Real-time optimal quantum control for atomic magnetometers with decoherence

Júlia Amorós Binefa

Abstract

Quantum entanglement can enhance the sensitivity of atomic sensors to static or slowly varying fields. But many important applications in fields such as medicine or navigation require tracking fast or transient signals. This presents unique challenges, one of them being that the potential benefits of entanglement in such cases are still not fully understood. To investigate this, we apply concepts from continuous quantum measurements and estimation theory to optical atomic magnetometers, aiming to accurately model these devices, interpret their measurement data, control their dynamics, and achieve optimal sensitivity.

Quantifying this optimal performance requires determining a fundamental quantum limit on sensitivity. The above bound imposed by noise is derived and shown to scale at best linearly with the sensing time and number of atoms N, ruling out any super-classical scaling. Moreover, this quantum limit is independent of the initial state, measurement, estimator, and measurement-based feedback, and depends only on the decoherence model and strength of the field fluctuations. Thus, finding an estimator that attains this limit proves a given sensing strategy optimal.

To approach this bound, we develop a quantum dynamical model scalable w.r.t. N, based on a co-moving Gaussian approximation of the stochastic master equation, which includes both measurement backaction and decoherence. This enables the construction of a real-time estimation and control architecture that integrates an extended Kalman filter (EKF) with a linear quadratic regulator (LQR).

By simulating the magnetometer with our model and our proposed EKF+LQR strategy, we show that quantum-limited tracking of constant and fluctuating fields is within reach of current atomic magnetometers. Strikingly, our sensing strategy can also track biological relevant signals, such as heartbeat-like waveforms. It can furthermore be used to drive the atomic ensemble into an entangled state, even when the measurement record is used for feedback but afterwards discarded.

Streszczenie

Splątanie kwantowe może zwiększyć czułość sensorów atomowych na pola statyczne lub wolnozmienne. Jednak wiele istotnych zastosowań w dziedzinach takich jak medycyna czy nawigacja wymaga śledzenia sygnałów szybko lub nagle zmieniających się. Wiąże się to ze szczególnymi wyzwaniami, biorąc pod uwagę brak wystarczającego zrozumienia potencjalnych korzyści ze splątania w takich przypadkach. Aby to zbadać, w pracy doktorskiej zastosowano narzędzia z zakresu kwantowych pomiarów ciągłych i teorii estymacji do optycznych magnetometrów atomowych, w celu dokładnego modelowania tych urządzeń, interpretacji danych pomiarowych, kontroli ich dynamiki oraz osiągnięcia optymalnej czułości.

Określenie tej optymalnej wydajności wymaga wyznaczenia fundamentalnej kwantowej granicy czułości. Powyższa granica, zdeterminowana przez szum, którą wyznaczamy w pracy, w najlepszym wypadku skaluje się liniowo z czasem pomiaru i liczbą atomów N, co wyklucza jakiekolwiek skalowanie lepsze niż klasyczne. Co więcej, ta kwantowa granica jest niezależna od stanu początkowego, pomiaru, estymatora oraz sprzężenia zwrotnego opartego na pomiarze i zależy wyłącznie od modelu dekoherencji oraz siły fluktuacji pola. Zatem znalezienie estymatora osiągającego tę granicę jest dowodem na optymalność danej strategii pomiarowej.

Aby zbliżyć się do tej granicy, opracowujemy kwantowy model dynamiczny dobrze skalujący się z N, oparty na współporuszającym się przybliżeniu Gaussowskim kwantowego równania "master", który uwzględnia zarówno działanie pomiaru, jak i dekoherencję. Umożliwia to stworzenie architektury estymacji i kontroli w czasie rzeczywistym, integrującej rozszerzony filtr Kalmana (EKF) z regulatorem liniowo-kwadratowym (LQR).

Poprzez symulację magnetometru z wykorzystaniem naszego modelu oraz proponowanej strategii EKF+LQR pokazujemy, że kwantowo-ograniczone śledzenie zarówno pól stałych, jak i fluktuujących, jest w zasięgu obecnych magnetometrów atomowych. Co istotne, nasza strategia pomiarowa umożliwia także śledzenie sygnałów istotnych biologicznie, takich jak sygnały podobne do tych generowanych przez serce, ale może być po prostu użyta do wprowadzenia układu atomowego w stan splątany, także kiedy dane pomiarowe są wywołane przez sprzężenie zwrotne ale później zapominane.

To my parents.

Acknowledgments

First and foremost, I would like to thank my supervisor, Janek Kołodyński, for his guidance and support throughout my PhD. Somehow, the vision he shared in our first meeting turned out to be pretty much what this thesis became, a rare thing! His positivity, even when results were uncertain, helped me stay motivated, and he never hesitated to dive into the nitty-gritty when we were stuck. I also deeply thank him for his efforts in teaching me how to write, proofreading my thesis, encouraging me to present at conferences, and ensuring consistent funding throughout my PhD.

I would also like to thank my other PhD supervisor, Prof. Konrad Banaszek, for his support in facilitating my doctoral studies as well as his efforts as the head of the QOT center in attracting funding.

I am also especially grateful to Francesco Albarelli, Marco G. Genoni and Matteo A. C. Rossi for their openness and generosity. Although we never formally collaborated, they helped me understand their code, answered my questions and shared their excellent notes on continuous measurements, which clarified lots of important concepts in my mind.

Another special thanks goes to our collaborator, Morgan W. Mitchell, for his invaluable expertise and guidance in applying the model and techniques developed in this thesis to experimentally realistic regimes. His insights were crucial in motivating our work and in bringing the theory closer to experiments.

Finally, I would like to thank my friends and colleagues: Lewis Clark, Mateusz Molenda, Rahul Deshpande, and Aditya Jain for reading parts of this thesis and providing valuable feedback. I am also thankful to Adam Parafiańczuk and Marcin Koźbiał for their help in translating the abstract into Polish.

List of Publications

J. Amorós-Binefa and J. Kołodyński, *"Noisy atomic magnetometry in real time,"* New J. Phys. 23, 123030 (2021).

J. Amorós-Binefa and J. Kołodyński, *"Noisy atomic magnetometry with Kalman filtering and measurement-based feedback,"* arXiv:2403.14764 (2024).

J. Amorós-Binefa, M. W. Mitchell, and J. Kołodyński, *"Tracking time-varying signals with quantum-enhanced atomic magnetometers,"* arXiv:2503.14793 (2025).

Contents

5   MODELS FOR NOISY REAL-TIME ATOMIC MAGNETOMETRY   **156**

6   CONCLUSIONS   **196**

APPENDIX A   APPENDIX FOR CHAP. 1   **198**

APPENDIX B   APPENDIX FOR CHAP. 2   **214**

APPENDIX C   APPENDIX FOR CHAP. 3   **218**

List of figures

List of Tables

Abbreviations

aMSE	averaged Mean Squared Error
BCRB	Bayesian Cramér-Rao Bound
CoG	Co-moving Gaussian approximation
CPTP	Completely Positive and Trace-Preserving map
CS	Classically-Simulated limit
CSS	Coherent Spin State
EKF	Extended Kalman Filter
EM	Euler-Maruyama method
FI	classical Fisher Information
GKSL	Gorini-Kosakowski-Sudarshan-Lindblad form
KF	Kalman Filter
LG	Linear and Gaussian
LQG	Linear-Quadratic-Gaussian control
LQR	Linear-Quadratic Regulator
MAP	Maximum A-Posteriori estimator
MCG	MagnetoCardioGram
ML	Maximum Likelihood estimator
MMSE	Minimum Mean Squared Error estimator
MSE	Mean Squared Error
OU	Ornstein-Uhlenbeck process
PDF	Probability Density Function
PMF	Probability Mass Function
POVM	Positive Operator-Valued Measure
RWA	Rotating-Wave Approximation
SDE	Stochastic Differential Equation
SME	Stochastic Master Equation
SQL	Standard Quantum Limit
SS	Steady-State
VdP	Van der Pol oscillator
w.r.t.	with respect to

0

Introduction

METROLOGY — THE SCIENCE OF MEASUREMENT — relies on classical estimation theory to infer unknown parameters from measurement data and quantify the uncertainty of these estimates. Perhaps most crucially, it also tells us how to identify optimal estimators that minimize this uncertainty, such as those that saturate the Cramêr–Rao bound [1, 2].

Even though classical estimation theory does not require the measurement outcomes to be probabilistic, we still use probabilities to express our uncertainty about the true value of a parameter. This uncertainty arises not from the theory itself, but from practical limitations such as environmental noise, sensor imperfections, or incomplete information about the system. In contrast, the randomness observed in measurements of quantum states is not due to external factors, but rather an intrinsic feature of quantum systems arising from the postulates of quantum mechanics. Even in ideal, noiseless conditions, the outcomes of quantum measurements are fundamentally stochastic, with their associated likelihoods dictated by Born's rule [3]. Quantum systems, moreover, have non-classical features such as coherence and entanglement, which can be harnessed to reduce measurement uncertainty beyond classical limits [4, 5, 6, 7]. This reduc-

tion in uncertainty improves estimation accuracy, which typically scales with the number of resources (e.g. the number of particles used, N). In classical strategies using uncorrelated particles, the estimation error scales as $1/\sqrt{N}$, known as the Standard Quantum Limit (SQL). However, quantum-enhanced strategies can achieve more favorable scalings, such as $1/N$, typically referred to as the Heisenberg Limit (HL) [8, 9, 10, 11, 12]. Even though numerous experiments have demonstrated surpassing the SQL [13, 14, 15, 16, 17], decoherence and other noise sources can significantly degrade entanglement and prevent achieving super-classical scalings [18, 19].

Much early work in quantum metrology focused on repeated measurements averaged over many trials [20, 21]. However, many real-world sensing tasks require tracking time-varying signals [22, 23, 24, 25, 26, 27, 28], where repeated measurements are not possible. In such cases, estimation must be performed in one shot. One way to achieve this is by employing Bayesian filtering, which continuously updates probabilistic estimates as new, noisy data becomes available [29, 30]. This real-time process involves two steps: (1) a predictive step, relying on a dynamical model of the system, and (2) a measurement update step, which refines the estimate based on incoming measurement outcomes [30]. A classical example is the Kalman filter (KF) [31, 32], optimal for linear systems with Gaussian noise [30]. For nonlinear quantum systems such as atomic sensors, methods like the extended Kalman filter (EKF) are better suited [29, 33].

Since Bayesian filtering relies on a model, it is essential to have a reliable description of both the system and the measurement process to implement any algorithm effectively [29, 30]. However, how can we model a quantum system that is being continuously monitored, when the very act of measurement inevitably disturbs its state [34, 35]? Fortunately, this evolution of the state conditional on the measurement outcomes is rigorously captured by the theory of continuous quantum measurements [36, 37, 38, 39]. In particular, the dynamical model is given by a stochastic master equation (SME), which correctly incorporates measurement backaction. This backaction can manifest itself as either discrete jumps when measuring photo-counts, or as a continuous, diffusive process when using homodyne detection [40, 41, 42]. Combining Bayesian filtering with this framework allows for real-time tracking while correctly accounting for measurement backaction. This has been experimentally implemented in Gaussian systems such as optomechanical resonators [43, 44, 45, 46, 47, 48] or levitated nanoparticles [49, 50].

Atomic sensors, particularly optical atomic magnetometers, can also benefit from these ideas. These devices rival state-of-the-art superconducting quantum interference devices (SQUIDs) in sensitivity, yet operate without the need for cryogenic cooling [51, 52] and can potentially be miniaturized to chip-scale sizes [53]. Although linear and Gaussian (LG) models are common

for atomic systems (as well as for optomechanical ones), atomic sensors are intrinsically nonlinear systems whose dynamics, including measurement backaction, can be rigorously described using SMEs [54, 55, 56]. That said, most atomic magnetometry experiments involve large atomic ensembles, where the sheer size of the Hilbert space makes exact quantum simulations of the SME intractable. Different approaches exist on how to deal with that: one can straight up ignore [57] or evade [58, 59] the measurement backaction, or instead restrict the analysis to the Gaussian regime [54, 60, 61, 56] or to low atomic numbers [62]. Despite these modeling challenges, a recent experiment with unpolarized atomic ensembles demonstrated that continuous measurement backaction can generate inter-atomic entanglement even without explicitly modeling the backaction [63]. This was possible because the spin dynamics remained within the LG regime. However, high-sensitivity spin-precession sensors typically use polarized atomic ensembles, whose spins evolve according to nonlinear dynamics [64]. This motivates the need for a realistic nonlinear model that captures both the measurement backaction and the decoherence effects in a way that is scalable to systems with large atom numbers.

To address this, we propose an approximate model referred to as the "co-moving Gaussian" picture. This model reduces the complexity of the full quantum dynamics while still accounting for dephasing and the measurement backaction that creates spin-squeezing [65, 66]. This model not only enables the design of an estimation and feedback scheme by combining an EKF with a linear quadratic regulator (LQR) [29, 55], but also allows us to simulate realistic atomic magnetometry experiments. Its accuracy is validated by comparing it against the exact simulation of the SME for moderate-sized systems [65].

Once the model is validated, the next step is to asses the performance of the sensor when tracking fluctuating fields by establishing fundamental limits on the estimation error. This is achieved by deriving a lower bound on the Bayesian Cramér-Rao bound (BCRB) [2, 67], which we refer to as classically-simulated (CS) limit or quantum limit [68, 19, 69, 65, 66]. Notably, this limit depends only on the dephasing rate and the strength of field fluctuations, and scales at best linearly with the atom number and sensing time, thereby precluding any possibility of surpassing the SQL [69, 65, 66]. Remarkably, our magnetometry setup attains this quantum limit, confirming that the entire sensing protocol is optimal under the given noise conditions, regardless of the initial state, measurement, estimator, or measurement-based feedback [66]. Beyond demonstrating optimal real-time sensing of fluctuating signals, we also apply our method to track signals commonly found in medicine and biology, such as magneto-cardiograms (MCG) [24, 66], which requires filtering to extract heartbeat-like signals from noisy backgrounds.

Finally, this framework can also be used for quantum state preparation [65]. Specifically, we show that the LQR not only enhances estimation accuracy but also steers the atomic ensemble into an entangled state without the need to store past measurement data. This makes the protocol a practical method for preparing entangled states in real time.

THIS THESIS IS ORGANIZED INTO SIX CHAPTERS. The first three chapters cover background material that is important for understanding the rest of the work. In particular, the first part of Chap. 1 reviews important topics like probability theory, stochastic processes and stochastic calculus. The second part of Chap. 1 introduces basic concepts of quantum mechanics, including angular momentum operators, the Wiener function and spin-squeezing. Then, Chap. 2 presents an in-depth discussion of Bayesian filtering, with complete treatments of the KF, EKF and LQR. In Chap. 3, we derive the SME for both photodetection and homodyne measurement. While these chapters aim to keep the thesis self-contained, my background in quantum physics inevitably influences what I consider to be introductory material, so readers with different expertise may find some sections more familiar than others. If you are already comfortable with the topics described above, feel free to skip ahead to Chap. 4 and Chap. 5, where the main results and contributions are discussed. In Chap. 4, we derive the quantum limit or classically-simulated limit on the estimation error via lower-bounding the BCRB. Finally, Chap. 5 applies these theoretical and numerical tools to atomic magnetometry. We introduce the "co-moving Gaussian" model to capture both the measurement backaction and decoherence in large atomic ensembles, and then put forward a complete estimation and control protocol. This protocol consists of an EKF combined with LQR, which achieves optimal performance and even enables real-time entanglement preparation. Finally, Chap. 6 summarizes the results and discusses potential directions for future research.

1

Preliminaries

T<small>HE MAIN FOCUS OF THIS THESIS</small> is the tracking of quantities that vary randomly over time. In many complex systems, such as those in physics, finance, or biology, the parameters of interested are not static but fluctuate continuously due to inherent randomness and external disturbances. These quantities are referred to as *stochastic processes*, and to capture their dynamics, it is essential to combine differential equations with probability theory.

I<small>N THE FIRST PART OF THIS OPENING CHAPTER</small>, we begin by reviewing a bit of probability theory: random variables, probability density functions, expectation and covariances and the Gaussian distribution. Next, we explain in Sec. 1.2 what a stochastic process is through some simple examples, and introduce in Sec. 1.3 the basics of stochastic calculus with topics like Itô's Lemma, stochastic differential equations and their corresponding numerical methods.

In the second part of this chapter, i.e. Sec. 1.4, we briefly introduce a few concepts related to the dynamics and properties of quantum systems, mostly for later reference. These include the position and momentum operators, the dynamics of open quantum systems, angular momen-

tum operators and coherent spin states. We also cover the Wigner quasiprobability distribution in Sec. 1.4.5 and how to map the Wigner function onto a sphere in Sec. 1.4.6, followed by a brief explanation of spin-squeezing in Sec. 1.4.7.

1.1 Fundamental concepts of probability theory

1.1.1 Random variables

In probability theory, a *random variable* assigns numerical values to the outcomes of a random process, each with an associated likelihood. Formally, it is a function mapping a sample space Ω to a subset of the real numbers $\mathbb{R}$. For example, when flipping a coin, the sample space Ω representing all possible outcomes of this experiment is $\Omega = \{\text{Heads}, \text{Tails}\}$. Now, let X^* be a random variable that assigns numerical values to these outcomes, e.g.:

$$\text{X}(\text{Heads}) = 1 \qquad \text{and} \qquad \text{X}(\text{Tails}) = 0, \tag{1.1}$$

where $x = 0$ and $x = 1$ are *realizations* of the random variable X. Random variables like the one above, or others such as the outcome of a dice roll, take on specific, countable values. Therefore, these are referred to as *discrete* random variables. Additionally, each possible outcome has an associated probability, the probability mass function (PMF), and these probabilities must sum to 1.

Definition 1.1 (Probability mass function). *For a discrete random variable X, let the PMF be a function $p : \mathbb{R} \to [0, 1]$ that gives the probability of the random variable X taking the specific value x, i.e.*

$$p(x) = \Pr[\text{X} = x], \tag{1.2}$$

where $p(x) \geq 0 \ \forall x$ and the sum of all probabilities must be equal to 1:

$$\sum_{x \in \mathcal{X}} p(x) = 1, \tag{1.3}$$

with $\mathcal{X}$ corresponding to the set of all the possible realizations x of the random variable X.

---

*In mathematics, random variables are typically denoted by a capital letter (e.g. X), and their possible values, a.k.a. realizations, by the corresponding lowercase letter (x). We will follow this convention here but use lowercase for both in later chapters.

For example, if X represents the outcome of rolling a fair six-sided dice, the PMF is:

$$p(x) = \frac{1}{6} \quad \text{for} \quad x \in \mathcal{X} = \Omega = \{1, 2, 3, 4, 5, 6\}. \tag{1.4}$$

In contrast, a *continuous random variable* takes values from a continuous range, such as time, position or temperature. Since a continuous random variable can have infinitely many values, the probability of it taking any specific value is zero. Therefore, probabilities are determined over intervals using the probability density function (PDF). The total probability for all values (the area under the PDF curve) must also sum to 1.

Definition 1.2 (Probability density function). *The PDF of a continuous-valued random variable* X *is denoted as* $p(x)$*. Its integration over an interval* $[a, b]$ *yields the probability of* $X \in [a, b]$*, i.e.*

$$\Pr[a \leq X \leq b] = \int_a^b p(x)\mathrm{d}x, \tag{1.5}$$

and thus, relating the PDF to the PMF. Additionally, the PDF must satisfy the following conditions

$$p(x) \geq 0 \; \forall x, \quad \text{and} \quad \int p(x)\,\mathrm{d}x = 1.$$

1.1.2 MULTIVARIATE RANDOM VARIABLES

Additionally, the definitions of PMF and PDF can be extended also to *multivariate random variables*, also referred to as *random vectors*. Formally, a multivariate random variable is a column vector $\mathbf{X} = (X_1, \ldots, X_n)^\mathrm{T}$, whose components are random variables.

Definition 1.3 (Multivariate PMF). *For a multivariate discrete random variable* X*, let the PMF be a function* $p : \mathbb{R}^n \to [0, 1]$ *that gives the probability of the random vector* $\mathbf{X}$ *taking the specific value* $\boldsymbol{x}$*, i.e.*

$$p(\boldsymbol{x}) = \Pr[\mathbf{X} = \boldsymbol{x}] \tag{1.6}$$

where $p(\boldsymbol{x}) \geq 0 \; \forall \boldsymbol{x}$ and the sum of all probabilities must be equal to 1:

$$\sum_{\boldsymbol{x} \in \mathcal{X}} p(\boldsymbol{x}) = 1 \tag{1.7}$$

with $\mathcal{X}$ corresponding to the set of all the elements $\boldsymbol{x}$ can take.

Definition 1.4 (Multivariate PDF). *The multivariate PDF of a continuous-valued random vector $\mathbf{X}$ is denoted as $p(\boldsymbol{x})$. Its integration over a domain $\mathcal{X}$ in the n-dimensional space of the values of $\mathbf{X}$ yields the probability of $\mathbf{X} \in \mathcal{X}$, i.e:*

$$\Pr[\mathbf{X} \in \mathcal{X}] = \int_{\mathcal{X}} p(\boldsymbol{x}) \, \mathcal{D}\boldsymbol{x}, \tag{1.8}$$

where $\int \mathcal{D}\boldsymbol{x}$ is shorthand for $\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} dx_1 \ldots dx_n$. Just like for the single-variable PDF, its multivariate counterpart must satisfy the following conditions

$$p(\boldsymbol{x}) \geq 0 \; \forall \boldsymbol{x}, \quad \text{and} \quad \int p(\boldsymbol{x}) \, \mathcal{D}\boldsymbol{x} = 1.$$

1.1.3 Joint probability functions

When working with more than one random variable, their combined behavior is described by *joint probability functions*, which capture the probability of two or more random variables taking specific values simultaneously.

Definition 1.5 (Joint PMF). *For discrete random variables, the joint behavior of X and Y is described by the joint PMF, which yields the probability of X taking the value x and Y taking the value y, simultaneously:*

$$p(x, y) = \Pr[\mathrm{X} = x, \mathrm{Y} = y]. \tag{1.9}$$

Definition 1.6 (Joint PDF). *For continuous random variables, the joint behavior of random variables X and Y is described by the joint PDF $p(x, y)$, which provides the relative likelihood of X taking a value near x and Y taking a value near y:*

$$\Pr[a \leq \mathrm{X} \leq b, c \leq \mathrm{Y} \leq d] = \int_a^b \int_c^d p(x, y) \, dy \, dx. \tag{1.10}$$

Both definitions can be easily extended to multivariate random variables.

1.1.4 INDEPENDENT AND DEPENDENT RANDOM VARIABLES

Two random variables are *independent* if neither of their probability functions depends on the value of the other variable. Their definition is formalized through the use of joint probability functions. In particular:

Definition 1.7 (Independent discrete random variables). *Let X and Y be two discrete random variables, with ranges $\mathcal{X}$ and $\mathcal{Y}$, respectively. We say that X and Y are independent if for all $x \in \mathcal{X}$ and $y \in \mathcal{Y}$, the joint PMF satisfies*

$$\Pr[X = x, Y = y] = \Pr[X = x] \cdot \Pr[Y = y] \quad \forall x, y \tag{1.11}$$

In other words, for every pair of possible outcomes x and y, the probability that X takes the value x and Y takes the value y is the product of the individual probabilities of those events. That can be extended to multivariate discrete random variables, as

$$\Pr[X_1 = x_1, \ldots, X_n = x_n] = \Pr[X_1 = x_1] \cdot \ldots \cdot \Pr[X_n = x_n], \quad \forall x_1, \ldots, x_n. \tag{1.12}$$

Additionally, a similar definition applies to independent continuous random variables, where independence is characterized by the joint PDF factoring into the product of their respective probability functions. Specifically,

Definition 1.8 (Independent continuous random variables). *Let X and Y be two continuous random variables. We say that X and Y are independent if for all $x \in [a, b]$ and $y \in [c, d]$, the joint PDF satisfies*

$$p(x, y) = p(x) \cdot p(y), \tag{1.13}$$

such that,

$$\Pr[X \in [a, b], Y \in [c, d]] = \int_a^b \int_c^d p(x, y) \, \mathrm{d}y \, \mathrm{d}x = \int_a^b \int_c^d p(x) p(y) \, \mathrm{d}y \, \mathrm{d}x$$

$$= \left(\int_a^b p(x) \, \mathrm{d}x \right) \left(\int_c^d p(y) \, \mathrm{d}y \right) = \Pr[X \in [a, b]] \cdot \Pr[Y \in [c, d]]. \tag{1.14}$$

In other words, the probability that X *falls in the interval* $[a, b]$ *as well as* Y *falls in* $[c, d]$ *is the product of the individual probabilities of these events. Just like in the discrete case, this definition can also be extended to multivariate continuous random variables, as*

$$p(\boldsymbol{x}) = p(x_1, \ldots, x_n) = p(x_1) \cdot \ldots \cdot p(x_n), \quad \forall x_1, \ldots, x_n. \tag{1.15}$$

On the other hand, two random variables are *dependent* when they have a joint probability density that cannot be factored into the product of their respective probability functions. Unlike independent random variables, where knowing the outcome of one variable tells us nothing about the other, the value of one dependent variable is influenced by the value of the other.

For continuous random variables, this dependency is best described through the use of the so-called *conditional PDFs* and *marginal PDFs*. Marginal PDFs are isolated densities of a single random variable where we have removed the influence of all the other variables. All these functions are central to the understanding of Bayesian filtering and estimation.

For that reason, let us start by defining the probability density for X if we were to know nothing about Y, i.e. the marginal of X, through the law of total probability. Namely,

Property 1.1 (Marginal PDF / The law of total probability / Sum rule). *Let* X *and* Y *be two random variables, such that* $Y \in \mathcal{Y}$, *then the law of total probability states that the marginal PDF reads as:*

$$p(x) = \int_{y \in \mathcal{Y}} p(x, y) \mathrm{d}y. \tag{1.16}$$

Thus, marginalizing over Y is simply integrating over all possible values of Y such that we are left with a PDF that depends only on X.

Now we have all the tools needed to define the conditional PDF, which in turn yields a very important property in probability theory: the product rule. A conditional distribution is the probability distribution of one random variable given that another has taken a specific value. It is derived from their joint distribution, since $p(x, y)$ with $y = a$ gives the relative probability for x given that $Y = a$. However, $p(x, y = a)$ is not normalized and therefore, we have to divide it by the integral over all values of x, i.e., the marginal of Y.

Property 1.2 (Conditional PDF / Product rule). *Let* X *and* Y *be two random variables with the joint PDF,* $p(x, y)$. *The conditional density function,* $p(x|y)$, *which reads as the probability density of* x *given* y, *is defined as*

$$p(x|y) = \frac{p(x, y)}{\int_{-\infty}^{\infty} p(x, y) \, \mathrm{d}x}, \tag{1.17}$$

which, since the denominator of Eq. (1.17) is the marginal of Y, it can be rewritten as

$$p(x|y) = \frac{p(x,y)}{p(y)}. \tag{1.18}$$

The definition of the conditional density function yields the product rule:

$$p(x,y) = p(x|y)p(y). \tag{1.19}$$

Note that $p(x|y)$ should be interpreted as a function of only x since y is a known value, i.e. we have no uncertainty in y.

1.1.5 EXPECTATION AND COVARIANCE

One of the most important distributions for both linear and nonlinear filtering is the Gaussian distribution, which can be described using only the mean vector (or expected value) and the co-variance matrix of the Gaussian random variable. And if our results are not Gaussian, we anyway tend to approximate them using the mean and covariance of this non-Gaussian distribution, even though these do not fully capture all the facets of our actual distribution. Therefore, let us first define what is an expected value and a covariance matrix.

Definition 1.9 (Expected value / mean). *The expected value (i.e. mean) of a continuous random variable X is defined as*

$$\mathbb{E}[X] := \int_{-\infty}^{\infty} x\, p(x)\, \mathrm{d}x. \tag{1.20}$$

The expectation of X is the integral of x weighted by the probability density of x. Thus, values of X with high probability density will influence our expected values more than values of X with low probability density. The expected value of X is also sometimes referred to as the first moment of X. We can further define higher order moments of X, such as the variance of X as:

Definition 1.10 (Variance). *The variance of a random variable X is the expected value of the squared deviation of X from the mean $\mathbb{E}[X]$:*

$$V[X] := \mathbb{E}\big[(X - \mathbb{E}[X])^2\big] = \mathbb{E}\big[X^2\big] - \mathbb{E}[X]^2. \tag{1.21}$$

Additionally, the definition of the mean and other moments can also be extended to random vectors. Specifically, the multivariate mean will read as:

Definition 1.11 (Multivariate mean). *A random vector* $\mathbf{X} = [X_1, X_2, \ldots, X_n]^\mathrm{T}$ *has an expected value (mean) given by*

$$\mathbb{E}[\mathbf{X}] := \int \boldsymbol{x}\, p(\boldsymbol{x})\, \mathcal{D}\boldsymbol{x}, \tag{1.22}$$

where $p(\boldsymbol{x})$ is the multivariate PDF of Definition 1.4.

We can also extend the definition of the variance to multivariate random variables by defining the covariance matrix of $\mathbf{X}$. For a random vector $\mathbf{X} = (X_1, X_2, \ldots, X_n)^\mathrm{T}$, the covariance matrix encodes how each pair of variables X_i and X_j vary together, revealing patterns of correlation and dependencies across the components of $\mathbf{X}$. Formally, the covariance matrix is defined as:

Definition 1.12 (Covariance matrix). *Let $\mathbf{X}$ be a random vector. Then, the covariance matrix is*

$$\mathrm{cov}[\mathbf{X}] := \mathbb{E}\left[(\mathbf{X} - \mathbb{E}[\mathbf{X}])(\mathbf{X} - \mathbb{E}[\mathbf{X}])^\mathrm{T} \right]. \tag{1.23}$$

We can view the factor $(\mathbf{X} - \mathbb{E}[\mathbf{X}])$ as the distance between the random vector $\mathbf{X}$ and its mean. That is, how much does it spread around its mean value. Additionally, note that the diagonal elements of $\mathrm{cov}[\mathbf{X}]$, i.e. $\mathrm{cov}[\mathbf{X}]_{ii}$ are the variances of each individual variable X_i, while the off-diagonal elements indicate the correlation between different variables.

1.1.6 Law of large numbers

Sometimes we are interested in finding the expected value of a random value, but an analytical solution is not available. Perhaps we are not able to solve the involved integral explicitly, or more commonly, because the exact form of the underlying distribution is not known. However, if instead we have access to a large number of samples from the random variable, we can use the law of large numbers and estimate the expected value numerically.

In particular, the law of large numbers states that, as the number of independent and identically distributed samples of a random variable increases, the sample average converges to the true expected value of that variable. In other words:

Theorem 1.1 (The law of large numbers). *Let X_1, $X_2, \ldots, X_n$ be a sequence of independent and identically distributed random variables. In other words, they are all distributed according to the same PDF, $p(x)$, and thus, have a finite expected value $\mu = \mathbb{E}[X_i]\ \forall i = 1, \ldots, n$. Then,*

$$\lim_{n \to \infty} \bar{X}_n = \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} X_i = \mu. \tag{1.24}$$

Or, in other words, the sample average $\bar{X}_n$ converges almost surely[†] to the expected value μ.

1.1.7 THE MULTIVARIATE GAUSSIAN DISTRIBUTION

Let $\mathbf{X} \sim \mathcal{N}(\boldsymbol{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma})$ denote a Gaussian random vector drawn from a multivariate PDF with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$. The PDF of $\boldsymbol{x}$ is known as a Gaussian distribution and defined as:

Definition 1.13 (Multivariate Gaussian distribution). *A Gaussian random vector $\mathbf{X} \in \mathbb{R}^n$ has a probability density*

$$\mathcal{N}(\boldsymbol{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{n/2}|\boldsymbol{\Sigma}|^{1/2}} \exp\left(-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{x} - \boldsymbol{\mu}) \right) \qquad (1.25)$$

where $|\cdot|$ denotes the determinant, $\boldsymbol{\mu} = \mathbb{E}[\mathbf{X}]$ is the mean of $\mathbf{X}$ and $\boldsymbol{\Sigma}$ represents the covariance of $\mathbf{X}$, i.e. $\boldsymbol{\Sigma} := \mathrm{cov}[\mathbf{X}] = \mathbb{E}\left[(\mathbf{X} - \mathbb{E}[\mathbf{X}])(\mathbf{X} - \mathbb{E}[\mathbf{X}])^{\mathrm{T}} \right]$.

Just as in the case of a Gaussian random variable, the PDF of a Gaussian random vector comes completely determined by its mean and covariance matrix. Furthermore, one of the most useful properties of Gaussian random variables is their linear combination property.

Property 1.3 (Linearity of Gaussian random vectors). *Let $\mathbf{X} \sim \mathcal{N}(\boldsymbol{x}|\boldsymbol{\mu}_x, \boldsymbol{\Sigma}_x)$ be a Gaussian random vector, as well as $\mathbf{Y} \sim \mathcal{N}(\boldsymbol{y}|\boldsymbol{\mu}_y, \boldsymbol{\Sigma}_y)$. Then, a linear combination of $\mathbf{X}$ and $\mathbf{Y}$ yields another Gaussian random vector $\mathbf{Z}$:*

$$\mathbf{Z} = \boldsymbol{A}\mathbf{X} + \boldsymbol{B}\mathbf{Y}, \qquad (1.26)$$

where $\mathbf{Z}$ has a mean and covariance:

$$\boldsymbol{\mu}_z = \mathbb{E}[\boldsymbol{A}\mathbf{X} + \boldsymbol{B}\mathbf{Y}] = \boldsymbol{A}\boldsymbol{\mu}_x + \boldsymbol{B}\boldsymbol{\mu}_y, \qquad (1.27)$$

$$\boldsymbol{\Sigma}_z = \mathrm{cov}[\boldsymbol{A}\mathbf{X} + \boldsymbol{B}\mathbf{Y}] = \mathrm{cov}[\boldsymbol{A}\mathbf{X}] + \mathrm{cov}[\boldsymbol{B}\mathbf{Y}] = \boldsymbol{A}\boldsymbol{\Sigma}_x\boldsymbol{A}^{\mathrm{T}} + \boldsymbol{B}\boldsymbol{\Sigma}_y\boldsymbol{B}^{\mathrm{T}}, \qquad (1.28)$$

with $\boldsymbol{A}$ and $\boldsymbol{B}$ being some deterministic matrices specifying the linear combination of $\mathbf{X}$ and $\mathbf{Y}$ that gives $\mathbf{Z}$.

[†]$\mathrm{Pr}[\lim_{n\to\infty} \bar{X}_n = \mu] = 1$

A *stochastic process* is defined as a collection of random variables

$$\{X(t), \text{ where } t \in T\}, \tag{1.29}$$

where each random variable is associated with a distinct point t in a parameter space T. The parameter space T, often interpreted as time, can be any set index, though it is commonly a subset of the real numbers $\mathbb{R}$.

We can sample the stochastic process $X(t)$ at times $t_0, t_1, \ldots, t_n$, ordered such that: $t_0 \leq t_1 \leq \ldots \leq t_n$. The evaluation of the stochastic process at these times yields a sequence of random variables $\mathbf{X}_{0:n} = \{X_0, X_1, \ldots, X_n\}$ with a joint probability density:

$$p(\boldsymbol{x}_{0:n}) = p(x_0, x_1, \ldots, x_n), \tag{1.30}$$

which can be related to the conditional PDF through the product rule:

$$p(\boldsymbol{x}_{0:n}) = p(\boldsymbol{x}_{0:n-1}, x_n) = p(x_n | \boldsymbol{x}_{0:n-1}) p(\boldsymbol{x}_{0:n-1}). \tag{1.31}$$

1.2.1   MARKOV PROCESS

A stochastic process is called a *Markov process* if it satisfies the Markov property, which states that the conditional probability distribution of future states of the process depends only upon the knowledge of the most recent realization, not on the sequence of events that preceded it.

Property 1.4 (Markov property). *A sequence of random variables $\mathbf{X}_{0:n} = \{X_k\}_{k=0,1,\ldots,n}$ form a Markov sequence (or Markov chain) if X_k given X_{k-1} is independent of all the other variables up to time k-1:*

$$p(x_k | \boldsymbol{x}_{0:k-1}) = p(x_k | x_{k-1}). \tag{1.32}$$

It follows that the form given in Eq. (1.31) can be simplified as:

$$p(\boldsymbol{x}_{0:n}) = p(x_0, x_1, \ldots, x_n) \underset{1.19}{=} p(x_n|\boldsymbol{x}_{0:n-1})p(\boldsymbol{x}_{0:n-1})$$

$$\underset{1.32}{=} p(x_n|x_{n-1})p(\boldsymbol{x}_{0:n-2}, x_{n-1})$$

$$\underset{1.19}{=} p(x_n|x_{n-1})p(x_{n-1}|\boldsymbol{x}_{0:n-2})p(\boldsymbol{x}_{0:n-2})$$

$$= p(x_n|x_{n-1})p(x_{n-1}|x_{n-2}) \cdot \ldots \cdot p(x_1|x_0)p(x_0), \qquad (1.33)$$

provided that the stochastic process $X(t)$ is evaluated at times $t_0, t_1, \ldots, t_n$ fulfilling $t_0 \leq t_1 \leq \ldots \leq t_n$.

Definition 1.14 (Chapman-Kolmogorov Equation). *The Chapman-Kolmogorov equation expresses the marginal transition probability between two states of a Markov process in terms of an intermediate state. Using the Markov property and the product rule, it takes the form*

$$p(x_k|x_{k-2}) = \int \mathrm{d}x_{k-1}\, p(x_k|x_{k-1})\, p(x_{k-1}|x_{k-2}), \qquad (1.34)$$

where the conditional independence $p(x_k|x_{k-1}, x_{k-2}) = p(x_k|x_{k-1})$ follows from the Markov assumption.

Proof. *The statement above can be quickly shown using the sum rule, the Markov assumption and the product rule:*

$$p(x_k|x_{k-2}) \underset{1.16}{=} \int \mathrm{d}x_{k-1} p(x_k, x_{k-1}|x_{k-2}) \underset{1.19}{=} \int \mathrm{d}x_{k-1} p(x_k|x_{k-1}, x_{k-2})p(x_{k-1}|x_{k-2})$$

$$\underset{1.32}{=} \int \mathrm{d}x_{k-1} p(x_k|x_{k-1})p(x_{k-1}|x_{k-2}). \qquad (1.35)$$

1.2.2 POISSON PROCESS

Definition 1.15 (Poisson process). *Let $\{N(t), t \in [0, \infty)\}$ be a stochastic process that counts the number of events occurring up to time t. The process $N(t)$ is said to be a Poisson process with rate $\lambda > 0$ if the following conditions are satisfied:*

*1. **Initial condition:** The process starts with zero events at time zero, i.e. $N(0) = 0$.*

2. **The process has independent increments:** *For any $0 \leq s < t$, the number of events occurring in the interval $[s, t]$ is given by $\Delta N(t - s) := N(t) - N(s)$, which is independent of the number of events occurring prior to s.*

3. **The process has stationary increments:** *For any $0 \leq s < t$, the number of events in $[s, t]$ depends only on the length of the interval $t - s$, and not on the specific values of s and t.*

4. **Continuity in probability:** *For any $\varepsilon > 0$ and $t \geq 0$, it holds that*

$$\lim_{\Delta t \to 0} \Pr[|N(t + \Delta t) - N(t)| > \varepsilon] = 0. \tag{1.36}$$

In other words, the Poisson process $N(t)$ has almost surely continuous trajectories. Crucially, this condition still allows the infinitesimal increments of $N(t)$ to be finite, i.e. for the process to jump.

5. **Poisson distribution for the increments:** *For each $t \geq 0$, the number of events in $[t, t + \Delta t]$ follows a distribution with rates:*

$$\Pr[\Delta N(\Delta t) := N(t + \Delta t) - N(t) = 1] = \lambda \Delta t + o(\Delta t), \tag{1.37}$$

$$\Pr[\Delta N(\Delta t) := N(t + \Delta t) - N(t) = 0] = 1 - \lambda \Delta t + o(\Delta t), \tag{1.38}$$

$$\Pr[\Delta N(\Delta t) := N(t + \Delta t) - N(t) > 1] = o(\Delta t), \tag{1.39}$$

where the probability of a jump occurring between t and $t + \Delta t$ vanishes as $\Delta t \to 0$. Here, the little-o notation $o(\Delta t)^{\ddagger}$ represents a function that fulfills $\lim_{\Delta t \to 0} \frac{o(\Delta t)}{\Delta t} = 0$.

Therefore, if we now take the limit of $\Delta t \to 0$, we can define an infinitesimal increment:

$$dN := N(t + dt) - N(t), \tag{1.41}$$

---

‡The notation $o(\Delta t)$ means that the additional terms vanish faster than Δt, i.e., they become negligible as $\Delta t \to 0$. This differs from Big-O notation, which describes the scaling behavior up to a constant factor. Formally:

$$f(\Delta t) = \begin{cases} o(\Delta t^k) & \Longleftrightarrow & \lim_{\Delta t \to 0} \frac{f(\Delta t)}{\Delta t^k} = 0 \\ O(\Delta t^k) & \Longleftrightarrow & \lim_{\Delta t \to 0} \frac{f(\Delta t)}{\Delta t^k} = \text{const} \end{cases} \tag{1.40}$$

that counts the number of events that occur in the interval $[t, t + dt]$. Namely,

$$\Pr[dN = 1] = \lambda dt + o(dt), \tag{1.42}$$

$$\Pr[dN = 0] = 1 - \lambda dt + o(dt), \tag{1.43}$$

$$\Pr[dN > 1] = o(dt). \tag{1.44}$$

In other words, in an infinitesimal interval $[t, t+dt]$, there can be either one jump with a vanishing probability of λdt, or no jumps, with a probability $1 - \lambda dt$. Therefore, since dN can only have values 0 or 1, it follows that

$$dN^2 = dN. \tag{1.45}$$

Additionally, both the expectation and variance of this increment, dN, are equal to λdt:

$$\mathbb{E}[dN] = \sum_{k=0}^{1} k \Pr[dN = k] = \lambda dt + o(dt), \tag{1.46}$$

$$\text{Var}[dN] = \mathbb{E}[dN^2] - \mathbb{E}[dN]^2 = \sum_{k=0}^{1} k^2 \Pr[dN = k] - (\lambda dt)^2 = \lambda dt + o(dt). \tag{1.47}$$

The distribution above can be shown to be a Poisson distribution by discretizing a total time evolution $[0, T]$ by time-steps Δt s.t. $n = t/\Delta t$ and calculating the probability that k jumps occur in k intervals in a total of n intervals. The number of different combinations of k elements from a total group of $k \leq n$ elements is given by:

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}. \tag{1.48}$$

Therefore, in the limit of $\Delta t \to 0$, or equivalently, $n \to \infty$, we get

$$\begin{aligned}
\Pr[(N(t) = k] &= \lim_{n \to \infty} \frac{n!}{k!(n-k)!} (\lambda \Delta t)^k (1 - \lambda \Delta t)^{n-k} \\
&= \lim_{n \to \infty} \frac{n!}{(n-k)! \, n^k} \frac{(\lambda t)^k (1 - \lambda t/n)^n}{k!(1 - \lambda t/n)^k} \\
&= \frac{(\lambda t)^k e^{-\lambda t}}{k!}, \quad \text{where} \quad k = 0, 1, 2, \ldots \tag{1.49}
\end{aligned}$$

where in the last step we have used the following identities:

$$\lim_{n\to\infty} (1 - \lambda t/n)^n = e^{-\lambda t}, \tag{1.50}$$

$$\lim_{n\to\infty} (1 - \lambda t/n)^k = 1, \tag{1.51}$$

$$\lim_{n\to\infty} \frac{n!}{(n-k)! n^k} = \lim_{n\to\infty} \frac{n}{n} \frac{n-1}{n} \cdot \ldots \cdot \frac{n-k+1}{n} = \lim_{n\to\infty} \frac{n^k}{n^k} = 1. \tag{1.52}$$

As stated in the third condition of the definition of a Poisson process $N(t)$, the process has stationary increments. Therefore, the increment $\Delta N(\Delta t) = N(t + \Delta t) - N(t)$ has the same distribution as $N(\Delta t)$. Then,

$$\Pr[\Delta N = k] = \Pr[(N(t + \Delta t) - N(t)) = k] = \Pr[N(\Delta t) = k]$$

$$= \frac{(\lambda \Delta t)^k}{k!} e^{-\lambda \Delta t} \quad \text{for} \quad k = 0, 1, 2, \ldots. \tag{1.53}$$

So to sum up, the Poisson increment, $\Delta N(\Delta t) \sim \text{Pois}(\lambda \Delta t)$, is drawn from a Poisson distribution with occurrences k, which in principle can be $0, 1, 2$, and so on. However, as $\Delta t \to 0$, $dN = \lim_{\Delta t \to 0} \Delta N(\Delta t)$, and the probability of dN taking values greater than 1 becomes negligible. In other words, dN is only either 0 or 1 almost surely[§], i.e. $\Pr[dN \in \{0, 1\}] = 1$.

These properties describe a discrete process where events occur one at a time, and with the time between events following an exponential distribution with the parameter λ.

1.2.3 WIENER PROCESS

Definition 1.16 (Wiener process). *Let $\{W(t), t \in [0, \infty)\}$ be a stochastic process representing the continuous evolution of a random variable over time. The process $W(t)$ is said to be a Wiener process if it satisfies the following conditions:*

1. ***Initial condition:*** $W(0) = 0$ *almost surely.*

2. ***Independent increments:*** *For any $0 \leq s < t$, the increment $\Delta W(t - s) := W(t) - W(s)$ is independent of the process history up to time s. Thus, evolution over any interval depends only on that interval, independent of past values.*

[§]An event happens *almost surely* when it happens with probability 1.

3. **Stationary increments:** *For any* $0 \leq s < t$, *the increment* $W(t) - W(s)$ *depends only on the length of the interval* $t - s$ *and not on the specific values s and t. This means that the distribution of* $\Delta W(t-s) = W(t) - W(s)$ *is identical to that of* $W(t-s)$, *i.e. the increment is just another Wiener process.*

4. **Continuity in probability:** *The process* $W(t)$ *is almost surely continuous in t. Specifically, for any $\varepsilon > 0$ and $t \geq 0$, it holds that*

$$\lim_{\Delta t \to 0} \Pr[|W(t + \Delta t) - W(t)| > \varepsilon] = 0 \tag{1.54}$$

5. **Normally distributed increments:** *For any* $0 \leq s < t$, *the increment* $W(t) - W(s)$ *is normally distributed with mean zero and variance $t - s$. Specifically,*

$$\Delta W(t - s) := W(t) - W(s) \sim \mathcal{N}(0, t - s), \tag{1.55}$$

where $\mathcal{N}(\,\cdot\,,\,\cdot\,)$ *is a Gaussian distribution, introduced in Eq. (1.25), with first and second moments:*

$$\mathbb{E}[\Delta W(t - s)] = 0, \tag{1.56}$$

$$\mathbb{E}\left[\Delta W(t - s)^2\right] = t - s. \tag{1.57}$$

Thus, if we take the limit $\Delta t \to 0$, we can define an infinitesimal increment

$$dW := W(t + dt) - W(t) \sim \mathcal{N}(0, dt), \tag{1.58}$$

which represents the change in the Wiener process over the infinitesimal interval $[t, t + dt]$ and has mean zero and variance dt:

$$\mathbb{E}[dW] = 0, \tag{1.59}$$

$$V[dW] = \mathbb{E}\left[dW^2\right] = dt. \tag{1.60}$$

The paths of the Wiener process are continuous but nowhere differentiable. In other words, although $W(t)$ has continuous sample paths, it has no well-defined slope at any point.

Stochastic calculus is a branch of mathematics that extends traditional calculus to handle systems driven by randomness. It addresses how to integrate and differentiate functions of stochastic processes, such as $f(X(t))$, with respect to other stochastic processes. One possible way of doing that is with the Itô integral, which integrates a given stochastic process $X(t)$ with respect to the Wiener process $W(t)$:

$$\int_0^T X(t)dW. \tag{1.61}$$

Since $W(t)$ has (random) continuous but nowhere differentiable paths, classical calculus techniques such as Riemann-Stieltjes integration is not applicable.

1.3.1   Itô integral

To define the Itô integral, let us consider a function $f(t, X(t))$ that depends on a time parameter t and some stochastic process $X(t)$ up to time t. It follows that the function $f(t, X(t))$ is in itself another stochastic process, and we have to further assume that $f(t, X(t))$ is a *non-anticipating* function (a.k.a. *non-anticipating* process or *adapted* process), i.e. $f(t, X(t))$ is independent of the behavior of $X(t)$ and the Wiener process $W(t)$ in the future of t. More rigorously:

Definition 1.17 (Non-anticipating function). *A function $f(t, X(t))$ is said to be non-anticipating (or adapted) if, $\forall \tau > t$, it is statistically independent of the future increment $W(\tau) - W(t)$ of the Wiener process.*

Additionally, $f(t, X(t))$ cannot grow too quickly since the average of its squared value must be finite. Then,

Definition 1.18 (The Itô integral). *The Itô integral of $f(t, X(t))$ with respect to the Wiener process $W(t)$ over an interval $[0, T]$ is defined as*

$$S = \int_0^T f(t, X(t))\, dW := \operatorname*{ms\text{-}lim}_{n \to \infty} S_n \tag{1.62}$$

where ms-lim *denotes the mean-squared limit of the approximating sum S_n:*

$$S_n = \sum_{i=1}^n f(t_{i-1}, X(t_{i-1}))\, \Delta W(\Delta t_i), \tag{1.63}$$
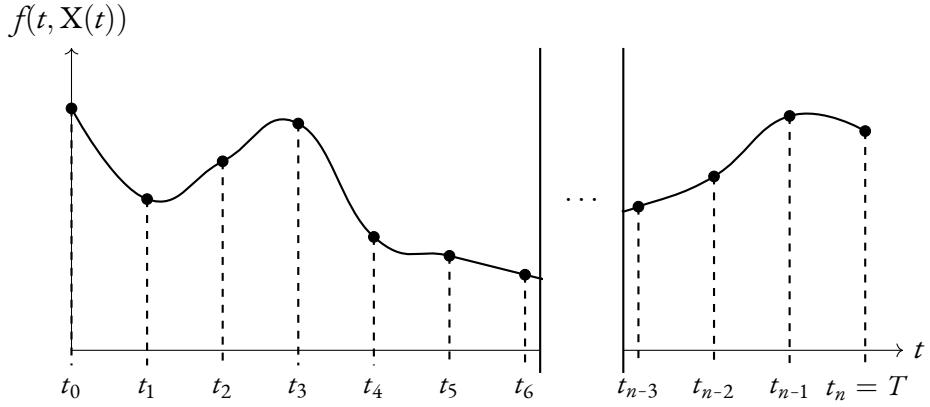
Figure 1.1: Discrete evaluation of the function $f(t, X(t))$. Visual representation of the partitioning of the time interval $[0, T]$ to evaluate the function $f(t, X(t))$ and define the Itô integral of $f(t, X(t))$ with respect to the noise process $W(t)$.

with the increment $\Delta W(\Delta t_i)$ given as $\Delta W(\Delta t_i) := W(t_i) - W(t_{i-1})$ and $\{t_i\}_{i=0}^{n}$ being a partition of the interval $[0, T]$ with $0 = t_0 \leq t_1 \leq \cdots \leq t_{n-1} \leq t_n = T$.

Note that the limit in Eq. (1.62) is taken in the mean square sense. Therefore, let us define what we mean by that:

Definition 1.19 (Mean-Squared Limit). *Let $\{S_n\}_{n=1}^{\infty}$ be a sequence of random variables and let S be another random variable. We say that S_n converges to S in the mean-squared sense, and write*

$$\text{ms-lim}_{n \to \infty} S_n = S, \tag{1.64}$$

if

$$\lim_{n \to \infty} \mathbb{E}\left[(S_n - S)^2\right] = 0, \tag{1.65}$$

where $\mathbb{E}[(S_n - S)^2]$ is what we call in statistics the mean squared error (MSE), which will be defined in more detail in the next chapter.

This type of convergence means that the approximating sum S_n will, on average, stay close to the Itô integral value S as we refine the partition, without requiring every individual path of the process S_n to converge exactly (pathwise) to the integral S. This might remind the reader of how, in statistics, we seek an estimator for a true quantity that minimizes the MSE. In that case, we easily recognize that even a good estimator may produce a poor estimate in a particular realization. What matters is that such deviations become increasingly rare on average as the approximation improves.

Besides the Itô integral, there are other ways to define a stochastic integral w.r.t. the Wiener process. Another approach, known as the *Stratonovich* integral, is defined as

$$\oint_0^T f(t, \mathrm{X}(t)) \, \mathrm{d}\mathrm{W} := \underset{n \to \infty}{\mathrm{ms\text{-}lim}} \sum_{i=1}^{n} f\left(\frac{t_{i\text{-}1} + t_i}{2}, \frac{\mathrm{X}(t_i) + \mathrm{X}(t_{i\text{-}1})}{2} \right) \Delta\mathrm{W}(\Delta t_i) \qquad (1.66)$$

where we superimpose $\mathcal{S}$ over the integral symbol to highlight that this integral is different from the Itô integral.

Example. (Itô integral of a Wiener process): Let us compute

$$\int_0^T \mathrm{W}(t) \, \mathrm{d}\mathrm{W}, \qquad (1.67)$$

where, for this stochastic integral, $f(t, \mathrm{X}(t)) = f(t, \mathrm{W}(t)) = \mathrm{W}(t)$. Recall the definition of an Itô integral as given in Eq. (1.62). In our case, S_n is of the form:

$$S_n = \sum_{i=1}^{n} \mathrm{W}(t_{i\text{-}1})\Delta\mathrm{W}(\Delta t_i) \qquad (1.68)$$

where $\Delta\mathrm{W}(\Delta t_i) := \mathrm{W}(t_i) - \mathrm{W}(t_{i\text{-}1})$. To simplify the sum, we can expand the summand as:

$$\mathrm{W}(t_{i\text{-}1})\Delta\mathrm{W}(\Delta t_i) = \frac{1}{2}\left(\mathrm{W}(t_i)^2 - \mathrm{W}(t_{i\text{-}1})^2 - \Delta\mathrm{W}(\Delta t_i)^2 \right), \qquad (1.69)$$

which makes evaluating the sum easier. In particular, substituting this identity into the sum S_n yields:

$$S_n = \frac{1}{2}\sum_{i=1}^{n}\left(\mathrm{W}(t_i)^2 - \mathrm{W}(t_{i\text{-}1})^2 \right) - \frac{1}{2}\sum_{i=1}^{n}\left(\Delta\mathrm{W}(\Delta t_i)^2 \right). \qquad (1.70)$$

Next, note that the first part of the sum S_n forms a so-called telescopic sum. This means that all intermediate terms cancel out, leaving only the boundary terms:

$$\sum_{i=1}^{n}\left(\mathrm{W}(t_i)^2 - \mathrm{W}(t_{i\text{-}1})^2 \right) = \mathrm{W}(T)^2 - \mathrm{W}(0)^2 = \mathrm{W}(T)^2, \qquad (1.71)$$

where in the last step we have used the initial condition of a Wiener process, $\mathrm{W}(0) = 0$ (see

Sec. *1.2.3*). Therefore, we can now rewrite S_n as:

$$S_n = \frac{1}{2}\left(W(T)^2 - \sum_{i=1}^{n}\Delta W(\Delta t_i)^2\right),\tag{1.72}$$

and hence computing the mean-squared limit of S_n

$$S = \underset{n\to\infty}{\text{ms-lim}}\, S_n = \frac{1}{2}\left(W(T)^2 - \underset{n\to\infty}{\text{ms-lim}}\sum_{i=1}^{n}\Delta W(\Delta t_i)^2\right)\tag{1.73}$$

reduces to computing the mean-squared limit of the sum of $\Delta W(\Delta t_i)^2$:

$$I = W(T)^2 - 2S = \underset{n\to\infty}{\text{ms-lim}}\, I_n = \underset{n\to\infty}{\text{ms-lim}}\sum_{i=1}^{n}\Delta W(\Delta t_i)^2.\tag{1.74}$$

In other words, we have now to find the Itô integral I s.t.

$$\lim_{n\to\infty}\mathbb{E}\left[(I_n - I)^2\right] = 0\tag{1.75}$$

is fulfilled. Our approach to tackle this problem is to give an ansatz to the solution of the Itô integral I and then check that it satisfies the mean-squared limit definition in Eq. (*1.75*). In particular, we choose the following ansatz:

$$I = T,\tag{1.76}$$

which is simply the mean of I_n. Namely,

$$\mathbb{E}[I_n] = \mathbb{E}\left[\sum_{i=1}^{n}\Delta W(\Delta t_i)^2\right] = \sum_{i=1}^{n}\mathbb{E}\left[\Delta W(\Delta t_i)^2\right] \underset{(1.57)}{=} \sum_{i=1}^{n}(t_i - t_{i-1}) = T,\tag{1.77}$$

where in the last step we have again a telescopic sum. Now that we know the reason behind our "inspired" ansatz, we just need to check whether I_n converges to T in the mean-squared sense:

$$\lim_{n\to\infty}\mathbb{E}\left[(I_n - I)^2\right] = \lim_{n\to\infty}\mathbb{E}\left[\left(\sum_{i=1}^{n}\Delta W(\Delta t_i)^2 - T\right)^2\right].\tag{1.78}$$

The first step is to expand the squared sum in the limit above:

$$\left(\sum_{i=1}^{n} \Delta W(\Delta t_i)^2 - T \right)^2 = \left(\sum_{i=1}^{n} \Delta W(\Delta t_i)^2 \right)^2 - 2T \sum_{i=1}^{n} \Delta W(\Delta t_i)^2 + T^2$$

$$= \sum_{i=1}^{n} \Delta W(\Delta t_i)^4 + 2 \sum_{i=1}^{n} \sum_{j=1}^{i-1} \Delta W(\Delta t_i)^2 \, \Delta W(\Delta t_j)^2 - 2T \sum_{i=1}^{n} \Delta W(\Delta t_i)^2 + T^2, \quad (1.79)$$

by employing the following identity:

$$\left(\sum_{i=1}^{n} x_i \right)^2 = \sum_{i=1}^{n} x_i^2 + 2 \sum_{i=1}^{n} \sum_{j=1}^{i-1} x_i \, x_j. \tag{1.80}$$

Then, we can take its expected value:

$$\mathbb{E}\left[\left(\sum_{i=1}^{n} \Delta W(\Delta t_i)^2 - T \right)^2 \right] = \sum_{i=1}^{n} \mathbb{E}\left[\Delta W(\Delta t_i)^4 \right] + 2 \sum_{i=1}^{n} \sum_{j=1}^{i-1} \mathbb{E}\left[\Delta W(\Delta t_i)^2 \, \Delta W(\Delta t_j)^2 \right]$$

$$- 2T \sum_{i=1}^{n} \mathbb{E}\left[\Delta W(\Delta t_i)^2 \right] + T^2, \tag{1.81}$$

and use the properties of the Wiener process (see Sec. 1.2.3) to calculate the higher moments of $\Delta W(\Delta t_i)$ *appearing above:*

$$\mathbb{E}\left[\Delta W(\Delta t_i)^4 \right] = 3\Delta t_i^2 = 3(t_i - t_{i-1})^2, \tag{1.82}$$

$$\mathbb{E}\left[\Delta W(\Delta t_i)^2 \Delta W(\Delta t_j)^2 \right] = \mathbb{E}\left[\Delta W(\Delta t_i)^2 \right] \mathbb{E}\left[\Delta W(\Delta t_j)^2 \right] = \Delta t_i \, \Delta t_j, \tag{1.83}$$

where the last expression uses the property of independent increments, since $i > j \quad \forall\, i, j.$
Therefore,

$$\mathbb{E}\left[\left(\sum_{i=1}^{n} \Delta W(\Delta t_i)^2 - T \right)^2 \right] = 3 \sum_{i=1}^{n} (t_i - t_{i-1})^2 + 2 \sum_{i=1}^{n} \sum_{j=1}^{i-1} (t_i - t_{i-1})(t_j - t_{j-1})$$

$$- 2T \sum_{i=1}^{n} (t_i - t_{i-1}) + T^2. \tag{1.84}$$

If we now simplify the telescopic sums

$$\sum_{i=1}^{n}(t_i - t_{i-1}) = T, \tag{1.85}$$

and employ the identity specified in Eq. (1.80) to roll back the term

$$2\sum_{i=1}^{n}\sum_{j=1}^{i-1}(t_i - t_{i-1})(t_j - t_{j-1}) = \left(\sum_{i=1}^{n}(t_i - t_{i-1})\right)^2 - \sum_{i=1}^{n}(t_i - t_{i-1})^2$$

$$= T^2 - \sum_{i=1}^{n}(t_i - t_{i-1})^2, \tag{1.86}$$

we can finally substitute all that into Eq. (1.84) and take its limit to verify that $I = T$. Namely,

$$\lim_{n\to\infty}\mathbb{E}\left[\left(\sum_{i=1}^{n}\Delta W(\Delta t_i)^2 - T\right)^2\right] = 2\lim_{n\to\infty}\sum_{i=1}^{n}(t_i - t_{i-1})^2 = 0. \tag{1.87}$$

Hence, it follows that

$$I = \operatorname*{ms\text{-}lim}_{n\to\infty} I_n = \operatorname*{ms\text{-}lim}_{n\to\infty}\sum_{i=1}^{n}\Delta W(\Delta t_i)^2 = T, \tag{1.88}$$

and thus,

$$S = \int_{0}^{T} W(t)dW = \operatorname*{ms\text{-}lim}_{n\to\infty}\sum_{i=1}^{n}W(t_{i-1})\Delta W(\Delta t_i) = \frac{1}{2}\left(W(T)^2 - \operatorname*{ms\text{-}lim}_{n\to\infty} I_n\right)$$

$$= \frac{1}{2}\left(W(T)^2 - T\right). \tag{1.89}$$

1.3.2 $dW^2 = dt$ AND $dW^n = 0$ FOR $n > 2$

The differential of a Wiener process, denoted as dW, behaves differently from ordinary differentials. In particular, a key result in stochastic calculus is that the square of dW is equal to the

differential of time, dt, while higher powers of dW vanish. More precisely, we want to prove that

$$(\mathrm{dW})^{2+N} = \begin{cases} \mathrm{d}t & \text{for} \quad N = 0, \\ \\ 0 & \text{for} \quad N > 0. \end{cases} \tag{1.90}$$

To do so, we have to use the Itô integral and show

$$\int_0^T f(t, \mathrm{X}(t)) \, (\mathrm{dW})^{2+N} := \underset{n \to \infty}{\text{ms-lim}} \sum_{i=1}^n f(t_{i\text{-}1}, \mathrm{X}(t_{i\text{-}1})) \, \Delta \mathrm{W}(\Delta t_i)^{2+N}$$

$$= \begin{cases} \int_0^T f(t, \mathrm{X}(t)) \, \mathrm{d}t & \text{for} \quad N = 0, \\ \\ 0 & \text{for} \quad N > 0 \end{cases}. \tag{1.91}$$

To tackle this proof, we need to divide it into two: first show that dW^2 = dt and then, demonstrate that dW^{2+N} = 0 for $N > 2$.

Proof. *To prove that* dW^2 = dt *we have to employ the definition of the Itô integral and show:*

$$\underset{n \to \infty}{\text{ms-lim}} \sum_{i=1}^n f(t_{i\text{-}1}, \mathrm{X}(t_{i\text{-}1})) \, \Delta \mathrm{W}(\Delta t_i)^2 = \int_0^T f(t, \mathrm{X}(t)) \, \mathrm{d}t. \tag{1.92}$$

However, to understand this limit, we must recall the definition of convergence in the mean-square sense, as introduced in Eq. (1.65). This means that, for the sequence of approximating sums S_n and the limiting value S, we require

$$\lim_{n \to \infty} \mathbb{E}\left[(S_n - S)^2\right] = 0, \tag{1.93}$$

where we define

$$S_n = \sum_{i=1}^n f(t_{i\text{-}1}, \mathrm{X}(t_{i\text{-}1})) \, \Delta \mathrm{W}(\Delta t_i)^2, \tag{1.94}$$

and

$$S = \int_0^T f(t, \mathrm{X}(t)) \, \mathrm{d}t = \lim_{n \to \infty} \sum_{i=1}^n f(t_{i\text{-}1}, \mathrm{X}(t_{i\text{-}1})) \, \Delta t_i. \tag{1.95}$$

Therefore, if we substitute the expressions for S_n and S, we get:

$$\lim_{n\to\infty} \mathbb{E}\left[(S_n - S)^2\right] = \lim_{n\to\infty} \mathbb{E}\left[\left(\sum_{i=1}^{n} f(t_{i-1}, X(t_{i-1}))\left(\Delta W(\Delta t_i)^2 - \Delta t_i\right)\right)^2\right]. \quad (1.96)$$

Note that the squared sum in the limit above can be expanded according to the identity in Eq. (1.80), i.e.:

$$\mathbb{E}\left[\left(\sum_{i=1}^{n} f(t_{i-1}, X(t_{i-1}))\left(\Delta W(\Delta t_i)^2 - \Delta t_i\right)\right)^2\right] = \sum_{i=1}^{n} \mathbb{E}\left[f_{i-1}^2\right]\mathbb{E}\left[\left(\Delta W(\Delta t_i)^2 - \Delta t_i\right)^2\right]$$

$$+ 2\sum_{i=1}^{n}\sum_{j=1}^{i-1} \mathbb{E}\left[f_{i-1} f_{j-1}\right]\left(\mathbb{E}\left[\Delta W(\Delta t_i)^2\right] - \Delta t_i\right)\left(\mathbb{E}[\Delta W(\Delta t_j)^2] - \Delta t_j\right), \quad (1.97)$$

where we simplify the notation by writing f_{i-1} instead of $f(t_{i-1}, X(t_{i-1}))$. Above we also use that $f_{i-1} = f(t_{i-1}, X(t_{i-1}))$ is independent of $\Delta W(\Delta t_i) = W(t_i) - W(t_{i-1})$. This follows from:

1. *The function f being a non-anticipating function, i.e. it does not depend on future values (recall Definition 1.17).*

2. *The property of stationary increments, fundamental to the definition of Lévy processes, and in particular, Wiener processes (if needed, refer back to Sec. 1.2.3).*

Even though $\Delta W(\Delta t_i)$ is defined in terms of $W(t_{i-1})$, its distribution is independent of $W(t_{i-1})$ because it is determined solely by the interval Δt_i. Namely, recall from Sec. 1.2.3 that $\Delta W(\Delta t_i) = W(t_i) - W(t_{i-1}) = W(\Delta t_i) \sim \mathcal{N}(0, \Delta t_i)$. This lack of dependency on the actual path of $W(t)$ up to t_{i-1} makes $\Delta W(\Delta t_i)$ statistically independent of $W(t_{i-1})$ despite being defined as $\Delta W(\Delta t_i) := W(t_i) - W(t_{i-1})$. Therefore, even if f was a function of the process $W(t)$, i.e. $f(t, W(t))$, or of a process $X(t)$ somehow correlated with $W(t)$, f_{i-1} is still independent of $\Delta W(\Delta t_i)$. Hence,

$$\mathbb{E}\left[f_{i-1}^2\left(\Delta W(\Delta t_i)^2 - \Delta t_i\right)^2\right] = \mathbb{E}\left[f_{i-1}^2\right]\mathbb{E}\left[\left(\Delta W(\Delta t_i)^2 - \Delta t_i\right)^2\right], \quad (1.98)$$

$$\mathbb{E}\left[f_{i-1} f_{j-1}\left(\Delta W(\Delta t_i)^2 - \Delta t_i\right)\left(\Delta W(\Delta t_j)^2 - \Delta t_j\right)\right] = \quad (1.99)$$

$$= \mathbb{E}\left[f_{i-1} f_{j-1}\right]\left(\mathbb{E}\left[\Delta W(\Delta t_i)^2\right] - \Delta t_i\right)\left(\mathbb{E}[\Delta W(\Delta t_j)^2] - \Delta t_j\right), \quad \text{since} \quad i > j \quad \forall i, j.$$

Note that in the last expression, we have used that

$$\mathbb{E}\left[\Delta W(\Delta t_i)\Delta W(\Delta t_j)\right] = \mathbb{E}[\Delta W(\Delta t_i)]\,\mathbb{E}\left[\Delta W(\Delta t_j)\right] \quad since \quad i > j, \tag{1.100}$$

which holds because the Wiener increments of different time steps are independent (again, see Sec. 1.2.3). Furthermore, given that each increment $\Delta W(\Delta t_i)$ is normally distributed, then

$$\mathbb{E}\left[\Delta W(\Delta t_i)^2\right] = \Delta t_i, \tag{1.101}$$

$$\mathbb{E}\left[\left(\Delta W(\Delta t_i)^2 - \Delta t_i\right)^2\right] = \mathbb{E}\left[\Delta W(\Delta t_i)^4 - 2\Delta W(\Delta t_i)^2\Delta t_i + \Delta t_i^2\right]$$

$$= 3\Delta t_i^2 - 2\Delta t_i^2 + \Delta t_i^2 = 2\Delta t_i^2. \tag{1.102}$$

When we apply these results to Eq. (1.97) and take the limit of $n \to \infty$, we obtain

$$\lim_{n\to\infty}\mathbb{E}\left[(S_n - S)^2\right] = \lim_{n\to\infty} 2\sum_{i=1}^{n}\mathbb{E}\left[f_{i-1}^2\right]\Delta t_i^2 = 0, \tag{1.103}$$

since Δt_i is squared. In other words, the average mean squared error of the difference between the "Riemann-sum" approximation S_n and the limiting stochastic integral S converges to zero as the partition is refined. Thus, we have shown that

$$\int_0^T f(t, X(t))(dW)^2 := \text{ms-}\lim_{n\to\infty}\sum_{i=1}^{n} f(t_{i\text{-}1}, X(t_{i\text{-}1}))\,\Delta W(\Delta t_i)^2 = \int_0^T f(t, X(t))\,dt, \tag{1.104}$$

and hence

$$dW^2 = dt. \tag{1.105}$$

The proofs of $(dW)^{2+N} = 0$ when $N > 2$ and $dWdt = 0$ follow very similar steps to the previous one, and are provided in App. A.1 and App. A.2. All these results are only valid for the Itô integral, since we have used that ΔW_i is independent of the non-anticipating function $f_{i\text{-}1}$. However, the integrand in the Stratonovich integral is evaluated at the midpoint of the interval, i.e. $f_{i\text{-}1} = f\left(\frac{1}{2}(t_{i\text{-}1} + t_i), \frac{1}{2}(X(t_i) + X(t_{i\text{-}1}))\right)$, while the increment remains defined as $\Delta W(\Delta t_i) = W(t_i) - W(t_{i\text{-}1})$. Hence, $\Delta W(\Delta t_i)$ and $f_{i\text{-}1}$ are not necessarily independent even though f is non-anticipating. Thus, $dW^2 = dt$ does not hold in Stratonovich calculus [70].

1.3.3 Itô's lemma

Let us now introduce Itô's lemma, the stochastic analog of the chain rule. It tells us how to differentiate functions of stochastic processes and is essential for formulating and solving stochastic differential equations (SDEs).

Lemma 1.2 (Itô's lemma). *Let $W(t)$ be a Wiener process and consider a function $f(t, W(t))$, twice differentiable w.r.t. $W(t)$ and once differentiable with respect to t. Then, Itô's lemma states:*

$$\mathrm{d}f(t, W(t)) = \left(\frac{\partial f}{\partial t} + \frac{1}{2}\frac{\partial^2 f}{\partial W^2} \right)\mathrm{d}t + \frac{\partial f}{\partial W}\mathrm{d}W, \tag{1.106}$$

where $\partial f/\partial W$ and $\partial^2 f/\partial W^2$ are the first and second partial derivatives of $f(t, W(t))$ w.r.t. $W(t)$, evaluated at $(t, W(t))$.

In this form, Itô's lemma shows that the differential of $f(t, W(t))$ has an additional term $1/2\,\partial^2 f/\partial W^2$ that accounts for the quadratic variation of the Wiener process $W(t)$.

Proof. *Consider the function $f(t + \Delta t, W(t + \Delta t))$, where Δt is a finite time-step s.t. $\Delta t > 0$. Additionally, let us define the Wiener increment over this small interval Δt as $\Delta W(\Delta t) := W(t + \Delta t) - W(t)$. Then, if we expand $f(t + \Delta t, W(t + \Delta t))$ using the Taylor series to second order in $W(t)$ and first in t, we get:*

$$f(t+\Delta t, W(t+\Delta t)) \approx f(t, W(t)) + \frac{\partial f}{\partial t}\Delta t + \frac{\partial f}{\partial W}\Delta W(\Delta t) + \frac{1}{2}\frac{\partial^2 f}{\partial W^2}(\Delta W(\Delta t))^2, \tag{1.107}$$

where higher-order terms in Δt and $\Delta W(\Delta t)$ are ignored since they will vanish as $\Delta t \to 0$. The key difference from ordinary calculus comes from the term including $(\Delta W(\Delta t))^2$, since as shown in the previous section, $\mathrm{d}W^2 = \mathrm{d}t$ in the limit of $\Delta t \to 0$.
Therefore, by now rearranging the terms above and taking the limit of $\Delta t \to 0$, we get

$$\mathrm{d}f(t, W(t)) = \left(\frac{\partial f}{\partial t} + \frac{1}{2}\frac{\partial^2 f}{\partial W^2} \right)\mathrm{d}t + \frac{\partial f}{\partial W}\mathrm{d}W. \tag{1.108}$$

Additionally, there is also a more general form for Itô's lemma, providing the differential of a function $f(t, X(t))$, where $X(t)$ is a general stochastic process. Namely,

Lemma 1.3 (General Itô's lemma). *Let $X(t)$ be an Itô process and let $f(t, X(t))$ be a twice-differentiable*

function of time and the process $X(t)$. *Then, Itô's lemma states:*

$$\mathrm{d}f(t, X(t)) = \frac{\partial f}{\partial t}\mathrm{d}t + \frac{\partial f}{\partial X}\mathrm{d}X + \frac{1}{2}\frac{\partial^2 f}{\partial X^2}\mathrm{d}X\mathrm{d}X, \qquad (1.109)$$

where $\mathrm{d}X\mathrm{d}X$ *represents the quadratic variation of* $X(t)$. *For* $X(t) = W(t)$, *the quadratic variation is* $\mathrm{d}t$.

Example. (How to apply the Itô lemma (I)): *Let us consider now* $f(t, W(t)) = \mu\,t + \sigma W(t)$, *where* $\mu,\ \sigma$ *are constant parameters. If now we evaluate the differential* $\mathrm{d}f(t, W(t))$ *using the Itô form, we get:*

$$\mathrm{d}f(t, W(t)) = \mu\,\mathrm{d}t + \sigma\,\mathrm{d}W. \qquad (1.110)$$

Note that a function of a stochastic process is simply another stochastic process, and thus, we can rename the function of $W(t)$ *as* $X(t) := f(t, W(t))$. *Then, we can view the Itô differential in Eq. (1.110) as a stochastic differential equation of the process* $X(t)$. *Namely,*

$$\mathrm{d}X(t) = \mu\,\mathrm{d}t + \sigma\,\mathrm{d}W. \qquad (1.111)$$

Example. (How to apply the Itô lemma (II)): *In this second example, let us consider a different function:* $f(t, W(t)) = W(t)^2$. *If we apply Itô's lemma to* $f(t, W(t))$, *we obtain*

$$\mathrm{d}f(t, W(t)) = \mathrm{d}\left(W(t)^2\right) = 0 \cdot \mathrm{d}t + 2W(t)\mathrm{d}W + \frac{1}{2}2\mathrm{d}t = \mathrm{d}t + 2W(t)\mathrm{d}W, \qquad (1.112)$$

since $\partial f/\partial t = 0$ *because f does not depend on t, and*

$$\frac{\partial f}{\partial W} = 2W(t), \qquad (1.113)$$

$$\frac{\partial^2 f}{\partial W^2} = 2. \qquad (1.114)$$

This result shows how the second derivative term in Itô's lemma influences the differential of functions that are at least quadratic in $W(t)$.

In ordinary calculus, the Leibniz product rule provides the derivative of a product of two or more functions. For stochastic processes, however, the corresponding rule for differentiating the product of two stochastic processes — or two functions of stochastic processes, which are themselves stochastic processes too — is known as the Itô-Leibniz product rule.

Rule. *(Itô-Leibniz product rule): Let $X(t)$ and $Y(t)$ be two Itô processes. Then, the differential of the product $Z(t) = X(t)Y(t)$ is given by:*

$$dZ(t) = d\left(X(t)Y(t)\right) = X(t)dY(t) + Y(t)dX(t) + dX(t)dY(t). \qquad (1.115)$$

Proof. To derive the Itô-Leibniz product rule, we can start considering the following increment of $Z(t) = X(t)Y(t)$:

$$\Delta Z(\Delta t) = Z(t + \Delta t) - Z(t) = X(t + \Delta t)Y(t + \Delta t) - X(t)Y(t), \qquad (1.116)$$

and then add and subtract the term $X(t)Y(t + \Delta t)$:

$$\begin{aligned}
\Delta Z(\Delta t) &= X(t + \Delta t)Y(t + \Delta t) + X(t)Y(t + \Delta t) - X(t)Y(t + \Delta t) - X(t)Y(t) \\
&= Y(t + \Delta t)\left(X(t + \Delta t) - X(t)\right) + X(t)\left(Y(t + \Delta t) - Y(t)\right). \qquad (1.117)
\end{aligned}$$

Next, we again add and subtract another term, $Y(t)\left(X(t + \Delta t) - X(t)\right)$:

$$\begin{aligned}
\Delta Z(\Delta t) &= Y(t + \Delta t)\left(X(t + \Delta t) - X(t)\right) - Y(t)\left(X(t + \Delta t) - X(t)\right) \\
&\quad + Y(t)\left(X(t + \Delta t) - X(t)\right) + X(t)\left(Y(t + \Delta t) - Y(t)\right) \\
&= \left(Y(t + \Delta t) - Y(t)\right)\left(X(t + \Delta t) - X(t)\right) \\
&\quad + Y(t)\left(X(t + \Delta t) - X(t)\right) + X(t)\left(Y(t + \Delta t) - Y(t)\right) \\
&= \Delta Y(\Delta t)\Delta X(\Delta t) + Y(t)\Delta X(\Delta t) + X(t)\Delta Y(\Delta t). \qquad (1.118)
\end{aligned}$$

If we now take the limit of $\Delta t \to 0$, we derive the Itô-Leibniz product rule:

$$dZ(t) = X(t)dY(t) + Y(t)dX(t) + dY(t)dX(t). \qquad (1.119)$$

Stochastic differential equations are ordinary differential equations with the right hand side perturbed with a random term. To mathematically define SDEs, we use Itô notation. An example of a simple SDE is given in Eq. (1.111), which its form in Itô notation is derived using Itô's lemma. More generally though, a differential equation perturbed by white Gaussian noise[¶] can be written in Itô form as:

$$dX(t) = \mu(X(t), t)dt + \sigma(X(t), t)dW, \tag{1.120}$$

where $X(t)$ is the stochastic process we aim to solve for, i.e. the state of the system at time t, and dW is the Wiener differential. The terms $\mu(X(t), t)$ and $\sigma(X(t), t)$ are the drift and diffusion term, representing the deterministic and random parts of the evolution, respectively.

Solutions to SDEs can be generally classified into two categories: analytical, with closed-form solutions, and numerical, which approximate the evolution of $X(t)$ through discretized steps. A simple example of an analytical closed-form solution is given by the SDE in Eq. (1.111), which has the solution $X(t) = \mu t + \sigma W(t)$.

Example. (Geometric Brownian motion solution): *Another well-known example is the geometric Brownian motion satisfying the following SDE:*

$$dX(t) = \mu X(t)dt + \sigma X(t)dW, \tag{1.121}$$

which has the solution

$$X(t) = X(0) \exp\left\{\left(\mu - \frac{1}{2}\sigma^2\right)t + \sigma W(t)\right\}, \tag{1.122}$$

where $X(0)$ is the initial value of the process $X(t)$.

Proof. *To find the solution of the geometric Brownian motion defined by Eq. (1.121), let us compute the differential of the natural logarithm of $X(t)$. In particular, since $\ln(X(t))$ is a function of the process $X(t)$, i.e. $f(t, X(t))$, we have to use Itô's lemma (1.109) to compute the*

[¶]White Gaussian noise is a random signal with a flat power spectral density and Gaussian amplitude distribution; it can be informally thought of as the derivative of the Wiener process. For a more in-depth explanation, see Sec. 1.3.8.

differential of $f(t, X(t)) = \ln(X(t))$. *Namely,*

$$d\left(\ln(X(t))\right) = \frac{1}{2}\left(\frac{\partial^2}{\partial X^2}\ln(X(t))\right)dX\,dX + \left(\frac{\partial}{\partial X}\ln(X(t))\right)dX, \qquad (1.123)$$

where $dX\,dX$ *is the quadratic variation of* $X(t)$:

$$dX\,dX = \mu^2 X^2(t)dt^2 + 2\sigma\mu X^2(t)\,dW dt + \sigma^2 X^2(t)\,dW^2 = \sigma^2 X^2(t)\,dt + O(dt^{3/2}). \quad (1.124)$$

Therefore, we can write the differential of $\ln(X(t))$ *as*

$$d\left(\ln(X(t))\right) = -\frac{1}{2}\frac{1}{X^2(t)}\sigma^2 X^2(t)dt + \frac{dX}{X(t)} = -\frac{1}{2}\sigma^2 dt + \frac{dX}{X(t)} =$$

$$= -\frac{1}{2}\sigma^2 dt + \mu\,dt + \sigma dW, \qquad (1.125)$$

where in the last step we used Eq. (1.121). Hence, by now integrating from 0 to t, we get

$$\ln(X(t)) - \ln(X(0)) = \ln\left(\frac{X(t)}{X(0)}\right) = \left(\mu - \frac{1}{2}\sigma^2\right)t + \sigma W(t), \qquad (1.126)$$

which, by exponentiating its both sides, yields an analytical form for $X(t)$:

$$X(t) = X(0)\exp\left\{\left(\mu - \frac{1}{2}\sigma^2\right)t + \sigma W(t)\right\}. \qquad (1.127)$$

1.3.6 Ornstein-Uhlenbeck process

The Ornstein-Uhlenbeck (OU) process is another process with Gaussian noise that models mean-reverting behavior, i.e. the system has a tendency to return to an equilibrium state. The dynamics of the OU process $X(t)$ are governed by the following SDE:

$$dX(t) = -\theta\left(X(t) - \mu\right)dt + \sigma\,dW, \qquad (1.128)$$

where $\theta > 0$, μ and σ are constants and dW is the Wiener differential. The deterministic term $-\theta X(t)$ pulls the process back towards a long-term mean, μ, while the stochastic term σdW introduces the random kicks.

Just like in the example of the geometric Brownian motion, the OU process (here we assume $\mu = 0$) also has a formal solution:

$$X(t) = X(0)e^{-\theta t} + \sigma \int_0^t e^{-\theta(t-\tau)} dW(\tau). \tag{1.129}$$

Proof. *Define a function of the stochastic process $X(t)$ as $f(X(t), t) := X(t)e^{\theta t}$, such that when computing its differential with Itô's lemma we get:*

$$d\left(X(t)e^{\theta t}\right) = \theta X(t)e^{\theta t}dt + e^{\theta t}dX(t) = \sigma dW(t)e^{\theta t}. \tag{1.130}$$

Hence, integrating both sides from 0 to t yields:

$$X(t)e^{\theta t} = X(0) + \sigma \int_0^t e^{\theta \tau} dW(\tau), \tag{1.131}$$

which can be simplified by taking the exponential term to the r.h.s. to reveal the formal solution of the process:

$$X(t) = X(0)e^{-\theta t} + \sigma \int_0^t e^{-\theta(t-\tau)} dW(\tau). \tag{1.132}$$

From this expression, one can derive the mean and the covariance of the process:

$$\mathbb{E}[X(t)] = X(0)e^{-\theta t}, \tag{1.133}$$

$$\begin{aligned}
\text{cov}[X(t)X(s)] &= \mathbb{E}[(X(t) - \mathbb{E}[X(t)])(X(s) - \mathbb{E}[X(s)])] \\
&= \sigma^2 e^{-\theta(s+t)} \mathbb{E}\left[\int_0^t e^{\theta \tau} dW(\tau) \int_0^s e^{\theta v} dW(v)\right] \\
&= \sigma^2 e^{-\theta(s+t)} \int_0^t \int_0^s e^{\theta \tau} e^{\theta v} \mathbb{E}[dW(\tau) dW(v)] \\
&= \sigma^2 e^{-\theta(s+t)} \int_0^t e^{\theta \tau} \left[\int_0^s e^{\theta v} \delta(\tau - v)dv\right] d\tau \\
&= \sigma^2 e^{-\theta(s+t)} \int_0^{\min(t,s)} e^{2\theta \tau} d\tau \\
&= \sigma^2 e^{-\theta(s+t)} \frac{1}{2\theta} \left(e^{2\theta \min(t,s)} - 1\right) = \frac{\sigma^2}{2\theta} \left(e^{-\theta|t-s|} - e^{-\theta(s+t)}\right)
\end{aligned} \tag{1.134}$$

where $\min(t, s)$ appears because $\mathrm{e}^{\theta \tau}$ is non-zero only when $0 \leq \tau \leq s$. It follows that the variance is:

$$\mathrm{V}[\mathrm{X}(t)] = \mathrm{cov}[\mathrm{X}(t)\mathrm{X}(t)] = \frac{\sigma^2}{2\theta} \left(1 - \mathrm{e}^{-2\theta t}\right), \tag{1.135}$$

which for small θ can be approximated as

$$\mathrm{V}[\mathrm{X}(t)] \approx \sigma^2 t. \tag{1.136}$$

The OU process can also be described in terms of a PDF $p(x, t)$, which describes the likelihood of the process $\mathrm{X}(t)$ being in a state x at a time t. This probability density evolves according to the Fokker-Planck equation, a partial differential equation that governs the time evolution of the probability distribution for a stochastic process. For the zero-mean ($\mu = 0$) OU process, the Fokker-Planck equation is given by:

$$\frac{\partial p}{\partial t} = \theta \frac{\partial}{\partial x}(x p) + \frac{\sigma^2}{2} \frac{\partial^2 p}{\partial x^2}. \tag{1.137}$$

Solving the Fokker-Planck equation with the initial condition $p(x, t_0) = \delta(x - x_0)$ yields the following transition probability function:

$$p(x, t | x_0, t_0) = \sqrt{\frac{\theta}{\pi \sigma^2 (1 - \mathrm{e}^{-2\theta (t - t_0)})}} \exp\left(-\frac{\theta}{\sigma^2} \frac{\left(x - x_0 \mathrm{e}^{-\theta (t - t_0)}\right)^2}{1 - \mathrm{e}^{-2\theta (t - t_0)}}\right), \tag{1.138}$$

where $t > t_0$, $x = \mathrm{X}(t)$ and $x_0 = \mathrm{X}(t_0)$. Note that the mean and variance of the OU process $\mathrm{X}(t)$, given in Eq. (1.133) and Eq. (1.135), respectively, can be easily inferred from the Gaussian form of the transition probability by setting the starting time $t_0 = 0$ in Eq. (1.138).

