classical communications: Quantum teleportation for the Quantum Internet," *IEEE Trans. Commun.*, vol. 68, no. 6, pp. 3808–3833, Jun. 2020.

[52] M. R. Bloch, "Covert communication over noisy channels: A resolvability perspective," *IEEE Trans. Inf. Theory*, vol. 62, no. 5, pp. 2334–2354, 2016.

[53] U. Keskin, "*Multipartite Quantum Clock Synchronization via Collective Symmetric States*," Master's thesis, Department of Aeronautics and Astronautics, Massachusetts Institute of Technology, Cambridge, MA, May 2025, thesis advisor: Professor Moe Z. Win.

[54] A. Conti, R. Malaney, and M. Z. Win, "Satellite-terrestrial quantum networks and the global quantum internet," *IEEE Commun. Mag.*, vol. 62, no. 10, pp. 34–39, Oct. 2024, special issue on *Integrated Non-Terrestrial and Terrestrial (NTN/TN) Quantum Networks*.

[55] L. Yang, Y. Zhao, H. Xu, L. Huang, and C. Qiao, "Dynamic entanglement routing based on stream processing for quantum networks," *IEEE/ACM Trans. Netw.*, vol. 33, no. 1, pp. 382–397, 2024.

[56] Z. Liu, S. Marano, and M. Z. Win, "Establishing high-fidelity entanglement in quantum repeater chains," *IEEE J. Sel. Areas Commun.*, vol. 42, no. 7, pp. 1763–1778, Jul. 2024, special issue on *The Quantum Internet: Principles, Protocols, and Architectures*.

[57] S. Marano and M. Z. Win, "Distributing quantum states with finite lifetime," *Phys. Rev. A*, vol. 107, p. 052413, May 2023.

[58] J. Illiano, M. Caleffi, A. Manzalini, and A. S. Cacciapuoti, "Quantum Internet protocol stack: A comprehensive survey," *Computer Networks*, vol. 213, p. 109092, Aug. 2022.

[59] A. Giani, M. Z. Win, and A. Conti, "Quantum sensing and communication via non-Gaussian states," *IEEE J. Sel. Areas Inf. Theory*, vol. 6, pp. 18–33, Mar. 2025, special issue dedicated to the memory of Toby Berger on *Data, Physics, and Life Through the Lens of Information Theory*.

[60] F. Poggiali, P. Cappellaro, and N. Fabbri, "Optimal control for one-qubit quantum sensing," *Phys. Rev. X*, vol. 8, no. 2, p. 021059, 2018.

[61] C. L. Degen, F. Reinhard, and P. Cappellaro, "Quantum sensing," *Rev. Mod. Phys.*, vol. 89, no. 3, p. 035002, 2017.

[62] M. Soltani and M. J. Balas, "On-line prediction of the quantum density matrix," *Quantum Reports*, vol. 8, no. 1, p. 1, 2025.

[63] M. Balas, V. P. Gehlot, and T. D. Griffith, "A note on the estimation of von Neumann and relative entropy via quantum state observers," *IEEE Contr. Syst. Lett.*, vol. 7, pp. 3807–3811, 2023.

[64] S. Hu, H. Ma, Y. Lee, D. Dong, and I. R. Petersen, "Control-enhanced quantum metrology guided by Pontryagin's minimum principle," in *Proc. IEEE Conf. Decis. and Contr.*, 2025, pp. 6724–6729.

[65] Q. Ansel, E. Dionis, and D. Sugny, "Optimal control strategies for parameter estimation of quantum systems," *SciPost Phys.*, vol. 16, no. 1, p. 013, 2024.

[66] C. Lin, Y. Ma, and D. Sels, "Application of Pontryagin's maximum principle to quantum metrology in dissipative systems," *Phys. Rev. A*, vol. 105, no. 4, p. 042621, 2022.

[67] ——, "Optimal control for quantum metrology via Pontryagin's principle," *Phys. Rev. A*, vol. 103, no. 5, p. 052607, 2021.

[68] J. Liu and H. Yuan, "Quantum parameter estimation with optimal control," *Phys. Rev. A*, vol. 96, no. 1, p. 012117, 2017.