1.3.7 Numerical methods for SDEs

For cases where closed-form solutions of SDEs are not attainable, numerical methods provide a practical way to approximate the behavior of stochastic processes over time. Numerical solutions to SDEs rely on time discretization and iterative computation of approximate values of the stochastic process. One of the most widely used methods for numerically solving SDEs is the *Euler-Maruyama* (EM) method, a straightforward extension of the classical Euler method for deterministic differential equations.

Consider a general SDE of the form specified in Eq. (1.120). Namely,

$$dX(t) = \mu\left(X(t), t\right) dt + \sigma\left(X(t), t\right) dW, \tag{1.139}$$

where $X(t)$ is the stochastic process for which we want to numerically solve the equation above over a time interval $[0, T]$. To implement the EM method, the time interval $[0, T]$ is divided into n discrete time steps of equal size $\Delta t = T/n$. Let $t_k = k\Delta t$ denote the discrete time points, with $k = 0, 1, 2, \ldots, n$. Then, the approximate solution at each time step, $X[k] \coloneqq X(t_k)$, is computed iteratively using the update formula:

$$X[k+1] = X[k] + \mu\left(t_k, X[k]\right) \Delta t + \sigma\left(t_k, X[k]\right) \Delta W_k, \tag{1.140}$$

where $\Delta W_k \coloneqq W(t_{k+1}) - W(t_k)$ is the increment of the Wiener process over the time step Δt, which is simulated by drawing from a Gaussian distribution with mean zero and variance Δt:

$$\Delta W_k \sim \mathcal{N}(0, \Delta t). \tag{1.141}$$

The EM method is a first-order method in both the time-step Δt and the Wiener increment ΔW_k. Specifically, it converges strongly with order $1/2$, meaning that the expected error between the true solution of X and the numerical solution scales as $O(\sqrt{\Delta t})$. For practical purposes, this implies that reducing size of the time step improves the accuracy of the method, but its precision is limited. Higher-order methods exist but they are often more complex and computationally expensive. In particular, when the drift and/or diffusion coefficients exhibit high nonlinearity or stiffness, then more sophisticated methods, such as the Milstein method or higher-order Runge-Kutta methods for SDEs may be necessary to achieve better accuracy.

1.3.8   WHITE NOISE

Ordinary differential equations can be extended to describe the dynamics of stochastic processes by adding a white noise term, such as:

$$\frac{dx}{dt} = \mu x + w(t), \tag{1.142}$$

where $w(t)^{\|}$ represents the white noise, also sometimes referred as a Langevin term. Even though we previously mentioned that white Gaussian noise $w(t)$ can be informally viewed as the formal derivative of a Wiener process, we have yet to rigorously define what we mean by white noise.

White noise processes differ significantly between discrete and continuous time, particularly in how they are defined and interpreted. In the discrete-time setting, white noise has an intuitive definition, involving sequences of uncorrelated (or independent) random variables with finite variance. In contrast, continuous-time white noise cannot be defined as a standard stochastic process, requiring more complicated mathematical machinery. Thus, we begin by introducing the simpler notion of discrete white noise and then move to continuous time.

Definition 1.20 (Discrete white noise). *Let $\{\boldsymbol{q}_k\}_{k\in\mathbb{Z}}$ be a real-valued discrete-time stochastic process. We say that it is a wide-sense white noise process if it satisfies [71]:*

$$\mathbb{E}[\boldsymbol{q}_k] = 0, \tag{1.143}$$

$$\mathrm{cov}[\boldsymbol{q}_k, \boldsymbol{q}_j] = \mathbb{E}\big[\boldsymbol{q}_k\boldsymbol{q}_j\big] = \boldsymbol{Q}_k\,\delta_{kj}, \tag{1.144}$$

where δ_{kj} is the Kronecker delta and $\boldsymbol{Q}_k > 0$ is the variance of $\boldsymbol{q}_k$. If, in addition, the random variables in $\{\boldsymbol{q}_k\}_{k\in\mathbb{Z}}$ are mutually independent, i.e.,

$$p(\boldsymbol{q}_1, \ldots, \boldsymbol{q}_k) = p(\boldsymbol{q}_1) \cdot \cdots \cdot p(\boldsymbol{q}_k) \tag{1.145}$$

then $\{\boldsymbol{q}_k\}_{k\in\mathbb{Z}}$ is also strict-sense white noise. Finally, when each element $\boldsymbol{q}_k$ of the white process is Gaussian, one speaks of a discrete Gaussian white noise:

$$\boldsymbol{q}_k \sim \mathcal{N}(0, \boldsymbol{Q}_k). \tag{1.146}$$

Unlike in discrete time, a continuous-time white noise process cannot be defined as a standard stochastic process that when evaluated point-wise yields a random variable with a finite variance. Instead, continuous white noise is treated as a *generalized* stochastic process, defined only through its action on test functions via integration:

Definition 1.21 (Continuous white noise). *A (generalized) continuous-time stochastic process*

$\|$In general, this noise in the so-called Langevin equation does not need to be white [70].

$\{\boldsymbol{w}(t), t \in [0, \infty)\}$ *taking values in* $\mathbb{R}^n$ *is called white noise if [72]:*

$$\mathbb{E}[\boldsymbol{w}(t)] = 0, \tag{1.147}$$

$$\mathrm{cov}[\boldsymbol{w}(t), \boldsymbol{w}(s)] = \mathbb{E}[\boldsymbol{w}(t)\boldsymbol{w}(s)] = \boldsymbol{Q}(t)\,\delta(t-s), \tag{1.148}$$

where $\delta(\,\cdot\,)$ *is the Dirac delta and* $\boldsymbol{Q}(t)$ *is the (continuous) covariance matrix, also known as the spectral density matrix [30].*

The term "white" in white noise refers to its flat power spectral density $S_w(\omega)$ over an infinite frequency range. To visualize this, consider the case of a white noise with a constant covariance, i.e. $Q(t) = Q$. Its power spectral density is given by:

$$S_w(\omega) = \int_{-\infty}^{\infty} \mathbb{E}[\boldsymbol{w}(t)\boldsymbol{w}(t+\tau)]\,\mathrm{e}^{i\omega\tau}\mathrm{d}\tau = Q. \tag{1.149}$$

Thus, this constant power spectral density across an infinite bandwidth implies that the generalized process has an infinite variance in the time domain. This is reflected by the Dirac delta $\delta(0)$ in the covariance function of Eq. (1.148) when $t = s$. Due to this infinite variance, $\boldsymbol{w}(t)$ cannot be treated as a standard stochastic process that can be evaluated point-wise yielding random variables with finite variance. Instead, it is referred to as a *generalized* process, just like the Dirac delta is called a generalized function: i.e. it is defined by multiplying it with a test function and then integrating:

$$\int_{-\infty}^{\infty} f(t)\delta(t)\mathrm{d}t = f(0) \iff \int_{-\infty}^{\infty} \boldsymbol{f}^{\mathrm{T}}(t)\boldsymbol{w}(t)\mathrm{d}t = \langle \boldsymbol{f}, \boldsymbol{w} \rangle \quad \text{is a well-defined random vector.} \tag{1.150}$$

In the case where the white noise is also Gaussian, it can be formally related to a Wiener process $\{\mathrm{W}(t), t \in [0, \infty)\}$ (introduced in Sec. 1.2.3) as:

$$w(t) = \frac{\mathrm{dW}(t)}{\mathrm{d}t}, \tag{1.151}$$

in a distributional sense. This means that for any suitable test function

$$\int f(t)w(t)\mathrm{d}t = \int f(t)\mathrm{dW}, \tag{1.152}$$

where the right-hand side is an Itô integral (see Sec. 1.3.1 for more on Itô integrals).

Another important point we should discuss is how to discretize continuous white noise. Since $w(t)$ is a generalized stochastic process whose properties are only well-defined under integration with a test function $f(t)$, we must utilize this and set $f(t) = 1$ in order to generate discrete random variables. Namely, over an interval $[t_{k-1}, t_k)$ of length Δt, the discretized white noise reads as

$$\boldsymbol{q}_k = \frac{1}{\Delta t} \int_{t_{k-1}}^{t_k} \boldsymbol{w}(\tau) \mathrm{d}\tau. \tag{1.153}$$

In the case where $\boldsymbol{w}(t)$ is a continuous *Gaussian* white noise, then its integral over the interval $[t_{k-1}, t_k)$ is simply the Wiener increment $\Delta \mathrm{W}_k$:

$$\Delta \mathrm{W}_k = \mathrm{W}(t_k) - \mathrm{W}(t_{k-1}) = \int_{t_{k-1}}^{t_k} \mathrm{dW}_\tau = \int_{t_{k-1}}^{t_k} \boldsymbol{w}(\tau) \mathrm{d}\tau, \tag{1.154}$$

such that

$$\boldsymbol{q}_k = \frac{\Delta \mathrm{W}_k}{\Delta t} = \frac{1}{\Delta t} \int_{t_{k-1}}^{t_k} \boldsymbol{w}(\tau) \mathrm{d}\tau, \tag{1.155}$$

where if the covariance of the continuous white Gaussian process is $\boldsymbol{Q}$, then this discretized white noise process is drawn from a Gaussian distribution with mean zero and covariance $\boldsymbol{Q}/\Delta t$:

$$\boldsymbol{q}_k \sim \mathcal{N}(0, \boldsymbol{Q}/\Delta t). \tag{1.156}$$

1.3.9 Discretization of a continuous linear and Gaussian system

Now that we know how to discretize white noise, let us explain how to discretize a LG system.

Proposition 1.4 (Discretization of a continuous-time LG system). *Let $\boldsymbol{x}(t)$ be the state of the system, $\boldsymbol{u}(t)$ the control input, and $\boldsymbol{y}(t)$ the measurement, and let $\boldsymbol{w}(t)$ and $\boldsymbol{v}(t)$ be zero-mean white Gaussian noises with covariances:*

$$\mathbb{E}\left[\boldsymbol{w}(t)\boldsymbol{w}(s)^{\mathrm{T}}\right] = \boldsymbol{Q}(t)\delta(t-s), \tag{1.157}$$

$$\mathbb{E}\left[\boldsymbol{v}(t)\boldsymbol{v}(s)^{\mathrm{T}}\right] = \boldsymbol{R}(t)\delta(t-s), \tag{1.158}$$

$$\mathbb{E}\left[\boldsymbol{w}(t)\boldsymbol{v}(s)^{\mathrm{T}}\right] = 0. \tag{1.159}$$

Consider the continuous-time LG model:

$$\dot{\boldsymbol{x}}(t) = \boldsymbol{F}(t)\,\boldsymbol{x}(t) + \boldsymbol{B}(t)\,\boldsymbol{u}(t) + \boldsymbol{G}(t)\,\boldsymbol{w}(t), \tag{1.160}$$

$$\boldsymbol{y}(t) = \boldsymbol{H}(t)\,\boldsymbol{x}(t) + \boldsymbol{v}(t). \tag{1.161}$$

Then, by discretizing the system above, we get:

$$\boldsymbol{x}_k = \boldsymbol{A}_{k\text{-}1}\boldsymbol{x}_{k\text{-}1} + \boldsymbol{B}_{k\text{-}1}\boldsymbol{u}_{k\text{-}1} + \boldsymbol{G}_{k\text{-}1}\boldsymbol{q}_{k\text{-}1}, \tag{1.162}$$

$$\boldsymbol{y}_k = \boldsymbol{H}_k\boldsymbol{x}_k + \boldsymbol{r}_k, \tag{1.163}$$

where $\boldsymbol{q}_{k-1} \sim \mathcal{N}(0, \boldsymbol{Q}_{k-1})$, $\boldsymbol{r}_k \sim \mathcal{N}(0, \boldsymbol{R}_k)$, $\mathbb{E}\big[\boldsymbol{q}_{k-1}\,\boldsymbol{r}_k^{\mathrm{T}}\big] = 0$, *and*

$$\boldsymbol{A}_{k\text{-}1} = \mathbb{I} + \boldsymbol{F}(t_{k\text{-}1})\,\Delta t, \quad \boldsymbol{B}_{k\text{-}1} = \boldsymbol{B}(t_{k\text{-}1})\,\Delta t, \quad \boldsymbol{G}_{k\text{-}1} = \boldsymbol{G}(t_{k\text{-}1})\,\Delta t,$$

$$\boldsymbol{H}_{k\text{-}1} = \boldsymbol{H}(t_{k\text{-}1}), \qquad \boldsymbol{Q}_{k\text{-}1} = \frac{\boldsymbol{Q}(t_{k\text{-}1})}{\Delta t}, \qquad \boldsymbol{R}_{k\text{-}1} = \frac{\boldsymbol{R}(t_{k\text{-}1})}{\Delta t}.$$
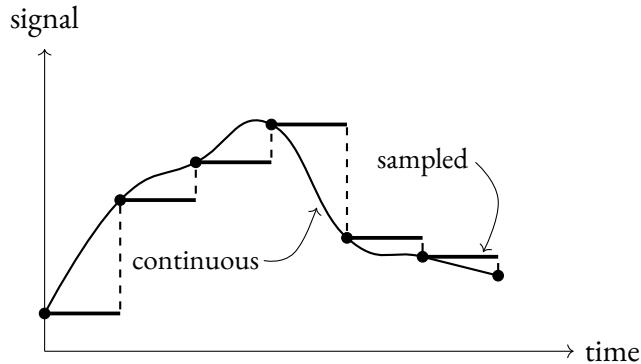
Figure 1.2: A continuous signal sampled using the zero-order hold assumption. To discretize a signal using zero-order hold, the time axis is divided in increments of Δt in order to evaluate the function at these steps: $k\Delta t$, with $k \in \mathbb{Z}$. The signal is then further assumed to maintain a constant value $f(k\Delta t)$ from time $k\Delta t$ to time $(k+1)\Delta t$.

> **Proof.** *Let us establish the equivalence between the continuous LG system and its discrete counterpart. The first step is to apply Theorem A.3 to solve Eq. (1.160), focusing on the interval*

between time points $t_{k-1} = (k-1)\Delta t$ and $t_k = k\Delta t$:

$$\boldsymbol{x}(t_k) = \boldsymbol{\Phi}(t_k, t_{k-1})\,\boldsymbol{x}(t_{k-1}) + \int_{t_{k-1}}^{t_k} \boldsymbol{\Phi}(t_k, \tau)\boldsymbol{B}(\tau)\boldsymbol{u}(\tau)\mathrm{d}\tau + \int_{t_{k-1}}^{t_k} \boldsymbol{\Phi}(t_k, \tau)\boldsymbol{G}(\tau)\boldsymbol{w}(\tau)\mathrm{d}\tau, \quad (1.164)$$

where $\boldsymbol{\Phi}(t_k, t_{k-1})$ is the state-transition matrix that satisfies

$$\frac{\mathrm{d}\boldsymbol{\Phi}(t_k, t_{k-1})}{\mathrm{d}t} = \boldsymbol{F}(t)\boldsymbol{\Phi}(t_k, t_{k-1}), \qquad \boldsymbol{\Phi}(t_{k-1}, t_{k-1}) = \mathbb{I}. \tag{1.165}$$

and fulfills the following properties for all $t_{k-1} \leq t_k \leq T$:

$$\boldsymbol{\Phi}(t_{k-1}, t_k) = \boldsymbol{\Phi}^{-1}(t_k, t_{k-1}), \tag{1.166}$$

$$\boldsymbol{\Phi}(T, t_{k-1}) = \boldsymbol{\Phi}(T, t_k)\boldsymbol{\Phi}(t_k, t_{k-1}). \tag{1.167}$$

To now find an expression for the transition matrix of the state, $\boldsymbol{\Phi}(t_k, t_{k-1})$, we apply the zero-order hold approximation. In other words, we assume a small enough time-step Δt during which, as depicted in Fig. 1.2, each deterministic continuous signal in Eq. (1.160) and Eq. (1.161) is constant within the time-step Δt:

$$\boldsymbol{F}(t) \approx \boldsymbol{F}(t_{k-1}), \quad \boldsymbol{B}(t) \approx \boldsymbol{B}(t_{k-1}), \quad \boldsymbol{G}(t) \approx \boldsymbol{G}(t_{k-1}), \quad \boldsymbol{H}(t) \approx \boldsymbol{H}(t_{k-1}), \quad \boldsymbol{u}(t) \approx \boldsymbol{u}(t_{k-1}),$$

with the only exception being the zero-mean Gaussian noise processes $\boldsymbol{w}(t)$ and $\boldsymbol{v}(t)$, as explained in Sec. 1.3.8. Then, the solution to Eq. (1.165) is simply

$$\boldsymbol{\Phi}(t_k, t_{k-1}) = \mathrm{e}^{\boldsymbol{F}(t_{k-1})(t_k - t_{k-1})} = \mathrm{e}^{\boldsymbol{F}(t_{k-1})\Delta t}, \tag{1.168}$$

and Eq. (1.164) becomes

$$\boldsymbol{x}(t_k) = \mathrm{e}^{\boldsymbol{F}(t_{k-1})\Delta t}\boldsymbol{x}(t_{k-1}) + \left[\int_{t_{k-1}}^{t_k} \mathrm{e}^{\boldsymbol{F}(t_{k-1})(t_k - \tau)}\boldsymbol{B}(\tau)\boldsymbol{u}(\tau)\mathrm{d}\tau\right] + \int_{t_{k-1}}^{t_k} \mathrm{e}^{\boldsymbol{F}(t_{k-1})(t_k - \tau)}\boldsymbol{G}(\tau)\boldsymbol{w}(\tau)\mathrm{d}\tau \quad (1.169)$$

$$= \mathrm{e}^{\boldsymbol{F}(t_{k-1})\Delta t}\boldsymbol{x}(t_{k-1}) + \left[\int_{t_{k-1}}^{t_k} \mathrm{e}^{\boldsymbol{F}(t_{k-1})(t_k - \tau)}\mathrm{d}\tau\right]\boldsymbol{B}(t_{k-1})\boldsymbol{u}(t_{k-1}) + \int_{t_{k-1}}^{t_k} \mathrm{e}^{\boldsymbol{F}(t_{k-1})(t_k - \tau)}\boldsymbol{G}(t_{k-1})\boldsymbol{w}(\tau)\mathrm{d}\tau. \quad (1.170)$$

If we now compare Eq. (1.170) with Eq. (1.162), we can identify:

$$\boldsymbol{x}_k = \boldsymbol{x}(t_k), \quad \boldsymbol{u}_{k-1} = \boldsymbol{u}(t_{k-1}), \quad \boldsymbol{B}_{k-1} = \left[\int_{t_{k-1}}^{t_k} e^{\boldsymbol{F}(t_{k-1})(t_k-\tau)} d\tau\right] \boldsymbol{B}(t_{k-1}), \tag{1.171}$$

$$\boldsymbol{A}_{k-1} = e^{\boldsymbol{F}(t_{k-1})\Delta t}, \quad \boldsymbol{G}_{k-1}\boldsymbol{q}_{k-1} = \int_{t_{k-1}}^{t_k} e^{\boldsymbol{F}(t_{k-1})(t_k-\tau)} \boldsymbol{G}(t_{k-1})\boldsymbol{w}(\tau) d\tau, \tag{1.172}$$

where we carefully maintain the order of matrix and vector operations, as they are not generally commutative. The last three terms can be further simplified by noting that the continuous equations are equivalent to their discrete counterparts only in the limit when the time-step Δt approaches zero (i.e., as $t_k \to t_{k-1}$). Consequently, we can assume the time-step Δt to be infinitesimally small and only keep the terms up to first-order in Δt:

$$\boldsymbol{B}_{k-1} = \left[\int_{t_{k-1}}^{t_k} e^{\boldsymbol{F}(t_{k-1})(t_k-\tau)} d\tau\right] \boldsymbol{B}(t_{k-1}) \approx {}^*\Delta t\, \boldsymbol{B}(t_{k-1}), \tag{1.173}$$

$$\boldsymbol{A}_{k-1} = e^{\boldsymbol{F}(t_{k-1})\Delta t} \approx \mathbb{I} + \boldsymbol{F}(t_{k-1})\Delta t, \tag{1.174}$$

$$\boldsymbol{G}_{k-1}\boldsymbol{q}_{k-1} = \int_{t_{k-1}}^{t_k} e^{\boldsymbol{F}(t_{k-1})(t_k-\tau)} \boldsymbol{G}(t_{k-1})\boldsymbol{w}(\tau) d\tau \approx {}^\dagger \boldsymbol{G}(t_{k-1})\int_{t_{k-1}}^{t_k} \boldsymbol{w}(\tau) d\tau. \tag{1.175}$$

From the last equation we can further identify that

$$\boldsymbol{G}_{k-1} = \boldsymbol{G}(t_{k-1})\Delta t, \quad and \quad \boldsymbol{q}_{k-1} = \frac{1}{\Delta t}\int_{t_{k-1}}^{t_k} \boldsymbol{w}(\tau) d\tau, \tag{1.176}$$

where the discrete white noise sample $\boldsymbol{q}_k$ comes from averaging the continuous white noise over the interval $[t_{k-1}, t_k)$, as explained in Eq. (1.155). Note that unlike for the state, control term and other functions in Eq. (1.164), we do not approximate the noise terms $\boldsymbol{w}(t)$ and $\boldsymbol{v}(t)$ using zero-order hold. In other words, we do not sample these continuous noises and approximate them with their value at the beginning of the interval, like in Fig. 1.2. The main reason behind this lays in the definition of continuous white Gaussian noise, which as discussed in Sec. 1.3.8, has a delta-correlated covariance. This simply means that the noise is completely uncorrelated from one time instant to another, but that it has infinite variance. If we try to approximate the process noise with the zero-order hold, i.e. $\boldsymbol{q}_{k-1} \approx \boldsymbol{w}(t_{k-1})$, and apply the

definition of the process covariance of Eq. (1.157):

$$\boldsymbol{Q}_k = \mathbb{E}\!\left[\boldsymbol{q}_k\,\boldsymbol{q}_k^{\mathrm{T}}\right] = \mathbb{E}\!\left[\boldsymbol{w}(t_k)\boldsymbol{w}^{\mathrm{T}}(t_k)\right] = \boldsymbol{Q}(t_k)\delta(0), \tag{1.177}$$

we obtain an unbounded discrete process noise covariance. This clearly illustrates why direct sampling of the noise, as required by the zero-order hold, is mathematically problematic. If instead we discretize the continuous process noise by averaging it over the infinitesimally small time-step Δt, as suggested by Eq. (1.176), we retrieve a mathematically sound expression relating $\boldsymbol{Q}_k$ and $\boldsymbol{Q}(t)$, which yields a well-behaved process:

$$\boldsymbol{Q}_k = \mathbb{E}\!\left[\boldsymbol{q}_k\,\boldsymbol{q}_k^{\mathrm{T}}\right] = \frac{1}{\Delta t^2}\int_{t_k}^{t_{k+1}}\!\!\int_{t_k}^{t_{k+1}} \mathbb{E}\!\left[\boldsymbol{w}(\tau)\boldsymbol{w}^{\mathrm{T}}(s)\right]\mathrm{d}\tau\,\mathrm{d}s = \frac{1}{\Delta t^2}\int_{t_k}^{t_{k+1}}\!\!\int_{t_k}^{t_{k+1}} \boldsymbol{Q}(\tau)\delta(\tau - s)\mathrm{d}\tau\,\mathrm{d}s$$

$$= \frac{1}{\Delta t^2}\int_{t_k}^{t_{k+1}} \boldsymbol{Q}(\tau)\mathrm{d}\tau \approx \frac{1}{\Delta t^2}\,\boldsymbol{Q}(t_k)\Delta t = \frac{\boldsymbol{Q}(t_k)}{\Delta t}, \tag{1.178}$$

where we have now applied zero-order hold to approximate $\boldsymbol{Q}(t) \approx \boldsymbol{Q}(t_k)$ inside the integral of $\boldsymbol{Q}(t)$ over an infinitesimal time-step Δt in the interval $[t_k, t_{k+1})$. It then follows that

$$\boldsymbol{G}_k\boldsymbol{Q}_k\boldsymbol{G}_k^{\mathrm{T}} = \Delta t\,\boldsymbol{G}(t_k)\boldsymbol{Q}(t_k)\boldsymbol{G}(t_k)^{\mathrm{T}}. \tag{1.179}$$

We can apply the same reasoning to the measurement process noise, since $\boldsymbol{v}(t)$ is a continuous Gaussian white noise. Namely, the discretized measurement noise is obtained by integrating $\boldsymbol{v}(t)$ over the interval Δt:

$$\boldsymbol{r}_k = \frac{1}{\Delta t}\int_{t_{k-1}}^{t_k} \boldsymbol{v}(\tau)\mathrm{d}\tau. \tag{1.180}$$

Then, since $\boldsymbol{y}(t)$ follows Eq. (1.161) and by applying the zero-hold assumption for the state $\boldsymbol{x}(t)$ and measurement matrix $\boldsymbol{H}(t)$, we get

$$\boldsymbol{y}_k := \frac{1}{\Delta t}\int_{t_k}^{t_{k+1}} \boldsymbol{y}(t)\mathrm{d}t = \frac{1}{\Delta t}\int_{t_k}^{t_{k+1}} \left(\boldsymbol{H}(t)\boldsymbol{x}(t) + \boldsymbol{v}(t)\right)\mathrm{d}t \tag{1.181}$$

$$\approx \boldsymbol{H}(t_k)\boldsymbol{x}(t_k) + \frac{1}{\Delta t}\int_{t_k}^{t_{k+1}} \boldsymbol{v}(t)\mathrm{d}t, \tag{1.182}$$

> *where the relation between the discrete and continuous measurement covariance, $\boldsymbol{R}_k$ and $\boldsymbol{R}(t_k)$, is:*
>
> $$\boldsymbol{R}_k = \mathbb{E}\left[\boldsymbol{r}_k \boldsymbol{r}_k^{\mathrm{T}}\right] = \frac{1}{\Delta t^2}\int_{t_k}^{t_{k+1}}\int_{t_k}^{t_{k+1}} \mathbb{E}\left[\boldsymbol{v}(t)\boldsymbol{v}^{\mathrm{T}}(\tau)\right]\,\mathrm{d}t\,\mathrm{d}\tau = \frac{1}{\Delta t^2}\int_{t_k}^{t_{k+1}} \boldsymbol{R}(t)\mathrm{d}t = \frac{\boldsymbol{R}(t_k)}{\Delta t}. \quad (1.183)$$
>
> ---
>
> [a] $\int_0^{\Delta t} f(t)\mathrm{d}t \approx f(0)\Delta t,$ when $\Delta t \to 0$
> [b] $\int_0^{\Delta t} f(t)w(t)\mathrm{d}t \approx f(0)\int_0^{\Delta t} w(t)\mathrm{d}t,$ when $\Delta t \to 0$

$\boldsymbol{x}_k = \boldsymbol{x}(t_k)$	$\boldsymbol{u}_k = \boldsymbol{u}(t_k)$
$\boldsymbol{A}_k = \mathbb{I} + \boldsymbol{F}(t_k)\Delta t$	$\boldsymbol{B}_k = \boldsymbol{B}(t_k)\Delta t$
$\boldsymbol{H}_k = \boldsymbol{H}(t_k)$	$\boldsymbol{G}_k = \boldsymbol{G}(t_k)\Delta t$
$\boldsymbol{r}_k = \frac{1}{\Delta t}\int_{t_{k-1}}^{t_k} \boldsymbol{v}(\tau)\mathrm{d}\tau$	$\boldsymbol{q}_k = \frac{1}{\Delta t}\int_{t_{k-1}}^{t_k} \boldsymbol{w}(\tau)\mathrm{d}\tau$
$\boldsymbol{R}_k = \frac{\boldsymbol{R}(t_k)}{\Delta t}$	$\boldsymbol{Q}_k = \frac{\boldsymbol{Q}(t_k)}{\Delta t}$

Table 1.1: Table summarizing how to discretize a continuous linear and Gaussian system. This table details how each vector, white noise component, and matrix defining the continuous LG system of Eqs. (1.160-1.161) is transformed to obtain the corresponding elements of the discrete LG system given in Eqs. (1.162-1.163).

1.4 (SOME) FUNDAMENTALS OF QUANTUM MECHANICS

1.4.1 POSITION AND MOMENTUM OPERATORS

The position and momentum operators, $\hat{X}$ and $\hat{P}$, can be written in terms of creation and anihilation operators as

$$\hat{X} = \sqrt{\frac{1}{2}}\left(\hat{a}^\dagger + \hat{a}\right), \qquad \hat{P} = i\sqrt{\frac{1}{2}}\left(\hat{a}^\dagger - \hat{a}\right), \qquad (1.184)$$

These operators satisfy the canonical commutation relation $[\hat{X}, \hat{P}] = i$, which follows directly from the algebra $[\hat{a}, \hat{a}^\dagger] = 1$.

The ground-state wavefunction of the quantum harmonic oscillator in the position eigenbasis is a Gaussian function:

$$\psi_0(x) = \langle x|0\rangle = \frac{1}{\pi^{1/4}} e^{-x^2/2}, \qquad (1.185)$$

which has been normalized, as required from a wavefunction.

> **Proof.** *From Eq. (1.184), we can write the annihilation and creation operators w.r.t. the position and momentum operators as*
>
> $$\hat{a} = \frac{1}{\sqrt{2}}(\hat{X} + i\hat{P}), \quad and \quad \hat{a}^\dagger = \frac{1}{\sqrt{2}}(\hat{X} - i\hat{P}). \qquad (1.186)$$
>
> *As standard, applying $\hat{a}$ to the ground state yields zero,*
>
> $$\hat{a}\,|0\rangle = 0. \qquad (1.187)$$
>
> *Thus, applying $\langle x|$ to the equation above, where $\langle x|$ is an eigenstate of $\hat{X}$, will also yield zero:*
>
> $$\langle x|\hat{a}|0\rangle = \frac{1}{\sqrt{2}}\langle x|(\hat{X} + i\hat{P})|0\rangle = 0. \qquad (1.188)$$
>
> *By using that the $\hat{P}$ operator represented in the position basis can be written as*
>
> $$\langle x|\hat{P}|0\rangle = -i\frac{d}{dx}\langle x|0\rangle, \qquad (1.189)$$
>
> *we find a first-order linear differential equation for the wavefunction $\psi_0(x) = \langle x|0\rangle$:*
>
> $$\langle x|\hat{X}|0\rangle + i\langle x|\hat{P}|0\rangle = x\langle x|0\rangle + \frac{d}{dx}\langle x|0\rangle = 0, \qquad (1.190)$$
>
> *which has a solution*
>
> $$\langle x|0\rangle = Ae^{-x^2/2}, \qquad (1.191)$$

with a normalization constant $A = 1/\pi^{1/4}$, since

$$1 = \int_{-\infty}^{\infty} |\langle x|0\rangle|^2 \, \mathrm{d}x = \int_{-\infty}^{\infty} |A|^2 \mathrm{e}^{-x^2} \, \mathrm{d}x = |A|^2 \sqrt{\pi}. \tag{1.192}$$

Therefore, the wavefunction of the ground state in the position eigenbasis reads as:

$$\langle x|0\rangle = \frac{1}{\pi^{1/4}} \mathrm{e}^{-x^2/2}. \tag{1.193}$$

1.4.2 Dynamics of open quantum systems

1.4.2.1 The Gorini–Kossakowski–Sudarshan–Lindblad generator

A quantum system interacting with a Markovian environment is governed by a master equation of the form

$$\frac{d\rho}{\mathrm{d}t} = \mathcal{L}\rho, \tag{1.194}$$

where the Gorini–Kossakowski–Sudarshan–Lindblad (GKSL) generator of the evolution [73, 74] reads as:

$$\mathcal{L}\rho = -i[\hat{H},\rho] + \sum_{k=1}^{K} \mathcal{D}[\hat{L}_k]\rho. \tag{1.195}$$

Here, $\hat{H}$ is a Hermitian operator representing the Hamiltonian of the system and $\{\hat{L}_j\}$ are a collection of operators, often referred to as Lindblad operators, that characterize the various irreversible processes via the (dissipative) superoperator:

$$\mathcal{D}[\hat{O}]\rho = \hat{O}\rho\hat{O}^\dagger - \frac{1}{2}(\hat{O}^\dagger\hat{O}\rho + \rho\hat{O}^\dagger\hat{O}). \tag{1.196}$$

The formal solution for a general master equation (1.194) with a time-dependent Lindblad form can be written as

$$\rho(t) = \mathcal{N}_t\rho(0) \tag{1.197}$$

where the superoperator $\mathcal{N}_t$ is defined as

$$\mathcal{N}_t = \mathcal{T}\left\{\mathrm{e}^{\int_0^t \mathcal{L}_\tau d\tau}\right\}, \tag{1.198}$$

with $\mathcal{T}\left\{\exp\left\{\cdot\right\}\right\}$ denoting the time-ordered exponential. In the special case when $\mathcal{L}$ is time-independent, the solution is a semigroup:

$$\mathcal{N}_t = e^{\mathcal{L}t}. \tag{1.199}$$

1.4.3 ANGULAR MOMENTUM

The three *total* angular momentum operators $\hat{J}_x$, $\hat{J}_y$ and $\hat{J}_z$ obey the commutation relation

$$[\hat{J}_j, \hat{J}_k] = i\varepsilon_{jkl}\hat{J}_l \tag{1.200}$$

where ε_{jkl} is the Levi-Civita symbol and the indeces $i, j, k = x, y, z$. In other words,

$$[\hat{J}_x, \hat{J}_y] = i\hat{J}_z \tag{1.201}$$
$$[\hat{J}_y, \hat{J}_z] = i\hat{J}_x \tag{1.202}$$
$$[\hat{J}_z, \hat{J}_x] = i\hat{J}_y. \tag{1.203}$$

The collection of these angular momentum operators form a total angular momentum *vector* operator:

$$\hat{\boldsymbol{J}} = (\hat{J}_x, \hat{J}_y, \hat{J}_z)^{\mathrm{T}}, \tag{1.204}$$

whose squared magnitude defines another operator:

$$\hat{J}^2 := \hat{J}_x^2 + \hat{J}_y^2 + \hat{J}_z^2. \tag{1.205}$$

The operators $\hat{J}^2$ and $\hat{J}_z$ have a common eigenbasis with eigenvectors labeled with quantum numbers j and m:

$$\hat{J}^2 \ket{j, m} = j(j+1) \ket{j, m},$$
$$\hat{J}_z \ket{j, m} = m \ket{j, m},$$

where $j \in \{0, \frac{1}{2}, 1, \frac{3}{2}, 2, \dots\}$, and for a given j, the value of m ranges as $m = -j, -j + 1, \dots, j - 1, j$, i.e. from $-j$ to j in steps of unity [75]. Using the operators $\hat{J}_x$ and $\hat{J}_y$, we define the

47

ladder operators for the angular momentum operators as:

$$\hat{J}_{\pm} = \hat{J}_x \pm i\hat{J}_y, \tag{1.206}$$

which fulfill $\hat{J}_+ = \hat{J}_-^\dagger$. These operators act on the eigenstates $|j, m\rangle$ by raising or lowering the eigenvalue m by one unit:

$$\hat{J}_{\pm} |j, m\rangle = \sqrt{j(j+1) - m(m \pm 1)} \, |j, m \pm 1\rangle . \tag{1.207}$$

If $\hat{\boldsymbol{J}}$ is the angular momentum vector of an ensemble of N two-level systems, then $j \leq N/2$ and $-j \leq m \leq j$, with j and m integer or half and $j_{\min} = 0, 1/2$ depending whether there is an even or odd number of two-level systems, i.e. whether N is even or odd [76]. Then, the collective angular momentum operators are defined as the sum over the individual contributions:

$$\hat{J}_x = \frac{1}{2} \sum_{k=1}^{N} \hat{\sigma}_x^{(k)}, \quad \hat{J}_y = \frac{1}{2} \sum_{k=1}^{N} \hat{\sigma}_y^{(k)}, \quad \hat{J}_z = \frac{1}{2} \sum_{k=1}^{N} \hat{\sigma}_z^{(k)}, \tag{1.208}$$

where the angular momentum operators for each individual two-level system are $\hat{s}_x^{(k)} = \frac{1}{2}\hat{\sigma}_x^{(k)}$, $\hat{s}_y^{(k)} = \frac{1}{2}\hat{\sigma}_y^{(k)}$ and $\hat{s}_z^{(k)} = \frac{1}{2}\hat{\sigma}_z^{(k)}$. The collective operators obey the same commutation relations and eigenvalue equations as those for a single system while encapsulating the macroscopic spin properties of the entire ensemble.

1.4.4   Coherent spin state

As might be deduced from the name, coherent spin states [77, 78] (also known as atomic coherent states) are simply an extension of the coherent states of a field first introduced by Glauber. A coherent spin state (CSS) is a product state, i.e. the tensor product of N qubits, e.g. spin-1/2, all aligned in the same direction $\boldsymbol{k} = (\sin \alpha \cos \beta, \sin \alpha \sin \beta, \cos \alpha)$ (see Fig. 1.3):

$$|\alpha, \beta\rangle = \bigotimes_{k=1}^{N} \left[\cos \frac{\alpha}{2} |0\rangle_k + e^{i\beta} \sin \frac{\alpha}{2} |1\rangle_k \right] . \tag{1.209}$$

As discussed in detail in Ref. [78], a CSS can also be written in terms of the angular momentum basis by rotating the ground state $|j, -j\rangle$ by an angle α about an axis $\boldsymbol{n} = (\sin \beta, -\cos \beta, 0)$ with

CSS convention [78, 79]
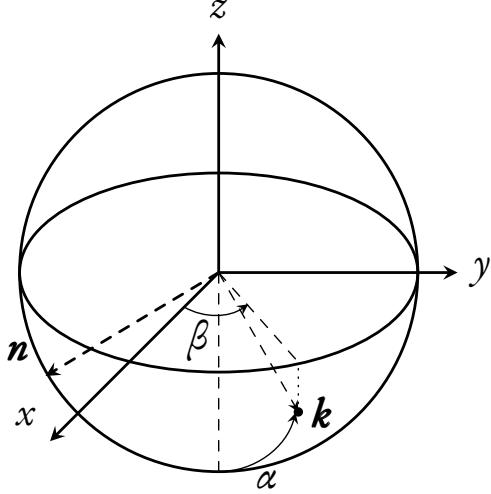
Standard (physics) convention [80, 12]
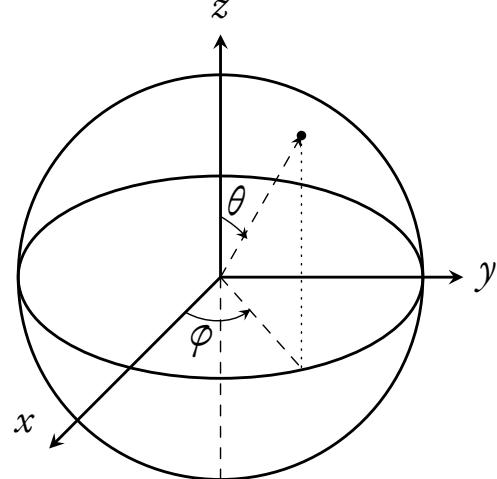
Figure 1.3: Spherical coordinates conventions. The choice of rotation angles β and α (see left sketch) define the rotation operator $R(\alpha, \beta)$ and thus, the axis along which the CSS is aligned (along $\boldsymbol{k}$). Given that the rotation operator around axis $\boldsymbol{n}$ is applied to the ground state of the ensemble $|j, -j\rangle$, by convention centered around the south pole, the rotation angle α is measured off the south pole [78, 79]. Thus, the standard spherical coordinates parametrizing a sphere, (θ, φ) (see right), relate to (α, β) as $\theta = \pi - \alpha$ and $\varphi = \beta$ [79]. When we discuss Wigner functions of CSS mapped onto the Bloch sphere, it is important to note that (α, β) simply define the direction of the CSS, whereas (θ, φ) map the Wigner function onto the 3D sphere.

the operator

$$R(\alpha, \beta) = e^{-i\alpha \left(\hat{J}_x \sin\beta - \hat{J}_y \cos\beta \right)} = e^{\xi \hat{J}_+ - \xi^* \hat{J}_-}, \tag{1.210}$$

where $\xi = \frac{1}{2}\alpha e^{-i\beta}$. Namely,

$$|\alpha, \beta\rangle := R(\alpha, \beta) |j, -j\rangle. \tag{1.211}$$

In Ref. [78] they further show how to rewrite Eq. (1.210) as

$$R(\alpha, \beta) = R(\eta) = e^{\eta \hat{J}_+} e^{\ln \left(1 + |\eta|^2 \right) \hat{J}_z} e^{-\eta^* \hat{J}_-}, \tag{1.212}$$

where

$$\eta = e^{-i\beta} \tan\frac{\alpha}{2}. \tag{1.213}$$

By now applying this form of the rotation operator to the ground state $|j, -j\rangle$, as indicated in

49

Eq. (1.211), we get

$$|\eta\rangle := R(\eta)\,|j,-j\rangle = \left(-\frac{1}{1+|\eta|^2}\right)^j e^{\eta \hat{J}_+}\,|j,-j\rangle, \qquad (1.214)$$

since

$$e^{-\eta^* \hat{J}_-}\,|j,-j\rangle = \mathbb{I}\,|j,-j\rangle \qquad (1.215)$$

$$e^{\ln\left(1+|\eta|^2\right)\hat{J}_z}\,|j,-j\rangle = e^{\ln\left(1+|\eta|^2\right)(-j)}\,|j,-j\rangle = e^{\ln\left(1+|\eta|^2\right)^{-j}}\,|j,-j\rangle$$

$$= \left(\frac{1}{1+|\eta|^2}\right)^j |j,-j\rangle. \qquad (1.216)$$

Next, we expand the exponent in Eq. (1.214) as

$$|\eta\rangle := R(\eta)\,|j,-j\rangle = \left(\frac{1}{1+|\eta|^2}\right)^j \sum_{k=0}^{2j} \frac{\eta^k}{k!} \hat{J}_+^k\,|j,-j\rangle, \qquad (1.217)$$

where the power series expansion of $e^{\eta \hat{J}_+}$ terminates at $k = 2j$, because $\hat{J}_+^k\,|j,j\rangle = 0$ for $k > 0$ and thus, $\hat{J}_+^k\,|j,-j\rangle = 0$ for $k > 2j$. By shifting k by j and redefining it as $k = m+j$, the summation becomes

$$|\eta\rangle = \left(\frac{1}{1+|\eta|^2}\right)^j \sum_{m=-j}^{j} \frac{\eta^{m+j}}{(m+j)!} \hat{J}_+^{m+j}\,|j,-j\rangle. \qquad (1.218)$$

By then applying the relation derived in Proposition A.4:

$$\binom{2j}{m+j}^{1/2} |j,m\rangle = \frac{1}{(m+j)!} \hat{J}_+^{m+j}\,|j,-j\rangle, \qquad (1.219)$$

we get the form of a CSS in the eigenbasis of the angular momentum

$$|\eta\rangle = (1+|\eta|^2)^{-j} \sum_{m=-j}^{j} \binom{2j}{m+j}^{1/2} \eta^{m+j}\,|j,m\rangle, \quad \eta \in \mathbb{C}, \qquad (1.220)$$

where $\eta = \tan\frac{\alpha}{2}e^{-i\beta}$ as given in Eq. (1.213). Assuming the CSS of the spin-1/2 ensemble points

in the *x*-direction s.t. the angles α and β have values:

$$j = \frac{N}{2}, \quad \alpha = \frac{\pi}{2}, \quad \text{and} \quad \beta = 0, \tag{1.221}$$

then, the value of η is 1 and the CSS, aligned along *x*, reads as:

$$|\eta\rangle = \frac{1}{2^{N/2}} \sum_{m=-N/2}^{N/2} \left(\frac{N}{\frac{N}{2} + m} \right)^{1/2} \left| \frac{N}{2}, m \right\rangle. \tag{1.222}$$

A CSS of that form has the following mean and variance:

$$\langle \hat{\boldsymbol{J}} \rangle_{\mathrm{CSS}_x} = \mathrm{Tr}\left\{ \rho_{\mathrm{CSS}_x} \hat{\boldsymbol{J}} \right\} = \left(\langle \hat{J}_x \rangle_{\mathrm{CSS}_x}, \ \langle \hat{J}_y \rangle_{\mathrm{CSS}_x}, \ \langle \hat{J}_z \rangle_{\mathrm{CSS}_x} \right)^{\mathrm{T}} = \left(\frac{N}{2}, \ 0, \ 0 \right)^{\mathrm{T}} \tag{1.223}$$

$$\langle \Delta^2 \hat{\boldsymbol{J}} \rangle_{\mathrm{CSS}_x} = \left(\Delta^2 \langle \hat{J}_x \rangle_{\mathrm{CSS}_x}, \ \Delta^2 \langle \hat{J}_y \rangle_{\mathrm{CSS}_x}, \ \Delta^2 \langle \hat{J}_z \rangle_{\mathrm{CSS}_x} \right)^{\mathrm{T}} = \left(0, \ \frac{N}{4}, \ \frac{N}{4} \right)^{\mathrm{T}} \tag{1.224}$$

with $\langle \Delta^2 \hat{J}_i \rangle = \langle \hat{J}_i^2 \rangle - \langle \hat{J}_i \rangle^2$, where $i = x$, y, or z. The values for the means and variances are calculated in Sec. A.4.1.

1.4.5 THE WIGNER QUASIPROBABILITY DISTRIBUTION

Definition 1.22 (Wigner distribution). *Let ρ be a mixed state, and x and p a pair of conjugate variables representing position and momentum. Then, the Wigner distribution is defined as:*

$$W_\rho(x, p) := \frac{1}{\pi} \int_{-\infty}^{\infty} \langle x - y | \rho | x + y \rangle e^{2ipy} \mathrm{d}y, \tag{1.225}$$

where we assume $\hbar = 1$. Equivalently, for a pure state $|\psi\rangle$, it can be written as:

$$W_\psi(x, p) := \frac{1}{\pi} \int_{-\infty}^{\infty} \psi^*(x + y) \psi(x - y) e^{2ipy} \mathrm{d}y. \tag{1.226}$$

The Wigner function fulfills the following properties:

1. $W_\rho(x, p)$ is a $\mathbb{R}$-valued function.

2. It may take on negative values. For continuous variables, this is often interpreted as a sign of nonclassical behavior [12].

3. It provides proper marginal distributions:

$$\langle x|\rho|x\rangle = \int dp\, W_\rho(x,p),\tag{1.227}$$

$$\langle p|\rho|p\rangle = \int dx\, W_\rho(x,p).\tag{1.228}$$

4. The state overlap of two pure states $|\psi\rangle$ and $|\varphi\rangle$ is calculated as

$$|\langle\psi|\varphi\rangle|^2 = 2\pi \int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} dp\, W_\psi(x,p) W_\varphi(x,p)\tag{1.229}$$

5. Operator averages are calculated as

$$\langle\hat{O}\rangle = \mathrm{Tr}\left\{\rho\,\hat{O}\right\} = \int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} dp\, W_\rho(x,p)\, f_{\hat{O}}(x,p),\tag{1.230}$$

where

$$f_{\hat{O}}(x,p) := \int_{-\infty}^{\infty} dy \left\langle x - \frac{y}{2}\left|\hat{O}\right|x - \frac{y}{2}\right\rangle e^{ipy}.\tag{1.231}$$

Example. (Wigner distribution of the vacuum state): *Let us explicitly derive the Wigner quasiprobability distribution for the vacuum state $|0\rangle$. Its wavefunction in the position eigenbasis, $\psi_0(x)$, is:*

$$\langle x|0\rangle = \psi_0(x) = \frac{1}{\pi^{1/4}} e^{-x^2/2},\tag{1.232}$$

as derived in Sec. 1.4.1.1. Therefore, its Wigner function can be written as:

$$W_0(x,p) = \frac{1}{\pi}\int_{-\infty}^{\infty} dy\, e^{2ipy}\frac{1}{\sqrt{\pi}}e^{-(x-y)^2/2}e^{-(x+y)^2/2}.\tag{1.233}$$

The sum of the exponents of the wavefunctions yields:

$$-\frac{(x-y)^2}{2} - \frac{(x+y)^2}{2} = -x^2 - y^2.\tag{1.234}$$

Therefore,

$$W_0(x,p) = \frac{1}{\pi^{3/2}} e^{-x^2} \int_{-\infty}^{\infty} dy\ e^{-y^2} e^{2ipy}. \tag{1.235}$$

Next we can evaluate the integral above by noting it is a standard Fourier transform of a Gaussian:

$$\int_{-\infty}^{\infty} dy\ e^{-y^2} e^{2ipy} = \mathcal{F}_y\left[e^{-y^2}\right](p) = \sqrt{\pi} e^{-p^2}. \tag{1.236}$$

Thus, the Wigner function of the vacuum state is

$$W_0(x,p) = \frac{1}{\pi} e^{x^2 + p^2}. \tag{1.237}$$

Furthermore, note that the Wigner distribution for the vacuum state $W_0(x,p)$ is invariant under rotations in the phase space, since the variables x and p define a radial distance $r^2 = x^2 + p^2$. Therefore, if we define new operators $\hat{X}_\varphi$ and $\hat{P}_\varphi$ rotated in phase space by φ through the following transformations:

$$\hat{X}_\varphi = \hat{X} \cos\varphi + \hat{P} \sin\varphi, \tag{1.238}$$

$$\hat{P}_\varphi = -\hat{X} \sin\varphi + \hat{P} \cos\varphi, \tag{1.239}$$

we can show that the wavefunction of the rotated eigenket $|x_\varphi\rangle$ is simply

$$\langle x_\varphi | 0 \rangle = \psi_0(x_\varphi) = \frac{1}{\pi^{1/4}} e^{-x_\varphi^2/2}, \tag{1.240}$$

since

$$|\langle x_\varphi | 0 \rangle|^2 = \int dp_\varphi W_0(x_\varphi, p_\varphi) = \frac{1}{\sqrt{\pi}} e^{-x_\varphi^2}, \tag{1.241}$$

because $x^2 + p^2 = x_\varphi^2 + p_\varphi^2$, i.e. the Wigner function of the vacuum state remains invariant under rotations in the phase space.

1.4.6   The Wigner function on a sphere

Computing the Wigner distribution and mapping it into the Bloch sphere is very useful for visualizing quantum states and how different operations affect them, specially in the case of atomic ensembles [12, 79, 80]. The Wigner quasiprobability distribution, mapped onto the generalized

Bloch sphere, is parametrized using the standard spherical convention with angles θ and φ [12] (see right in Fig. 1.3):

$$\mathrm{W}_\rho(\theta, \varphi) = \sqrt{\frac{N+1}{4\pi}} \sum_{k=0}^{N} \sum_{q=-k}^{k} \rho_{kq} Y_k^q(\theta, \varphi) \tag{1.242}$$

where $Y_k^q(\theta, \varphi)$ are the complex spherical harmonics [81]. These functions are defined on the surface of a unit sphere, mapping $Y_k^q : S^2 \to \mathbb{C}$, where S^2 is the 2D surface of the sphere. Spherical harmonics provide a complete, orthonormal basis for square-integrable functions defined on S^2. As a result, each function on the surface of a sphere can be written as a weighted sum of these spherical harmonics. In a way, spherical harmonics generalize the Fourier series from periodic functions on a circle (S^1) to functions on a sphere (S^2). While the Fourier series decomposes a periodic function into a sum of sines and cosines, spherical harmonics extend this concepts to two dimensions by employing both azimuthal (φ) and polar (θ) angular dependencies. Here, θ and φ follow the standard convention, with θ being the polar angle measured off the $+z$-axis [79, 80] (see right in Fig. 1.3).

The coefficients ρ_{kq} of the spherical harmonic decomposition of the Wigner function,

$$\rho_{kq} = \sum_{m_1,m_2=-J}^{J} \rho_{m_1,m_2} t_{kq}^{m_1 m_2}, \tag{1.243}$$

are determined by the part of the density matrix supported by the *totally symmetric subspace*, in particular, its elements $\rho_{m_1,m_2} := \langle J, m_1 | \rho | J, m_2 \rangle$, written in the angular momentum basis for the maximal total spin $J = N/2$, as well as the coefficients $t_{kq}^{m_1 m_2}$ dictating the transformation from the Dicke space to the k-space [80, 82]:

$$t_{kq}^{m_1 m_2} := (-1)^{J-m_1-q} \langle J, m_1; J, -m_2 | k, q \rangle, \tag{1.244}$$

where $\langle J, m_1; J, -m_2 | k, q \rangle$ are the Clebsch-Gordan coefficients. Note that the exact density matrix is needed to generate the Wigner quasiprobability distribution.

1.4.7 SPIN SQUEEZING

Spin-squeezed states are defined as states in which the variance of one collective spin component is reduced below that of a CSS, at the expense of increased variance along an orthogonal direction.

Consider an ensemble of N two-level atoms described by collective spin operators. The Wineland spin-squeezing parameter in Ref. [5, 6] is defined as

$$\xi^2(t) := \frac{V_\perp(t)}{\langle \hat{J}_s(t) \rangle^2} \left(\frac{V_\perp^{\text{CSS}}}{\langle \hat{J}_s \rangle_{\text{CSS}}^2} \right)^{-1} = \frac{\langle \hat{J}_s(t) \rangle^2}{N V_\perp(t)}, \qquad (1.245)$$

where s is the mean spin direction, and $\perp$ is the perpendicular direction along the ensemble is being squeezed. For an ideal CSS where all atoms are uncorrelated, $\xi^2(t) = 1$. Hence, when $\xi^2(t) < 1$, the state is (metrologically) spin squeezed. It is important to note that the definition of spin-squeezing is not unique. Another widely used definition, introduced by Kitagawa and Ueda [83], is inspired by photon squeezing and focuses on the minimum variance of a spin component orthogonal to the mean spin direction:

$$\xi_S^2 = \frac{4 \min_\perp (\Delta \hat{J}_\perp)^2}{N}, \qquad (1.246)$$

where just like before, $\xi_S^2 < 1$ defines the spin-squeezing condition. This parameter quantifies the reduction in quantum noise along a particular spin direction and is directly related to the metrological spin-squeezing parameter through the following inequality:

$$\xi_S^2 \leq \xi^2. \qquad (1.247)$$

In other words, metrological spin squeezing ($\xi^2 < 1$) implies spin squeezing in the sense of Kitagawa and Ueda ($\xi_S^2 < 1$), but the converse is not necessarily true. That is, while a state with reduced spin variance ($\xi_S^2 < 1$) may exhibit quantum correlations, it does not automatically guarantee an improvement in metrological sensitivity.

2

Bayesian estimation and control

BAYESIAN INFERENCE is a statistical framework in which parameters are treated as random variables [2, 29]. In this approach, probabilities quantify our uncertainty or beliefs about these parameters or hypothesis, and are systematically updated as new data becomes available [2, 29, 30]. That contrasts with the frequentist approach, where parameters are constant and probabilities are interpreted as the proportion of outcomes in a large number of repeated identical experiments [1, 2].

How to update our knowledge by incorporating new observations is described by the Bayes' theorem[84, 30]. It combines prior information, represented by a *prior* distribution, with the statistical model of the observations, given by the *likelihood* distribution, to produce an updated probability known as the *posterior* distribution:

$$p(\theta|y) \propto p(\theta|y)p(\theta) \quad \longleftrightarrow \quad \text{posterior} \propto \text{likelihood} \cdot \text{prior}, \quad (2.1)$$

where θ is the parameter, and y represents the observed data.

Bayesian inference is especially well suited for problems such as optimal filtering, where the

parameters of the system or state evolve over time, requiring continuous updates as new observations are performed [30]. Filtering involves estimating the time-varying state of a system using noisy observations when the true state is not directly observable. This estimation process is typically implemented in two stages: a prediction step, which uses a model of the dynamics of the system to propagate in time the state distribution, and an update step, which refines this estimation based on the latest observation [29, 30]. This recursive process ensures that the state estimate improves as more data becomes available. The Kalman filter (KF) is a well-known example, providing an optimal solution when the system is linear and the noise is Gaussian [31, 32, 1, 29, 30]. For nonlinear or non-Gaussian scenarios, extensions like the extended Kalman filter (EKF) are used, which linearize the system dynamics around the current estimate to handle mild nonlinearities[33, 29]. In more challenging situations, which are beyond the scope of this thesis, more general methods such as particle filtering can be employed, using a set of particles to approximate the posterior distribution [30].

Besides estimating parameters in real time, we are often also interested in how to optimally steer the system into a state more advantageous for metrology, e.g. more sensitive to the parameter of interest or more robust to noise. To define what we mean by optimal control, we must first specify a control cost, i.e. a function that balances two competing objectives: (1) achieving the desired control task with high accuracy, and (2) minimizing the cost or effort required by the actuator [85]. This cost function typically includes adjustable parameters, or "knobs", that allow us to trade off estimation performance against control effort[29, 85, 55]. The optimal control law is then the one that minimizes this total cost. For linear and Gaussian (LG) systems, the optimal control is given by the linear-quadratic Gaussian (LQG) controller, which combines a KF with a linear-quadratic regulator (LQR) [29, 85]. A complete and rigorous proof on why LQG is optimal requires knowledge of dynamical programming and the broader field of optimal control. To avoid delving into that but still provide convincing arguments, we present simpler — albeit less rigorous — proofs that show how both LQR and LQG minimize their respective control costs[29, 85].

The results and derivations presented in this chapter are not original contributions. Rather, they are re-derivations and explanations in my own words, inspired by and adapted from well-established treatments in the literature, particularly: Särkkä [30], Crassidis and Junkins [29], Simon [33], and Kolosov [85]. Much of the material on Bayesian filtering and the KF is drawn from Särkkä [30] and Crassidis and Junkins [29]; the EKF section closely follows the treatment in Simon [33]; and the discussion on control, including the LQR and LQG controller, is based

primarily on Crassidis and Junkins [29] and Kolosov [85].

THIS CHAPTER is organized as follows: we begin in Sec. 2.1.1 with a brief overview of Bayesian inference, introducing Bayes' rule and common estimators such as the maximum likelihood estimator, the maximum a-priori (MAP) estimator, and the minimum mean squared error (MMSE) estimator. In Sec. 2.2, we define the Fisher information and introduce the Bayesian Cramér-Rao bound (BCRB), a lower bound on the average mean squared error (aMSE). We then move to Bayesian filtering in Sec. 2.3, where we highlight the importance of modeling both the system dynamics and the measurement process in Sec. 2.3.1. These models define the probability distributions used to track the state of the system over time, and they form the foundation of the two key steps in any Bayesian filter: the prediction and measurement update, which are detailed in Sec. 2.3.2.

In Sec. 2.4, we derive the discrete-time KF under the assumption of uncorrelated system and measurement noise. This is the simplest form of the KF and a particularly clear example of how Bayesian filtering works when the system is linear and all noise is Gaussian. It provides closed-form equations for the estimate and its covariance, which correspond to the mean and variance of the (Gaussian) posterior distributions. This eliminates the need to track full probability distributions explicitly since updating the mean and covariance is sufficient. Moreover, because the mean of a Gaussian posterior is the MMSE estimator, the KF yields the optimal estimate in this setting.

In Sec. 2.5, we extend the KF to the continuous-time setting. To do so, we had address already in Sec. 1.3.9 how to discretize LG systems with white Gaussian noise, which is not a conventional Gaussian random variable but the formal derivative of a Wiener process. Properly discretizing this noise to avoid pathological behavior requires integrating it over a finite time step, which yields a Wiener increment. With this clear, we then take the continuous-time limit in Sec. 2.5 to arrive at the continuous-time KF. The same steps are later followed in Sec. 2.6, where we extend the KF to handle correlated process and measurement noise. This is the form necessary for continuously monitored atomic magnetometers, where correlations between the system and the measurement naturally arise due to measurement back-action. Since the systems we are interested in are often nonlinear, we derive the EKF in Sec. 2.7 by linearizing the dynamics around the current estimate. Finally, in Sec. 2.8, we turn to optimal control. In Sec. 2.8.1, we study the LQR for systems where the state is directly accessible. We then combine this with state estimation using the KF to derive the LQG controller in Sec. 2.8.2, which handles noisy measurements and provides optimal

feedback control.

2.1 Introduction to Bayesian inference

2.1.1 Bayesian statistics

Suppose we want to use Bayesian inference to estimate some quantity of interest, θ, given a measurement y. Any Bayesian method has three key steps:

1. **Modelling.** Model what we know about θ even before making any measurements (using a prior $p(\theta)$), as well as how the measurements y relate to θ (using a probability density $p(y|\theta)$, also known as a likelihood function).

2. **Measurement update.** Combine what we know before (the prior) with our measurement (i.e., with the likelihood $p(y|\theta)$) in order to get a function of θ, i.e. the posterior distribution $p(\theta|y)$, which takes into account our updated knowledge on θ after performing a measurement y.

3. **Decision making.** We can now combine what we know about θ, i.e., the posterior $p(\theta|y)$, with a cost function, in order to perform an *optimal decision*.

2.1.2 Prior and likelihood

As hinted at in the previous section, Bayesian methods rely on three important components: the prior, the likelihood and the posterior. For now, let us define and discuss the role of the prior and the likelihood when analyzing an unknown parameter $\theta \in \Theta$ based on some observed data y. The posterior, which requires introducing Bayes' rule, will be discussed in the subsequent section.

Definition 2.1 (The likelihood function)**.** *The first key assumption in any Bayesian model is that the observed data, y, given that the parameter is θ, follows the distribution*

$$y \sim p(y|\theta), \tag{2.2}$$

which is referred to as the likelihood. The model given by the likelihood $p(y|\theta)$ describes how we expect our observations y to behave when we know that the true value of the parameter of interest is θ. This distribution is often easier to compute than the probability of θ conditioned w.r.t. y. Since y

is observed, $p(y|\theta)$ can be viewed as a function of θ,

$$I(\theta|y) = p(y|\theta), \tag{2.3}$$

where $I(\theta|y)$ is also called the likelihood function. Note that the likelihood is not a PDF w.r.t. θ, but w.r.t. y. In other words, if we integrate $p(y|\theta)$ w.r.t. $\theta \in (-\infty, \infty)$, it does not yield one, unlike when integrating w.r.t. $y \in (-\infty, \infty)$.

Definition 2.2 (Prior). *Another fundamental component of many Bayesian methods is the prior distribution, where the term "prior" simply means earlier or before:*

$$p(\theta). \tag{2.4}$$

This probability distribution expresses our a-priori knowledge concerning the parameter θ, i.e., the knowledge held before any data y is taken into account.

2.1.3 BAYES' RULE AND THE POSTERIOR

The Bayes' rule is the cornerstone of any Bayesian method because it enables us to compute the posterior $p(\theta|y)$ from the likelihood $p(y|\theta)$ and $p(\theta)$. It straightforwardly follows from Property 1.2 and Property 1.1. Specifically, the product rule implies the following equivalence:

$$p(\theta, y) = p(\theta|y)p(y) = p(y|\theta)p(\theta). \tag{2.5}$$

Then, we can simply write the posterior as:

$$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{p(y)}, \tag{2.6}$$

where $p(\theta)$ is our initial knowledge of the parameter, $p(y|\theta)$ is given by the model and $p(y)$ is the marginal PDF given in Property 1.1, i.e. $p(y) = \int p(y|\theta)p(\theta)d\theta$.

Thus, by employing Bayes' theorem as derived above to combine the prior and likelihood, we get the *posterior*, $p(\theta|y)$, one of the main goals of Bayesian statistics.

Definition 2.3 (Posterior). *The posterior distribution describes what we know about θ after observ-*

ing y, and can be easily derived from the likelihood and prior using Bayes' rule:

$$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{p(y)} \propto I(\theta|y)p(\theta). \tag{2.7}$$

2.1.4 Bayesian decision theory

Next comes the question of how we can use the posterior, $p(\theta|y)$, to make decisions. Typically, in Bayesian statistics, the goal is to minimize some function known as the cost, $C(\theta, \alpha)$, where θ is the quantity of interest and α denotes the decision. Since we do not know the true value of θ, we cannot directly minimize the actual cost $C(\theta, \alpha)$. Instead, we consider the expected cost under our current beliefs about θ, which are encoded in the posterior distribution $p(\theta|y)$. Therefore, an optimal decision α is one that minimizes the average cost we would expect to incur, weighted by the probabilities given by the posterior:

$$\alpha_{\mathrm{opt}} = \arg\min_{\alpha} \mathbb{E}_{p(\theta|y)}[C(\theta, \alpha)] = \arg\min_{\alpha} \int_{\Theta} C(\theta, \alpha)p(\theta|y)\mathrm{d}\theta. \tag{2.8}$$

where the expectation is taken w.r.t. to the posterior $p(\theta|y)$, for which y is given and $\theta \in \Theta$ is a random variable. Note that this process follows all the steps outlined in Sec. 2.1.1.

2.1.4.1 Cost functions for estimation problems

Let us now introduce two commonly used cost functions for estimation problems and then derive their respective estimators. In the context of estimation, choosing a decision α corresponds to selecting a specific value as our best guess for the unknown parameter θ. Therefore, the decision α is simply our estimator of θ, denoted from now on as $\tilde{\theta}$. Importantly, the estimator $\tilde{\theta}$ is a function of the observed measurements y, reflecting how our estimate depends on the data collected. However, not every choice of estimator $\tilde{\theta}(y)$ is guaranteed to be optimal. When an estimator is optimal with respect to a chosen cost function, we will denote it with a subscript, such as $\tilde{\theta}_{\mathrm{opt}}(y)$ for a general optimal estimator, or $\tilde{\theta}_{\mathrm{MMSE}}(y)$ specifically when it is the optimal estimator minimizing the mean squared error (MSE).

QUADRATIC COST AND THE MMSE ESTIMATE   One of the most common cost functions in estimation theory is the quadratic cost or the *squared error*:

$$C(\theta, \tilde{\theta}) = (\theta - \tilde{\theta}(y))^{\mathrm{T}}(\theta - \tilde{\theta}(y)). \tag{2.9}$$

Its corresponding optimal estimator is known as the minimum mean squared error (MMSE) estimator. To derive it, we need to minimize the posterior expected cost as described in Eq. (2.8) for the case of a quadratic cost:

$$\tilde{\theta}_{\mathrm{MMSE}}(y) = \arg\min_{\tilde{\theta}} \mathbb{E}_{p(\theta|y)}\left[C(\theta, \tilde{\theta})\right] = \arg\min_{\tilde{\theta}} \mathbb{E}_{p(\theta|y)}\left[(\theta - \tilde{\theta}(y))^{\mathrm{T}}(\theta - \tilde{\theta}(y))\right] \tag{2.10}$$

$$= \arg\min_{\tilde{\theta}} \int_{\Theta} (\theta - \tilde{\theta}(y))^{\mathrm{T}}(\theta - \tilde{\theta}(y))\, p(\theta|y)\, \mathrm{d}\theta. \tag{2.11}$$

Note that the quantity we are minimizing, i.e. the posterior expected quadratic cost, is precisely the *mean squared error* (MSE), a term likely more familiar to the reader. We can formally define the MSE as follows:

Definition 2.4 (Mean squared error). *The MSE of an estimator $\tilde{\theta}(y)$ for a parameter $\theta \in \Theta$, given observations y, is defined as*

$$\mathbb{E}_{p(\theta|y)}\left[(\theta - \tilde{\theta}(y))^{\mathrm{T}}(\theta - \tilde{\theta}(y))\right] = \int_{\Theta} (\theta - \tilde{\theta}(y))^{\mathrm{T}}(\theta - \tilde{\theta}(y))\, p(\theta|y)\, \mathrm{d}\theta \tag{2.12}$$

Here, the "mean" (or average) is taken with respect to the posterior distribution, $p(\theta|y)$.

Later, we will discuss the *average mean squared error* (aMSE), which refers to the squared error or quadratic cost averaged with respect to the joint PDF of the parameter and the data.

The optimal estimatior that solves the above minimization problem can be shown to be the mean of the posterior, which is therefore referred to as the MMSE estimate. Namely,

$$\tilde{\theta}_{\mathrm{MMSE}}(y) = \arg\min_{\tilde{\theta}} \mathbb{E}_{p(\theta|y)}\left[(\theta - \tilde{\theta}(y))^{\mathrm{T}}(\theta - \tilde{\theta}(y))\right] = \bar{\theta}, \tag{2.13}$$

where $\bar{\theta}$ denotes the mean of θ with respect to the posterior:

$$\bar{\theta} = \int_{\Theta} \theta\, p(\theta|y)\mathrm{d}\theta. \tag{2.14}$$

To visualize the MMSE estimator, see Fig. 2.1, where it is marked as a black dot.

Proof. *Let us show that the optimal estimator of θ is the mean of the posterior. To do so, let us first rewrite the cost function $C(\theta, \tilde{\theta})$ as follows:*

$$C(\theta, \tilde{\theta}) = (\theta - \tilde{\theta}(y))^{\mathrm{T}}(\theta - \tilde{\theta}(y)) = (\theta - \bar{\theta} + \bar{\theta} - \tilde{\theta}(y))^{\mathrm{T}}(\underbrace{\theta - \bar{\theta}}_{\text{zero mean}} + \underbrace{\bar{\theta} - \tilde{\theta}(y)}_{\text{deterministic}}), \quad (2.15)$$

where we have added and subtracted the mean of the posterior $\bar{\theta}$. That splits the factor $\theta - \tilde{\theta}(y)$ into two terms, one still probabilistic but with zero mean, $\theta - \bar{\theta}$, and another completely deterministic, $\bar{\theta} - \tilde{\theta}(y)$. If we now expand the quadratic term and take the expectation value w.r.t. to the posterior $p(\theta|y)$, we obtain:

$$\mathbb{E}_{p(\theta|y)}\left[C(\theta, \tilde{\theta})\right] = \mathbb{E}_{p(\theta|y)}\left[(\theta - \bar{\theta})^{\mathrm{T}}(\theta - \bar{\theta})\right] + \overbrace{\mathbb{E}_{p(\theta|y)}\left[(\theta - \bar{\theta})^{\mathrm{T}}\right]}^{0}(\bar{\theta} - \tilde{\theta}(y))$$

$$+ (\bar{\theta} - \tilde{\theta}(y))^{\mathrm{T}}\underbrace{\mathbb{E}_{p(\theta|y)}\left[(\theta - \bar{\theta})\right]}_{0} + \mathbb{E}_{p(\theta|y)}\left[(\bar{\theta} - \tilde{\theta}(y))^{\mathrm{T}}(\bar{\theta} - \tilde{\theta}(y))\right]$$

$$= \mathbb{E}_{p(\theta|y)}\left[(\theta - \bar{\theta})^{\mathrm{T}}(\theta - \bar{\theta})\right] + (\bar{\theta} - \tilde{\theta}(y))^{\mathrm{T}}(\bar{\theta} - \tilde{\theta}(y)). \quad (2.16)$$

Thus, it follows that the argument of $\tilde{\theta}(y)$ that minimizes the quantity above can only be $\bar{\theta}$. Therefore, the optimal estimator that minimizes the expectation of the squared error w.r.t. the posterior, $p(\theta|y)$, is the mean of the posterior, i.e. the MMSE estimator given in Eq. (2.13).

The 0-1 cost and the MAP estimator   Let us now consider the following cost function:

$$C(\theta, \tilde{\theta}) = -\delta(\theta - \tilde{\theta}(y)), \quad (2.17)$$

where $\delta(\,\cdot\,)$ represents the Dirac delta. This cost function, also referred to as the 0-1 loss function, picks only the exact value of θ and disregards all the other values of $\tilde{\theta}(y)$. It is an idealized cost function, since it penalizes any deviation from the true value equally and maximally.

To find the optimal estimator for this cost function, we compute the expected value of $C(\theta, \tilde{\theta})$ w.r.t. the posterior:

$$\mathbb{E}_{p(\theta|y)}\left[C(\theta, \tilde{\theta})\right] = -\int_{\theta \in \mathbb{R}^n} \delta(\theta - \tilde{\theta}(y))p(\theta|y)\mathrm{d}\theta = -p(\theta|y)|_{\theta = \tilde{\theta}}, \quad (2.18)$$
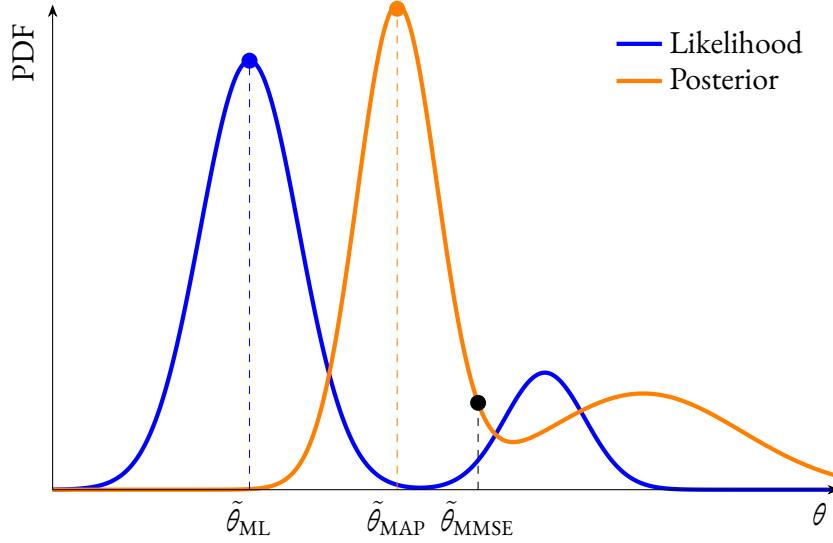
Figure 2.1: Comparison of the ML, MAP and MMSE estimates. Visual representation of the maximum likelihood estimator ($\tilde{\theta}_{\mathrm{ML}}(y)$, blue dot), the maximum a-priori estimator ($\tilde{\theta}_{\mathrm{MAP}}(y)$, yellow dot) and the minimum mean squared error estimator ($\tilde{\theta}_{\mathrm{MMSE}}(y)$, black dot).

where in the last step we have used the Dirac delta property: $\int_{-\infty}^{\infty} f(x)\delta(x - \tilde{\theta}(y))\mathrm{d}x = f(\tilde{\theta}(y))$. If now we minimize over all possible arguments of $\tilde{\theta}(y)$, we get the optimal estimator $\tilde{\theta}(y)$:

$$\tilde{\theta}_{\mathrm{MAP}}(y) = \arg \min_{\tilde{\theta}} \left[-p(\theta|y)|_{\theta=\tilde{\theta}}\right] = \arg \max_{\theta} p(\theta|y). \qquad (2.19)$$

As shown by a yellow dot in Fig. 2.1, the optimal estimator for the 0-1 cost is the θ that maximizes the posterior distribution, which we call the maximum a-posteriori (MAP) estimator.

While the Bayesian approach allows for the incorporation of prior information through $p(\theta)$, sometimes prior knowledge is not available or we ignore it altogether. In that case, the goal is to directly maximize the likelihood function $p(y|\theta)$. This leads to the maximum likelihood (ML) estimator:

$$\tilde{\theta}_{\mathrm{ML}}(y) = \arg \max_{\theta} p(y|\theta), \qquad (2.20)$$

which is marked in Fig. 2.1 with a blue dot, i.e. the maximum of the likelihood. The ML estimator can be seen as a method that focuses entirely on the data, with no consideration of prior beliefs.

2.2 Fisher information and estimation theory

Having established the foundations of classical Bayesian estimation, we now turn our attention to a quantity crucial for assessing the quality of an estimator: the Fisher information (FI). Besides quantifying the amount of information that a dataset contains about an unknown parameter, the FI also lower bounds the variance of unbiased estimators, serving as an estimation benchmark. This bound, referred to as Cramér-Rao bound, can be extended to a Bayesian setting by incorporating prior information. Known as the Bayesian Cramér-Rao bound (BCRB), it establishes a lower bound on the average mean squared error (aMSE) or average quadratic cost of all (biased or unbiased) estimators, given a well-behaved prior.

2.2.1 Fisher information

Let y be a random variable representing our observations, and θ an unknown parameter, for which we can construct the likelihood $p(y|\theta)$. Then, the so-called FI is defined as:

$$F[p(\theta|y)] = \int dy \frac{1}{p(\theta|y)} \left(\frac{\partial p(\theta|y)}{\partial \theta} \right)^2 \tag{2.21}$$

$$= \mathbb{E}_{p(y|\theta)} \left[\left(\frac{\partial}{\partial \theta} \ln p(\theta|y) \right)^2 \right] \tag{2.22}$$

$$= -\mathbb{E}_{p(y|\theta)} \left[\frac{\partial^2}{\partial \theta^2} \ln p(\theta|y) \right] . \tag{2.23}$$

2.2.2 Cramér-Rao bound

Under the conditions of regularity and local unbiasedness, i.e.,

$$\int \tilde{\theta}(y) \frac{\partial}{\partial \theta} p(y|\theta) dy = 1, \qquad \text{and} \qquad \int dy \frac{\partial}{\partial \theta} p(y|\theta) = 0, \tag{2.24}$$

for any locally unbiased estimator $\tilde{\theta}$, its MSE is lower bounded by the inverse of the FI. Namely,

$$\mathbb{E}_{p(\theta|y)} \left[\Delta^2 \tilde{\theta} \right] \geq \frac{1}{F[p(\theta|y)]}, \tag{2.25}$$

where the formula for the FI is given in Eqs. (2.21-2.23) and $\Delta^2\tilde{\theta}$ is the squared error:

$$\Delta^2\tilde{\theta} := (\theta - \tilde{\theta}(y))^2. \tag{2.26}$$

2.2.3 Bayesian Cramér-Rao bound

In the Bayesian framework, the parameter θ is treated as a random variable with a prior distribution $p(\theta)$, specified before observing any data. Therefore, an optimal estimator in this setting should not only minimize the MSE with respect to the posterior, but also account for which values of θ are more likely. To reflect this, the average mean squared error (aMSE) is defined as:

Definition 2.5 (Average mean squared error). *Let $\theta \in \Theta$ be a parameter, $\tilde{\theta}$ its estimate and y the observed data with a joint PDF $p(\theta, y)$. The aMSE of the estimator $\tilde{\theta}$ is defined as*

$$\mathbb{E}\left[\Delta^2\tilde{\theta}\right] := \mathbb{E}_{p(\theta,y)}\left[(\theta - \tilde{\theta}(y))^2\right] = \int_Y \int_\Theta (\theta - \tilde{\theta}(y))^2 p(\theta, y) \, \mathrm{d}\theta \, \mathrm{d}y \tag{2.27}$$

$$= \int_\Theta \mathrm{d}\theta \, p(\theta) \int_Y \mathrm{d}y \, p(y|\theta)(\theta - \tilde{\theta}(y))^2. \tag{2.28}$$

The aMSE can be lower bounded by different classes of Bayesian bounds, which are equivalent when working in the LG regime [67]. In this thesis, we focus on the so-called marginal unconditional BCRB [86, 67]:

$$\mathbb{E}\left[\Delta^2\tilde{\theta}\right] \geq \frac{1}{J_B}, \tag{2.29}$$

where J_B is the Bayesian information (BI) [2],

$$J_B = \mathbb{E}_{p(\theta,y)}\left[(\partial_\theta \log p(\theta, y))^2\right]. \tag{2.30}$$

The BI can be split into two terms, $J_B = J_P + J_M$. The first term, J_P, represents the contribution of our prior knowledge about θ,

$$J_P = \mathrm{F}[p(\theta)] = \mathbb{E}_{p(\theta)}\left[(\partial_\theta \log p(\theta))^2\right]. \tag{2.31}$$

The second term, namely the contribution of the measurement records, or J_M, can be understood

as averaging the FI of the likelihood over the prior distribution, i.e.,

$$J_M = \mathbb{E}_{p(\theta,y)}\left[(\partial_\theta \log p(y,\theta))^2\right] = \tag{2.32}$$

$$= \int d\theta\, p(\theta)\, \mathrm{F}[p(y|\theta)], \tag{2.33}$$

with

$$\mathrm{F}[p(y|\theta)] = \mathbb{E}_{p(y|\theta)}\left[(\partial_\theta \log p(y|\theta))^2\right] \tag{2.34}$$

being the FI of the likelihood $p(y|\theta)$ of observing a measurement trajectory y given that the field (or more precisely, the Larmor frequency) has a value θ, as introduced in Eqs. (2.21-2.23).

2.3 Bayesian filtering

Bayesian filtering provides a recursive framework for estimating the state parameters of a system based on indirect, noisy measurements over time. To formulate this problem, we begin by discretizing time s.t. $t = k\Delta t$, where Δt is the time step and $k \in \mathbb{N}$. With this discretization, the evolution of parameters and observations can be written as follows:

- x_k: the state vector at time k, representing the unknown parameters we aim to estimate.

- y_k: the measurement vector at time k.

All the state parameters and measurements *up to* time k can be collected into:

- $x_{0:k} = \{x_0, x_1, \ldots, x_k\}$: the state trajectory up to time k,

- $y_{0:k} = \{y_0, y_1, \ldots, y_k\}$: the measurement trajectory up to time k.

2.3.1 State space model

An essential step to formulate any filtering problem is to model the system and measurement by creating a state space model. These models describe how the state vector evolves recursively, as well as how it relates to the observations. In particular, for a state vector x_k and a measurement vector y_k, where k denotes time, their respective dynamics can be modeled with either equations or probability distributions:

$$x_k = f_{k-1}[x_{k-1}, u_{k-1}, q_{k-1}] \qquad \Longleftrightarrow \qquad p(x_k|x_{0:k-1}, y_{0:k}; u_{0:k-1}) \tag{2.35}$$

$$y_k = b_k[x_k, r_k] \qquad \Longleftrightarrow \qquad p(y_k|x_{0:k}, y_{0:k}) \tag{2.36}$$

where $\boldsymbol{u}_k$ is an external known control vector. Furthermore, we assume the state noise $\boldsymbol{q}_k$ and the measurement noise $\boldsymbol{r}_k$ to be zero-mean white noise vectors, with known probability distributions which are not necessarily Gaussian (see Sec. 1.3.8).

The initial value for the state vector is drawn from a prior distribution $\boldsymbol{x}_0 \sim p(\boldsymbol{x}_0)$. The first equation describes how the system evolves in time and updates the state vector from time k-1 to time k. The measurement model given by Eq. (2.36) relates the state vector to the observation vector and outlines how the measurement updates the state of the system.

Thus far, we have made no assumptions regarding the noise, besides it being white. In what follows, we assume that the process and measurement noise are uncorrelated. This is an essential condition to ensure that the state follows a Markov evolution and that the measurements are conditionally independent of past measurements and states. We will later revisit the case of correlated noise and discuss how to de-correlate it when necessary.

2.3.1.1 UNCORRELATED PROCESS AND MEASUREMENT NOISE

From the system and measurement model in Eqs. (2.35-2.36), it follows that the state vector $\boldsymbol{x}_k$ depends solely on the previous state $\boldsymbol{x}_{k-1}$ and control input $\boldsymbol{u}_{k-1}$, while the measurement $\boldsymbol{y}_k$ depends only on the current state $\boldsymbol{x}_k$. Assuming that the process noise $\boldsymbol{q}_{k-1}$ and measurement noise $\boldsymbol{r}_k$ are white (i.e., temporally uncorrelated), as well as mutually uncorrelated, there is no mechanism through which they can introduce dependencies between $\boldsymbol{x}_k$ and earlier states $\boldsymbol{x}_{0:k-2}$ or past measurements $\boldsymbol{y}_{0:k-1}$. Thus, the probability distributions of Eqs. (2.35-2.36) can be simplified to

$$\boldsymbol{x}_k = \boldsymbol{f}_{k-1}[\boldsymbol{x}_{k-1}, \boldsymbol{u}_{k-1}, \boldsymbol{q}_{k-1}] \qquad \Longleftrightarrow \qquad p(\boldsymbol{x}_k|\boldsymbol{x}_{k-1}; \boldsymbol{u}_{k-1}) \qquad (2.37)$$

$$\boldsymbol{y}_k = \boldsymbol{h}_k[\boldsymbol{x}_k, \boldsymbol{r}_k] \qquad \Longleftrightarrow \qquad p(\boldsymbol{y}_k|\boldsymbol{x}_k) \qquad (2.38)$$

$$\text{when} \quad \text{cov}\,[\boldsymbol{q}_k, \boldsymbol{r}_s] = 0 \ \ \forall \, k, s. \qquad (2.39)$$

Note that Eq. (2.37) and Eq. (2.38) fulfill two very important properties: the Markovianity of states and the conditional independence of measurements. In particular, the system function $\boldsymbol{f}_{k-1}$ in Eq. (2.35) depends only on $\boldsymbol{x}_{k-1}$ and $\boldsymbol{u}_{k-1}$, and not on any measurement $\boldsymbol{y}_{0:k}$ (not even indirectly since the system and measurement noise are uncorrelated). Hence, the transition probability $p(\boldsymbol{x}_k|\boldsymbol{x}_{0:k-1}, \boldsymbol{y}_{0:k}; \boldsymbol{u}_{0:k-1})$ in Eq. (2.35) is conditionally dependent only on $\boldsymbol{x}_{k-1}$ and $\boldsymbol{u}_{k-1}$, i.e. $p(\boldsymbol{x}_k|\boldsymbol{x}_{k-1}; \boldsymbol{u}_{k-1})$, which means that Eq. (2.37) obeys the Markov property defined in Property 1.4.

Additionally, the likelihood in Eq. (2.38) is conditionally dependent only on $\boldsymbol{x}_k$, and independent on other measurements and state vectors. Therefore, it fulfills the following property:

Property 2.1 (Conditional independence of measurements). *The current measurement $\boldsymbol{y}_k$ given the current state $\boldsymbol{x}_k$ is conditionally independent of the measurement and state histories:*

$$p(\boldsymbol{y}_k|\boldsymbol{x}_{0:k}, \boldsymbol{y}_{0:k-1}) = p(\boldsymbol{y}_k|\boldsymbol{x}_k). \tag{2.40}$$

Moving forward we will drop the dependence of the transition probability on the control law $\boldsymbol{u}_{k-1}$ for brevity.

2.3.2 Bayesian optimal filtering

The goal of Bayesian optimal filtering is to compute the posterior density of the state $\boldsymbol{x}_k$ given the history of measurements up to time k, i.e. $p(\boldsymbol{x}_k|\boldsymbol{y}_{0:k})$. To do so, we can either use a brute force approach where we find the posterior by simply applying Bayes' rule and then marginalizing w.r.t. $\boldsymbol{x}_{0:k-1}$, or use a more efficient recursive approach whose complexity does not grow with k. In particular, we can recursively compute the posterior density at time k, i.e. $p(\boldsymbol{x}_k|\boldsymbol{y}_{0:k})$, from the posterior density at the previous time, $p(\boldsymbol{x}_{k-1}|\boldsymbol{y}_{0:k-1})$, as summarized in Fig. 2.2. First, we assume that in the previous step, k-1, we have computed $p(\boldsymbol{x}_{k-1}|\boldsymbol{y}_{0:k-1})$. Next, we take this density, which summarizes our knowledge of the state $\boldsymbol{x}_{k-1}$ given the measurement trajectory $\boldsymbol{y}_{0:k-1}$, and use it to predict the state at time k given the measurement trajectory up to time k-1. Namely, we use the system or process model to predict the probability density $p(\boldsymbol{x}_k|\boldsymbol{y}_{0:k-1})$. Then, we update this probability distribution using the measurement model, which gives us the current measurement $\boldsymbol{y}_k$, and allows us to derive the posterior for the state $\boldsymbol{x}_k$: $p(\boldsymbol{x}_k|\boldsymbol{y}_{0:k})$. From this posterior, we can then compute the relevant quantities of interest, such as the estimator of $\boldsymbol{x}_k$ that minimizes the quadratic cost or MSE.

Let us now write a detailed step-by-step guide on how to recursively update the Bayesian filter:

(i) **Initialize.** The first step of the recursion is the prior $p(\boldsymbol{x}_0)$.

(ii) **Predict.** Compute $p(\boldsymbol{x}_k|\boldsymbol{y}_{0:k-1})$ from $p(\boldsymbol{x}_{k-1}|\boldsymbol{y}_{0:k-1})$. Namely,

$$p(\boldsymbol{x}_k|\boldsymbol{y}_{0:k-1}) \underset{1.1}{=} \int p(\boldsymbol{x}_k, \boldsymbol{x}_{k-1}|\boldsymbol{y}_{0:k-1})\mathrm{d}\boldsymbol{x}_{k-1} \tag{2.41}$$

$$\underset{1.2}{=} \int p(\boldsymbol{x}_k|\boldsymbol{x}_{k-1}, \boldsymbol{y}_{0:k-1})p(\boldsymbol{x}_{k-1}|\boldsymbol{y}_{0:k-1})\mathrm{d}\boldsymbol{x}_{k-1}. \tag{2.42}$$

However, if now we recall Property 1.4, the probability density $p(\boldsymbol{x}_k|\boldsymbol{x}_{k-1}, \boldsymbol{y}_{0:k-1})$ can be simplified to $p(\boldsymbol{x}_k|\boldsymbol{x}_{k-1})$ because the current state $\boldsymbol{x}_k$ is conditionally independent of the
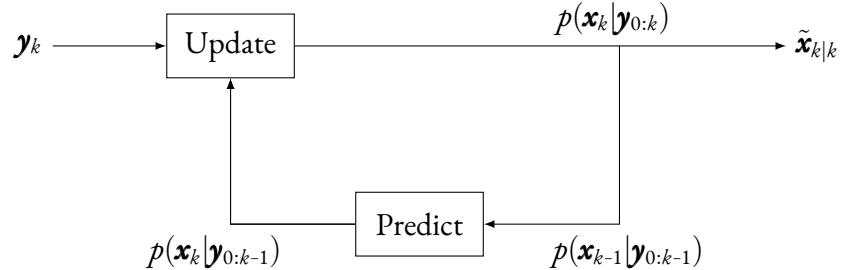
Figure 2.2: A scheme illustrating the recursive algorithm of a Bayesian filter. In the "Predict" box, a model is used to propagate the *update PDF* from step $k-1$ to the *prediction PDF* $p(\boldsymbol{x}_k|\boldsymbol{y}_{0:k-1})$. Then, in the "Update" box, a new measurement $\boldsymbol{y}_k$ is incorporated into the algorithm to update the *prediction PDF* into a new *update PDF* for time k: $p(\boldsymbol{x}_k|\boldsymbol{y}_{0:k})$. Finally, before repeating this process again, an estimate at time k is computed from the just-updated posterior: $\tilde{\boldsymbol{x}}_{k|k}$.

measurements $\boldsymbol{y}_{0:k-1}$. Thus, we get

$$\text{prediction}: p(\boldsymbol{x}_k|\boldsymbol{y}_{0:k-1}) = \int p(\boldsymbol{x}_k|\boldsymbol{x}_{k-1})p(\boldsymbol{x}_{k-1}|\boldsymbol{y}_{0:k-1})\mathrm{d}\boldsymbol{x}_{k-1}, \tag{2.43}$$

which is also known as the Chapman-Kolmogorov equation, introduced in Definition 1.14.

(iii) **Update.** Compute $p(\boldsymbol{x}_k|\boldsymbol{y}_{0:k})$ from $p(\boldsymbol{x}_k|\boldsymbol{y}_{0:k-1})$. To do so, we want to update our knowledge about $\boldsymbol{x}_k$ using the new measurement $\boldsymbol{y}_k$:

$$p(\boldsymbol{x}_k|\boldsymbol{y}_{0:k}) = p(\boldsymbol{x}_k|\boldsymbol{y}_k,\boldsymbol{y}_{0:k-1}) = \frac{p(\boldsymbol{y}_k|\boldsymbol{x}_k,\boldsymbol{y}_{0:k-1})p(\boldsymbol{x}_k|\boldsymbol{y}_{0:k-1})}{p(\boldsymbol{y}_k|\boldsymbol{y}_{0:k-1})}. \tag{2.44}$$

This equation can be further simplified if we recall that our measurement model fulfills Property 2.1: the likelihood of measuring $\boldsymbol{y}_k$ is given only by $\boldsymbol{x}_k$ and independent of previous measurements. Namely,

$$p(\boldsymbol{y}_k|\boldsymbol{x}_k,\boldsymbol{y}_{0:k-1}) = p(\boldsymbol{y}_k|\boldsymbol{x}_k). \tag{2.45}$$

Therefore,

$$\text{update}: p(\boldsymbol{x}_k|\boldsymbol{y}_{0:k}) = \frac{1}{p(\boldsymbol{y}_k|\boldsymbol{y}_{0:k-1})}p(\boldsymbol{y}_k|\boldsymbol{x}_k)p(\boldsymbol{x}_k|\boldsymbol{y}_{0:k-1}), \tag{2.46}$$

where $p(\boldsymbol{x}_k|\boldsymbol{y}_{0:k-1})$ is the prediction PDF, and $p(\boldsymbol{y}_k|\boldsymbol{y}_{0:k-1})$ is the likelihood of detecting $\boldsymbol{y}_k$

given a past measurement trajectory $\boldsymbol{y}_{0:k-1}$:

$$p(\boldsymbol{y}_k|\boldsymbol{y}_{0:k-1}) = \int \mathrm{d}\boldsymbol{x}_k\, p(\boldsymbol{y}_k, \boldsymbol{x}_k|\boldsymbol{y}_{0:k-1}) = \int \mathrm{d}\boldsymbol{x}_k\, p(\boldsymbol{y}_k|\boldsymbol{x}_k) p(\boldsymbol{x}_k|\boldsymbol{y}_{0:k-1}). \qquad (2.47)$$

Note that these expressions are general and provide a recursive solution to any filtering problem. In the case of LG systems, these equations can be solved exactly. When the system is nonlinear and/or non-Gaussian, good approximations can also be found.

2.4 The discrete uncorrelated Kalman filter

The Kalman filter (KF) is one of the few filters that, under certain modeling assumptions, yields an exact solution. Additionally, it is also the foundation of more advanced filters that provide approximate solutions to the Bayesian filtering equations, enabling them to handle more general models.

The starting assumption to find a close form solution to the Bayesian filtering equations is that the system and measurement models are both LG. Namely,

$$\boldsymbol{x}_k = \boldsymbol{A}_{k-1}\boldsymbol{x}_{k-1} + \boldsymbol{B}_{k-1}\boldsymbol{u}_{k-1} + \boldsymbol{G}_{k-1}\boldsymbol{q}_{k-1}, \qquad (2.48)$$

$$\boldsymbol{y}_k = \boldsymbol{H}_k\boldsymbol{x}_k + \boldsymbol{r}_k, \qquad (2.49)$$

where $\boldsymbol{x}_k$ is the state vector, $\boldsymbol{y}_k$ the measurement. Assuming the model to be Gaussian implies that the process noise $\boldsymbol{q}_{k-1} \sim \mathcal{N}(0, \boldsymbol{Q}_{k-1})$ and the measurement noise $\boldsymbol{r}_k \sim \mathcal{N}(0, \boldsymbol{R}_k)$ are both mutually uncorrelated white Gaussian noises with zero mean and covariances $\boldsymbol{Q}_{k-1}$ and $\boldsymbol{R}_k$, respectively. The dependence of the model and the measurement on the state is clearly linear, and it is guided by a transition matrix $\boldsymbol{A}_{k-1}$ and by a measurement model matrix $\boldsymbol{H}_k$, respectively. Then, the model described by Eqs. (2.48-2.49) can be equivalently written in probabilistic terms as

$$\boldsymbol{x}_k \sim p(\boldsymbol{x}_k|\boldsymbol{x}_{k-1}) = \mathcal{N}(\boldsymbol{x}_k|\boldsymbol{A}_{k-1}\boldsymbol{x}_{k-1} + \boldsymbol{B}_{k-1}\boldsymbol{u}_{k-1}, \boldsymbol{G}_{k-1}\boldsymbol{Q}_{k-1}\boldsymbol{G}_{k-1}^{\mathrm{T}}), \qquad (2.50)$$

$$\boldsymbol{y}_k \sim p(\boldsymbol{y}_k|\boldsymbol{x}_k) = \mathcal{N}(\boldsymbol{y}_k|\boldsymbol{H}_k\boldsymbol{x}_k, \boldsymbol{R}_k), \qquad (2.51)$$

where the notation $\mathcal{N}(\boldsymbol{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma})$ simply denotes a multivariate Gaussian PDF of a state $\boldsymbol{x}$ with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$.

The KF recursively computes the prediction and update probability densities of Eq. (2.43) and

Eq. (2.46), which now, given that our model is LG, will also be Gaussian:

$$\text{prediction} \quad \implies \quad p(\boldsymbol{x}_k|\boldsymbol{y}_{0:k\text{-}1}) := \mathcal{N}(\boldsymbol{x}_k|\tilde{\boldsymbol{x}}_{k|k\text{-}1}, \boldsymbol{\Sigma}_{k|k\text{-}1}) \tag{2.52}$$

$$\text{update} \quad \implies \quad p(\boldsymbol{x}_k|\boldsymbol{y}_{0:k}) := \mathcal{N}(\boldsymbol{x}_k|\tilde{\boldsymbol{x}}_{k|k}, \boldsymbol{\Sigma}_{k|k}) \tag{2.53}$$

where $\tilde{\boldsymbol{x}}_{k|k\text{-}1}$ and $\tilde{\boldsymbol{x}}_{k|k}$ are their respective means, and $\boldsymbol{\Sigma}_{k|k\text{-}1}$ and $\boldsymbol{\Sigma}_{k|k}$, their covariances. Additionally, given that $\tilde{\boldsymbol{x}}_{k|k\text{-}1}$ and $\tilde{\boldsymbol{x}}_{k|k}$ are the means of the prediction and update *posterior* PDFs, they are also the optimal estimators of $\boldsymbol{x}_k$ that minimize the quadratic cost or MSE, as shown in Sec. 2.1.4.1. Namely,

$$\tilde{\boldsymbol{x}}_{k|k\text{-}1} = \int \boldsymbol{x}_k \, p(\boldsymbol{x}_k|\boldsymbol{y}_{0:k\text{-}1}) \mathrm{d}\boldsymbol{x}_k \quad = \quad \textit{a-priori} \text{ estimate,} \tag{2.54}$$

$$\tilde{\boldsymbol{x}}_{k|k} = \int \boldsymbol{x}_k \, p(\boldsymbol{x}_k|\boldsymbol{y}_{0:k}) \mathrm{d}\boldsymbol{x}_k \quad = \quad \textit{a-posteriori} \text{ estimate.} \tag{2.55}$$

Note that both $\tilde{\boldsymbol{x}}_{k|k\text{-}1}$ and $\tilde{\boldsymbol{x}}_{k|k}$ are estimates of the same quantity, i.e. $\boldsymbol{x}_k$, with the crucial difference that the *a-posteriori* estimate incorporates the latest measurement, $\boldsymbol{y}_k$, while the *a-priori* estimate does not. Therefore, it is to be expected that the *a-posteriori* estimate is more accurate than the *a-priori* one.

In Sec. 2.3.2 we discussed how the posterior distribution of $\boldsymbol{x}_k$ conditioned on measurements up to time k: $p(\boldsymbol{x}_k|\boldsymbol{y}_{0:k})$, can be computed recursively starting from $p(\boldsymbol{x}_0)$ using recursive Bayesian optimal filtering. In the case of LG systems, we will show that the posterior, a.k.a. update density of Eq. (2.53), remains fully Gaussian at all times. Then, given that Gaussian distributions are uniquely determined by their mean and covariance, the filtering problem reduces to recursively updating the moments of Eqs. (2.52-2.53):

(i) The **Prediction step** is defined by the following equations

$$\tilde{\boldsymbol{x}}_{k|k\text{-}1} = \boldsymbol{A}_{k\text{-}1} \tilde{\boldsymbol{x}}_{k\text{-}1|k\text{-}1} + \boldsymbol{B}_{k\text{-}1} \boldsymbol{u}_{k\text{-}1}, \tag{2.56}$$

$$\boldsymbol{\Sigma}_{k|k\text{-}1} = \boldsymbol{A}_{k\text{-}1} \boldsymbol{\Sigma}_{k\text{-}1|k\text{-}1} \boldsymbol{A}_{k\text{-}1}^{\mathrm{T}} + \boldsymbol{G}_{k\text{-}1} \boldsymbol{Q}_{k\text{-}1} \boldsymbol{G}_{k\text{-}1}^{\mathrm{T}}. \tag{2.57}$$

(ii) The **Update step** is given by

$$\tilde{\boldsymbol{x}}_{k|k} = \tilde{\boldsymbol{x}}_{k|k\text{-}1} + \boldsymbol{K}_k(\boldsymbol{y}_k - \boldsymbol{H}_k \tilde{\boldsymbol{x}}_{k|k\text{-}1}), \tag{2.58}$$

$$\boldsymbol{\Sigma}_{k|k} = (\mathbb{I} - \boldsymbol{K}_k \boldsymbol{H}_k) \boldsymbol{\Sigma}_{k|k\text{-}1}, \tag{2.59}$$

where

$$K_k = \Sigma_{k|k-1} H_k^{\mathrm{T}} T_k^{-1}, \qquad (2.60)$$

$$T_k = H_k \Sigma_{k|k-1} H_k^{\mathrm{T}} + R_k. \qquad (2.61)$$

Furthermore, let us highlight that T_k is not just a handy definition to shorten our notation but actually is the *innovation covariance*. Namely,

$$T_k = \mathbb{E}\left[(\boldsymbol{y}_k - \tilde{\boldsymbol{y}}_k)(\boldsymbol{y}_k - \tilde{\boldsymbol{y}}_k)^{\mathrm{T}}\right], \qquad (2.62)$$

where $\boldsymbol{y}_k - \tilde{\boldsymbol{y}}_k$ is the so-called innovation, which is the difference between the measurement and its prediction

$$\tilde{\boldsymbol{y}}_k = H_k \tilde{\boldsymbol{x}}_{k|k-1}. \qquad (2.63)$$

The relationship between Eq. (2.61) and Eq. (2.62) can be established by:

$$T_k = \mathbb{E}\left[(H_k \boldsymbol{x}_k + \boldsymbol{r}_k - H_k \tilde{\boldsymbol{x}}_{k|k-1})(H_k \boldsymbol{x}_k + \boldsymbol{r}_k - H_k \tilde{\boldsymbol{x}}_{k|k-1})^{\mathrm{T}}\right] \qquad (2.64)$$

$$= \mathbb{E}\left[(H_k(\boldsymbol{x}_k - \tilde{\boldsymbol{x}}_{k|k-1}) + \boldsymbol{r}_k)(H_k(\boldsymbol{x}_k - \tilde{\boldsymbol{x}}_{k|k-1}) + \boldsymbol{r}_k)^{\mathrm{T}}\right] \qquad (2.65)$$

$$= H_k \mathbb{E}\left[(\boldsymbol{x}_k - \tilde{\boldsymbol{x}}_{k|k-1})(\boldsymbol{x}_k - \tilde{\boldsymbol{x}}_{k|k-1}))^{\mathrm{T}}\right] H_k^{\mathrm{T}} + \mathbb{E}\left[\boldsymbol{r}_k \boldsymbol{r}_k^{\mathrm{T}}\right] \qquad (2.66)$$

$$= H_k \Sigma_k H_k^{\mathrm{T}} + R_k, \qquad (2.67)$$

where the cross terms vanish due to the assumption that the state prediction error and the measurement noise are uncorrelated: $\mathbb{E}\left[(\boldsymbol{x}_k - \tilde{\boldsymbol{x}}_{k|k-1})\boldsymbol{r}_k^{\mathrm{T}}\right] = \mathbb{E}\left[\boldsymbol{r}_k(\boldsymbol{x}_k - \tilde{\boldsymbol{x}}_{k|k-1})^{\mathrm{T}}\right] = 0$. Additionally, by taking Eq. (2.59):

$$\Sigma_{k|k-1} - \Sigma_{k|k} = K_k H_k \Sigma_{k|k-1}, \qquad (2.68)$$

and substituting it into Eq. (2.60) and Eq. (2.61), one can rewrite the Kalman gain:

$$\Sigma_{k|k-1} H_k^{\mathrm{T}} = K_k T_k = K_k(H_k \Sigma_{k|k-1} H_k^{\mathrm{T}} + R_k) = (\Sigma_{k|k-1} - \Sigma_{k|k})H_k^{\mathrm{T}} + K_k R_k, \qquad (2.69)$$

such that

$$K_k = \Sigma_{k|k} H_k^{\mathrm{T}} R_k^{-1}. \qquad (2.70)$$

Let us now show how to derive the prediction and update equations through inductive hy-

pothesis.

Proof. *The outline of the proof is as follows: we assume that the posterior at time $k-1$ holds, and then show that both the prediction and update posteriors are also Gaussians with means and covariances following the equations given above.*

(i) **Predict.** *To find an expression for Eq. (2.52), we assume that the step $k-1$ holds, i.e. the posterior at time $k-1$ is a Gaussian PDF of the form:*

$$p(\boldsymbol{x}_{k-1}|\boldsymbol{y}_{0:k-1}) = \mathcal{N}(\boldsymbol{x}_{k-1}|\tilde{\boldsymbol{x}}_{k-1|k-1}, \boldsymbol{\Sigma}_{k-1|k-1}), \tag{2.71}$$

Then, we follow the steps of Eq. (2.41) to find an expression for the prediction PDF:

$$p(\boldsymbol{x}_k|\boldsymbol{y}_{0:k-1}) \underset{1.1}{=} \int p(\boldsymbol{x}_k, \boldsymbol{x}_{k-1}|\boldsymbol{y}_{0:k-1}) \mathrm{d}\boldsymbol{x}_{k-1}$$

$$\underset{1.2}{=} \int p(\boldsymbol{x}_k|\boldsymbol{x}_{k-1}, \boldsymbol{y}_{0:k-1}) p(\boldsymbol{x}_{k-1}|\boldsymbol{y}_{0:k-1}) \mathrm{d}\boldsymbol{x}_{k-1}$$

$$\underset{1.4}{=} \int p(\boldsymbol{x}_k|\boldsymbol{x}_{k-1}) p(\boldsymbol{x}_{k-1}|\boldsymbol{y}_{0:k-1}) \mathrm{d}\boldsymbol{x}_{k-1}$$

$$= \int \mathcal{N}(\boldsymbol{x}_k|\boldsymbol{A}_{k-1}\boldsymbol{x}_{k-1} + \boldsymbol{B}_{k-1}\boldsymbol{u}_{k-1}, \boldsymbol{G}_{k-1}\boldsymbol{Q}_{k-1}\boldsymbol{G}_{k-1}^{\mathrm{T}}) \mathcal{N}(\boldsymbol{x}_{k-1}|\tilde{\boldsymbol{x}}_{k-1|k-1}, \boldsymbol{\Sigma}_{k-1|k-1}) \mathrm{d}\boldsymbol{x}_{k-1}$$

$$\underset{B.1}{=} \int \mathcal{N}\left(\begin{pmatrix} \boldsymbol{x}_{k-1} \\ \boldsymbol{x}_k \end{pmatrix} \middle| \boldsymbol{m}_{k|k-1}, \boldsymbol{P}_{k|k-1} \right) \mathrm{d}\boldsymbol{x}_{k-1}$$

$$\underset{B.2}{=} \mathcal{N}(\boldsymbol{x}_k|\boldsymbol{A}_{k-1}\tilde{\boldsymbol{x}}_{k-1|k-1} + \boldsymbol{B}_{k-1}\boldsymbol{u}_{k-1}, \boldsymbol{A}_{k-1}\boldsymbol{\Sigma}_{k-1|k-1}\boldsymbol{A}_{k-1}^{\mathrm{T}} + \boldsymbol{G}_{k-1}\boldsymbol{Q}_{k-1}\boldsymbol{G}_{k-1}^{\mathrm{T}}) \tag{2.72}$$

where in the last step we computed the marginal of the joint probability density, with mean and covariance

$$\boldsymbol{m}_{k|k-1} = \begin{pmatrix} \tilde{\boldsymbol{x}}_{k-1|k-1} \\ \boldsymbol{A}_{k-1}\tilde{\boldsymbol{x}}_{k-1|k-1} + \boldsymbol{B}_{k-1}\boldsymbol{u}_{k-1} \end{pmatrix},$$

$$\boldsymbol{P}_{k|k-1} = \begin{pmatrix} \boldsymbol{\Sigma}_{k-1|k-1} & \boldsymbol{\Sigma}_{k-1|k-1}\boldsymbol{A}_{k-1}^{\mathrm{T}} \\ \boldsymbol{A}_{k-1}\boldsymbol{\Sigma}_{k-1|k-1} & \boldsymbol{A}_{k-1}\boldsymbol{\Sigma}_{k-1|k-1}\boldsymbol{A}_{k-1}^{\mathrm{T}} + \boldsymbol{G}_{k-1}\boldsymbol{Q}_{k-1}\boldsymbol{G}_{k-1}^{\mathrm{T}} \end{pmatrix}. \tag{2.73}$$

Therefore, the prediction density at time k is also a Gaussian:

$$p(\boldsymbol{x}_k|\boldsymbol{y}_{0:k-1}) := \mathcal{N}(\boldsymbol{x}_k|\tilde{\boldsymbol{x}}_{k|k-1}, \boldsymbol{\Sigma}_{k|k-1}) \tag{2.74}$$

$$= \mathcal{N}(\boldsymbol{x}_k|\boldsymbol{A}_{k-1}\tilde{\boldsymbol{x}}_{k-1|k-1}+\boldsymbol{B}_{k-1}\boldsymbol{u}_{k-1}, \boldsymbol{A}_{k-1}\boldsymbol{\Sigma}_{k-1|k-1}\boldsymbol{A}_{k-1}^{\mathrm{T}} + \boldsymbol{G}_{k-1}\boldsymbol{Q}_{k-1}\boldsymbol{G}_{k-1}^{\mathrm{T}}), \tag{2.75}$$

such that

$$\tilde{\boldsymbol{x}}_{k|k-1} = \boldsymbol{A}_{k-1}\tilde{\boldsymbol{x}}_{k-1|k-1}+\boldsymbol{B}_{k-1}\boldsymbol{u}_{k-1}, \tag{2.76}$$

$$\boldsymbol{\Sigma}_{k|k-1} = \boldsymbol{A}_{k-1}\boldsymbol{\Sigma}_{k-1|k-1}\boldsymbol{A}_{k-1}^{\mathrm{T}} + \boldsymbol{G}_{k-1}\boldsymbol{Q}_{k-1}\boldsymbol{G}_{k-1}^{\mathrm{T}}. \tag{2.77}$$

(ii) **Update.** *The update density of Eq. (2.53) can be derived by first employing Lemma B.1 to find an expression for the joint PDF $p(\boldsymbol{x}_k, \boldsymbol{y}_k|\boldsymbol{y}_{0:k-1})$. Then, using Lemma B.2, one can derive the conditional density $\boldsymbol{x}_k|\boldsymbol{y}_k \sim p(\boldsymbol{x}_k|\boldsymbol{y}_k, \boldsymbol{y}_{0:k-1})$. The joint PDF, conditioned on the measurement outcomes up to time $k-1$, is:*

$$p(\boldsymbol{x}_k, \boldsymbol{y}_k|\boldsymbol{y}_{0:k-1}) = p(\boldsymbol{y}_k|\boldsymbol{x}_k, \boldsymbol{y}_{0:k-1})p(\boldsymbol{x}_k|\boldsymbol{y}_{0:k-1}) \underset{2.1}{=} p(\boldsymbol{y}_k|\boldsymbol{x}_k)p(\boldsymbol{x}_k|\boldsymbol{y}_{0:k-1})$$

$$= \mathcal{N}(\boldsymbol{y}_k|\boldsymbol{H}_k\boldsymbol{x}_k, \boldsymbol{R}_k)\mathcal{N}(\boldsymbol{x}_k|\tilde{\boldsymbol{x}}_{k|k-1}, \boldsymbol{\Sigma}_{k|k-1})$$

$$= \mathcal{N}\left(\begin{pmatrix}\boldsymbol{x}_k\\\boldsymbol{y}_k\end{pmatrix}\middle|\begin{pmatrix}\tilde{\boldsymbol{x}}_{k|k-1}\\\boldsymbol{H}_k\tilde{\boldsymbol{x}}_{k|k-1}\end{pmatrix}, \begin{pmatrix}\boldsymbol{\Sigma}_{k|k-1} & \boldsymbol{\Sigma}_{k|k-1}\boldsymbol{H}_k^{\mathrm{T}}\\\boldsymbol{H}_k\boldsymbol{\Sigma}_{k|k-1} & \boldsymbol{H}_k\boldsymbol{\Sigma}_{k|k-1}\boldsymbol{H}_k^{\mathrm{T}} + \boldsymbol{R}_k\end{pmatrix}\right), \tag{2.78}$$

where in the last step we have applied Lemma B.1. Now, we can compute the conditional PDF $p(\boldsymbol{x}_k|\boldsymbol{y}_{0:k})$ from Eq. (2.78) using Lemma B.2:

$$p(\boldsymbol{x}_k|\boldsymbol{y}_{0:k}) = p(\boldsymbol{x}_k|\boldsymbol{y}_k, \boldsymbol{y}_{0:k-1}) = \mathcal{N}(\boldsymbol{x}_k|\tilde{\boldsymbol{x}}_{k|k}, \boldsymbol{\Sigma}_{k|k}), \tag{2.79}$$

$$\tilde{\boldsymbol{x}}_{k|k} = \tilde{\boldsymbol{x}}_{k|k-1} + \boldsymbol{K}_k(\boldsymbol{y}_k - \boldsymbol{H}_k\tilde{\boldsymbol{x}}_{k|k-1}), \tag{2.80}$$

$$\boldsymbol{\Sigma}_{k|k} = \boldsymbol{\Sigma}_{k|k-1} - \boldsymbol{K}_k\boldsymbol{T}_k\boldsymbol{K}_k^{\mathrm{T}} \tag{2.81}$$

where

$$\boldsymbol{T}_k = \boldsymbol{H}_k\boldsymbol{\Sigma}_{k|k-1}\boldsymbol{H}_k^{\mathrm{T}} + \boldsymbol{R}_k, \tag{2.82}$$

$$\boldsymbol{K}_k = \boldsymbol{\Sigma}_{k|k-1}\boldsymbol{H}_k^{\mathrm{T}}\boldsymbol{T}_k^{-1}. \tag{2.83}$$

Model	$\boldsymbol{x}_k = \boldsymbol{A}_{k\text{-}1} \boldsymbol{x}_{k-1} + \boldsymbol{B}_{k\text{-}1} \boldsymbol{u}_{k-1} + \boldsymbol{G}_{k\text{-}1} \boldsymbol{q}_{k-1}, \quad \boldsymbol{q}_k \sim \mathcal{N}(0, \boldsymbol{Q}_k)$ $\boldsymbol{y}_k = \boldsymbol{H}_k \boldsymbol{x}_k + \boldsymbol{r}_k, \quad \boldsymbol{r}_k \sim \mathcal{N}(0, \boldsymbol{R}_k)$
Initialize	$\tilde{\boldsymbol{x}}(t_0) = \tilde{\boldsymbol{x}}_{0|0}$ $\boldsymbol{\Sigma}_{0|0} = E\left\{(\tilde{\boldsymbol{x}}_{0|0} - \boldsymbol{x}_0)(\tilde{\boldsymbol{x}}_{0|0} - \boldsymbol{x}_0)^T\right\}$
Predict	$\tilde{\boldsymbol{x}}_{k|k\text{-}1} = \boldsymbol{A}_{k\text{-}1} \tilde{\boldsymbol{x}}_{k-1|k\text{-}1} + \boldsymbol{B}_{k\text{-}1} \boldsymbol{u}_{k-1}$ $\boldsymbol{\Sigma}_{k|k\text{-}1} = \boldsymbol{A}_{k\text{-}1} \boldsymbol{\Sigma}_{k-1|k\text{-}1} \boldsymbol{A}_{k\text{-}1}^T + \boldsymbol{G}_{k\text{-}1} \boldsymbol{Q}_{k-1} \boldsymbol{G}_{k\text{-}1}^{\mathrm{T}}$
Gain	$\boldsymbol{K}_k = \boldsymbol{\Sigma}_{k|k\text{-}1} \boldsymbol{H}_k^T \left(\boldsymbol{H}_k \boldsymbol{\Sigma}_{k|k\text{-}1} \boldsymbol{H}_k^T + \boldsymbol{R}_k\right)^{-1}$
Update	$\tilde{\boldsymbol{x}}_{k|k} = \tilde{\boldsymbol{x}}_{k|k\text{-}1} + \boldsymbol{K}_k \left(\boldsymbol{y}_k - \boldsymbol{H}_k \tilde{\boldsymbol{x}}_{k|k\text{-}1}\right)$ $\boldsymbol{\Sigma}_{k|k} = \left(\mathbb{I} - \boldsymbol{K}_k \boldsymbol{H}_k\right) \boldsymbol{\Sigma}_{k|k\text{-}1}$

Table 2.1: Summary of the discrete uncorrelated Kalman filter.

2.5 THE CONTINUOUS UNCORRELATED KALMAN FILTER

To transform the discrete KF to a continuous one, we have to take the limit of $\Delta t \to 0$. Therefore, we do not need anymore to distinguish between predict and update steps and can combine their equations for both the estimator and the covariance, summarized in Table 2.1, into a joint *a-priori* recursive form of the KF:

$$\tilde{\boldsymbol{x}}_{k+1} = \boldsymbol{A}_k \tilde{\boldsymbol{x}}_k + \boldsymbol{B}_k \boldsymbol{u}_k + \boldsymbol{A}_k \boldsymbol{K}_k \left(\boldsymbol{y}_k - \boldsymbol{H}_k \tilde{\boldsymbol{x}}_k\right), \tag{2.85}$$

$$\boldsymbol{K}_k = \boldsymbol{\Sigma}_k \boldsymbol{H}_k^T \left(\boldsymbol{H}_k \boldsymbol{\Sigma}_k \boldsymbol{H}_k^T + \boldsymbol{R}_k\right)^{-1}, \tag{2.86}$$

$$\boldsymbol{\Sigma}_{k+1} = \boldsymbol{A}_k \boldsymbol{\Sigma}_k \boldsymbol{A}_k^T - \boldsymbol{A}_k \boldsymbol{K}_k \boldsymbol{H}_k \boldsymbol{\Sigma}_k \boldsymbol{A}_k^T + \boldsymbol{G}_k \boldsymbol{Q}_k \boldsymbol{G}_k^{\mathrm{T}}. \tag{2.87}$$

First, let us derive the expression for the continuous Kalman gain starting from its discrete version in Eq. (2.86) and substituting the first-order approximations for $\boldsymbol{H}_k$ and $\boldsymbol{R}_k$ summarized in Table 1.1:

$$\boldsymbol{K}_k = \boldsymbol{\Sigma}_k \boldsymbol{H}_k^{\mathrm{T}} \left(\boldsymbol{H}_k \boldsymbol{\Sigma}_k \boldsymbol{H}_k^{\mathrm{T}} + \boldsymbol{R}_k \right)^{-1} = \boldsymbol{\Sigma}_k \boldsymbol{H}^{\mathrm{T}}(t_k) \left(\boldsymbol{H}(t_k) \boldsymbol{\Sigma}_k \boldsymbol{H}^{\mathrm{T}}(t_k) + \frac{\boldsymbol{R}(t_k)}{\Delta t} \right)^{-1}, \qquad (2.88)$$

which, since $\boldsymbol{R}(t_k)/\Delta t \gg \boldsymbol{H}(t_k) \boldsymbol{\Sigma}_k \boldsymbol{H}^{\mathrm{T}}(t_k)$ when $\Delta t \to 0$, then it can be simplified to

$$\boldsymbol{K}_k = \boldsymbol{\Sigma}_k \boldsymbol{H}^{\mathrm{T}}(t_k) \boldsymbol{R}(t_k)^{-1} \Delta t. \qquad (2.89)$$

By now dividing the discrete Kalman gain $\boldsymbol{K}_k$ by Δt, we define the continuous Kalman gain as

$$\boldsymbol{K}(t_k) := \frac{\boldsymbol{K}_k}{\Delta t}, \qquad (2.90)$$

which by taking the limit of $\Delta t \to 0$, yields the following expression:

$$\boldsymbol{K}(t) = \boldsymbol{\Sigma}(t) \boldsymbol{H}^{\mathrm{T}}(t) \boldsymbol{R}^{-1}(t). \qquad (2.91)$$

Next, we turn to the derivation of the Riccati equation, i.e. the continuous differential equation for the covariance. The first step is to substitute the first-order approximation of $\boldsymbol{A}_k = (\mathbb{I} + \boldsymbol{F}(t_k)\Delta t)$ from Proposition 1.4, and note that both terms $\boldsymbol{G}_k \boldsymbol{Q}_k \boldsymbol{G}_k^{\mathrm{T}}$ and $\boldsymbol{K}_k$ are of order Δt, as indicated in Eq. (1.179) and Eq. (2.90), respectively. Then,

$$\begin{aligned}
\boldsymbol{\Sigma}_{k+1} &= (\mathbb{I} + \boldsymbol{F}(t_k)\Delta t) \boldsymbol{\Sigma}_k (\mathbb{I} + \boldsymbol{F}^{\mathrm{T}}(t_k)\Delta t) - (\mathbb{I} + \boldsymbol{F}(t_k)\Delta t) \boldsymbol{K}(t_k)\Delta t \boldsymbol{H}(t_k) \boldsymbol{\Sigma}_k (\mathbb{I} + \boldsymbol{F}^{\mathrm{T}}(t_k)\Delta t) \\
&\quad + \boldsymbol{G}(t_k) \boldsymbol{Q}(t_k) \boldsymbol{G}^{\mathrm{T}}(t_k)\Delta t = \boldsymbol{\Sigma}_k + \boldsymbol{F}(t_k)\boldsymbol{\Sigma}_k\Delta t + \boldsymbol{\Sigma}_k \boldsymbol{F}^{\mathrm{T}}(t_k)\Delta t - \boldsymbol{K}(t_k)\boldsymbol{H}(t_k)\boldsymbol{\Sigma}_k\Delta t \\
&\quad + \boldsymbol{G}(t_k) \boldsymbol{Q}(t_k) \boldsymbol{G}^{\mathrm{T}}(t_k)\Delta t. \qquad (2.92)
\end{aligned}$$

By now rearranging terms and taking the limit of $\Delta t \to 0$, we retrieve a Riccati equation for the covariance:

$$\lim_{\Delta t \to 0} \frac{\boldsymbol{\Sigma}_{k+1} - \boldsymbol{\Sigma}_k}{\Delta t} = \frac{\mathrm{d}\boldsymbol{\Sigma}(t)}{\mathrm{d}t} = \boldsymbol{F}(t)\boldsymbol{\Sigma}(t) + \boldsymbol{\Sigma}(t)\boldsymbol{F}^{\mathrm{T}}(t) - \boldsymbol{K}(t)\boldsymbol{H}(t)\boldsymbol{\Sigma}(t) + \boldsymbol{G}(t)\boldsymbol{Q}(t)\boldsymbol{G}^{\mathrm{T}}(t). \quad (2.93)$$

Similarly, we can derive the so-called Kalman-Bucy equation from Eq. (2.85) by substituting the

corresponding first-order approximations:

$$\tilde{\boldsymbol{x}}(t_{k+1}) = (\mathbb{I} + \boldsymbol{F}(t_k)\Delta t)\,\tilde{\boldsymbol{x}}(t_k) + \boldsymbol{B}(t_k)\boldsymbol{u}(t_k)\Delta t + (\mathbb{I} + \boldsymbol{F}(t_k)\Delta t)\,\boldsymbol{K}(t_k)\Delta t\,(\boldsymbol{y}_k - \boldsymbol{H}(t_k)\tilde{\boldsymbol{x}}(t_k)), \quad (2.94)$$

and then rearranging terms to take the limit $\Delta t \to 0$:

$$\begin{aligned}
\frac{\mathrm{d}\tilde{\boldsymbol{x}}}{\mathrm{d}t} &= \lim_{\Delta t \to 0} \frac{\tilde{\boldsymbol{x}}_{k+1} - \tilde{\boldsymbol{x}}_k}{\Delta t} \\
&= \lim_{\Delta t \to 0} \left[\boldsymbol{F}(t_k)\tilde{\boldsymbol{x}}(t_k) + \boldsymbol{B}(t_k)\boldsymbol{u}(t_k) + \boldsymbol{K}(t_k)\left(\frac{1}{\Delta t}\int_{t_{k-1}}^{t_k} \boldsymbol{y}(t)\mathrm{d}t - \boldsymbol{H}(t_k)\tilde{\boldsymbol{x}}(t_k)\right) \right] \\
&= \boldsymbol{F}(t)\tilde{\boldsymbol{x}}(t) + \boldsymbol{B}(t)\boldsymbol{u}(t) + \boldsymbol{K}(t)\left(\lim_{\Delta t \to 0}\frac{1}{\Delta t}\int_{t_{k-1}}^{t_k} \boldsymbol{y}(t)\mathrm{d}t - \boldsymbol{H}(t)\tilde{\boldsymbol{x}}(t)\right) \\
&= {}^* \boldsymbol{F}(t)\tilde{\boldsymbol{x}}(t) + \boldsymbol{B}(t)\boldsymbol{u}(t) + \boldsymbol{K}(t)(\boldsymbol{y}(t) - \boldsymbol{H}(t)\tilde{\boldsymbol{x}}(t))\,. \qquad (2.95)
\end{aligned}$$

This yields a differential equation describing the evolution of the continuous-time estimate.

Model	$\dot{\boldsymbol{x}}(t) = \boldsymbol{F}(t)\boldsymbol{x}(t) + \boldsymbol{B}(t)\boldsymbol{u}(t) + \boldsymbol{G}(t)\boldsymbol{w}(t), \quad \boldsymbol{w}(t) \sim \mathcal{N}(0, \boldsymbol{Q}(t))$ $\boldsymbol{y}(t) = \boldsymbol{H}(t)\boldsymbol{x}(t) + \boldsymbol{v}(t), \quad \boldsymbol{v}(t) \sim \mathcal{N}(0, \boldsymbol{R}(t))$
Initialize	$\tilde{\boldsymbol{x}}(t_0) = \tilde{\boldsymbol{x}}_0$ $\boldsymbol{\Sigma}_0 = \mathbb{E}\left[(\tilde{\boldsymbol{x}}_0 - \boldsymbol{x}_0)(\tilde{\boldsymbol{x}}_0 - \boldsymbol{x}_0)^{\mathrm{T}}\right]$
Gain	$\boldsymbol{K}(t) = \boldsymbol{\Sigma}(t)\,\boldsymbol{H}^{\mathrm{T}}(t)\boldsymbol{R}^{-1}(t)$
Covariance	$\dot{\boldsymbol{\Sigma}}(t) = \boldsymbol{F}(t)\boldsymbol{\Sigma}(t) + \boldsymbol{\Sigma}(t)\,\boldsymbol{F}^{\mathrm{T}}(t) - \boldsymbol{K}(t)\boldsymbol{H}(t)\boldsymbol{\Sigma}(t) + \boldsymbol{G}(t)\boldsymbol{Q}(t)\boldsymbol{G}^{\mathrm{T}}(t)$
Estimate	$\dot{\tilde{\boldsymbol{x}}}(t) = \boldsymbol{F}(t)\tilde{\boldsymbol{x}}(t) + \boldsymbol{B}(t)\boldsymbol{u}(t) + \boldsymbol{K}(t)(\boldsymbol{y}(t) - \boldsymbol{H}(t)\tilde{\boldsymbol{x}}(t))$

Table 2.2: Summary of the continuous uncorrelated Kalman filter.

${}^*\lim_{\Delta t \to 0} \frac{1}{\Delta t}\int_{t_k - \Delta t}^{t_k} \boldsymbol{y}(t)\mathrm{d}t = \boldsymbol{y}(t_k).$

In many systems, the measurement noise is uncorrelated with the process noise. However, in quantum systems, the measurement can act back onto the system — an effect referred to as measurement back-action — and introduce correlation between the noises. As a result, the standard KF theory has to be extended to properly account for correlated measurement and process noises [29]. This requires allowing the zero-mean Gaussian noise processes, $\boldsymbol{w}(t)$ and $\boldsymbol{v}(t)$, present in the LG model of Eqs. (1.160-1.161) to be cross-correlated:

$$\mathbb{E}\big[\boldsymbol{w}(t)\boldsymbol{v}^{\mathrm{T}}(s)\big] = \boldsymbol{S}(t)\delta(t-s), \tag{2.96}$$

where the matrix $\boldsymbol{S}(t)$ is not necessarily symmetric. To derive the KF equations for a system with correlated measurement and process noises, we proceed in the same way as we did for the case of uncorrelated noise, with the exception that first we must de-correlate the measurement and process noise [29, 87].

In Sec. 1.3.9 we established the equivalency between the continuous model in Eqs. (1.160-1.161) and the following discrete model, which now includes correlated process and measurement noise:

$$\begin{cases} \boldsymbol{x}_k = \boldsymbol{A}_{k-1}\boldsymbol{x}_{k-1} + \boldsymbol{B}_{k-1}\boldsymbol{u}_{k-1} + \boldsymbol{G}_{k-1}\boldsymbol{q}_{k-1}, \\ \boldsymbol{y}_k = \boldsymbol{H}_k\boldsymbol{x}_k + \boldsymbol{r}_k, \end{cases} \tag{2.97}$$

$$\mathbb{E}\big[\boldsymbol{q}_k\boldsymbol{q}_\ell^{\mathrm{T}}\big] = \boldsymbol{Q}_k\delta_{k\ell}, \tag{2.98}$$

$$\mathbb{E}\big[\boldsymbol{r}_k\boldsymbol{r}_\ell^{\mathrm{T}}\big] = \boldsymbol{R}_k\delta_{k\ell}, \tag{2.99}$$

$$\mathbb{E}\big[\boldsymbol{q}_{k-1}\boldsymbol{r}_\ell^{\mathrm{T}}\big] = \boldsymbol{S}_k\delta_{k\ell}. \tag{2.100}$$

Starting from this discrete model, we de-correlate the measurement and process equations by introducing the measurement term into the process equation:

$$\boldsymbol{x}_k = \boldsymbol{A}_{k-1}\boldsymbol{x}_{k-1} + \boldsymbol{B}_{k-1}\boldsymbol{u}_{k-1} + \boldsymbol{G}_{k-1}\boldsymbol{q}_{k-1} + \boldsymbol{D}_{k-1}\left(\boldsymbol{y}_{k-1} - \boldsymbol{H}_{k-1}\boldsymbol{x}_{k-1} - \boldsymbol{r}_{k-1}\right) \tag{2.101}$$

$$= \left(\boldsymbol{A}_{k-1} - \boldsymbol{D}_{k-1}\boldsymbol{H}_{k-1}\right)\boldsymbol{x}_{k-1} + \boldsymbol{B}_{k-1}\boldsymbol{u}_{k-1} + \boldsymbol{D}_{k-1}\boldsymbol{y}_{k-1} + \boldsymbol{Z}_{k-1}, \tag{2.102}$$

where $\boldsymbol{D}_{k-1}$ is an arbitrary matrix, since $\boldsymbol{y}_{k-1} - \boldsymbol{H}_{k-1}\boldsymbol{x}_{k-1} - \boldsymbol{r}_{k-1} = 0$. Moreover, we define $\boldsymbol{Z}_{k-1} :=$ $\boldsymbol{G}_{k-1}\boldsymbol{q}_{k-1} - \boldsymbol{D}_{k-1}\boldsymbol{r}_{k-1}$ to be the effective process noise. In order to ensure that the correlation be-

tween the measurement $\boldsymbol{r}_k$ and the new process noise $\boldsymbol{Z}_{k-1}$ vanishes, we set

$$\boldsymbol{D}_k := \boldsymbol{G}_k \boldsymbol{S}_k \boldsymbol{R}_k^{-1} \tag{2.103}$$

such that

$$\mathbb{E}\left[\boldsymbol{Z}_{k-1}\boldsymbol{r}_k^{\mathrm{T}}\right] = \left(\boldsymbol{G}_{k-1}\boldsymbol{S}_{k-1} - \boldsymbol{D}_{k-1}\boldsymbol{R}_{k-1}\right) = 0. \tag{2.104}$$

Then, $\boldsymbol{Z}_{k-1}$ can be rewritten as:

$$\boldsymbol{Z}_{k-1} = \boldsymbol{G}_{k-1}\boldsymbol{q}_{k-1} - \boldsymbol{G}_{k-1}\boldsymbol{S}_{k-1}\boldsymbol{R}_{k-1}^{-1}\boldsymbol{r}_{k-1} = \boldsymbol{G}_{k-1}\left(\boldsymbol{q}_{k-1} - \boldsymbol{S}_{k-1}\boldsymbol{R}_{k-1}^{-1}\boldsymbol{r}_{k-1}\right) = \boldsymbol{G}_{k-1}\boldsymbol{U}_{k-1}, \tag{2.105}$$

with $\boldsymbol{U}_{k-1} := \boldsymbol{q}_{k-1} - \boldsymbol{S}_{k-1}\boldsymbol{R}^{-1}\boldsymbol{r}_{k-1}$. This results in an uncorrelated model:

$$\boldsymbol{x}_k = \left(\boldsymbol{A}_{k-1} - \boldsymbol{G}_{k-1}\boldsymbol{S}_{k-1}\boldsymbol{R}_{k-1}^{-1}\boldsymbol{H}_{k-1}\right)\boldsymbol{x}_{k-1} + \boldsymbol{B}_{k-1}\boldsymbol{u}_{k-1} + \boldsymbol{G}_{k-1}\boldsymbol{S}_{k-1}\boldsymbol{R}_{k-1}^{-1}\boldsymbol{y}_{k-1} + \boldsymbol{G}_{k-1}\boldsymbol{U}_{k-1} \tag{2.106}$$

$$\boldsymbol{y}_k = \boldsymbol{H}_k\boldsymbol{x}_k + \boldsymbol{r}_k, \tag{2.107}$$

where, in order for the process noise $\boldsymbol{U}_k$ to be uncorrelated w.r.t. the measurement noise, the process must now depend on the measurement $\boldsymbol{y}_{k-1}$. Additionally, the variance of the process noise $\boldsymbol{U}_k$ now reads as:

$$\mathbb{E}\left[\boldsymbol{U}_k\boldsymbol{U}_k^{\mathrm{T}}\right] = \boldsymbol{Q}_{k-1} - \boldsymbol{S}_{k-1}\boldsymbol{R}_{k-1}^{-1}\boldsymbol{S}_{k-1}^{\mathrm{T}}. \tag{2.108}$$

To derive the prediction and update probability densities, $p(\boldsymbol{x}_k|\boldsymbol{y}_{0:k-1})$ and $p(\boldsymbol{x}_k|\boldsymbol{y}_{0:k})$, for the system described by Eqs. (2.106-2.107), we follow the approach of Sec. 2.4. As in that derivation, both probabilities turn out to be Gaussian, with means and covariances following predict and update rules. These equations describe the optimal evolution of the estimator and its covariance over time, since the mean of the posterior is the optimal estimator minimizing the MSE.

Additionally, it is important to note that the de-correlation trick exclusively affects the process equation, i.e. Eq. (2.106). Thus, we expect only the prediction step to be modified, leaving the update step unchanged. In particular, the predict step, i.e. Eq. (2.106), now depends on the measurement outcome $\boldsymbol{y}_{k-1}$. That might seem counter-intuitive, as measurement outcomes typically only affect the update step, not the prediction. Nonetheless, this is irrelevant for the derivation of the continuous correlated KF, since the equations for the predict and update step have to be merged.

With this understanding, let us derive an expression for the Gaussian probability density for the prediction step, $p(\boldsymbol{x}_k|\boldsymbol{y}_{0:k-1}) := \mathcal{N}(\boldsymbol{x}_k|\tilde{\boldsymbol{x}}_{k|k-1}, \boldsymbol{\Sigma}_{k|k-1})$. This derivation essentially mirrors the

one in Eq. (2.72) with a slight modification:

$$p(\boldsymbol{x}_k|\boldsymbol{y}_{0:k-1}) \underset{1.1}{=} \int p(\boldsymbol{x}_k, \boldsymbol{x}_{k-1}|\boldsymbol{y}_{0:k-1})\mathrm{d}\boldsymbol{x}_{k-1}$$

$$\underset{1.2}{=} \int p(\boldsymbol{x}_k|\boldsymbol{x}_{k-1}, \boldsymbol{y}_{0:k-1})p(\boldsymbol{x}_{k-1}|\boldsymbol{y}_{0:k-1})\mathrm{d}\boldsymbol{x}_{k-1}$$

$$= \int p(\boldsymbol{x}_k|\boldsymbol{x}_{k-1}, \boldsymbol{y}_{k-1})p(\boldsymbol{x}_{k-1}|\boldsymbol{y}_{0:k-1})\mathrm{d}\boldsymbol{x}_{k-1}. \qquad (2.109)$$

Here, $p(\boldsymbol{x}_k|\boldsymbol{x}_{k-1}, \boldsymbol{y}_{0:k-1})$ is simplified to $p(\boldsymbol{x}_k|\boldsymbol{x}_{k-1}, \boldsymbol{y}_{k-1})$ instead of $p(\boldsymbol{x}_k|\boldsymbol{x}_{k-1})$, since unlike in Eq. (2.72), the state $\boldsymbol{x}_k$ not only depends on $\boldsymbol{x}_{k-1}$ but also on $\boldsymbol{y}_{k-1}$, as evident from Eq. (2.106) [87]. Fortunately, this still fulfills Markovianity and $p(\boldsymbol{x}_k|\boldsymbol{x}_{k-1}, \boldsymbol{y}_{k-1})$ is still a Gaussian distribution:

$$p(\boldsymbol{x}_k|\boldsymbol{x}_{k-1}, \boldsymbol{y}_{k-1}) =$$
$$= \mathcal{N}(\boldsymbol{x}_k|\boldsymbol{A}'_{k-1}\boldsymbol{x}_{k-1} + \boldsymbol{B}_{k-1}\boldsymbol{u}_{k-1} + \boldsymbol{G}_{k-1}\boldsymbol{S}_{k-1}\boldsymbol{R}_{k-1}^{-1}\boldsymbol{y}_{k-1}, \boldsymbol{G}_{k-1}(\boldsymbol{Q}_{k-1} - \boldsymbol{S}_{k-1}\boldsymbol{R}_{k-1}^{-1}\boldsymbol{S}_{k-1}^{\mathrm{T}})\boldsymbol{G}_{k-1}), \qquad (2.110)$$

where $\boldsymbol{A}'_{k-1} = \boldsymbol{A}_{k-1} - \boldsymbol{G}_{k-1}\boldsymbol{S}_{k-1}\boldsymbol{R}_{k-1}^{-1}\boldsymbol{H}_{k-1}$. Assuming that the posterior at the previous time step is Gaussian, as given in Eq. (2.71): $p(\boldsymbol{x}_{k-1}|\boldsymbol{y}_{0:k-1}) = \mathcal{N}(\boldsymbol{x}_{k-1}|\tilde{\boldsymbol{x}}_{k-1|k-1}, \boldsymbol{\Sigma}_{k-1|k-1})$, we can use Lemma B.1 to write the joint probability distribution $p(\boldsymbol{x}_k, \boldsymbol{x}_{k-1}|\boldsymbol{y}_{0:k-1})$ as:

$$p(\boldsymbol{x}_k, \boldsymbol{x}_{k-1}|\boldsymbol{y}_{0:k-1}) = \mathcal{N}\left(\begin{pmatrix} \boldsymbol{x}_{k-1} \\ \boldsymbol{x}_k \end{pmatrix}\middle| \boldsymbol{m}_{k|k-1}, \boldsymbol{P}_{k|k-1}\right), \qquad (2.111)$$

where

$$\boldsymbol{m}_{k|k-1} = \begin{pmatrix} \tilde{\boldsymbol{x}}_{k-1|k-1} \\ \boldsymbol{A}'_{k-1}\tilde{\boldsymbol{x}}_{k-1|k-1} + \boldsymbol{B}_{k-1}\boldsymbol{u}_{k-1} + \boldsymbol{G}_{k-1}\boldsymbol{S}_{k-1}\boldsymbol{R}_{k-1}^{-1}\boldsymbol{y}_{k-1} \end{pmatrix}, \qquad (2.112)$$

and

$$\boldsymbol{P}_{k|k-1} = \begin{pmatrix} \boldsymbol{\Sigma}_{k-1|k-1} & \boldsymbol{\Sigma}_{k-1|k-1}\boldsymbol{A}_{k-1}'^{\mathrm{T}} \\ \boldsymbol{A}'_{k-1}\boldsymbol{\Sigma}_{k-1|k-1} & \boldsymbol{A}'_{k-1}\boldsymbol{\Sigma}_{k-1|k-1}\boldsymbol{A}_{k-1}'^{\mathrm{T}} + \boldsymbol{G}_{k-1}(\boldsymbol{Q}_{k-1} - \boldsymbol{S}_{k-1}\boldsymbol{R}_{k-1}^{-1}\boldsymbol{S}_{k-1}^{\mathrm{T}})\boldsymbol{G}_{k-1}^{\mathrm{T}} \end{pmatrix}. \qquad (2.113)$$

To now obtain the predict probability density $p(\boldsymbol{x}_k|\boldsymbol{y}_{0:k-1})$ from the joint probability $p(\boldsymbol{x}_k, \boldsymbol{x}_{k-1}|\boldsymbol{y}_{0:k-1})$,

we compute its marginal following Lemma B.2:

$$p(\boldsymbol{x}_k|\boldsymbol{y}_{0:k-1}) = \mathcal{N}(\boldsymbol{x}_k|\tilde{\boldsymbol{x}}_{k|k-1}, \boldsymbol{\Sigma}_{k|k-1}), \tag{2.114}$$

$$\tilde{\boldsymbol{x}}_{k|k-1} = \left(\boldsymbol{A}_{k-1} - \boldsymbol{G}_{k-1}\boldsymbol{S}_{k-1}\boldsymbol{R}_{k-1}^{-1}\boldsymbol{H}_{k-1}\right)\tilde{\boldsymbol{x}}_{k-1|k-1} + \boldsymbol{B}_{k-1}\boldsymbol{u}_{k-1} + \boldsymbol{G}_{k-1}\boldsymbol{S}_{k-1}\boldsymbol{R}_{k-1}^{-1}\boldsymbol{y}_{k-1}, \tag{2.115}$$

$$\boldsymbol{\Sigma}_{k|k-1} = \left(\boldsymbol{A}_{k-1} - \boldsymbol{G}_{k-1}\boldsymbol{S}_{k-1}\boldsymbol{R}_{k-1}^{-1}\boldsymbol{H}_{k-1}\right)\boldsymbol{\Sigma}_{k-1|k-1}\left(\boldsymbol{A}_{k-1} - \boldsymbol{G}_{k-1}\boldsymbol{S}_{k-1}\boldsymbol{R}_{k-1}^{-1}\boldsymbol{H}_{k-1}\right)^{\mathrm{T}}$$
$$+ \boldsymbol{G}_{k-1}\left(\boldsymbol{Q}_{k-1} - \boldsymbol{S}_{k-1}\boldsymbol{R}_{k-1}^{-1}\boldsymbol{S}_{k-1}^{\mathrm{T}}\right)\boldsymbol{G}_{k-1}^{\mathrm{T}}, \tag{2.116}$$

whose mean also depends on the measurement outcome $\boldsymbol{y}_{k-1}$.

Since the measurement equation remains unchanged when de-correlating the process and measurement noise, the update probability density matches that of the uncorrelated case in Eq. (2.79):

$$p(\boldsymbol{x}_k|\boldsymbol{y}_{0:k}) = \mathcal{N}\left(\boldsymbol{x}_k|\tilde{\boldsymbol{x}}_{k|k}, \boldsymbol{\Sigma}_{k|k}\right), \tag{2.117}$$

$$\tilde{\boldsymbol{x}}_{k|k} = \tilde{\boldsymbol{x}}_{k|k-1} + \boldsymbol{K}_k(\boldsymbol{y}_k - \boldsymbol{H}_k\tilde{\boldsymbol{x}}_{k|k-1}), \tag{2.118}$$

$$\boldsymbol{\Sigma}_{k|k} = (\mathbb{I} - \boldsymbol{K}_k\boldsymbol{H}_k)\boldsymbol{\Sigma}_{k|k-1}, \tag{2.119}$$

$$\boldsymbol{K}_k = \boldsymbol{\Sigma}_{k|k-1}\boldsymbol{H}_k^{\mathrm{T}}\left(\boldsymbol{H}_k\boldsymbol{\Sigma}_{k|k-1}\boldsymbol{H}_k^{\mathrm{T}} + \boldsymbol{R}_k\right)^{-1}. \tag{2.120}$$

Let us now merge the predict and update equations for the estimate, i.e. Eq. (2.118) and Eq. (2.115), into a joint recursive form to derive its continuous counterpart:

$$\tilde{\boldsymbol{x}}_{k+1|k} = \left(\boldsymbol{A}_k - \boldsymbol{G}_k\boldsymbol{S}_k\boldsymbol{R}_k^{-1}\boldsymbol{H}_k\right)\tilde{\boldsymbol{x}}_{k|k} + \boldsymbol{B}_k\boldsymbol{u}_k + \boldsymbol{G}_k\boldsymbol{S}_k\boldsymbol{R}_k^{-1}\boldsymbol{y}_k$$
$$= \left(\boldsymbol{A}_k - \boldsymbol{G}_k\boldsymbol{S}_k\boldsymbol{R}_k^{-1}\boldsymbol{H}_k\right)\left(\tilde{\boldsymbol{x}}_{k|k-1} + \boldsymbol{K}_k(\boldsymbol{y}_k - \boldsymbol{H}_k\tilde{\boldsymbol{x}}_{k|k-1})\right) + \boldsymbol{B}_k\boldsymbol{u}_k + \boldsymbol{G}_k\boldsymbol{S}_k\boldsymbol{R}_k^{-1}\boldsymbol{y}_k$$
$$= \boldsymbol{A}_k\tilde{\boldsymbol{x}}_{k|k-1} + \boldsymbol{B}_k\boldsymbol{u}_k + \left(\boldsymbol{A}_k\boldsymbol{K}_k - \boldsymbol{G}_k\boldsymbol{S}_k\boldsymbol{R}_k^{-1}\left(\boldsymbol{H}_k\boldsymbol{K}_k - \mathbb{I}\right)\right)\left(\boldsymbol{y}_k - \boldsymbol{H}_k\tilde{\boldsymbol{x}}_{k|k-1}\right). \tag{2.121}$$

Next, if we substitute the first-order approximations in Δt summarized in Table 1.1 into the equation above, we retrieve:

$$\tilde{\boldsymbol{x}}(t_{k+1}) = \left(\mathbb{I} + \boldsymbol{F}(t_k)\Delta t\right)\tilde{\boldsymbol{x}}(t_k) + \boldsymbol{B}(t_k)\boldsymbol{u}(t_k)\Delta t + \left(\frac{1}{\Delta t}\int_{t_{k-1}}^{t_k}\boldsymbol{y}(t)\mathrm{d}t - \boldsymbol{H}(t_k)\tilde{\boldsymbol{x}}(t_k)\right) \times$$
$$\times \left(\left(\mathbb{I} + \boldsymbol{F}(t_k)\Delta t\right)\boldsymbol{K}(t_k)\Delta t - \boldsymbol{G}(t_k)\Delta t\frac{\boldsymbol{S}(t_k)}{\Delta t}\boldsymbol{R}^{-1}(t_k)\Delta t\left(\boldsymbol{H}(t_k)\boldsymbol{K}(t_k)\Delta t - \mathbb{I}\right)\right)$$
$$= \tilde{\boldsymbol{x}}(t_k) + \boldsymbol{F}(t_k)\tilde{\boldsymbol{x}}(t_k)\,\Delta t + \boldsymbol{B}(t_k)\boldsymbol{u}(t_k)\Delta t + \left(\frac{1}{\Delta t}\int_{t_{k-1}}^{t_k}\boldsymbol{y}(t)\mathrm{d}t - \boldsymbol{H}(t_k)\tilde{\boldsymbol{x}}(t_k)\right) \times$$
$$\times \left(\boldsymbol{K}(t_k) - \boldsymbol{G}(t_k)\boldsymbol{S}(t_k)\boldsymbol{R}^{-1}(t_k)\right)\Delta t + \mathcal{O}(\Delta t). \tag{2.122}$$

82

Then, by taking the limit of $\Delta t \to 0$, we derive the Kalman-Bucy equation for a correlated system:

$$
\begin{aligned}
\frac{\mathrm{d}\tilde{\boldsymbol{x}}}{\mathrm{d}t} &= \lim_{\Delta t \to 0} \frac{\tilde{\boldsymbol{x}}(t_{k+1}) - \tilde{\boldsymbol{x}}(t_k)}{\Delta t} \\
&= \boldsymbol{F}(t)\tilde{\boldsymbol{x}}(t) + \boldsymbol{B}(t)\boldsymbol{u}(t) + \left(\boldsymbol{K}(t) - \boldsymbol{G}(t)\boldsymbol{S}(t)\boldsymbol{R}^{-1}(t)\right)\left(\boldsymbol{y}(t) - \boldsymbol{H}(t)\tilde{\boldsymbol{x}}(t)\right),
\end{aligned} \tag{2.123}
$$

leading us to redefine the Kalman gain as

$$
\boldsymbol{K}(t) := \boldsymbol{K}(t) - \boldsymbol{G}(t)\boldsymbol{S}(t)\boldsymbol{R}^{-1}(t) = \left(\boldsymbol{\Sigma}(t)\boldsymbol{H}^{\mathrm{T}}(t) - \boldsymbol{G}(t)\boldsymbol{S}(t)\right)\boldsymbol{R}^{-1}(t). \tag{2.124}
$$

Next, we substitute Eq. (2.119) into Eq. (2.116) in order to obtain a joint recursive equation for the covariance $\boldsymbol{\Sigma}_k$:

$$
\begin{aligned}
\boldsymbol{\Sigma}_{k+1|k} &= \left(\boldsymbol{A}_k - \boldsymbol{G}_k\boldsymbol{S}_k\boldsymbol{R}_k^{-1}\boldsymbol{H}_k\right)\left(\mathbb{I} - \boldsymbol{K}_k\boldsymbol{H}_k\right)\boldsymbol{\Sigma}_{k|k-1}\left(\boldsymbol{A}_k - \boldsymbol{G}_k\boldsymbol{S}_k\boldsymbol{R}_k^{-1}\boldsymbol{H}_k\right)^{\mathrm{T}} \\
&\quad + \boldsymbol{G}_k\left(\boldsymbol{Q}_k - \boldsymbol{S}_k\boldsymbol{R}_k^{-1}\boldsymbol{S}_k^{\mathrm{T}}\right)\boldsymbol{G}_k^{\mathrm{T}}.
\end{aligned} \tag{2.125}
$$

If now we substitute each of its terms by its corresponding first-order approximations:

$$
\begin{aligned}
\boldsymbol{\Sigma}(t_{k+1}) &= \left(\mathbb{I} + \boldsymbol{F}'(t_k)\Delta t\right)\boldsymbol{\Sigma}(t_k)\left(\mathbb{I} + \boldsymbol{F}'^{\mathrm{T}}(t_k)\Delta t\right) \\
&\quad + \left(\mathbb{I} + \boldsymbol{F}'(t_k)\Delta t\right)\boldsymbol{\Sigma}(t_k)\boldsymbol{H}^{\mathrm{T}}(t_k)\boldsymbol{R}^{-1}(t_k)\Delta t\,\boldsymbol{H}(t_k)\boldsymbol{\Sigma}(t_k)\left(\mathbb{I} + \boldsymbol{F}'^{\mathrm{T}}(t_k)\Delta t\right) \\
&\quad + \boldsymbol{G}(t_k)\left(\boldsymbol{Q}(t_k) - \boldsymbol{S}(t_k)\boldsymbol{R}^{-1}(t_k)\boldsymbol{S}^{\mathrm{T}}(t_k)\right)\boldsymbol{G}^{\mathrm{T}}(t_k)\Delta t \\
&= \boldsymbol{\Sigma}(t_k) + \boldsymbol{F}'(t_k)\boldsymbol{\Sigma}(t_k)\,\Delta t + \boldsymbol{\Sigma}(t_k)\boldsymbol{F}'^{\mathrm{T}}(t_k)\Delta t \\
&\quad + \boldsymbol{\Sigma}(t_k)\boldsymbol{H}^{\mathrm{T}}(t_k)\boldsymbol{R}^{-1}(t_k)\boldsymbol{H}(t_k)\boldsymbol{\Sigma}(t_k)\Delta t \\
&\quad + \boldsymbol{G}(t_k)\left(\boldsymbol{Q}(t_k) - \boldsymbol{S}(t_k)\boldsymbol{R}^{-1}(t_k)\boldsymbol{S}^{\mathrm{T}}(t_k)\right)\boldsymbol{G}^{\mathrm{T}}(t_k)\Delta t + \mathcal{O}(\Delta t),
\end{aligned} \tag{2.126}
$$

where $\boldsymbol{F}'(t_k) = \boldsymbol{F}(t_k) - \boldsymbol{G}(t_k)\boldsymbol{S}(t_k)\boldsymbol{R}^{-1}(t_k)\boldsymbol{H}(t_k)$, and take the limit of Δt approaching zero, we retrieve

$$
\begin{aligned}
\frac{\mathrm{d}\boldsymbol{\Sigma}}{\mathrm{d}t} &= \lim_{\Delta t \to 0} \frac{\boldsymbol{\Sigma}(t_{k+1}) - \boldsymbol{\Sigma}(t_k)}{\Delta t} = \left(\boldsymbol{F}(t) - \boldsymbol{G}(t)\boldsymbol{S}(t)\boldsymbol{R}^{-1}(t)\boldsymbol{H}(t)\right)\boldsymbol{\Sigma}(t) \\
&\quad + \boldsymbol{\Sigma}(t)\left(\boldsymbol{F}(t) - \boldsymbol{G}(t)\boldsymbol{S}(t)\boldsymbol{R}^{-1}(t)\boldsymbol{H}(t)\right)^{\mathrm{T}} + \boldsymbol{\Sigma}(t)\boldsymbol{H}^{\mathrm{T}}(t)\boldsymbol{R}^{-1}(t)\boldsymbol{H}(t)\boldsymbol{\Sigma}(t) \\
&\quad + \boldsymbol{G}(t)\left(\boldsymbol{Q}(t) - \boldsymbol{S}(t)\boldsymbol{R}^{-1}(t)\boldsymbol{S}^{\mathrm{T}}(t)\right)\boldsymbol{G}^{\mathrm{T}}(t),
\end{aligned} \tag{2.127}
$$

which can be rewritten as

$$\frac{\mathrm{d}\boldsymbol{\Sigma}}{\mathrm{d}t} = \boldsymbol{F}(t)\boldsymbol{\Sigma}(t) + \boldsymbol{\Sigma}(t)\boldsymbol{F}^{\mathrm{T}}(t) - \boldsymbol{K}(t)\boldsymbol{R}(t)\boldsymbol{K}^{\mathrm{T}}(t) + \boldsymbol{G}(t)\boldsymbol{Q}(t)\boldsymbol{G}^{\mathrm{T}}(t), \tag{2.128}$$

with the Kalman gain being $\boldsymbol{K}(t) = \left(\boldsymbol{\Sigma}(t)\boldsymbol{H}^{\mathrm{T}}(t) - \boldsymbol{G}(t)\boldsymbol{S}(t)\right)\boldsymbol{R}^{-1}(t)$, as redefined in Eq. (2.124).

Model	$\dot{\boldsymbol{x}}(t) = \boldsymbol{F}(t)\boldsymbol{x}(t) + \boldsymbol{B}(t)\boldsymbol{u}(t) + \boldsymbol{G}(t)\boldsymbol{w}(t), \quad \boldsymbol{w}(t) \sim \mathcal{N}(0, \boldsymbol{Q}(t))$ $\boldsymbol{y}(t) = \boldsymbol{H}(t)\boldsymbol{x}(t) + \boldsymbol{v}(t), \quad \boldsymbol{v}(t) \sim \mathcal{N}(0, \boldsymbol{R}(t))$ $\mathbb{E}\left[\boldsymbol{w}(t)\boldsymbol{v}^{\mathrm{T}}(s)\right] = \boldsymbol{S}(t)\delta(t-s)$
Initialize	$\tilde{\boldsymbol{x}}(t_0) = \tilde{\boldsymbol{x}}_0$ $\boldsymbol{\Sigma}_0 = \mathbb{E}\left[(\tilde{\boldsymbol{x}}_0 - \boldsymbol{x}_0)(\tilde{\boldsymbol{x}}_0 - \boldsymbol{x}_0)^{\mathrm{T}}\right]$
Gain	$\boldsymbol{K}(t) = \left(\boldsymbol{\Sigma}(t)\boldsymbol{H}^{\mathrm{T}}(t) - \boldsymbol{G}(t)\boldsymbol{S}(t)\right)\boldsymbol{R}^{-1}(t)$
Covariance	$\dot{\boldsymbol{\Sigma}}(t) = \boldsymbol{F}(t)\boldsymbol{\Sigma}(t) + \boldsymbol{\Sigma}(t)\boldsymbol{F}^{\mathrm{T}}(t) - \boldsymbol{K}(t)\boldsymbol{R}(t)\boldsymbol{K}^{\mathrm{T}}(t) + \boldsymbol{G}(t)\boldsymbol{Q}(t)\boldsymbol{G}^{\mathrm{T}}(t)$
Estimate	$\dot{\tilde{\boldsymbol{x}}}(t) = \boldsymbol{F}(t)\tilde{\boldsymbol{x}}(t) + \boldsymbol{B}(t)\boldsymbol{u}(t) + \boldsymbol{K}(t)\left(\boldsymbol{y}(t) - \boldsymbol{H}(t)\tilde{\boldsymbol{x}}(t)\right)$

Table 2.3: Summary of the continuous correlated Kalman filter.

2.6.1 ORTHOGONALITY PRINCIPLE

An important property used later to prove the optimality of the linear-quadratic Gaussian controller is the orthogonality between the MMSE estimate and its error, which also applies to the KF estimate since its the MMSE estimator for LG systems.

Property 2.2 (Orthogonality Principle of the MMSE estimate). *Let $\boldsymbol{x}(t)$ be the true state of a system and $\tilde{\boldsymbol{x}}(t)$ its MMSE estimate. Define the estimation error as:*

$$\boldsymbol{e}(t) = \tilde{\boldsymbol{x}}(t) - \boldsymbol{x}(t). \tag{2.129}$$

Then, the orthogonality principle states that the error $\boldsymbol{e}(t)$ is uncorrelated with the estimate $\tilde{\boldsymbol{x}}(t)$,

expressed by:

$$\mathbb{E}\left[\tilde{\boldsymbol{x}}(t)\,\boldsymbol{e}^{\mathrm{T}}(t)\right] = 0. \tag{2.130}$$

In other words, the expectation value of the cross-correlation between the estimate and its error is zero.

Proof. *To start, let us first recall that the optimal estimate minimizing the MSE or quadratic cost function is the mean of the posterior* (2.14), *i.e. the MMSE estimator:*

$$\tilde{\boldsymbol{x}}(t) = \int \boldsymbol{x}(t)\,p(\boldsymbol{x}|\boldsymbol{y}_{\leq t})\mathrm{d}\boldsymbol{x}. \tag{2.131}$$

Therefore, we can rewrite the expectation product of the

$$\mathbb{E}\left[\tilde{\boldsymbol{x}}(t)\,\boldsymbol{e}^{\mathrm{T}}(t)\right] = \mathbb{E}\left[\tilde{\boldsymbol{x}}(t)\,\tilde{\boldsymbol{x}}^{\mathrm{T}}(t)\right] - \mathbb{E}\left[\tilde{\boldsymbol{x}}(t)\,\boldsymbol{x}^{\mathrm{T}}(t)\right] \tag{2.132}$$

$$= \mathbb{E}\left[\tilde{\boldsymbol{x}}(t)\left(\int \boldsymbol{x}(t)\,p(\boldsymbol{x}|\boldsymbol{y}_{\leq t})\mathrm{d}\boldsymbol{x}\right)^{\mathrm{T}}\right] - \mathbb{E}\left[\tilde{\boldsymbol{x}}(t)\,\boldsymbol{x}^{\mathrm{T}}(t)\right]. \tag{2.133}$$

Note that the estimator $\tilde{\boldsymbol{x}}$ is ultimately a function of $\boldsymbol{y}$. Therefore,

$$\mathbb{E}\left[\tilde{\boldsymbol{x}}(t)\,\boldsymbol{e}^{\mathrm{T}}(t)\right] = \mathbb{E}\left[\tilde{\boldsymbol{x}}(t)\left(\int \boldsymbol{x}(t)\,p(\boldsymbol{x}|\boldsymbol{y}_{\leq t})\mathrm{d}\boldsymbol{x}\right)^{\mathrm{T}}\right] - \mathbb{E}\left[\tilde{\boldsymbol{x}}(t)\,\boldsymbol{x}^{\mathrm{T}}(t)\right] \tag{2.134}$$

$$= \int \tilde{\boldsymbol{x}}(t)\,\boldsymbol{x}^{\mathrm{T}}(t)\,p(\boldsymbol{x},\boldsymbol{y}_{\leq t})\,\mathrm{d}\boldsymbol{x}\,\mathrm{d}\boldsymbol{y}_{\leq t} - \int \tilde{\boldsymbol{x}}(t)\,\boldsymbol{x}^{\mathrm{T}}(t)p(\boldsymbol{x}|\boldsymbol{y}_{\leq t})p(\boldsymbol{y}_{\leq t})\,\mathrm{d}\boldsymbol{x}\,\mathrm{d}\boldsymbol{y}_{\leq t} = 0. \tag{2.135}$$

2.7 Extended Kalman filter

The derivation of the KF so far has relied on the assumption that the model is linear. If that is not the case and the model is nonlinear, a Gaussian input does not necessarily produce a Gaussian output, unlike in the linear scenario. As a result, the linear KF may no longer be the optimal estimator, as there might be nonlinear filters that produce a better estimate. There are a wide range of nonlinear filters one could choose from, but the simplest, most natural step is to consider nonlinear extensions of the KF, such as the linearized or extended Kalman filter (EKF), even though they are not guaranteed to provide an optimal estimate.

To derive the linearized KF equations, we must first consider the nonlinear model given by a set of coupled nonlinear stochastic equations of the form:

$$\dot{x}(t) = f[x(t), u(t), w(t), t], \tag{2.136}$$

$$y(t) = h[x(t), v(t), t], \tag{2.137}$$

where f and h are both continuously differentiable nonlinear functions, $x(t)$ and $u(t)$ are the state and control vectors, and $w(t)$ and $v(t)$ denote Langevin-noise terms, whose noise covariances fulfill the same relations as the ones specified for a linear and correlated system, detailed in Eqs. (1.157-1.158) and Eq. (2.96).

Next, we expand the process and measurement equations around a nominal trajectory:

$$(x_0(t), u_0(t), w_0(t), v_0(t)), \tag{2.138}$$

i.e. nominal values for the state, control and noises. A nominal trajectory is simply *a-priori* guess of what the system trajectory might look like, e.g. either a pre-planned trajectory, such as a flight trajectory, or the actual KF estimate. Crucially, the nominal trajectory should be as close as possible to the real trajectory so the following linearization approximately holds:

$$
\begin{aligned}
\dot{x}(t) &\approx f[x_0(t), u_0(t), w_0(t), t] + \nabla_x f|_0 (x(t) - x_0(t)) \\
&\quad + \nabla_u f|_0 (u(t) - u_0(t)) + \nabla_w f|_0 (w(t) - w_0(t)) \\
&= f[x_0(t), u_0(t), w_0(t), t] + \nabla_x f|_0 \Delta x(t) + \nabla_u f|_0 \Delta u(t) + \nabla_w f|_0 \Delta w(t), \quad (2.139) \\
y(t) &\approx h[x_0(t), v_0(t), t] + \nabla_x h|_0 (x(t) - x_0(t)) + \nabla_v h|_0 (v(t) - v_0(t)) = \\
&= h[x_0(t), v_0(t), t] + \nabla_x h|_0 \Delta x(t) + \nabla_v h|_0 \Delta v(t), \quad (2.140)
\end{aligned}
$$

where $\nabla_x f|_0$ denotes the Jacobian of the function $f[x(t), u(t), w(t), t]$ with respect to the variable x and evaluated at the nominal values $(x_0(t), u_0(t), w_0(t))$, or for the case of the measurement function $h(x(t), v(t))$, at $(x_0(t), v_0(t))$. Additionally, note that we have introduced the notation $\Delta x(t), \Delta u(t), \Delta w(t)$, and $\Delta v(t)$ to represent the deviation of the real trajectory from the nominal trajectory. Since it is reasonable to assume that the control $u(t)$ is known at all times, we can set the nominal control to the actual value $u_0(t) = u(t)$ s.t. $\Delta u(t) = 0$. Furthermore, we assume that the nominal noise for the process and measurement are both zero at all times: $w_0(t) = 0$ and $v_0(t) = 0$, s.t. $\Delta w(t) = w(t)$ and $\Delta v(t) = v(t)$. Thus, from now on, the nominal values

considered are $(\boldsymbol{x_0}(t), \boldsymbol{u_0}(t), \boldsymbol{w_0}(t), \boldsymbol{v_0}(t)) = (\boldsymbol{x_0}(t), \boldsymbol{u}(t), 0, 0)$.

Starting from Eqs. (2.139-2.140), we can write an equivalent system of SDEs for the deviation of the state from its nominal value, $\Delta\boldsymbol{x}(t)$, as well as the deviation of the measurement, $\Delta\boldsymbol{y}(t)$:

$$\Delta\dot{\boldsymbol{x}}(t) = \dot{\boldsymbol{x}}(t) - \dot{\boldsymbol{x}}_0(t) = \nabla_{\boldsymbol{x}}\boldsymbol{f}|_0\,\Delta\boldsymbol{x}(t) + \nabla_{\boldsymbol{w}}\boldsymbol{f}|_0\,\boldsymbol{w}(t) = \nabla_{\boldsymbol{x}}\boldsymbol{f}|_0\,\Delta\boldsymbol{x}(t) + \boldsymbol{q}(t), \qquad (2.141)$$

$$\Delta\boldsymbol{y}(t) = \boldsymbol{y}(t) - \boldsymbol{y}_0(t) = \nabla_{\boldsymbol{x}}\boldsymbol{h}|_0\,\Delta\boldsymbol{x}(t) + \nabla_{\boldsymbol{v}}\boldsymbol{h}|_0\,\boldsymbol{v}(t) = \nabla_{\boldsymbol{x}}\boldsymbol{h}|_0\,\Delta\boldsymbol{x}(t) + \boldsymbol{r}(t), \qquad (2.142)$$

where we have employed that the nominal trajectory fulfills: $\dot{\boldsymbol{x}}_0(t) = \boldsymbol{f}[\boldsymbol{x_0}(t), \boldsymbol{u_0}(t), \boldsymbol{w_0}(t), t]$ and $\boldsymbol{y}_0(t) = \boldsymbol{h}[\boldsymbol{x_0}(t), \boldsymbol{v_0}(t), t]$. Moreover, both noise terms have been relabeled as:

$$\boldsymbol{q}(t) := \nabla_{\boldsymbol{w}}\boldsymbol{f}|_0\,\boldsymbol{w}(t), \quad \text{and} \quad \boldsymbol{r}(t) := \nabla_{\boldsymbol{v}}\boldsymbol{h}|_0\,\boldsymbol{v}(t), \qquad (2.143)$$

with covariances

$$\mathbb{E}\big[\boldsymbol{q}(t)\boldsymbol{q}^{\mathrm{T}}(s)\big] = \tilde{\boldsymbol{Q}}(t)\delta(t-s)\mathrm{d}t, \qquad \text{where} \qquad \tilde{\boldsymbol{Q}}(t) := \nabla_{\boldsymbol{w}}\boldsymbol{f}|_0\,\boldsymbol{Q}(t)\nabla_{\boldsymbol{w}}\boldsymbol{f}^{\mathrm{T}}|_0, \qquad (2.144)$$

$$\mathbb{E}\big[\boldsymbol{r}(t)\boldsymbol{r}^{\mathrm{T}}(s)\big] = \tilde{\boldsymbol{R}}(t)\delta(t-s)\mathrm{d}t, \qquad \text{where} \qquad \tilde{\boldsymbol{R}}(t) := \nabla_{\boldsymbol{v}}\boldsymbol{h}|_0\,\boldsymbol{R}(t)\nabla_{\boldsymbol{v}}\boldsymbol{h}^{\mathrm{T}}|_0, \qquad (2.145)$$

$$\mathbb{E}\big[\boldsymbol{q}(t)\boldsymbol{r}^{\mathrm{T}}(s)\big] = \tilde{\boldsymbol{S}}(t)\delta(t-s)\mathrm{d}t, \qquad \text{where} \qquad \tilde{\boldsymbol{S}}(t) := \nabla_{\boldsymbol{w}}\boldsymbol{f}|_0\,\boldsymbol{S}(t)\nabla_{\boldsymbol{v}}\boldsymbol{h}^{\mathrm{T}}|_0. \qquad (2.146)$$

Crucially, Eqs. (2.141-2.142) are linear with respect to the state deviation $\Delta\boldsymbol{x}$, which is now the state variable. Therefore, we can use the standard KF to find an estimate for $\Delta\boldsymbol{x}(t)$, i.e. $\Delta\tilde{\boldsymbol{x}}(t)$:

$$\Delta\dot{\tilde{\boldsymbol{x}}}(t) = \boldsymbol{F}(t)\Delta\tilde{\boldsymbol{x}} + \boldsymbol{K}(t)\left(\Delta\boldsymbol{y} - \boldsymbol{H}(t)\Delta\tilde{\boldsymbol{x}}\right), \qquad (2.147)$$

$$\dot{\boldsymbol{\Sigma}}(t) = \boldsymbol{F}(t)\boldsymbol{\Sigma}(t) + \boldsymbol{\Sigma}(t)\boldsymbol{F}^{\mathrm{T}}(t) - \boldsymbol{K}(t)\tilde{\boldsymbol{R}}(t)\boldsymbol{K}^{\mathrm{T}}(t) + \tilde{\boldsymbol{Q}}(t), \qquad (2.148)$$

$$\boldsymbol{K}(t) = \left(\boldsymbol{\Sigma}(t)\boldsymbol{H}^{\mathrm{T}}(t) - \tilde{\boldsymbol{S}}(t)\right)\tilde{\boldsymbol{R}}^{-1}(t), \qquad (2.149)$$

where $\boldsymbol{\Sigma}(t)$ is the covariance matrix defined as $\boldsymbol{\Sigma}(t) = \mathbb{E}\big[(\Delta\boldsymbol{x}(t) - \Delta\tilde{\boldsymbol{x}}(t))(\Delta\boldsymbol{x}(t) - \Delta\tilde{\boldsymbol{x}}(t))^{\mathrm{T}}\big]$, and the matrices given by the model which define the KF are:

$$\boldsymbol{F}(t) := \nabla_{\boldsymbol{x}}\boldsymbol{f}|_0, \quad \boldsymbol{H}(t) := \nabla_{\boldsymbol{x}}\boldsymbol{h}|_0, \quad \text{and} \quad \boldsymbol{G}(t) := \nabla_{\boldsymbol{w}}\boldsymbol{f}|_0, \qquad (2.150)$$

with the equations for the covariances $\tilde{\boldsymbol{Q}}(t)$, $\tilde{\boldsymbol{R}}(t)$ and $\tilde{\boldsymbol{S}}(t)$, given in Eqs. (2.144-2.146). It follows from the definition of $\Delta\tilde{\boldsymbol{x}}$ that the estimate of $\boldsymbol{x}$ is simply the nominal state trajectory plus the

deviation of the estimate from this trajectory:

$$\tilde{\boldsymbol{x}}(t) = \boldsymbol{x_0}(t) + \Delta\tilde{\boldsymbol{x}}(t). \tag{2.151}$$

Therefore, we can combine the nominal state trajectory with the Kalman-Bucy equation to obtain an update rule for the estimate of $\boldsymbol{x}$. Namely,

$$\dot{\tilde{\boldsymbol{x}}}(t) = \dot{\boldsymbol{x}}_0(t) + \Delta\dot{\tilde{\boldsymbol{x}}}(t)$$
$$= \boldsymbol{f}[\boldsymbol{x_0}(t), \boldsymbol{u}(t), 0, t] + \nabla_{\boldsymbol{x}}\boldsymbol{f}|_0\Delta\tilde{\boldsymbol{x}}(t) + \boldsymbol{K}(t)(\boldsymbol{y}(t) - \boldsymbol{h}[\boldsymbol{x_0}(t), 0, t] - \nabla_{\boldsymbol{x}}\boldsymbol{h}|_0\Delta\tilde{\boldsymbol{x}}(t)), \tag{2.152}$$

where we have used that $\dot{\boldsymbol{x}}_0(t) = \boldsymbol{f}[\boldsymbol{x_0}(t), \boldsymbol{u_0}(t), \boldsymbol{w_0}(t), t], \boldsymbol{y_0}(t) = \boldsymbol{h}[\boldsymbol{x_0}(t), \boldsymbol{v_0}(t), t]$, and previous assumptions $\boldsymbol{w_0}(t) = \boldsymbol{v_0}(t) = 0$ and $\boldsymbol{u_0}(t) = \boldsymbol{u}(t)$.