[69] S. Pang and A. N. Jordan, "Optimal adaptive control for quantum metrology with time-dependent Hamiltonians," *Nat. Commun.*, vol. 8, no. 1, p. 14695, 2017.

[70] J. Liu and H. Yuan, "Control-enhanced multiparameter quantum estimation," *Phys. Rev. A*, vol. 96, no. 4, p. 042114, 2017.

[71] M. M. Wilde, *Quantum Information Theory*, 2nd ed. Cambridge, UK: Cambridge University Press, 2017.

[72] Y. Polyanskiy and Y. Wu, *Information theory: From coding to learning*, 1st ed. Cambridge University Press, 2025.

[73] M. Clouâtré, B. Kartal, S. Marano, P. L. Falb, and M. Z. Win, "Finite-time quantum dissipation engineering," in *Proc. IEEE Int. Conf. Commun.*, Montreal, Canada, Jun. 2025, pp. 6808–6813.

[74] A. S. Holevo, *Probabilistic and statistical aspects of quantum theory*. Springer Science & Business Media, 2011, vol. 1.

[75] R. Durrett, *Probability: theory and examples*. Cambridge university press, 2019, vol. 49.

[76] A. S. Holevo, *Quantum systems, channels, and information: A mathematical introduction*. De Gruyter, 2012.

[77] W. Lohmiller and J.-J. E. Slotine, "On contraction analysis for non-linear systems," *Automatica*, vol. 34, no. 6, pp. 683–696, 1998.

[78] K. Kraus, "General state changes in quantum theory," *Annal Phys.*, vol. 64, no. 2, pp. 311–335, 1971.

[79] D. Pérez-García, M. M. Wolf, D. Petz, and M. B. Ruskai, "Contractivity of positive and trace-preserving maps under l_p norms," *J. Multivariate Anal.*, vol. 47, no. 8, 2006.

[80] G. Lindblad, "On the generators of quantum dynamical semigroups," *Comm. Math. Phys.*, vol. 48, no. 2, pp. 119–130, 1976.

[81] J. Watrous, *The theory of quantum information*. Cambridge University Press, 2018, pre-publication copy.

[82] M. Girard, D. Leung, J. Levick, C.-K. Li, V. Paulsen, Y. T. Poon, and J. Watrous, "On the mixed-unitary rank of quantum channels," *Comm. Math. Phys.*, vol. 394, no. 2, pp. 919–951, 2022.

[83] A. W. Marshall, I. Olkin, and B. C. Arnold, *Inequalities: theory of majorization and its applications*, 2nd ed. Springer, 2011.

[84] H. Yuan, "Characterization of majorization monotone quantum dynamics," *IEEE Trans. Autom. Control*, vol. 55, no. 4, pp. 955–959, 2010.

[85] R. A. Horn and C. R. Johnson, *Matrix Analysis*, 2nd ed. New York, NY: Cambridge University Press, 2013.

[86] P. Krantz, M. Kjaergaard, F. Yan, T. P. Orlando, S. Gustavsson, and W. D. Oliver, "A quantum engineer's guide to superconducting qubits," *Appl. Phys. Rev.*, vol. 6, no. 2, 2019.

[87] D. Liberzon, *Calculus of variations and optimal control theory: a concise introduction*. Princeton university press, 2011.

[88] S. J. Glaser, U. Boscain, T. Calarco, C. P. Koch, W. Köckenberger, R. Kosloff, I. Kuprov, B. Luy, S. Schirmer, T. Schulte-Herbrüggen, D. Sugny, and F. K. Wilhelm, "Training Schrödinger's cat: Quantum optimal control: Strategic report on current status, visions and goals for research in Europe," *European Phys. J. D*, vol. 69, pp. 1–24, 2015.

[89] H. P. Cartan, *Differential calculus on normed spaces: A course in analysis*. Independent publishing, 2017.

[90] A. W. Naylor and G. R. Sell, *Linear Operator Theory in Engineering and Science*. New York: Holt, Rinehart and Winston, Inc., 1971.

[91] W. Rudin, *Principles of Mathematical Analysis*, 3rd ed. New York: McGraw-Hill, 1976.

[92] D. W. Stroock, *Essentials of integration theory for analysis*. Springer, 2020.

[93] D. P. Bertsekas, *Nonlinear Programming*, 3rd ed. Belmont, MA: Athena Scientific, 2016.

[94] G. Casella and R. L. Berger, *Statistical Inference*, 2nd ed. Pacific Grove, CA, USA: Duxbury Press, 2002.

[95] P. J. Bickel and K. Doksum, *Mathematical Statistics: Basic Ideas and Selected Topics*, 1st ed. Oakland, California: Holden-Day, Inc., 1977.

[96] H. L. Van Trees, *Detection, estimation, and modulation theory, part I: Detection, estimation, and linear modulation theory*. John Wiley & Sons, 1968.

[97] M. G. A. Paris, "Quantum estimation for quantum technology," *International Journal of Quantum Information*, vol. 7, no. supp01, pp. 125–137, 2009.

[98] S. Gudder, "Combinations of quantum observables and instruments," *J. Phys. A: Math. Theor.*, vol. 54, no. 36, p. 364002, 2021.

[99] K. Reuer, J. Landgraf, T. Fösel, J. O'Sullivan, L. Beltrán, A. Akin, G. J. Norris, A. Remm, M. Kerschbaum, J.-C. Besse, F. Marquardt, A. Wallraff, and C. Eichler, "Realizing a deep reinforcement learning agent for real-time quantum feedback," *Nat. Commun.*, vol. 14, no. 1, p. 7138, 2023.

[100] C. A. Ryan, B. R. Johnson, D. Ristè, B. Donovan, and T. A. Ohki, "Hardware for dynamic quantum computing," *Rev. Sci. Instruments*, vol. 88, no. 10, 2017.

[101] A. Wald, "Note on the consistency of the maximum likelihood estimate," *Annals Math. Stat.*, vol. 20, no. 4, pp. 595–601, 1949.

[102] S. Kullback and R. A. Leibler, "On information and sufficiency," *Annals Math. Stat.*, vol. 22, no. 1, pp. 79–86, 1951.

[103] M. J. Wainwright, *High-dimensional statistics: A non-asymptotic viewpoint*. Cambridge university press, 2019, vol. 48.

[104] P. Chaudhuri and P. A. Mykland, "On efficient designing of nonlinear experiments," *Statistica Sinica*, pp. 421–440, 1995.

[105] S. E. Rasmussen, K. S. Christensen, S. P. Pedersen, L. B. Kristensen, T. Baekkegaard, N. J. S. Loft, and N. T. Zinner, "Superconducting circuit companion—an introduction with worked examples," *PRX Quantum*, vol. 2, no. 4, p. 040204, 2021.

[106] B. Bürgler, T. F. Sjolander, O. Brinza, A. Tallaire, J. Achard, and P. Maletinsky, "All-optical nuclear quantum sensing using nitrogen-vacancy centers in diamond," *npj Quantum Inf.*, vol. 9, no. 1, p. 56, 2023.

[107] J. F. Barry, J. M. Schloss, E. Bauch, M. J. Turner, C. A. Hart, L. M. Pham, and R. L. Walsworth, "Sensitivity optimization for NV-diamond magnetometry," *Rev. Mod. Phys.*, vol. 92, no. 1, p. 015004, 2020.

[108] P. Rembold, N. Oshnik, M. M. Müller, S. Montangero, T. Calarco, and E. Neu, "Introduction to quantum optimal control for quantum sensing with nitrogen-vacancy centers in diamond," *AVS Quantum Sci.*, vol. 2, no. 2, 2020.

[109] R. Schirhagl, K. Chang, M. Loretz, and C. L. Degen, "Nitrogen-vacancy centers in diamond: nanoscale sensors for physics and biology," *Ann. Rev. Phys. Chem.*, vol. 65, no. 1, pp. 83–105, 2014.

[110] S. Hong, M. S. Grinolds, L. M. Pham, D. Le Sage, L. Luan, R. L. Walsworth, and A. Yacoby, "Nanoscale magnetometry with NV centers in diamond," *MRS bulletin*, vol. 38, no. 2, pp. 155–161, 2013.

[111] J. M. Taylor, P. Cappellaro, L. Childress, L. Jiang, D. Budker, P. R. Hemmer, A. Yacoby, R. Walsworth, and M. D. Lukin, "High-sensitivity diamond magnetometer with nanoscale resolution," *Nat. Physics*, vol. 4, no. 10, pp. 810–816, 2008.