A common issue when employing a linearized KF is that finding a nominal state trajectory is not straightforward. Therefore, a standard workaround is to assume that the KF estimate itself is the trajectory around which we linearize the system, i.e. $\boldsymbol{x_0}(t) := \tilde{\boldsymbol{x}}(t)$. Then, the update rule for the estimate of $\boldsymbol{x}(t)$ becomes:

$$\dot{\tilde{\boldsymbol{x}}}(t) = \boldsymbol{f}[\tilde{\boldsymbol{x}}(t), \boldsymbol{u}(t), 0, t] + \boldsymbol{K}(t)\left(\boldsymbol{y}(t) - \boldsymbol{h}[\tilde{\boldsymbol{x}}(t), 0, t]\right), \tag{2.153}$$

since $\Delta\tilde{\boldsymbol{x}}(t) = 0$, due to our redefinition of the nominal state trajectory as the estimate trajectory. The Kalman gain and the covariance $\boldsymbol{\Sigma}(t) = \mathbb{E}\left[(\boldsymbol{x}(t) - \tilde{\boldsymbol{x}}(t))(\boldsymbol{x}(t) - \tilde{\boldsymbol{x}}(t))^{\mathrm{T}}\right]$ follow the same equations as in Eqs. (2.148-2.149).

2.8 Control

So far, we have explored how to estimate time-varying parameters in a Bayesian setting, with a special emphasis on how to update their estimates as more data becomes available. However, what if we not only aim to estimate the state of our system but also to control it?

When the state of the system is fully observable, then the challenge is to devise an optimal control law to steer the state towards the desired target. For LG systems, the optimal control strategy is given by the linear-quadratic regulator (LQR).

If we have no access to the state of the system but only to indirect measurements, then accurate estimation becomes essential for successful control. For LG systems, the optimal solution is to combine the LQR with a KF, resulting in the linear-quadratic-Gaussian (LQG) control.

Model	$\dot{\boldsymbol{x}}(t) = \boldsymbol{f}[\boldsymbol{x}(t), \boldsymbol{u}(t), \boldsymbol{w}(t), t], \quad \boldsymbol{w}(t) \sim \mathcal{N}(0, \boldsymbol{Q}(t))$ $\boldsymbol{y}(t) = \boldsymbol{h}[\boldsymbol{x}(t), \boldsymbol{v}(t), t], \quad \boldsymbol{v}(t) \sim \mathcal{N}(0, \boldsymbol{R}(t))$ $\mathbb{E}\left[\boldsymbol{w}(t)\boldsymbol{v}^{\mathrm{T}}(s)\right] = \boldsymbol{S}(t)\delta(t-s)$
Initialize	$\tilde{\boldsymbol{x}}(t_0) = \tilde{\boldsymbol{x}}_0$ $\boldsymbol{\Sigma}_0 = \mathbb{E}\left[(\boldsymbol{x}_0 - \tilde{\boldsymbol{x}}_0)(\boldsymbol{x}_0 - \tilde{\boldsymbol{x}}_0)^{\mathrm{T}}\right]$
Gradients	$\boldsymbol{F}(t) := \nabla_{\boldsymbol{x}}\boldsymbol{f}\|_{\boldsymbol{0}}, \qquad \boldsymbol{H}(t) := \nabla_{\boldsymbol{x}}\boldsymbol{h}\|_{\boldsymbol{0}}, \qquad \boldsymbol{G}(t) := \nabla_{\boldsymbol{w}}\boldsymbol{f}\|_{\boldsymbol{0}}$ $\tilde{\boldsymbol{Q}}(t) := \nabla_{\boldsymbol{w}}\boldsymbol{f}\|_{\boldsymbol{0}}\,\boldsymbol{Q}(t)\nabla_{\boldsymbol{w}}\boldsymbol{f}^{\mathrm{T}}\|_{\boldsymbol{0}}, \quad \tilde{\boldsymbol{R}}(t) = \nabla_{\boldsymbol{v}}\boldsymbol{h}\|_{\boldsymbol{0}}\,\boldsymbol{R}(t)\nabla_{\boldsymbol{v}}\boldsymbol{h}^{\mathrm{T}}\|_{\boldsymbol{0}},$ $\tilde{\boldsymbol{S}}(t) = \nabla_{\boldsymbol{w}}\boldsymbol{f}\|_{\boldsymbol{0}}\,\boldsymbol{S}(t)\nabla_{\boldsymbol{v}}\boldsymbol{h}^{\mathrm{T}}\|_{\boldsymbol{0}}$
Evaluate	$\forall \dots : \quad \dots\|_{\boldsymbol{0}} := \dots\|_{(\tilde{\boldsymbol{x}}(t), \boldsymbol{u}(t), \boldsymbol{w}=0, \boldsymbol{v}=0)}$
Gain	$\boldsymbol{K}(t) = \left(\boldsymbol{\Sigma}(t)\boldsymbol{H}^{\mathrm{T}}(t) - \tilde{\boldsymbol{S}}(t)\right)\tilde{\boldsymbol{R}}^{-1}(t)$
Covariance	$\dot{\boldsymbol{\Sigma}}(t) = \boldsymbol{F}(t)\boldsymbol{\Sigma}(t) + \boldsymbol{\Sigma}(t)\boldsymbol{F}^{\mathrm{T}}(t) - \boldsymbol{K}(t)\tilde{\boldsymbol{R}}(t)\boldsymbol{K}^{\mathrm{T}}(t) + \tilde{\boldsymbol{Q}}(t)$
Estimate	$\dot{\tilde{\boldsymbol{x}}}(t) = \boldsymbol{f}[\tilde{\boldsymbol{x}}(t), \boldsymbol{u}(t), 0, t] + \boldsymbol{K}(t)\left(\boldsymbol{y}(t) - \boldsymbol{h}[\tilde{\boldsymbol{x}}(t), 0, t]\right)$

Table 2.4: Summary of the continuous correlated extended Kalman filter.

Complete and rigorous derivations of the LQR and LQG control, as well as proofs of their optimality, emerge from Bellman's equation [29, 85, 88]. However, since dynamical programming is beyond the scope of this thesis, we instead try to provide simpler yet insightful derivations, avoiding the need to explore an entirely new mathematical field. Thus, this section is intended as an introduction to the complex field of optimal control theory, with more comprehensive treatments available in textbooks like Crassidis and Junkins [29] and Kolosov [85].

Let us assume we have full access to the state of the system, $\boldsymbol{x}(t)$. Then, the simplest control law we can consider is to feed back this state $\boldsymbol{x}(t)$ into the system:

$$\boldsymbol{u}(t) = -\boldsymbol{K}_c(t)\boldsymbol{x}(t), \tag{2.154}$$

where $\boldsymbol{K}_c(t)$ is a gain matrix. Now, the key question is: what is the optimal value of $\boldsymbol{K}_c(t)$ such that the system behaves as desired? This is where the LQR comes in: it provides a systematic way to find the optimal feedback gain while balancing system performance with control effort. The first assumption, as suggested by the term *linear* in LQR, requires the system to be modeled using linear equations:

$$\dot{\boldsymbol{x}}(t) = \boldsymbol{F}(t)\boldsymbol{x}(t) + \boldsymbol{B}(t)\boldsymbol{u}(t), \tag{2.155}$$

$$\boldsymbol{y}(t) = \boldsymbol{H}(t)\boldsymbol{x}(t), \tag{2.156}$$

where $\boldsymbol{x}(t)$ is the state of the system, $\boldsymbol{y}(t)$ the measurement, and $\boldsymbol{u}(t)$ the control law. For full-state feedback, we further assume that the output of the system is the state, such that $\boldsymbol{y}(t) = \boldsymbol{x}(t)$ (i.e., $\boldsymbol{H}(t) = 1$). Therefore, only the matrices $\boldsymbol{F}(t)$ and $\boldsymbol{B}(t)$ are relevant for this problem.
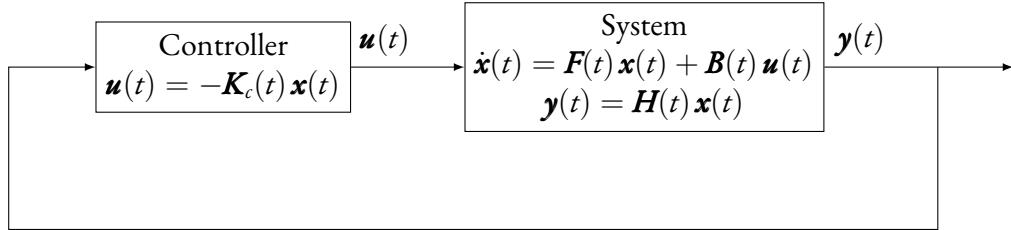
Figure 2.3: A scheme illustrating the feedback loop of a linear-quadratic regulator. When the full state of the system is accessible, it can be steered using a controller with a control law $\boldsymbol{u}(t)$ proportional to the state $\boldsymbol{x}(t)$ through a control gain $\boldsymbol{K}_c(t)$. The optimal value of $\boldsymbol{K}_c(t)$ is determined by solving the LQR minimization problem.

The *quadratic* aspect of the LQR refers to how we assess the quality of our control. Namely, through a quadratic cost function that we aim to minimize:

$$J = \int_0^\infty \left(\boldsymbol{x}^{\mathrm{T}}(t)\,\boldsymbol{\mathcal{P}}(t)\,\boldsymbol{x}(t) + \boldsymbol{u}^{\mathrm{T}}(t)\,\boldsymbol{\mathcal{V}}(t)\,\boldsymbol{u}(t) \right)\,\mathrm{d}t. \tag{2.157}$$

Here, the integration is carried out over time from zero to infinity, since we want the controller

to continuously operate, i.e. the control horizon is infinite. In the cost function J, there are two control "knobs": the positive semi-definite matrix $\boldsymbol{P}(t) \geq 0$, which penalizes deviations in the state of the system, and $\boldsymbol{V}(t) > 0$, a positive definite matrix that penalizes deviations in the control effort. By minimizing J, the controller aims to balance two competing objectives: driving the system state as close to zero as possible while minimizing the actuator effort required to achieve this. There are various methods to solve this minimization problem, including learning algorithms like gradient descent. However, we can leverage the constraints of the problem, such as the assumption of linear dynamics and quadratic cost function, to analytically solve it.

Theorem 2.1 (Linear-quadratic regulator). *Consider the following minimization problem:*

$$J_{min} = \arg\min_{\boldsymbol{u}} J = \arg\min_{\boldsymbol{u}} \int_0^\infty \left(\boldsymbol{x}^{\mathrm{T}}(t)\, \boldsymbol{P}(t)\, \boldsymbol{x}(t) + \boldsymbol{u}^{\mathrm{T}}(t)\, \boldsymbol{V}(t)\, \boldsymbol{u}(t) \right) \mathrm{d}t \qquad (2.158)$$

$$\textit{subject to} \quad \dot{\boldsymbol{x}}(t) = \boldsymbol{F}(t)\, \boldsymbol{x}(t) + \boldsymbol{B}(t)\, \boldsymbol{u}(t), \qquad (2.159)$$

where $\boldsymbol{x}(t)$ is the state vector, $\boldsymbol{P}(t) \geq 0$, and $\boldsymbol{V}(t) > 0$. The optimal control law $\boldsymbol{u}(t)$ that solves this optimization problem is known as the LQR:

$$\boldsymbol{u}(t) = -\boldsymbol{K}_c(t)\boldsymbol{x}(t), \qquad (2.160)$$

$$\boldsymbol{K}_c(t) = \boldsymbol{V}^{-1}(t)\boldsymbol{B}^{\mathrm{T}}(t)\boldsymbol{\Lambda}(t), \qquad (2.161)$$

$$-\dot{\boldsymbol{\Lambda}}(t) = \boldsymbol{F}^{\mathrm{T}}(t)\boldsymbol{\Lambda}(t) + \boldsymbol{\Lambda}(t)\, \boldsymbol{F}(t) + \boldsymbol{P}(t) - \boldsymbol{\Lambda}(t)\, \boldsymbol{B}(t)\, \boldsymbol{V}^{-1}(t)\boldsymbol{B}^{\mathrm{T}}(t)\boldsymbol{\Lambda}(t), \qquad (2.162)$$

where $\boldsymbol{K}_c(t)$ is the control gain, and Eq. (2.162) is a Riccati equation with terminal condition $\boldsymbol{\Lambda}(\infty) = 0$ to ensure that the system is continuously controlled over an infinite-time horizon.

Proof. *To show that the optimal control law given by Eqs. (2.160-2.162) solves the optimization problem in Eqs. (2.158-2.159), let us start by introducing a symmetric matrix $\boldsymbol{\Lambda}(t) = \boldsymbol{\Lambda}^{\mathrm{T}}(t)$ and rewriting the quadratic cost in Eq. (2.158) as:*

$$J = \boldsymbol{x}^{\mathrm{T}}(0)\boldsymbol{\Lambda}(0)\, \boldsymbol{x}(0) - \boldsymbol{x}^{\mathrm{T}}(0)\boldsymbol{\Lambda}(0)\, \boldsymbol{x}(0)$$

$$+ \int_0^\infty \left(\boldsymbol{x}^{\mathrm{T}}(t)\boldsymbol{P}(t)\, \boldsymbol{x}(t) + \boldsymbol{u}^{\mathrm{T}}(t)\boldsymbol{V}(t)\, \boldsymbol{u}(t) \right) \mathrm{d}t = \boldsymbol{x}^{\mathrm{T}}(0)\boldsymbol{\Lambda}(0)\, \boldsymbol{x}(0)$$

$$+ \int_0^\infty \left(\frac{\mathrm{d}}{\mathrm{d}t}\left(\boldsymbol{x}^{\mathrm{T}}(t)\boldsymbol{\Lambda}(t)\, \boldsymbol{x}(t) \right) + \boldsymbol{x}^{\mathrm{T}}(t)\boldsymbol{P}(t)\, \boldsymbol{x}(t) + \boldsymbol{u}^{\mathrm{T}}(t)\boldsymbol{V}(t)\, \boldsymbol{u}(t) \right) \mathrm{d}t, \qquad (2.163)$$

where $\boldsymbol{x}(0)$ is the initial state of our system. Furthermore, we assume the state to be stable and go to zero at infinity. Then, we can bring the term $-\boldsymbol{x}(0)\boldsymbol{\Lambda}(0)\,\boldsymbol{x}(0)$ into the integral because $\boldsymbol{x}^{\mathrm{T}}(t)\boldsymbol{\Lambda}(t)\,\boldsymbol{x}(t)|_0^\infty = 0 - \boldsymbol{x}^{\mathrm{T}}(0)\boldsymbol{\Lambda}(0)\,\boldsymbol{x}(0)$. The derivative in Eq. (2.163) can be expanded using the state space model as follows:

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\boldsymbol{x}^{\mathrm{T}}(t)\boldsymbol{\Lambda}(t)\,\boldsymbol{x}(t)\right) = \dot{\boldsymbol{x}}^{\mathrm{T}}(t)\boldsymbol{\Lambda}(t)\boldsymbol{x}(t) + \boldsymbol{x}^{\mathrm{T}}(t)\dot{\boldsymbol{\Lambda}}(t)\boldsymbol{x}(t) + \boldsymbol{x}^{\mathrm{T}}(t)\boldsymbol{\Lambda}(t)\dot{\boldsymbol{x}}(t) \tag{2.164}$$

$$= \boldsymbol{x}^{\mathrm{T}}(t)\dot{\boldsymbol{\Lambda}}(t)\boldsymbol{x}(t) + \left(\boldsymbol{F}(t)\boldsymbol{x}(t) + \boldsymbol{B}(t)\boldsymbol{u}(t)\right)^{\mathrm{T}}\boldsymbol{\Lambda}(t)\,\boldsymbol{x}(t)$$

$$+ \boldsymbol{x}^{\mathrm{T}}(t)\boldsymbol{\Lambda}(t)\left(\boldsymbol{F}(t)\boldsymbol{x}(t) + \boldsymbol{B}(t)\boldsymbol{u}(t)\right). \tag{2.165}$$

If now we plug this into Eq. (2.163) and group the $\boldsymbol{x}^{\mathrm{T}}(t)\,\cdot\,\boldsymbol{x}(t)$ terms, we get:

$$J = \boldsymbol{x}^{\mathrm{T}}(0)\boldsymbol{\Lambda}(0)\,\boldsymbol{x}(0) + \int_0^\infty \Big(\boldsymbol{x}^{\mathrm{T}}(t)\dot{\boldsymbol{\Lambda}}(t)\boldsymbol{x}(t) + \left(\boldsymbol{F}(t)\boldsymbol{x}(t) + \boldsymbol{B}(t)\boldsymbol{u}(t)\right)^{\mathrm{T}}\boldsymbol{\Lambda}(t)\,\boldsymbol{x}(t)$$

$$+ \boldsymbol{x}^{\mathrm{T}}(t)\boldsymbol{\Lambda}(t)\left(\boldsymbol{F}(t)\boldsymbol{x}(t) + \boldsymbol{B}(t)\boldsymbol{u}(t)\right) + \boldsymbol{x}^{\mathrm{T}}(t)\boldsymbol{\mathcal{P}}(t)\,\boldsymbol{x}(t) + \boldsymbol{u}^{\mathrm{T}}(t)\boldsymbol{\mathcal{V}}(t)\,\boldsymbol{u}(t)\Big)\mathrm{d}t \tag{2.166}$$

$$= \boldsymbol{x}^{\mathrm{T}}(0)\boldsymbol{\Lambda}(t)\,\boldsymbol{x}(0) + \int_0^\infty \Big(\boldsymbol{x}^{\mathrm{T}}(t)\left(\dot{\boldsymbol{\Lambda}}(t) + \boldsymbol{F}^{\mathrm{T}}(t)\boldsymbol{\Lambda}(t) + \boldsymbol{\Lambda}(t)\boldsymbol{F}(t) + \boldsymbol{\mathcal{P}}(t)\right)\boldsymbol{x}(t)$$

$$+ \boldsymbol{u}^{\mathrm{T}}(t)\boldsymbol{\mathcal{V}}(t)\,\boldsymbol{u}(t) + \boldsymbol{x}^{\mathrm{T}}(t)\boldsymbol{\Lambda}(t)\boldsymbol{B}(t)\boldsymbol{u}(t) + \boldsymbol{u}^{\mathrm{T}}(t)\boldsymbol{B}^{\mathrm{T}}(t)\boldsymbol{\Lambda}(t)\,\boldsymbol{x}(t)\Big)\mathrm{d}t. \tag{2.167}$$

Recall that the objective of this optimization task is to find the control $\boldsymbol{u}(t)$ that minimizes the cost J. Therefore, we focus on the last three terms of Eq. (2.167), since they are the only ones that depend on $\boldsymbol{u}(t)$. In particular, by completing the square, we can rewrite them as follows:

$$\boldsymbol{u}^{\mathrm{T}}(t)\boldsymbol{\mathcal{V}}(t)\boldsymbol{u}(t) + \boldsymbol{x}^{\mathrm{T}}(t)\boldsymbol{\Lambda}(t)\boldsymbol{B}(t)\boldsymbol{u}(t) + \boldsymbol{u}^{\mathrm{T}}(t)\boldsymbol{B}^{\mathrm{T}}(t)\boldsymbol{\Lambda}(t)\,\boldsymbol{x}(t) \tag{2.168}$$

$$= \left(\boldsymbol{u}(t) + \boldsymbol{\mathcal{V}}(t)^{-1}\boldsymbol{B}^{\mathrm{T}}(t)\boldsymbol{\Lambda}(t)\,\boldsymbol{x}(t)\right)^{\mathrm{T}}\boldsymbol{\mathcal{V}}(t)\left(\boldsymbol{u}(t) + \boldsymbol{\mathcal{V}}(t)^{-1}\boldsymbol{B}^{\mathrm{T}}(t)\boldsymbol{\Lambda}(t)\,\boldsymbol{x}(t)\right)$$

$$- \boldsymbol{x}^{\mathrm{T}}(t)\left(\boldsymbol{\Lambda}(t)\,\boldsymbol{B}(t)\boldsymbol{\mathcal{V}}(t)^{-1}\boldsymbol{B}^{\mathrm{T}}(t)\boldsymbol{\Lambda}(t)\right)\boldsymbol{x}(t), \tag{2.169}$$

and substitute them back into the cost function J:

$$J = \boldsymbol{x}^{\mathrm{T}}(0)\boldsymbol{\Lambda}(0)\,\boldsymbol{x}(0)$$

$$+ \int_0^\infty \Big(\boldsymbol{x}^{\mathrm{T}}(t)\left(\dot{\boldsymbol{\Lambda}}(t) + \boldsymbol{F}^{\mathrm{T}}(t)\boldsymbol{\Lambda}(t) + \boldsymbol{\Lambda}(t)\boldsymbol{F}(t) + \boldsymbol{\mathcal{P}}(t) - \boldsymbol{\Lambda}(t)\boldsymbol{B}(t)\boldsymbol{\mathcal{V}}^{-1}(t)\boldsymbol{B}^{\mathrm{T}}(t)\boldsymbol{\Lambda}(t)\right)\boldsymbol{x}(t)$$

$$+ \left(\boldsymbol{u}(t) + \boldsymbol{\mathcal{V}}^{-1}(t)\boldsymbol{B}^{\mathrm{T}}(t)\boldsymbol{\Lambda}(t)\boldsymbol{x}(t)\right)^{\mathrm{T}}\boldsymbol{\mathcal{V}}(t)\left(\boldsymbol{u}(t) + \boldsymbol{\mathcal{V}}^{-1}(t)\boldsymbol{B}^{\mathrm{T}}(t)\boldsymbol{\Lambda}(t)\,\boldsymbol{x}(t)\right)\Big)\mathrm{d}t. \tag{2.170}$$

Now, by inspection of the expression above, we can see that the cost function can be minimized by choosing the control $\boldsymbol{u}(t)$ to be:

$$\boldsymbol{u}(t) = -\boldsymbol{K}_c(t)\,\boldsymbol{x}(t), \tag{2.171}$$

where

$$\boldsymbol{K}_c(t) = \boldsymbol{\mathcal{V}}^{-1}(t)\boldsymbol{B}^{\mathrm{T}}(t)\boldsymbol{\Lambda}(t), \tag{2.172}$$

as well as solving for $\boldsymbol{\Lambda}(t)$ such that the first term of Eq. (2.170) is also zero, i.e.:

$$-\dot{\boldsymbol{\Lambda}}(t) = \boldsymbol{F}^{\mathrm{T}}(t)\boldsymbol{\Lambda}(t) + \boldsymbol{\Lambda}(t)\,\boldsymbol{F}(t) + \boldsymbol{\mathcal{P}}(t) - \boldsymbol{\Lambda}(t)\,\boldsymbol{B}(t)\,\boldsymbol{\mathcal{V}}^{-1}(t)\boldsymbol{B}^{\mathrm{T}}(t)\boldsymbol{\Lambda}(t). \tag{2.173}$$

In other words, if we can find a matrix $\boldsymbol{\Lambda}(t)$ such that the algebraic Riccati equation in (2.173) holds, then the optimal control $\boldsymbol{u}(t)$ that minimizes the cost function J is a full-state feedback term of the form $\boldsymbol{u}(t) = -\boldsymbol{\mathcal{V}}^{-1}(t)\boldsymbol{B}^{\mathrm{T}}(t)\boldsymbol{\Lambda}(t)\,\boldsymbol{x}(t)$.

2.8.2 THE LINEAR-QUADRATIC-GAUSSIAN CONTROLLER

Unfortunately, we often do not have access to the state $\boldsymbol{x}(t)$. While LQR feedback requires full state knowledge, this assumption might unrealistic due to the presence of noise. In such cases, we need to infer the state of our system from indirect noisy observations. A natural solution is to use the KF to provide state estimates, which can be later used instead of the true state to construct the LQR. For LG systems, the optimal strategy is to combine a KF with a LQR, referred to as the linear-quadratic-Gaussian (LQG) controller [29, 85].

Theorem 2.2 (Linear-quadratic-Gaussian controller)**.** *The objective is to minimize the quadratic cost function:*

$$J = \mathbb{E}\left[\int_0^\infty \left(\boldsymbol{x}^{\mathrm{T}}(t)\boldsymbol{\mathcal{P}}(t)\boldsymbol{x}(t) + \boldsymbol{u}^{\mathrm{T}}(t)\boldsymbol{\mathcal{V}}(t)\boldsymbol{u}(t)\right)\,\mathrm{d}t\right], \tag{2.174}$$

subject to a linear system driven by white Gaussian noise:

$$\dot{\boldsymbol{x}}(t) = \boldsymbol{F}(t)\,\boldsymbol{x}(t) + \boldsymbol{B}(t)\,\boldsymbol{u}(t) + \boldsymbol{G}(t)\,\boldsymbol{w}(t), \tag{2.175}$$

$$\boldsymbol{y}(t) = \boldsymbol{H}(t)\,\boldsymbol{x}(t) + \boldsymbol{v}(t), \tag{2.176}$$

where $\boldsymbol{x}(t)$ is the state vector, $\boldsymbol{u}(t)$ is the control input and $\boldsymbol{y}(t)$ is the measurement output. The zero-mean Gaussian noise process $\boldsymbol{w}(t) \sim \mathcal{N}(0, \boldsymbol{Q}(t))$ and $\boldsymbol{v}(t) \sim \mathcal{N}(0, \boldsymbol{R}(t))$ are cross-correlated through the matrix $\boldsymbol{S}(t)$, i.e. $\mathbb{E}\left[\boldsymbol{w}(t)\boldsymbol{v}^{\mathrm{T}}(s)\right] = \boldsymbol{S}(t)\delta(t - s)$. The new cost function is defined as the expected value of the cost function in Eq. (2.157), and just as in the case of the LQR, $\boldsymbol{\mathcal{P}}(t) \geq 0$, and $\boldsymbol{\mathcal{V}}(t) > 0$. The control law $\boldsymbol{u}(t)$ that minimizes the cost function is

$$\boldsymbol{u}(t) = -\boldsymbol{K}_c(t)\tilde{\boldsymbol{x}}(t), \tag{2.177}$$

where now the control is not proportional to the state but rather to its estimate, $\tilde{\boldsymbol{x}}(t)$. The control gain
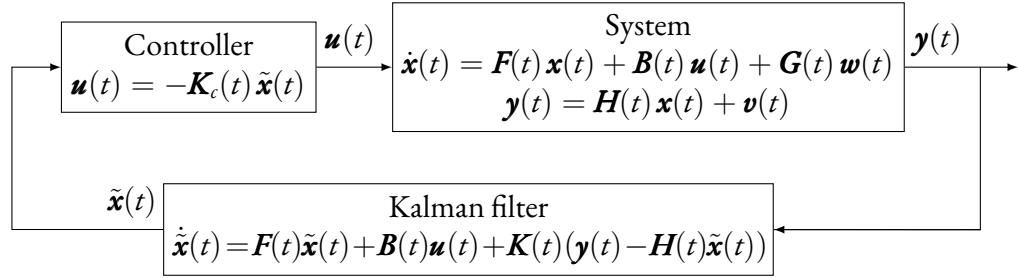
Figure 2.4: **Block diagram of the feedback scheme of a linear-quadratic Gaussian controller.** The controller computes the control input $\boldsymbol{u}(t)$ based on the estimated state $\tilde{\boldsymbol{x}}(t)$, which is provided by the KF. The system dynamics are influenced by both a process noise $\boldsymbol{w}(t)$ and a measurement noise $\boldsymbol{v}(t)$, preventing full access to the state of the system. For that reason, the KF is needed in order to estimate the state from the measurements $\boldsymbol{y}(t)$. The control law is then generated as $\boldsymbol{u}(t) = -\boldsymbol{K}_c(t)\tilde{\boldsymbol{x}}(t)$, where the control gain $\boldsymbol{K}_c(t)$ is obtained by solving the LQR problem.

regulating the feedback is given by:

$$\boldsymbol{K}_c(t) = \boldsymbol{\mathcal{V}}^{-1}(t)\boldsymbol{B}^{\mathrm{T}}(t)\boldsymbol{\Lambda}(t), \tag{2.178}$$

$$-\dot{\boldsymbol{\Lambda}}(t) = \boldsymbol{F}^{\mathrm{T}}(t)\boldsymbol{\Lambda}(t) + \boldsymbol{\Lambda}(t)\boldsymbol{F}(t) + \boldsymbol{\mathcal{P}}(t) - \boldsymbol{\Lambda}(t)\boldsymbol{B}(t)\boldsymbol{\mathcal{V}}^{-1}(t)\boldsymbol{B}^{\mathrm{T}}(t)\boldsymbol{\Lambda}(t), \tag{2.179}$$

$$\boldsymbol{\Lambda}(\infty) = 0, \tag{2.180}$$

where the matrix $\boldsymbol{\Lambda}(t)$ is the solution to a Riccati equation with an infinite control horizon. The

estimator of $\boldsymbol{x}(t)$ in Eq. (2.177) is given by the Kalman-Bucy filter:

$$\dot{\tilde{\boldsymbol{x}}}(t) = \boldsymbol{F}(t)\tilde{\boldsymbol{x}}(t) + \boldsymbol{B}(t)\boldsymbol{u}(t) + \boldsymbol{K}(t)(\boldsymbol{y}(t) - \boldsymbol{H}(t)\tilde{\boldsymbol{x}}(t)), \tag{2.181}$$

$$\tilde{\boldsymbol{x}}(0) = \mathbb{E}\big[\boldsymbol{x}(0)\,\boldsymbol{x}^{\mathrm{T}}(0)\big], \tag{2.182}$$

$$\boldsymbol{K}(t) = \big(\boldsymbol{\Sigma}(t)\,\boldsymbol{H}^{\mathrm{T}}(t) - \boldsymbol{G}(t)\boldsymbol{S}(t)\big)\,\boldsymbol{R}^{-1}(t), \tag{2.183}$$

$$\dot{\boldsymbol{\Sigma}}(t) = \boldsymbol{F}(t)\boldsymbol{\Sigma}(t) + \boldsymbol{\Sigma}(t)\boldsymbol{F}^{\mathrm{T}}(t) - \boldsymbol{K}(t)\boldsymbol{R}(t)\boldsymbol{K}^{\mathrm{T}}(t) + \boldsymbol{G}(t)\boldsymbol{Q}(t)\boldsymbol{G}^{\mathrm{T}}(t),$$

$$\boldsymbol{\Sigma}(0) = \mathbb{E}\big[\boldsymbol{x}(0)\,\boldsymbol{x}^{\mathrm{T}}(0)\big], \tag{2.184}$$

where to compute the Kalman gain $\boldsymbol{K}(t)$ one also needs to solve the dual (estimation) Riccati problem.

Proof. *First, let us focus on the term $\mathbb{E}\big[\boldsymbol{x}^{\mathrm{T}}(t)\boldsymbol{P}(t)\boldsymbol{x}(t)\big]$ in Eq. (2.174) and rewrite it in terms of the estimation error*

$$\boldsymbol{e}(t) = \tilde{\boldsymbol{x}}(t) - \boldsymbol{x}(t). \tag{2.185}$$

Namely,

$$\mathbb{E}\big[\boldsymbol{x}^{\mathrm{T}}(t)\boldsymbol{P}(t)\boldsymbol{x}(t)\big] = \mathbb{E}\big[(\tilde{\boldsymbol{x}}(t) - \boldsymbol{e}(t))^{\mathrm{T}}\boldsymbol{P}(t)(\tilde{\boldsymbol{x}}(t) - \boldsymbol{e}(t))\big] = \mathbb{E}\big[\tilde{\boldsymbol{x}}^{\mathrm{T}}(t)\boldsymbol{P}(t)\tilde{\boldsymbol{x}}(t)\big]$$
$$- 2\mathbb{E}\big[\mathrm{Tr}\big[\boldsymbol{P}(t)\boldsymbol{e}(t)\tilde{\boldsymbol{x}}^{\mathrm{T}}(t)\big]\big] + \mathbb{E}\big[\mathrm{Tr}\big[\boldsymbol{P}(t)\boldsymbol{e}(t)\boldsymbol{e}^{\mathrm{T}}(t)\big]\big] \tag{2.186}$$

where in the last step we have applied following property of the trace: $\mathrm{Tr}\big[\boldsymbol{A}\,\boldsymbol{x}\boldsymbol{x}^{\mathrm{T}}\big] = \boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}$. Note that by applying the orthogonality principle, stated in Eq. (2.130), we can simplify the expression above, since

$$\mathbb{E}\big[\mathrm{Tr}\big[\boldsymbol{P}(t)\boldsymbol{e}(t)\tilde{\boldsymbol{x}}^{\mathrm{T}}(t)\big]\big] = \mathrm{Tr}\big[\boldsymbol{P}(t)\mathbb{E}\big[\boldsymbol{e}(t)\tilde{\boldsymbol{x}}^{\mathrm{T}}(t)\big]\big] = 0. \tag{2.187}$$

Additionally, by employing the definition of the covariance $\boldsymbol{\Sigma}(t) = \mathbb{E}\big[\boldsymbol{e}(t)\boldsymbol{e}^{\mathrm{T}}(t)\big]$, we rewrite Eq. (2.186) as

$$\mathbb{E}\big[\boldsymbol{x}^{\mathrm{T}}(t)\boldsymbol{P}(t)\boldsymbol{x}(t)\big] = \mathbb{E}\big[\tilde{\boldsymbol{x}}^{\mathrm{T}}(t)\boldsymbol{P}(t)\tilde{\boldsymbol{x}}(t)\big] + \mathrm{Tr}[\boldsymbol{P}(t)\boldsymbol{\Sigma}(t)], \tag{2.188}$$

such that then, the cost function can be expressed in terms of the estimator $\tilde{\boldsymbol{x}}(t)$ and covariance matrix $\boldsymbol{\Sigma}(t)$:

$$J = \mathbb{E}\left[\int_0^\infty \left(\tilde{\boldsymbol{x}}^{\mathrm{T}}(t)\boldsymbol{P}(t)\tilde{\boldsymbol{x}}(t) + \boldsymbol{u}^{\mathrm{T}}(t)\boldsymbol{V}(t)\boldsymbol{u}(t)\right)\mathrm{d}t\right] + \int_0^\infty \mathrm{Tr}[\boldsymbol{P}(t)\boldsymbol{\Sigma}(t)]\mathrm{d}t, \qquad (2.189)$$

subject to a new constraint:

$$\dot{\tilde{\boldsymbol{x}}}(t) = \boldsymbol{F}(t)\tilde{\boldsymbol{x}}(t) + \boldsymbol{B}(t)\boldsymbol{u}(t) + \boldsymbol{K}(t)\left(\boldsymbol{y}(t) - \boldsymbol{H}(t)\tilde{\boldsymbol{x}}\right), \qquad (2.190)$$

i.e. the Kalman-Bucy filter. The problem of optimizing the cost J of Eq. (2.189) is equivalent to minimizing the first term of Eq. (2.189), since the second term, $\mathrm{Tr}[\boldsymbol{P}(t)\boldsymbol{\Sigma}(t)]$, does not depend on $\boldsymbol{u}(t)$. Namely,

$$J_{min} = \arg\min_{\boldsymbol{u}} J = \arg\min_{\boldsymbol{u}} J' \qquad (2.191)$$

where

$$J' = \mathbb{E}\left[\int_0^\infty \left(\tilde{\boldsymbol{x}}^{\mathrm{T}}(t)\boldsymbol{P}(t)\tilde{\boldsymbol{x}}(t) + \boldsymbol{u}^{\mathrm{T}}(t)\boldsymbol{V}(t)\boldsymbol{u}(t)\right)\mathrm{d}t\right]. \qquad (2.192)$$

Now, instead of the optimization problem being constrained by a differential equation on the state $\boldsymbol{x}(t)$, it is subjected to the Kalman-Bucy equation of Eq. (2.190), since we have rewritten the cost function w.r.t. the KF estimate instead of the state. Then, just like in the case of the LQR, we can use complete-the-squares to find the control $\boldsymbol{u}(t)$ that minimizes the cost,

$$\begin{aligned}J' &= \mathbb{E}\left[\tilde{\boldsymbol{x}}^{\mathrm{T}}(0)\boldsymbol{\Lambda}(0)\tilde{\boldsymbol{x}}(0)\right] - \mathbb{E}\left[\tilde{\boldsymbol{x}}^{\mathrm{T}}(0)\boldsymbol{\Lambda}(0)\tilde{\boldsymbol{x}}(0)\right] + \mathbb{E}\left[\int_0^\infty \left(\tilde{\boldsymbol{x}}^{\mathrm{T}}(t)\boldsymbol{P}(t)\tilde{\boldsymbol{x}}(t) + \boldsymbol{u}^{\mathrm{T}}(t)\boldsymbol{V}(t)\boldsymbol{u}(t)\right)\mathrm{d}t\right] \\ &= \mathbb{E}\left[\tilde{\boldsymbol{x}}^{\mathrm{T}}(0)\boldsymbol{\Lambda}(0)\tilde{\boldsymbol{x}}(0)\right] + \mathbb{E}\left[\int_0^\infty \mathrm{d}\left(\tilde{\boldsymbol{x}}^{\mathrm{T}}(t)\boldsymbol{\Lambda}(t)\tilde{\boldsymbol{x}}(t)\right) + \int_0^\infty \left(\tilde{\boldsymbol{x}}^{\mathrm{T}}(t)\boldsymbol{P}(t)\tilde{\boldsymbol{x}}(t) + \boldsymbol{u}^{\mathrm{T}}(t)\boldsymbol{V}(t)\boldsymbol{u}(t)\right)\mathrm{d}t\right] \\ &= \mathbb{E}\left[\tilde{\boldsymbol{x}}^{\mathrm{T}}(0)\boldsymbol{\Lambda}(0)\tilde{\boldsymbol{x}}(0)\right] + \int_0^\infty \mathbb{E}\left[\mathrm{d}\left(\tilde{\boldsymbol{x}}^{\mathrm{T}}(t)\boldsymbol{\Lambda}(t)\tilde{\boldsymbol{x}}(t)\right)\right] \\ &\quad + \int_0^\infty \left(\mathbb{E}\left[\tilde{\boldsymbol{x}}^{\mathrm{T}}(t)\boldsymbol{P}(t)\tilde{\boldsymbol{x}}(t)\right] + \mathbb{E}\left[\boldsymbol{u}^{\mathrm{T}}(t)\boldsymbol{V}(t)\boldsymbol{u}(t)\right]\right)\mathrm{d}t. \end{aligned} \qquad (2.193)$$

Importantly, the KF estimate is a stochastic process that follows a SDE: the Kalman-Bucy equation. Therefore, to evaluate $\mathrm{d}\big(\tilde{\boldsymbol{x}}^{\mathrm{T}}(t)\boldsymbol{\Lambda}(t)\tilde{\boldsymbol{x}}(t)\big)$, we have to use Itô calculus:

$$\mathrm{d}\big(\tilde{\boldsymbol{x}}^{\mathrm{T}}(t)\boldsymbol{\Lambda}(t)\tilde{\boldsymbol{x}}(t)\big) = \tilde{\boldsymbol{x}}^{\mathrm{T}}(t)(\mathrm{d}\boldsymbol{\Lambda}(t))\,\tilde{\boldsymbol{x}}(t) + \mathrm{d}\tilde{\boldsymbol{x}}^{\mathrm{T}}(t)\boldsymbol{\Lambda}(t)\tilde{\boldsymbol{x}}(t) + \tilde{\boldsymbol{x}}^{\mathrm{T}}(t)\boldsymbol{\Lambda}(t)\mathrm{d}\tilde{\boldsymbol{x}}(t)$$
$$+ \mathrm{d}\tilde{\boldsymbol{x}}^{\mathrm{T}}(t)\boldsymbol{\Lambda}(t)\mathrm{d}\tilde{\boldsymbol{x}}(t). \tag{2.194}$$

Then, the KF equation can be written in Itô form as

$$\mathrm{d}\tilde{\boldsymbol{x}} = \boldsymbol{F}(t)\tilde{\boldsymbol{x}}(t)\mathrm{d}t + \boldsymbol{B}(t)\boldsymbol{u}(t)\mathrm{d}t + \boldsymbol{K}(t)\mathrm{d}\boldsymbol{I}, \tag{2.195}$$
$$\mathrm{d}\boldsymbol{I} = \mathrm{d}\boldsymbol{y} - \boldsymbol{H}\tilde{\boldsymbol{x}}\,\mathrm{d}t = \boldsymbol{H}(\boldsymbol{x}-\tilde{\boldsymbol{x}})\mathrm{d}t + \mathrm{d}\boldsymbol{v} = -\boldsymbol{H}\boldsymbol{e}\,\mathrm{d}t + \mathrm{d}\boldsymbol{v} \tag{2.196}$$

where $\mathrm{d}\boldsymbol{v} := \boldsymbol{v}(t)\mathrm{d}t \sim \mathcal{N}(0, \boldsymbol{R}\mathrm{d}t)$, s.t. $\mathrm{d}\boldsymbol{I} \sim \mathcal{N}(0, \boldsymbol{R}\mathrm{d}t)$, with

$$\mathbb{E}[\mathrm{d}\boldsymbol{I}] = 0, \tag{2.197}$$
$$\mathbb{E}\big[\mathrm{d}\boldsymbol{I}\mathrm{d}\boldsymbol{I}^{\mathrm{T}}\big] = \boldsymbol{R}\mathrm{d}t, \tag{2.198}$$
$$\mathbb{E}\big[\tilde{\boldsymbol{x}}\,\mathrm{d}\boldsymbol{I}^{\mathrm{T}}\big] = 0. \tag{2.199}$$

If we apply these relationships to Eq. (2.194), we get

$$\mathrm{d}\big(\tilde{\boldsymbol{x}}^{\mathrm{T}}(t)\boldsymbol{\Lambda}(t)\tilde{\boldsymbol{x}}(t)\big) = \tilde{\boldsymbol{x}}^{\mathrm{T}}(t)(\mathrm{d}\boldsymbol{\Lambda}(t))\tilde{\boldsymbol{x}}(t) + \big(\tilde{\boldsymbol{x}}^{\mathrm{T}}(t)\boldsymbol{F}^{\mathrm{T}}(t)\mathrm{d}t + \boldsymbol{u}^{\mathrm{T}}(t)\boldsymbol{B}^{\mathrm{T}}(t)\mathrm{d}t + \mathrm{d}\boldsymbol{I}^{\mathrm{T}}\boldsymbol{K}^{\mathrm{T}}(t)\big)\boldsymbol{\Lambda}(t)\tilde{\boldsymbol{x}}(t)$$
$$+ \tilde{\boldsymbol{x}}^{\mathrm{T}}(t)\boldsymbol{\Lambda}(t)\big(\boldsymbol{F}(t)\tilde{\boldsymbol{x}}(t)\mathrm{d}t + \boldsymbol{B}(t)\boldsymbol{u}(t)\mathrm{d}t + \boldsymbol{K}(t)\mathrm{d}\boldsymbol{I}\big)$$
$$+ \mathrm{d}\boldsymbol{I}^{\mathrm{T}}(t)\boldsymbol{K}^{\mathrm{T}}(t)\boldsymbol{\Lambda}(t)\boldsymbol{K}(t)\mathrm{d}\boldsymbol{I}(t), \tag{2.200}$$

the expectation value of which reads as:

$$\mathbb{E}\big[\mathrm{d}\big(\tilde{\boldsymbol{x}}^{\mathrm{T}}(t)\boldsymbol{\Lambda}(t)\tilde{\boldsymbol{x}}(t)\big)\big] = \mathbb{E}\big[\tilde{\boldsymbol{x}}^{\mathrm{T}}(t)\big(\mathrm{d}\boldsymbol{\Lambda}(t) + \boldsymbol{F}^{\mathrm{T}}(t)\boldsymbol{\Lambda}(t)\mathrm{d}t + \boldsymbol{\Lambda}(t)\boldsymbol{F}(t)\mathrm{d}t\big)\tilde{\boldsymbol{x}}(t)\big]$$
$$+ \mathbb{E}\big[\boldsymbol{u}^{\mathrm{T}}(t)\boldsymbol{B}^{\mathrm{T}}(t)\boldsymbol{\Lambda}(t)\tilde{\boldsymbol{x}}(t)\mathrm{d}t\big] + \mathbb{E}\big[\tilde{\boldsymbol{x}}^{\mathrm{T}}(t)\boldsymbol{\Lambda}(t)\boldsymbol{B}(t)\boldsymbol{u}(t)\mathrm{d}t\big]$$
$$+ \mathbb{E}\big[\mathrm{d}\boldsymbol{I}^{\mathrm{T}}\boldsymbol{K}^{\mathrm{T}}(t)\boldsymbol{\Lambda}(t)\tilde{\boldsymbol{x}}(t) + \tilde{\boldsymbol{x}}^{\mathrm{T}}(t)\boldsymbol{\Lambda}(t)\boldsymbol{K}(t)\mathrm{d}\boldsymbol{I}\big] + \mathbb{E}\big[\mathrm{d}\boldsymbol{I}^{\mathrm{T}}(t)\boldsymbol{K}^{\mathrm{T}}(t)\boldsymbol{\Lambda}(t)\boldsymbol{K}(t)\mathrm{d}\boldsymbol{I}(t)\big]. \tag{2.201}$$

The last two terms of the expansion above can be simplified since $\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{z} = \mathrm{Tr}[\boldsymbol{A}\boldsymbol{z}\boldsymbol{x}^{\mathrm{T}}]$ when $\boldsymbol{x}$ and $\boldsymbol{z}$ have the same dimensions, which is the case for $\mathrm{d}\boldsymbol{I}$ and $\tilde{\boldsymbol{x}}(t)$. Therefore,

$$\mathbb{E}\big[\mathrm{d}\boldsymbol{I}^{\mathrm{T}}\boldsymbol{K}^{\mathrm{T}}(t)\boldsymbol{\Lambda}(t)\tilde{\boldsymbol{x}}(t) + \tilde{\boldsymbol{x}}^{\mathrm{T}}(t)\boldsymbol{\Lambda}(t)\boldsymbol{K}(t)\mathrm{d}\boldsymbol{I}\big] = \mathbb{E}\big[\mathrm{Tr}\big[\boldsymbol{K}^{\mathrm{T}}(t)\boldsymbol{\Lambda}(t)\tilde{\boldsymbol{x}}(t)\mathrm{d}\boldsymbol{I}^{\mathrm{T}}\big]\big] \tag{2.202}$$

$$+ \mathbb{E}\big[\mathrm{Tr}\big[\boldsymbol{\Lambda}(t)\boldsymbol{K}(t)\mathrm{d}\boldsymbol{I}\tilde{\boldsymbol{x}}^{\mathrm{T}}(t)\big]\big] = \mathrm{Tr}\big[\boldsymbol{K}^{\mathrm{T}}(t)\boldsymbol{\Lambda}(t)\mathbb{E}\big[\tilde{\boldsymbol{x}}(t)\mathrm{d}\boldsymbol{I}^{\mathrm{T}}\big]\big] \tag{2.203}$$

$$+ \mathrm{Tr}\big[\boldsymbol{\Lambda}(t)\boldsymbol{K}(t)\mathbb{E}\big[\mathrm{d}\boldsymbol{I}\tilde{\boldsymbol{x}}^{\mathrm{T}}(t)\big]\big] = 0, \tag{2.204}$$

and

$$\mathbb{E}\big[\mathrm{d}\boldsymbol{I}^{\mathrm{T}}(t)\boldsymbol{K}^{\mathrm{T}}(t)\boldsymbol{\Lambda}(t)\boldsymbol{K}(t)\mathrm{d}\boldsymbol{I}(t)\big] = \mathbb{E}\big[\mathrm{Tr}\big[\boldsymbol{K}^{\mathrm{T}}(t)\boldsymbol{\Lambda}(t)\boldsymbol{K}(t)\mathrm{d}\boldsymbol{I}(t)\mathrm{d}\boldsymbol{I}^{\mathrm{T}}(t)\big]\big] \tag{2.205}$$

$$= \mathrm{Tr}\big[\boldsymbol{K}^{\mathrm{T}}(t)\boldsymbol{\Lambda}(t)\boldsymbol{K}(t)\mathbb{E}\big[\mathrm{d}\boldsymbol{I}(t)\mathrm{d}\boldsymbol{I}^{\mathrm{T}}(t)\big]\big] = \mathrm{Tr}\big[\boldsymbol{K}^{\mathrm{T}}(t)\boldsymbol{\Lambda}(t)\boldsymbol{K}(t)\boldsymbol{R}(t)\big]\mathrm{d}t. \tag{2.206}$$

Thus, Eq. (2.201) reduces to

$$\mathbb{E}\big[\mathrm{d}\big(\tilde{\boldsymbol{x}}^{\mathrm{T}}(t)\boldsymbol{\Lambda}(t)\tilde{\boldsymbol{x}}(t)\big)\big] = \mathbb{E}\big[\tilde{\boldsymbol{x}}^{\mathrm{T}}(t)\big(\mathrm{d}\boldsymbol{\Lambda}(t) + \boldsymbol{F}^{\mathrm{T}}(t)\boldsymbol{\Lambda}(t)\mathrm{d}t + \boldsymbol{\Lambda}(t)\boldsymbol{F}(t)\mathrm{d}t\big)\tilde{\boldsymbol{x}}(t)\big]$$

$$+ \mathbb{E}\big[\boldsymbol{u}^{\mathrm{T}}(t)\boldsymbol{B}^{\mathrm{T}}(t)\boldsymbol{\Lambda}(t)\tilde{\boldsymbol{x}}(t)\mathrm{d}t + \tilde{\boldsymbol{x}}^{\mathrm{T}}(t)\boldsymbol{\Lambda}(t)\boldsymbol{B}(t)\boldsymbol{u}(t)\mathrm{d}t\big] + \mathrm{Tr}\big[\boldsymbol{K}^{\mathrm{T}}(t)\boldsymbol{\Lambda}(t)\boldsymbol{K}(t)\boldsymbol{R}(t)\mathrm{d}t\big], \tag{2.207}$$

which can finally be inserted back into Eq. (2.193):

$$\boldsymbol{J}' = \mathbb{E}\big[\tilde{\boldsymbol{x}}^{\mathrm{T}}(0)\boldsymbol{\Lambda}(0)\tilde{\boldsymbol{x}}(0)\big] + \int_0^\infty \mathbb{E}\big[\tilde{\boldsymbol{x}}^{\mathrm{T}}(t)\big(\mathrm{d}\boldsymbol{\Lambda}(t) + \boldsymbol{F}^{\mathrm{T}}(t)\boldsymbol{\Lambda}(t)\mathrm{d}t + \boldsymbol{\Lambda}(t)\boldsymbol{F}(t)\mathrm{d}t + \boldsymbol{\mathcal{P}}(t)\mathrm{d}t\big)\tilde{\boldsymbol{x}}(t)\big]$$

$$+ \int_0^\infty \big(\mathbb{E}\big[\boldsymbol{u}^{\mathrm{T}}(t)\boldsymbol{B}^{\mathrm{T}}(t)\boldsymbol{\Lambda}(t)\tilde{\boldsymbol{x}}(t)\mathrm{d}t + \tilde{\boldsymbol{x}}^{\mathrm{T}}(t)\boldsymbol{\Lambda}(t)\boldsymbol{B}(t)\boldsymbol{u}(t)\mathrm{d}t + \boldsymbol{u}^{\mathrm{T}}(t)\boldsymbol{\mathcal{V}}(t)\boldsymbol{u}(t)\mathrm{d}t\big]\big)$$

$$+ \int_0^\infty \mathrm{Tr}\big[\boldsymbol{K}^{\mathrm{T}}(t)\boldsymbol{\Lambda}(t)\boldsymbol{K}(t)\boldsymbol{R}(t)\big]\mathrm{d}t. \tag{2.208}$$

Given that also the last term, $\mathrm{Tr}\big[\boldsymbol{K}^{\mathrm{T}}(t)\boldsymbol{\Lambda}(t)\boldsymbol{K}(t)\boldsymbol{R}(t)\big]$, does not depend on $\boldsymbol{u}(t)$, we can equivalently write the minimization problem as

$$\boldsymbol{J}_{min} = \arg\min_{\boldsymbol{u}} \boldsymbol{J} = \arg\min_{\boldsymbol{u}} \boldsymbol{J}'', \tag{2.209}$$

where

$$J'' = \int_0^\infty \mathbb{E}\left[\tilde{\boldsymbol{x}}^{\mathrm{T}}(t)\left(\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{\Lambda}(t)+\boldsymbol{F}^{\mathrm{T}}(t)\boldsymbol{\Lambda}(t)+\boldsymbol{\Lambda}(t)\boldsymbol{F}(t)+\boldsymbol{\mathcal{P}}(t)-\boldsymbol{\Lambda}(t)\boldsymbol{B}(t)\boldsymbol{\mathcal{V}}^{-1}(t)\boldsymbol{B}^{\mathrm{T}}(t)\boldsymbol{\Lambda}(t)\right)\tilde{\boldsymbol{x}}(t)\right]\mathrm{d}t$$

$$+ \int_0^\infty \mathbb{E}\left[\left(\boldsymbol{u}(t)+\boldsymbol{\mathcal{V}}^{-1}(t)\boldsymbol{B}^{\mathrm{T}}(t)\boldsymbol{\Lambda}(t)\boldsymbol{x}(t)\right)^{\mathrm{T}}\boldsymbol{\mathcal{V}}(t)\left(\boldsymbol{u}(t)+\boldsymbol{\mathcal{V}}^{-1}(t)\boldsymbol{B}^{\mathrm{T}}(t)\boldsymbol{\Lambda}(t)\,\boldsymbol{x}(t)\right)\right]\mathrm{d}t. \qquad (2.210)$$

Then, just like in the LQR case, the control function $\boldsymbol{u}(t)$ *that minimizes the cost* $\boldsymbol{J}$ *is*

$$\boldsymbol{u}(t) = -\boldsymbol{K}_c(t)\tilde{\boldsymbol{x}}(t), \qquad (2.211)$$

where the control gain is:

$$\boldsymbol{K}_c(t) = \boldsymbol{\mathcal{V}}^{-1}(t)\boldsymbol{B}^{\mathrm{T}}(t)\boldsymbol{\Lambda}(t). \qquad (2.212)$$

To determine $\boldsymbol{\Lambda}(t)$, *we require the first term of Eq. (2.170) to be zero, resulting in the following equation:*

$$-\dot{\boldsymbol{\Lambda}}(t) = \boldsymbol{F}^{\mathrm{T}}(t)\boldsymbol{\Lambda}(t)+\boldsymbol{\Lambda}(t)\,\boldsymbol{F}(t)+\boldsymbol{\mathcal{P}}(t)-\boldsymbol{\Lambda}(t)\,\boldsymbol{B}(t)\,\boldsymbol{\mathcal{V}}^{-1}(t)\boldsymbol{B}^{\mathrm{T}}(t)\boldsymbol{\Lambda}(t), \qquad (2.213)$$

where $\boldsymbol{\Lambda}(\infty) = 0$ *because the actuator is intended to continuously control the system over an infinite time horizon.*

3

Continuously monitored quantum systems

To learn something about a quantum system, like where a particle is, how fast it is moving or how much energy it has, we must perform a measurement. Measurements in quantum mechanics are fundamentally different form those in classical systems: instead of simply observing the system, they interact with it in a way that disturbs its state.

The way quantum measurements are usually introduced in a first Quantum Mechanics course follows the framework of projective measurements developed in the 1930s by Heisenberg, Dirac and von Neumann [89, 90, 91]. Von Neumann described measurements using observables, i.e. mathematical objects called self-adjoint operators that live in the Hilbert space of the system. When an observable is measured, the possible outcomes correspond to its eigenvalues, and the quantum state undergoes a discontinuous, probabilistic, and non-unitary transformation, often referred to as "collapse" or "projection", into an eigenstate associated with the observed eigenvalue. Once this measurement is performed, there is no ambiguity on what the value of the observable is, since the state has been projected onto an eigenstate. While this framework works well for some idealized scenarios: perfect, instantaneous measurements on isolated systems, real-life

experiments often deviate from these assumptions. For instance, measurements can be imperfect, i.e. extract only partial information, and non-instantaneous, as any measurement device inevitably requires a finite amount of time to detect and output a measurement value.

To model non-ideal measurements, Davies [36] and Kraus [92] extended von Neumann's framework to include positive operator-valued measures (POVMs). Unlike the sharp "collapse" described by von Neumann, POVMs allow for imperfect measurements where the state of the system only partially collapses upon measurement.

This framework was sufficient for most quantum experiments until the 1980s, when breakthroughs in laser sources as well as trapping and cooling techniques enabled experimentalists to probe quantum systems in ways that highlighted the need for a continuous description of measurement [93, 94]. In particular, systems like atoms or ions were observed transitioning between discrete energy levels [95, 96, 97]. In other words, an atom could "jump" between states under continuous observation, and this *quantum jump* could be influenced by how the system was being measured [98, 99, 100, 101].

Inspired by the observation of these quantum jumps, a new mathematical theory for *continuous measurements* was developed [36, 37, 38, 39], describing the evolution of a quantum system under continuous monitoring, as opposed to instantaneous, projective measurements. In classical statistical physics, a system is described by an ensemble of noisy trajectories generated by a set of SDEs. Analogously, in quantum physics, the state of a system conditioned on the measurement record follows a *quantum trajectory* [38]. The evolution of such a *conditional state* [38] is generated by a stochastic master equation (SME) [40, 41, 42]. Furthermore, the trajectory of such a state can be controlled by feeding back the measurement outcomes (or an appropriate function of them) as they are registered [102, 103]. A quantum theory of feedback, which naturally arises from and requires a continuous measurement formalism, can be similarly described using SMEs [104, 105].

These SMEs, central to the theory of continuous measurement, were shown to connect directly to quantum stochastic calculus [106, 107, 108, 109, 110] and quantum filtering [111, 112, 113, 114]. Quantum stochastic calculus, a non-commutative analogue of Itô's stochastic calculus, introduces "white noise" Bose fields, $\hat{b}_t$ and $\hat{b}_t^\dagger$, satisfying canonical commutation relations. In quantum optics, these fields approximate the electromagnetic field and serve as the foundation for a consistent theory of photodetection [110, 115, 116]. Using quantum stochastic calculus, Belavkin and Barchielli extended Bayesian filtering into the quantum domain to describe how quantum states are *conditioned* by continuous measurements, thereby establishing the field of

quantum filtering [111, 112, 114]. Turns out that the SME is nothing but the representation of their quantum filter in its adjoint density form [117].

IN THIS CHAPTER, we revisit and adapt established derivations from the literature to the context of this thesis. In particular, we rederive the SME for an ensemble of two-level systems continuously monitored through the polarization of a probing light field. To do so, we start by assuming a very general setup: that of a bath (probe) interacting with a system. We first detail in Sec. 3.1 all the relevant approximations as laid out by Gross et al. (2018) [118]. Next, we derive the SME for both photodetection (in Sec. 3.2.1) and homodyne measurements (in Sec. 3.2.2), following the steps of Albarelli et al. (2024) [119]. Then, in Sec. 3.2.3, inspired by the work of Deutsch et al. (2010) [120], we explain how a system monitored with polarization spectroscopy can be described by the same SME as one measured by homodyne detection. Last but not least, we also briefly introduce in Sec. 3.3 Markovian and Bayesian feedback, as they will also be of interest in later chapters. If the reader wants to dive deeper into this topic, many other detailed sources exist, such as the works of Carmichael [38], Gardiner and Zoller [121], Breuer and Petruccione [122], Wiseman and Milburn [42], and Jacobs [123]. Those specifically interested in quantum filtering might find the review by van Handel et al. [117] particularly insightful.

3.1   INTRODUCING THE SYSTEM-BATH SETUP

Consider the case of a system $\mathcal{S}$ coupled to a bath $\mathcal{B}$ corresponding to a continuum of bosonic modes, as depicted in Fig. 3.1. The total Hamiltonian describing the evolution of the joint system and bath is given by

$$\hat{H} = \hat{H}_{\mathcal{S}} + \hat{H}_{\mathcal{B}} + \hat{H}_{\text{int}}, \tag{3.1}$$

with $\hat{H}_{\mathcal{S}}$ representing the Hamiltonian of the system, and $\hat{H}_{\text{int}}$ accounting for the interaction between the system and the bath, with the bath modeled as an infinite collection of bosonic modes whose Hamiltonian $\hat{H}_{\mathcal{B}}$ reads as:

$$\hat{H}_{\mathcal{B}} = \int_0^\infty d\omega\, \omega\, \hat{b}^\dagger(\omega)\hat{b}(\omega), \tag{3.2}$$
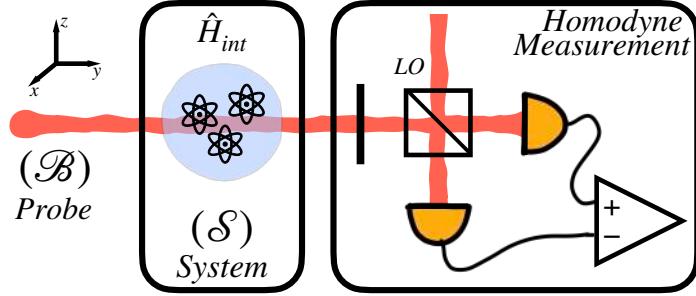
Figure 3.1: Diagram of the setup. The setup consists of a main system (labeled with $\mathcal{S}$) continuously monitored by a probe ($\mathcal{B}$) that interacts with the system and is later measured by, in this depiction, a homodyne detector. In this section, though, we consider other types of measurement such as photodetection. The interaction between the system and the probe (also sometimes referred as bath) is governed by the interaction Hamiltonian $\hat{H}_{\text{int}}$.

where ω is the frequency of each bosonic mode, with its corresponding creation and anihilation operators, $\hat{b}^\dagger(\omega)$ and $\hat{b}(\omega)$, fulfilling the commutation relationships

$$[\hat{b}(\omega), \hat{b}^\dagger(\omega')] = 2\pi\delta(\omega - \omega') \quad \text{and} \quad [\hat{b}(\omega), \hat{b}(\omega')] = 0. \tag{3.3}$$

This bath, which we label with $\mathcal{B}$, is in our case the probe that the experimenter wants to use to measure the state of the system. In other words, the probe or bath interacts with the system and it is later measured. Therefore, in this section we will use interchangeably bath, probe and in a lesser extend, environment, all referring to the subspace $\mathcal{B}$.

Additionally, the interaction Hamiltonian is assumed to be linear w.r.t. the bosonic operators $\hat{b}$:

$$\hat{H}_{\text{int}} = i \int_0^\infty d\omega \, \frac{\sqrt{\kappa(\omega)}}{2\pi} \left(\hat{L} \otimes \hat{b}^\dagger(\omega) - \hat{L}^\dagger \otimes \hat{b}(\omega) \right), \tag{3.4}$$

where $\hat{L}$ corresponds to a given system operator and $\kappa(\omega)$ is the coupling strength of the system to the bath mode with frequency ω. Throughout this derivation, we assume the system $\mathcal{S}$ to have only one characteristic frequency Ω. Additionally, we will assume that when going into the interaction picture of $\hat{H}_0 := \hat{H}_\mathcal{S} + \hat{H}_\mathcal{B}$, the system operator acquires a time dependent phase $\hat{L}(t) = \hat{L}e^{-i\Omega(t-t_0)}$, with t_0 denoting the initial time.

Even then, the interaction Hamiltonian is still too complex and further approximations must be performed. Throughout this section, each important approximation will be marked with a dot, for easier identification. The first one relies on the Markovian properties of the bath:

- *First Markov approximation:* This approximation relies on the timescale of the system's dynamics being much slower than the timescale of the bath's memory, characterized by

its correlation time τ_c. The key idea is that the bath rapidly "forgets" its interaction with the system system. In other words, any perturbation in the bath caused by its interaction with the system dissipates quickly by propagating away into the bath. From the system's perspective, the excitation it created in the bath has effectively vanished, and the bath has lost all memory of the initial disturbance. This rapid decay of the bath's memory often leads to it being described as "memoryless".

Furthermore, a bath that forgets its past very quickly indicates that its fluctuations are also rapid. In particular, they are much faster than the comparatively slow response of the system, which therefore cannot "resolve" them. In other words, the system "sees" only the average effect of these fluctuations over longer timescales, i.e., it is no longer sensitive to specific frequencies ω, and therefore, the interaction strength is constant in a (large) frequency bandwidth $\mathcal{W}$. Namely, $\kappa(\omega) = \kappa$ for $\omega \in \mathcal{W} = [\Omega - \theta, \Omega + \theta]$, and zero outside of it, where Ω is the characteristic frequency of the system.

Then, the Hamiltonian becomes:

$$\hat{H}_{\text{int}} = i\sqrt{\kappa} \int_{\mathcal{W}} d\omega \, \frac{1}{2\pi} \left(\hat{L} \otimes \hat{b}^\dagger(\omega) - \hat{L}^\dagger \otimes \hat{b}(\omega) \right). \tag{3.5}$$

We can take this Hamiltonian as the starting point [115] or derive it from a standard dipole coupling by performing a rotating-wave approximation (RWA) and keeping the slow-varying energy-conserving terms of Eq. (3.5). If the latter is performed, the timescales we consider, i.e. Δt, have to be much longer than the characteristic time of the system, $1/\Omega$. Namely, if the RWA is performed, then

$$\Omega^{-1} \ll \Delta t \tag{3.6}$$

Additionally, another assumption intrinsic in the form of Eq. (3.5) is that the system is small enough, i.e. $\Delta x \ll c\Delta t$, such that the spatial integration has been substituted by a point interaction [118].

If now we go into the interaction frame of $\hat{H}_0 = \hat{H}_S + \hat{H}_B$, then, the interaction Hamiltonian

in the time domain becomes:

$$
\begin{aligned}
\hat{H}_I(t) &:= e^{i\hat{H}_0(t-t_0)} \hat{H}_{\text{int}} e^{-i\hat{H}_0(t-t_0)} \\
&= i\frac{\sqrt{\kappa}}{2\pi} \int_{\mathcal{W}} d\omega\, e^{i\hat{H}_0(t-t_0)} \left(\hat{L} \otimes \hat{b}^\dagger(\omega) - \hat{L}^\dagger \otimes \hat{b}(\omega) \right) e^{-i\hat{H}_0(t-t_0)} \\
^{*\dagger} &= i\frac{\sqrt{\kappa}}{2\pi} \int_{\mathcal{W}} d\omega\, \left(\hat{L}e^{-i\Omega(t-t_0)} \otimes \hat{b}^\dagger(\omega)e^{i\omega(t-t_0)} - \hat{L}^\dagger e^{i\Omega(t-t_0)} \otimes \hat{b}(\omega)e^{-i\omega(t-t_0)} \right) \\
&= i\frac{\sqrt{\kappa}}{2\pi} \int_{\mathcal{W}} d\omega\, \left(\hat{L} \otimes \hat{b}^\dagger(\omega)e^{i(\omega-\Omega)(t-t_0)} - \hat{L}^\dagger \otimes \hat{b}(\omega)e^{-i(\omega-\Omega)(t-t_0)} \right) \\
&= i\sqrt{\kappa} \left(\hat{L} \otimes \frac{1}{2\pi}\int_{\mathcal{W}} d\omega\, \hat{b}^\dagger(\omega)\,e^{i(\omega-\Omega)(t-t_0)} - \hat{L}^\dagger \otimes \frac{1}{2\pi}\int_{\mathcal{W}} d\omega\, \hat{b}(\omega)e^{-i(\omega-\Omega)(t-t_0)} \right) \\
&= i\sqrt{\kappa} \left(\hat{L} \otimes \hat{b}^\dagger(t) - \hat{L}^\dagger \otimes \hat{b}(t) \right),
\end{aligned}
\tag{3.7}
$$

where in the last step we have made use of the definition of the so-called input modes $\hat{b}(t)$

$$
\hat{b}(t) = \frac{1}{2\pi}\int_{\mathcal{W}} d\omega\, \hat{b}(\omega)e^{-i(\omega-\Omega)(t-t_0)},
\tag{3.8}
$$

which one might notice resembles the Fourier transform of the frequency-domain operators but with frequencies restricted to the set $\mathcal{W} = [\Omega - \theta, \Omega + \theta]$.

- *Weak coupling:* In our simplified treatment, the system has only one characteristic or transition frequency Ω with a decay κ due to the interaction with the bath. Hence, it stands to reason to expect [*] its spectrum to be centered at Ω with a broadening proportional to κ, with its tails going to zero within a few linewidths κ of Ω, as depicted in Fig. 3.2. Earlier we defined the interaction bandwidth or the effective bath frequencies that the system sees as a set $\mathcal{W} = [\Omega - \theta, \Omega + \theta]$. Given that we assume the spectrum of the system to be centered around Ω with a broadening of κ, we must then pick θ to be much larger than κ,

---

[*]Given our initial assumption that $\hat{H}_\mathcal{B}$ has the form specified in Eq. (3.2), by using the Baker-Campbell-Hausdorff formula, it follows that $e^{i\hat{H}_\mathcal{B}(t-t_0)}\hat{b}(\omega)e^{-i\hat{H}_\mathcal{B}(t-t_0)} = \hat{b}(\omega)e^{-i\omega(t-t_0)}$, since $\hat{H}_\mathcal{B}$ is time-independent and the modes obey the commutation relationships of (3.3). In particular, the key step in this calculation is the derivation of $[\hat{H}_\mathcal{B}, \hat{b}(\omega)] = -\omega\hat{b}(\omega)$. Other work circumvents defining $\hat{H}_\mathcal{B}$ and instead require the field operators to acquire a time dependence at the frequency of the mode [118].

[†]Earlier in the definition of $\hat{L}$ we stated that when going into the interaction picture $\hat{H}_0 := \hat{H}_\mathcal{S} + \hat{H}_\mathcal{B}$, the system operator acquires a time dependent phase $\hat{L}(t) = \hat{L}e^{-i\Omega(t-t_0)}$, with t_0 denoting the initial time.

[*]For example, for a signal $S(t) \sim \cos\Omega t\, e^{-kt}$, its power spectral density is a Lorentzian centered around Ω and with a width of k, i.e. $\propto 1/\left((\omega - \Omega)^2 + k^2\right)$
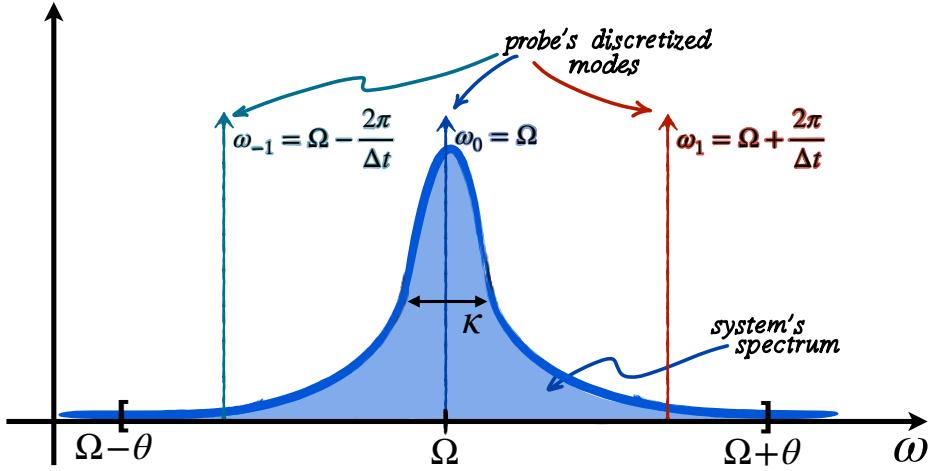
Figure 3.2: Interaction between the system and the probe in the frequency domain. This figure illustrates how the spectrum of the system, centered around Ω and broadened by a decay κ, aligns with the discretized frequency modes of the probe, in particular ω_0, ω_1 and ω_{-1}, which are spaced by $\frac{2\pi}{\Delta t}$. As detailed in the main text, an important condition when picking the Δt is that $\Delta t \ll \kappa^{-1}$, or equivalently, $1/\Delta t \gg \kappa$, ensuring that modes such as $\Omega \pm \frac{2\pi}{\Delta t}$ are sufficiently detuned from the spectrum's core to be negligible in system interactions. The parameter θ defines the interaction bandwidth $\mathcal{W}$ for which we assume constant system-probe interaction, i.e. $\kappa(\omega) = \kappa$ if $\omega \in \mathcal{W} = [\Omega - \theta, \Omega + \theta]$. It is evident also from this depiction why is necessary that $\theta \gg \kappa$.

i.e. $\theta \gg \kappa$, to ensure the whole spectrum falls within $\mathcal{W}$. Then, we can extend the integral of Eq. (3.8) over the whole real axis $\mathbb{R}$. However, for the set $\mathcal{W} = [\Omega - \theta, \Omega + \theta]$ to be properly defined, the following condition must hold: $\Omega - \theta > 0$. Thus, since $\theta \gg \kappa$, then $\Omega \gg \kappa$. In other words, the system is weakly coupled to the bath, since $\kappa \ll \Omega$.

Now, thanks to the weak-coupling approximation, we can extend $\mathcal{W}$ to the whole real axis $\mathbb{R}$. In this case, the interaction Hamiltonian has the same form as in Eq. (3.7):

$$\hat{H}_I(t) = i\sqrt{\kappa} \left(\hat{L} \otimes \hat{b}^\dagger(t) - \hat{L}^\dagger \otimes \hat{b}(t) \right), \tag{3.9}$$

but with $\hat{b}(t)$ and $\hat{b}^\dagger(t)$ becoming instantaneous temporal input modes, i.e., the Fourier transform of the frequency modes $\hat{b}(\omega)$ and $\hat{b}^\dagger(\omega)$. Namely,

$$\hat{b}(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} d\omega \, \hat{b}(\omega) e^{-i(\omega - \Omega)(t - t_0)}, \tag{3.10}$$

following the standard commutation relationship

$$[\hat{b}(t), \hat{b}^\dagger(t')] = \delta(t - t'). \tag{3.11}$$
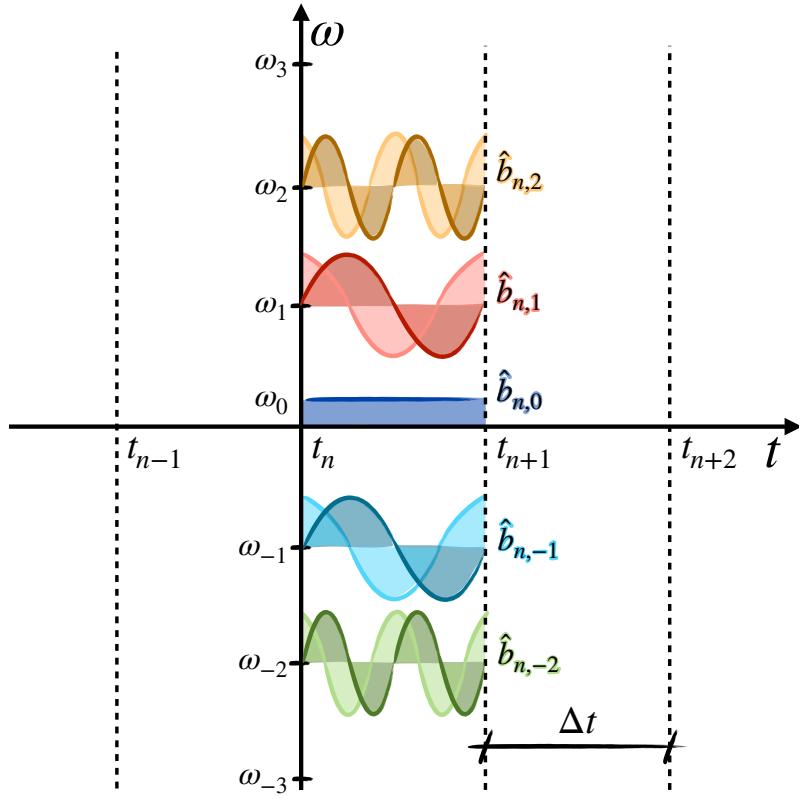
Figure 3.3: Frequency and time discretization of the probe modes. This illustration depicts the time discretization of the probe beam into intervals of Δt, with discretized times labeled as t_n with $n = -\infty, \ldots, \infty \in \mathbb{Z}$. At t_n, the on-resonant and first four discrete modes are represented, with their frequencies also discretized as $\omega_k = \Omega + \frac{2\pi}{\Delta t}k$, with $k = -2, -1, 0, 1, 2$.

So far we are in the interacting frame and the Hamiltonian governing the dynamics of the joint state of the system and the probe is given in Eq. (3.9), where the probe modes $\hat{b}(t)$ and $\hat{b}^\dagger(t)$ evolve in time.

To further simplify this description, we now discretize the probe field into time intervals of duration Δt. The time instances are defined as $t_n = n\Delta t$, with $n \in \mathbb{Z}$. Each corresponding segmented mode is labeled as $\hat{b}_n(t)$, with $n \in \mathbb{Z}$. Consequently, the temporal mode $\hat{b}(t)$ can be written as a sum of all the discrete time modes $\hat{b}_n(t)$ as:

$$\hat{b}(t) = \sqrt{\Delta t} \sum_{n=-\infty}^{\infty} \hat{b}_n(t)\Theta(t - t_n), \tag{3.12}$$

such that $\hat{b}(t)$ is still continuous but has been structured into (continuous) intervals of size Δt, utilizing the Heaviside function defined as $\Theta(u) = 1$ for $0 \le u < \Delta t$ and zero otherwise. This

segmentation ensures that each segment, or "discrete time mode", $\hat{b}_n(t)$, while being a continuous mode within its respective time interval $[t_n, t_n + \Delta t)$, is distinctly associated with the time index n. This index n highlights that $\hat{b}_n(t)$ contributes to the overall mode only during its specific interval from t_n to $t_n + \Delta t$. The factor $\sqrt{\Delta t}$ has been added in order to maintain the correct commutation relations among discrete modes, akin to those in continuous modes. This will be demonstrated in detail later in the section. This factor also ensures that the total energy, expressed in continuous time modes as $\int_{-\infty}^{\infty} \hat{b}^\dagger(t)\hat{b}(t)\mathrm{d}t$, matches that calculated using discrete modes, $\sum_n \hat{b}_n^\dagger \hat{b}_n \Delta t$.

Additionally, the segmented time modes $\hat{b}_n(t)$ can be expressed as a linear combination of discrete frequency modes $\hat{b}_{n,k}$, with the discretized frequencies ranging as: $\omega_k = \Omega + \frac{2\pi}{\Delta t}k$ with $k = -\infty, \ldots, \infty \in \mathbb{Z}$. In particular, we can use the discrete equivalent of the Fourier transform of Eq. (3.10) to expand $\hat{b}_n(t)$ as:

$$\hat{b}_n(t) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \hat{b}_{n,k}\, \mathrm{e}^{-i2\pi k t/\Delta t} \Delta\omega = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \hat{b}_{n,k}\, \mathrm{e}^{-i2\pi k t/\Delta t} \frac{2\pi}{\Delta t}, \qquad (3.13)$$

where we have set $t_0 = 0$ for simplicity, and $\Delta\omega = 2\pi/\Delta t$ is the spacing between consecutive frequency samples in the discretized frequency domain. Note that Eq. (3.13) is simply the Fourier series of $\hat{b}_n(t)$, and $\hat{b}_{n,k}$ are its Fourier coefficients defined by the integrals:

$$\hat{b}_{n,k} = \int_0^{\Delta t} \hat{b}_n(t)\, \mathrm{e}^{-i2\pi n t/\Delta t}\, \mathrm{d}t. \qquad (3.14)$$

Note, however, that the integral above defines the Fourier coefficients in terms of $\hat{b}_n(t)$, i.e. a segmented version of $\hat{b}(t)$, rather than $\hat{b}(t)$ itself.

For practical purposes, we require a direct transformation of the continuous mode $\hat{b}(t)$ into the discrete modes or Fourier coefficients $\hat{b}_{n,k}$. This transformation can be achieved by deriving the Fourier coefficients from a *double series expansion* of $\hat{b}(t)$ w.r.t. time and frequency.

By applying the Fourier expansion of the mode $\hat{b}_n(t)$ given in Eq. (3.13) to Eq. (3.12), the continuous mode $\hat{b}(t)$ can be expressed as a series expansion in both frequency and time as follows [118]:

$$\hat{b}(t) = \frac{1}{\sqrt{\Delta t}} \sum_{n=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \hat{b}_{n,k}\, \Theta(t - t_n)\, \mathrm{e}^{-i2\pi k t/\Delta t}, \qquad (3.15)$$

where the labeling n emphasizes that each frequency mode $\hat{b}_k$ is associated with a specific time interval $[t_n, t_n + \Delta t)$ indexed by n. As derived in Proposition C.1, the "Fourier" or discretized

modes $\hat{b}_{n,k}$ for Eq. (3.15) are:

$$\hat{b}_{n,k} = \frac{1}{\sqrt{\Delta t}} \int_{t_n}^{t_n+\Delta t} \hat{b}(t) e^{i2\pi kt/\Delta t} \, dt, \tag{3.16}$$

which one might notice to highly resemble the Fourier series and the integral form for its coefficients (see Table 3.1).

Type of Series	Windowed Complex Exponential Series	Fourier Series
Time-Domain Signal	$\hat{b}(t) = \sum\limits_{n=-\infty}^{\infty} \sum\limits_{k=-\infty}^{\infty} \frac{1}{\sqrt{\Delta t}} \hat{b}_{n,k} \Theta(t-t_n) e^{-i2\pi kt/\Delta t}$	$f(t) = \frac{1}{T} \sum\limits_{n=-\infty}^{\infty} c_n \, e^{-i\frac{2\pi n}{T}t}$
Frequency-Domain Coefficients	$\hat{b}_{n,k} = \frac{1}{\sqrt{\Delta t}} \int_{t_n}^{t_n+\Delta t} \hat{b}(t) e^{i2\pi kt/\Delta t} \, dt$	$c_n = \int_0^T f(t) e^{i\frac{2\pi n}{T}t} \, dt$

Table 3.1: Comparison of Signal Representations

As a sanity check, we make sure that the discretized modes $\hat{b}_{n,k}$ obey the discrete canonical commutation relation:

$$
\begin{aligned}
[\hat{b}_{n,k}, \hat{b}_{m,\ell}^\dagger] &= \frac{1}{\Delta t} \int_{t_n}^{t_n+\Delta t} \int_{t_m}^{t_m+\Delta t} [\hat{b}(t), \hat{b}^\dagger(t')] e^{i2\pi kt/\Delta t} e^{-i2\pi \ell t'/\Delta t} dt \, dt' \\
&= \frac{1}{\Delta t} \int_{t_n}^{t_n+\Delta t} \int_{t_m}^{t_m+\Delta t} \delta(t-t') e^{i2\pi kt/\Delta t} e^{-i2\pi \ell t'/\Delta t} dt \, dt' \\
&= \frac{1}{\Delta t} \int_{t_n}^{t_n+\Delta t} e^{i2\pi kt/\Delta t} \int_{t_m}^{t_m+\Delta t} e^{-i2\pi \ell t'/\Delta t} \delta(t-t') \, dt' \, dt \\
&= \begin{cases} \frac{1}{\Delta t} \int_{t_n}^{t_n+\Delta t} e^{i2\pi kt/\Delta t} e^{-i2\pi \ell t/\Delta t} \, dt & \text{if } t_m \le t \le t_m + \Delta t, \\ 0 & \text{otherwise,} \end{cases} \\
&= \frac{1}{\Delta t} \delta_{n,m} \int_{t_n}^{t_n+\Delta t} e^{i2\pi kt/\Delta t} e^{-i2\pi \ell t/\Delta t} \, dt \\
&= \frac{1}{\Delta t} \delta_{n,m} \Delta t \, \delta_{k,\ell} = \delta_{n,m} \delta_{k,\ell},
\end{aligned}
$$

where in the last line we have used the orthogonality condition of integration over the period Δt given in Eq. (C.4). Note the analogy with the commutation relationship of Eq. (3.11), which

holds due to writing the time expansion of $\hat{b}(t)$ in Eq. (3.12) with a factor Δt.

As depicted in Fig. 3.3, in each time step, there is an infinite collection of modes $\hat{b}_{n,k}$ with which the system can interact. Let us perform one more approximation in order to ensure that the system interacts only with the main on-resonant mode $\hat{b}_{n,0}$. Namely:

- *Quasimonochromatic approximation:* Let us pick a time-step Δt much smaller than the time the state of the system and bath take to "couple", i.e. $\Delta t \ll 1/\kappa$, so that the system-probe interaction is weak during this time-step Δt. Now, recall that the mode frequencies are discretized as $\omega_k = \Omega + \frac{2\pi}{\Delta t}k$, with $k \in \mathbb{Z}$. Since we have just set the condition that $1/\Delta t \gg \kappa$, that means that all probe modes except the on-resonant one will be far away from the main part of the spectrum of the system (see Fig. 3.2).

Hence, only the on-resonant mode will interact with the system ($k = 0$) and thus we can disregard the rest of the modes ($k \neq 0$) from all our calculations moving forward. Namely, the mode $\hat{b}(t)$ can now be expanded as:

$$\hat{b}(t) = \frac{1}{\sqrt{\Delta t}} \sum_{n=-\infty}^{\infty} \hat{b}_n \Theta(t - t_n) \tag{3.17}$$

where

$$\hat{b}_n := \hat{b}_{n,0} = \frac{1}{\sqrt{\Delta t}} \int_{t_n}^{t_n + \Delta t} \hat{b}(t)\, dt, \tag{3.18}$$

which is obtained by simply setting $k = 0$ in Eq. (3.16). Note that so far in this derivation, the mode at time n can be correlated with past or future modes. Thus, to further simplify the treatment, we perform yet another assumption: that the time Δt is longer than the correlation time of the bath τ_c such that different segmented modes are uncorrelated with other modes:

- *Born-Markov approximation:* As explained in the first Markov approximation, the correlation time τ_c quantifies how rapidly the bath (a.k.a. the probe) "forgets" its interaction with the system. In the setup considered so far, the probe is divided into a series of discrete temporal modes $\hat{b}_n(t)$ with $n \in \mathbb{Z}$, each interacting with the system over a distinct time interval $[t_n, t_n + \Delta t)$. When the condition $\tau_c \ll \Delta t$ is satisfied, the memory of the bath decays much faster than the duration of each probe segment. This means that the influence of the system on any given probe segment $\hat{b}_n(t)$ does not propagate forward to affect subsequent probe segments. Importantly, the state of each probe segment after interacting with the system is not "lost". Instead, each segment leaves the interaction region

with its new state preserved, i.e. dissipating away from the system, potentially to be measured later. However, the assumption of memorylessness means that when a new probe segment $\hat{b}_{n+1}(t)$ arrives, it is initialized in the same state as the earlier segments before their interactions, typically $|0\rangle\langle 0|$. This ensures that the joint state of the system and the incoming probe prior to interaction is separable and can be written as $\rho[n] \otimes |0\rangle\langle 0|$, where $\rho[n]$ describes the state of the system before the interaction with the n-th segment.

This discretization of the probe, combined with the Born-Markov approximation, yields a "conveyor belt"-type interaction [118, 119]: for instance, at time $t_n = 0$, modes $\hat{b}_{n<0}$ have not yet interacted with the system and modes $\hat{b}_{n\geq 0}$ may be correlated with the system but will not interact with it ever again. We can view this from the perspective of Hilbert spaces: each of these discrete modes exist in a distinct Hilbert space $\hat{b}_n(t) \in \mathcal{H}_n^{\mathcal{B}}$ such that the total subspace of the bath is $\mathcal{H}^{\mathcal{B}} = \bigotimes_{n\in\mathbb{Z}} \mathcal{H}_n^{\mathcal{B}}$. Furthermore, the joint system-bath state will live in an even larger Hilbert space $\mathcal{H}^{\mathcal{SB}} = \mathcal{H}^{\mathcal{S}} \bigotimes_{n\in\mathbb{Z}} \mathcal{H}^{\mathcal{B}}$. However, at each time t_n, it is not necessary to consider the whole Hilbert space but rather the reduced space $\mathcal{H}^{\mathcal{S}} \otimes \mathcal{H}_n^{\mathcal{B}}$.

At each time, the Hamiltonian living in $\mathcal{H}^{\mathcal{S}} \otimes \mathcal{H}_n^{\mathcal{B}}$ and describing the interaction between system and probe at time t_n will also be labeled with n. To derive its expression, we plug in the expression of the segmented probe field given in Eq. (3.17) into Eq. (3.9):

$$\hat{H}_I(t) = \sum_{n=-\infty}^{\infty} i\sqrt{\frac{\kappa}{\Delta t}} \left(\hat{L} \otimes \hat{b}_n^\dagger - \hat{L}^\dagger \otimes \hat{b}_n \right) \Theta(t - t_n) := \sum_{n=-\infty}^{\infty} \hat{H}_I^{(n)} \Theta(t - t_n)$$

where the interaction Hamiltonian acting from time t_n to time t_{n+1} is:

$$\hat{H}_I^{(n)} := i\sqrt{\frac{\kappa}{\Delta t}} \left(\hat{L} \otimes \hat{b}_n^\dagger - \hat{L}^\dagger \otimes \hat{b}_n \right),$$

which, importantly, is inversely proportional to $\sqrt{\Delta t}$. It then follows that the unitary evolution governed by the interaction Hamiltonian from time t_n to $t_n + \Delta t$ has the form:

$$\begin{aligned}
\hat{U}_{\Delta t}^{(n)} = \hat{U}_{\Delta t}(t_n) = \mathrm{e}^{-i\hat{H}_I^{(n)}\Delta t} &= \exp\left[-i\left(i\sqrt{\frac{\kappa}{\Delta t}} \left(\hat{L} \otimes \hat{b}_n^\dagger - \hat{L}^\dagger \otimes \hat{b}_n \right) \right) \Delta t \right] \\
&= \exp\left[\sqrt{\kappa \Delta t} \left(\hat{L} \otimes \hat{b}_n^\dagger - \hat{L}^\dagger \otimes \hat{b}_n \right) \right],
\end{aligned} \tag{3.19}$$

with the exponent proportional to $\sqrt{\Delta t}$ instead of Δt. By taking the limit $\Delta t \to 0$, the unitary

responsible for the evolution of the joint system and probe from t to $t + \mathrm{d}t$ can be written as:

$$\hat{U}_{\mathrm{d}t}(t) = \mathrm{e}^{-i\hat{H}_I(t)\mathrm{d}t} = \exp\left[\sqrt{\kappa\,\mathrm{d}t}\left(\hat{L}\otimes\hat{b}_t^\dagger - \hat{L}^\dagger\otimes\hat{b}_t\right)\right], \tag{3.20}$$

where $\mathrm{d}t$ is an infinitesimally small time increment. Note that the input modes are written as $\hat{b}_t$ (not as $\hat{b}(t)$) to highlight that t is a label identifying a particular input operator in the collection $\{\hat{b}_t\}_{t\in\mathbb{R}}$, rather than a variable.

The derivation of the unitary in Eq. (3.20) is the main result of this section. Now that we understand how the probe interacts with the system through a series of "conveyor belt"-like interactions, represented by sequential joint unitaries, we will summarize each approximation performed on the way:

- *Rotating-Wave Approximation (when necessary)*: If the RWA is performed, then it is important that $\Delta t \gg 1/\Omega$, so that the interaction of the system with the probe is averaged over many oscillations of the system during the interval Δt.

- *The First Markov approximation*, argues that $\kappa(\omega) = \kappa$ for $\omega \in \mathcal{W} = [\Omega - \theta, \Omega + \theta]$ and zero elsewhere, since the bath fluctuates so quickly that the system cannot resolve them and sees a uniform coupling.

- *Weak-coupling approximation*: the state of the bath is weakly coupled to the system, $\Omega \gg \kappa$.

- *Quasimonochromatic approximation*: If the system interacts weakly enough with the probe during a time interval Δt, i.e., $\Delta t \ll 1/\kappa$, then we can ignore all modes with $k \neq 0$ since their frequencies would fall outside the spectrum of the system and therefore not interact with it.

- *Born-Markov approximation*: The correlation time τ_c quantifies how quickly the bath (or probe) "forgets" its interaction with the system. When $\tau_c \ll \Delta t$, where Δt is the duration of each probe segment, the effect of the system on one probe segment does not propagate to subsequent segments. In other words, each discrete temporal mode $\hat{b}_n(t)$ is uncorrelated from other modes, both in its past and future. Then, each new probe segment starts in the same initial state, typically $|0\rangle\langle 0|$, ensuring that the joint state of the system and probe before interaction is separable: $\rho[n] \otimes |0\rangle\langle 0|$, where $\rho[n]$ represents the state of the system before interacting with the corresponding probe segment at time t_n. After interaction,

each probe segment retains its state and is not "lost", but simply propagates away from the system, remaining available for future measurements.

To summarize, the conditions that Δt has to fulfill are [118]:

$$\Delta x \ll c\Delta t,$$

$$\tau_c \ll \Delta t,$$

$$\frac{1}{\Omega} \ll \Delta t \ll \frac{1}{\kappa}.$$

3.2 Derivation of the stochastic master equation

Until now we have considered a general system interacting solely with a probe in a "conveyor belt" fashion. This means that the probe is discretized in time so that, at each time step Δt, its mode $\hat{b}_n$ (with $n = t/\Delta t$) interacts with the system and is then measured, yielding an outcome y_n. The effect of this interaction and subsequent measurement is captured by the following measurement operator:

$$\hat{E}_{y_n} = \langle y_n | \hat{U}^{(n)}_{\Delta t} | 0 \rangle, \tag{3.21}$$

which is obtained by evolving the initial probe state, assumed to be $|0\rangle\langle 0|$ due to the Born-Markov approximation, via the unitary interaction of Eq. (3.19). Finally, the evolved state is projected onto the eigenstate associated to the measurement outcome y_n.

However, in more general models including, the system may also dissipate into an additional, unmonitored environment. In this case, its evolution is no longer governed exclusively by a Hamiltonian but rather by a completely positive and trace-preserving (CPTP) map, Φ_n, acting from t_{n-1} to t_n.

Thus, the discretized evolution of the state consists of two alternating processes: (1) its "internal" evolution as described by the CPTP map Φ_n that accounts for both the Hamiltonian dynamics and the coupling to an unmonitored environment, and (2) the measurement update via the operator $\hat{E}_{y_n}$ corresponding to the outcome y_n obtained from the probe. Collecting these outcomes gives a discretised measurement record $\boldsymbol{y}_{0:n} = \{y_0, y_1, \ldots, y_n\}$, with $n = t/\Delta t$, which, in the continuous limit of $n \to \infty$ as $\Delta t \to 0$ becomes a continuous measurement record $\boldsymbol{y}_t$. Since the measurement update at each time step depends on the specific outcome y_n, the evolution of the state is conditional on the measurement record. Thus, different measurement trajectories lead to different state evolutions. For instance, in photodetection the outcomes are inherently

stochastic: each time step yields either a photoclick $y_n = 1$ or no click $y_n = 0$. Thus, taking the continuous-time limit of the discretized evolution results in a (stochastic) master equation, with its state evolving along a (stochastic) trajectory fully conditioned on the measurement outcomes.

We therefore begin by writing down the full discretized conditional evolution:

$$\rho[n|\mathbf{y}_{0:n}] = \frac{\Phi_n\left[\hat{E}_{y_n}\ldots\Phi_1\left[\hat{E}_{y_1}\Phi_0\left[\hat{E}_{y_0}\rho_0\hat{E}_{y_0}^\dagger\right]\hat{E}_{y_1}^\dagger\right]\ldots\hat{E}_{y_n}^\dagger\right]}{\mathrm{Tr}\left\{\Phi_n\left[\hat{E}_{y_n}\ldots\Phi_1\left[\hat{E}_{y_1}\rho_0\hat{E}_{y_1}^\dagger\right]\ldots\hat{E}_{y_n}^\dagger\right]\right\}}, \tag{3.22}$$

where the brackets $[\cdot]$ emphasize that this is the *discrete* conditional state, which at time $n = t/\Delta t$ reads as:

$$\rho[n|\mathbf{y}_{0:n}] := \rho(t_n|\mathbf{y}_{t_n}) := \rho_{(c)}(t_n). \tag{3.23}$$

It follows from Eq. (3.22) that any consecutive conditional states are related as:

$$\rho[n|\mathbf{y}_{0:n}] = \frac{\tilde{\rho}[n|\mathbf{y}_{0:n}]}{\mathrm{Tr}\{\tilde{\rho}[n|\mathbf{y}_{0:n}]\}} = \frac{\Phi_n\left[\hat{E}_{y_n}\rho[n-1|\mathbf{y}_{0:n-1}]\hat{E}_{y_n}^\dagger\right]}{\mathrm{Tr}\left\{\Phi_n\left[\hat{E}_{y_n}\rho[n-1|\mathbf{y}_{0:n-1}]\hat{E}_{y_n}^\dagger\right]\right\}}, \tag{3.24}$$

where $\tilde{\rho}$ denotes the unnormalized state. This iterative rule relating the state at time $n-1$ with the one at time n can be further split into two steps: first a measurement update and then an "internal" evolution. Specifically, we can write

$$\rho[n|\mathbf{y}_{0:n}] = \frac{\Phi_n\left[\rho[n-1|\mathbf{y}_{0:n}]\right]}{\mathrm{Tr}\{\Phi_n\left[\rho[n-1|\mathbf{y}_{0:n}]\right]\}} \tag{3.25}$$

where $\rho[n-1|\mathbf{y}_{0:n}]$ is the updated state after performing the measurement but before evolving the system under the internal map $\Phi_n[\cdot]$, i.e.,

$$\rho[n-1|\mathbf{y}_{0:n}] = \frac{\tilde{\rho}[n-1|\mathbf{y}_{0:n}]}{\mathrm{Tr}\{\tilde{\rho}[n-1|\mathbf{y}_{0:n}]\}} = \frac{\hat{E}_{y_n}\rho[n-1|\mathbf{y}_{0:n-1}]\hat{E}_{y_n}^\dagger}{\mathrm{Tr}\left\{\hat{E}_{y_n}\rho[n-1|\mathbf{y}_{0:n-1}]\hat{E}_{y_n}^\dagger\right\}}, \tag{3.26}$$

with

$$p_{\Delta t}(y_n|\mathbf{y}_{0:n-1}) = \mathrm{Tr}\{\tilde{\rho}[n-1|\mathbf{y}_{0:n}]\} \tag{3.27}$$

representing the probability of obtaining the outcome y_n given a state evolved according to a previous measurement record $\mathbf{y}_{0:n-1}$. This two-step procedure: first updating the state by the

measurement and then evolving it with the internal CPTP map, is analogous to the two-step process of a discrete KF. Accordingly, we label the states as follows:

$$
\begin{aligned}
\text{update}: \rho[n{-}1|\mathbf{y}_{0:n}] \quad &= \quad \text{state after measurement update but before internal evolution} \\
\text{predict}: \rho[n|\mathbf{y}_{0:n}] \quad &= \quad \text{state after measurement update and internal evolution.} \quad (3.28)
\end{aligned}
$$

To explicitly derive the measurement update, let us now examine up to first order in Δt the effect of the probe-system interaction and subsequent measurement. Specifically, we have to expand the unnormalized post-measurement state to first order in Δt:

$$
\tilde{\rho}[n{-}1|\mathbf{y}_{0:n}] = \hat{E}_{y_n}\rho[n{-}1|\mathbf{y}_{0:n-1}]\hat{E}_{y_n}^{\dagger}. \tag{3.29}
$$

To do so, we need to write this expression in terms of the unitary interaction of the system and the probe, and thus, their joint state. Recalling the form of the measurement operators $\hat{E}_{y_n}$ described in Eq. (3.21), we can rewritte the expression above as

$$
\begin{aligned}
\tilde{\rho}[n{-}1|\mathbf{y}_{0:n}] &= \hat{E}_{y_n}\rho[n{-}1|\mathbf{y}_{0:n-1}]\hat{E}_{y_n}^{\dagger} \\
&= \langle y_n|\hat{U}_{\Delta t}^{(n)}|0\rangle\rho[n{-}1|\mathbf{y}_{0:n-1}]\langle 0|\hat{U}_{\Delta t}^{(n)\dagger}|y_n\rangle \tag{3.30} \\
&= \langle y_n|\,\hat{U}_{\Delta t}^{(n)}\big(\rho[n{-}1|\mathbf{y}_{0:n-1}]\otimes|0\rangle\langle 0|\big)\hat{U}_{\Delta t}^{(n)\dagger}\,|y_n\rangle\,, \tag{3.31}
\end{aligned}
$$

where the joint system–probe state before measurement is separable under the Born–Markov approximation. To proceed, we expand the unitary operator $\hat{U}_{\Delta t}^{(n)}$ of Eq. (3.19) to first order in Δt:

$$
\begin{aligned}
\hat{U}_{\Delta t}^{(n)} = \mathbb{I}\otimes\mathbb{I} &+ \Big(\hat{L}\otimes\hat{b}_n^{\dagger}-\hat{L}^{\dagger}\otimes\hat{b}_n\Big)\sqrt{\kappa\,\Delta t} \\
&+ \frac{1}{2}\Big(\hat{L}^2\otimes(\hat{b}_n^{\dagger})^2 -\hat{L}^{\dagger}\hat{L}\otimes\hat{b}_n\hat{b}_n^{\dagger}-\hat{L}\hat{L}^{\dagger}\otimes\hat{b}_n^{\dagger}\hat{b}_n+(\hat{L}^{\dagger})^2\otimes(\hat{b}_n)^2\Big)\kappa\,\Delta t + o(\Delta t). \tag{3.32}
\end{aligned}
$$

Then, it follows that the joint state after the probe-system interaction reads as:

$$
\begin{aligned}
\hat{U}_{\Delta t}^{(n)}\big(\rho[n-1|\boldsymbol{y}_{0:n-1}] \otimes |0\rangle\langle 0|\big)\hat{U}_{\Delta t}^{(n)\dagger} &= \rho[n-1|\boldsymbol{y}_{0:n-1}]\otimes|0\rangle\langle 0| + \kappa\,\hat{L}\rho[n-1|\boldsymbol{y}_{0:n-1}]\hat{L}^{\dagger}\otimes|1\rangle\langle 1|\,\Delta t \\
&+ \Big(\hat{L}\rho[n-1|\boldsymbol{y}_{0:n-1}]\otimes|1\rangle\langle 0| + \rho[n-1|\boldsymbol{y}_{0:n-1}]\hat{L}^{\dagger}\otimes|0\rangle\langle 1|\Big)\sqrt{\kappa\,\Delta t} \\
&+ \frac{\kappa}{2}\Big(\sqrt{2}\hat{L}^{2}\rho[n-1|\boldsymbol{y}_{0:n-1}]\otimes|2\rangle\langle 0| - \hat{L}^{\dagger}\hat{L}\rho[n-1|\boldsymbol{y}_{0:n-1}]\otimes|0\rangle\langle 0| \\
&+ \sqrt{2}\rho[n-1|\boldsymbol{y}_{0:n-1}](\hat{L}^{\dagger})^{2}\otimes|0\rangle\langle 2| - \rho[n-1|\boldsymbol{y}_{0:n-1}]\hat{L}^{\dagger}\hat{L}\otimes|0\rangle\langle 0|\Big)\Delta t + o(\Delta t).
\end{aligned}
\tag{3.33}
$$

Thus, different SMEs will be obtained depending on our choice of measurement, i.e. the projector $|y_n\rangle\langle y_n|$, where y_n can for instance be the number of photons in the case of photodetection, or a Gaussian photocurrent in the case of homodyne measurement.

3.2.1 Photodetection

Let us consider the case where the probe is continuously measured with a photodetector, i.e., by projecting it on the Fock basis $|n\rangle\langle n|$. It is obvious from the form of the joint state before the measurement given in Eq. (3.33) that only when projecting the joint state onto the vacuum state $|0\rangle\langle 0|$ or the single photon state $|1\rangle\langle 1|$, a non-zero result will be obtained. Thus, let us consider each event, no-detection and detection, separately. If no photon is detected, then the unnormalized conditional state is:

$$
\tilde{\rho}[n-1|\{0,\boldsymbol{y}_{0:n-1}\}] = \langle 0|\hat{U}_{\Delta t}^{(n)}\big(\rho[n-1|\boldsymbol{y}_{0:n-1}] \otimes |0\rangle\langle 0|\big)\hat{U}_{\Delta t}^{(n)\dagger}|0\rangle \tag{3.34}
$$

$$
= \rho[n-1|\boldsymbol{y}_{0:n-1}] - \frac{\kappa}{2}\{\hat{L}^{\dagger}\hat{L}\,,\,\rho[n-1|\boldsymbol{y}_{0:n-1}]\}\,\Delta t + o(\Delta t), \tag{3.35}
$$

where $\rho[n-1|\boldsymbol{y}_{0:n-1}]$ is the state conditioned by the measurement at the previous time step. Then, the probability of actually detecting no photons after the state evolves for a time-step Δt is:

$$
p_{\Delta t}(0|\boldsymbol{y}_{0:n-1}) = \mathrm{Tr}\{\tilde{\rho}[n-1|\{0,\boldsymbol{y}_{0:n-1}\}]\} = 1 - \kappa\langle\hat{L}^{\dagger}\hat{L}\rangle\Delta t + o(\Delta t) \tag{3.36}
$$

where $\langle\hat{L}^{\dagger}\hat{L}\rangle = \mathrm{Tr}\{\rho[n-1|\boldsymbol{y}_{0:n-1}]\hat{L}^{\dagger}\hat{L}\}$. By now expanding the inverse of $p_{\Delta t}(0|\boldsymbol{y}_{0:n-1})$ to first order Δt,

$$
p_{\Delta t}(0|\boldsymbol{y}_{0:n-1})^{-1} = 1 + \kappa\langle\hat{L}^{\dagger}\hat{L}\rangle\Delta t + o(\Delta t), \tag{3.37}
$$

we can calculate, to first order Δt, the normalized state conditioned on the measurement outcome 0 w.r.t. the conditional state at the previous time $\rho[n-1|\boldsymbol{y}_{0:n-1}] := \rho_{(c)}(t_{n-1})$:

$$
\begin{aligned}
\rho[n-1|\{0,\boldsymbol{y}_{0:n-1}\}] &= \frac{\tilde{\rho}[n-1|\{0,\boldsymbol{y}_{0:n-1}\}]}{p_{\Delta t}(0|\boldsymbol{y}_{0:n-1})} \\
&= \left(\rho_{(c)}(t_{n-1}) - \frac{\kappa}{2}\{\hat{L}^\dagger\hat{L}, \rho_{(c)}(t_{n-1})\}\,\Delta t + o(\Delta t)\right)\left(1 + \kappa\langle\hat{L}^\dagger\hat{L}\rangle\Delta t + o(\Delta t)\right) \\
&= \rho_{(c)}(t_{n-1}) - \frac{\kappa}{2}\{\hat{L}^\dagger\hat{L}, \rho_{(c)}(t_{n-1})\}\Delta t + \kappa\langle\hat{L}^\dagger\hat{L}\rangle\rho_{(c)}(t_{n-1})\Delta t + o(\Delta t) \\
&= \rho_{(c)}(t_{n-1}) - \frac{\kappa}{2}\{\hat{L}^\dagger\hat{L}, \rho_{(c)}(t_{n-1})\}\Delta t + \frac{\kappa}{2}\langle\hat{L}^\dagger\hat{L} + \hat{L}^\dagger\hat{L}\rangle\rho_{(c)}(t_{n-1})\Delta t + o(\Delta t) \\
&= \rho_{(c)}(t_{n-1}) - \frac{\kappa}{2}\mathcal{H}[\hat{L}^\dagger\hat{L}]\rho_{(c)}(t_{n-1})\Delta t + o(\Delta t),
\end{aligned}
\tag{3.38}
$$

where we introduce the nonlinear superoperator:

$$
\mathcal{H}[\hat{O}]\,\bullet\, = \hat{O}\,\bullet\, + \,\bullet\,\hat{O}^\dagger - \mathrm{Tr}\left\{(\hat{O}+\hat{O}^\dagger)\,\bullet\,\right\}\,\bullet\,.
\tag{3.39}
$$

If instead a photon is detected, the unnormalized conditional state becomes

$$
\tilde{\rho}[n-1|\{1,\boldsymbol{y}_{0:n-1}\}] = \langle 1|\hat{U}^{(n)}_{\Delta t}(\rho[n-1|\boldsymbol{y}_{0:n-1}]\otimes|0\rangle\langle 0|)\hat{U}^{(n)\dagger}_{\Delta t}|1\rangle = \kappa\hat{L}\rho_{(c)}(t_{n-1})\hat{L}^\dagger\Delta t + o(\Delta t), \tag{3.40}
$$

from which we can calculate the associated probability:

$$
p_{\Delta t}(1|\boldsymbol{y}_{0:n-1}) = \mathrm{Tr}\{\tilde{\rho}[n-1|\{1,\boldsymbol{y}_{0:n-1}\}]\} = \kappa\langle\hat{L}^\dagger\hat{L}\rangle\Delta t + o(\Delta t).
\tag{3.41}
$$

By dividing the unnormalized updated state by the probability of measuring one photon, we retrieve:

$$
\rho[n-1|\{1,\boldsymbol{y}_{0:n-1}\}] = \frac{\tilde{\rho}[n-1|\{1,\boldsymbol{y}_{0:n-1}\}]}{p_{\Delta t}(1|\boldsymbol{y}_{0:n-1})} = \frac{\hat{L}\rho_{(c)}(t_{n-1})\hat{L}^\dagger}{\langle\hat{L}^\dagger\hat{L}\rangle} + o(\Delta t).
\tag{3.42}
$$

To sum up, at each time step, the photodetector yields either 0 or 1 with probabilities:

$$
y_n = 0 \sim p_{\Delta t}(0|\boldsymbol{y}_{0:n-1}) = 1 - \kappa\langle\hat{L}^\dagger\hat{L}\rangle\Delta t + o(\Delta t)
\tag{3.43}
$$

$$
y_n = 1 \sim p_{\Delta t}(1|\boldsymbol{y}_{0:n-1}) = \kappa\langle\hat{L}^\dagger\hat{L}\rangle\Delta t + o(\Delta t)
\tag{3.44}
$$

If one recalls the definition of a Poisson increment in Sec. 1.2.2, it comes natural to model the $\{0,1\}$ output within the timestep Δt as a Poisson increment $\Delta \mathrm{N} \sim \mathrm{Pois}(\lambda\Delta t)$, where as $\Delta t \to$

$0, \Pr[\Delta N \in \{0, 1\}] \to 1$. Thus, the Poisson increment we introduce has a mean

$$\mathbb{E}[\Delta N] = 0 \cdot p_{\Delta t}(0|\boldsymbol{y}_{0:n-1}) + 1 \cdot p_{\Delta t}(1|\boldsymbol{y}_{0:n-1}) = \kappa \langle \hat{L}^\dagger \hat{L} \rangle \Delta t + o(\Delta t) \tag{3.45}$$

which completely describes ΔN, as discussed in Sec. 1.2.2. Finally, if we apply the "internal" evolution map Φ_n, we get the unnormalized conditional state at time $t = n\Delta t$:

$$\text{if } y_n = 1 \implies \tilde{\rho}[n|\{1, \boldsymbol{y}_{0:n-1}\}] = \Phi_n \left[\rho[n-1|\{1, \boldsymbol{y}_{0:n-1}\}] \right] \tag{3.46}$$

$$\text{if } y_n = 0 \implies \tilde{\rho}[n|\{0, \boldsymbol{y}_{0:n-1}\}] = \Phi_n \left[\rho[n-1|\{0, \boldsymbol{y}_{0:n-1}\}] \right] \tag{3.47}$$

Let us now consider a measurement-based feedback map $\Phi_n[\,\cdot\,]$ where the whole history of measurement results $\boldsymbol{y}_{0:n}$ affects the Lindbladian governing the evolution of the state as

$$\Phi_n[\,\cdot\,] = e^{\mathcal{L}_{\boldsymbol{y}_{0:n}} \Delta t}[\,\cdot\,] = \cdot + \mathcal{L}_{\boldsymbol{y}_{0:n}} \cdot \Delta t + o(\Delta t), \tag{3.48}$$

where we assume $\mathcal{L}_{\boldsymbol{y}_{0:n}} = O(1)^*$. Then, the unnormalized conditional state becomes:

$$\text{if } y_n = 1 \implies \tilde{\rho}[n|\{1, \boldsymbol{y}_{0:n-1}\}] = \Phi_n \left[\frac{\hat{L} \rho_{(c)}(t_{n-1}) \hat{L}^\dagger}{\langle \hat{L}^\dagger \hat{L} \rangle} + o(\Delta t) \right]$$

$$= \frac{\hat{L} \rho_{(c)}(t_{n-1}) \hat{L}^\dagger}{\langle \hat{L}^\dagger \hat{L} \rangle} + O(\Delta t) \tag{3.49}$$

$$\text{if } y_n = 0 \implies \tilde{\rho}[n|\{0, \boldsymbol{y}_{0:n-1}\}] = \Phi_n \left[\rho_{(c)}(t_{n-1}) - \frac{\kappa}{2} \mathcal{H}[\hat{L}^\dagger \hat{L}] \rho_{(c)}(t_{n-1}) \Delta t + o(\Delta t) \right]$$

$$= \rho_{(c)}(t_{n-1}) + \mathcal{L}_{\boldsymbol{y}_{0:n}} \rho_{(c)}(t_{n-1}) \Delta t - \frac{\kappa}{2} \mathcal{H}[\hat{L}^\dagger \hat{L}] \rho_{(c)}(t_{n-1}) \Delta t + o(\Delta t), \tag{3.50}$$

where for the case of $y_n = 1$, we expand only to zeroth order. Since $\text{Tr}\left\{ \mathcal{H}[\hat{L}^\dagger \hat{L}] \rho \right\} = 0$ and $\text{Tr}\{\mathcal{L}\rho\} = 0$, the trace of both unnormalized states above is 1. In other words, the states are

---

*Big-O notation, $O(\Delta t)$, describes terms that scale at most linearly with Δt, i.e., they may be proportional to Δt or smaller (e.g. $\Delta t^{3/2}, \Delta t^2, \ldots$). Little-o, $o(\Delta t)$, includes only terms that vanish faster than Δt as $\Delta t \to 0$. For instance, $\Delta t^{3/2} \in o(\Delta t)$, while $\Delta t \in O(\Delta t)$ but not in $o(\Delta t)$.

already normalized:

$$\rho[n|\{1,\mathbf{y}_{0:n-1}\}] = \frac{\hat{L}\rho_{(c)}(t_{n-1})\hat{L}^\dagger}{\langle \hat{L}^\dagger \hat{L}\rangle} + O(\Delta t) \tag{3.51}$$

$$\rho[n|\{0,\mathbf{y}_{0:n-1}\}] = \rho_{(c)}(t_{n-1}) + \mathcal{L}_{\mathbf{y}_{0:n}}\rho_{(c)}(t_{n-1})\Delta t - \frac{\kappa}{2}\mathcal{H}[\hat{L}^\dagger \hat{L}]\rho_{(c)}(t_{n-1})\Delta t + o(\Delta t). \tag{3.52}$$

Since ΔN indicates whether a particular quantum event within the interval Δt has occurred, i.e.,

$$\Delta N = 1 \qquad \text{if the event occurs (measurement outcome 1)}, \tag{3.53}$$

$$\Delta N = 0 \qquad \text{if the event does not occur (measurement outcome 0)}, \tag{3.54}$$

then the change on the state when the event occurs v.s. when it does not can be written as

$$\Delta N \left(\rho[n|\{1,\mathbf{y}_{0:n-1}\}] - \rho[n\text{-}1|\mathbf{y}_{0:n-1}]\right) \qquad \text{if the event occurs (outcome 1)}, \tag{3.55}$$

$$(1 - \Delta N)\left(\rho[n|\{0,\mathbf{y}_{0:n-1}\}] - \rho[n\text{-}1|\mathbf{y}_{0:n-1}]\right) \quad \text{if the event does not occur (outcome 0)}, \tag{3.56}$$

where the state $\rho[n\text{-}1|\mathbf{y}_{0:n-1}] := \rho_{(c)}(t_{n-1})$ is the conditional state of the system at the previous step. Note that ΔN and $1 - \Delta N$ are included when describing the change in the state because they indicate which of the outcomes, i.e. 1 or 0, actually occur within the interval Δt. Thus, the overall change in $\rho_{(c)}(t_{n-1})$ can be written as:

$$\Delta\rho_{(c)}(t_n) = \rho_{(c)}(t_n) - \rho_{(c)}(t_{n-1}) \tag{3.57}$$
$$= \Delta N\left(\rho[n|\{1,\mathbf{y}_{0:n-1}\}] - \rho_{(c)}(t_{n-1})\right) + (1 - \Delta N)\left(\rho[n|\{0,\mathbf{y}_{0:n-1}\}] - \rho_{(c)}(t_{n-1})\right) + o(\Delta t),$$

where we can disregard any terms $\Delta t \Delta N$ because they represent higher-order infinitesimal contributions that are negligible in this analysis. In particular, for any calculus involving averages or integrations over time, the terms with $\Delta t \Delta N$ become a second-order small term, akin to Δt^2, since for a general Poisson increment with a rate $\lambda > 0$:

$$\mathbb{E}[\Delta t\,\Delta N] = \Delta t\,\mathbb{E}[\Delta N] = \Delta t\,\lambda\Delta t = \lambda\Delta t^2. \tag{3.58}$$

Therefore, when substituting Eq. (3.38) and Eq. (3.42) into Eq. (3.57), it becomes

$$
\Delta\rho_{(c)}(t_n) = \Delta N \left(\frac{\hat{L}\rho_{(c)}(t_{n-1})\hat{L}^\dagger}{\langle \hat{L}^\dagger \hat{L} \rangle} - \rho_{(c)}(t_{n-1}) \right) + (1 - \Delta N) \left(\mathcal{L}_{\mathbf{y}_{0:n}}\rho_{(c)}(t_{n-1})\Delta t - \frac{\kappa}{2}\mathcal{H}[\hat{L}^\dagger\hat{L}]\rho_{(c)}(t_{n-1})\Delta t \right)
$$

$$
= \mathcal{L}_{\mathbf{y}_{0:n}}\rho_{(c)}(t_{n-1})\Delta t - \frac{\kappa}{2}\mathcal{H}[\hat{L}^\dagger\hat{L}]\rho_{(c)}(t_{n-1})\Delta t + \left(\frac{\hat{L}\rho_{(c)}(t_{n-1})\hat{L}^\dagger}{\mathrm{Tr}\left\{ \hat{L}\rho_{(c)}(t_{n-1})\hat{L}^\dagger \right\}} - \rho_{(c)}(t_{n-1}) \right)\Delta N, \qquad (3.59)
$$

where $\mathcal{H}[\,\bullet\,]$ is the nonlinear operator defined in Eq. (3.39) and the jumps governed by the Poisson increment ΔN are modulated by a nonlinear quantity. As we will also see in the case of the homodyne measurement, these nonlinearities arise as a consequence of the normalization of the conditional state after conditioning due to the measurement. By finally taking the limit of $\Delta t \to 0$, we retrieve the SME for photodetection:

$$
\mathrm{d}\rho_{(c)}(t) = \mathcal{L}_{\mathbf{y}_t}\rho_{(c)}(t)\mathrm{d}t - \frac{\kappa}{2}\mathcal{H}[\hat{L}^\dagger\hat{L}]\rho_{(c)}(t)\mathrm{d}t + \left(\frac{\hat{L}\rho_{(c)}(t)\hat{L}^\dagger}{\mathrm{Tr}\left\{ \hat{L}\rho_{(c)}(t)\hat{L}^\dagger \right\}} - \rho_{(c)}(t) \right)\mathrm{d}N. \qquad (3.60)
$$

3.2.2 Homodyne measurement

Next, we explore monitoring the quantum system $\mathcal{S}$ via a homodyne measurement instead of with photodetection. This method involves the projection of the probe state onto the eigenstates $\{|x_n^\varphi\rangle\}$ of the general quadrature operator $\hat{X}_n^\varphi$ defined as

$$
\hat{X}_n^\varphi := \frac{\hat{b}_n\, \mathrm{e}^{\mathrm{i}\varphi} + \hat{b}_n^\dagger \mathrm{e}^{-\mathrm{i}\varphi}}{\sqrt{2}}, \qquad (3.61)
$$

which corresponds to the position operator $\hat{X}$ defined in Eq. (1.184) when $\varphi = 0$, and the momentum operator $\hat{P}$ when $\varphi = \pi/2$. The set of eigenstates $\{|x_n^\varphi\rangle\}$ related to $\hat{X}_n^\varphi$ are known as quadrature eigenstates and measure the amplitude or phase quadrature when $\varphi = 0$ and $\varphi = \pi/2$, respectively. The statistics of the homodyne outcomes x_n^φ depend on both the state of the probe being measured and the choice of φ. For specific quantum states, such as the vacuum, coherent and squeezed states, the quadrature distribution is Gaussian, characterized by its mean and variance. The mean reflects the displacement of the state, while the variance captures quantum noise, including any squeezing effects. For example, the vacuum state yields a Gaussian distribution centered at zero with variance $1/2$, representing the fundamental quantum noise, or

shot noise.

Since the probe segments are initially assumed to be in the vacuum state (recall Eq. (3.21)), performing a homodyne measurement on the probe before its interaction with the system yields a Gaussian distribution:

$$
\begin{aligned}
p_0(x_n^\varphi | \boldsymbol{x}_{0:n-1}^\varphi) &= \mathrm{Tr}\Big\{ \langle x_n^\varphi | \hat{U}_0^{(n)} |0\rangle \rho_{(c)}(t_{n-1}) \langle 0| \hat{U}_0^{(n)} |x_n^\varphi\rangle \Big\} \\
&= \mathrm{Tr}\Big\{ \langle x_n^\varphi | \left(\rho_{(c)}(t_{n-1}) \otimes |0\rangle\langle 0| \right) |x_n^\varphi\rangle \Big\} \\
&= \mathrm{Tr}\Big\{ \rho_{(c)}(t_{n-1}) \Big\} \, |\langle 0|x_n^\varphi\rangle|^2 = |\langle 0|x_n^\varphi\rangle|^2 = \frac{1}{\sqrt{\pi}} e^{-x_n^{\varphi 2}},
\end{aligned}
\tag{3.62}
$$

where in the last step we have used the fact that the ground state wavefunction of the 1D harmonic oscillator – with its energy eigenstates being the Fock states – in the (rotated) position representation is a Gaussian function, as derived in Eq. (1.240).

If, instead, we first evolve the joint system-probe state with the interaction unitary before performing the homodyne measurement as described above, the probability distribution of the measurement outcome x_n^φ will also be Gaussian, but with a shifted mean and potentially modified variance due to the system-probe interaction. To derive this result, we need to consider the trace of the unnormalized state after the measurement, which determines he probability distribution of the measurement outcome:

$$
p_{\Delta t}(x_n^\varphi | \boldsymbol{x}_{0:n-1}^\varphi) = \mathrm{Tr}\big\{ \tilde{\rho}[n-1|\boldsymbol{x}_{0:n}^\varphi] \big\}
\tag{3.63}
$$

where

$$
\begin{aligned}
\tilde{\rho}[n-1|\boldsymbol{x}_{0:n}^\varphi] &= \langle x_n^\varphi | \, \hat{U}_{\Delta t}^{(n)} \left(\rho[n-1|\boldsymbol{x}_{0:n-1}^\varphi] \otimes |0\rangle\langle 0| \right) \hat{U}_{\Delta t}^{(n)\dagger} \, |x_n^\varphi\rangle \\
&= \rho[n-1|\boldsymbol{x}_{0:n-1}^\varphi] |\langle 0|x_n^\varphi\rangle|^2 + \kappa \, \hat{L}\rho[n-1|\boldsymbol{x}_{0:n-1}^\varphi]\hat{L}^\dagger |\langle 1|x_n^\varphi\rangle|^2 \, \Delta t \\
&\quad + \left(\hat{L}\rho[n-1|\boldsymbol{x}_{0:n-1}^\varphi]\langle x_n^\varphi|1\rangle\langle 0|x_n^\varphi\rangle + \rho[n-1|\boldsymbol{x}_{0:n-1}^\varphi]\hat{L}^\dagger \langle x_n^\varphi|0\rangle\langle 1|x_n^\varphi\rangle \right) \sqrt{\kappa \, \Delta t} \\
&\quad + \frac{\kappa}{2}\Big(\sqrt{2}\hat{L}^2\rho[n-1|\boldsymbol{x}_{0:n-1}^\varphi]\langle x_n^\varphi|2\rangle\langle 0|x_n^\varphi\rangle - \hat{L}^\dagger\hat{L}\rho[n-1|\boldsymbol{x}_{0:n-1}^\varphi]|\langle 0|x_n^\varphi\rangle|^2 \\
&\quad + \sqrt{2}\rho[n-1|\boldsymbol{x}_{0:n-1}^\varphi](\hat{L}^\dagger)^2\langle x_n^\varphi|0\rangle\langle 2|x_n^\varphi\rangle - \rho[n-1|\boldsymbol{x}_{0:n-1}^\varphi]\hat{L}^\dagger\hat{L}|\langle 0|x_n^\varphi\rangle|^2 \Big)\Delta t + o(\Delta t).
\end{aligned}
\tag{3.64}
$$

The terms $\langle x_n^\varphi|1\rangle$ and $\langle x_n^\varphi|2\rangle$ can be written in terms of $\langle x_n^\varphi|0\rangle$:

$$\langle x_n^\varphi|1\rangle = \langle x_n^\varphi|\hat{b}_n^\dagger|0\rangle = \sqrt{2}e^{i\varphi}\left\langle x_n^\varphi\left|\frac{\hat{b}_n\,e^{i\varphi}+\hat{b}_n^\dagger e^{-i\varphi}}{\sqrt{2}}\right|0\right\rangle = \sqrt{2}e^{i\varphi}\langle x_n^\varphi|\hat{X}_n^\varphi|0\rangle$$

$$= \sqrt{2}e^{i\varphi}x_n^\varphi\langle x_n^\varphi|0\rangle, \tag{3.65}$$

$$\sqrt{2}\langle x_n^\varphi|2\rangle = (2x_n^{\varphi 2}-1)e^{2i\varphi}\langle x_n^\varphi|0\rangle, \tag{3.66}$$

which we derived using the fact that the second identity follows from:

$$x_n^{\varphi 2}\langle x_n^\varphi|0\rangle = \langle x_n^\varphi|\hat{X}_n^{\varphi 2}|0\rangle = \left\langle x_n^\varphi\left|\frac{\hat{b}_n\,e^{i\varphi}+\hat{b}_n^\dagger e^{-i\varphi}}{\sqrt{2}}\frac{\hat{b}_n\,e^{i\varphi}+\hat{b}_n^\dagger e^{-i\varphi}}{\sqrt{2}}\right|0\right\rangle$$

$$= \frac{1}{2}\langle x_n^\varphi|(\hat{b}_n^\dagger)^2 e^{-2i\varphi}+\hat{b}_n\,\hat{b}_n^\dagger|0\rangle = \frac{1}{2}\left(\sqrt{2}e^{-2i\varphi}\langle x_n^\varphi|2\rangle+\langle x_n^\varphi|0\rangle\right). \tag{3.67}$$

By now substituting Eq. (3.65) and Eq. (3.66) into Eq. (3.64), and recalling that $|\langle x_n^\varphi|0\rangle|^2 = p_0(x_n^\varphi|\boldsymbol{x}_{0:n-1}^\varphi)$, then the unnormalized updated state becomes

$$\tilde{\rho}[n-1|\boldsymbol{x}_{0:n}^\varphi] = p_0(x_n^\varphi|\boldsymbol{x}_{0:n-1}^\varphi)\left\{\rho_{(c)}(t_{n-1})+\left(\hat{L}\rho_{(c)}(t_{n-1})e^{i\varphi}+\rho_{(c)}(t_{n-1})\hat{L}^\dagger e^{-i\varphi}\right)\sqrt{2\kappa\Delta t}\,x_n^\varphi+\right.$$

$$+\kappa\hat{L}\rho\hat{L}^\dagger\,2x_n^{\varphi 2}\Delta t+\frac{\kappa}{2}\left(\hat{L}^2\rho_{(c)}(t_{n-1})(2x_n^{\varphi 2}-1)e^{2i\varphi}-\hat{L}^\dagger\hat{L}\rho_{(c)}(t_{n-1})\right.$$

$$\left.+\rho_{(c)}(t_{n-1})(\hat{L}^\dagger)^2(2x_n^{\varphi 2}-1)e^{-2i\varphi}-\rho_{(c)}(t_{n-1})\hat{L}^\dagger\hat{L}\right)\Delta t\right\}+o(\Delta t). \tag{3.68}$$

By now taking the trace of the unconditional state as described in Eq. (3.63), we get the probability of obtaining the measurement outcome x_n^φ after the system and probe interact for a time Δt:

$$p_{\Delta t}(x_n^\varphi|\boldsymbol{x}_{0:n-1}^\varphi) = p_0(x_n^\varphi|\boldsymbol{x}_{0:n-1}^\varphi)\left(1+x_n^\varphi\langle\hat{L}e^{i\varphi}+\hat{L}^\dagger e^{-i\varphi}\rangle\sqrt{2\kappa\Delta t}+O(\Delta t)\right), \tag{3.69}$$

which, up to order $\sqrt{\Delta t}$, follows a Gaussian distribution, since any Gaussian function can be expanded as:

$$g(t) = Ae^{-(x-\mu\sqrt{t/2})^2} = g(0)\left\{1+x\mu\sqrt{2t}+O(t)\right\}. \tag{3.70}$$

Thus, the probability of the homodyne measurement yielding an outcome x_n^φ at time t is given,

up to leading order in Δt, by the following Gaussian distribution:

$$p_{\Delta t}(x_n^\varphi | \boldsymbol{x}_{0:n-1}^\varphi) = \frac{1}{\sqrt{\pi}} \exp\left\{ -\left(x_n^\varphi - \langle \hat{L}e^{i\varphi} + \hat{L}^\dagger e^{-i\varphi} \rangle \sqrt{\frac{\kappa \Delta t}{2}} \right)^2 \right\} + O(\Delta t), \qquad (3.71)$$

with a mean $\sqrt{\kappa \Delta t / 2} \langle \hat{L}e^{i\varphi} + \hat{L}^\dagger e^{-i\varphi} \rangle = \sqrt{\kappa \Delta t / 2} \mathrm{Tr}\{ (\hat{L}e^{i\varphi} + \hat{L}^\dagger e^{-i\varphi}) \rho_{(c)}(t_{n-1}) \}$ and variance $1/2$. We can now introduce a new stochastic increment Δy_n by multiplying x_n^φ by $\sqrt{2\Delta t}$:

$$\Delta y_n := x_n^\varphi \sqrt{2\Delta t}, \qquad (3.72)$$

which, therefore, has a mean $\sqrt{\kappa} \langle \hat{L}e^{i\varphi} + \hat{L}^\dagger e^{-i\varphi} \rangle \Delta t$ and a variance Δt. Namely,

$$\Delta y_n := x_n^\varphi \sqrt{2\Delta t} = \sqrt{\kappa} \langle \hat{L}e^{i\varphi} + \hat{L}^\dagger e^{-i\varphi} \rangle \Delta t + \Delta W, \qquad (3.73)$$

where $\Delta W \sim \mathcal{N}(0, \Delta t)$ denotes the Wiener increment introduced in Sec. 1.2.3. Physically, the derivative $I(t) := \lim_{\Delta t \to 0} \Delta y_n / \Delta t$ corresponds to the stochastically fluctuating photocurrent being measured in real time in a homodyne setup. Next, by using $x_n^\varphi \sqrt{2\Delta t} = \Delta y_n$ and $2x_n^{\varphi 2} \Delta t = \Delta y_n^2 = \Delta t + o(\Delta t)$, we can rewrite Eq. (3.68) it into the following, more familiar, form:

$$\begin{aligned}
\tilde{\rho}[n-1|\boldsymbol{x}_{0:n}^\varphi] &= p_0(x_n^\varphi | \boldsymbol{x}_{0:n-1}^\varphi) \left\{ \rho_{(c)}(t_{n-1}) + \sqrt{\kappa} \left(\hat{L}\rho_{(c)}(t_{n-1})e^{i\varphi} + \rho_{(c)}(t_{n-1})\hat{L}^\dagger e^{-i\varphi} \right) \Delta y_n + \kappa \hat{L}\rho\hat{L}^\dagger \Delta t \right.\\
&\quad \left. - \frac{\kappa}{2} \left(\hat{L}^\dagger \hat{L}\rho_{(c)}(t_{n-1}) + \rho_{(c)}(t_{n-1})\hat{L}^\dagger \hat{L} \right) \Delta t + o(\Delta t) \right\} = p_0(x_n^\varphi | \boldsymbol{x}_{0:n-1}^\varphi) \left\{ \rho_{(c)}(t_{n-1}) \right.\\
&\quad \left. + \kappa \mathcal{D}[\hat{L}]\rho_{(c)}(t_{n-1})\Delta t + \sqrt{\kappa} \left(\hat{L}\rho_{(c)}(t_{n-1})e^{i\varphi} + \rho_{(c)}(t_{n-1})\hat{L}^\dagger e^{-i\varphi} \right) \Delta y_n + o(\Delta t) \right\}, \quad (3.74)
\end{aligned}$$

where the superoperator representing the measurement-induced decoherence has the same form as the one defined in Eq. (1.196). To normalize this state, we divide it by $p_{\Delta t}(x_n^\varphi | \boldsymbol{x}_{0:n-1}^\varphi)$ whilst

keeping terms to first order in Δt and lower:

$$
\begin{aligned}
\rho[n-1|\boldsymbol{x}_{0:n}^{\varphi}] = \frac{\tilde{\rho}[n-1|\boldsymbol{x}_{0:n}^{\varphi}]}{p_{\Delta t}(x_n^{\varphi}|\boldsymbol{x}_{0:n-1}^{\varphi})} &= \Big(\rho_{(c)}(t_{n-1}) + \kappa\mathcal{D}[\hat{L}]\rho_{(c)}(t_{n-1})\Delta t \\
&+ \sqrt{\kappa}\big(\hat{L}\rho_{(c)}(t_{n-1})\mathrm{e}^{\mathrm{i}\varphi} + \rho_{(c)}(t_{n-1})\hat{L}^{\dagger}\mathrm{e}^{-\mathrm{i}\varphi}\big)\Delta y_n\Big)\Big(1 - \sqrt{\kappa}\langle\hat{L}\mathrm{e}^{\mathrm{i}\varphi} + \hat{L}^{\dagger}\mathrm{e}^{-\mathrm{i}\varphi}\rangle\Delta y_n \\
&+ \kappa\langle\hat{L}\mathrm{e}^{\mathrm{i}\varphi} + \hat{L}^{\dagger}\mathrm{e}^{-\mathrm{i}\varphi}\rangle^2\Delta t\Big) + o(\Delta t) = \rho_{(c)}(t_{n-1}) + \kappa\mathcal{D}[\hat{L}]\rho_{(c)}(t_{n-1})\Delta t \\
&+ \sqrt{\kappa}\big(\hat{L}\rho_{(c)}(t_{n-1})\mathrm{e}^{\mathrm{i}\varphi} + \rho_{(c)}(t_{n-1})\hat{L}^{\dagger}\mathrm{e}^{-\mathrm{i}\varphi}\big)\Delta y_n - \sqrt{\kappa}\langle\hat{L}\mathrm{e}^{\mathrm{i}\varphi} + \hat{L}^{\dagger}\mathrm{e}^{-\mathrm{i}\varphi}\rangle\rho_{(c)}(t_{n-1})\Delta y_n \\
&- \kappa\big(\hat{L}\rho_{(c)}(t_{n-1})\mathrm{e}^{\mathrm{i}\varphi} + \rho_{(c)}(t_{n-1})\hat{L}^{\dagger}\mathrm{e}^{-\mathrm{i}\varphi}\big)\langle\hat{L}\mathrm{e}^{\mathrm{i}\varphi} + \hat{L}^{\dagger}\mathrm{e}^{-\mathrm{i}\varphi}\rangle\Delta t + \kappa\langle\hat{L}\mathrm{e}^{\mathrm{i}\varphi} \\
&+ \hat{L}^{\dagger}\mathrm{e}^{-\mathrm{i}\varphi}\rangle^2\Delta t + o(\Delta t).
\end{aligned}
\tag{3.75}
$$

where we have used that the inverse of the conditional probability to leading order Δt is:

$$
\begin{aligned}
\frac{1}{p_{\Delta t}(x_n^{\varphi}|\boldsymbol{x}_{0:n-1}^{\varphi})} &= \frac{1}{p_0(x_n^{\varphi}|\boldsymbol{x}_{0:n-1}^{\varphi})}\left(1 + x_n^{\varphi}\langle\hat{L}\mathrm{e}^{\mathrm{i}\varphi} + \hat{L}^{\dagger}\mathrm{e}^{-\mathrm{i}\varphi}\rangle\sqrt{2\kappa\Delta t} + o(\Delta t)\right)^{-1} \\
&= \frac{1}{p_0(x_n^{\varphi}|\boldsymbol{x}_{0:n-1}^{\varphi})}\left(1 + \sqrt{\kappa}\langle\hat{L}\mathrm{e}^{\mathrm{i}\varphi} + \hat{L}^{\dagger}\mathrm{e}^{-\mathrm{i}\varphi}\rangle\Delta y_n + o(\Delta t)\right)^{-1} \\
&= \frac{1}{p_0(x_n^{\varphi}|\boldsymbol{x}_{0:n-1}^{\varphi})}\left(1 - \sqrt{\kappa}\langle\hat{L}\mathrm{e}^{\mathrm{i}\varphi} + \hat{L}^{\dagger}\mathrm{e}^{-\mathrm{i}\varphi}\rangle\Delta y_n + \kappa\langle\hat{L}\mathrm{e}^{\mathrm{i}\varphi} + \hat{L}^{\dagger}\mathrm{e}^{-\mathrm{i}\varphi}\rangle^2\Delta t + o(\Delta t)\right).
\end{aligned}
\tag{3.76}
$$

By now inserting the expression of Δy_n defined in Eq. (3.73), we can simplify the state in Eq. (3.75) and get

$$
\rho[n-1|\boldsymbol{x}_{0:n}^{\varphi}] = \rho_{(c)}(t_{n-1}) + \kappa\mathcal{D}[\hat{L}]\rho_{(c)}(t_{n-1})\Delta t + \sqrt{\kappa}\mathcal{H}[\hat{L}\mathrm{e}^{\mathrm{i}\varphi}]\rho_{(c)}(t_{n-1})\Delta\mathrm{W} + o(\Delta t)
\tag{3.77}
$$

with the nonlinear superoperator modulating the stochastic (Gaussian) kicks defined in Eq. (3.39). Then, if we evolve the state with the "internal" map Φ_n, we get the conditional state at time $t = n\Delta t$

$$
\begin{aligned}
\rho_{(c)}(t_n) = \rho[n|\boldsymbol{x}_{0:n}^{\varphi}] &= \frac{\tilde{\rho}[n|\boldsymbol{x}_{0:n}^{\varphi}]}{\mathrm{Tr}\{\tilde{\rho}[n|\boldsymbol{x}_{0:n}^{\varphi}]\}} \\
&= \frac{\Phi_n\Big[\rho_{(c)}(t_{n-1}) + \kappa\mathcal{D}[\hat{L}]\rho_{(c)}(t_{n-1})\Delta t + \sqrt{\kappa}\mathcal{H}[\hat{L}\mathrm{e}^{\mathrm{i}\varphi}]\rho_{(c)}(t_{n-1})\Delta\mathrm{W} + o(\Delta t)\Big]}{\mathrm{Tr}\{\tilde{\rho}[n|\boldsymbol{x}_{0:n}^{\varphi}]\}},
\end{aligned}
\tag{3.78}
$$

which, when the CPTP map $\Phi_n[\,\cdot\,]$ has the same form as the one considered in Eq. (3.48), the

conditional state at time t_n becomes

$$\rho_{(c)}(t_n) = \rho_{(c)}(t_{n-1}) + \mathcal{L}_{\boldsymbol{y}_{0:n}}\rho_{(c)}(t_{n-1})\Delta t + \kappa \mathcal{D}[\hat{L}]\rho_{(c)}(t_{n-1})\Delta t + \sqrt{\kappa}\mathcal{H}[\hat{L}e^{i\varphi}]\rho_{(c)}(t_{n-1})\Delta W + o(\Delta t), \quad (3.79)$$

since $\mathrm{Tr}\{\tilde{\rho}[n|\boldsymbol{x}_{0:n}^\varphi]\} = 1$. Finally, to wrap up this derivation, we just need to identify the change in the conditional state as $\mathrm{d}\rho_{(c)}(t_n) = \rho_{(c)}(t_n) - \rho_{(c)}(t_{n-1})$ and keep the terms up to first order in $\mathrm{d}t$ to get the final form of the SME:

$$\mathrm{d}\rho_{(c)}(t) = \mathcal{L}_{\boldsymbol{y}_t}\rho_{(c)}(t)\mathrm{d}t + \kappa \mathcal{D}[\hat{L}]\rho_{(c)}(t)\mathrm{d}t + \sqrt{\kappa}\mathcal{H}[\hat{L}e^{i\varphi}]\rho_{(c)}(t)\mathrm{d}W, \quad (3.80)$$

with an associated measurement:

$$\mathrm{d}y = \sqrt{\kappa}\langle \hat{L}e^{i\varphi} + \hat{L}^\dagger e^{-i\varphi}\rangle \mathrm{d}t + \mathrm{d}W. \quad (3.81)$$

Note here that the Lindbladian depends only on the measurement results $\boldsymbol{y}_t$ and not, importantly, on the derivatives of the measurement results, i.e., $I(t) = \frac{\mathrm{d}y_t}{\mathrm{d}t} = \lim_{\Delta t \to 0} \frac{\Delta y_n}{\Delta t}$.

3.2.3 Polarimetric measurement

Consider a quasimonochromatic probe beam propagating in the y-direction, such that a general quantized multi-mode electric field [124] can be approximated as a single traveling-wave spacial mode with two orthogonal polarizations $\{\hat{a}_H, \hat{a}_V\}$ [120, 125]. Then, the positive frequency component for the monochromatic quantized electric field is [120, 125]:

$$\hat{\boldsymbol{E}}^{(+)} = \sqrt{\frac{2\pi\hbar\omega}{Ac\Delta t}}\left(\boldsymbol{e}_H\hat{a}_H + \boldsymbol{e}_V\hat{a}_V\right), \quad (3.82)$$

and $\hat{\boldsymbol{E}}^{(-)} = \hat{\boldsymbol{E}}^{(+)\dagger}$, A is the area of the beam and c the speed of light. Thus, to describe the beam it is sufficient with the photon annihilation operators $\{\hat{a}_H, \hat{a}_V\}$, associated with horizontal and vertical polarizations, respectively. Similarly as before, for each polarization component, the beam has been divided into a train of modes of duration Δt, each segment interacting for a time Δt with an atomic cloud of length Δx such that $c\Delta t \gg \Delta x$ [125].

After the segmented electric field at time $t_k = k\Delta t$ has interacted with the cloud for a time Δt, we perform a measurement of the transmitted light using a polarimeter: a polarization beam splitter and a differential photo detector [126, 127] made up of two photodiodes counting the

number of horizontal polarized photons $\langle \hat{N}_H \rangle$ and the number of vertical ones $\langle \hat{N}_V \rangle$, respectively [128]. The difference between the number of horizontal photons and the vertical ones is the expected value of the first Stokes vector. The Stokes operators are defined using the polarization modes $\{\hat{a}_H, \hat{a}_V\}$, and the position of their vector $\boldsymbol{S} = \left(\hat{S}_1, \hat{S}_2, \hat{S}_3 \right)$ on the Poincaré sphere represents the polarization state of the field. These operators can be introduced from the classical Stokes parameters [129] by changing intensities to photon-number operators:

$$\hat{S}_1 = \frac{1}{2} \left(\hat{N}_H - \hat{N}_V \right) = \frac{1}{2} \left(\hat{a}_H^\dagger \hat{a}_H - \hat{a}_V^\dagger \hat{a}_V \right) \tag{3.83}$$

$$\hat{S}_2 = \frac{1}{2} \left(\hat{N}_{H'} - \hat{N}_{V'} \right) = \frac{1}{2} \left(\hat{a}_{H'}^\dagger \hat{a}_{H'} - \hat{a}_{V'}^\dagger \hat{a}_{V'} \right) = \frac{1}{2} \left(\hat{a}_H^\dagger \hat{a}_V + \hat{a}_V^\dagger \hat{a}_H \right) \tag{3.84}$$

$$\hat{S}_3 = \frac{1}{2} \left(\hat{N}_+ - \hat{N}_- \right) = \frac{1}{2} \left(\hat{a}_+^\dagger \hat{a}_+ - \hat{a}_-^\dagger \hat{a}_- \right) = \frac{1}{2i} \left(\hat{a}_H^\dagger \hat{a}_V - \hat{a}_V^\dagger \hat{a}_H \right), \tag{3.85}$$

which satisfy the $SU(2)$ algebra $[\hat{S}_i, \hat{S}_j] = i\varepsilon_{ijk}\hat{S}_k$. We define the total photon number operator as $\hat{S}_0 = \hat{a}_H^\dagger \hat{a}_H + \hat{a}_V^\dagger \hat{a}_V$. The Stokes operator $\hat{S}_1$ represents linearly polarized light in either the horizontal or vertical direction, quantified through the annihilation and creation operators $\{\hat{a}_H, \hat{a}_V\}$ and $\{\hat{a}_H^\dagger, \hat{a}_V^\dagger\}$. Thus, note how the expected value of $\hat{S}_1$ yields the difference between the number of horizontal v.s. vertical photons. The operator $\hat{S}_2$ measures the difference in population of linearly polarized modes rotated in the ± 45 degree direction, and $\hat{S}_3$, of left and right circularly polarized light:

$$\hat{a}_{H'} = \frac{1}{\sqrt{2}} \left(\hat{a}_H + \hat{a}_V \right), \qquad\qquad \hat{a}_{V'} = \frac{1}{\sqrt{2}} \left(-\hat{a}_H + \hat{a}_V \right), \tag{3.86}$$

$$\hat{a}_+ = \frac{1}{\sqrt{2}} \left(\hat{a}_H - i\hat{a}_V \right), \qquad\qquad \hat{a}_- = \frac{1}{\sqrt{2}} \left(-\hat{a}_H - i\hat{a}_V \right). \tag{3.87}$$

Two polarizations perpendicular in real space will be represented by two vectors pointing in the opposite directions in the Poincaré sphere, and for example, measuring $|H\rangle$ or $|V\rangle$ with $\hat{S}_1$ will yield $+1$ and -1, respectively (see Fig. 3.4).

In general, the Hamiltonian governing the interaction of the atomic cloud with a monochromatic probe detuned off-resonance is given by:

$$\hat{H}_I = \sum_{g,e} \hat{\boldsymbol{E}}^{(-)} \cdot \frac{\boldsymbol{\alpha}_{g,e}}{\hbar \Delta_{g,e}} \cdot \hat{\boldsymbol{E}}^{(+)}, \tag{3.88}$$

where $\Delta_{g,e}$ is the detuning of the probe from the $g \to e$ transition (e.g. for a cloud of Rubidium
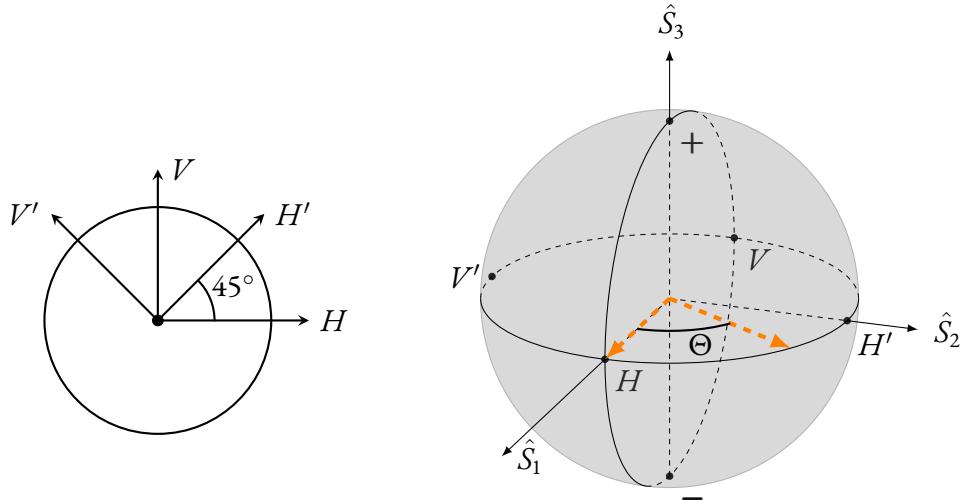
Figure 3.4: Stokes vectors on the Pointcaré sphere. (left) Definition of the $\pm 45°$ polarization basis $\{H', V'\}$ w.r.t. the linear polarization $\{H, V\}$. (right) Pointcaré sphere, a graphical representation describing the polarization of light. Points on the sphere correspond to different polarization states, with the equator representing linear polarization, and the poles representing circular polarization. Note that the rotation of linear polarization is confined to the equatorial plane of the Poincaré sphere. A rotation of the polarization by an angle $\Theta/2$ corresponds to a rotation of Θ in the Poincaré sphere. This can be observed from the state H', which can be generated from H by rotating the polarization by 45 degrees; as a result, in the Poincaré sphere, H' lies at a 90-degree angle from H.

^{87}Rb, that would be the transition of the D_2 line [130]). The term $\boldsymbol{\alpha}_{g,e}$ is the *atomic polarizability* tensor of that same transition and has the form:

$$\boldsymbol{\alpha}_{g,e} = \hat{P}_g \boldsymbol{d} \hat{P}_e \boldsymbol{d}^\dagger \hat{P}_g, \tag{3.89}$$

where $\boldsymbol{d}$ corresponds to the vector of dipole operators, and $\hat{P}_g$ and $\hat{P}_e$ are the projectors for the ground and excited states, respectively [120]. Ultimately, the Hamiltonian in Eq. (3.88) describes the interaction and eventual transmission of the light through the sample: a photon is annihilated from the probe field through $\hat{\boldsymbol{E}}^{(+)}$, which brings the atom from its ground state to its excited state via the dipole raising operator $\boldsymbol{d}^\dagger$. Then the excited atom returns to a ground state via $\boldsymbol{d}$ by emitting a photon into a transmitted probe mode through $\hat{\boldsymbol{E}}^{(-)}$ [125].

The atomic polarizability tensor of Eq. (3.89) can be decomposed under the group of rotations into three irreducible components: $\boldsymbol{\alpha}_{g,e} = \boldsymbol{\alpha}_{g,e}^{(0)} + \boldsymbol{\alpha}_{g,e}^{(1)} + \boldsymbol{\alpha}_{g,e}^{(2)}$, and thus, so can the Hamiltonian:

$$\hat{H}_I = \hat{H}_I^{(0)} + \hat{H}_I^{(1)} + \hat{H}_I^{(2)}, \tag{3.90}$$

where each term corresponds to a scalar, vectorial and tensorial component, respectively [125, 130, 131, 120]. The scalar or rank-0 component is invariant under the group of SO(3) rotations, and since it is state independent, it can be dropped. The vectorial term transforms like a rank-1 vector, and hence, it can be written as a linear combination of components of some vector like the spin $\hat{\boldsymbol{S}}$. Finally, the tensorial term, a rank-2 tensor, can also be neglected by further increasing the probe detuning [130, 131]. Therefore, the Hamiltonian reduces to:

$$\hat{H}_I = \hat{H}_I^{(1)} = \chi\, \hat{J}_y \hat{S}_3, \tag{3.91}$$

where the light, which propagates along the y-direction, couples to the y component of the angular momentum of the atoms through the linear Stokes component $\hat{S}_3$, since a rotation of the linear polarization by an angle $\Theta/2$ corresponds to a rotation on the Pointcaré sphere of an angle Θ about $\hat{S}_3$ (see Fig. 3.4).

Thus, thanks to Hamiltonian engineering [131, 132], the continuous non-demolition measurement required to monitor a system is implemented through a rotation of the polarization of the off-resonant probe beam proportional to the collective angular momentum of the atoms at an angle $\Theta \sim \chi\langle\hat{J}_y\rangle(t)$ [133, 120, 63], a magneto-optical effect known as Faraday rotation [134, 120].

Similarly to the approximation of the generalized Bloch sphere as a phase-plane in the linear-Gaussian (or Holstein-Primakoff) approximation, the Poincaré sphere can also be approximated as a plane under the conditions of a large photon number, an initial polarization aligned with $\hat{S}_1$, and small-angle Faraday rotation during transmission through the atomic cloud [120]. Then, the Stokes vectors can be approximated as:

$$\hat{S}_1 \sim \sqrt{N_{ph}/2}, \quad \hat{X}_{ph} = \hat{S}_2/\sqrt{\frac{N_{ph}}{2}}, \quad \hat{P}_{ph} = \hat{S}_3/\sqrt{\frac{N_{ph}}{2}}, \tag{3.92}$$

where N_{ph} is the number of photons in the probe. Now, the interaction Hamiltonian becomes,

$$\hat{H}_I = \chi\sqrt{N_{ph}/2}\, \hat{J}_y \hat{P}_{ph}, \tag{3.93}$$

and a rotation of the polarization along the axis $\hat{S}_3$ becomes a displacement along $\hat{X}_{ph}$. Hence, measurement of the Faraday rotation of the probe is fully characterized by the eigenvectors of $\hat{X}_{ph}$, i.e. $\left|\hat{X}_{ph}\right\rangle$.

Therefore, if now we discretize the system-probe evolution just like in Sec. 3.1, the interaction Hamiltonian for a segment Δt is:

$$\hat{H}_I^{(n)} := i\chi\sqrt{\frac{N_{ph}}{4\Delta t}}\,\hat{J}_y\left(\hat{b}_n^\dagger - \hat{b}_n\right),\tag{3.94}$$

where we have written the momentum operator $\hat{P}_{ph}$ in terms of creation and annihilation operators (see Eq. (1.184)) and then discretized the modes according to Eq. (3.18). Hence, the unitary evolution operator taking the state of the system from t_n to $t_n + \Delta t$ reads as

$$\hat{U}_{\Delta t}^{(n)} = e^{-i\hat{H}_I^{(n)}\Delta t} = \exp\left(\sqrt{\kappa\Delta t}\,\hat{J}_y\left(\hat{b}_n^\dagger - \hat{b}_n\right)\right),\tag{3.95}$$

where $\kappa = \chi^2 N_{ph}/4$. Since angular momentum operators are Hermitian, the unitary form above matches Eq. (3.19) for $\hat{L} := \hat{J}_y$. Furthermore, since the measurement of the Faraday rotation of the polarization in the Holstein-Primakoff approximation corresponds to a measurement of the displacement along $\hat{X}_{ph}$, which is equivalent to Eq. (3.61) but with $\varphi = 0$, then this type of polarimetric measurement is equivalent to a homodyne measurement. And thus, we can write the conditional state evolution as:

$$d\rho_{(c)}(t) = \mathcal{L}_{y}\rho_{(c)}(t)dt + \kappa\mathcal{D}[\hat{J}_y]\rho_{(c)}(t)dt + \sqrt{\kappa}\mathcal{H}[\hat{J}_y]\rho_{(c)}(t)dW,\tag{3.96}$$

with a continuous measurement output:

$$dy = 2\sqrt{\kappa}\langle\hat{J}_y\rangle dt + dW.\tag{3.97}$$

3.3 MEASUREMENT-BASED FEEDBACK

There are two types of measurement-based feedback one might consider. The first is feedback that utilizes the entire history of measurement outcomes, often referred to as *state-based* or *Bayesian* feedback [119]. The second is *Markovian* feedback [104], where we feed back in an instantaneous measurement signal, e.g. a photocurrent, modeled at each time step as a stochastic increment.

3.3.1 Bayesian feedback

In the previous sections, we derived the SME for homodyne and photodetection measurements, accounting for the possibility of measurement-based feedback. In particular, we considered a specific class of feedback in which the entire history of measurement outcomes $\boldsymbol{y}_{\leq t}$ enters the dynamics only through terms proportional to $\mathrm{d}t$, such that $\mathcal{L}_{\boldsymbol{y}_{\leq t}}\mathrm{d}t = O(\mathrm{d}t)$, and not through stochastic contributions like $\mathrm{d}W$ or the photocurrent $\mathrm{d}y$. We refer to this subclass of measurement-based feedback as Bayesian feedback following the terminology of Ref. [119].

Because the GKSL generator appears at order $\mathrm{d}t$, the corresponding feedback-modified evolution map can be expanded as shown in Eq. (3.48). This leads directly to the final form of the SME for homodyne detection with Bayesian feedback, given in Eq. (3.80). If now we consider the case where the GKSL generator takes the form

$$\mathcal{L}_{\boldsymbol{y}_t}\rho_{(c)}(t) = -\mathrm{i}\left[\hat{H} + u(t|\boldsymbol{y}_t)\hat{F}, \rho_{(c)}(t)\right], \tag{3.98}$$

where the control function $u(t|\boldsymbol{y}_t)$ depends on the measurement history $\boldsymbol{y}_t$ but not on its derivative, then the SME becomes

$$\mathrm{d}\rho_{(c)}(t) = -\mathrm{i}\left[\hat{H} + u(t|\boldsymbol{y}_t)\hat{F}, \rho_{(c)}(t)\right]\mathrm{d}t + \kappa\mathcal{D}[\hat{L}]\rho_{(c)}(t)\mathrm{d}t + \sqrt{\kappa}\mathcal{H}[\hat{L}]\rho_{(c)}(t)\mathrm{d}W. \tag{3.99}$$

3.3.2 Markovian feedback

We now contrast Bayesian feedback with another common class of measurement-based control: Markovian feedback. In Bayesian control, the feedback appears only at order $\mathrm{d}t$, as part of a Lindbladian contribution $\mathcal{L}_{\boldsymbol{y}_{\leq t}}\mathrm{d}t = O(\mathrm{d}t)$. In contrast, Markovian feedback modifies the SME through stochastic terms, with the feedback entering directly through the homodyne signal $\mathrm{d}y$. As a result, the evolution now includes terms of order $\sqrt{\mathrm{d}t}$ and the feedback cannot be described as a GKSL map at order $\mathrm{d}t$. In other words, to include Markovian feedback, the map in Eq. (3.48) has to be modified as:

$$\Phi_n = \exp\left\{\mathcal{L}_0\Delta t + \mathcal{L}_Y\Delta y_n\right\}, \tag{3.100}$$

where now the dependence on the measurement outcomes is limited to the term $\mathcal{L}_Y\Delta y_n$, with Δy_n being the discretized photocurrent $\mathrm{d}y$ at time $n = t/\Delta t$ and $\mathcal{L}_Y$ its associated Lindbladian. The other term, $\mathcal{L}_0\Delta t$, contains the internal dynamics of the atoms, e.g. Larmor precession and

dephasing. Expanding the map above yields:

$$
\begin{aligned}
\Phi_n &= \mathbb{I} + \mathcal{L}_0 \Delta t + \mathcal{L}_Y \Delta y_n + \frac{1}{2} \left(\mathcal{L}_Y\right)^2 \left(\Delta y_n\right)^2 + o(\Delta t) \\
&= \left(\mathbb{I} + \mathcal{L}_0 \Delta t + o(\Delta t)\right) \left(\mathbb{I} + \Delta y_n \mathcal{L}_Y + \frac{1}{2} \left(\Delta y_n\right)^2 \left(\mathcal{L}_Y\right)^2 + o(\Delta t)\right) \\
&= \Xi_0 \circ \mathfrak{F}_n,
\end{aligned}
\tag{3.101}
$$

where the last two expressions correspond to the lowest-order Suzuki–Trotter decomposition into deterministic and stochastic components, with

$$
\Xi_0 := \exp\left\{\mathcal{L}_0 \Delta t\right\} \quad \text{and} \quad \mathfrak{F}_n := \exp\left\{\mathcal{L}_Y \Delta y_n\right\}.
\tag{3.102}
$$

Thus, similarly to Eq. (3.24), the conditional state evolves from time $t_{n-1} = t - \mathrm{d}t$ to $t_n = t$ according to:

$$
\rho_{(c)}(t_n) = \frac{\mathfrak{F}_n\left[\Xi_0\left[\hat{E}_{y_n} \rho_{(c)}(t_{n-1}) \hat{E}_{y_n}^\dagger\right]\right]}{\mathrm{Tr}\left\{\mathfrak{F}_n\left[\Xi_0\left[\hat{E}_{y_n} \rho_{(c)}(t_{n-1}) \hat{E}_{y_n}^\dagger\right]\right]\right\}},
\tag{3.103}
$$

where $\hat{E}_{y_n}$ is a general operator representing the weak measurement. In the case of a photocount measurement, the operators read $\hat{E}_{y_n} = \langle y_n | \hat{U}_{\Delta t} |0\rangle$, where $y_n = 0, 1$. For a homodyne measurement, the measurement operators are $\hat{E}_{y_n} = \langle x_n^\varphi | \hat{U}_{\Delta t} |0\rangle$, where $|x_n^\varphi\rangle$ is the eigenstate of Eq. (3.61). As shown in Sec. 3.2.2, the eigenvalue x_n^φ is a Gaussian random variable, which defines the homodyne signal $\Delta y_n \propto x_n^\varphi$.

In the formulation introduced by Wiseman [104, 105], Markovian feedback is realized via a unitary Hamiltonian evolution driven by the instantaneous measurement record:

$$
\hat{H}_f = \hat{F}\mathrm{d}y,
\tag{3.104}
$$

where $\hat{F}$ is a fixed Hermitian feedback operator and $\mathrm{d}y$ is given by Eq. (3.81). Therefore, the feedback map $\mathfrak{F}_n$ is actually an unitary evolution that reads as:

$$
\mathfrak{F}_n := \hat{U}_f \cdot \hat{U}_f^\dagger = \exp\left\{-\mathrm{i}\hat{F}\Delta y_n\right\} \cdot \exp\left\{\mathrm{i}\hat{F}\Delta y_n\right\},
\tag{3.105}
$$

where Δy_n follows Eq. (3.73). Furthermore, this map can be expanded to first order in Δt as

$$\mathfrak{F}_n = \hat{U}_f \cdot \hat{U}_f^\dagger = \mathbb{I} - \mathrm{i}\,\Delta y_n \left[\hat{F}, \cdot \right] + \mathcal{D}[\hat{F}] \cdot \Delta t + o(\mathrm{d}t). \qquad (3.106)$$

Therefore, to derive a SME which accounts for Markovian feedback, we simply have to apply the feedback unitary to the continuous measurement map, where we have set $\varphi = 0$:

$$\begin{aligned}
\rho_{(c)}(t_n) &= \hat{U}_f \frac{\Xi_0 \left[\hat{E}_{y_n} \rho_{(c)}(t_{n-1}) \hat{E}_{y_n}^\dagger \right]}{\mathrm{Tr}\left\{ \Xi_0 \left[\hat{E}_{y_n} \rho_{(c)}(t_{n-1}) \hat{E}_{y_n}^\dagger \right] \right\}} \hat{U}_f^\dagger \\
&= \hat{U}_f \left[\rho_{(c)}(t_{n-1}) + \mathcal{L}_0 \rho_{(c)}(t_{n-1}) \Delta t + \kappa \mathcal{D}[\hat{L}] \rho_{(c)}(t_{n-1}) \Delta t + \sqrt{\kappa} \mathcal{H}[\hat{L}] \rho_{(c)}(t_{n-1}) \Delta W \right] \hat{U}_f^\dagger \\
&= \rho_{(c)}(t_{n-1}) + \mathcal{L}_0 \rho_{(c)}(t_{n-1}) \Delta t + \kappa \mathcal{D}[\hat{L}] \rho_{(c)}(t_{n-1}) \Delta t + \mathcal{D}[\hat{F}] \rho_{(c)}(t_{n-1}) \Delta t - \mathrm{i}\,\Delta W [\hat{F}, \rho_{(c)}(t_{n-1})] \\
&\quad - \mathrm{i}\sqrt{\kappa} \left[\hat{F}, \hat{L} \rho_{(c)}(t_{n-1}) + \rho_{(c)}(t_{n-1}) \hat{L}^\dagger \right] \Delta t + \sqrt{\kappa} \mathcal{H}[\hat{L}] \rho_{(c)}(t_{n-1}) \Delta W + o(\Delta t). \qquad (3.107)
\end{aligned}$$

The expression above can be reformulated using the following identities:

$$\begin{aligned}
\kappa \mathcal{D}[\hat{L}] \rho_{(c)}(t) + \mathcal{D}[\hat{F}] \rho_{(c)}(t) &- \mathrm{i}\sqrt{\kappa} \left[\hat{F}, \hat{L} \rho_{(c)}(t) + \rho_{(c)}(t) \hat{L}^\dagger \right] = \\
&= -\mathrm{i}\sqrt{\kappa} \left[\frac{\hat{L}^\dagger \hat{F} + \hat{F} \hat{L}}{2}, \rho_{(c)}(t) \right] + \mathcal{D}[\sqrt{\kappa}\hat{L} - \mathrm{i}\hat{F}] \rho_{(c)}(t), \qquad (3.108)
\end{aligned}$$

$$\sqrt{\kappa} \mathcal{H}[\hat{L}] \rho_{(c)}(t) - \mathrm{i} \left[\hat{F}, \rho_{(c)}(t) \right] = \mathcal{H}[\sqrt{\kappa}\hat{L} - \mathrm{i}\hat{F}] \rho_{(c)}(t). \qquad (3.109)$$

Namely,

$$\begin{aligned}
\rho_{(c)}(t_n) &= \rho_{(c)}(t_{n-1}) + \mathcal{L}_{\boldsymbol{y}_{0:n}} \rho_{(c)}(t_{n-1}) \Delta t - \mathrm{i}\sqrt{\kappa} \left[\frac{\hat{L}^\dagger \hat{F} + \hat{F} \hat{L}}{2}, \rho_{(c)}(t_{n-1}) \right] \Delta t \\
&\quad + \mathcal{D}[\sqrt{\kappa}\hat{L} - \mathrm{i}\hat{F}] \rho_{(c)}(t_{n-1}) \Delta t + \mathcal{H}[\sqrt{\kappa}\hat{L} - \mathrm{i}\hat{F}] \rho_{(c)}(t_{n-1}) \Delta W + o(\mathrm{d}t). \qquad (3.110)
\end{aligned}$$

Finally, if the internal evolution represented by the Lindbladian $\mathcal{L}_0$ is of the form:

$$\mathcal{L}_0 \rho_{(c)}(t) = -\mathrm{i} \left[\hat{H}, \rho_{(c)}(t) \right], \qquad (3.111)$$

then, Eq. (3.110) in the limit of $\Delta t \to 0$, can be written as

$$d\rho_{(c)}(t) = -i \left[\hat{H}_{\text{eff}}, \rho_{(c)}(t) \right] dt + \mathcal{D}[\hat{L}_{\text{eff}}]\rho_{(c)}(t) + \mathcal{H}[\hat{L}_{\text{eff}}]\rho_{(c)}(t)dW, \qquad (3.112)$$

where

$$\hat{H}_{\text{eff}} = \hat{H} + \frac{\sqrt{\kappa}}{2} \left(\hat{L}^\dagger \hat{F} + \hat{F}\hat{L} \right), \qquad (3.113)$$

$$\hat{L}_{\text{eff}} = \sqrt{\kappa}\hat{L} - i\hat{F}. \qquad (3.114)$$

4

Ultimate precision limits in noisy systems

I N   Q U A N T U M   M E T R O L O G Y , finding the fundamental limits to precision is essential for verifying optimal sensing strategies. Thus, this chapter focuses on the derivation and analysis of a lower bound on the BCRB introduced in Sec. 2.2.3 for frequency estimation in the presence of atomic dephasing and field fluctuations. This bound is referred throughout this thesis as either the classically-simulated (CS) limit or the quantum limit. The term *classically-simulated* stems from its derivation via the decomposition of quantum channels as a convex mixture of unitaries, while the term *quantum limit* highlights that, given a particular form and strength of noise, no strategy involving any possible quantum effects may surpass it.

Importantly, as will become clear through the derivation, this bound is entirely independent of the choice of initial quantum state, measurement, or measurement-based feedback. It depends only on the noise model of the system (in our case, local and collective dephasing along the field direction) and the fluctuating strength of the signal we aim to track. As such, attaining this limit would certify that the entire sensing protocol is optimal, i.e. that our particular choice of initial state, measurement, estimator and measurement-based control yields the best possible sensitivity.

THIS CHAPTER, which one can view as an extended derivation or proof, is organized as follows: we begin in Sec. 4.1 by introducing the most general discrete-time evolution of a quantum state, alternating between internal system dynamics and measurement updates, which is connected to the conditional evolution of Chap. 3. Next, we specify the types of quantum channels used in our model: the feedback map, and the collective and local dephasing maps.

In Sec. 4.2.1, we turn to the central step of this proof: the convex decomposition of the likelihood $p(\boldsymbol{y}_{0:k}|\boldsymbol{\omega}_{0:k})$, which describes the probability of observing measurement outputs $\boldsymbol{y}_{0:k}$ given a signal trajectory $\boldsymbol{\omega}_{0:k}$. This likelihood is rewritten as a mixture of two other conditional probabilities: one that contains the ω-dependence via a classical mixing distribution, and another "fictitious" likelihood encoding all the measurement record, independent of the signal. To enable this decomposition, both collective and local dephasing channels must be expressed as Gaussian-weighted integrals over unitary operations (see Sec. 4.2.2 and Sec. 4.2.3). This allows us to reformulate $p(\boldsymbol{y}_{0:k}|\boldsymbol{\omega}_{0:k})$ accordingly, which in turn enables us to upper-bound the FI of the marginal likelihood $p(\boldsymbol{y}_{0:k}|\omega_k)$ in Sec. 4.2.5.

This upper-bound on the FI can be analytically derived (see Sec. 4.2.7), since it ultimately involves computing the FI of several Gaussian distributions, each given by the inverse of their variance. In the case of fluctuating fields, finding such a variance requires solving a recursive relation, which can be done explicitly for the problem at hand. Taking the continuous-time limit of the resulting expression in Sec. 4.2.8, yields the CS limit or quantum limit. Finally, in Sec. 4.2.9 we extend the analysis to account for scenarios where the number of atoms N fluctuates between experimental runs, and show how the bounds can be appropriately modified.

4.1   DISCRETE-TIME PICTURE OF MEASUREMENT-BASED FEEDBACK SCHEME

Consider the discrete evolution of a quantum state that alternates between its intrinsic dynamics and measurement updates, as shown in Eq. (3.22). In this picture, each measurement is modeled by a set of measurement Kraus operators $\hat{E}_{y_k}$, which form a POVM $\{\hat{E}_{y_k}^\dagger \hat{E}_{y_k}\}_k$ whose elements fulfill $\sum_k \hat{E}_{y_k}^\dagger \hat{E}_{y_k} = \mathbb{I}$ and whose outcome y_k forms part of the discretized measurement record $\boldsymbol{y}_{0:k} = \{y_0, y_1, \ldots, y_k\}$. The quantum channel governing the evolution of the state in between measurements is a generic map

$$\Phi_k := \Phi_{\Delta t}(\boldsymbol{y}_{0:k}, \omega_k), \qquad (4.1)$$

which acts on the state for a time Δt and depends on a parameter ω_k, as well as, potentially, on all previous measurement records $\boldsymbol{y}_{0:k} = \{y_0, y_1, \ldots, y_k\}$. The frequency ω_k is itself also a time-

discretized element from a frequency trajectory $\boldsymbol{\omega}_{0:k} = \{\omega_0, \omega_1, \ldots, \omega_k\}$, whose elements will later be assumed to be drawn from a probability distribution at each time-step. Therefore, the state at time $t = k\Delta t$, conditional on the measurement outcomes $\boldsymbol{y}_{0:k} = \{y_0, y_1, \ldots, y_k\}$, reads as

$$\rho[k|\boldsymbol{y}_{0:k}] = \frac{\Phi_k\left[\hat{E}_{y_k}\Phi_{k-1}\left[\hat{E}_{y_{k-1}}\ldots\Phi_1\left[\hat{E}_{y_1}\Phi_0\left[\hat{E}_{y_0}\rho_0\,\hat{E}^\dagger_{y_0}\right]\hat{E}^\dagger_{y_1}\right]\ldots\hat{E}^\dagger_{y_{k-1}}\right]\hat{E}^\dagger_{y_k}\right]}{p(\boldsymbol{y}_{0:k}|\boldsymbol{\omega}_{0:k})}, \tag{4.2}$$

where we have used a similar notation to Eq. (3.23), $\rho[k|\boldsymbol{y}_{0:k}] := \rho_{(c)}(k\Delta t) = \rho(k\Delta t|\boldsymbol{y}_{k\Delta t})$, to refer to the discretized conditional state at time $t = k\Delta t$. Note that ρ_0 denotes the initial state of the atoms before any operation is applied, and hence, is different from the state $\rho[0|\boldsymbol{y}_0]$, which is obtained after getting the first outcome y_0 and evolving the state with the map Φ_0, dependent on the measurement y_0 and the frequency ω_0. The denominator is the discretized version of the likelihood $p(\boldsymbol{y}_t|\boldsymbol{\omega}_t)$, i.e. the probability of measuring $\boldsymbol{y}_{0:k} = \{y_0, y_1, \ldots, y_k\}$ given field inputs $\boldsymbol{\omega}_{0:k} = \{\omega_0, \omega_1, \ldots, \omega_k\}$. Namely,

$$p(\boldsymbol{y}_{0:k}|\boldsymbol{\omega}_{0:k}) = \text{Tr}\left\{\Phi_k\left[\hat{E}_{y_k}\Phi_{k-1}\left[\hat{E}_{y_{k-1}}\ldots\Phi_1\left[\hat{E}_{y_1}\Phi_0\left[\hat{E}_{y_0}\rho_0\,\hat{E}^\dagger_{y_0}\right]\hat{E}^\dagger_{y_1}\right]\ldots\hat{E}^\dagger_{y_{k-1}}\right]\hat{E}^\dagger_{y_k}\right]\right\}, \tag{4.3}$$

which will later play a crucial role when bounding the aMSE, introduced in Definition 2.5.