[112] Y. Polyanskiy and S. Verdú, "Arimoto channel coding converse and Rényi divergence," in *Allerton Conf. Commun., Control, Comput.* IEEE, 2010, pp. 1327–1333.

[113] Y. Polyanskiy, H. V. Poor, and S. Verdú, "Channel coding rate in the finite blocklength regime," *IEEE Trans. Inf. Theory*, vol. 56, no. 5, pp. 2307–2359, 2010.

[114] I. Csiszár, "Eine informationstheoretische ungleichung und ihre anwendung auf den beweis der ergodizität von markoffschen ketten," *A Magyar Tudományos Akadémia Matematikai Kutató Intézetének Közleményei*, vol. 8, no. 1-2, pp. 85–108, 1963.

[115] ——, "Information measures: A critical survey," in *Trans. Seventh Prague Conf. Inf. Theory, Stat. Decis. Fun., Rand. Process.*, 1974, pp. 73–86.

[116] N. Sharma and N. A. Warsi, "Fundamental bound on the reliability of quantum information transmission," *Physical review letters*, vol. 110, no. 8, p. 080501, 2013.

[117] ——, "On the strong converses for the quantum channel capacity theorems," *arXiv preprint arXiv:1205.1712*, 2012.

[118] G. Y. Tham, R. Nair, and M. Gu, "Quantum limits of covert target detection," *Phys. Rev. Lett.*, vol. 133, no. 11, p. 110801, 2024.

[119] S. Guerrini, M. Z. Win, M. Chiani, and A. Conti, "Quantum discrimination of noisy photon-added coherent states," *IEEE J. Sel. Areas Inf. Theory*, vol. 1, no. 2, pp. 469–479, Aug. 2020, special issue on *Quantum Information Science*.

[120] C. Wittmann, M. Takeoka, K. N. Cassemiro, M. Sasaki, G. Leuchs, and U. L. Andersen, "Demonstration of near-optimal discrimination of optical coherent states," *Phys. Rev. Lett.*, vol. 101, no. 21, p. 210501, 2008.

[121] S. Axler, *Linear algebra done right*, 4th ed. Springer Nature, 2024.

[122] M. Ringbauer, M. Meth, L. Postler, R. Stricker, R. Blatt, P. Schindler, and T. Monz, "A universal qudit quantum processor with trapped ions," *Nat. Physics*, vol. 18, no. 9, pp. 1053–1057, 2022.

[123] L. Mirsky, "The spread of a matrix," *Mathematika*, vol. 3, no. 2, pp. 127–130, 1956.

[124] H. K. Khalil, *Nonlinear systems*. Pearson India, 2015, vol. 3.

[125] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, UK: Cambridge University Press, 2004.

[126] J.-S. Li and N. Khaneja, "Control of inhomogeneous quantum ensembles," *Phys. Rev. A*, vol. 73, no. 3, p. 030302, 2006.

[127] ——, "Ensemble control of Bloch equations," *IEEE Trans. Autom. Control*, vol. 54, no. 3, pp. 528–536, 2009.

[128] A. Conti, A. Giani, and M. Z. Win, "State engineering for quantum networks," *IEEE Network*, pp. 1–7, 2026, to appear.

[129] F. Forzano, A. Giani, S. Marano, M. Z. Win, and A. Conti, "Quadrature measurement characterization for single-mode photon-varied Gaussian states," in *Proc. IEEE Int. Conf. on Computer Commun.*, London, UK, May 2025, pp. 1–6.

[130] N. Dunford and J. T. Schwartz, *Linear operators, Part 1: General theory*. John Wiley & Sons, 1958.

[131] N. Weaver, *Lipschitz algebras*, 2nd ed. World Scientific, 2018.

[132] L. Mirsky, "On the trace of matrix products," *Mathematische Nachrichten*, vol. 20, no. 3-6, pp. 171–174, 1959.

[133] C. D. Aliprantis and K. C. Border, *Infinite dimensional analysis: a hitchhiker's guide*. Springer, 2006.

[134] A. F. Filippov, *Differential equations with discontinuous righthand sides*. Springer Science & Business Media, 1988.