If we now focus on a single time interval Δt in the discrete evolution described above, we can relate the two time-consecutive conditional states, $\rho[k\text{-}1|\boldsymbol{y}_{0:k-1}]$ and $\rho[k|\boldsymbol{y}_{0:k}]$ as:

$$\rho[k|\boldsymbol{y}_{0:k}] = \frac{\Phi_k\left[\hat{E}_{y_k}\rho[k\text{-}1|\boldsymbol{y}_{0:k-1}]\hat{E}^\dagger_{y_k}\right]}{\text{Tr}\left\{\Phi_k\left[\hat{E}_{y_k}\rho[k\text{-}1|\boldsymbol{y}_{0:k-1}]\hat{E}^\dagger_{y_k}\right]\right\}}, \tag{4.4}$$

with the first time step defined as

$$\rho[0|\boldsymbol{y}_0] = \frac{\Phi_0\left[\hat{E}_{y_0}\rho_0\,\hat{E}^\dagger_{y_0}\right]}{\text{Tr}\left\{\Phi_0\left[\hat{E}_{y_0}\rho_0\,\hat{E}^\dagger_{y_0}\right]\right\}}. \tag{4.5}$$

As discussed in Sec. 3.3.1, we focus here on Bayesian feedback[*], in which case, with $\mathcal{L}^\omega_{\boldsymbol{y}_{0:k}}\Delta t =$

---

[*]Nonetheless, our analysis can be straightforwardly generalized to also Markovian feedback or any other form of measurement-based feedback, even with general Lévy processes [65].
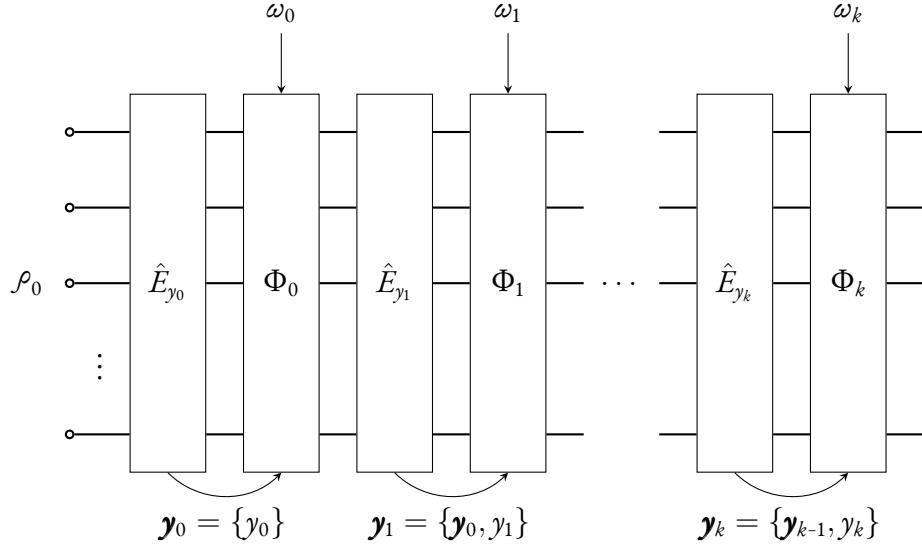
Figure 4.1: Quantum circuit representation of conditional evolution via sequential measurements. Scheme depicting a quantum circuit representing a sequential measurement on a system of N atoms with an initial density matrix ρ_0. Each step involves a POVM $\hat{E}_{y_k}$, which depends on a measurement outcome y_k, followed by an evolution through a quantum channel Φ_k. This map depends on an input parameter ω_k and all prior measurement outcomes $\boldsymbol{y}_{0:k}$, which are progressively collected as the system evolves.

$O(\Delta t)$:

$$\Phi_k = e^{\mathcal{L}^{\omega}_{\boldsymbol{y}_{0:k}} \Delta t}. \tag{4.6}$$

The overall dynamical generator $\mathcal{L}^{\omega}_{\boldsymbol{y}_{0:k}}$ depends on all previous outcomes $\boldsymbol{y}_{0:k}$ through the measurement-based feedback, as well as accounting for ω-encoding but also importantly, decoherence. Nonetheless, it can always be decomposed into two parts: one corresponding to the dynamics generated by the ω-encoding, and the other corresponding to the feedback based on previous measurements by Trotter-Suzuki arguments as $\Delta t \to 0$:

$$\Phi_k = e^{(\mathcal{L}_{\omega_k} + \mathcal{L}^f_{\boldsymbol{y}_{0:k}})\Delta t} = e^{\mathcal{L}_{\omega_k} \Delta t} \circ e^{\mathcal{L}^f_{\boldsymbol{y}_{0:k}} \Delta t} + O(\Delta t^2) \tag{4.7}$$

$$= \Xi_{\omega_k} \circ \mathfrak{F}_{\boldsymbol{y}_{0:k}} + O(\Delta t^2)$$

where Ξ_{ω_k} represents the portion of the evolution due to the ω-encoding and decoherence (i.e. the intrinsic, noisy dynamics), while $\mathfrak{F}_{\boldsymbol{y}_{0:k}}$ accounts for the measurement-based feedback.

As a result, the discrete conditional evolution given in Eq. (4.2) can be equivalently written in a form that explicitly separates the state evolution into three parts: the measurement update, the

feedback and the remaining internal dynamics:

$$\rho[k|\boldsymbol{y}_{0:k}] = \frac{\Xi_{\omega_k}\Big[\mathfrak{F}_{\boldsymbol{y}_{0:k}}\Big[\hat{E}_{y_k}\ldots\Xi_{\omega_1}\Big[\mathfrak{F}_{\boldsymbol{y}_{0:1}}\Big[\hat{E}_{y_1}\Xi_{\omega_0}\Big[\mathfrak{F}_{\boldsymbol{y}_0}\Big[\hat{E}_{y_0}\rho_0\hat{E}_{y_0}^\dagger\Big]\Big]\ldots\hat{E}_{y_1}^\dagger\Big]\Big]\ldots\hat{E}_{y_k}^\dagger\Big]\Big]}{p(\boldsymbol{y}_{0:k}|\boldsymbol{\omega}_{0:k})}, \tag{4.8}$$

with its likelihood now being:

$$p(\boldsymbol{y}_{0:k}|\boldsymbol{\omega}_{0:k}) =$$
$$= \mathrm{Tr}\Big\{\Xi_{\omega_k}\Big[\mathfrak{F}_{\boldsymbol{y}_{0:k}}\Big[\hat{E}_{y_k}\ldots\Xi_{\omega_1}\Big[\mathfrak{F}_{\boldsymbol{y}_{0:1}}\Big[\hat{E}_{y_1}\Xi_{\omega_0}\Big[\mathfrak{F}_{\boldsymbol{y}_0}\Big[\hat{E}_{y_0}\rho_0\hat{E}_{y_0}^\dagger\Big]\Big]\ldots\hat{E}_{y_1}^\dagger\Big]\Big]\ldots\hat{E}_{y_k}^\dagger\Big]\Big]\Big\}. \tag{4.9}$$

Hence, the proof that follows, which is fully based on the form of the map Ξ_{ω_k} responsible for noisy ω-encoding, applies to any form of measurement-based feedback.

4.2 PRECISION BOUND FOR ANY PROTOCOL WITH LOCAL AND GLOBAL NOISE

We may further assume that the internal dynamics of the system consists of an ω-encoded unitary evolution, collective dephasing and local dephasing. The map representing this internal evolution, Ξ_ω, can therefore be decomposed into two additional maps,

$$\Xi_\omega = \Omega \circ \Lambda_\omega, \tag{4.10}$$

where Ω denotes the non-unitary evolution arising in between measurements due to the collective decoherence (of strength κ_c), and the channel Λ_ω accounts for both the unitary frequency-encoding and the non-unitary local decoherence (of strength κ_ℓ). Even for non-commuting maps, we can apply the Suzuki-Trotter expansion to first order in Δt and split up the map Ω and Λ_ω as required. Then, Eq. (4.8) becomes

$$\rho[k|\boldsymbol{y}_{0:k}] =$$
$$= \frac{\Omega_k\Big[\Lambda_{\omega_k}\Big[\mathfrak{F}_{\boldsymbol{y}_{0:k}}\Big[\hat{E}_{y_k}\ldots\Omega_1\Big[\Lambda_{\omega_1}\Big[\mathfrak{F}_{\boldsymbol{y}_{0:1}}\Big[\hat{E}_{y_1}\Omega_0\Big[\Lambda_{\omega_0}\Big[\mathfrak{F}_{\boldsymbol{y}_0}\Big[\hat{E}_{y_0}\rho_0\hat{E}_{y_0}^\dagger\Big]\Big]\hat{E}_{y_1}^\dagger\Big]\Big]\ldots\hat{E}_{y_k}^\dagger\Big]\Big]}{p(\boldsymbol{y}_{0:k}|\boldsymbol{\omega}_{0:k})}, \tag{4.11}$$

such that

$$
\begin{aligned}
p(\boldsymbol{y}_{0:k}|\boldsymbol{\omega}_{0:k}) &= \\
&= \mathrm{Tr}\Big\{\Omega_k\Big[\Lambda_{\omega_k}\Big[\mathfrak{F}\boldsymbol{y}_{0:k}\Big[\hat{E}_{y_k}\dots\Omega_1\Big[\Lambda_{\omega_1}\Big[\mathfrak{F}\boldsymbol{y}_{0:1}\Big[\hat{E}_{y_1}\Omega_0\Big[\Lambda_{\omega_0}\Big[\mathfrak{F}\boldsymbol{y}_0\Big[\hat{E}_{y_0}\rho_0\hat{E}_{y_0}^\dagger\Big]\Big]\hat{E}_{y_1}^\dagger\Big]\Big]\Big]\dots\hat{E}_{y_k}^\dagger\Big]\Big]\Big]\Big\}\,. \quad (4.12)
\end{aligned}
$$
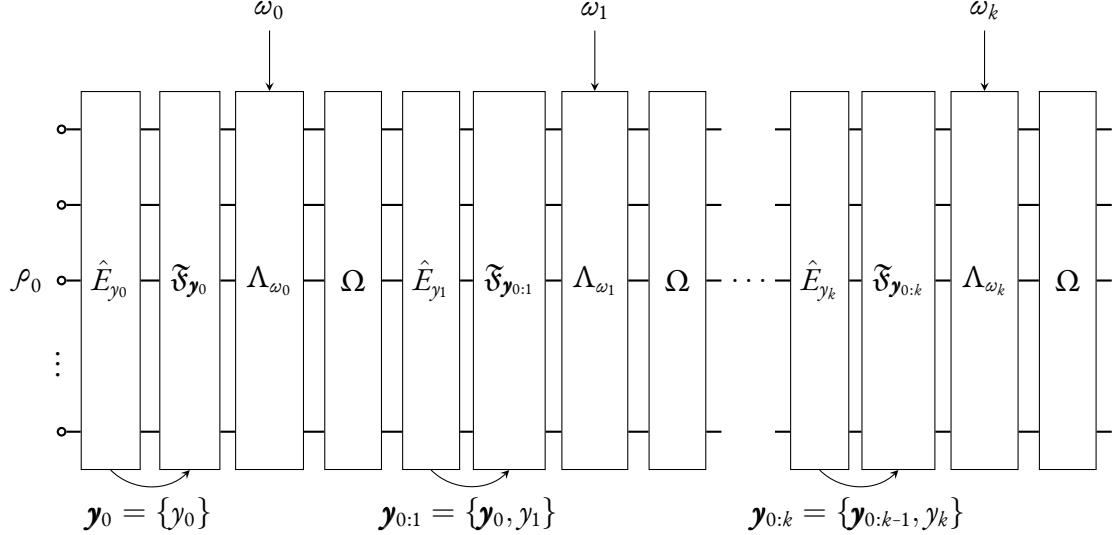
Figure 4.2: **Scheme illustrating a quantum circuit with sequential measurements, feedback and dephasing on a multi-qubit system.** The state evolves step-by step through a repeating sequence of operations: a measurement operator $\hat{E}_{y_j}$, a feedback map $\mathfrak{F}\boldsymbol{y}_{0:j}$ conditioned on past outcomes, and an internal dynamics channels Ω and Λ_{ω_j}. Here, Ω represents a collective dephasing map, while Λ_{ω_k} encodes both local dephasing and the unitary ω_k-dependent evolution. Measurement outcomes $\boldsymbol{y}_{0:k}$ are collected progressively and used to inform future feedback steps, allowing the system to evolve adaptively under both unitary and non-unitary processes.

4.2.1 CONVEX DECOMPOSITION OF THE LIKELIHOOD

As in our earlier work [69], which dealt only with collective decoherence, our motivation is to find a convex decomposition of the effective noisy ω-encoding map [65, 66], i.e. $\Omega\,[\Lambda_\omega\,[\,\cdot\,]]$ in Eq. (4.10), so that the discretised likelihood (4.12) can be decomposed as follows:

$$
p(\boldsymbol{y}_{0:k}|\boldsymbol{\omega}_{0:k}) = \int \mathcal{D}\boldsymbol{Z}_{0:k}\, q(\boldsymbol{Z}_{0:k}|\boldsymbol{\omega}_{0:k})\, p(\boldsymbol{y}_{0:k}|\boldsymbol{Z}_{0:k}) \quad (4.13)
$$

where $\boldsymbol{Z}_{0:k} = \{\boldsymbol{\zeta}_0, \boldsymbol{\zeta}_1, \dots, \boldsymbol{\zeta}_k\}$ is a sequence of sets, each containing N auxiliary frequency-like random variables. For instance, $\boldsymbol{\zeta}_\ell = \{\zeta_\ell^{(1)}, \zeta_\ell^{(2)}, \dots, \zeta_\ell^{(N)}\}$ indicates that within the ℓth step, the

first probe undergoes the Larmor precession for Δt with frequency $\zeta^{(1)}$, the second probe with $\zeta^{(2)}$, etc [65, 66].

While $q(\boldsymbol{\mathcal{Z}}_{0:k}|\boldsymbol{\omega}_{0:k})$ in Eq. (4.13) represents the mixing distribution, which contains all the dependence on the frequency trajectory $\boldsymbol{\omega}_{0:k}$, the term $p(y_{0:k}|\boldsymbol{\mathcal{Z}}_{0:k})$ can be interpreted as a *fictitious* likelihood of obtaining the measurement record $\boldsymbol{y}_{0:k} = \{y_j\}_{j=0}^{k}$. Here, the sequence $\boldsymbol{\mathcal{Z}}_{0:k}$ specifies a series of unitary maps that encode the frequency information in between discretized measurements and feedback operations:

$$p(\boldsymbol{y}_{0:k}|\boldsymbol{\mathcal{Z}}_{0:k}) =$$
$$= \mathrm{Tr}\left\{ \mathcal{U}_{\boldsymbol{\zeta}_k}\left[\mathfrak{F}_{\boldsymbol{y}_{0:k}}\left[\hat{E}_{y_k} \ldots \mathcal{U}_{\boldsymbol{\zeta}_1}\left[\mathfrak{F}_{\boldsymbol{y}_{0:1}}\left[\hat{E}_{y_1} \mathcal{U}_{\boldsymbol{\zeta}_0}\left[\mathfrak{F}_{\boldsymbol{y}_0}\left[\hat{E}_{y_0} \rho_0 \hat{E}_{y_0}^{\dagger} \right] \right] \hat{E}_{y_1}^{\dagger} \right] \right] \ldots \hat{E}_{y_k}^{\dagger} \right] \right\}. \qquad (4.14)$$

To justify the form of the convex decomposition in (4.13), we seek to represent the overall noisy encoding channel $\Omega\left[\Lambda_{\omega}[\,\cdot\,] \right]$ as a probabilistic mixture of unitaries. Rather than decomposing the full channel all at once, we handle its components separately: the collective map $\Omega[\cdot]$, which acts globally, and the local channel, which describes local dynamics and exhibits a tensor product structure with local dephasing and unitary evolution acting independently on each two-level system.

4.2.2   The map Ω as a convex mixture of unitaries

The channel Ω, which represents the evolution of the atomic state under collective dephasing, admits a representation as a convex mixture of unitaries. This follows from the following result established in [69]:

Theorem 4.1 (Map as a convex mixture of unitaries). *Given a unitary evolution governed by a Hamiltonian $\xi\hat{H}$*

$$\mathcal{U}_{\xi,\tau}[\,\cdot\,] = \mathrm{e}^{-\mathrm{i}\xi\hat{H}\tau} \cdot \mathrm{e}^{\mathrm{i}\xi\hat{H}\tau}, \qquad (4.15)$$

whose scalar encoding $\xi \in \mathbb{R}$ (frequency) is randomly distributed according to a Gaussian probability density

$$\xi \sim p_{\mu,\sigma}(\xi) = \mathcal{N}(\mu(\tau), \sigma^2(\tau)) = \frac{1}{\sqrt{2\pi\sigma^2(\tau)}} \exp\left\{ -\frac{(\xi - \mu(\tau))^2}{2\sigma^2(\tau)} \right\}, \qquad (4.16)$$

then, the quantum map Ω can be written as a convex mixture of these unitaries as:

$$\rho(\tau) = \Omega[\rho(0)] = \mathbb{E}_{p(\xi)}[\mathcal{U}_{\xi,\tau}[\rho(0)]] = \int \mathrm{d}\xi \, p_{\mu,\sigma}(\xi) \, e^{-i\xi\hat{H}\tau}\rho(0)e^{i\xi\hat{H}\tau}, \qquad (4.17)$$

if $\rho(\tau)$ corresponds to the solution of the following master equation:

$$\frac{\mathrm{d}\rho(\tau)}{\mathrm{d}\tau} = -i\omega(\tau)[\hat{H},\rho(\tau)] + \Gamma(\tau)\left(\hat{H}\rho(\tau)\hat{H} - \frac{1}{2}\{\hat{H}^2,\rho(\tau)\}\right) \qquad (4.18)$$

$$= -i\omega(\tau)[\hat{H},\rho(\tau)] - \frac{1}{2}\Gamma(\tau)\left[\hat{H},[\hat{H},\rho(\tau)]\right] \qquad (4.19)$$

with the time-dependent frequency and decay parameters being

$$\omega(\tau) = \mu(\tau) + \tau\dot{\mu}(\tau) \quad \text{and} \quad \Gamma(\tau) = 2\sigma^2(\tau)\tau\left(1 + \frac{\dot{\sigma}(\tau)}{\sigma(\tau)}\tau\right). \qquad (4.20)$$

Proof. *Available in App. D.1.*

Now consider a system of N spin$-1/2$ particles evolving for a time Δt under the dynamics of Eq. (D.5):

$$\frac{\mathrm{d}\rho(\Delta t)}{\mathrm{d}(\Delta t)} = -\frac{1}{2}\kappa_{\mathrm{c}}\left[\hat{J}_z,\left[\hat{J}_z,\rho(\Delta t)\right]\right], \qquad (4.21)$$

with frequency $\omega(\Delta t) = 0$, $\Gamma(\Delta t) = \kappa_{\mathrm{c}}$ and $\hat{H} = \hat{J}_z$. This particular choice of $\hat{H} = \hat{J}_z$ anticipates the structure of the SME considered in the next chapter, where the atomic sensor undergoes collective dephasing along the z-axis.

Then, the effective map Ω describing the evolution from $\rho(0)$ to $\rho(\Delta t)$ can be written as a mixture of unitary channels. Specifically, it takes the form:

$$\Omega[\,\cdot\,] = \int \mathrm{d}\xi \, p_{\mathrm{c}}(\xi) \, e^{-i\xi\hat{J}_z\Delta t} \cdot e^{i\xi\hat{J}_z\Delta t}, \qquad (4.22)$$

where the mixing probability of Eq. (D.2) is a Gaussian distribution $p_{\mathrm{c}}(\xi) = \mathcal{N}(0, V_{\mathrm{c}})$ with zero mean and variance:

$$V_{\mathrm{c}} := \kappa_{\mathrm{c}}/\Delta t. \qquad (4.23)$$

On the other hand, the overall map associated with the local dephasing is given by

$$\Lambda_\omega = e^{t\mathcal{L}}, \tag{4.24}$$

where $\mathcal{L}$ is the GKSL generator defined in Eq. (1.195), i.e.:

$$\frac{d\rho}{dt} = \mathcal{L}\rho = -i[\hat{H}, \rho(t)] + \sum_{i=1}^{N} \mathcal{D}[\hat{L}_i]\rho(t). \tag{4.25}$$

If we choose the Hamiltonian as $\hat{H} = \omega \hat{J}_z$, with $\hat{J}_z = \frac{1}{2}\sum_{i=1}^{N} \hat{\sigma}_z^{(i)}$ being the collective angular momentum operator in the z-direction, and define the local collapse operators as $\hat{L}_i = \sqrt{\kappa_\ell/2}\hat{\sigma}_z^{(i)}$, then the master equation becomes

$$\begin{aligned}
\frac{d\rho}{dt} &= -i\omega[\hat{J}_z, \rho(t)] + \frac{\kappa_\ell}{2}\sum_{i=1}^{N} \mathcal{D}[\hat{\sigma}_z^{(i)}]\rho(t) \\
&= \left(-i\frac{\omega}{2}\sum_{i=1}^{N}[\hat{\sigma}_z^{(i)}, \ \cdot \ ] + \frac{\kappa_\ell}{2}\sum_{i=1}^{N} \mathcal{D}[\hat{\sigma}_z^{(i)}] \ \cdot \ \right)\rho(t) \\
&= \left[\bigoplus_{i=1}^{N} \mathcal{L}^{(i)} \right]\rho, \tag{4.26}
\end{aligned}$$

where the subscript (i) denoting the position of $\hat{\sigma}_z$ in the tensor-product structure. Just as in the previous section, the choice of $\hat{H}$ and $\hat{L}_i$ reflects our atomic sensor model, where both the field and dephasing occur along the z-axis. More details regarding this model will be presented in the next chapter, i.e. Chap. 5. We then define the local generator for each spin as

$$\mathcal{L}^{(i)} = -i\frac{\omega}{2}[\hat{\sigma}_z^{(i)}, \ \cdot \ ] + \frac{\kappa_\ell}{2}\mathcal{D}[\hat{\sigma}_z^{(i)}] \ \cdot \ , \tag{4.27}$$

so that the overall map $\mathcal{L}$ is simply the direct sum of the individual contributions $\mathcal{L}^{(i)}$. The formal solution to the master equation above is given by the collective map Λ_ω, which can be written as a tensor product of individual CPTP maps acting on each atom:

$$\Lambda_\omega = e^{\bigoplus_{i} t\mathcal{L}^{(i)}} = \bigotimes_{i=1}^{N} e^{t\mathcal{L}^{(i)}} = \bigotimes_{i=1}^{N} \Lambda_\omega^{(i)}, \tag{4.28}$$

with the semigroup map $\Lambda_\omega^{(i)} = e^{t\mathcal{L}^{(i)}}$ defined by the GKSL generator $\mathcal{L}^{(i)}$ representing the unconditional evolution of the ith atom, i.e.:

$$\frac{\mathrm{d}\rho_i(t)}{\mathrm{d}t} = -i\omega[\hat{s}_z^{(i)}, \rho_i(t)] + 2\kappa_\ell \mathcal{D}[\hat{s}_z^{(i)}]\rho_i(t), \tag{4.29}$$

where $\hat{s}_z = \frac{1}{2}\hat{\sigma}_z$, and $\rho_i(t) = \mathrm{Tr}_{\forall \neq i}\{\rho(t)\}$ is the reduced state of the ith atom, i.e. the state after tracing out all atoms except the ith one. Applying again the result from Theorem 4.1, each local map $\Lambda_\omega^{(i)}$ can be expressed as a convex mixture of unitary channels:

$$\Lambda_\omega^{(i)}[\,\cdot\,] = \int \mathrm{d}v^{(i)}\, p_\ell(v^{(i)}|\omega)\, \mathcal{U}_{v^{(i)},\Delta t}[\,\cdot\,], \tag{4.30}$$

where the auxiliary variable $v^{(i)}$ is distributed according to $p_\ell(v^{(i)}|\omega) = \mathcal{N}(\omega, \mathrm{V}_\ell)$, i.e. a Gaussian with mean ω and variance

$$\mathrm{V}_\ell := 2\kappa_\ell/\Delta t. \tag{4.31}$$

The corresponding unitary channel $\mathcal{U}_{v^{(i)},\Delta t}[\,\cdot\,]$ is also parametrized w.r.t. the auxiliary variable $v^{(i)}$:

$$\mathcal{U}_{v^{(i)},\Delta t}[\,\cdot\,] = e^{-i\,v^{(i)}\hat{s}_z^{(i)}\Delta t} \,\cdot\, e^{i\,v^{(i)}\hat{s}_z^{(i)}\Delta t}. \tag{4.32}$$

Putting everything together, the overall local map Λ_ω defined in Eq. (4.24) is equivalent to a convex combination of tensor products of unitary maps:

$$\Lambda_\omega[\,\cdot\,] = \bigotimes_{i=1}^N \Lambda_\omega^{(i)}[\,\cdot\,] = \int \mathcal{D}\boldsymbol{v}\, \wp_\ell(\boldsymbol{v}|\omega) \bigotimes_{i=1}^N \mathcal{U}_{v^{(i)},\Delta t}[\,\cdot\,] = \int \mathcal{D}\boldsymbol{v}\, \wp_\ell(\boldsymbol{v}|\omega)\mathcal{U}_{\boldsymbol{v}}[\,\cdot\,], \tag{4.33}$$

where $\boldsymbol{v} = (v^{(1)}, \ldots, v^{(i)}, \ldots, v^{(N)})$, the integration measure is $\mathcal{D}\boldsymbol{v} = \prod_{i=1}^N \mathrm{d}v^{(i)}$, and the joint product distribution is $\wp_\ell(\boldsymbol{v}|\omega) = \prod_{i=1}^N p_\ell(v^{(i)}|\omega)$. Note that since $\exp(A)\otimes\exp(B) = \exp(A \oplus B)$, then the overall unitary channel takes the explicit form:

$$\mathcal{U}_{\boldsymbol{v}}[\,\cdot\,] := \bigotimes_{i=1}^N \mathcal{U}_{v^{(i)},\Delta t}[\,\cdot\,] = e^{-i\Delta t \sum_{i=1}^N v^{(i)}\hat{j}_z^{(i)}} \,\cdot\, e^{i\Delta t \sum_{i=1}^N v^{(i)}\hat{j}_z^{(i)}} \tag{4.34}$$

where $\hat{j}_z^{(i)} = \underbrace{\mathbb{I} \otimes \cdots \otimes \mathbb{I}}_{i-1} \otimes \hat{s}_z^{(i)} \otimes \underbrace{\mathbb{I} \otimes \cdots \otimes \mathbb{I}}_{N-i}$ denotes the appropriately embedded single-site spin operator $\hat{s}_z^{(i)} = \frac{1}{2}\hat{\sigma}_z^{(i)}$ in the full tensor-product Hilbert space.

The overall evolution of the system under ω-encoding, including both collective and local decoherence effects, is described by the combined map $\Omega[\Lambda_\omega[\,\cdot\,]]$. Both the collective map $\Omega[\,\cdot\,]$ and the local channel $\Lambda_\omega[\,\cdot\,]$ have already been represented as a convex combination of unitaries in Eq. (4.22) and Eq. (4.33), respectively. To now represent the overall map $\Omega[\Lambda_\omega[\,\cdot\,]]$ as a convex mixture of unitaries, the first step is to combine the collective and local channels as:

$$\Omega[\Lambda_\omega[\,\cdot\,]] = \int \mathrm{d}\xi\, p_{\mathrm{c}}(\xi) \int \mathcal{D}\boldsymbol{v}\, \wp_\ell(\boldsymbol{v}|\omega)\, \mathcal{U}_{\xi,\boldsymbol{v}}[\,\cdot\,] \tag{4.35}$$

where

$$\mathcal{U}_{\xi,\boldsymbol{v}}[\,\cdot\,] = \mathrm{e}^{-i\Delta t \sum_{i=1}^{N}(\xi+v^{(i)})\hat{\jmath}_z^{(i)}} \cdot \mathrm{e}^{i\Delta t \sum_{i=1}^{N}(\xi+v^{(i)})\hat{\jmath}_z^{(i)}}. \tag{4.36}$$

Here, $p_{\mathrm{c}}(\xi)$ is the Gaussian probability density associated with the collective dephasing, while $\wp_\ell(\boldsymbol{v}|\omega) = \prod_{i=1}^{N} p_\ell(v^{(i)}|\omega)$ is the product of local densities.

In the above expression, the unitary operator depends on the sum $\xi + v^{(i)}$, meaning that the collective parameter ξ and the local variables $v^{(i)}$ appear together. To recast the overall map into a form that is more amenable to classical simulation, we redefine the integration variables by setting

$$v^{(i)} \rightarrow \zeta^{(i)} - \xi, \tag{4.37}$$

for each i. This change of variables incorporates the contribution of ξ into a new effective variable $\zeta^{(i)}$ for each particle. Consequently, the unitary operator becomes solely a function of $\zeta^{(i)}$, thereby simplifying the overall map into a convex combination of tensor products of local unitaries:

$$\Omega[\Lambda_\omega[\,\cdot\,]] = \int \mathcal{D}\boldsymbol{\zeta} \left[\int \mathrm{d}\xi\, p_{\mathrm{c}}(\xi) \prod_{i=1}^{N} p_\ell(\zeta^{(i)} - \xi|\omega) \right] \mathcal{U}_{\boldsymbol{\zeta}}[\,\cdot\,]$$

$$= \int \mathcal{D}\boldsymbol{\zeta} \left[\frac{1}{(2\pi \mathrm{V_c})^{1/2}} \frac{1}{(2\pi \mathrm{V}_\ell)^{N/2}} \int \mathrm{d}\xi\, \exp\left\{ -\frac{\xi^2}{2\mathrm{V_c}} \right\} \exp\left\{ -\sum_{i=1}^{N} \frac{(\zeta^{(i)} - \xi - \omega)^2}{2\mathrm{V}_\ell} \right\} \right] \mathcal{U}_{\boldsymbol{\zeta}}[\,\cdot\,] \tag{4.38}$$

where the vector $\boldsymbol{\zeta} = (\zeta^{(1)}, \ldots, \zeta^{(i)}, \ldots, \zeta^{(N)})$ collects the N auxiliary frequencies acting on each

particle. The corresponding unitary map is defined as:

$$\mathcal{U}_{\boldsymbol{\zeta}}[\,\cdot\,] = e^{-i\,\Delta t \sum_{i=1}^{N} \zeta^{(i)} \hat{\jmath}_z^{(i)}} \;\cdot\; e^{i\,\Delta t \sum_{i=1}^{N} \zeta^{(i)} \hat{\jmath}_z^{(i)}}. \tag{4.39}$$

By evaluating the integral in Eq. (4.38) using standard results for Gaussian integrals (see Lemma D.2 in Appendices), we obtain the final expression for the overall map:

$$\Omega[\Lambda_\omega[\,\cdot\,]] = \frac{1}{\sqrt{2\pi\,(V_c + V_\ell/N)}} \int \mathcal{D}\boldsymbol{\zeta}\, f(\boldsymbol{\zeta})\, \exp\left\{-\frac{(\bar{\zeta}-\omega)^2}{2(V_c + V_\ell/N)}\right\} \mathcal{U}_{\boldsymbol{\zeta}}[\,\cdot\,] \tag{4.40}$$

where $\bar{\zeta}$ is the average (mean) of the components of $\boldsymbol{\zeta}$,

$$\bar{\zeta} := \frac{1}{N} \sum_{i=1}^{N} \zeta^{(i)}, \tag{4.41}$$

and

$$f(\boldsymbol{\zeta}) = \sqrt{\frac{1}{N(2\pi V_\ell)^{N-1}}}\, \exp\left\{-\frac{1}{2V_\ell}\left(\sum_{i=1}^{N}(\zeta^{(i)})^2 - N\bar{\zeta}^2\right)\right\}. \tag{4.42}$$

4.2.5 Upper-bounding the Fisher information

The main goal of this chapter is to find an analytical lower bound for the aMSE of the estimator for ω at the time-step $t = k\Delta t$. To achieve this, we rely on a family of BCRBs [2, 67], and, in particular, choose the one that lower bounds the aMSE of the estimator at the last step k [67]:

$$\mathbb{E}\left[\Delta^2 \tilde{\omega}_k\right] \geq \frac{1}{F[p(\omega_k)] + \int d\omega_k\, p(\omega_k) F[p(\boldsymbol{y}_{0:k}|\omega_k)]}, \tag{4.43}$$

where $p(\omega_k)$ represents the prior knowledge about the frequency at time $t = k\Delta t$, and $p(\boldsymbol{y}_{0:k}|\omega_k)$ is the probability of observing a measurement trajectory $\boldsymbol{y}_{0:k}$ given that the frequency at time $t = kdt$ is ω_k. Here, the FI $F[\,\cdot\,]$ is defined just like in Sec. 2.2.1:

$$F[p(\boldsymbol{y}_{0:k}|\omega_k)] = \mathbb{E}_{p(\boldsymbol{y}_{0:k}|\omega_k)}\left[(\partial_{\omega_k} \log p(\boldsymbol{y}_{0:k}|\omega_k))^2\right] \tag{4.44}$$

$$= \mathbb{E}_{p(\boldsymbol{y}_{0:k}|\omega_k)}\left[-\partial_{\omega_k}^2 \log p(\boldsymbol{y}_{0:k}|\omega_k)\right]. \tag{4.45}$$

However, for our problem there is no analytical solution for $F[p(\boldsymbol{y}_{0:k}|\omega_k)]$. Therefore, to avoid the computationally demanding task of calculating $F[p(\boldsymbol{y}_{0:k}|\omega_k)]$, we opt to derive an upper bound by using another probability for which the FI can be analytically determined. To do so, we begin by rewriting the conditional probability of interest $p(\boldsymbol{y}_{0:k}|\omega_k)$ using the Bayes' rule:

$$p(\boldsymbol{y}_{0:k}|\omega_k) = \frac{p(\boldsymbol{y}_{0:k}, \omega_k)}{p(\omega_k)} = \frac{1}{p(\omega_k)}\int \mathcal{D}\boldsymbol{\omega}_{0:k-1}\, p(\boldsymbol{y}_{0:k}, \boldsymbol{\omega}_{0:k})$$
$$= \frac{1}{p(\omega_k)}\int \mathcal{D}\boldsymbol{\omega}_{0:k-1}\, p(\boldsymbol{\omega}_{0:k}) p(\boldsymbol{y}_{0:k}|\boldsymbol{\omega}_{0:k}), \tag{4.46}$$

which allows us to establish a connection between $p(\boldsymbol{y}_{0:k}|\omega_k)$, i.e. the probability of observing the vector of outcomes $\boldsymbol{y}_{0:k}$ conditioned on the last parameter ω_k, and $p(\boldsymbol{y}_{0:k}|\boldsymbol{\omega}_{0:k})$, i.e. the probability of detecting a measurement trajectory $\boldsymbol{y}_{0:k}$ given that the parameter to estimate has followed a trajectory $\boldsymbol{\omega}_{0:k}$. Then, it is possible to apply Eq. (4.13) to Eq. (4.46), which reveals a decomposition analogous to Eq. (4.13) but now for $p(\boldsymbol{y}_{0:k}|\omega_k)$:

$$p(\boldsymbol{y}_{0:k}|\omega_k) = \int \mathcal{D}\boldsymbol{\mathcal{Z}}_{0:k}\, p(\boldsymbol{y}_{0:k}|\boldsymbol{\mathcal{Z}}_{0:k}) \left[\frac{1}{p(\omega_k)}\int \mathcal{D}\boldsymbol{\omega}_{0:k-1}\, p(\boldsymbol{\omega}_{0:k}) q(\boldsymbol{\mathcal{Z}}_{0:k}|\boldsymbol{\omega}_{0:k})\right]$$
$$= \int \mathcal{D}\boldsymbol{\mathcal{Z}}_{0:k}\, p(\boldsymbol{y}_{0:k}|\boldsymbol{\mathcal{Z}}_{0:k})\, \mathbb{P}_{\omega_k}(\boldsymbol{\mathcal{Z}}_{0:k}) = \mathcal{S}_{\boldsymbol{\mathcal{Z}}_{0:k}\to\boldsymbol{y}_{0:k}}[\mathbb{P}_{\omega_k}(\boldsymbol{\mathcal{Z}}_{0:k})] \tag{4.47}$$

where we identify $\mathcal{S}_{\boldsymbol{\mathcal{Z}}_{0:k}\to\boldsymbol{y}_{0:k}}[\ \cdot\ ] = \int \mathcal{D}\boldsymbol{\mathcal{Z}}_{0:k}\, p(\boldsymbol{y}_{0:k}|\boldsymbol{\mathcal{Z}}_{0:k})\ \cdot\ $ as a stochastic map independent of the parameter ω_k, and the probability distribution $\mathbb{P}_{\omega_k}(\boldsymbol{\mathcal{Z}}_{0:k})$ as

$$\mathbb{P}_{\omega_k}(\boldsymbol{\mathcal{Z}}_{0:k}) = \frac{1}{p(\omega_k)}\int \mathcal{D}\boldsymbol{\omega}_{0:k-1}\, p(\boldsymbol{\omega}_{0:k}) q(\boldsymbol{\mathcal{Z}}_{0:k}|\boldsymbol{\omega}_{0:k}), \tag{4.48}$$

which contains the information on ω_k. As the FI is always nonincreasing under the action of any stochastic map, we can now upper-bound $F[p(\boldsymbol{y}_{0:k}|\omega_k)]$ as

$$F[p(\boldsymbol{y}_{0:k}|\omega_k)] = F[\mathcal{S}_{\boldsymbol{\mathcal{Z}}_{0:k}\to\boldsymbol{y}_{0:k}}[\mathbb{P}_{\omega_k}(\boldsymbol{\mathcal{Z}}_{0:k})]] \le F[\mathbb{P}_{\omega_k}(\boldsymbol{\mathcal{Z}}_{0:k})]. \tag{4.49}$$

Thus, the problem of lower-bounding the BCRB in Eq. (4.43) now reduces to evaluating the FI of $\mathbb{P}_{\omega_k}(\boldsymbol{\mathcal{Z}}_{0:k})$.

4.2.6 Analytical form of $\mathbb{P}_{\omega_k}(\mathcal{Z}_{0:k})$

The probability distribution in Eq. (4.48) is made up of three different probability components: the marginal probability distribution $p(\omega_k)$, the prior $p(\boldsymbol{\omega}_{0:k})$ and the CS likelihood or mixing distribution $q(\mathcal{Z}_{0:k}|\boldsymbol{\omega}_{0:k})$. To find an analytical expression for Eq. (4.48), we first need to elaborate on the exact forms of each probability component, which, in turn, depend on the stochastic process governing $\omega(t)$. Specifically, throughout this chapter, we assume that $\omega(t)$ follows an OU process.

4.2.6.1 Prior contribution

As mentioned in the beginning of this section, we wish to track the trajectory of a parameter or frequency $\boldsymbol{\omega}_{0:k} = \{\omega_0, \omega_1, \ldots, \omega_k\}$, with each element ω_k drawn from a probability distribution $p(\omega_k)$:

$$p(\omega_k) = \int \mathcal{D}\boldsymbol{\omega}_{0:k-1}\, p(\boldsymbol{\omega}_{0:k}) \tag{4.50}$$

If we choose ω_k to be the time-discretized version of the process $\omega(t)$ following the OU equation

$$\mathrm{d}\omega(t) = -\chi\omega(t)\mathrm{d}t + \mathrm{d}W_\omega \tag{4.51}$$

where $\chi > 0$ and $q_\omega > 0$ parametrize the decay and volatility of the process, and $\mathrm{d}W_\omega$ denotes the Wiener differential with mean $\mathbb{E}[\mathrm{d}W_\omega] = 0$ and variance $\mathbb{E}\left[\mathrm{d}W_\omega^2\right] = q_\omega \mathrm{d}t$, then the probability of the process transitioning from ω_{k-1} at time $(k-1)\Delta t$ to ω_k at $k\Delta t$ is given by

$$p(\omega_k|\omega_{k-1}) = \sqrt{\frac{1}{2\pi V_\mathrm{P}}} \exp\left\{-\frac{(\omega_k - \omega_{k-1}\mathrm{e}^{-\chi\Delta t})^2}{2V_\mathrm{P}}\right\} \tag{4.52}$$

with variance

$$V_\mathrm{P} = \frac{q_\omega}{2\chi}(1 - \mathrm{e}^{-2\chi\Delta t}). \tag{4.53}$$

We choose to consider the OU process in Eq. (4.51) instead of the more general process $\mathrm{d}\omega(t) = -\chi\left(\omega(t) - \bar{\omega}\right)\mathrm{d}t + \mathrm{d}W_\omega$ because the constant shift $\bar{\omega}$ preserves the aMSE $\mathbb{E}\left[\Delta^2\tilde{\nu}(t)\right] = \mathbb{E}\left[\Delta^2\tilde{\omega}(t)\right]$ [66], so that the simplified OU process is sufficient for our purposes.

Since the OU process is a Markov process (see Sec. 1.2.1), the probability of the process $\omega(t)$

of following a discrete trajectory $\boldsymbol{\omega}_{0:k} = \{\omega_0, \omega_1, \ldots, \omega_k\}$ is given by

$$p(\boldsymbol{\omega}_{0:k}) = \prod_{i=1}^{k} p(\omega_i|\omega_{i-1})p(\omega_0), \tag{4.54}$$

where we assume $p(\omega_0)$ to be a Gaussian prior with mean zero and variance σ_0^2, i.e. $p(\omega_0) = \mathcal{N}(0, \sigma_0^2)$. From that, we can compute the probability of the frequency taking the value ω_k at time $k\Delta t$, irrespective of the previous values of ω:

$$p(\omega_k) = \frac{1}{\sqrt{2\pi V_P^{(k)}}} \exp\left(-\frac{\omega_k^2}{2V_P^{(k)}}\right), \tag{4.55}$$

with variance

$$V_P^{(k)} = \sigma_0^2 e^{-2k\chi\Delta t} + \frac{q_\omega}{2\chi}(1 - e^{-2k\chi\Delta t}). \tag{4.56}$$

4.2.6.2 Classically-simulated contribution

If now we substitute the CS form of the joint map $\Omega[\Lambda_\omega[~\cdot~]]$ of Eq. (4.40) into the likelihood $p(\boldsymbol{y}_{0:k}|\boldsymbol{\omega}_{0:k})$ in Eq. (4.12), we retrieve the desired decomposition of Eq. (4.13). Namely,

$$p(\boldsymbol{y}_{0:k}|\boldsymbol{\omega}_{0:k}) = \int \mathcal{D}\boldsymbol{Z}_{0:k}\left[\prod_{j=0}^{k}f(\boldsymbol{\zeta}_j)\frac{1}{\sqrt{2\pi(V_c + V_\ell/N)}}\exp\left\{-\frac{(\bar{\zeta}_j - \omega_j)^2}{2(V_c + V_\ell/N)}\right\}\right]p(\boldsymbol{y}_{0:k}|\boldsymbol{Z}_{0:k})$$

$$= \int \mathcal{D}\boldsymbol{Z}_{0:k}\, q(\boldsymbol{Z}_{0:k}|\boldsymbol{\omega}_{0:k})\, p(\boldsymbol{y}_{0:k}|\boldsymbol{Z}_{0:k}), \tag{4.57}$$

with $p(\boldsymbol{y}_{0:k}|\boldsymbol{Z}_{0:k})$ consistent with the form given in Eq. (4.14). Thus, we identify the conditional distribution $q(\boldsymbol{Z}_{0:k}|\boldsymbol{\omega}_{0:k})$ as a product of independent likelihood terms:

$$q(\boldsymbol{Z}_{0:k}|\boldsymbol{\omega}_{0:k}) = \prod_{j=0}^{k}q(\boldsymbol{\zeta}_j|\omega_j) = \prod_{j=0}^{k}f(\boldsymbol{\zeta}_j)\frac{1}{\sqrt{2\pi(V_c + V_\ell/N)}}\exp\left\{-\frac{(\bar{\zeta}_j - \omega_j)^2}{2(V_c + V_\ell/N)}\right\}$$

$$= \prod_{j=0}^{k}f(\boldsymbol{\zeta}_j)\mathcal{Q}(\bar{\zeta}_j|\omega_j) = f(\boldsymbol{Z}_{0:k})\mathcal{Q}(\bar{\boldsymbol{\zeta}}_{0:k}|\boldsymbol{\omega}_{0:k}), \tag{4.58}$$

where we define the prefactor $f(\boldsymbol{\mathcal{Z}}_{0:k}) = \prod_{j=0}^{k} f(\boldsymbol{\zeta}_j)$ as the product of normalization constants (which do not depend on $\boldsymbol{\omega}_{0:k}$), and

$$\mathcal{Q}(\bar{\boldsymbol{\zeta}}_{0:k}|\boldsymbol{\omega}_{0:k}) = \prod_{j=0}^{k} \mathcal{Q}(\bar{\zeta}_j|\omega_j) = \prod_{j=0}^{k} \frac{1}{\sqrt{2\pi(V_c + V_\ell/N)}} \exp\left\{ -\frac{(\bar{\zeta}_j - \omega_j)^2}{2(V_c + V_\ell/N)} \right\} \quad (4.59)$$

is a product of $k + 1$ Gaussians $\mathcal{Q}(\bar{\zeta}_j|\omega_j)$ each centered around ω_j with a common variance

$$V_Q = V_c + \frac{V_\ell}{N} = \frac{\kappa_c}{\Delta t} + \frac{2\kappa_\ell}{N\Delta t}, \quad (4.60)$$

which uses the definitions of the collective and local variances in Eq. (4.23) and Eq. (4.31), respectively.

4.2.6.3 Integrated form of $\mathbb{P}_{\omega_k}(\boldsymbol{\mathcal{Z}}_{0:k})$

Once each contribution to $\mathbb{P}_{\omega_k}(\boldsymbol{\mathcal{Z}}_{0:k})$ has been established, we can bring them all together and rearrange its integral form in Eq. (4.48) as a set of nested integrals:

$$\mathbb{P}_{\omega_k}(\boldsymbol{\mathcal{Z}}_{0:k}) = \frac{1}{p(\omega_k)} \int \mathcal{D}\boldsymbol{\omega}_{0:k-1} \, p(\boldsymbol{\omega}_{0:k}) q(\boldsymbol{\mathcal{Z}}_{0:k}|\boldsymbol{\omega}_{0:k}) = \frac{f(\boldsymbol{\mathcal{Z}}_{0:k})}{p(\omega_k)} \int \mathcal{D}\boldsymbol{\omega}_{0:k-1} \, p(\boldsymbol{\omega}_{0:k}) \mathcal{Q}(\bar{\boldsymbol{\zeta}}_{0:k}|\boldsymbol{\omega}_{0:k})$$

$$= \frac{f(\boldsymbol{\mathcal{Z}}_{0:k})}{p(\omega_k)} \int \mathcal{D}\boldsymbol{\omega}_{0:k-1} \prod_{j=1}^{k} p(\omega_j|\omega_{j-1}) \mathcal{Q}(\bar{\zeta}_j|\omega_j) \, p(\omega_0) \mathcal{Q}(\bar{\zeta}_0|\omega_0)$$

$$= \frac{f(\boldsymbol{\mathcal{Z}}_{0:k})}{p(\omega_k)} \mathcal{Q}(\bar{\zeta}_k|\omega_k) \int d\omega_{k-1} \, p(\omega_k|\omega_{k-1}) \mathcal{Q}(\bar{\zeta}_{k-1}|\omega_{k-1}) \ldots \int d\omega_0 \, p(\omega_1|\omega_0) \mathcal{Q}(\bar{\zeta}_0|\omega_0) p(\omega_0). \quad (4.61)$$

Since all functions within the integrals are Gaussian, this set of nested integrals admits a closed-form recursive solution, as established in Lemma D.3 in Appendices. In particular, by identifying the variances V_P and V_Q in Eq. (D.29) with those defined in Eq. (4.53) and Eq. (4.60), respectively, we can directly state that:

$$\mathbb{P}_{\omega_k}(\boldsymbol{\mathcal{Z}}_{0:k}) = \frac{f(\boldsymbol{\mathcal{Z}}_{0:k})}{p(\omega_k)} \mathcal{Q}(\bar{\zeta}_k|\omega_k) \mathcal{P}_k(\omega_k), \quad (4.62)$$

where $\mathcal{P}_k(\omega_k)$ is a Gaussian distribution of the form

$$\mathcal{P}_k(\omega_k) = C_k \exp\left\{-\left(\frac{(\omega_k - \mu_k)^2}{2\mathrm{V}_k^{\mathrm{CS}}}\right)\right\} \tag{4.63}$$

with a variance that evolves recursively according to

$$\mathrm{V}_k^{\mathrm{CS}} = \mathrm{V}_{\mathrm{P}} + \frac{\mathrm{V}_{\mathrm{Q}} \mathrm{V}_{k-1}^{\mathrm{CS}}}{\mathrm{V}_{\mathrm{Q}} + \mathrm{V}_{k-1}^{\mathrm{CS}}} \tag{4.64}$$

starting from an initial value $\mathrm{V}_0 = \sigma_0^2$.

4.2.7 Fisher information of $\mathbb{P}_{\omega_k}(\mathbf{Z}_{0:k})$

Now that we have a closed-form expression for $\mathbb{P}_{\omega_k}(\mathbf{Z}_{0:k})$, as given in Eq. (4.62), we can move to computing its FI using the definition provided in Eq. (4.45). Namely,

$$\mathrm{F}[\mathbb{P}_{\omega_k}(\mathbf{Z}_{0:k})] = \int \mathcal{D}\mathbf{Z}_{0:k}\, \mathbb{P}_{\omega_k}(\mathbf{Z}_{0:k})\left[-\partial_{\omega_k}^2 \log\left(\frac{f(\mathbf{Z}_{0:k})}{p(\omega_k)}\mathcal{Q}(\bar{\zeta}_k|\omega_k)\mathcal{P}_k(\omega_k)\right)\right]$$

$$= \int \mathcal{D}\mathbf{Z}_{0:k}\, \mathbb{P}_{\omega_k}(\mathbf{Z}_{0:k})\left[-\partial_{\omega_k}^2 \log f(\mathbf{Z}_{0:k})\right] \tag{4.65}$$

$$- \int \mathcal{D}\mathbf{Z}_{0:k}\, \mathbb{P}_{\omega_k}(\mathbf{Z}_{0:k})\left[-\partial_{\omega_k}^2 \log p(\omega_k)\right] \tag{4.66}$$

$$+ \int \mathcal{D}\mathbf{Z}_{0:k}\, \mathbb{P}_{\omega_k}(\mathbf{Z}_{0:k})\left[-\partial_{\omega_k}^2 \log \mathcal{Q}(\bar{\zeta}_k|\omega_k)\right] \tag{4.67}$$

$$+ \int \mathcal{D}\mathbf{Z}_{0:k}\, \mathbb{P}_{\omega_k}(\mathbf{Z}_{0:k})\left[-\partial_{\omega_k}^2 \log \mathcal{P}_k(\omega_k)\right] \tag{4.68}$$

$$= -\frac{1}{\mathrm{V}_{\mathrm{P}}^{(k)}} + \frac{1}{\mathrm{V}_{\mathrm{Q}}} + \frac{1}{\mathrm{V}_k^{\mathrm{CS}}}, \tag{4.69}$$

where in Eq. (4.65) we have used that $f(\mathbf{Z}_{0:k})$ is independent of ω_k, and therefore, does not contribute to the FI. Meanwhile, the non-zero contributions arising from Eqs. (4.66-4.68) follow from the identity:

$$\mathrm{F}[\mathcal{N}(\mu, \mathrm{V})] = -\partial_{\omega_k}^2 \log \exp\left\{-\frac{(\omega_k - \mu)^2}{2\mathrm{V}}\right\} = \mathrm{V}^{-1}, \tag{4.70}$$

where we have used the definition of a Gaussian distribution and the FI of Eq. (4.45). While we already have a closed-form expression for both $\mathrm{V}_{\mathrm{P}}^{(k)}$ and V_{Q}, obtaining $\mathrm{V}_k^{\mathrm{CS}}$ requires solving the

recursive relation in Eq. (4.64). Fortunately, despite being somewhat lengthy, an explicit expression for V_k^{CS} exists:

$$V_k^{CS} = \frac{W_+ V_+^k + W_- V_-^k}{U_- V_-^k + U_+ V_+^k},$$

(4.71)

with terms W_+, V_+, U_+, W_-, V_-, and U_- given by

$$W_+ = 2V_P V_Q + \sigma_0^2 V_P + \sigma_0^2 \sqrt{V_P(4V_Q + V_P)}$$

(4.72)

$$W_- = -2V_P V_Q - \sigma_0^2 V_P + \sigma_0^2 \sqrt{V_P(4V_Q + V_P)}$$

(4.73)

$$U_+ = -V_P + 2\sigma_0^2 + \sqrt{V_P(4V_Q + V_P)}$$

(4.74)

$$U_- = V_P - 2\sigma_0^2 + \sqrt{V_P(4V_Q + V_P)}$$

(4.75)

$$V_+ = 2V_Q + V_P + \sqrt{V_P(4V_Q + V_P)}$$

(4.76)

$$V_- = 2V_Q + V_P - \sqrt{V_P(4V_Q + V_P)}$$

(4.77)

where V_Q is the variance given in Eq. (4.60) and V_P is the variance of the process when transitioning from ω_{k-1} to ω_k, i.e. Eq. (4.53).

4.2.8 THE CONTINUOUS-TIME LIMIT

If now we take the continuous-time limit of $\Delta t \to 0$, the term $1/V_Q$ in Eq. (4.69) goes to zero since V_Q is inversely proportional to Δt. The other terms become,

$$V_P^{(t)} = \lim_{\Delta t \to 0} V_P^{(k)} = \sigma_0^2 e^{-2\chi t} + \frac{q_\omega}{2\chi}(1 - e^{-2\chi t})$$

(4.78)

and

$$V^{CS}(t) = \lim_{\Delta t \to 0} V_k^{CS} = \frac{\sqrt{q_\omega \kappa_Q(N)}\, \sigma_0^2 \cosh\left(t\sqrt{\frac{q_\omega}{\kappa_Q(N)}}\right) + q_\omega \kappa_Q(N) \sinh\left(t\sqrt{\frac{q_\omega}{\kappa_Q(N)}}\right)}{\sqrt{q_\omega \kappa_Q(N)} \cosh\left(t\sqrt{\frac{q_\omega}{\kappa_Q(N)}}\right) + \sigma_0^2 \sinh\left(t\sqrt{\frac{q_\omega}{\kappa_Q(N)}}\right)}$$

(4.79)

where $\kappa_Q(N) = \kappa_c + 2\kappa_\ell/N$. Therefore, the BCRB of Eq. (4.43) in the continuous-time limit can be now bounded as follows:

$$
\begin{aligned}
\mathbb{E}\big[\Delta^2\tilde{\omega}(t)\big] &\geq \frac{1}{F[p(\omega(t))] + \int d\omega\, p(\omega(t)) F[p(\boldsymbol{y}_t|\omega(t))]} \\
&\geq \frac{1}{F[p(\omega(t))] + \int d\omega\, p(\omega(t)) F[\mathbb{P}_\omega(\boldsymbol{\mathcal{Z}}_t)]} \\
&= \frac{1}{\dfrac{1}{V_P^{(t)}} - \dfrac{1}{V_P^{(t)}} + \dfrac{1}{V^{CS}(t)}} = V^{CS}(t)
\end{aligned}
\tag{4.80}
$$

where $\boldsymbol{y}_t = \{y(\tau) : 0 \leq \tau \leq t\}$ is the realization of the measurement process, and we have used Eq. (4.70) to write $F[p(\omega(t))] = 1/V_P^{(t)}$. Hence, in its most general form, the aMSE in the continuous limit of $\Delta t \to 0$ is lower bounded by

$$
\mathbb{E}\big[\Delta^2\tilde{\omega}(t)\big] \geq V_{\sigma_0}^{CS}(t) = \frac{\sqrt{q_\omega\kappa_Q(N)}\,\sigma_0^2\cosh\left(t\sqrt{\frac{q_\omega}{\kappa_Q(N)}}\right) + q_\omega\kappa_Q(N)\sinh\left(t\sqrt{\frac{q_\omega}{\kappa_Q(N)}}\right)}{\sqrt{q_\omega\kappa_Q(N)}\cosh\left(t\sqrt{\frac{q_\omega}{\kappa_Q(N)}}\right) + \sigma_0^2\sinh\left(t\sqrt{\frac{q_\omega}{\kappa_Q(N)}}\right)},
\tag{4.81}
$$

which in the limit of $t \to \infty$, i.e. the steady-state, simplifies to:

$$
\mathbb{E}\big[\Delta^2\tilde{\omega}(t)\big] \geq V_{\sigma_0}^{CS}(t \to \infty) = \sqrt{q_\omega\kappa_Q(N)} = \sqrt{q_\omega\left(\kappa_c + \frac{2\kappa_\ell}{N}\right)},
\tag{4.82}
$$

since $\coth x = 1$ when $x \to \infty$. The general form of Eq. (4.81) can be simplified when considering different regimes: (1) an infinitely wide prior, (2) no field fluctuations and (3) a combination of both cases. In particular, when $\sigma_0 \to \infty$, then Eq. (4.81) reduces to

$$
\begin{aligned}
\mathbb{E}\big[\Delta^2\tilde{\omega}(t)\big] \geq V_\infty^{CS}(t) &= \sqrt{q_\omega\,\kappa_Q(N)}\coth\left(t\sqrt{\frac{q_\omega}{\kappa_Q(N)}}\right) \\
&= \sqrt{q_\omega\left(\kappa_c + \frac{2\kappa_\ell}{N}\right)}\coth\left(t\sqrt{q_\omega\left(\kappa_c + \frac{2\kappa_\ell}{N}\right)^{-1}}\right).
\end{aligned}
\tag{4.83}
$$
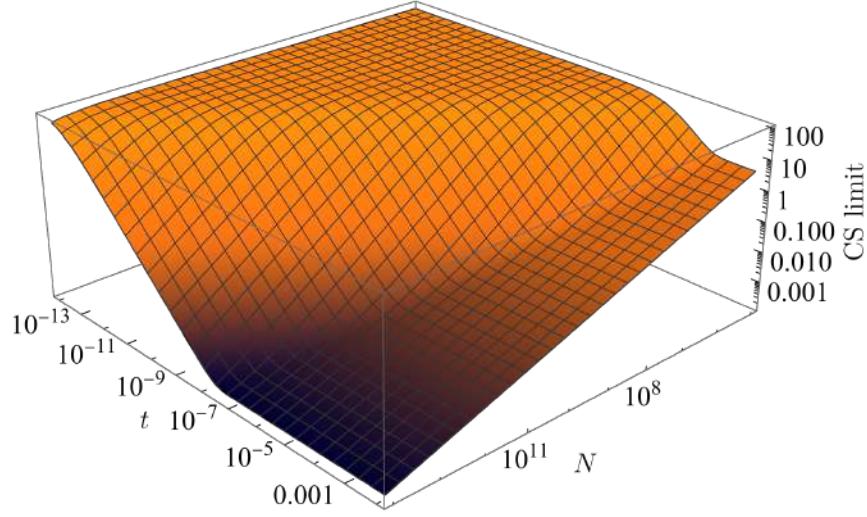
Figure 4.3: 3D Plot of the CS limit with finite σ_0 w.r.t. N and t. Log-log plot of the CS limit in Eq. (4.81) as a function of N and t, showcasing the behavior of the function over several orders of magnitude. The color gradient indicates the magnitude of the function, transitioning from high (bright) values to low (dark) values. The parameters used to generate this figure: $\sigma_0 = 10\,\mathrm{rads}^{-1}$, $q_\omega = 10^4\mathrm{rad}^2\mathrm{s}^{-3}$, $\kappa_c = 0$, $\kappa_\ell = 100\,\mathrm{Hz}$.

If instead we take the limit of $q_\omega \to 0$ of Eq. (4.81), it becomes

$$\mathbb{E}\left[\Delta^2\tilde{\omega}(t)\right] \geq \mathrm{V}_{\sigma_0}^{\mathrm{CS}}(t, q_\omega \to 0) = \frac{1}{\dfrac{1}{\sigma_0^2} + \dfrac{t}{\kappa_Q(N)}} = \frac{1}{\dfrac{1}{\sigma_0^2} + \dfrac{t}{\kappa_c + \dfrac{2\kappa_\ell}{N}}}, \qquad (4.84)$$

which exhibits the standard quantum limit (SQL) when considering an infinitely wide prior:

$$\mathbb{E}\left[\Delta^2\tilde{\omega}(t)\right] \geq \mathrm{V}_{\infty}^{\mathrm{CS}}(t, q_\omega \to 0) = \frac{\kappa_Q(N)}{t} = \frac{\kappa_c}{t} + \frac{2\kappa_\ell}{Nt}. \qquad (4.85)$$

As briefly hinted at in the introduction, we refer to this lower bound on the BCRB, $\mathrm{V}_{\sigma_0}^{\mathrm{CS}}(t, N)$, as either the CS limit or the quantum limit. After going through its derivation, it is hopefully clear that the term "classically-simulated" is used because the bound is derived by expressing the noisy quantum evolution as a convex combination of unitary channels (recall Eq. (4.22) or Eq. (4.33)), which can be efficiently simulated using classical methods. At the same time, we also call it the "quantum limit" because it represents the general bound on sensitivity that cannot be surpassed by any strategy involving any quantum effects. Although this limit sets a lower bound on the
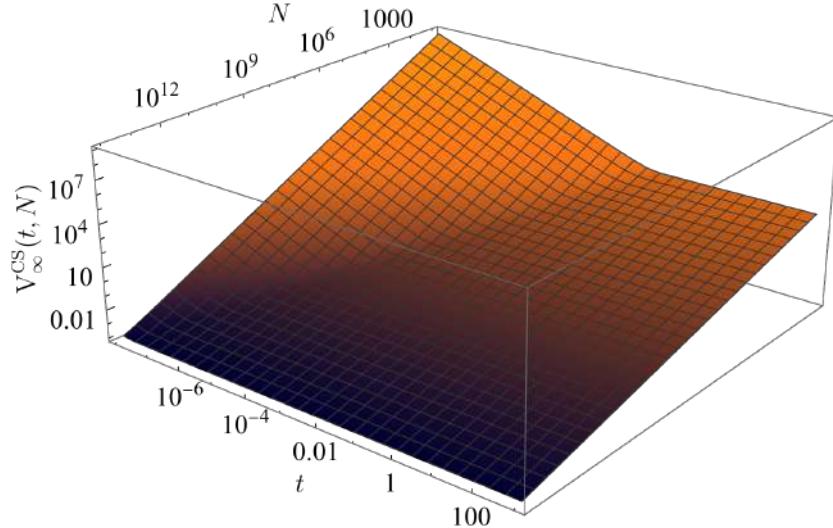
Figure 4.4: 3D Plot of the CS limit with infinitely wide prior w.r.t. N and t. Log-log plot of the CS limit in Eq. (4.83) as a function of N and t over several orders of magnitude showing its convexity over this specific range. The color gradient indicates the magnitude of the function: bright is high, and dark is low. The parameters used are: $q_\omega = 10^4 \mathrm{rad}^2\mathrm{s}^{-3}$, $\kappa_c = 0$, $\kappa_\ell = 100$ Hz.

aMSE, it is not guaranteed to be tight; that is, there is no guarantee that there exists an estimator can attain this bound. However, the CS limit still disproves the possibility of attaining super-classical scalings of N^2 and t^3 in the presence of dephasing [54, 69, 65].

Perhaps most crucially, this quantum limit serves as a powerful benchmark. Since it depends solely on the noise model and field fluctuations and is entirely *independent* on the initial state, measurement and measurement-based feedback, attaining this bound would demonstrate that the chosen sensing strategy is optimal; namely, that the state preparation, measurement, estimation and control all collectively yield the best possible precision in estimating the fluctuating field.

4.2.9 ATOM NUMBER FLUCTUATIONS AND THE CONVEXITY OF THE CS LIMIT

In practice, the exact number of atoms N in the ensemble may vary from shot to shot. We model this uncertainty by assuming $N \sim p(N) = \mathcal{N}(\bar{N}, \sigma^2)$. As a result, the bound given in Eq. (4.80) must be averaged over the distribution $p(N)$, yielding

$$\mathbb{E}_{p(N)}\left[\mathbb{E}\left[\Delta^2 \tilde{\omega}(t)\right]\right] \geq \mathbb{E}_{p(N)}\left[V_{\sigma_0}^{\mathrm{CS}}(t, N)\right], \tag{4.86}$$

where we emphasize the dependence of $V_{\sigma_0}^{\mathrm{CS}}(t, N)$ w.r.t N.

Importantly, we focus on the CS limit for an infinitely wide prior (see Eq. (4.83)) such that

the influence of the prior is effectively negligible, allowing us to isolate the effects of the noise model on the estimation process. In this case, $V_\infty^{CS}(t, N)$ is a convex function of N, which we verify by evaluating the second derivative of Eq. (4.83) w.r.t. N and checking that it is positive for all $N > 0$, $t > 0$, and q_ω, κ_ℓ, $\kappa_c > 0^\dagger$ (see also Fig. 4.4 for a more graphical confirmation). Then, by applying Jensen's inequality we obtain

$$\mathbb{E}_{p(N)}\big[V_\infty^{CS}(t, N)\big] \geq V_\infty^{CS}(t, \mathbb{E}_{p(N)}[N]) = V_\infty^{CS}(t, \bar{N}). \tag{4.87}$$

Hence, for an uninformative prior and a system with fluctuating atomic number, the aMSE of the estimate ω is lower bounded by:

$$\mathbb{E}_{p(N)}\big[\mathbb{E}\big[\Delta^2 \tilde{\omega}(t)\big]\big] \geq \sqrt{q_\omega \, \kappa_Q(\bar{N})} \, \coth\left(t\sqrt{\frac{q_\omega}{\kappa_Q(\bar{N})}}\right) \tag{4.88}$$

where $\kappa_Q(\bar{N}) = \kappa_c + 2\kappa_\ell/\bar{N}$.

---

†We omit the second derivative with respect to N in the CS limit with an infinitely wide prior due to its length, but it can be readily computed using a symbolic algebra tool such as *Mathematica*.

5

Models for Noisy Real-Time Atomic Magnetometry

TRACKING A TIME-VARYING MAGNETIC FIELD IN REAL TIME WITH HIGH SENSITIVITY is essential for applications ranging non-invasive medical diagnostics like magnetocardiography [22, 23, 24, 25, 26] or magnetoencephalography [27, 135], to navigation in GPS-denied environments [28]. Since in these situations the field varies over time (sometimes even randomly), we cannot rely on repeated runs to lower the estimation error; and instead, we must explore other strategies. Besides classical methods, such as boosting signal strength or reducing noise, quantum resources offer another way to enhance sensor performance [8, 10]. In particular, interatomic entanglement in the form of spin-squeezing can improve their sensitivity beyond the SQL [13, 14, 15, 16, 17].

While quantum enhancement has been demonstrated in repeated measurements using spin-squeezing [14, 16, 17], its application to single-shot estimation of time-varying signals [136] requires further investigation, particularly for atomic magnetometers. Specifically, having an accu-

rate model of the system dynamics that incorporates measurement backaction [34, 35, 112, 40] is essential for understanding how spin-squeezing can enhance real-time estimation. This has been achieved for mechanical oscillators using quantum Gaussian stochastic models, enabling accurate state estimation [44], even at the quantum limit [43], as well as cooling and controlling of both optomechanical resonators [45, 46, 48, 47, 137] and levitated nanoparticles [49, 50]. Notably, experiments on unpolarized atomic ensembles with Gaussian unconditional spin dynamics have also shown that continuous quantum backaction can be used to generate multipartite entanglement [63], even without a fully explicit conditional dynamical model.

In contrast, high-sensitivity spin-precession sensors used in optical magnetometry rely on polarized atomic ensembles [64], which generally require a nonlinear (and thus non-Gaussian) description. This more complex description is given by a stochastic master equation (SME), as derived in Chap. 3, which we now apply in Sec. 5.1 to the particular case of an optical atomic magnetometer, detailing both the model and its dynamics. The SME automatically provides a rigorous model of continuously monitored atomic ensembles [55, 56], incorporating both the measurement backaction [34, 35, 112, 40] and environmental decoherence due to the unavoidable noise present in these magnetometers [51, 15, 138, 57, 139, 63]. However, solving it becomes infeasible for the typical sizes of high-performance magnetometers, where $N \sim 10^6 - 10^{15}$, as explained in Sec. 5.2.1. As a result, most approaches either (1) ignore [57] or (2) evade [58, 59] the measurement backaction, (3) simulate the SME for moderately sized ensembles [62] or (4) adopt approximate linearized Gaussian models [54, 60, 61, 56, 69], as discussed in Sec. 5.2.2.1.

Instead, we address this challenge by developing a nonlinear dynamical model for a continuously measured atomic spin-1/2 ensemble that includes both measurement-based feedback and dephasing, without relying on linear approximations [65, 66] in Sec. 5.2.2.2. This nonlinear model, referred to as the co-moving Gaussian (CoG) approximation, allows us to simulate large ensembles $N \sim 10^6 - 10^{15}$. We validate this dynamical model by comparing it to the exact SME solution for moderate-sized ensembles ($N \sim 100$), demonstrating its accuracy and scalability. Besides having a reliable and scalable quantum model for the atomic magnetometer, we also address the problem of optimally estimating the field and controlling the sensor in Sec. 5.3. To this end, we propose combining an extended Kalman filter (EKF) with a linear-quadratic regulator (LQR), both introduced in Chap. 2, which are designed using the CoG nonlinear model. The optimality of the entire sensing protocol, which includes the initial state, measurement, estimate and control, is verified by deriving and attaining bounds on precision applicable to any sensing scheme involving measurement-based feedback, as established in Chap. 4.
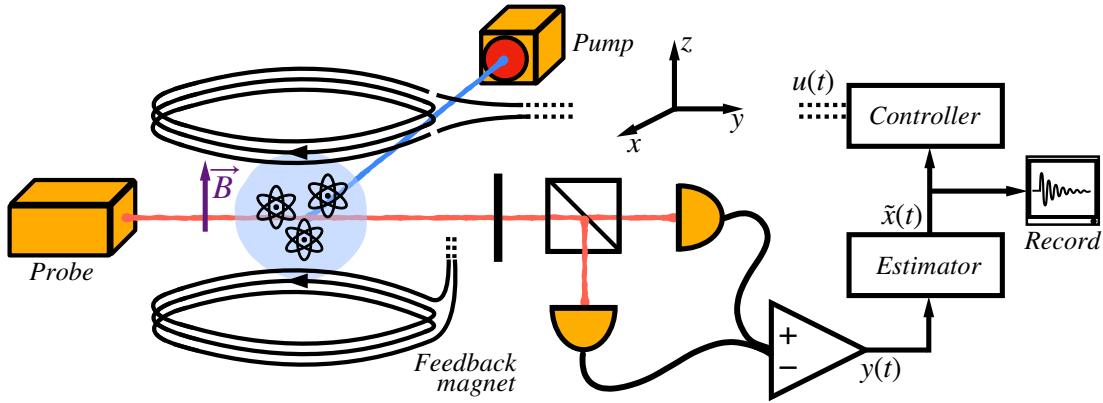
Figure 5.1: Geometry of the atomic magnetometer. (a) The *magnetometry scheme* involves an atomic ensemble optically pumped along the *x*-direction (*red arrow*) into a coherent spin-state (CSS). The magnetic field being sensed is directed along *y*, while the Faraday-rotation-based continuous measurement is performed by using the light-probe propagating along *z* (*blue arrow*), which yields a photocurrent signal $y(t)$ being recorded. (b) Bloch sphere representation of the *angular momentum of the ensemble* prepared in a CSS along *x* (with *red* and *blue arrows* indicating the pumping and probing directions, respectively).

First, we demonstrate how to reach the quantum limit in the Gaussian regime (i.e., the weak-field regime) in Sec. 5.4.1. We then focus on the more general—and practically important—case of precession-inducing fields. These field can be (1) constant (see Sec. 5.4.2.1), (2) fluctuating stochastically (see Sec. 5.4.2.2), or (3) determined by a continuously varying waveform, a magnetocardiogram (MCG) (see Sec. 5.4.2.3), which is distorted by stochastic noise that should be filtered out rather than tracked. The two last signals are tracked by an optical atomic magnetometer with realistic parameters taken from the experimental setting of Ref. [63] but ignoring spin-exchange atomic collisions [140, 141]. For both constant and fluctuating fields, we show that the magnetometer operates at the quantum limit, and that it generates conditional spin-squeezing.

Finally, in Sec. 5.4.3, we resort back to the exact SME in order to demonstrate that the proposed sensing scheme steers the atomic ensemble into a state that exhibits unconditional spin-squeezing [12], i.e. the state is entangled even when not recording the measurement data.

5.1 ATOMIC MAGNETOMETER MODEL

5.1.1 SETUP AND ITS DYNAMICS

At its core, an optical atomic magnetometer extracts information about a field from light that is scattered after interacting with a collection of N atoms (as depicted in Fig. 3.1). Importantly, these atoms should be sensitive to the time-varying and/or stochastic magnetic field $B(t)$ we aim

to sense.

Thus, to monitor the changes in the atomic state caused by the interaction with the magnetic field, the magnetometer must: (1) prepare the atoms into a suitable initial state, (2) allow their state to evolve due to its interaction with the magnetic field, and (3) simultaneously probe that state with a continuous non-demolition measurement [138]. Additionally, one might devise a control law that (4) steers the atoms towards a state optimized for magnetometry.

5.1.1.1 PREPARATION

The ensemble of N atoms is initialized by pumping it with circularly polarized light along the x-direction (see Fig. 5.1) such that only two energy levels of each atom contributes to the field and probing interactions [142, 134]. Thus, we treat the ensemble as a collection of N spin-1/2 particles. Then, the evolution of the total spin can be described through *collective* angular momentum operators:

$$\hat{J}_\alpha = \sum_{i=1}^{N} \frac{\hat{\sigma}_\alpha^{(i)}}{2}, \quad \text{with} \quad \alpha = x, y, z, \tag{5.1}$$

defining a vector of collective angular momentum operators $\hat{\boldsymbol{J}} = (\hat{J}_x, \hat{J}_y, \hat{J}_z)^{\mathrm{T}}$. Optical pumping of the atoms along the x-direction initializes a *coherent spin state* (CSS) along the same direction [143] (see also Sec. 1.4.4), so that the initial mean and variance of the ensemble angular momentum operators, denoted by the vector $\hat{\boldsymbol{J}}(t) = (\hat{J}_x(t), \hat{J}_y(t), \hat{J}_z(t))^{\mathrm{T}}$ in the Heisenberg picture, read $\langle \hat{\boldsymbol{J}}(0) \rangle_{\mathrm{CSS}} = (J, 0, 0)^{\mathrm{T}}$ and $\langle \Delta^2 \hat{\boldsymbol{J}}(0) \rangle_{\mathrm{CSS}} = (0, J/2, J/2)^{\mathrm{T}}$, respectively, where $J = N/2$. As shown in Fig. 5.2a, one may visualize the distribution of the angular momentum of a CSS with the help of the Wigner distribution, which is mapped onto the Bloch sphere (see Sec. 1.4.6).

5.1.1.2 EVOLUTION

Once pumped, the total spin of the polarized atoms starts to precess around the magnetic field axis (see Fig. 5.2a), assumed hereafter to be the z-axis, at a Larmor frequency $\omega(t) = \gamma B(t)$, with γ being the effective (constant) gyromagnetic ratio. This translates into a unitary evolution of the state of the atoms:

$$\mathrm{d}\rho(t) = -\mathrm{i}\omega(t) \left[\hat{J}_z, \rho(t) \right] \mathrm{d}t, \tag{5.2}$$

where $\rho(t)$ is the density matrix of the atoms. Thanks to this Zeeman term, it becomes clear how the atomic state can be monitored to indirectly track the magnetic field, which is the main goal of
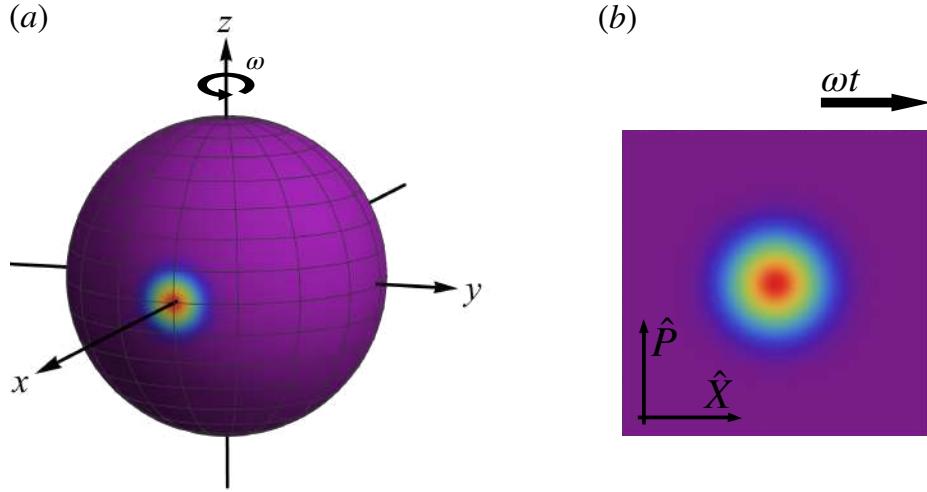
Figure 5.2: Bloch sphere representation of a CSS and phase representation of a coherent state. (a) Visual representation onto a generalized Bloch sphere of the Wigner function of a CSS of 100 spin-$1/2$. Warmer colors near the center represent higher values of the Wigner distribution, while cooler colors farther away indicate its gradual decay. To generate this figure, we have used the spherical harmonic expansion described in Sec. 1.4.6. (b) Depiction of a coherent state in the phase space.

any magnetometer. In this thesis, we consider the signal to track, i.e. $B(t)$ (or equivalently, $\omega(t)$), to follow three different profiles:

1. **Constant:** The Larmor frequency does not change in time, i.e. $\omega(t) = \omega$ s.t.,

$$\mathrm{d}\omega(t) = 0. \tag{5.3}$$

2. **Fluctuating:** The stochastic field we consider is an OUP (see Sec. 1.3.6), whose evolution is governed by the following SDE:

$$\mathrm{d}\omega(t) = -\chi(\omega(t) - \bar{\omega})\mathrm{d}t + \sqrt{q_\omega}\mathrm{d}\mathrm{W}_\omega, \tag{5.4}$$

where $\mathrm{d}\mathrm{W}_\omega$ is a Wiener differential, with mean zero and variance $\mathbb{E}\left[\mathrm{d}\mathrm{W}_\omega^2\right] = \mathrm{d}t$. The drift and volatility parameters: $\chi \geq 0$ and $q_\omega \geq 0$, are constant. The long-term mean towards which the process reverts, $\bar{\omega}$, is also positive.

3. **Time-varying:** The last magnetic field profile we aim to track is a time-varying but deterministic signal resembling a heartbeat, i.e. a MCG. This MCG-like signal can be modeled

using the dynamics of a *filtered Van der Pol* (VdP) oscillator:

$$d\nu(t) = -p\,\omega(t)\,dt \tag{5.5}$$

$$d\omega(t) = \frac{k}{m}\nu(t)\,dt + 2\frac{c}{m}(1 - \upsilon(t))\,\omega(t)\,dt \tag{5.6}$$

$$d\upsilon(t) = \frac{|\nu(t)| - \upsilon(t)}{2T}\,dt - \frac{\upsilon(t)}{T}\,dt, \tag{5.7}$$

where the frequency of interest, $\omega(t)$, is part of a larger state vector describing the signal evolution: $\boldsymbol{x}_s(t) = (\nu(t), \omega(t), \upsilon(t))$. The parameters specifying the MCG-like signal are all positive constants: $p, k, m, c, T > 0$. Additionally, we distort the time-varying signal $\omega(t)$ by adding white noise. This noise is not a feature we aim to track and therefore, should be filtered out without resorting to time-averaging [22, 144].

5.1.1.3 MEASUREMENT

In order to track the magnetic field in real time, we must continuously measure the state of the atoms. There are various continuous measurements that can be implemented, but here we consider a continuous polarimetric measurement [133, 13, 130, 125, 120, 63] (see Sec. 3.2.3). As depicted in Fig. 5.1, a probe beam traveling along y with linearly polarized light is transmitted through an atomic cloud. The interaction with the atoms rotates the polarization of the probe by angle $\Theta \propto \left\langle \hat{J}_y(t) \right\rangle_{(c)}$, due to the Faraday effect [134]. This change in polarization, and thus, the variation in $\left\langle \hat{J}_y(t) \right\rangle_{(c)}$, is later measured with a polarizing beam splitter and two photodetectors [130]. The effect of such a measurement on the state of the atoms can be described through a SME, as derived in detail in Sec. 3.2.3. Hence, the master equation in Eq. (5.2) becomes a SME of the form:

$$d\rho_{(c)}(t) = -i\omega(t)\left[\hat{J}_z, \rho_{(c)}(t)\right]dt + M\mathcal{D}[\hat{J}_y]\rho_{(c)}(t)dt + \sqrt{\eta M}\mathcal{H}[\hat{J}_y]\rho_{(c)}(t)dW, \tag{5.8}$$

with an associated measurement:

$$dy = 2\eta\sqrt{M}\left\langle \hat{J}_y(t) \right\rangle_{(c)}dt + \sqrt{\eta}dW, \tag{5.9}$$

where M is the measurement strength, $\eta \in [0,1]$ the detection efficiency, dW is the Wiener differential (see Sec. 1.2.3) and $\left\langle \hat{J}_y(t) \right\rangle_{(c)} := \text{Tr}\left\{\rho_{(c)}(t)\hat{J}_y\right\}$. Note that throughout this thesis, the detection efficiency is assumed to be perfect (i.e. $\eta = 1$), and thus, in Sec. 3.2 we have not derived

the SME for inefficient detection as given in Eq. (5.8). A detailed derivation yielding the SME of Eq. (5.8) can be found in Ref. [119]. The sub-(c) notation simply indicates that the density matrix of the atoms is now dependent, i.e. conditional, on the recorded measurement trajectory $\mathbf{y}_t = \{y(\tau)\}_{\tau \leq t}$. Namely,

$$\rho_{(c)}(t) := \rho(t|\mathbf{y}_t), \tag{5.10}$$

which is simply Eq. (3.23) but written in continuous time. The two terms in Eq. (5.8) describing the measurement back-action arise naturally when discretizing the interaction of the system with a conveyor-belt like probe (see Sec. 3.2). The first term is a dissipative one, with the superoperator $\mathcal{D}[\ \bullet\ ]\ \bullet$ given by Eq. (1.196). The second term is stochastic as well as nonlinear w.r.t. $\rho_{(c)}(t)$, where the superoperator $\mathcal{H}[\ \bullet\ ]\ \bullet$ is defined in Eq. (3.39), and is correlated with the measurement outcomes dy through the Wiener differential dW. This last term is the one responsible for the creation of conditional spin squeezing [125, 69].

5.1.1.4 MODELING NOISE

Noise in quantum systems arise from the interaction of the system with an unmonitored environment. In our work, we incorporate local and global dephasing terms into Eq. (5.8):

$$d\rho_{(c)}(t) = - i\omega(t) \left[\hat{J}_z, \rho_{(c)}(t)\right] dt + \frac{\kappa_\ell}{2} \sum_{j=1}^{N} \mathcal{D}[\hat{\sigma}_z^{(j)}]\rho_{(c)}(t)dt + \kappa_c \mathcal{D}[\hat{J}_z]\rho_{(c)}(t)dt$$

$$+ M\mathcal{D}[\hat{J}_y]\rho_{(c)}(t)dt + \sqrt{\eta M}\mathcal{H}[\hat{J}_y]\rho_{(c)}(t)dW, \tag{5.11}$$

$$dy = 2\eta\sqrt{M}\langle \hat{J}_y \rangle_{(c)}(t)dt + \sqrt{\eta}dW, \tag{5.12}$$

which model effects such as collisions, stray fields, and laser instabilities occurring along the z-direction, i.e. the direction along which the B-field is applied. Local dephasing, as its name suggests, acts independently on each individual atom at a rate κ_ℓ [145]. Similar mechanisms affecting the entire ensemble uniformly can instead be modeled with collective dephasing, at a rate κ_c [146, 147].

5.1.1.5 FEEDBACK

Finally, we also consider controlling the system by feeding back estimates constructed from the measurement record through, for instance, coils placed around the atomic ensemble (see Fig. 5.1). In particular, by applying the control law $u(t) := u(\mathbf{y}(t))$ along the direction of the field $B(t)$,

the Larmor frequency at time t will be modified to $\omega(t) \rightarrow \omega(t) + u(t)$. Importantly, because of the addition of a control field which depends on the whole measurement record $\boldsymbol{y}(t)$ up to time t, the dynamics of the atomic ensemble at each time step are now dependent on the *entire* measurement history, instead of the last measurement increment. Since the control is applied as a magnetic field along z, it comes up in the SME of Eq. (5.11) as a unitary term:

$$d\rho_{(c)}(t) = -i\left(\omega(t) + u(t)\right)\left[\hat{J}_z, \rho_{(c)}(t)\right]dt + \frac{\kappa_\ell}{2}\sum_{j=1}^{N}\mathcal{D}[\hat{\sigma}_z^{(j)}]\rho_{(c)}(t)dt + \kappa_c\mathcal{D}[\hat{J}_z]\rho_{(c)}(t)dt$$

$$+ M\mathcal{D}[\hat{J}_y]\rho_{(c)}(t)dt + \sqrt{\eta M}\mathcal{H}[\hat{J}_y]\rho_{(c)}(t)dW, \tag{5.13}$$

$$y(t)dt = 2\eta\sqrt{M}\langle\hat{J}_y(t)\rangle_{(c)}dt + \sqrt{\eta}dW. \tag{5.14}$$

This is the final form of the SME that we consider throughout this chapter. One important question that remains to be answered is how to design the control law $u(t)$. As hinted earlier and also depicted in Fig. 5.1, the control $u(t)$ will be a function of estimates constructed from the measurement record $\boldsymbol{y}(t) = \{y(\tau)\}_{0 \leq \tau \leq t}$, applied back into the system through a closed-loop structure of the form shown in Fig. 5.3. The exact form of the controller and estimator will be discussed later.
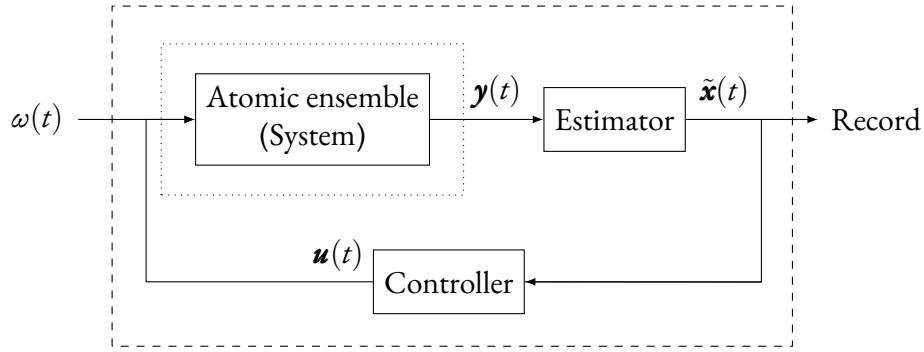

Figure 5.3: Block diagram of the optical atomic magnetometer. Structure of a closed-loop control feeding back to the atoms a control function $u(t)$ devised from estimators computed from measurement data $\boldsymbol{y}(t)$ provided by a system (in this case, an atomic ensemble). The control law at time t is constructed from the whole measurement history because the estimator at time t depends on the estimator at the previous time step and the measurement at time t. The photocurrent $y(t)$ is computed following Eq. (5.14), with the conditional mean of $\hat{J}_y$ provided either by the conditional state evolved according to Eq. (5.13) or by an approximate dynamical model containing first and second order moments. Assessing the accuracy of such an approximation in replicating the evolution of the system can be done at two different stages: comparing the estimator $\tilde{\omega}$ w.r.t. the real Larmor frequency ω (output of dashed box v.s. input), or comparing comparing the first and second moments given by the approximation v.s. the moments provided by the evolution of the conditional density matrix (output of dashed box depending on how we evolve the system).

In the absence of an experiment providing us with real-time measurements $y(t)$, the next best thing is simulating Eqs. (5.13-5.14) to generate the measurement trajectories $y(t)$. To explicitly simulate this SME, we adapt recent numerical methods [148] to incorporate estimation and feedback. However, a direct numerical simulation is feasible only for atomic ensembles of small to moderate size. This limitation motivates us to develop an effective model for the dynamics that applies to relevant experiments [13, 15, 63]. Moreover, this effective dynamical model can be then used to design the building blocks of an estimation+control scheme. Nevertheless, numerical simulations remain crucial for validating our approach and studying unconditional spin-squeezing (see Sec. 5.4.3 for more details). In particular, by benchmarking against "brute-force" numerics, we ensure that our model can be extrapolated to larger ensemble sizes beyond the reach of direct simulation.

5.2.1 EXACT NUMERICAL SOLUTION OF THE SME

Simulating the full ensemble dynamics of a typical optically pumped magnetometer, with sizes ranging in between $N \approx 10^5 - 10^{13}$ [13, 14, 15, 16, 17, 63], is computationally prohibitive, since the dimension of the underlying Hilbert space scales exponentially with N, i.e. as 2^N for two-level systems such spin-1/2 atoms. Fortunately, the size of the density matrix can be reduced to scale only polynomially with N when the system maintains permutational invariance throughout its evolution, i.e., when any two atoms in the ensemble are indistinguishable. In particular, the complexity of a collection of spin-1/2 atoms scales as $O(N^3)$ [149, 150, 76], since the density matrix is now a block-diagonal matrix with each block corresponding to a spin-number j ranging from 0 ($\frac{1}{2}$) to $N/2$ for even (odd) N. Additionally, if the evolution is exclusively governed by collective operators, which are themselves also permutationally invariant, any state initially living within the totally symmetric subspace (with $j = N/2$), e.g. CSS, evolves within it, further reducing the complexity to $O(N)$ [149, 150, 76].

Turns out, that the SME in Eq. (5.13) preserves the permutational symmetry and, for the case of $\kappa_\ell = 0$, it is even sufficient to study the evolution of the density matrix supported by the totally symmetric subspace (with $j = N/2$). Although this symmetry greatly simplifies the simulation of the SME, it still cannot be solved for ensembles with $N \approx 10^5 - 10^{13}$ but rather for ensembles of more moderate sizes of around $N \approx 100$ [76, 62]. For these type of systems, we employ the code of Rossi *et al.* [148] to numerically integrate the SME by exploiting the symmetries described

above. In particular, it constructs Kraus operators of the weak measurement at each time-step, while also guaranteeing the positivity of the density matrix [151, 152]. We extend it to perform estimation and control tasks in order to implement the closed-open loop described in Fig. 5.3.

5.2.2 Gaussian approximations

Since the exploitation of permutational symmetry is not sufficient to simulate the system in the experimentally relevant values of N, an alternative approach is to derive from Eq. (5.13) a system of SDEs of the moments of the collective angular momentum operators $\hat{J}_x$, $\hat{J}_y$ and $\hat{J}_z$. Although the density matrix $\rho_{(c)}(t)$ encodes all the statistical information about the system, one may alternatively focus on the evolution of the moments of the system's observables [38, 153, 121]. By taking expectation values and higher-order correlations from the SME, it is possible to derive a set of SDEs for these moments (or cumulants). In general, the evolution of a lower-order moment (such as the mean value of an observable) depends on higher-order moments (such as the variance), resulting in an infinite hierarchy of coupled SDEs. To manage this complexity, one can approximate the higher-order products by neglecting correlations beyond a certain order, effectively truncating the hierarchy [154]. In certain special cases, such as when the state of the system is Gaussian, the hierarchy closes; that is, all higher-order moments can be expressed in terms of the first and second moments. Consequently, only a finite number of equations are required to capture the system's dynamics completely [54, 155].

5.2.2.1 Linear and Gaussian regime

It is evident from Eq. (5.13) that the evolution of the state is not inherently linear, meaning that the atomic state is not guaranteed to remain Gaussian over time. However, under certain approximations, the SME in Eq. (5.13) can be reduced to a closed, linear system of SDEs of the first and second moments. Specifically, by neglecting feedback and local noise effects, and assuming the magnetic field B is sufficiently small, then at short enough timescales, $t \lesssim 1/(M + \kappa_c)$, we can approximate $\langle \hat{J}_x(t) \rangle_{(c)}$ with its unconditional average value $\langle \hat{J}_x(t) \rangle = J e^{-(M+\kappa_c)t/2}$ [69], where $J = N/2$. If the $B-$field is not constant and instead follows an OU process like in Eq. (5.4), one must ensure that the process obeys the constraints $\chi t \lesssim 1$ and $q_\omega t^3 \lesssim 1$ for any time $t < (M + \kappa_c)^{-1}$ (see App. E.1 for a detailed derivation [*]).

---

[*]In it we also discuss why the system is unaffected by dephasing acting along the x or y directions
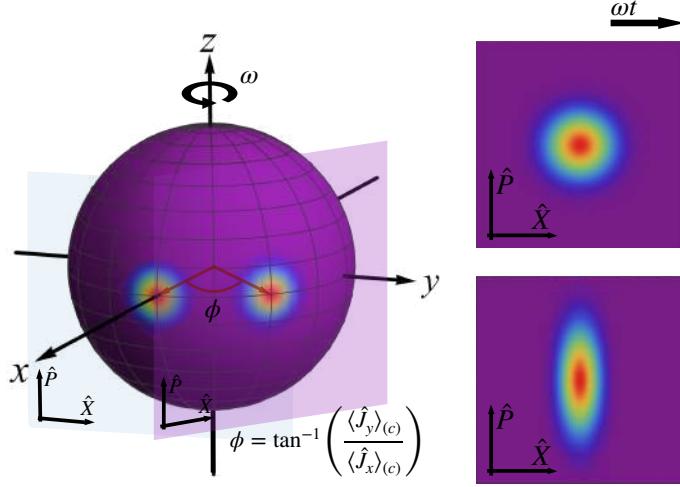
Figure 5.4: Bloch sphere representation of the atomic state: Linear and Gaussian and co-moving Gaussian approximations. The 3D-plot on the *left* shows the Wigner function of a CSS aligned along the x-axis, represented on a generalized Bloch sphere for $N = 100$ particles. When the atomic ensemble is large ($N \gg 1$), the local curvature of the sphere near the peak of the CSS can be approximated by a tangent plane, often referred to as the LG-plane or Holstein-Primakoff plane (in light blue or light pink), in which the effective phase-space quadratures $\hat{X}$ and $\hat{P}$ are defined as in Eq. (5.15). The continuous measurement of the y spin-component induces spin-squeezing of the atomic state, which in this planar approximation, corresponds to the (anti)squeezing of the $(\hat{P})\hat{X}$ quadrature (see *right* side). To preserve this Gaussian approximation over longer timescales, the LG-plane is allowed to co-rotate with the spin at the Larmor frequency ω, an approach referred to as the as the CoG approximation.

Under these approximations, the collective angular momentum vector $\hat{\boldsymbol{J}}(t)$ remains predominantly aligned with the x-axis, with only small deviations. In that case, and for sufficiently large N, the surface of the generalized Bloch sphere can thus be approximated by a two-dimensional phase-plane perpendicular to the vector $\hat{\boldsymbol{J}}(t)$ [54, 56, 69] (see Fig. 5.4 for a depiction). This plane then defines an effective phase space with position and momentum operators given by:

$$\hat{X} := \hat{J}_y / \sqrt{|\langle \hat{J}_x(t) \rangle|}, \quad \text{and} \quad \hat{P} := \hat{J}_z / \sqrt{|\langle \hat{J}_x(t) \rangle|}, \tag{5.15}$$

which satisfy the canonical commutation relation $[\hat{X}, \hat{P}] \approx i$, as long as $\hat{J}_x \approx \left| \langle \hat{J}_x(t) \rangle \right| \mathbb{I}$ for sufficiently large N [60, 61] and $(M+\kappa_c)t \lesssim 1, q_\omega t^3 \lesssim 1$, and $\chi t \lesssim 1$. With these approximations in place, the SME (5.13) reduces to a linear set of SDEs for the first and second moments of the

quadratures (5.15), as well as in the magnetic field B [60, 61]:

$$y(t)\mathrm{d}t = 2\eta\sqrt{M}\big\langle \hat{J}_y(t)\big\rangle_{(c)}\mathrm{d}t + \sqrt{\eta}\mathrm{d}W, \tag{5.16}$$

$$\mathrm{d}\omega(t) = -\chi\omega(t)\mathrm{d}t + \sqrt{q_\omega}\mathrm{d}W_\omega, \tag{5.17}$$

$$\mathrm{d}\big\langle \hat{J}_y(t)\big\rangle_{(c)} = \omega(t)J\mathrm{e}^{-(M+\kappa_c)t/2}\mathrm{d}t + 2\sqrt{\eta M}\mathrm{V}_y(t)\mathrm{d}W, \tag{5.18}$$

$$\mathrm{d}\mathrm{V}_y(t) = -4\eta M\mathrm{V}_y(t)\mathrm{d}t + \kappa_c J^2\,\mathrm{e}^{-(M+\kappa_c)t}\mathrm{d}t, \tag{5.19}$$

where we have dropped the sub-index (c) because the variance is no longer conditional on the measurement outcomes. In fact, Eq. (5.19) is completely decoupled of other state variables and can, therefore, be solved independently (see App. E.2). Therefore, Eqs. (5.16-5.18) define a continuous state-space model that describes both the evolution of the state vector and its relationship to observations (i.e., the measurement vector $\boldsymbol{y}_{\leq t}$. In this case, the state vector is constructed from the variables $\big\langle \hat{J}_y(t)\big\rangle_{(c)}$ and $\omega(t)$ as follows:

$$\boldsymbol{x}(t) = (\,\big\langle \hat{J}_y(t)\big\rangle_{(c)}, \omega(t)\,)^{\mathrm{T}}. \tag{5.20}$$

Following the framework of state-space models provided in Sec. 2.3.1, we identify the state and measurement noise vectors as $\mathrm{d}\boldsymbol{w}(t) = (\,\mathrm{d}W, \mathrm{d}W_\omega\,)^{\mathrm{T}}$ and $\mathrm{d}\boldsymbol{v}(t) = \sqrt{\eta}\,\mathrm{d}W$, respectively. Using these definitions, the system of SDEs given by Eqs. (5.16-5.18) can be expressed as:

$$\mathrm{d}\boldsymbol{x}(t) = \boldsymbol{F}(t)\,\boldsymbol{x}(t)\,\mathrm{d}t + \boldsymbol{G}(t)\,\mathrm{d}\boldsymbol{w}(t), \tag{5.21}$$

$$\mathrm{d}\boldsymbol{y}(t) = \boldsymbol{H}(t)\,\boldsymbol{x}(t)\,\mathrm{d}t + \mathrm{d}\boldsymbol{v}(t), \tag{5.22}$$

where the matrices are defined as:

$$\boldsymbol{F}(t) = \begin{pmatrix} 0 & J\mathrm{e}^{-rt/2} \\ 0 & -\chi \end{pmatrix}, \quad \boldsymbol{G}(t) = \begin{pmatrix} 2\sqrt{\eta M}\mathrm{V}_y(t) & 0 \\ 0 & \sqrt{q_\omega} \end{pmatrix}, \quad \boldsymbol{H}(t) = 2\sqrt{\eta M}\begin{pmatrix} 1 & 0 \end{pmatrix}, \tag{5.23}$$

with the noise correlations defined as $\mathbb{E}\big[\mathrm{d}\boldsymbol{w}(t)\mathrm{d}\boldsymbol{w}^{\mathrm{T}}(s)\big] = \boldsymbol{Q}(t)\,\delta(t-s)\,\mathrm{d}t$, $\mathbb{E}\big[\mathrm{d}\boldsymbol{v}(t)\mathrm{d}\boldsymbol{v}^{\mathrm{T}}(s)\big] = \boldsymbol{R}(t)\,\delta(t-s)\,\mathrm{d}t$ and $\mathbb{E}\big[\mathrm{d}\boldsymbol{w}(t)\mathrm{d}\boldsymbol{v}^{\mathrm{T}}(s)\big] = \boldsymbol{S}(t)\,\delta(t-s)\,\mathrm{d}t$. Based on the noise vector definitions, the covariance matrix for the system noise is $\boldsymbol{Q}(t) = \mathbb{I}$, the measurement noise has a scalar variance $\boldsymbol{R}(t) = \eta$, and the cross-correlation matrix between the system and measurement is $\boldsymbol{S}(t) = (\,\sqrt{\eta}\,,\,0\,)^{\mathrm{T}}$. Notably, the field and atomic noises are uncorrelated: $\mathbb{E}[\mathrm{d}W_\omega\mathrm{d}W] = \mathbb{E}[\mathrm{d}W\mathrm{d}W_\omega] = 0$. Furthermore, since $\mathbb{E}\big[\mathrm{d}W^2\big] = \mathbb{E}\big[\mathrm{d}W_\omega^2\big] = \mathrm{d}t \geq 0$, the covariance matrix

$Q(t)$ satisfies semi-positivity (i.e., $Q \geq 0$). Finally, observe that both $d\boldsymbol{w}(t)$ and $d\boldsymbol{v}(t)$ are partially-correlated Gaussian noises and $\boldsymbol{F}(t)\boldsymbol{x}(t)$ and $\boldsymbol{H}(t)\boldsymbol{x}(t)$ are both linear functions w.r.t. the state vector. This is consistent with Eqs. (5.16-5.18), which involve Gaussian stochastic terms and are linear with respect to one another. Therefore, a state-space model of the form in Eqs. (5.21-5.22) (or Eqs. (5.16-5.19)) is referred to as *linear and Gaussian* (LG) [69].

ANALYTICAL SOLUTION OF THE SPIN-SQUEEZING PARAMETER IN THE LG REGIME   The dynamics of $V_y(t)$ can be found analytically by solving Eq. (5.19), since it is a differential equation fully decoupled from the rest. Due to its complexity, the complete analytical solution is presented in detail in App. E.2 [69]. However, the expression for the variance of $\hat{J}_y$ greatly simplifies when dividing it into a short-time and long-time regime:

$$V_y(t) = \begin{cases} V_{<t^*}(t) = \dfrac{J}{2}\dfrac{1+2Jt\kappa_c}{1+2JtM\eta}e^{-(M+\kappa_c)t/2}, & \text{if} \quad 0 \leq t \ll t^* \qquad (5.24a) \\[3mm] V_{>t^*}(t) = \dfrac{J}{2}\sqrt{\dfrac{\kappa_c}{\eta M}}\,e^{-(M+\kappa_c)t/2}, & \text{if} \quad t \gg t^* \qquad (5.24b) \end{cases}$$

where initially $V_y(0) = J/2$, and $t^* = (2J\sqrt{M\kappa_c\eta})^{-1}$ is the transition time between the two regimes. Importantly, note that $t \ll t^*$ implies $2Jt\kappa_c \ll \sqrt{\kappa_c/(\eta M)}$. Then, if also $\kappa_c < \eta M$, we may infer $2Jt\kappa_c \ll 1$ and approximate $1 + 2Jt\kappa_c \approx 1$ in (5.24a). As a result, we then recover the known noiseless ($\kappa_c = 0$) solution for the variance within the short-time regime [54]:

$$V_y(t) = \frac{J}{2}\frac{1+2Jt\kappa_c}{1+2JtM\eta}e^{-(M+\kappa_c)t/2} \approx \frac{J}{2+4JtM\eta}, \qquad (5.25)$$

despite collective dephasing being present, i.e. $\kappa_c > 0$. In fact, we prove also in App. E.2 that $V_y(t)$ is a non-decreasing function at $t \approx 0$ if $\kappa_c \geq \eta M$. Hence, the noise may be considered negligible at small times *only if* $\kappa_c < \eta M$.

Finding an analytical solution for the variance $V_y(t)$ is crucial because of its close relationship with the spin-squeezing parameter of the atomic ensemble [143] introduced by Wineland *et al.* [6]. In the experimental setup of Fig. 5.1, a continuous measurement performed along y squeezes the variance of $\hat{J}_y$, $V_y(t)$, in detriment of the variance of $\hat{J}_z$. Given that we want to use this state to sense small variations in ω, it follows that the state should have maximal polarization along x and minimal variance along y (or maximal squeezing). How closely such a state resembles a CSS pointing along x is therefore quantified by the spin-squeezing parameter ξ_y^2 [83, 6, 12] (see
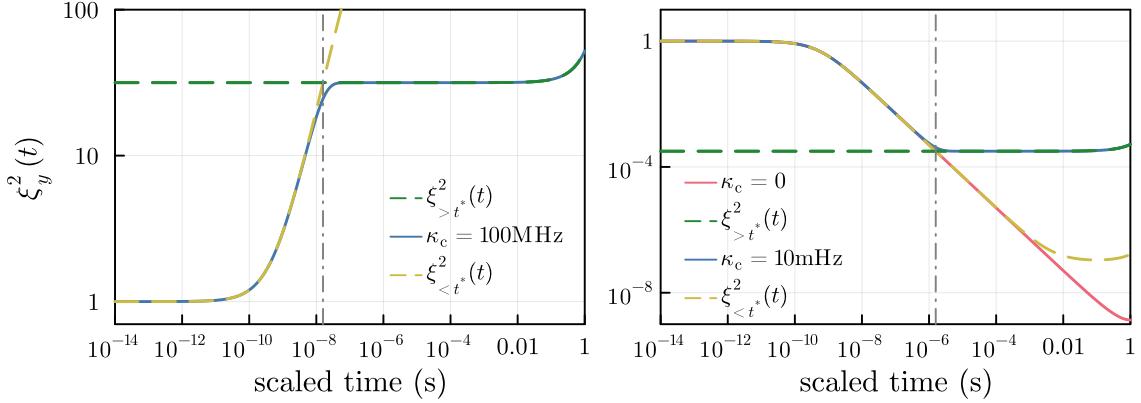
Figure 5.5: Strong and weak dephasing regimes for the squeezing parameter. Both figures plot the *squeezing parameter* $\xi_y^2(t)$ as defined in equation (5.26) as a function of scaled time $t_s = (M + \kappa_c)t$, for the case of $\kappa_c = 10\,\mathrm{mHz} < M = 100\,\mathrm{kHz}$ *(right)* and $\kappa_c = 100\,\mathrm{MHz} > M = 100\,\mathrm{kHz}$ *(left)*. The exact function $\xi_y^2(t)$ (solid, blue) is compared with its two different regimes $\xi_{<t^*}^2(t)$ and $\xi_{>t^*}^2(t)$ (dashed, yellow and green, respectively), as well as the noiseless solution when $\kappa_c < M$ (dashed, red). The gray dash-dot line indicates the inflection point t^*, where the transition between the two regimes occurs. The other parameters used to generate the plots are $J = 10^9$, $\gamma = 1\,\mathrm{kHz/mG}$, and $\eta = 1$. The time has been rescaled to $t_S = (M + \kappa_c)t$ so that its range is therefore limited to $t_S \in [0, 1]$, where the LG approximation holds.

Sec. 1.4.7):

$$\xi_y^2(t) := \frac{V_y(t)}{\langle \hat{J}_x(t) \rangle^2} \left(\frac{V_y^{\mathrm{CSS}}}{\langle \hat{J}_x \rangle_{\mathrm{CSS}}^2} \right)^{-1} = \frac{N\,V_y(t)}{\langle \hat{J}_x(t) \rangle^2}, \qquad (5.26)$$

since $\xi_y^2(t) < 1$ indicates a gain in squeezing relative to the CSS [6], whereas $\xi_y^2(t) \geq 1$ implies the absence of spin squeezing and, consequently, the inability to certify multi-particle entanglement [12].

Next, given that $V_y(t)$ has two distinct regimes, as presented in Eq. (5.24), the spin-squeezing parameter can be similarly split as

$$\xi_y^2(t) \approx \frac{2V_y(t)}{J e^{-(M+\kappa_c)t}} = \begin{cases} \xi_{<t^*}^2(t) = \dfrac{1 + 2Jt\kappa_c}{1 + 2JtM\eta} e^{(M+\kappa_c)t/2}, & \text{if} \quad 0 \leq t \ll t^*, \qquad (5.27a) \\[4mm] \xi_{>t^*}^2(t) = \sqrt{\dfrac{\kappa_c}{\eta M}} e^{(M+\kappa_c)t/2}, & \text{if} \quad t \gg t^*. \qquad (5.27b) \end{cases}$$

In Fig. 5.5 we present explicitly the exact dynamics of the squeezing parameter (5.26) in the LG regime for the two important cases: *(right)*, when $\kappa_c < \eta M$ and the spin-squeezing ($\xi_y^2(t) < 1$) occurs within a finite-time window (see Fig. 5.5 *(right)*); and *(left)*, when $\kappa_c \geq \eta M$ for which spin-squeezing is forbidden (i.e. $\xi_y^2(t) \geq 1$, as shown in Fig. 5.5 *(left)*). In both cases, it is evident that the exact solution for $\xi_y^2(t)$ very closely follows the two-regime behavior in (5.27), with a

clear transition at $t \approx t^*$. Moreover, as seen explicitly from the two-regime solution (5.27) (and the exact solution that can be straightforwardly derived based on the explicit variance evolution of App. E.2), the dynamics of the squeezing parameter are specified solely by the properties of the continuous measurement (η and M), collective decoherence (κ_c), and the number of atoms ($J = N/2$).

CO-MOVING GAUSSIAN APPROXIMATION

In practical magnetometers [13, 14, 15, 16, 17, 63], the atomic spin must precess multiple times during detection to collect a sufficient signal. Since the LG approximation assumes that the angular momentum vector does not precess, it is no longer applicable. To then maintain an approximately Gaussian description of the system at all times, we allow the LG-plane (see Fig. 5.4) to precess with the mean angular-momentum vector $\langle \hat{\boldsymbol{J}}(t) \rangle$ at the frequency ω [156]. This approach is referred to as the *co-moving Gaussian* (CoG) approximation [65], and it holds under the following conditions: (1) the ensemble size is large, i.e., $N \gg 1$; and (2) the squeezing arising from the continuous measurement is not strong enough to wrap the Wigner function around the Bloch sphere.

Specifically, by analyzing the conditional evolution in the Heisenberg picture for the mean angular momenta $\langle \hat{J}_\alpha(t) \rangle_{(c)}$, $\alpha = x, y, z$, and their corresponding covariance matrix $C^{(c)}_{\alpha\beta}(t) := \frac{1}{2} \left(\langle \{\hat{J}_\alpha(t), \hat{J}_\beta(t)\} \rangle_{(c)} - 2 \langle \hat{J}_\alpha(t) \rangle_{(c)} \langle \hat{J}_\beta(t) \rangle_{(c)} \right)$ with diagonal elements $V^{(c)}_\alpha(t) := C^{(c)}_{\alpha\alpha}(t)$ ($\alpha, \beta = x, y, z$), we derive a set of coupled SDEs based on the SME (5.13) (details in App. E.3). For sim-

plicity, the explicit time dependence of all the quantities is omitted below:

$$d\langle \hat{J}_x \rangle_{(c)} = -(\omega(t)+u(t))\langle \hat{J}_y \rangle_{(c)}dt - \frac{1}{2}(\kappa_c+2\kappa_\ell+M)\langle \hat{J}_x \rangle_{(c)}dt + 2\sqrt{\eta M}C_{xy}^{(c)}\,dW \qquad (5.28a)$$

$$d\langle \hat{J}_y \rangle_{(c)} = (\omega(t)+u(t))\langle \hat{J}_x \rangle_{(c)}dt - \frac{1}{2}(\kappa_c+2\kappa_\ell)\langle \hat{J}_y \rangle_{(c)}dt + 2\sqrt{\eta M}V_y^{(c)}\,dW \qquad (5.28b)$$

$$dV_x^{(c)} = -2(\omega(t)+u(t))C_{xy}^{(c)}dt + \kappa_c\left(V_y^{(c)} + \langle \hat{J}_y \rangle_{(c)}^2 - V_x^{(c)}\right)dt + \kappa_\ell\left(\frac{N}{2}-2V_x^{(c)}\right)dt$$

$$+ M\left(V_z^{(c)} - V_x^{(c)} - 4\eta C_{xy}^{(c)2}\right)dt \qquad (5.28c)$$

$$dV_y^{(c)} = 2(\omega(t)+u(t))C_{xy}^{(c)}dt + \kappa_c\left(V_x^{(c)} + \langle \hat{J}_x \rangle_{(c)}^2 - V_y^{(c)}\right)dt + \kappa_\ell\left(\frac{N}{2}-2V_y^{(c)}\right)dt$$

$$- 4\eta MV_y^{(c)2}dt \qquad (5.28d)$$

$$dV_z^{(c)} = M\left(V_x^{(c)} + \langle \hat{J}_x \rangle_{(c)}^2 - V_z^{(c)}\right)dt \qquad (5.28e)$$

$$dC_{xy}^{(c)} = (\omega(t)+u(t))\left(V_x^{(c)} - V_y^{(c)}\right)dt - \kappa_c\left(2C_{xy}^{(c)} + \langle \hat{J}_x \rangle_{(c)}\langle \hat{J}_y \rangle_{(c)}\right)dt - 2\kappa_\ell C_{xy}^{(c)}dt$$

$$- \frac{1}{2}MC_{xy}^{(c)}\left(1 + 8\eta V_y^{(c)}\right)dt \qquad (5.28f)$$

$$d\omega = -\chi\omega(t)\,dt + \sqrt{q_\omega}\,dW_\omega, \qquad (5.28g)$$

where we have ignored all the (stochastic) contributions that involve the third-order moments (as discussed in App. E.3).

This system of SDE gives us a state-space model, a continuous-time version of the discrete-time state-space model introduced in Sec. 2.3.1. If we write Eq. (5.28) in Langevin form instead of Itô, then:

$$\dot{\boldsymbol{x}}(t) = \boldsymbol{f}(\boldsymbol{x}(t), u(t), \boldsymbol{q}(t), t), \qquad (5.29)$$

$$y(t) = h(\boldsymbol{x}(t), r(t), t), \qquad (5.30)$$

where $\boldsymbol{x}(t) = (\langle \hat{J}_x \rangle_{(c)}, \langle \hat{J}_y \rangle_{(c)}, V_x^{(c)}, V_y^{(c)}, V_z^{(c)}, C_{xy}^{(c)}, \omega)^{\mathrm{T}}$ is the state vector, $u(t)$ is the control field, and $\boldsymbol{q}(t)$ and $r(t)$ are the state and measurement noises. The function $\boldsymbol{f}(\boldsymbol{x}(t), u(t), \boldsymbol{q}(t), t)$ is

nonlinear w.r.t. $\boldsymbol{x}(t)$ and given by the SDE system in Eq. (5.28):

$$\boldsymbol{f}(\boldsymbol{x}(t), u(t), \boldsymbol{q}(t), t) = \tag{5.31}$$

$$\begin{pmatrix} -(\omega(t)+u(t))\langle \hat{J}_y \rangle_{(c)} - \frac{1}{2}(\kappa_c + 2\kappa_\ell + M)\langle \hat{J}_x \rangle_{(c)} + 2\sqrt{\eta M}\mathrm{C}_{xy}^{(c)}\,\xi \\ (\omega(t)+u(t))\langle \hat{J}_x \rangle_{(c)} - \frac{1}{2}(\kappa_c + 2\kappa_\ell)\langle \hat{J}_y \rangle_{(c)} + 2\sqrt{\eta M}\mathrm{V}_y^{(c)}\,\xi \\ -2(\omega(t)+u(t))\mathrm{C}_{xy}^{(c)} + \kappa_c\left(\mathrm{V}_y^{(c)} + \langle \hat{J}_y \rangle_{(c)}^2 - \mathrm{V}_x^{(c)}\right) + \kappa_\ell\left(\frac{N}{2} - 2\mathrm{V}_x^{(c)}\right) + M\left(\mathrm{V}_z^{(c)} - \mathrm{V}_x^{(c)} - 4\eta \mathrm{C}_{xy}^{(c)2}\right) \\ 2(\omega(t)+u(t))\mathrm{C}_{xy}^{(c)} + \kappa_c\left(\mathrm{V}_x^{(c)} + \langle \hat{J}_x \rangle_{(c)}^2 - \mathrm{V}_y^{(c)}\right) + \kappa_\ell\left(\frac{N}{2} - 2\mathrm{V}_y^{(c)}\right) - 4\eta M\mathrm{V}_y^{(c)2} \\ M\left(\mathrm{V}_x^{(c)} + \langle \hat{J}_x \rangle_{(c)}^2 - \mathrm{V}_z^{(c)}\right) \\ (\omega(t)+u(t))\left(\mathrm{V}_x^{(c)} - \mathrm{V}_y^{(c)}\right) - \kappa_c\left(2\mathrm{C}_{xy}^{(c)} + \langle \hat{J}_x \rangle_{(c)}\langle \hat{J}_y \rangle_{(c)}\right) - 2\kappa_\ell\mathrm{C}_{xy}^{(c)} - \frac{1}{2}M\mathrm{C}_{xy}^{(c)}\left(1 + 8\eta\mathrm{V}_y^{(c)}\right) \\ -\chi\omega(t) + \sqrt{q_\omega}\,\xi_\omega \end{pmatrix}.$$

On the other hand, $h(\boldsymbol{x}(t), r(t), t)$ is actually a linear function, since the measurement model in Eq. (5.14) is linear w.r.t. the state $\boldsymbol{x}(t)$:

$$y(t) = 2\eta\sqrt{M}\langle \hat{J}_y(t) \rangle_{(c)} + \sqrt{\eta}\xi = h(\boldsymbol{x}(t), r(t), t) = \boldsymbol{H}\boldsymbol{x}(t) + r(t) \tag{5.32}$$

where $\boldsymbol{H} = 2\eta\sqrt{M}\,(0\ 1\ 0\ 0\ 0\ 0\ 0)$ and the state and measurement noises are:

$$\boldsymbol{q}(t) = \begin{pmatrix} \xi \\ \xi_\omega \end{pmatrix}, \qquad \text{and} \qquad r(t) = \sqrt{\eta}\xi, \tag{5.33}$$

where ξ and ξ_ω are Langevin noises defined as $\xi := \mathrm{d}W/\mathrm{d}t$ and $\xi_\omega := \mathrm{d}W_\omega/\mathrm{d}t$, respectively.

In order to validate the CoG approximation, we compare it against the exact SME solution for simulatable ensemble sizes. Our results indicate that while the first and second moments of the system (e.g., $\langle \hat{J}_x \rangle_{(c)}$, $\langle \hat{J}_y \rangle_{(c)}$, and $\mathrm{V}_y^{(c)}$) show good agreement between the exact and approximate approaches, the real-time estimation of the Larmor frequency $\tilde{\omega}(t)$ using the CoG-based dynamics aligns even more closely with the one generated by the exact SME. This accurate generation of an estimate of $\omega(t)$ without having to resort to a full SME simulation, reinforces the usefulness of the proposed CoG approximation, especially in the experimentally relevant regimes of large N. A comprehensive error analysis is presented in App. E.4.

So far, we have outlined both the physical dynamics and numerical solution of the system—that is, the atomic ensemble within the optical atomic magnetometer. In particular, we discussed two complementary simulation strategies: one that exploits the permutational symmetry of the ensemble to numerically solve the SME for moderate system sizes, and another based on Gaussian approximations that captures the essential dynamics in larger, experimentally relevant regimes. However, as illustrated in Fig. 5.3, this represents only part of the complete picture. Our next goal is to explore state estimation and control methods that use the real-time measurement data to monitor and steer the system continuously, thereby enhancing the performance of the magnetometer in tracking the magnetic field.

5.3.1   Kalman filter

For an atomic magnetometer operated in the LG regime, like the one in Sec. 5.2.2.1, the optimal estimator $\tilde{\boldsymbol{x}}(t) = (\,\langle \hat{\tilde{J}}_y(t) \rangle_{(\mathrm{c})},\, \tilde{\omega}(t)\,)^{\mathrm{T}}$ is given by the *correlated* Kalman filter (KF) [29] (see also Sec. 2.6):

$$\dot{\tilde{\boldsymbol{x}}}(t) = \boldsymbol{F}(t)\tilde{\boldsymbol{x}}(t) + \boldsymbol{K}(t)\big(y(t) - \boldsymbol{H}(t)\tilde{\boldsymbol{x}}(t)\big)\,, \tag{5.34}$$

$$\boldsymbol{K}(t) = \big(\boldsymbol{\Sigma}(t)\boldsymbol{H}^{\mathrm{T}}(t) - \boldsymbol{G}(t)\boldsymbol{S}(t)\big)\,\boldsymbol{R}^{-1}(t), \tag{5.35}$$

$$\dot{\boldsymbol{\Sigma}}(t) = \boldsymbol{F}(t)\boldsymbol{\Sigma}(t) + \boldsymbol{\Sigma}(t)\boldsymbol{F}^{\mathrm{T}}(t) - \boldsymbol{K}(t)\boldsymbol{R}(t)\boldsymbol{K}^{\mathrm{T}}(t) + \boldsymbol{G}(t)\boldsymbol{Q}(t)\boldsymbol{G}^{\mathrm{T}}(t), \tag{5.36}$$

where the differential equation for the covariance matrix $\boldsymbol{\Sigma}(t) = \mathbb{E}\big[\boldsymbol{x}(t) - \tilde{\boldsymbol{x}}(t))(\boldsymbol{x}(t) - \tilde{\boldsymbol{x}}(t))^{\mathrm{T}}\big]$, the Riccati equation, yields the (minimal) estimator error of $\boldsymbol{x}(t)$, with initial conditions $\boldsymbol{\Sigma}(0) = \mathrm{Diagonal}\,[0, \sigma_0^2]$, where σ_0^2 is the variance of the prior distribution $p(\omega_0)$. The rest of the matrices, $\boldsymbol{F}(t)$, $\boldsymbol{G}(t)$, $\boldsymbol{H}(t)$, $\boldsymbol{Q}(t)$, $\boldsymbol{S}(t)$ and $\boldsymbol{R}(t)$ are provided by the model in Eq. (5.23), and the photocurrent $y(t)$ is given by Eq. (5.22).

SOLUTION IN THE ABSENCE OF DECOHERENCE AND FIELD FLUCTUATIONS   In the absence of dephasing and field fluctuations ($\kappa_{\mathrm{c}} = \kappa_{\ell} = 0$, and $q_\omega = \chi = 0$), the Ricatti equation in (5.36) can be explicitly solved [54] by redefining the covariance matrix as $\boldsymbol{\Sigma}(t) = \boldsymbol{Y}(t)\boldsymbol{X}(t)^{-1}$, and

thus decoupling the Riccati equation as:

$$\dot{\boldsymbol{X}}(t) = -\left(\boldsymbol{F}(t) - \boldsymbol{G}(t)\boldsymbol{S}(t)\boldsymbol{R}^{-1}(t)\boldsymbol{H}(t)\right)^{\mathrm{T}}\boldsymbol{X}(t) + \boldsymbol{H}^{\mathrm{T}}(t)\boldsymbol{R}^{-1}(t)\boldsymbol{H}(t)\boldsymbol{Y}(t), \tag{5.37}$$

$$\dot{\boldsymbol{Y}}(t) = \left(\boldsymbol{F}(t) - \boldsymbol{G}(t)\boldsymbol{S}(t)\boldsymbol{R}^{-1}(t)\boldsymbol{H}(t)\right)\boldsymbol{Y}(t) + \boldsymbol{G}(t)\left(\boldsymbol{Q}(t) - \boldsymbol{S}(t)\boldsymbol{R}^{-1}(t)\boldsymbol{S}^{\mathrm{T}}(t)\right)\boldsymbol{B}^{\mathrm{T}}(t)\boldsymbol{X}(t), \tag{5.38}$$

with initial conditions $\boldsymbol{X}(0) = \mathbb{I}$ and $\boldsymbol{Y}(0) = \boldsymbol{\Sigma}(0)$. Since $\kappa_{\mathrm{c}} = 0$, the solution to the variance differential equation in Eq. (5.19) is

$$\mathrm{V}_y(t) = \frac{J}{2 + 4M\eta Jt}, \tag{5.39}$$

as shown in App. E.2. Moreover, since no fluctuations of the field are being considered, the volatility matrix is now simply $\boldsymbol{G}(t) = \mathrm{Diag}[2\sqrt{\eta M}\mathrm{V}_y(t), 0]$. Under these conditions, the decoupled system of differential equations introduced in Eqs. (5.37-5.38), given initial conditions $\boldsymbol{X}(0) = \mathbb{I}$ and $\boldsymbol{Y}(0) = \mathrm{Diag}[0, \sigma_0^2]$, can be analytically solved. The solution for $\boldsymbol{\Sigma}_{\omega,\omega}$, which for a LG system matches the aMSE, $\mathbb{E}\left[\Delta^2\tilde{\omega}(t)\right]$, is:

$$\mathbb{E}\left[\Delta^2\tilde{\omega}^{\mathrm{HS}}(t, \sigma_0)\right] = \mathrm{A}\frac{(1 + 2JMt\eta)}{a(t)\,\mathrm{e}^{-Mt} + 4\,(1 + 4J\eta)\,\mathrm{e}^{-Mt/2} + b(t)}, \tag{5.40}$$

where

$$\mathrm{A} = M^2/(16\eta J^2), \tag{5.41}$$

$$a(t) = -(1 + 2\eta J(4 + Mt)), \tag{5.42}$$

$$b(t) = \frac{M^2}{16\eta J^2\sigma_0^2} + \frac{M^3 t}{8J\sigma_0^2} + (Mt - 3) + 2\eta J(Mt - 4) \tag{5.43}$$

$$= \frac{M^2}{16\eta J^2\sigma_0^2} + \frac{M^3 t}{8J\sigma_0^2} + c(t), \tag{5.44}$$

with the subscript HS highlighting the super-classical scaling of the error in both t and N (Heisenberg scaling). Note that for $\sigma_0^2 \to \infty$, the functions $b(t)$ and $c(t)$ are equal and $\mathbb{E}\left[\Delta^2\tilde{\omega}^{\mathrm{HS}}(t, \infty)\right]$ matches consistently the solution of Geremia *et al.* [54]. In order to highlight its non-classical

174

scaling with t and N, the aMSE can be approximated as

$$
\mathbb{E}\left[\Delta^2\tilde{\omega}^{\mathrm{HS}}(t,\infty)\right] \approx
\begin{cases}
\dfrac{1}{N^2}\dfrac{3}{\eta M t^3}, & \text{for } t \ll (NM)^{-1}, \\[2ex]
\dfrac{4}{N^2}\dfrac{3}{\eta M t^3}, & \text{for } (NM)^{-1} \ll t\ (< M^{-1}),
\end{cases}
\tag{5.45}
$$

when assuming no initial knowledge of the field ($\sigma_0^2 \to \infty$). The first term in (5.45) is derived by Taylor expanding to leading order in time the solution $\mathbb{E}\left[\Delta^2\tilde{\omega}^{\mathrm{HS}}(t,\infty)\right]$. The second term is obtained by expanding $\mathbb{E}\left[\Delta^2\tilde{\omega}^{\mathrm{HS}}(t,\infty)\right]$ to leading order around $(NMt)^{-1}$.

STEADY STATE SOLUTION OF THE KALMAN FILTER   The KF of Eqs. (5.34-5.36) which tracks an OU process has an analytical steady solution, both in the case of $\chi = 0$ and $\chi \neq 0$. Below we discuss the steady-state solution for the case of a pure Wiener noise (i.e. $\chi = 0$), with the general solution for $\chi \neq 0$ presented in App. E.5.

To find the steady-state solution of the Riccati differential equation one must set $\mathrm{d}\boldsymbol{\Sigma}(t) = 0$, which is greatly simplified by noting that the variance of $\hat{J}_y$ at $t \gg t^*$ is $\mathrm{V}_{>t^*}(t)$ (5.24b). Then, the steady-state solution for $\boldsymbol{\Sigma}_{\omega,\omega}$, which in the case of a KF coincides with the minimal aMSE $\mathbb{E}\left[\Delta^2\tilde{\omega}(t)\right]$, can be shown to be:

$$
\mathbb{E}\left[\Delta^2\tilde{\omega}^{\mathrm{SS}}(t)\right] = \left(q_\omega\,\kappa_{\mathrm{c}} + \frac{2}{N}\sqrt{\frac{q_\omega^3}{M\eta}}\,\mathrm{e}^{(M+\kappa_{\mathrm{c}})t/2}\right)^{1/2},
\tag{5.46}
$$

where second term, which survives in the $\kappa_{\mathrm{c}} \to 0$ limit, had been derived previously in Ref. [55]. By accounting for collective dephasing in our analysis, an additional and important term appears, which dominates for large N. Namely, for

$$
N > \frac{2}{\kappa_{\mathrm{c}}}\sqrt{\frac{q_\omega}{\eta M}}\,\mathrm{e}^{(M+\kappa_{\mathrm{c}})t/2},
\tag{5.47}
$$

the steady-state solution of the Riccati equation simplifies to

$$
\mathbb{E}\left[\Delta^2\tilde{\omega}^{\mathrm{SS}}(t)\right] = \sqrt{q_\omega\kappa_{\mathrm{c}}} \quad \text{for} \quad N > \frac{2}{\kappa_{\mathrm{c}}}\sqrt{\frac{q_\omega}{\eta M}}\,\mathrm{e}^{(M+\kappa_{\mathrm{c}})t/2}
\tag{5.48}
$$

which coincides with the form of the CS limit for an OU process (4.82) when $t \gtrsim \sqrt{\kappa_c/q_\omega}$, i.e.:

$$\mathbb{E}\left[\Delta^2 \tilde{\omega}^{SS}(t)\right] = \lim_{t \to \infty} V_{\sigma_0}^{CS}(t) = \sqrt{q_\omega \kappa_c}, \quad \text{for } t \gtrsim \sqrt{\frac{\kappa_c}{q_\omega}} \text{ and } N \gg \frac{2}{\kappa_c}\sqrt{\frac{q_\omega}{\eta M}}. \tag{5.49}$$

Hence, this proves that the chosen set of measurements and initial state, i.e. a continuous ho-
modyne measurement and CSS, give us the best possible precision in the steady state when the
atomic ensemble is large (i.e. $N \gg 1$, which is also required for the LG approximation to hold).

5.3.2 EXTENDED KALMAN FILTER

When our system is no longer LG but rather described by a nonlinear dynamical model like the
one in Sec. 5.2.2.2, we need an estimator capable of handing nonlinearities, such as an extended
Kalman filter (EKF). As discussed in Sec. 2.7, an EKF estimates the state vector $\boldsymbol{x}(t)$ in real time
from the noisy measurement record using a state-space model defined by *nonlinear* state and mea-
surement functions, $\boldsymbol{f}$ and $\boldsymbol{h}$, as well as their associated noise vectors. Namely, at each time t, the
EKF produces the state estimate $\tilde{\boldsymbol{x}}(t)$ and its error covariance $\boldsymbol{\Sigma}(t)$ by integrating the differential
equations along the photocurrent record $\boldsymbol{y}(t) = \{y(t)\}_{\leq t}$ [29]:

$$\dot{\tilde{\boldsymbol{x}}}(t) = \boldsymbol{f}(\tilde{\boldsymbol{x}}(t), u(t), 0, t) + \boldsymbol{K}(t)(y(t) - h(\tilde{\boldsymbol{x}}(t), 0, t)) \tag{5.50a}$$

$$\dot{\boldsymbol{\Sigma}}(t) = (\boldsymbol{F}(t) - \boldsymbol{G}(t)\boldsymbol{S}\boldsymbol{R}^{-1}\boldsymbol{H})\boldsymbol{\Sigma}(t) + \boldsymbol{\Sigma}(t)(\boldsymbol{F}(t) - \boldsymbol{G}(t)\boldsymbol{S}\boldsymbol{R}^{-1}\boldsymbol{H})^{\mathrm{T}} +$$
$$+ \boldsymbol{G}(t)(\boldsymbol{Q} - \boldsymbol{S}\boldsymbol{R}^{-1}\boldsymbol{S}^{\mathrm{T}})\boldsymbol{G}(t)^{\mathrm{T}} - \boldsymbol{\Sigma}(t)\boldsymbol{H}^{\mathrm{T}}\boldsymbol{R}^{-1}\boldsymbol{H}\boldsymbol{\Sigma}(t), \tag{5.50b}$$

where in the state update-predict equation for the estimate (5.50a), the term

$$y(t) - h(\tilde{\boldsymbol{x}}(t), 0, t), \tag{5.51}$$

represents the innovation, i.e. the difference between the actual measurement (provided by the
model in Eq. (5.32)) and the predicted measurement based on the current state estimate. This
innovation is scaled by the Kalman gain

$$\boldsymbol{K}(t) := (\boldsymbol{\Sigma}(t)\boldsymbol{H}^{\mathrm{T}} - \boldsymbol{G}(t)\boldsymbol{S})\boldsymbol{R}^{-1}, \tag{5.52}$$

which couples the state and covariance $\boldsymbol{\Sigma}(t)$ in Eq. (5.50b). The matrices $\boldsymbol{F}(t)$, $\boldsymbol{G}(t)$, and $\boldsymbol{H}$ are the
Jacobians of the nonlinear functions, determined by the dynamical model in Eq. (5.28). Specifi-

cally, we have:

$$\boldsymbol{F}(t) := \nabla_{\boldsymbol{x}} \boldsymbol{f}|_{(\tilde{\boldsymbol{x}}, u, 0)}, \quad \boldsymbol{G}(t) := \nabla_{\xi} \boldsymbol{f}|_{\tilde{\boldsymbol{x}}} \quad \text{and} \quad H := \nabla_{\boldsymbol{x}} h, \tag{5.53}$$

with their exact form provided in Sec. E.8. Because the entries of these matrices depend on the current state estimate $\tilde{\boldsymbol{x}}(t)$, they must be recomputed at each EKF iteration (see Sec. 2.7 for an introduction to the EKF). In contrast, the noise matrices are predetermined:

$$\boldsymbol{Q} := \mathbb{E}\big[\boldsymbol{q}(t)\boldsymbol{q}^{\mathrm{T}}(t)\big] = \mathbb{I}, \quad R := \mathbb{E}\big[r^2(t)\big] = \eta, \quad \text{and} \quad \boldsymbol{S} := \mathbb{E}[\boldsymbol{q}(t)r(t)] = (\sqrt{\eta},\, 0)^{\mathrm{T}}. \tag{5.54}$$

The initial estimates, $\tilde{\boldsymbol{x}}(0)$, and its error covariance, $\boldsymbol{\Sigma}(0) = \mathbb{E}\big[\Delta^2 \tilde{\boldsymbol{x}}(0)\big]$, are chosen based on the assumed prior distribution [29].

5.3.3 Linear-quadratic regulator

A naive approach to controlling the state of the atoms is to directly compensate the estimated frequency by setting

$$u(t) = -\tilde{\omega}(t), \tag{5.55}$$

which we refer to as *field compensation*. In principle, this should cancel the precession; however, as the EKF estimate $\tilde{\omega}(t)$ is only approximate, small errors accumulate over time and the resulting control becomes suboptimal, leading to residual precession that degrades overall performance. To build upon the frequency cancellation achieved by the initial control, we design a more refined control law using linear-quadratic regulator (LQR) theory (introduced in Sec. 2.8.1). In our approach, the overall control input is expressed as

$$u(t) = -\tilde{\omega}(t) - v(t) \tag{5.56}$$

where the term $\tilde{\omega}(t)$ provides the primary compensation for the rapid Larmor precession, and the additional correction $v(t)$ is determined via LQR to minimize a prescribed quadratic cost function. Once the frequency-cancellation part of the control counteracts the rapid Larmor precession, which is responsible for driving the state away from the x-axis, the system dynamics slow down and become nearly linear. Then, we can assume that the system resides in the LG regime

and describe the system with a reduced state vector

$$\boldsymbol{z}(t) = (\langle \hat{J}_y \rangle, \omega)^{\mathrm{T}}, \tag{5.57}$$

which simplifies the dynamics and allows us to apply the optimal control strategy derived from LQR theory. Under a linearization of the full dynamical model at short timescales [54, 69], the state vector evolves according to Eqs. (5.21-5.23), which can be further approxmated as:

$$\dot{\boldsymbol{z}}(t) = \boldsymbol{A}\,\boldsymbol{z}(t) + \boldsymbol{B}\,u(t) + \boldsymbol{G}(t)\,\boldsymbol{q}(t), \tag{5.58}$$

where

$$\boldsymbol{A}(t) = \begin{pmatrix} 0 & J \\ 0 & -\chi \end{pmatrix}, \quad \boldsymbol{G}(t) = \begin{pmatrix} 2\sqrt{\eta M}\mathrm{V}_y(t) & 0 \\ 0 & \sqrt{q_\omega} \end{pmatrix}, \quad \boldsymbol{B}(t) = \begin{pmatrix} J \\ 0 \end{pmatrix}, \tag{5.59}$$

with $\boldsymbol{q} = (\xi, \xi_\omega)^{\mathrm{T}}$ being the same stochastic term as in Eq. (5.33), such that $\mathrm{dW} = \xi\mathrm{d}t$. Notably, in the LG regime, the variance of $\hat{J}_y$, $\mathrm{V}_y(t)$, is a deterministic function and can be computed analytically [55, 69] (see App. E.2). The LQR design is based on minimizing a *quadratic* cost function [29, 55]:

$$I(u) = \int_0^\infty \mathrm{d}t \ \left[\boldsymbol{z}^{\mathrm{T}}(t)\boldsymbol{P}\boldsymbol{z}(t) + u(t)Vu(t)\right] \tag{5.60}$$

$$= \int_0^\infty \mathrm{d}t \ \left[p_J\langle \hat{J}_y \rangle_{(c)}^2 + p_\omega \omega^2 + \nu u^2(t)\right], \tag{5.61}$$

where, following Ref. [55], we choose the cost matrices $P \geq 0$ and $V > 0$ to take the form

$$\boldsymbol{P} = \begin{pmatrix} p_J & 0 \\ 0 & p_\omega \end{pmatrix} \quad \text{and} \quad V = \nu, \tag{5.62}$$

with $p_J, p_\omega \geq 0$ and $\nu > 0$. This cost function not only penalizes deviations of $\omega(t)$ from zero (thereby counteracting the Larmor precession) but also drives the angular-momentum component $\langle \hat{J}_y \rangle_{(c)}$ towards zero.

Since the dynamics in Eq. (5.58) are linear, the optimal control law can be obtained by neglecting the stochastic terms, which only increase the attainable minimal cost (5.60) on average [29]. Moreover, when the state-space model is linear, as in Eqs. (5.21-5.22), the optimal control design can be decoupled from the state estimation [29] (see also Sec. 2.8.2). The resulting LQR solution

takes the form:

$$u(t) = -\boldsymbol{K}_C \tilde{\boldsymbol{z}}(t), \tag{5.63}$$

$$\boldsymbol{K}_C := V^{-1} \boldsymbol{B}^{\mathrm{T}} \boldsymbol{\Lambda}, \tag{5.64}$$

$$0 = \boldsymbol{A}^{\mathrm{T}} \boldsymbol{\Lambda} + \boldsymbol{\Lambda} \boldsymbol{A} + \boldsymbol{P} - \boldsymbol{\Lambda} \boldsymbol{B} V^{-1} \boldsymbol{B}^{\mathrm{T}} \boldsymbol{\Lambda} \tag{5.65}$$

where the optimal control field $u(t)$ is linearly related to the state estimator $\tilde{\boldsymbol{z}}(t)$ by the gain matrix $\boldsymbol{K}_C$, and the gain, defined in Eq. (5.64), couples to the control "covariance" $\boldsymbol{\Lambda}$, the steady-state solution of the algebraic Ricatti equation in Eq. (5.65). Since the matrices A and B in Eq. (5.58) are time-independent, $\boldsymbol{\Lambda}$ (and consequently $\boldsymbol{K}_C$ and $u(t)$) can be determined analytically [55]. A detailed derivation, as provided in App. E.6, shows that the gain matrix simplifies so that the control law can be recast as

$$u(t) = -\tilde{\omega}(t) - \lambda \left\langle \tilde{\hat{J}}_y(t) \right\rangle_{(\mathrm{c})}, \tag{5.66}$$

where $\lambda := \sqrt{p_J/\nu}$ is a constant parameter functioning like a control "knob". Notably, by setting $\lambda = 0$, then Eq. (E.75) we recover the field compensation strategy considered initially. In order to apply this result to the setting of Sec. 5.2.2.2, one has to simply combine the EKF that estimates the full state $\boldsymbol{x}(t) = (\langle \hat{J}_x \rangle_{(\mathrm{c})}, \langle \hat{J}_y \rangle_{(\mathrm{c})}, V_x^{(\mathrm{c})}, V_y^{(\mathrm{c})}, V_z^{(\mathrm{c})}, C_{xy}^{(\mathrm{c})}, \omega)^{\mathrm{T}}$ with the LQR, along the lines of the linear-quadratic Gaussian controller discussed in Sec. 2.8.2. In particular, to apply the LQR strategy in this broader context, we generalize the control law of Eq. (5.63) to

$$u(t) = -\boldsymbol{K}_C \boldsymbol{\Xi} \tilde{\boldsymbol{x}}(t) \tag{5.67}$$

where

$$\boldsymbol{\Xi} := \begin{pmatrix} 0 & 1 & 0 & \ldots & 0 & 0 \\ 0 & 0 & 0 & \ldots & 0 & 1 \end{pmatrix} \tag{5.68}$$

is just a matrix that extracts the relevant state components of $\boldsymbol{x}(t)$ that appear in $\boldsymbol{z}(t)$, while $\tilde{\boldsymbol{x}}(t)$ is now the EKF estimator provided by Eq. (5.50a) [65].

5.4 Results

In the previous sections, we detailed the model of the optical atomic magnetometer and its numerical simulation, including its atomic ensemble dynamics, simulation methods, and the estimation and control strategies that enhance its performance. We now present our main results,

which we organize as follows: in Sec. 5.4.1, we begin by examining the response of the magnetometer under weak-field sensing conditions, where it exhibits LG dynamics. Sec. 5.4.2 then details the real-time sensing of a precession-inducing field, with separate analyses for constant, fluctuating, and nonlinear MCG-like field profiles. Finally, Sec. 5.4.3 demonstrates the generation of multiparticle entanglement by evaluating conditional as well as unconditional spin squeezing.

5.4.1 WEAK-FIELD SENSING

When employing large ensembles to track weak fields, which is the relevant regime when discussing detection limits, one can safely assume that the system is LG for times within the coherence time of the system and measurement (i.e. $t < (M + \kappa_c)^{-1}$ (see Sec. 5.2.2.1). In this regime, the KF provides an *optimal*, recursive strategy (see Sec. 5.3.1). By continually incorporating new measurement data, the filter updates both the state estimate and its error covariance, ensuring that the mean-squared error is minimized at every step. Crucially, the KF not only provides a real-time estimate of the magnetic field but also yields a *quantifiable metric of its precision* through the covariance matrix, which for the case of a KF, coincides with the aMSE (see Sec. 2.6).

In Fig. 5.6 we show the minimal averaged error (solid, red) for the estimator of the Larmor frequency (the squared root of the aMSE, $\sqrt{\mathbb{E}[\Delta^2\tilde{\omega}(t)]} = \sqrt{\mathbf{\Sigma}_{\omega,\omega}}$). The top plots display the error as a function of time for two atomic ensembles of different sizes: $N = 10^9$ (a) and $N = 10^5$ (b), and the bottom plots show the error as a function of the number of atoms for two different time slices: $t = 10^{-4}$s (c) and $t = 10^{-2}$s (d). In all cases, the error is obtained by numerically solving the Riccati equation of Eq. (5.36), with the model matrices defined in Eq. (5.23) and the exact form for the variance of $\hat{J}_y$ provided in Eq. (E.26). In Fig. 5.6 we further analyze the different behaviors of the error of the KF (solid, red), by comparing it to the analytical solutions of the KF derived in Sec. 5.3.1 for (1) the steady state (dashed, green) and (2) the noiseless regime (dashed, blue). In particular, for short times or low atomic numbers, the error of $\tilde{\omega}$ behaves like in the ideal case of negligible dephasing and field fluctuations, revealing a non-classical scaling of the estimation error with both time and the number of atoms:

$$\mathbb{E}\left[\Delta^2\tilde{\omega}(t)\right] \propto \frac{1}{N^2 t^3}, \tag{5.69}$$

highlighting the quantum-enhanced sensitivity achievable in our setup[54]. Over time, however, the precision scaling deteriorates, as shown in the two top plots. This occurs either when the
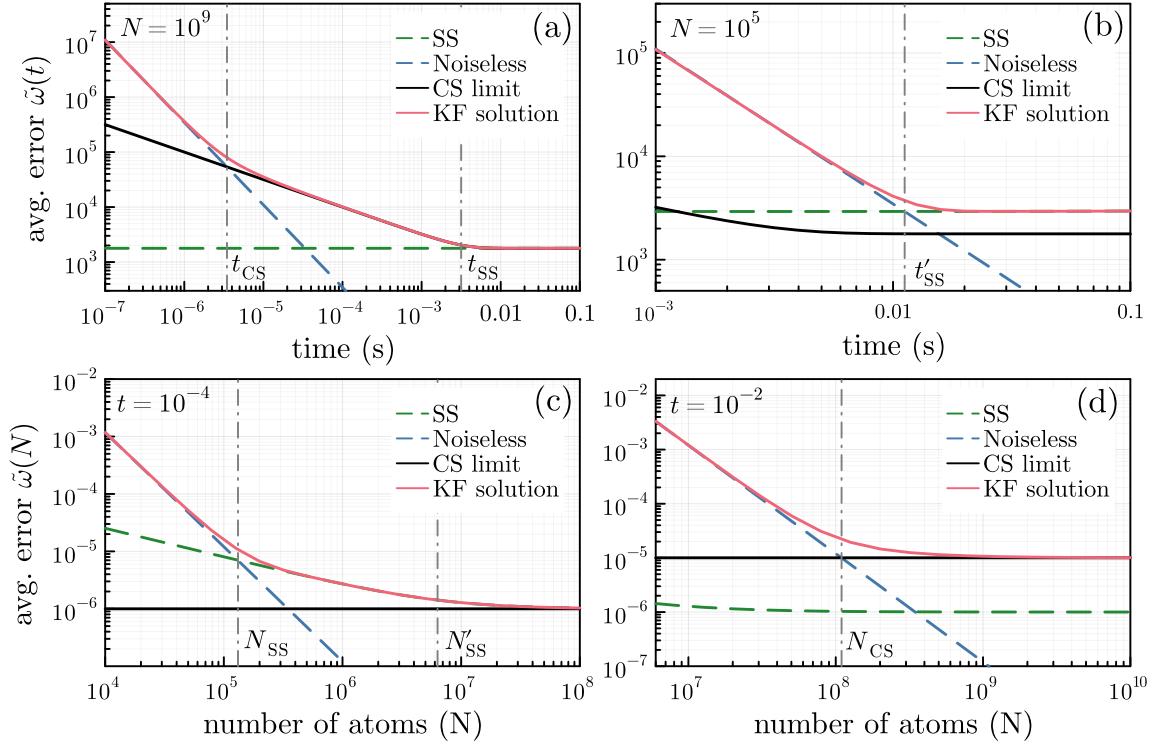
Figure 5.6: The minimal estimation error as a function of time. Subplots (a) and (b) show the time evolution of the averaged error of estimating $\omega(t)$, $\sqrt{\mathbb{E}[\Delta^2\tilde{\omega}(t)]}$, for large ($N = 10^9$) and small ($N = 10^5$) ensembles, respectively. Subplots (c) and (d) present the dependence of the error on the number of atoms at short ($t = 10^{-4}$s, (c)) and long ($t = 10^{-2}$s, (d)) measurement times. Other parameters used to generate this figure are: $M = 100\,\text{kHz}$, $q_\omega = 10^{14}\text{rad}^2\,\text{s}^{-3}$, $\kappa_c = 100\,\text{mHz}$, $\chi = 0$ and $\eta = 1$. Colored dashed lines serve as references, representing different scaling behaviors of the Riccati solution: dashed green corresponds to the steady-state (SS) solution of the KF, dashed blue represents the analytical solution obtained in an ideal noiseless scenario, and finally, solid black corresponds to the quantum limit imposed by dephasing and field fluctuations, i.e. the CS limit. This quantum limit is attained for large N when $t > t_{CS}$, proving in this regime the optimality of our measurement and initial state.

system reaches a steady state (Fig. 5.6 (b)) at time

$$t'_{SS} = 3^{1/3}\left(\frac{4}{N^2\eta M q_\omega}\right), \tag{5.70}$$

or by first attaining the CS limit of Eq. (4.83) (see Fig. 5.6 (a)) at time

$$t_{CS} = \frac{2}{N}\sqrt{\frac{3}{\eta M \kappa_c}}, \tag{5.71}$$

which also later coincides with the KF steady state after $t = t_{SS}$. One might therefore realize that

the CS bound in the limit of an infinitely wide prior, as presented in Eq. (4.83), has two distinct regimes:

$$V_\infty^{\text{CS}}(t) = \sqrt{q_\omega \kappa_c} \coth\left(t\sqrt{\frac{q_\omega}{\kappa_c}}\right) = \begin{cases} \dfrac{\kappa_c}{t} & \text{for } t < t_{\text{SS}} \\ \sqrt{q_\omega \kappa_c} & \text{for } t \geq t_{\text{SS}} \end{cases} \qquad (5.72)$$

where

$$t_{\text{SS}} = \sqrt{\frac{\kappa_c}{q_\omega}}. \qquad (5.73)$$

Fig. 5.6 (c) and (d) illustrate that while small ensembles may achieve a precision scaling of $1/N^2$ in estimating $\omega(t)$, this advantage diminishes as the system size increases. As a result, larger atomic ensembles do not provide a net gain in precision. That occurs at

$$N_{\text{CS}} = \frac{2}{t}\sqrt{\frac{3}{\eta M \kappa_c}}, \qquad \text{and} \quad N'_{\text{SS}} = \frac{2}{\kappa_c}\sqrt{\frac{q_\omega}{\eta M}}. \qquad (5.74)$$

For a certain range of ensemble sizes, i.e. $[N_{\text{SS}}, N'_{\text{SS}}]$, where

$$N_{\text{SS}} = \frac{2\,3^{2/3}}{t^2 \sqrt{\eta M q_\omega}} \qquad (5.75)$$

and for short times, the KF error follows the steady-state solution (dashed, green) of Eq. (5.46), which scales as $\propto 1/N^{1/2}$ (and thus, the error displayed in Fig. 5.6 (c) scales as $\propto 1/N^{1/4}$).

The most important message that Fig. 5.6 aims to convey is that for large ensembles (i.e. $N = 10^9$) the aMSE of the KF attains the lower bound on the error dictated by dephasing and fluctuations, referred throughout this thesis as Classical Simulation (CS) limit. Even though we already know that for a LG system, the KF is the optimal estimator, Fig. 5.6 (a) also confirms that for large atomic ensembles, our choice of measurement and initial state are also optimal, since the CS limit is independent of our choice of measurement and initial state (see Chap. 4). In summary, applying the KF to a system initialized in a CSS and continuously monitored through homodyne measurement not only resolves the estimation error in real time but also establishes an optimal framework for high-precision sensing of weak magnetic fields.

5.4.2 Real-time sensing of a larger frequency

We now turn our attention to the practical task of real-time sensing of a precession-inducing magnetic field. As before, our main goal is to estimate the Larmor frequency and benchmark the performance of our proposed EKF+LQR strategy against other estimation and control approaches as well as the quantum limit dictated by dephasing and field fluctuations derived in Chap. 4. To explore the role of quantum effects in enhancing estimation, we go beyond analyzing the average error of the frequency estimate, $\sqrt{\mathbb{E}[\Delta^2\tilde{\omega}(t)]}$, and also examine the averaged evolution of the spin-squeezing parameter $\mathbb{E}[\xi_y^2(t)]$ (defined in Sec. 1.4.7), along with the averaged ensemble polarization, $\mathbb{E}[\langle \hat{J}_x(t)\rangle_{(c)}]$.

5.4.2.1 Constant field

In what follows, we investigate the real-time tracking of the simplest type of field: a constant magnetic field. First, we do so by using the full SME of Eq. (5.13) to model the evolution of the system. Using this "brute-force" approach we perform two important tasks: (1) benchmarking of the EKF+LQR strategy against alternative methods and (2) validation of the CoG approximation introduced in Sec. 5.2.2.2. Once validated, we employ this approximation to study larger atomic ensembles, determining whether our approach reaches the quantum limit imposed by dephasing, and hence, testing the optimality of our magnetometry setup.

Low atomic numbers   Fig. 5.7 *(top)* presents the real-time estimation of a constant magnetic field, simulated via the full SME of Eq. (5.13) for a system of $N = 200$ atoms. As time progresses, the continuous incorporation of measurement data reduces the error (green shaded area), while the estimate (solid, red) converges towards the true value (solid, blue). The middle panel illustrates that spin squeezing (solid, blue) quickly emerges and is maintained throughout the experiment, which can be correctly estimated with the EKF (dashed, red) for times within the LG region. Finally, the bottom plot shows that the averaged error approaches the quantum limit dictated by dephasing (i.e., the CS limit, in solid, black):

$$\mathbb{E}\big[\Delta^2\tilde{\omega}(t)\big] \geq \frac{1}{\dfrac{1}{\sigma_0^2} + \dfrac{t}{\kappa_{\mathrm{c}}}}, \tag{5.76}$$
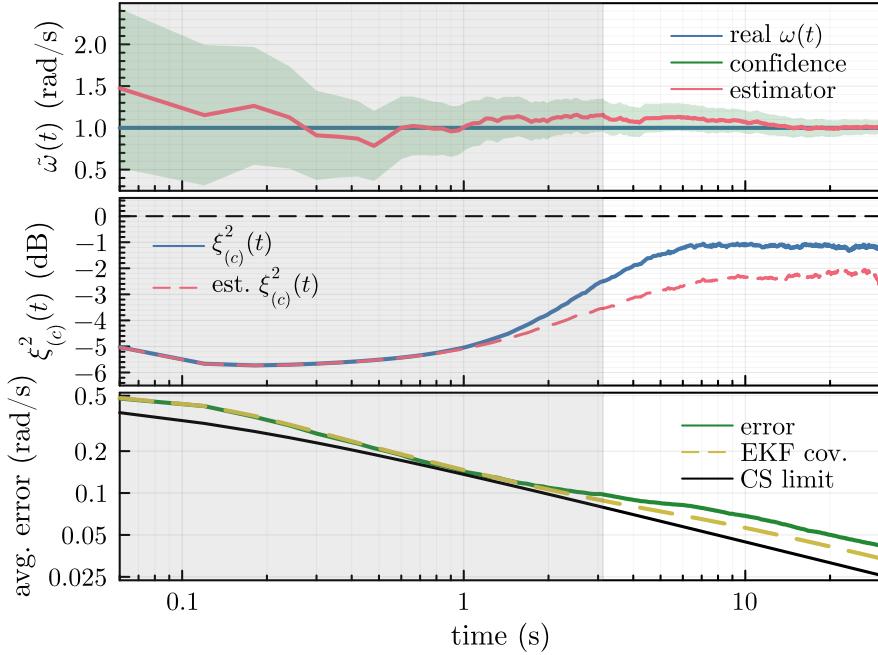
Figure 5.7: Quantum-enhanced tracking of a constant magnetic field with $N = 200$**.** *Top*: Constant field (solid, blue) tracked in real time by the EKF estimate (solid, red). The shaded area represents the confidence band limited by the averaged error, i.e. $\pm 2\sqrt{\mathbb{E}[\Delta^2\tilde{\omega}(t)]}$. *Middle*: Averaged evolution of the spin-squeezing parameter (solid, blue), compared with its real-time EKF prediction (dashed, red). When the parameter is below the dashed black line of $\xi^2_{y,(c)}(t) = 1$, it indicates squeezing and thus, the presence of multipartite entanglement. *Bottom*: Evolution of the average error (solid, green) in estimating the fluctuating field, $\sqrt{\mathbb{E}[\Delta^2\tilde{\omega}(t)]}$, which attains the quantum limit imposed by dephasing (solid, black), as correctly predicted by the EKF covariance (dashed, yellow), which matches the averaged error within the LG regime. In all plots, averaging was performed over 1000 atom stochastic trajectories, and the parameters used are: $N = 200, \kappa_c = 0.02, \kappa_\ell = 0, M = 0.3$, $\omega = 1, \eta = 1$, and $\sigma_0 = 0.5$.

where σ_0 is the standard deviation of the initial prior of the Larmor frequency. As expected, the EKF covariance (its squared root in dashed yellow) provides a good estimate of the error within the LG regime, although it becomes slightly optimistic beyond the $(M + \kappa_c)^{-1}$ mark.

We further compare the performance of the EKF+LQR strategy against less sophisticated methods, including the KF with field compensation (green), the EKF without control (yellow), and the EKF with field compensation (blue). In the left column of Fig. 5.8, we keep the estimator as the EKF and switch the type of controller: none (yellow), frequency compensation (blue) and LQR (red). In contrast, in the right column we fix the control as LQR and compare different estimators: either KF (green) or EKF (red). This comparison in Fig. 5.8 (a)-(b) under controlled conditions, e.g. no local dephasing such that we can simulate for larger N, highlights the supe-
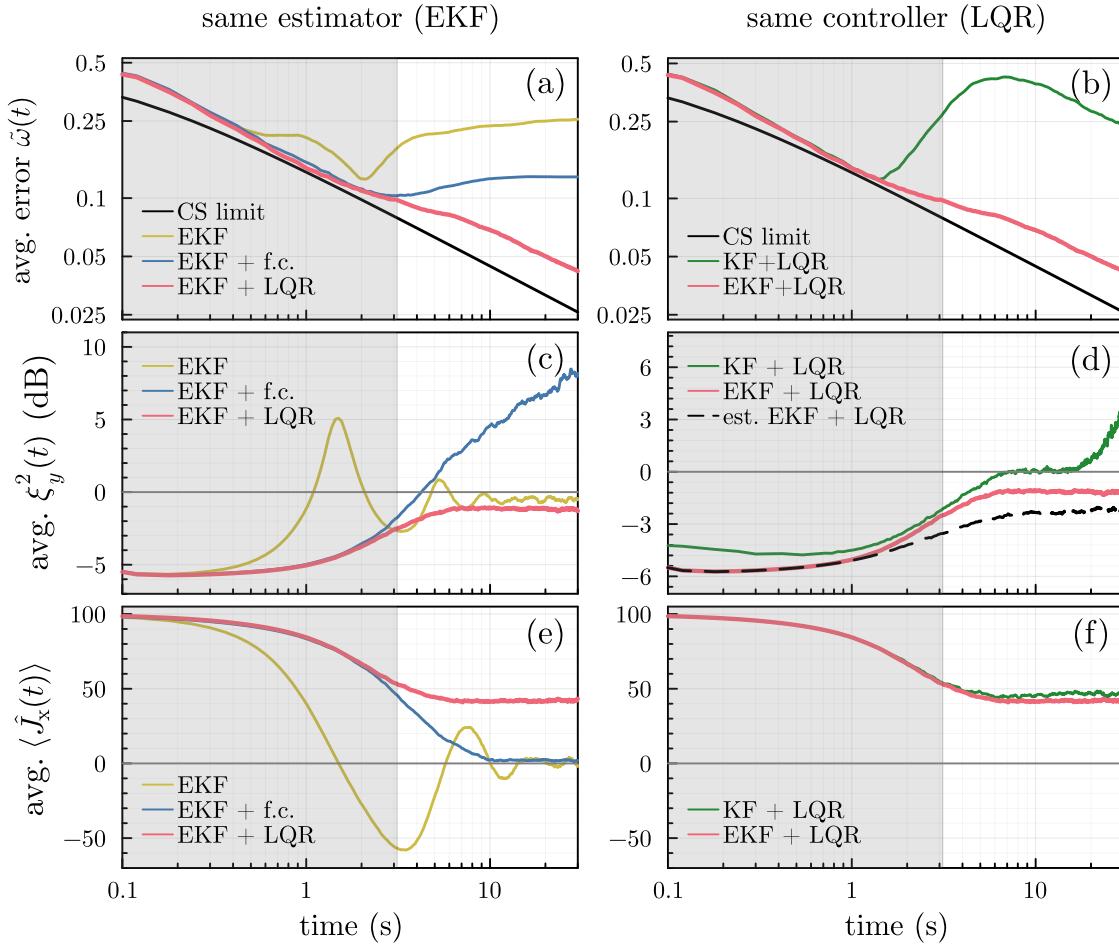
Figure 5.8: Performance of different estimation and control strategies. The top row shows the averaged estimation error, the middle row displays the spin-squeezing parameter (as defined in (5.26)), and the bottom row illustrates the ensemble polarization, $\left\langle \hat{J}_x \right\rangle_{(c)}$. In the left column (subplots (a), (c), and (e)), the estimator is fixed as the EKF while three control methods are compared: no control (yellow), field compensation (blue), and LQR (red). In the right column (subplots (b), (d), and (f)), the control method is fixed to LQR while two different estimators are considered: KF (green) and EKF (red). In both subplot (a) and (b), the EKF+LQR strategy (red) is benchmarked against the CS limit (black), which sets the quantum limit imposed by dephasing on the attainable error. Notably, the EKF+LQR approach outperforms the other methods by achieving the lowest estimation error, which continues to decrease even beyond the LG regime (shaded grey). Subplots (c)–(f) demonstrate that only the combination of an EKF with an LQR maintains both spin squeezing and polarization throughout the experiment. Additionally, in subplot (d), the black dashed line indicates the overoptimistic predictions from the EKF, as shown already in Fig. 5.7. The parameters used in the SME (5.13) for simulations are: $N = 200$, $\kappa_c = 0.02$, $\kappa_\ell = 0$, $M = 0.3$, $\omega = 1$ and $\eta = 1$. The KF and EKF estimators are initialized with the mean $\tilde{\boldsymbol{x}}(0) = (N/2, 0, 0, N/4, N/4, 0, \mu_0)^{\mathrm{T}}$ and covariance $\boldsymbol{\Sigma}(0) = \mathrm{Diag}[0, 0, 0, 0, 0, 0, \sigma_0^2]$ dictated by the initial CSS state of the atoms, and the Gaussian prior distribution for $\omega \sim \mathcal{N}(\mu_0, \sigma_0^2)$. All results are obtained after averaging over $\nu = 1000$ measurement trajectories, whereas ω-averaging is avoided by choosing its true value $\omega = 1$ for a prior with $\mu_0 = \omega + \sigma_0 = 1.5$ and $\sigma_0 = 0.5$.

riority of combining an EKF with an LQR, since it yields the lowest average error (red) closest to the quantum limit dictated by dephasing (i.e. the CS limit in black of Eq. (5.76)). The other estimation and control approaches also fail to yield both spin-squeezing and a non-vanishing polarization for the full duration of the experiment ($T = 30\,\text{s} > (M + \kappa_c)^{-1} \approx 3\,\text{s}$), as illustrated in Fig. 5.8 (c)-(d) for squeezing and (e)-(f) for the polarization.

Lastly, simulating the full SME of Eq. (5.13) allows us to verify the accuracy of the CoG approximation in simulating our system. As briefly discussed in Sec. 5.2.2.2 and extensively in App. E.4, we compare the Larmor frequency estimates obtained using two methods: one by solving the full SME and the other by solving a system of SDEs given by the CoG approximation (see Sec. 5.2.2.2). In both cases, the estimation and control is carried out using an EKF combined with an LQR, since Fig. 5.8 established it as the best known strategy. The results, shown in the first column of Fig. E.3, demonstrate that the relative error remains below 1% at all times and decreases with increasing N. This confirms that the CoG approximation is valid and can be confidently used in the subsequent analysis.

HIGH ATOMIC NUMBERS   For larger ensembles ($N \sim 10^5 - 10^{15}$ c.f. [157, 158]), simulating the exact SME becomes computationally prohibitive. In this regime, we rely on the CoG approximation, which accurately captures the conditional evolution of the first and second moments of the angular-momentum operators.

The *top row* of Fig. 5.9 demonstrates that, for $N = 10^5$ atoms, the EKF+LQR strategy achieves optimal performance when its averaged estimation error (solid, red) reaches the quantum limit imposed by dephasing Eq. (4.84) (solid, black). This optimal behavior is observed under both collective (a) and local (b) decoherence. Although the quantum limit is reached only briefly within the LG regime under local dephasing, using an EKF is essential, as the KF is not applicable in this scenario (see Sec. 5.2.2.1). For collective dephasing, the EKF+LQR still outperforms the KF (solid, green), even within the LG regime (shaded grey area). Attaining the quantum limit ensures that the measurement, feedback, and initial state are optimal during that period, thereby addressing the open question posed in Ref. [62]. Moreover, the EKF covariance (its squared root in dashed, blue) closely tracks the averaged estimation error (solid, red), confirming that, despite the nonlinearity of the CoG model (5.28), the EKF provides reliable trajectory-dependent error estimates.

The *middle row* of Fig. 5.9 shows the averaged spin-squeezing parameter in decibels, while the *bottom row* displays the averaged ensemble polarization. In all cases, the EKF predictions (dashed,
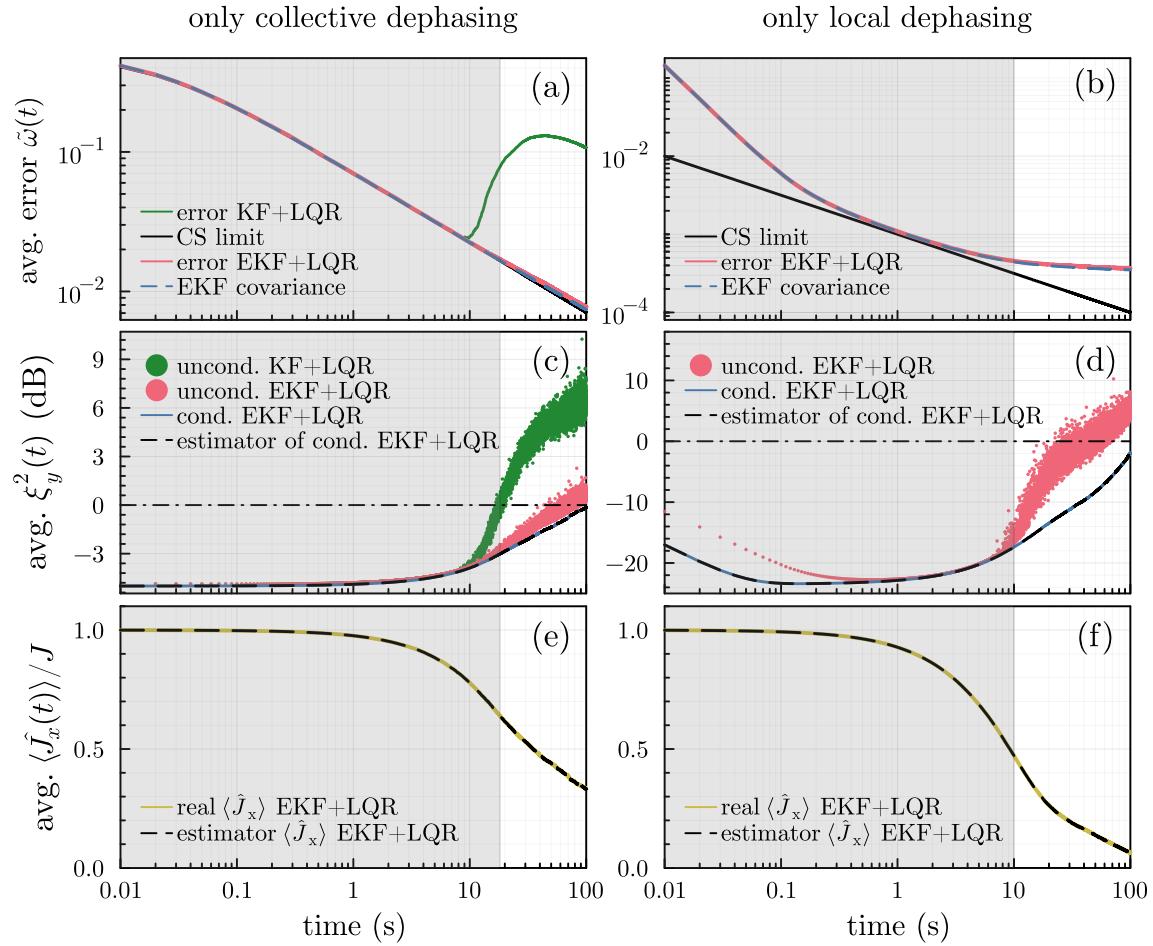
Figure 5.9: Performance in estimation and spin-squeezing extrapolated to large atomic ensembles ($N = 10^5$). Subplots (*a*) and (*b*) (*upper row*) depict the case of pure collective decoherence $\kappa_c = 0.005$, whereas subplots (*c*) and (*d*) (*lower row*) deal with pure local decoherence $\kappa_\ell = 0.05$. *Left column*: (*a*) and (*c*) show the error (aMSE) attained by the EKF+LQR strategy (red dots) when estimating ω and its average prediction by the EKF (blue line), $\mathbb{E}[\mathbf{\Sigma}_{\omega\omega}]$, both being lower-bounded by the CS limit (5.76) (black line). For collective decoherence, the performance of KF+LQR strategy is also included (grey dots) to emphasize its failure beyond the LG regime (pink shading in all subplots). *Right column*: (*b*) and (*d*) illustrate the evolution of the spin-squeezing parameter (5.26) for conditional (blue line) and unconditional (red dots) dynamics, as compared with its classical threshold (horizontal dash-dotted line). The evolution of the ensemble polarization $\langle \hat{J}_x \rangle = \mathbb{E}[\langle \hat{J}_x \rangle_{(c)}]$ (green line) is also shown in both cases in extra lower plots. Both the conditional spin-squeezing and the polarization in (*b*) and (*d*) are estimated very accurately by the EKF on average (superimposed dashed black lines). The above data is simulated employing the CoG model (5.28) with other parameters set to: $M = 0.05$, $\omega = 1$ and $\eta = 1$. As in Fig. 5.8, the KF and EKF estimators are initialized with the mean $\tilde{\boldsymbol{x}}(0) = (N/2, 0, 0, N/4, N/4, 0, \mu_0)^{\mathrm{T}}$ and covariance $\mathbf{\Sigma}(0) = \mathrm{Diag}[0, 0, 0, 0, 0, 0, \sigma_0^2]$ dictated by the initial CCS state and the Gaussian prior distribution for $\omega \sim \mathcal{N}(\mu_0, \sigma_0^2)$. All results are obtained after averaging over $\nu = 20000$ measurement trajectories, while ω-averaging is avoided by choosing the prior with $\sigma_0 = 0.5$ and $\mu_0 = \omega + \sigma_0 = 1.5$.

black) closely match the simulation results: the averaged conditional spin squeezing (solid, blue) and the polarization or averaged $\langle \hat{J}_x \rangle_{(c)}(t)$ (solid, yellow). This excellent agreement holds as long as the CoG approximation is reliable, and contrasts with our earlier findings in Fig. 5.8(d), where the EKF predictions overestimated the spin-squeezing parameter. Moreover, subplots (c) and (d) confirm that the EKF+LQR strategy not only generates conditional spin-squeezing, as already shown in Fig. 5.8(c) and (d) for $N = 200$, but also yields significant unconditional spin-squeezing (red dots) above the classical limit (horizontal dash-dot black line). Additionally, subplot (c) demonstrates that the EKF outperforms the KF (green dots) in preserving an unconditional multiparticle entangled state beyond the LG regime, though it also eventually degrades after approximately $\sim 50\,\text{s}$.

5.4.2.2 Fluctuating field

In this subsection, we extend our discussion to the case where the Larmor frequency fluctuates as an Orstein-Uhlenbeck process (OUP) [55, 159] (see Sec. 1.3.6):

$$\mathrm{d}\omega(t) = -\chi\omega(t)\mathrm{d}t + \sqrt{q_\omega}\mathrm{d}W_\omega, \tag{5.77}$$

where $\mathrm{d}W_\omega$ is a (new) Wiener increment, $\chi > 0$ is the decay rate, and $q_\omega > 0$ sets the strength of fluctuations, which are chosen to be $\chi = 0.01\,\text{s}^{-1}$ and $q_\omega = 10^4\text{rad}^2\text{s}^{-3}$ for the simulations presented throughout this section. The results discussed here and in the following sections are obtained using experimentally realistic parameters, chosen to reflect conditions achievable in state-of-the-art setups. These parameter values are inspired by Ref. [63], though their adaptation to our setting required interpreting experimental quantities such as the measurement strength, which are not directly provided but must be inferred from various experimental parameters. A detailed explanation is given in App. E.7. Given these parameters, the standard deviation of the fluctuations over the magnetometer coherence time T_2 (here 10 ms) is approximately 0.1% of the initial Larmor frequency $\omega_0 = 10^4\text{rad s}^{-1}$ (recall from Eq. (1.136) that the variance of an OU process can be approximated as $q_\omega t$ at short times $t \lesssim 1/\chi$). Fig. 5.10 (*top*) illustrates how the proposed EKF+LQR strategy is capable of tracking the OUP-induced fluctuations in real time for an experiment with realistic parameters (with the EKF equations explicitly written in App. E.8). Additionally, the EKF also provides an accurate estimate of the atomic spin-squeezing parame-
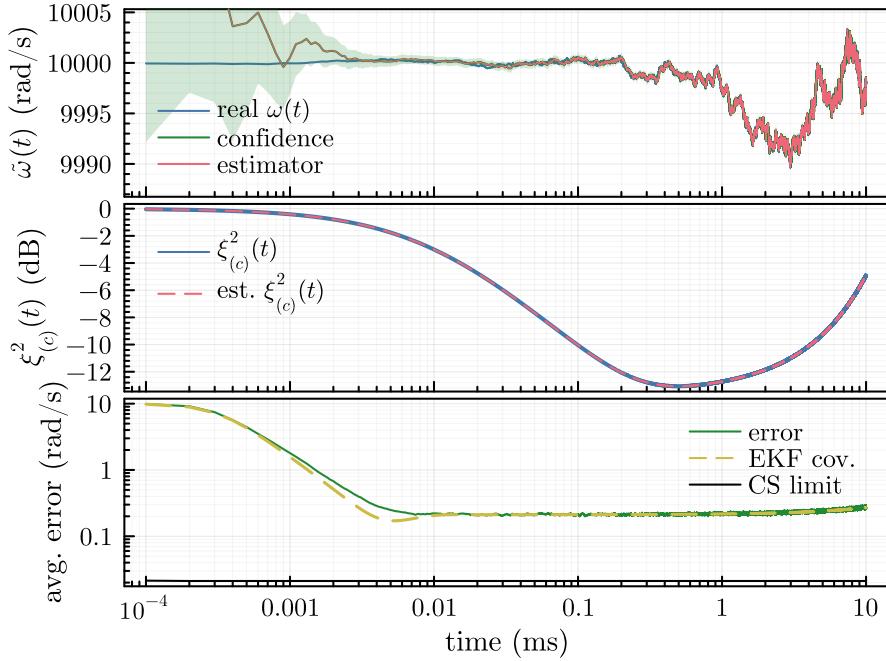
Figure 5.10: Quantum-enhanced tracking of a fluctuating magnetic field. *Top*: Efficiently tracking in real time of field fluctuations (solid, blue) by its corresponding EKF estimate (solid, red) after gathering only ≈ 0.01 ms of photocurrent data. The shaded area represents the confidence band limited by $\pm 2\sqrt{\mathbb{E}[\Delta^2 \tilde{\omega}(t)]}$, i.e. the error obtained upon averaging. *Middle*: Evolution of the (average) spin-squeezing parameter (solid, blue) in dB, compared with its real-time prediction by the EKF (dashed, red). *Bottom*: The averaged error (solid, green) in estimating the fluctuating field, $\sqrt{\mathbb{E}[\Delta^2 \tilde{\omega}(t)]}$, which stabilizes at $\approx 0.21 \, \text{rads}^{-1}$, as correctly predicted by the EKF covariance (grey). The quantum limit imposed by local dephasing (a.k.a. CS limit, in dashed, black) is not attained due to insufficient measurement strength ($\mathcal{M} = 10^{-8}$Hz [63], see also Fig. E.4) [65]. Averaging was performed over 1000 field+atom stochastic trajectories.

ter [143] (*middle*) that reaches $\gtrsim 13$ dB† at around 0.5 ms, being induced purely by the measurement backaction emerging at ≈ 0.01 ms. The magnetometer reaches its best resolution at times $0.01 \, \text{ms} \lesssim t \lesssim T_2$, where it tracks field fluctuations in real time with an error of $0.21 \, \text{rad s}^{-1}$ (*bottom*). The minimal averaged error (solid, green) is correctly predicted by the EKF covariance (its squared root in dashed, yellow). While the quantum limit imposed by local dephasing (Eq. (4.81) with $\kappa_c = 0$) sets a fundamental lower bound of $0.021 \, \text{rad s}^{-1}$ (solid, black), this limit is not reached under the current measurement strength. However, increasing $\mathcal{M}$ can, in principle, allow the system to attain this bound [65] (see Fig. E.4), though the required measurement strengths may not be experimentally feasible.

---

†The value of ≈ 2 dB reached in Ref. [63] should not be directly compared, as therein an unpolarised ensemble in the spin-exchange relaxation-free regime was considered.
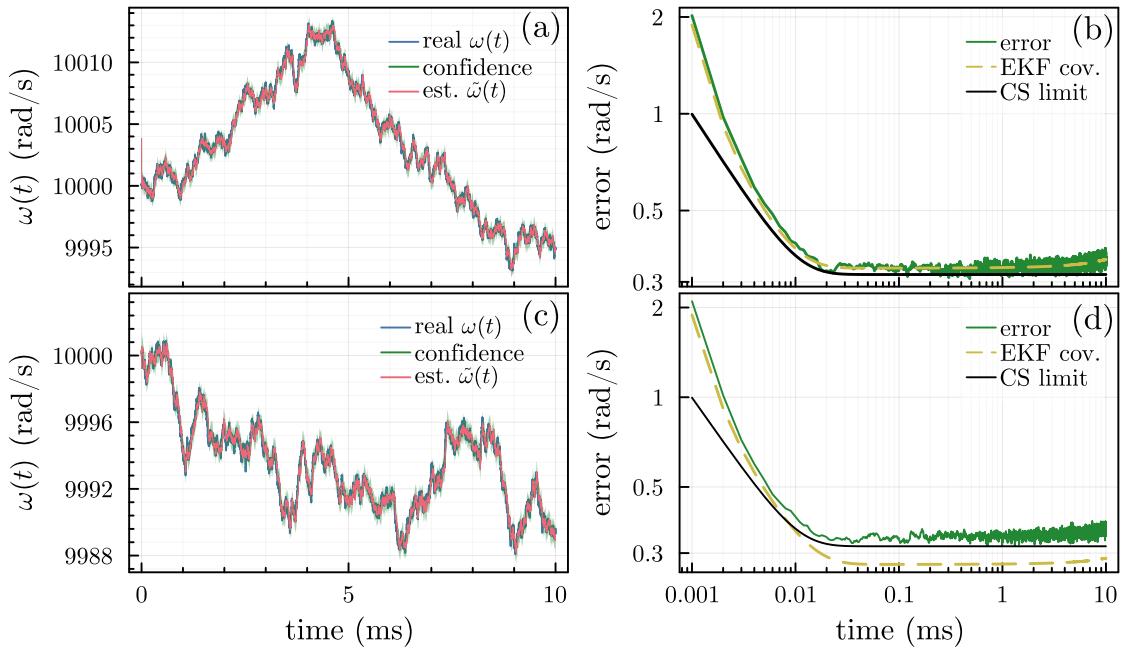
Figure 5.11: Tracking field fluctuations at the quantum limit (4.81). As in the bottom plot of Fig. 5.10, the true average error, $\sqrt{\mathbb{E}[\Delta^2\tilde{\omega}(t)]}$ (solid, green), is compared against the error predicted by the EKF (dashed, yellow), i.e. the covariance, and the quantum limit imposed by dephasing, as given in Eq. (4.81), now including collective dephasing $\kappa_c = 1\,\mu\text{Hz}$. Remarkably, the magnetometer operates at the quantum limit regardless of whether the EKF is provided with the exact OUP dynamics (5.77) (a-b) or a mismatched version (c-d). In the latter case, despite the EKF assuming a fluctuation strength that is half as large ($q_K = q_\omega/2$) and much faster decay ($\chi_K = 10\chi$), the average error (solid, green) still reaches the quantum limit (solid, black), while the EKF covariance (dashed, yellow) underestimates the error. Both plots were obtained by averaging over 1000 field+atom stochastic trajectories. Additionally, subplots (a) and (c) show a representative field trajectory (blue) alongside its EKF real-time estimate (red), which remains well within the confidence interval $\pm 2\sqrt{\mathbb{E}[\Delta^2\tilde{\omega}(t)]}$ (shaded green).

Another scenario where the magnetometer operates at the quantum limit, i.e. the CS limit of Eq. (4.81), is when a small amount of collective dephasing is present. In Fig. 5.11, we add $\kappa_c = 1\,\mu\text{Hz}\,(1/\kappa_c \approx 11\,\text{days})$ alongside the local dephasing of $\kappa_\ell = 100\,\text{Hz}$. As shown in Fig. 5.11 (b), while collective dephasing worsens the estimation error and raises the quantum limit, it also leads to a favorable outcome: the averaged estimation error (solid, green) saturates the quantum limit at $\approx 0.32\,\text{rad}\,\text{s}^{-1}$ (solid, black), indicating that the magnetometer operates in an optimal regime. This holds true even though spin-squeezing is no longer present, as collective dephasing dominates over measurement-induced correlations ($\kappa_c > \eta M$), preventing the generation of spin-squeezed states [69] (see also Fig. 5.5). Additionally, the quantum limit (solid, black) is not only reached when the EKF is provided with the exact OU field dynamics (subplot (b)), i.e. the parameters defining Eq. (5.77), but also when the field decay and fluctuating strength

are mismatched (subplot (d)). In this later case, the EKF operates assuming that the fluctuation strength is half as strong, $q_K = q_\omega/2$, and the field decay is ten times faster, $\chi_K = 10\chi$. Although under these mismatched conditions the EKF covariance (dashed, yellow) is no longer reliable, the averaged error (solid, green) remains unaffected and still attains the quantum limit. This result confirms that even with imperfect knowledge of the OU process parameters, the magnetometer can still operate optimally and yield field estimates at the quantum limit of precision, even though the EKF covariance in each realization no longer provides an exact measure of the error.

5.4.2.3 NONLINEAR MCG-LIKE FIELD

Similarly to electrocardiography, atomic magnetocardiography (MCG) [23, 24, 25, 26] is a non-invasive technique for imaging the magnetic fields generated by the electrical activity of the heart [22]. The goal in MCG is to reconstruct in real time the full magnetic waveform produced by the heart, capturing the characteristic P-wave, QRS-complex and T-wave [160], while filtering out unwanted stochastic noise [144] without resorting to extensive time-averaging [22].

As in previous sections, we use the EKF+LQR feedback loop to estimate in real time the cardiac magnetic signal from the homodyne photocurrent, since this type of recurrent estimation and control is well suited for tracking such complex, time-varying signals. To simulate a realistic MCG signal, we use a filtered VdP oscillator [161, 162] as defined in Eq. (5.5):

$$d\nu(t) = -p\,\omega(t)\,dt \tag{5.78}$$

$$d\omega(t) = \frac{k}{m}\nu(t)\,dt + 2\frac{c}{m}(1 - \nu(t))\,\omega(t)\,dt \tag{5.79}$$

$$dv(t) = \frac{|\nu(t)| - \nu(t)}{2T}\,dt - \frac{v(t)}{T}\,dt, \tag{5.80}$$

with parameters $p, k, m, c, T > 0$. The VdP model effectively reproduces the P-wave and QRS complex, though it does not fully capture the T-wave. We then superimpose white noise onto the VdP-generated signal to mimic the stochastic disturbances observed in practice. The simulation parameters are adapted from Ref. [63] for our setting, as discussed in App. E.7, with the EKF parameters for the VdP estimate matching the ones used to generate the clean VdP signal. Namely, $p = p_K = 10^3, k = k_K = 1, m = m_K = 0.00098, c = c_K = 1$ and $T = T_K = 0.003$, with initial values: $\nu(0) = \omega(0) = v(0) = 0.0045$. The exact expressions for the EKF gradients $\boldsymbol{F}(t), \boldsymbol{G}(t)$ and $\boldsymbol{H}(t)$ needed to construct the EKF, are given in App. E.9.

Fig. 5.12 *(left)* displays a cyclic MCG-like signal with a period of approximately 20 ms and a
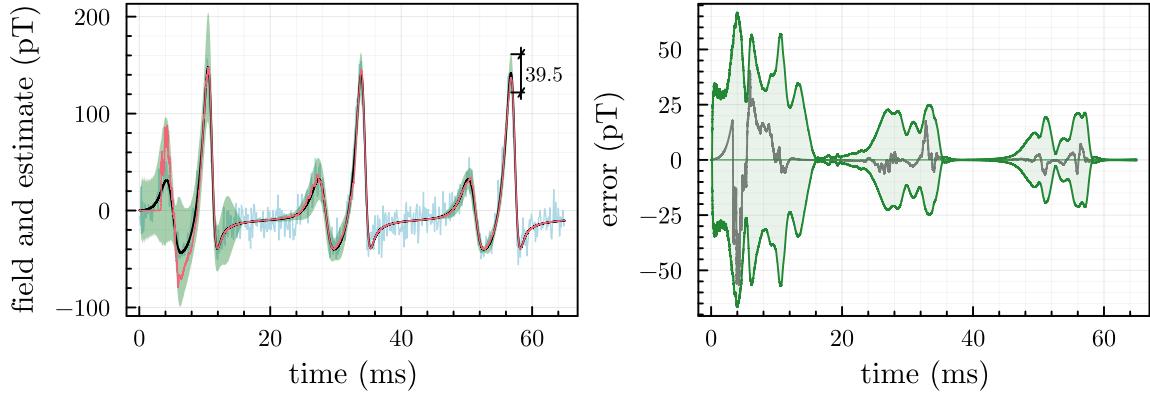
Figure 5.12: Tracking a MCG-like signal. Under the same conditions as in Fig. 5.11, the magnetometer tracks a MCG signal in the pT range [23, 25]. Although the atoms sense the noisy raw signal (blue), our goal is to recover the clean, noise-free waveform (black). The EKF, configured to expect a VdP–type signal [161, 162], produces an estimate (red) that converges to the clean waveform after one MCG cycle ($\approx$ 20 ms), as shown in the *left* subplot. In it, the green-shaded area represents the $\pm 3\sqrt{\mathbb{E}[\Delta^2\tilde{\omega}(t)]}$ confidence band, obtained after averaging over 1000 trajectories. Notably, at the R-wave in the third cycle, this band spans approximately $40\,\mathrm{pT}$, implying an average error of roughly $6.6\,\mathrm{pT}$. The subplot on the *right* displays the difference (gray) between the clean waveform and the EKF estimate, with the same confidence bounds clearly delineating the error magnitude (green).

field $B(t)$ ranging over $[-14.2\,\mathrm{pT}, 42.6\,\mathrm{pT}]$, compatible with human-heart fields [23, 25] (where we have used the Rb-87 ground state hyperfine gyromagnetic ratio, $2\pi \times 7\,\mathrm{GHz\,T^{-1}}$). After an initial transient, the EKF estimate (red) closely follows the true waveform (black), despite the raw signal that the magnetometer actually senses (blue) being contaminated by white noise with a strength of $q_\omega = 2.5 \times 10^5\,\mathrm{rad^2\,s^{-3}}$.

In Fig. 5.12 *(right)*, we further assess the tracking performance by plotting the difference between the true Larmor frequency and its EKF estimate (grey). This error is bounded by $\pm 3\sqrt{\mathbb{E}[\Delta^2\tilde{\omega}(t)]}$ (green), confirming that the estimation remains robust under noisy conditions.

5.4.3 MULTIPARTICLE ENTANGLEMENT: CONDITIONAL V.S. UNCONDITIONAL SPIN SQUEEZING

So far, our primary focus has been estimation, where conditional spin-squeezing, i.e. squeezing of the atomic state dependent on the measurement record $\rho_{(c)}(t) \equiv \rho(t|\mathbf{y}_{\leq t})$, has potential for enhancing precision beyond classical limits. However, instead of viewing the magnetometer solely as a sensing device, we can ask whether it can also function as a mechanism for preparing the system in a multipartite entangled state *independent* of the measurement trajectory.

In that case, evaluating the spin-squeezing parameter along a specific measurement trajectory

(i.e. *conditional*), is no longer sufficient. Instead, we should quantify the entanglement of the state independent of our observations (i.e. *unconditional*). While the *conditional* state of the system $\rho_{(c)}(t) \equiv \rho(t|\mathbf{y}_{\leq t})$ is understood as the one most closely describing the state given a particular measurement record $\mathbf{y}_{\leq t}$, an *unconditional* state $\rho(t)$ describes the system when we discard, or do not have access to, the measurement outcomes, what formally corresponds to averaging the conditional state over all the possible past measurement trajectories:

$$\rho(t) = \mathbb{E}_{p(\mathbf{y}_{\leq t})}\left[\rho_{(c)}(t)\right]. \tag{5.81}$$

Without feedback, the only effect experienced by the unconditional state $\rho(t)$ is a collective dephasing induced by the continuous measurement. For example, in the SME of Eq. (5.13) with $u(t) = 0$, averaging over measurement records simply introduces an extra $\mathcal{MD}[\hat{J}_y]$-term. However, once feedback is turned on, determining an effective master equation for the unconditional dynamics is far less straightforward. In general, feedback-driven evolution relies on Markovianity assumptions [104, 163], which are not satisfied in our LQR-based control scheme (see Sec. 5.3.3). Nevertheless, as in Markovian feedback scenarios where the system is unconditionally driven into a spin-squeezed state [104, 164], we demonstrate that this is also the case under the non-Markovian LQR control considered here.

Already in Fig. 5.9 we showed that an EKF combined with a LQR generates unconditional spin-squeezing, even in the presence of collective dephasing (red dots in Fig. 5.9(c)) or local dephasing (red dots in Fig. 5.9(d)). However, these results rely on the CoG approximation and the solution of the SDE system in Eq. (5.28) to compute the unconditional moments of angular momentum, and thus, the unconditional spin squeezing. Since the CoG approximation circumvents the need for the full density matrix, it does not allow direct visualization of how continuous measurement and feedback squeeze the atoms. To obtain the Wigner function and map it into the generalized Bloch sphere, a full numerical solution of the SME in Eq. (5.13) is required. For this reason, in Fig. 5.13 we present results from the exact SME (5.13) under purely collective decoherence ($N = 100$, $\kappa_c = 0.005$) to compare the EKF+LQR strategy *(left column)* with the naive field compensation *(right column)* in generating conditional as well as unconditional spin-squeezing. As in Fig. 5.9(c-d), Fig. 5.13 shows the average conditional spin-squeezing parameter (solid, blue) alongside the unconditional spin-squeezing of the average state (red, dots). Notably, only the EKF+LQR strategy generates an averaged state with unconditional squeezing that consistently breaches the classical value (dash-dot black horizontal line). This is further confirmed
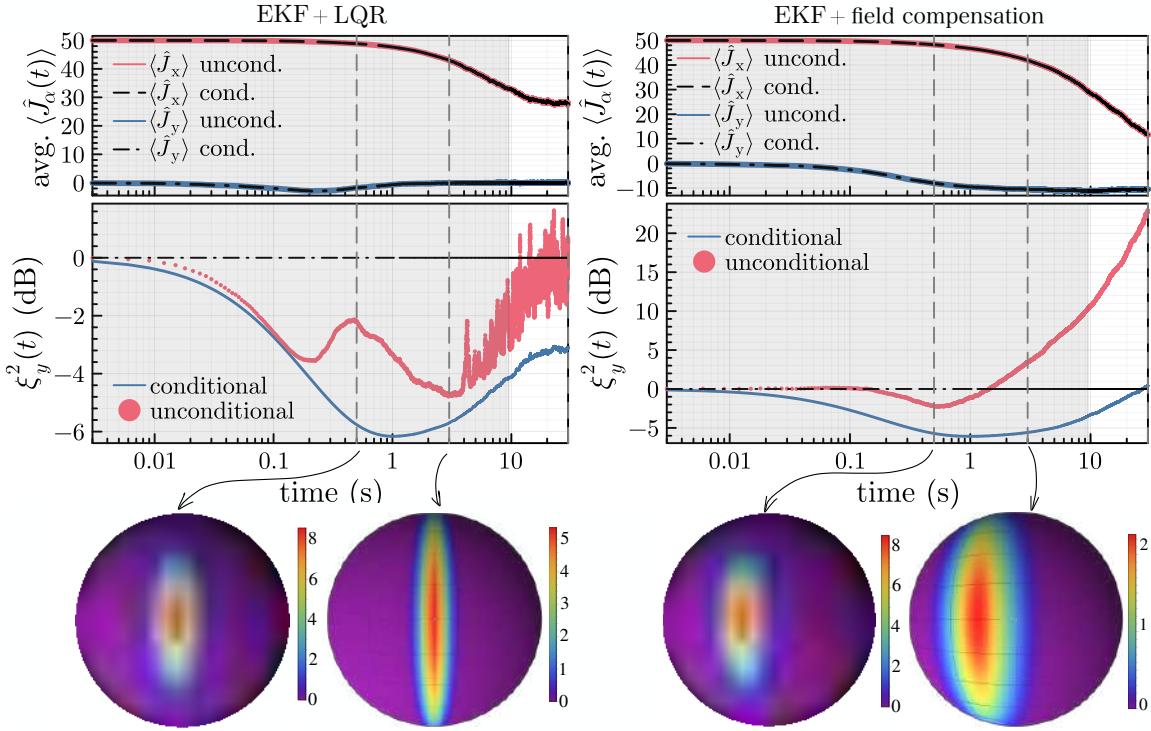
Figure 5.13: Conditional v.s. unconditional spin-squeezing. The exact ($N = 100$) spin-squeezing dynamics with collective decoherence is shown depending on the control strategy: LQR (*left column*) vs (naive) field compensation (*right column*). *Top row:* Subplots (*a*) and (*c*) depict evolution of angular momentum components in $\hat{J}_x$ (red) and $\hat{J}_y$ (blue) directions, in particular their conditional and unconditional means that consistently match. *Middle row:* Subplots (*b*) and (*d*) compare the dynamics of the average unconditional (in red) and conditional (in blue) spin-squeezing parameters (5.26), also verifying whether they surpass the classical value (horizontal black line). Vertical dashed grey lines mark the relevant times for which we explicitly plot the spherical Wigner functions (*bottom row*) representing the instantaneous unconditional state. Note that for the LQR control (*left*), even though the width along y of the distribution progressively narrows with time, the amplitude of the Wigner function also decays. The other parameters used in the SME (**??**) for simulations read: $\kappa_c = 0.005, \kappa_\ell = 0, M = 0.1$, $\omega = 1, \eta = 1$. The EKF is initialized with the mean $\tilde{\boldsymbol{x}}(0) = (N/2, 0, 0, N/4, N/4, 0, \mu_0)^{\mathrm{T}}$ and covariance $\Sigma(0) = \mathrm{Diag}[0, 0, 0, 0, 0, 0, \sigma_0^2]$ dictated by the initial CCS state, and the Gaussian prior distribution for $\omega \sim \mathcal{N}(\mu_0, \sigma_0^2)$. All results are obtained after averaging over $\nu = 500$ measurement trajectories, while ω-averaging is avoided by choosing the prior with $\sigma_0 = 0.5$ and $\mu_0 = \omega + \sigma_0 = 1.5$.

by the (spherical) Wigner distribution plots (snapshots at $t = 0.5, 3$), where the EKF+LQR strategy maintains steady squeezing along the y-direction over time. In contrast, under the naive field-compensation strategy, the Wigner distribution begins to lose its structure as early as $t = 3$ within the LQ regime.

As the ω-estimate of the EKF is initially set in Fig. 5.13 to $\tilde{\omega}(0) > \omega$, the control initially over-compensates for the Larmor precession and rotates the spin in the counter-clockwise direction when viewed along z. This is reflected in the spin components, with $\langle \hat{J}_y \rangle = \mathbb{E}[\langle \hat{J}_y \rangle_{(c)}]$ acquir-

ing negative values in both *top* plots, and in the corresponding Wigner function being shifted leftward at $\approx t = 0.1$. A similar effect would occur when choosing $\tilde{\omega}(0) < \omega$, in which case the control operation would initially undercompensate the Larmor precession, causing an initial clockwise spin rotation around z (Wigner function shifts to the right). Over time, the LQR control corrects the counter-rotation and stabilizes the spin along x *(left)*, whereas under naive field compensation *(right)*, stability is eventually lost.

6
Conclusions

In this thesis, we have studied how to track fluctuating or time-varying signals with optical atomic magnetometers at the quantum limit of precision. A key result of our work is the derivation of this quantum limit by establishing a lower bound on the BCRB [2, 67, 69]. Notably, the bound scales at best linearly with the number of atoms and sensing time, thus precluding any super-classical scaling [65, 66]. Also sometimes referred to as the classically-simulated limit, the quantum limit is independent of the choices of initial state, measurement, estimator, or measurement-based feedback, relying solely on the decoherence model and strength of field fluctuations. Thus, by developing a measurement, estimation, and control strategy that saturates this limit, we demonstrate that the sensing protocol is optimal.

But how can we develop an optimal sensing protocol? First, we need to be able to simulate the optical atomic magnetometer under typical experimental conditions using a quantum dynamical model that is scalable w.r.t. N. This model must rigorously incorporate both the measurement back-action and the environmental decoherence [57, 58, 63]. Second, we must select an estimator and controller that, when combined with the typical measurement (polarization spectroscopy [128, 126, 125, 120]) and initial state (coherent spin state) used in optical atomic

magnetometers, will enable us to attain the quantum limit.

The dynamics of an optical atomic magnetometer are rigorously described by a stochastic master equation [40, 55, 56], but it is not scalable w.r.t. N. To overcome this, we introduce the co-moving Gaussian approximation, which yields a nonlinear dynamical model that scales with N while accurately capturing the effects of measurement backaction and dephasing [65]. To validate this approximation, we compared it to the solution of the exact SME, which is numerically tractable for up to $N \approx 100$ atoms [65]. Building on this model, we have designed an integrated estimation and control scheme that couples an extended Kalman filter (EKF) with a linear quadratic regulator (LQR) [55, 65, 66].

We show that this sensing protocol achieves real-time tracking of constant and fluctuating fields at the quantum limit of precision [65, 66]. It also can track more complex waveforms such as those encountered in magneto-cardiography [66]. Furthermore, we also demonstrate that the continuous measurement and feedback generate inter-atomic entanglement, manifest as conditional spin-squeezing [65, 66]. Remarkably, even if the measurement data is not stored, the sensor is driven by the LQR feedback into an entangled state [65].

In this work, we have aimed to bridge the gap between mathematical formulations of continuous quantum measurement and estimation theory, and experiments in optical atomic magnetometry. Our results mark a significant step towards real-time quantum-limited metrology, even in the presence of decoherence, by leveraging continuous measurement, feedback, and spin-squeezing. To apply our framework to experiments with orientation-based magnetometers in the Bell-Bloom configuration [165, 58, 59], or alignment-based magnetometers [166, 167, 168], the dynamical model must be extended to incorporate spin-exchange collisions [140, 141] and other relevant sources of atomic decoherence [169]. These additions would require more general techniques to bound the Fisher information [170], beyond those used in this thesis.

A

Appendix for Chap. 1

A.1 Proof of $(dW(t))^{2+N} = 0$

Proposition A.1 (Proof of $(dW(t))^{2+N} = 0$)**.** *Consider a standard Wiener process* $W(t)$, *then for any* $N > 0$, *we have:*

$$(dW(t))^{2+N} = 0. \tag{A.1}$$

To interpret this rigorously, we need to write it in terms of the Itô stochastic integral. Specifically, let $f(t, X(t))$ *be a non-anticipating function of time and* $X(t)$, *a stochastic process. Then, the following holds*

$$\int_0^T f(t, X(t))\, dW(t)^{2+N} = \underset{n \to \infty}{\text{ms-lim}} \sum_{i=1}^n f(t_{i-1}, X(t_{i-1}))\, \Delta W_i^{2+N} = 0 \ \ for \ \ N > 0. \tag{A.2}$$

Proof. *To prove Eq. (A.2), we have to start with the definition of the mean-squared limit and*

show:

$$\lim_{n\to\infty} \mathbb{E}\left[(S_n - S)^2\right] = \lim_{n\to\infty} \mathbb{E}\left[S_n^2\right] = 0, \tag{A.3}$$

where $S = 0$ and

$$S_n = \sum_{i=1}^{n} f(t_{i\text{-}1}, X(t_{i\text{-}1}))\, \Delta W_i^{2+N}. \tag{A.4}$$

Thus, let us start by expanding S_n^2 using Eq. (1.80):

$$\lim_{n\to\infty} \mathbb{E}\left[S_n^2\right] = \lim_{n\to\infty} \mathbb{E}\left[\left(\sum_{i=1}^{n} f(t_{i\text{-}1}, X(t_{i\text{-}1}))\, \Delta W_i^{2+N}\right)^2\right]$$

$$= \lim_{n\to\infty} \mathbb{E}\left[\sum_{i=1}^{n} \left(f_{i\text{-}1} \Delta W_i^{2+N}\right)^2 + 2\sum_{i=1}^{n}\sum_{j=1}^{i-1} f_{i\text{-}1} f_{j\text{-}1} \Delta W_i^{2+N} \Delta W_j^{2+N}\right]$$

$$= \lim_{n\to\infty} \left\{\sum_{i=1}^{n} \mathbb{E}\left[f_{i\text{-}1}^2\right] \mathbb{E}\left[(\Delta W_i^2)^{2+N}\right] + 2\sum_{i=1}^{n}\sum_{j=1}^{i-1} \mathbb{E}\left[f_{i\text{-}1} f_{j\text{-}1}\right] \mathbb{E}\left[\Delta W_i^{2+N}\right] \mathrm{E}\left[\Delta W_j^{2+N}\right]\right\}.$$

Let us now recall that the higher order moments of a Gaussian random variable can be written as:

$$\mathbb{E}\left[\left(\Delta W_i^2\right)^{2+N}\right] = \Delta t_i^{2+N}, \tag{A.5}$$

$$\mathbb{E}\left[\Delta W_i^{2+N}\right] = \begin{cases} 0 & \text{for} \quad N \text{ odd}, \\ (N+1)!!\, \Delta t_i^{\frac{N}{2}+1} & \text{for} \quad N \text{ even}. \end{cases} \tag{A.6}$$

Note that for any value of N, the expectation values of ΔW_i^{2+N} and $(\Delta W_i^2)^{2+N}$ will be either zero or at least Δt_i^2. Therefore, $\forall N$, in the limit of $n \to \infty$, $\mathbb{E}[S_n^2]$ goes to zero:

$$\lim_{n\to\infty} \mathbb{E}\left[S_n^2\right] = 0, \tag{A.7}$$

and thus

$$\int_0^T f(t, X(t))\, dW(t)^{2+N} = 0 \;\; \text{for} \;\; N > 0, \tag{A.8}$$

s.t.

$$dW(t)^{2+N} = 0. \tag{A.9}$$

A.2 Proof of $dW(t)dt = 0$

Proposition A.2 (Proof of $dW(t)dt = 0$). *Consider a standard Wiener process $W(t)$, then we have:*

$$dW(t)dt = 0. \tag{A.10}$$

Similarly as before, we write this it in terms of the Itô stochastic integral:

$$\int_0^T f(t, X(t)) \, dW(t)dt = \underset{n \to \infty}{\text{ms-lim}} \sum_{i=1}^n f(t_{i-1}, X(t_{i-1})) \, \Delta W_i \, \Delta t_i = 0, \tag{A.11}$$

where $f(t, X(t))$ is a non-anticipating function of time and $X(t)$ is a general a stochastic process.

Proof. *To prove Eq. (A.11), we have to follow the same steps as in App. A.1. Since the stochastic integral S is defined as the mean-square limit of the Riemann sum S_n, we have to start from the definition of convergence in the mean-squared sense:*

$$\lim_{n \to \infty} \mathbb{E}\left[(S_n - S)^2\right] = 0, \tag{A.12}$$

and identify S and S_n. It is straightforward to see that for Eq. (A.11), $S = 0$ and

$$S_n = \sum_{i=1}^n f(t_{i-1}, X(t_{i-1})) \, \Delta W_i \, \Delta t_i. \tag{A.13}$$

Then, it follows that Eq. (A.12) is simply

$$\lim_{n \to \infty} \mathbb{E}\left[S_n^2\right] = 0, \tag{A.14}$$

which we can easily show by expanding its l.h.s. using Eq. (1.80):

$$\lim_{n\to\infty} \mathbb{E}\left[S_n^2\right] = \lim_{n\to\infty} \mathbb{E}\left[\left(\sum_{i=1}^n f(t_{i\text{-}1}, X(t_{i\text{-}1}))\,\Delta W_i\,\Delta t_i\right)^2\right]$$

$$= \lim_{n\to\infty} \mathbb{E}\left[\sum_{i=1}^n (f_{i\text{-}1}\Delta W_i\,\Delta t_i)^2 + 2\sum_{i=1}^n\sum_{j=1}^{i-1} f_{i\text{-}1}f_{j\text{-}1}\Delta W_i\,\Delta t_i\Delta W_j\,\Delta t_j\right]$$

$$= \lim_{n\to\infty}\left\{\sum_{i=1}^n \mathbb{E}\left[f_{i\text{-}1}^2\right]\mathbb{E}\left[(\Delta W_i^2)\right]\Delta t_i^2 + 2\sum_{i=1}^n\sum_{j=1}^{i-1}\mathbb{E}\left[f_{i\text{-}1}f_{j\text{-}1}\right]\mathbb{E}[\Delta W_i]\,\mathrm{E}\left[\Delta W_j\right]\Delta t_i\Delta t_j\right\}$$

$$= \lim_{n\to\infty}\sum_{i=1}^n \mathbb{E}\left[f_{i\text{-}1}^2\right]\Delta t_i^3 + 0 = 0. \tag{A.15}$$

Thus

$$\int_0^T f(t, X(t))\,\mathrm{d}W(t)\mathrm{d}t = 0, \tag{A.16}$$

s.t.

$$\mathrm{d}W(t)\mathrm{d}t = 0. \tag{A.17}$$

A.3 Solution to forced linear stochastic systems

Theorem A.3 (General solution of time-varying inhomogeneous linear stochastic differential equations). *Consider the following time-varying inhomogeneous stochastic differential equation*

$$\dot{\boldsymbol{x}}(t) = \boldsymbol{F}(t)\,\boldsymbol{x}(t) + \boldsymbol{B}(t)\,\boldsymbol{u}(t) + \boldsymbol{G}(t)\,\boldsymbol{w}(t) \tag{A.18}$$

where $\boldsymbol{F}(t)$ is the system matrix, $\boldsymbol{x}(t)$ is the state vector, $\boldsymbol{B}(t)$ is the control matrix, $\boldsymbol{u}(t)$ is the control vector, $\boldsymbol{G}(t)$ is the time-varying matrix modifying the stochastic term, and $\boldsymbol{w}(t)$ is the white noise with mean zero and covariance $\mathbb{E}\left[\boldsymbol{w}(t)\boldsymbol{w}(s)^\mathrm{T}\right] = \boldsymbol{Q}(t)\delta(t-s)$. The solution is

$$\boldsymbol{x}(t) = \boldsymbol{\Phi}(t, t_0)\,\boldsymbol{x}(t_0) + \int_{t_0}^t \boldsymbol{\Phi}(t, \tau)\boldsymbol{B}(\tau)\boldsymbol{u}(\tau)\mathrm{d}\tau + \int_{t_0}^t \boldsymbol{\Phi}(t, \tau)\boldsymbol{G}(\tau)\boldsymbol{w}(\tau)\mathrm{d}\tau, \tag{A.19}$$

where $\mathbf{\Phi}(t, t_0)$ is the state-transition matrix that satisfies

$$\frac{\mathrm{d}\mathbf{\Phi}(t, t_0)}{\mathrm{d}t} = \boldsymbol{F}(t)\mathbf{\Phi}(t, t_0), \qquad \mathbf{\Phi}(t_0, t_0) = \mathbb{I}. \tag{A.20}$$

and fulfills the following properties for all $t_0 \leq t \leq T$:

$$\mathbf{\Phi}(t_0, t) = \mathbf{\Phi}^{-1}(t, t_0), \tag{A.21}$$

$$\mathbf{\Phi}(T, t_0) = \mathbf{\Phi}(T, t)\mathbf{\Phi}(t, t_0). \tag{A.22}$$

Proof. *The standard approach to solving any forced linear system is to first derive the homogeneous solution. Namely, solving the homogeneous system of the form*

$$\frac{\mathrm{d}\boldsymbol{x}}{\mathrm{d}t} = \boldsymbol{F}(t)\,\boldsymbol{x}(t), \quad \boldsymbol{x}(t_0) = \boldsymbol{x}_0. \tag{A.23}$$

Let us assume then that the system's response in the absence of external forces is given by

$$\boldsymbol{x}(t) = \mathbf{\Phi}(t, t_0)\boldsymbol{x}(t_0), \tag{A.24}$$

where $\mathbf{\Phi}(t, t_0)$ is a so-called transition matrix. By substituting this solution into the homogeneous differential equation in (A.23), we obtain a differential equation for the transition matrix with its corresponding initial conditions:

$$\frac{\mathrm{d}\mathbf{\Phi}(t, t_0)}{\mathrm{d}t} = \boldsymbol{F}(t)\mathbf{\Phi}(t, t_0), \quad \mathbf{\Phi}(t_0, t_0) = \mathbb{I}. \tag{A.25}$$

Let us now consider the full dynamical model

$$\dot{\boldsymbol{x}}(t) = \boldsymbol{F}(t)\,\boldsymbol{x}(t) + \boldsymbol{B}(t)\,\boldsymbol{u}(t) + \boldsymbol{G}(t)\,\boldsymbol{w}(t), \tag{A.26}$$

for which, following the method of variation of parameters, we assume a solution of the form:

$$\boldsymbol{x}(t) = \mathbf{\Phi}(t, t_0)\boldsymbol{z}(t), \quad \boldsymbol{z}(t_0) = \boldsymbol{x}(t_0), \tag{A.27}$$

where $\boldsymbol{z}(t)$ is a vector of unknown functions that modifies the homogeneous solution to account

for the non-homogeneous terms in the full differential equation. If now we differentiate Eq. (A.27) we will get

$$\dot{\boldsymbol{x}}(t) = \boldsymbol{\Phi}(t, t_0) \dot{\boldsymbol{z}}(t) + \dot{\boldsymbol{\Phi}}(t, t_0) \boldsymbol{z}(t) = \boldsymbol{\Phi}(t, t_0) \dot{\boldsymbol{z}}(t) + \boldsymbol{F}(t) \boldsymbol{\Phi}(t, t_0) \boldsymbol{z}(t). \tag{A.28}$$

Let us now substitute Eq. (A.27) and in the lhs and rhs of Eq. (A.26):

$$\boldsymbol{\Phi}(t, t_0) \dot{\boldsymbol{z}}(t) + \boldsymbol{F}(t) \boldsymbol{\Phi}(t, t_0) \boldsymbol{z}(t) = \boldsymbol{F}(t) \, \boldsymbol{\Phi}(t, t_0) \boldsymbol{z}(t) + \boldsymbol{B}(t) \boldsymbol{u}(t) + \boldsymbol{G}(t) \boldsymbol{w}(t) \tag{A.29}$$

such that

$$\dot{\boldsymbol{z}}(t) = \boldsymbol{\Phi}^{-1}(t, t_0) \boldsymbol{B}(t) \boldsymbol{u}(t) + \boldsymbol{\Phi}^{-1}(t, t_0) \boldsymbol{G}(t) \boldsymbol{w}(t). \tag{A.30}$$

By now integrating this expression and recalling that $\boldsymbol{z}(t_0) = \boldsymbol{x}(t_0)$, we obtain

$$\int_{t_0}^{t} \dot{\boldsymbol{z}}(t) = \boldsymbol{z}(t) - \boldsymbol{z}(t_0) = \boldsymbol{z}(t) - \boldsymbol{x}(t_0) = \boldsymbol{\Phi}^{-1}(\tau, t_0) \boldsymbol{x}(t) - \boldsymbol{x}(t_0)$$

$$= \int_{t_0}^{t} \boldsymbol{\Phi}^{-1}(\tau, t_0) \boldsymbol{B}(\tau) \boldsymbol{u}(\tau) \, \mathrm{d}\tau + \int_{t_0}^{t} \boldsymbol{\Phi}^{-1}(\tau, t_0) \boldsymbol{G}(\tau) \boldsymbol{w}(\tau) \, \mathrm{d}\tau \tag{A.31}$$

such that

$$\boldsymbol{x}(t) = \boldsymbol{\Phi}(t, t_0) \boldsymbol{x}(t_0) + \boldsymbol{\Phi}(t, t_0) \int_{t_0}^{t} \boldsymbol{\Phi}^{-1}(\tau, t_0) \boldsymbol{B}(\tau) \boldsymbol{u}(\tau) \, \mathrm{d}\tau \tag{A.32}$$

$$+ \boldsymbol{\Phi}(t, t_0) \int_{t_0}^{t} \boldsymbol{\Phi}^{-1}(\tau, t_0) \boldsymbol{G}(\tau) \boldsymbol{w}(\tau) \, \mathrm{d}\tau. \tag{A.33}$$

By now using Eq. (A.22), and first multiplying from the right $\boldsymbol{\Phi}^{-1}(\tau, t_0)$ and then from the left $\boldsymbol{\Phi}^{-1}(t, t_0)$:

$$\boldsymbol{\Phi}^{-1}(t, t_0) \boldsymbol{\Phi}(t, t_0) \boldsymbol{\Phi}^{-1}(\tau, t_0) = \boldsymbol{\Phi}^{-1}(t, t_0) \boldsymbol{\Phi}(t, \tau) \boldsymbol{\Phi}(\tau, t_0) \boldsymbol{\Phi}^{-1}(\tau, t_0) \tag{A.34}$$

we obtain

$$\boldsymbol{\Phi}^{-1}(\tau, t_0) = \boldsymbol{\Phi}^{-1}(t, t_0) \boldsymbol{\Phi}(t, \tau) \tag{A.35}$$

which when substituted into Eq. (A.31) yields the final expression of Eq. (A.19):

$$\boldsymbol{x}(t) = \boldsymbol{\Phi}(t, t_0)\,\boldsymbol{x}(t_0) + \int_{t_0}^{t} \boldsymbol{\Phi}(t, \tau)\boldsymbol{B}(\tau)\boldsymbol{u}(\tau)\mathrm{d}\tau + \int_{t_0}^{t} \boldsymbol{\Phi}(t, \tau)\boldsymbol{G}(\tau)\boldsymbol{w}(\tau)\mathrm{d}\tau. \qquad (A.36)$$

A.4 (SOME) ANGULAR MOMENTUM PROPERTIES

Proposition A.4 (Generation of angular momentum eigenstates via the raising operator). *The angular momentum eigenstate $|j, m\rangle$ where j is the total angular momentum quantum number and m the magnetic quantum number, can be generated from applying the raising operator $\hat{J}_+$ to the lowest eigenstate $|j, -j\rangle$ as follows [78]:*

$$|j, m\rangle = \frac{1}{(m+j)!}\binom{2j}{m+j}^{-1/2}\hat{J}_+^{m+j}|j, -j\rangle, \qquad (A.37)$$

where $\hat{J}_+$ is the angular momentum raising operator, defined as $\hat{J}_+ = \hat{J}_x + i\hat{J}_y$.

Proof. *Recall how $\hat{J}_+$ acts on $|j, m\rangle$:*

$$\hat{J}_+|j, m\rangle = \sqrt{j(j+1) - m(m+1)}\,|j, m+1\rangle. \qquad (A.38)$$

Note that the factor inside the square root can be rewritten as:

$$j(j+1) - m(m+1) = (j+m+1)(j-m). \qquad (A.39)$$

Therefore, applying the operator $\hat{J}_+$, $j + m$ times, in order to raise the state $|j, -j\rangle$ to $|j, m\rangle$, as sketched in Table A.1, yields:

$$\hat{J}_+^{m+j}|j, -j\rangle = \sqrt{(j+m)!\frac{(2j)!}{(j-m)!}}\,|j, m\rangle, \qquad (A.40)$$

which can be rewritten using a binomial notation, since

$$(j+m)! \binom{2j}{m+j}^{1/2} = (j+m)! \sqrt{\frac{(2j)!}{(j+m)!(j-m)!}} = \sqrt{(j+m)! \frac{(2j)!}{(j-m)!}}. \quad (A.41)$$

Hence,

$$|j, m\rangle = \frac{1}{(j+m)!} \binom{2j}{m+j}^{-1/2} \hat{J}_+^{m+j} |j, -j\rangle. \quad (A.42)$$

(a)

$|j,j\rangle$

$|j, m\rangle$

$\hat{J}_+^{j+m}$

$|j, -j\rangle$
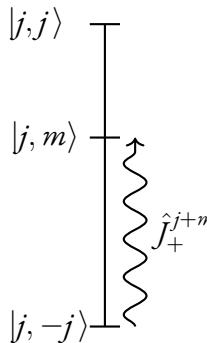
(b)

m	$(j+m+1)(j-m)$
$-j$	$1 \cdot 2j$
$-j+1$	$2(2j-1)$
$-j+2$	$3(2j-2)$
$\vdots$	$\vdots$
$m-1$	$(j+m)(j-m+1)$

Table A.1: Raising and lowering operators for the angular momentum. (a) Scheme depicting the transition from state $|j, -j\rangle$ to $|j, m\rangle$ by applying the raising operator $\hat{J}_+$ $j + m$ times. (b) Table summarizing the factors that appear every time we apply the $\hat{J}_+$ operator to state $|j, -j\rangle$ and subsequent states.

A.4.1 COHERENT SPIN STATE MEANS AND VARIANCES

A.4.1.1 MEANS

First, we will compute the term $\langle \hat{J}_x \rangle_{\mathrm{CSS}_x}$. To do so, let us write $\hat{J}_x$ in terms of the ladder operators, $\hat{J}_x = \frac{1}{2}(\hat{J}_+ + \hat{J}_-)$, such that:

$$\langle \hat{J}_x \rangle_{\mathrm{CSS}_x} = \langle \eta | \hat{J}_x | \eta \rangle = \frac{1}{2} \left(\langle \eta | \hat{J}_+ | \eta \rangle + \langle \eta | \hat{J}_- | \eta \rangle \right). \quad (A.43)$$

Then, the expected value of $\hat{J}_+$ w.r.t a CSS can be computed to be $N/4$:

$$\langle \eta | \hat{J}_+ | \eta \rangle =$$

$$= \frac{1}{2^{N+1}} \sum_{n,m=-N/2}^{N/2} \binom{N}{\frac{N}{2}+n}^{1/2} \binom{N}{\frac{N}{2}+m}^{1/2} \sqrt{\frac{N}{2}\left(\frac{N}{2}+1\right) - m(m+1)} \left\langle n, \frac{N}{2} \middle| \frac{N}{2}, m+1 \right\rangle$$

$$= \frac{1}{2^{N+1}} \sum_{m=-N/2}^{N/2} \binom{N}{\frac{N}{2}+m+1}^{1/2} \binom{N}{\frac{N}{2}+m}^{1/2} \sqrt{\frac{N}{2}\left(\frac{N}{2}+1\right) - m(m+1)}$$

$$\stackrel{A.1}{=} \frac{1}{2^{N+1}} \sum_{m=-N/2}^{N/2} \left(\frac{N}{2}-m\right) \binom{N}{\frac{N}{2}+m} \stackrel{A.2 \& A.4}{=} \frac{1}{2^{N+1}} N 2^{N-1} = \frac{N}{4}, \tag{A.44}$$

and similarly for $\hat{J}_-$:

$$\langle \eta | \hat{J}_- | \eta \rangle =$$

$$= \frac{1}{2^{N+1}} \sum_{n,m=-N/2}^{N/2} \binom{N}{\frac{N}{2}+n}^{1/2} \binom{N}{\frac{N}{2}+m}^{1/2} \sqrt{\frac{N}{2}\left(\frac{N}{2}+1\right) - m(m-1)} \left\langle n, \frac{N}{2} \middle| \frac{N}{2}, m-1 \right\rangle$$

$$= \frac{1}{2^{N+1}} \sum_{m=-N/2}^{N/2} \binom{N}{\frac{N}{2}+m-1}^{1/2} \binom{N}{\frac{N}{2}+m}^{1/2} \sqrt{\frac{N}{2}\left(\frac{N}{2}+1\right) - m(m-1)}$$

$$\stackrel{A.1}{=} \frac{1}{2^{N+1}} \sum_{m=-N/2}^{N/2} \left(\frac{N}{2}+m\right) \binom{N}{\frac{N}{2}-m} \stackrel{A.2 \& A.4}{=} \frac{1}{2^{N+1}} N 2^{N-1} = \frac{N}{4}. \tag{A.45}$$

Hence,

$$\langle \hat{J}_x \rangle_{\text{CSS}_x} = \frac{N}{2}. \tag{A.46}$$

The other terms $\langle \hat{J}_y \rangle_{\text{CSS}_x}$ and $\langle \hat{J}_z \rangle_{\text{CSS}_x}$ are straightforward:

$$\langle \hat{J}_y \rangle_{\text{CSS}_x} = \frac{1}{2i} \left(\langle \eta | \hat{J}_+ | \eta \rangle - \langle \eta | \hat{J}_- | \eta \rangle \right) = 0, \tag{A.47}$$

and

$$\langle \hat{J}_z \rangle_{CSS_x} = \langle \eta | \hat{J}_z | \eta \rangle = \frac{1}{2^N} \sum_{n=-N/2}^{N/2} \sum_{m=-N/2}^{N/2} \binom{N}{\frac{N}{2}+n}^{1/2} \binom{N}{\frac{N}{2}+m}^{1/2} m \left\langle n, \frac{N}{2} \Big| \frac{N}{2}, m \right\rangle$$

$$= \frac{1}{2^N} \sum_{m=-N/2}^{N/2} m \binom{N}{\frac{N}{2}+m} \stackrel{A.4}{=} 0. \tag{A.48}$$

A.4.1.2   VARIANCES

Let us start with $\langle \Delta^2 \hat{J}_x \rangle_{CSS_x} = \langle \eta | \hat{J}_x^2 | \eta \rangle - \langle \eta | \hat{J}_x | \eta \rangle^2$. Note that,

$$\langle \Delta^2 \hat{J}_x \rangle_{CSS_x} = \langle \eta | \hat{J}_x^2 | \eta \rangle - \langle \eta | \hat{J}_x | \eta \rangle^2 = \langle \eta | \hat{J}_x^2 | \eta \rangle - \frac{N^2}{4}. \tag{A.49}$$

Thus, it only remains to compute the term $\langle \eta | \hat{J}_x^2 | \eta \rangle$, i.e.,

$$\langle \eta | \hat{J}_x^2 | \eta \rangle = \frac{1}{4} \left(\langle \eta | \hat{J}_+ \hat{J}_+ | \eta \rangle + \langle \eta | \hat{J}_+ \hat{J}_- | \eta \rangle + \langle \eta | \hat{J}_- \hat{J}_+ | \eta \rangle + \langle \eta | \hat{J}_- \hat{J}_- | \eta \rangle \right). \tag{A.50}$$

To calculate the variances of a CSS polarized along the x-axis, the first term, when evaluated for a CSS as defined in Eq. (1.222), reads as:

$$\langle \eta | \hat{J}_+ \hat{J}_+ | \eta \rangle = \frac{1}{2^N} \sum_{n,m=-N/2}^{N/2} \binom{N}{\frac{N}{2}+n}^{1/2} \binom{N}{\frac{N}{2}+m}^{1/2} \left\langle n, \frac{N}{2} \Big| \hat{J}_+ \hat{J}_+ \Big| \frac{N}{2}, m \right\rangle. \tag{A.51}$$

Note that $\hat{J}_+$ applied to the bra representing the angular state with magnetic quantum number n, can be rewritten as:

$$\left\langle n, \frac{N}{2} \Big| \hat{J}_+ = \left\langle n, \frac{N}{2} \Big| \hat{J}_-^\dagger = \left(\hat{J}_- \Big| \frac{N}{2}, n \right\rangle \right)^\dagger = \left(\sqrt{\frac{N}{2} \left(\frac{N}{2} + 1 \right) - n(n-1)} \Big| \frac{N}{2}, n-1 \right\rangle \right)^\dagger$$

$$= \left\langle n-1, \frac{N}{2} \Big| \sqrt{\frac{N}{2} \left(\frac{N}{2} + 1 \right) - n(n-1)}. \tag{A.52}$$

Therefore,

$$
\langle \eta | \hat{J}_+ \hat{J}_+ | \eta \rangle = \frac{1}{2^N} \sum_{n,m=-N/2}^{N/2} \binom{N}{\frac{N}{2}+n}^{1/2} \binom{N}{\frac{N}{2}+m}^{1/2} \left\langle n, \frac{N}{2} \middle| \hat{J}_+ \hat{J}_+ \middle| \frac{N}{2}, m \right\rangle =
$$

$$
= \frac{1}{2^N} \sum_{n,m=-N/2}^{N/2} \binom{N}{\frac{N}{2}+n}^{1/2} \binom{N}{\frac{N}{2}+m}^{1/2} \sqrt{\frac{N}{2}\left(\frac{N}{2}+1\right) - n(n-1)} \times
$$

$$
\times \sqrt{\frac{N}{2}\left(\frac{N}{2}+1\right) - m(m+1)} \left\langle n-1, \frac{N}{2} \middle| \frac{N}{2}, m+1 \right\rangle =
$$

$$
= \frac{1}{2^N} \sum_{m=-N/2}^{N/2} \binom{N}{\frac{N}{2}+m+2}^{1/2} \binom{N}{\frac{N}{2}+m}^{1/2} \sqrt{\frac{N}{2}\left(\frac{N}{2}+1\right) - (m+2)(m+1)} \times
$$

$$
\times \sqrt{\frac{N}{2}\left(\frac{N}{2}+1\right) - m(m+1)} = \frac{1}{2^N} \sum_{m=-N/2}^{N/2} \frac{N!}{\left(\frac{N}{2}+m\right)!\left(\frac{N}{2}-m-2\right)!} =
$$

$$
= \frac{1}{2^N} \sum_{m=-N/2}^{N/2} \left(\frac{N}{2}-m\right)\left(\frac{N}{2}-m-1\right)\binom{N}{\frac{N}{2}+m} =
$$

$$
= \frac{1}{2^{N+2}} N(N-2) \sum_{m=-N/2}^{N/2} \binom{N}{\frac{N}{2}+m} + \frac{1}{2^N} \sum_{m=-N/2}^{N/2} m^2 \binom{N}{\frac{N}{2}+m} =
$$

$$
= N(N-2)2^{-2} + N2^{-2} = N(N-2+1)2^{-2} = \frac{N}{4}(N-1). \qquad (A.53)
$$

The rest of the terms are:

$$
\langle \eta | \hat{J}_+ \hat{J}_- | \eta \rangle = \frac{1}{2^N} \sum_{n=-N/2}^{N/2} \binom{N}{\frac{N}{2}+n}^{1/2} \binom{N}{\frac{N}{2}+m}^{1/2} \sqrt{\frac{N}{2}\left(\frac{N}{2}+1\right) - n(n-1)} \times
$$

$$
\times \sqrt{\frac{N}{2}\left(\frac{N}{2}+1\right) - m(m-1)} \left\langle n-1, \frac{N}{2} \middle| \frac{N}{2}, m-1 \right\rangle =
$$

$$
= \frac{1}{2^N} \sum_{m=-N/2}^{N/2} \binom{N}{\frac{N}{2}+m} \left(\frac{N}{2}\left(\frac{N}{2}+1\right) - m(m-1)\right) =
$$

$$
= \frac{1}{2^N} N(N+2)\, 2^{N-2} - \frac{1}{2^N} N 2^{N-2} = \frac{N}{4}(N+1), \qquad (A.54)
$$

which yields the same value as:

$$\langle\eta|\hat{J}_-\hat{J}_+|\eta\rangle = \frac{1}{2^N}\sum_{n=-N/2}^{N/2}\binom{N}{\frac{N}{2}+n}^{1/2}\binom{N}{\frac{N}{2}+m}^{1/2}\sqrt{\frac{N}{2}\left(\frac{N}{2}+1\right)-n(n+1)}\times$$

$$\times\sqrt{\frac{N}{2}\left(\frac{N}{2}+1\right)-m(m+1)}\left\langle n+1,\frac{N}{2}\bigg|\frac{N}{2},m+1\right\rangle =$$

$$=\frac{1}{2^N}\sum_{m=-N/2}^{N/2}\binom{N}{\frac{N}{2}+m}\left(\frac{N}{2}\left(\frac{N}{2}+1\right)-m(m+1)\right)=$$

$$=\frac{N}{4}(N+1),\tag{A.55}$$

with the expected value of $\hat{J}_-^2$ being:

$$\langle\eta|\hat{J}_-\hat{J}_-|\eta\rangle = \frac{1}{2^N}\sum_{n=-N/2}^{N/2}\binom{N}{\frac{N}{2}+n}^{1/2}\binom{N}{\frac{N}{2}+m}^{1/2}\sqrt{\frac{N}{2}\left(\frac{N}{2}+1\right)-n(n+1)}\times$$

$$\times\sqrt{\frac{N}{2}\left(\frac{N}{2}+1\right)-m(m-1)}\left\langle n+1,\frac{N}{2}\bigg|\frac{N}{2},m-1\right\rangle =$$

$$=\frac{1}{2^N}\sum_{m=-N/2}^{N/2}\binom{N}{\frac{N}{2}+m-2}^{1/2}\binom{N}{\frac{N}{2}+m}^{1/2}\sqrt{\frac{N}{2}\left(\frac{N}{2}+1\right)-(m-2)(m-1)}\times$$

$$\times\sqrt{\frac{N}{2}\left(\frac{N}{2}+1\right)-m(m-1)}=\frac{1}{2^N}\sum_{m=-N/2}^{N/2}\frac{N!}{\left(\frac{N}{2}+m-2\right)!\left(\frac{N}{2}-m\right)!}=$$

$$=\frac{1}{2^N}\sum_{m=-N/2}^{N/2}\left(\frac{N}{2}+m\right)\left(\frac{N}{2}+m-1\right)\binom{N}{\frac{N}{2}-m}=$$

$$=\frac{1}{2^{N+2}}N(N-2)\sum_{m=-N/2}^{N/2}\binom{N}{\frac{N}{2}+m}+\frac{1}{2^N}\sum_{m=-N/2}^{N/2}m^2\binom{N}{\frac{N}{2}+m}=$$

$$=N(N-2)2^{-2}+N2^{-2}=N(N-2+1)2^{-2}=\frac{N}{4}(N-1).\tag{A.56}$$

Hence, the second moment of $\hat{J}_x$ w.r.t. $|\eta\rangle$ is:

$$\langle\eta|\hat{J}_x^2|\eta\rangle = \frac{1}{4}\left(\frac{N}{2}(N+1) + \frac{N}{2}(N-1)\right) = \frac{N^2}{4}. \tag{A.57}$$

We know from previous calculations that $\langle\eta|\hat{J}_x|\eta\rangle = \frac{N}{2}$. Hence,

$$\langle\Delta^2\hat{J}_x\rangle_{\text{CSS}_x} = \langle\eta|\hat{J}_x^2|\eta\rangle - \langle\eta|\hat{J}_x|\eta\rangle^2 = \frac{N^2}{4} - \left(\frac{N}{2}\right)^2 = 0. \tag{A.58}$$

The variances for the rest of the components are,

$$\begin{aligned}
\langle\Delta^2\hat{J}_y\rangle_{\text{CSS}_x} &= \langle\eta|\hat{J}_y^2|\eta\rangle - \langle\eta|\hat{J}_y|\eta\rangle^2 = \langle\eta|\hat{J}_y^2|\eta\rangle = -\frac{1}{4}\langle\eta|(\hat{J}_+ - \hat{J}_-)(\hat{J}_+ - \hat{J}_-)|\eta\rangle \\
&= -\frac{1}{4}\left(\langle\eta|\hat{J}_+\hat{J}_+|\eta\rangle - \langle\eta|\hat{J}_+\hat{J}_-|\eta\rangle - \langle\eta|\hat{J}_-\hat{J}_+|\eta\rangle + \langle\eta|\hat{J}_-\hat{J}_-|\eta\rangle\right) \\
&= -\frac{1}{4}\left(\frac{N}{4}(N-1) - \frac{N}{4}(N+1) - \frac{N}{4}(N+1) + \frac{N}{4}(N-1)\right) = \frac{N}{4},
\end{aligned} \tag{A.59}$$

and,

$$\begin{aligned}
\langle\Delta^2\hat{J}_z\rangle_{\text{CSS}_x} &= \frac{1}{2^N}\sum_{n=-N/2}^{N/2}\sum_{m=-N/2}^{N/2}\binom{N}{\frac{N}{2}+n}^{1/2}\binom{N}{\frac{N}{2}+m}^{1/2}\left\langle n,\frac{N}{2}\middle|\hat{J}_z^2\middle|m,\frac{N}{2}\right\rangle = \\
&= \frac{1}{2^N}\sum_{n=-N/2}^{N/2}\sum_{m=-N/2}^{N/2}\binom{N}{\frac{N}{2}+n}^{1/2}\binom{N}{\frac{N}{2}+m}^{1/2}n\,m\left\langle n,\frac{N}{2}\middle|m,\frac{N}{2}\right\rangle = \\
&= \frac{1}{2^N}\sum_{m=-N/2}^{N/2}m^2\binom{N}{\frac{N}{2}+m} = \frac{1}{2^N}N2^{N-2} = \frac{N}{4}.
\end{aligned} \tag{A.60}$$

A.5   Binomial identities

Theorem A.5 (Binomial theorem). *The expansion of the binomial $x + y$ to the power n, where n is any nonnegative integer, is given by*

$$(x+y)^n = \sum_{k=0}^{n}\binom{n}{k}x^k y^{n-k}, \tag{A.61}$$

which in the case of y = 1, reads as

$$(1+x)^n = \sum_{k=0}^{n} \binom{n}{k} x^k. \tag{A.62}$$

Identity A.1.

$$\binom{N}{\frac{N}{2}+m\pm1}^{1/2} \binom{N}{\frac{N}{2}+m}^{1/2} = \frac{\left(\frac{N}{2}\mp m\right)}{\sqrt{\frac{N}{2}\left(\frac{N}{2}+1\right)-m(m\pm1)}} \binom{N}{\frac{N}{2}\pm m} \tag{A.63}$$

Proof. *It is straightforward to show using factorial notation:*

$$\binom{N}{\frac{N}{2}+m\pm1}^{1/2} \binom{N}{\frac{N}{2}+m}^{1/2} =$$

$$= \left(\frac{N!}{\left(\frac{N}{2}+m\pm1\right)!\left(\frac{N}{2}-m\mp1\right)!}\right)^{1/2} \left(\frac{N!}{\left(\frac{N}{2}+m\right)!\left(\frac{N}{2}-m\right)!}\right)^{1/2}$$

$$= \frac{1}{\sqrt{\left(\frac{N}{2}\pm m+1\right)\left(\frac{N}{2}\mp m\right)}} \frac{N!}{\left(\frac{N}{2}\pm m\right)!\left(\frac{N}{2}\mp m-1\right)!}$$

$$= \frac{\left(\frac{N}{2}\mp m\right)}{\sqrt{\frac{N}{2}\left(\frac{N}{2}+1\right)-m(m\pm1)}} \binom{N}{\frac{N}{2}\pm m}$$

Identity A.2. *Follows straightforwardly from the binomial formula of Eq. (A.62) by setting $x = 1$ and changing the summand index k to $\frac{N}{2}+m$:*

$$\sum_{m=-N/2}^{N/2} \binom{N}{\frac{N}{2}+m} = 2^N. \tag{A.64}$$

Identity A.3.

$$\sum_{m=-N/2}^{N/2} \binom{N}{\frac{N}{2}\pm m}\left(\frac{N}{2}\pm m\right) = N2^{N-1}. \tag{A.65}$$

Proof. *The closed forms for the above sums can be found using the derivative of the binomial formula given in Eq. (A.62). Namely,*

$$\frac{\mathrm{d}}{\mathrm{d}x}(x+1)^N = N(x+1)^{N-1} = \frac{\mathrm{d}}{\mathrm{d}x}\sum_{k=0}^{N}\binom{N}{k}x^k = \sum_{k=0}^{N}\binom{N}{k}k\,x^{k-1}. \tag{A.66}$$

Then, Eq. (A.65) is obtained by setting $x = 1$ and changing the summation limits.

Identity A.4.

$$\sum_{m=-N/2}^{N/2} m\binom{N}{\frac{N}{2}+m} = N2^{N-1} - N2^{N-1} = 0, \tag{A.67}$$

which follows trivially from Identity A.2 and Identity A.3.

Identity A.5.

$$\sum_{k=-N/2}^{N/2} m^2\binom{N}{\frac{N}{2}+m} = N2^{N-2}. \tag{A.68}$$

Proof. *To find the close form expression of Identity A.5, we take the second order derivative of the binomial of Eq. (A.62) multiplied by x:*

$$x\,(x+1)^N = \sum_{k=0}^{N}\binom{N}{k}x^{k+1}. \tag{A.69}$$

To do so, we first compute the second derivative of the l.h.s. is

$$\frac{\mathrm{d}^2}{\mathrm{d}x^2}\left(x(x+1)^N\right) = 2N(x+1)^{N-1} + Nx\,(N-1)(x+1)^{N-2}, \tag{A.70}$$

and of the r.h.s.:

$$\frac{\mathrm{d}^2}{\mathrm{d}x^2}\sum_{k=0}^{N}\binom{N}{k}x^{k+1} = \sum_{k=0}^{N}\binom{N}{k}(k+1)k\,x^{k-1}. \tag{A.71}$$

If we now set $x = 1$ and change the summation limits by redefining the index k as

$k = \frac{N}{2} + m$, then, Eq. (A.69) can be written as

$$\sum_{m=-N/2}^{N/2} \left(\frac{N}{2} + m + 1\right) \left(\frac{N}{2} + m\right) \binom{N}{\frac{N}{2} + m} = N2^N + N(N-1)2^{N-2}. \qquad (A.72)$$

If now we massage the l.h.s. a bit more:

$$\sum_{m=-N/2}^{N/2} \left(\frac{N}{2} + m + 1\right) \left(\frac{N}{2} + m\right) \binom{N}{\frac{N}{2} + m} =$$

$$= \sum_{m=-N/2}^{N/2} \left(\frac{N}{2}\left(\frac{N}{2} + 1\right) + m^2 + m(N+1)\right) \binom{N}{\frac{N}{2} + m}$$

$$\underset{A.4}{=} N(N+2)2^{-2} \sum_{m=-N/2}^{N/2} \binom{N}{\frac{N}{2} + m} + \sum_{m=-N/2}^{N/2} m^2 \binom{N}{\frac{N}{2} + m}$$

$$\underset{A.2}{=} N(N+2)2^{N-2} + \sum_{m=-N/2}^{N/2} m^2 \binom{N}{\frac{N}{2} + m}, \qquad (A.73)$$

we can finally rewrite Eq. (A.72) as

$$N(N+2)2^{N-2} + \sum_{m=-N/2}^{N/2} m^2 \binom{N}{\frac{N}{2} + m} = N2^N + N(N-1)2^{N-2}, \qquad (A.74)$$

and thus, reach the final form of Eq. (A.68):

$$\sum_{m=-N/2}^{N/2} m^2 \binom{N}{\frac{N}{2} + m} = N2^N + N(N-1)2^{N-2} - N(N+2)2^{N-2}$$

$$= N\left(4 + (N-1) - (N+2)\right)2^{N-2} = N2^{N-2}. \qquad (A.75)$$

B

Appendix for Chap. 2

B.1 Gaussian properties I

Property B.1 (Inversion of a 2x2 Block matrix). *Consider a Block matrix partitioned into four blocks where A and D are square blocks of arbitrary size. Then,*

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix}^{-1} = \begin{pmatrix} A^{-1} + A^{-1}B(D - CA^{-1}B)^{-1}CA^{-1} & -A^{-1}B(D - CA^{-1}B)^{-1} \\ -(D - CA^{-1}B)^{-1}CA^{-1} & (D - CA^{-1}B)^{-1} \end{pmatrix}. \tag{B.1}$$

Property B.2 (Determinant of a 2x2 Block matrix). *Consider the following Block matrix*

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix}. \tag{B.2}$$

If A is invertible, then the determinant is given by:

$$\det\left[\begin{pmatrix} A & B \\ C & D \end{pmatrix} \right] = \det[A] \det[D - CA^{-1}B] \tag{B.3}$$

Lemma B.1 (The joint distribution of Gaussian variables). *Let $\boldsymbol{x} \in \mathbb{R}^n$ and $\boldsymbol{y} \in \mathbb{R}^m$ be two Gaussian random variables such that:*

$$\boldsymbol{x} \sim p(\boldsymbol{x}) = \mathcal{N}(\boldsymbol{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}), \tag{B.4}$$

$$\boldsymbol{y}|\boldsymbol{x} \sim p(\boldsymbol{y}|\boldsymbol{x}) = \mathcal{N}(\boldsymbol{y}|\boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{u}, \boldsymbol{G}\boldsymbol{Q}\boldsymbol{G}^{\mathrm{T}}) \tag{B.5}$$

where $\boldsymbol{A}$ is a transition matrix, $\boldsymbol{B}$ is the control input matrix, $\boldsymbol{u}$ is an external control, and $\boldsymbol{\Sigma}$ and $\boldsymbol{Q}$ are positive-definite covariance matrices. Then, the joint distribution of $(\boldsymbol{x}, \boldsymbol{y})$ is a multivariate Gaussian given by:

$$\begin{pmatrix} \boldsymbol{x} \\ \boldsymbol{y} \end{pmatrix} \sim p(\boldsymbol{x}, \boldsymbol{y}) = \mathcal{N}\left(\begin{pmatrix} \boldsymbol{x} \\ \boldsymbol{y} \end{pmatrix} \middle| \begin{pmatrix} \boldsymbol{\mu} \\ \boldsymbol{A}\boldsymbol{\mu} + \boldsymbol{B}\boldsymbol{u} \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Sigma} & \boldsymbol{\Sigma}\boldsymbol{A}^{\mathrm{T}} \\ \boldsymbol{A}\boldsymbol{\Sigma} & \boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A}^{\mathrm{T}} + \boldsymbol{G}\boldsymbol{Q}\boldsymbol{G}^{\mathrm{T}} \end{pmatrix}\right) \tag{B.6}$$

Proof. *The joint distribution of $(\boldsymbol{x}, \boldsymbol{y})$ can be expressed using the product rule (Property 1.2):*

$$p(\boldsymbol{x}, \boldsymbol{y}) = p(\boldsymbol{y}|\boldsymbol{x})p(\boldsymbol{x}) = \mathcal{N}(\boldsymbol{y}|\boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{u}, \boldsymbol{G}\boldsymbol{Q}\boldsymbol{G}^{\mathrm{T}})\, \mathcal{N}(\boldsymbol{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}). \tag{B.7}$$

Let us now work backwards by assuming that Eq. (B.6) holds and see if we can split the joint distribution into the following two explicit forms of the Gaussian distributions of $\boldsymbol{x}$ and $\boldsymbol{y}|\boldsymbol{x}$:

$$\mathcal{N}(\boldsymbol{y}|\boldsymbol{A}\boldsymbol{x}, \boldsymbol{G}\boldsymbol{Q}\boldsymbol{G}^{\mathrm{T}}) = \frac{1}{\sqrt{(2\pi)^m |\boldsymbol{G}\boldsymbol{Q}\boldsymbol{G}^{\mathrm{T}}|}} \exp\left(-\frac{1}{2}(\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x} - \boldsymbol{B}\boldsymbol{u})^{\mathrm{T}}(\boldsymbol{G}\boldsymbol{Q}\boldsymbol{G}^{\mathrm{T}})^{-1}(\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x} - \boldsymbol{B}\boldsymbol{u})\right), \tag{B.8}$$

$$\mathcal{N}(\boldsymbol{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{\sqrt{(2\pi)^n |\boldsymbol{\Sigma}|}} \exp\left(-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{x} - \boldsymbol{\mu})\right). \tag{B.9}$$

Namely, if

$$p(\boldsymbol{x}, \boldsymbol{y}) = \mathcal{N}\left(\begin{pmatrix} \boldsymbol{x} \\ \boldsymbol{y} \end{pmatrix} \middle| \boldsymbol{m}, \boldsymbol{P}\right) \tag{B.10}$$

$$= \mathcal{N}\left(\begin{pmatrix} \boldsymbol{x} \\ \boldsymbol{y} \end{pmatrix} \middle| \begin{pmatrix} \boldsymbol{\mu} \\ \boldsymbol{A}\boldsymbol{\mu} + \boldsymbol{B}\boldsymbol{u} \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Sigma} & \boldsymbol{\Sigma}\boldsymbol{A}^{\mathrm{T}} \\ \boldsymbol{A}\boldsymbol{\Sigma} & \boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A}^{\mathrm{T}} + \boldsymbol{G}\boldsymbol{Q}\boldsymbol{G}^{\mathrm{T}} \end{pmatrix}\right) \tag{B.11}$$

$$= \frac{1}{\sqrt{(2\pi)^{n+m} |\boldsymbol{\Sigma}||\boldsymbol{G}\boldsymbol{Q}\boldsymbol{G}^{\mathrm{T}}|}} \exp\left(-\frac{1}{2}\left(\begin{pmatrix} \boldsymbol{x} \\ \boldsymbol{y} \end{pmatrix} - \begin{pmatrix} \boldsymbol{\mu} \\ \boldsymbol{A}\boldsymbol{\mu} + \boldsymbol{B}\boldsymbol{u} \end{pmatrix}\right)^{\mathrm{T}} \boldsymbol{P}^{-1}\left(\begin{pmatrix} \boldsymbol{x} \\ \boldsymbol{y} \end{pmatrix} - \begin{pmatrix} \boldsymbol{\mu} \\ \boldsymbol{A}\boldsymbol{\mu} + \boldsymbol{B}\boldsymbol{u} \end{pmatrix}\right)\right)$$

where we have used Property B.1 and Property B.2 to compute the inverse and determinant of $\boldsymbol{P}$ *since* $\boldsymbol{\Sigma}$ *is a positive-definite symmetric matrix and thus, invertible:*

$$
\boldsymbol{P}^{-1} = \begin{pmatrix} \boldsymbol{\Sigma} & \boldsymbol{\Sigma}\boldsymbol{A}^{\mathrm{T}} \\ \boldsymbol{A}\boldsymbol{\Sigma} & \boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A}^{\mathrm{T}} + \boldsymbol{G}\boldsymbol{Q}\boldsymbol{G}^{\mathrm{T}} \end{pmatrix}^{-1} = \begin{pmatrix} \boldsymbol{\Sigma}^{-1} + \boldsymbol{A}^{\mathrm{T}}(\boldsymbol{G}\boldsymbol{Q}\boldsymbol{G}^{\mathrm{T}})^{-1}\boldsymbol{A} & -\boldsymbol{A}^{\mathrm{T}}(\boldsymbol{G}\boldsymbol{Q}\boldsymbol{G}^{\mathrm{T}})^{-1} \\ -(\boldsymbol{G}\boldsymbol{Q}\boldsymbol{G}^{\mathrm{T}})^{-1}\boldsymbol{A} & (\boldsymbol{G}\boldsymbol{Q}\boldsymbol{G}^{\mathrm{T}})^{-1} \end{pmatrix},
$$
(B.12)

$$
\det \boldsymbol{P} = \det\left[\begin{pmatrix} \boldsymbol{\Sigma} & \boldsymbol{\Sigma}\boldsymbol{A}^{\mathrm{T}} \\ \boldsymbol{A}\boldsymbol{\Sigma} & \boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A}^{\mathrm{T}} + \boldsymbol{G}\boldsymbol{Q}\boldsymbol{G}^{\mathrm{T}} \end{pmatrix} \right] = \det\left[\boldsymbol{\Sigma}\right]\det\left[\boldsymbol{G}\boldsymbol{Q}\boldsymbol{G}^{\mathrm{T}}\right].
$$
(B.13)

If now we expand the exponent of Eq. (B.10), we get

$$
\begin{pmatrix} \boldsymbol{x} - \boldsymbol{\mu} \\ \boldsymbol{y} - \boldsymbol{A}\boldsymbol{\mu} \end{pmatrix}^{\mathrm{T}} \begin{pmatrix} \boldsymbol{\Sigma}^{-1} + \boldsymbol{A}^{\mathrm{T}}(\boldsymbol{G}\boldsymbol{Q}\boldsymbol{G}^{\mathrm{T}})^{-1}\boldsymbol{A} & -\boldsymbol{A}^{\mathrm{T}}(\boldsymbol{G}\boldsymbol{Q}\boldsymbol{G}^{\mathrm{T}})^{-1} \\ -(\boldsymbol{G}\boldsymbol{Q}\boldsymbol{G}^{\mathrm{T}})^{-1}\boldsymbol{A} & (\boldsymbol{G}\boldsymbol{Q}\boldsymbol{G}^{\mathrm{T}})^{-1} \end{pmatrix} \begin{pmatrix} \boldsymbol{x} - \boldsymbol{\mu} \\ \boldsymbol{y} - \boldsymbol{A}\boldsymbol{\mu} \end{pmatrix}
$$

$$
= \begin{pmatrix} \boldsymbol{x} - \boldsymbol{\mu} \\ \boldsymbol{y} - \boldsymbol{A}\boldsymbol{\mu} \end{pmatrix}^{\mathrm{T}} \begin{pmatrix} (\boldsymbol{\Sigma}^{-1} + \boldsymbol{A}^{\mathrm{T}}(\boldsymbol{G}\boldsymbol{Q}\boldsymbol{G}^{\mathrm{T}})^{-1}\boldsymbol{A})(\boldsymbol{x} - \boldsymbol{\mu}) - \boldsymbol{A}^{\mathrm{T}}(\boldsymbol{G}\boldsymbol{Q}\boldsymbol{G}^{\mathrm{T}})^{-1}(\boldsymbol{y} - \boldsymbol{A}\boldsymbol{\mu}) \\ -(\boldsymbol{G}\boldsymbol{Q}\boldsymbol{G}^{\mathrm{T}})^{-1}\boldsymbol{A}(\boldsymbol{x} - \boldsymbol{\mu}) + (\boldsymbol{G}\boldsymbol{Q}\boldsymbol{G}^{\mathrm{T}})^{-1}(\boldsymbol{y} - \boldsymbol{A}\boldsymbol{\mu}) \end{pmatrix}
$$

$$
= \begin{pmatrix} \boldsymbol{x} - \boldsymbol{\mu} \\ \boldsymbol{y} - \boldsymbol{A}\boldsymbol{\mu} \end{pmatrix}^{\mathrm{T}} \begin{pmatrix} \boldsymbol{\Sigma}^{-1}(\boldsymbol{x} - \boldsymbol{\mu}) + \boldsymbol{A}^{\mathrm{T}}(\boldsymbol{G}\boldsymbol{Q}\boldsymbol{G}^{\mathrm{T}})^{-1}(\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y}) \\ -(\boldsymbol{G}\boldsymbol{Q}\boldsymbol{G}^{\mathrm{T}})^{-1}(\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y}) \end{pmatrix}
$$

$$
= (\boldsymbol{x} - \boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{x} - \boldsymbol{\mu}) + (\boldsymbol{x} - \boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}(\boldsymbol{G}\boldsymbol{Q}\boldsymbol{G}^{\mathrm{T}})^{-1}(\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y}) + (\boldsymbol{y} - \boldsymbol{A}\boldsymbol{\mu})^{\mathrm{T}}(\boldsymbol{G}\boldsymbol{Q}\boldsymbol{G}^{\mathrm{T}})^{-1}(\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x})
$$

$$
= (\boldsymbol{x} - \boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{x} - \boldsymbol{\mu}) + (\boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}} - \boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}} + \boldsymbol{y}^{\mathrm{T}} - \boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{A}^{T})(\boldsymbol{G}\boldsymbol{Q}\boldsymbol{G}^{\mathrm{T}})^{-1}(\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x})
$$

$$
= (\boldsymbol{x} - \boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{x} - \boldsymbol{\mu}) + (\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x})^{\mathrm{T}}(\boldsymbol{G}\boldsymbol{Q}\boldsymbol{G}^{\mathrm{T}})^{-1}(\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}).
$$
(B.14)

It follows then that Eq. (B.10) can indeed be split into the two Gaussian densities of Eq. (B.4) and Eq. (B.5):

$$
p(\boldsymbol{x}, \boldsymbol{y}) = \frac{1}{\sqrt{(2\pi)^{n+m}|\boldsymbol{\Sigma}||\boldsymbol{G}\boldsymbol{Q}\boldsymbol{G}^{\mathrm{T}}|}} \exp\left(-\frac{1}{2}\left((\boldsymbol{x} - \boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{x} - \boldsymbol{\mu}) + (\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x})^{\mathrm{T}}(\boldsymbol{G}\boldsymbol{Q}\boldsymbol{G}^{\mathrm{T}})^{-1}(\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x})\right)\right)
$$

$$
= \frac{1}{\sqrt{(2\pi)^n|\boldsymbol{\Sigma}|}} e^{-\frac{1}{2}\left((\boldsymbol{x} - \boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{x} - \boldsymbol{\mu})\right)} \frac{1}{\sqrt{(2\pi)^m|\boldsymbol{G}\boldsymbol{Q}\boldsymbol{G}^{\mathrm{T}}|}} e^{-\frac{1}{2}\left((\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x})^{\mathrm{T}}(\boldsymbol{G}\boldsymbol{Q}\boldsymbol{G}^{\mathrm{T}})^{-1}(\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x})\right)}
$$

$$
= \mathcal{N}(\boldsymbol{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma})\mathcal{N}(\boldsymbol{y}|\boldsymbol{A}\boldsymbol{x}, \boldsymbol{G}\boldsymbol{Q}\boldsymbol{G}^{\mathrm{T}}).
$$
(B.15)

Lemma B.2 (Marginal and conditional distributions of a joint Gaussian probability density). *Let*

two random variables $\boldsymbol{x}$ and $\boldsymbol{y}$ have the joint Gaussian probability density

$$\begin{pmatrix} \boldsymbol{x} \\ \boldsymbol{y} \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \boldsymbol{x} \\ \boldsymbol{y} \end{pmatrix} \middle| \begin{pmatrix} \boldsymbol{\mu}_x \\ \boldsymbol{\mu}_y \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Sigma}_{x,x} & \boldsymbol{\Sigma}_{x,y} \\ \boldsymbol{\Sigma}_{y,x} & \boldsymbol{\Sigma}_{y,y} \end{pmatrix} \right) \tag{B.16}$$

where $\boldsymbol{\Sigma}_{x,y}^{\mathrm{T}} = \boldsymbol{\Sigma}_{y,x}$. Then, the marginal and conditional probability densities of $\boldsymbol{x}$ and $\boldsymbol{y}$ are

$$\boldsymbol{x} \sim p(\boldsymbol{x}) = \mathcal{N}(\boldsymbol{x}|\boldsymbol{\mu}_x, \boldsymbol{\Sigma}_{x,x}), \tag{B.17}$$

$$\boldsymbol{y} \sim p(\boldsymbol{y}) = \mathcal{N}(\boldsymbol{y}|\boldsymbol{\mu}_y, \boldsymbol{\Sigma}_{y,y}), \tag{B.18}$$

$$\boldsymbol{x}|\boldsymbol{y} \sim p(\boldsymbol{x}|\boldsymbol{y}) = \mathcal{N}(\boldsymbol{x}|\boldsymbol{\mu}_x + \boldsymbol{\Sigma}_{x,y}\boldsymbol{\Sigma}_{y,y}^{-1}(\boldsymbol{y} - \boldsymbol{\mu}_y), \boldsymbol{\Sigma}_{x,x} - \boldsymbol{\Sigma}_{x,y}\boldsymbol{\Sigma}_{y,y}^{-1}\boldsymbol{\Sigma}_{x,y}^{\mathrm{T}}), \tag{B.19}$$

$$\boldsymbol{y}|\boldsymbol{x} \sim p(\boldsymbol{y}|\boldsymbol{x}) = \mathcal{N}(\boldsymbol{y}|\boldsymbol{\mu}_y + \boldsymbol{\Sigma}_{x,y}^{\mathrm{T}}\boldsymbol{\Sigma}_{x,x}^{-1}(\boldsymbol{x} - \boldsymbol{\mu}_x), \boldsymbol{\Sigma}_{y,y} - \boldsymbol{\Sigma}_{x,y}^{\mathrm{T}}\boldsymbol{\Sigma}_{x,x}^{-1}\boldsymbol{\Sigma}_{x,y}) \tag{B.20}$$

C

Appendix for Chap. 3

C.1 FOURIER COEFFICIENTS OF A DOUBLE SERIES EXPANSION

Proposition C.1 (Double time–frequency Fourier series expansion of a continuous mode). *Let $\hat{b}(t)$ be a continuous mode defined over time t. By segmenting time into intervals of duration Δt and applying a Fourier expansion, the continuous mode $\hat{b}(t)$ can be written as a double series expansion in both time and frequency:*

$$\hat{b}(t) = \frac{1}{\sqrt{\Delta t}} \sum_{n=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \hat{b}_{n,k}\, \Theta(t - t_n)\, \mathrm{e}^{-i2\pi k t/\Delta t}, \tag{C.1}$$

where $t_n = n\Delta t$ is the discretized time step, $\Theta(t - t_n)$ is the Heaviside function defined as $\Theta(u) = 1$ for $0 \leq u < \Delta t$ and zero otherwise, and $\hat{b}_{n,k}$ are the "Fourier" coefficients of the double series, given by:

$$\hat{b}_{n,k} = \frac{1}{\sqrt{\Delta t}} \int_{t_n}^{t_n+\Delta t} \hat{b}(t)\mathrm{e}^{i2\pi k t/\Delta t}\, \mathrm{d}t. \tag{C.2}$$

Proof. *Let us start the proof by multiplying both sides of Eq. (C.1) by* $e^{i2\pi\ell t/\Delta t}$, *as well as integrating them w.r.t. time from* t_m *to* $t_m + \Delta t$:

$$\int_{t_m}^{t_m+\Delta t} \hat{b}(t)e^{i2\pi\ell t/\Delta t}\, dt = \int_{t_m}^{t_m+\Delta t} \frac{1}{\sqrt{\Delta t}} \sum_{n=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \hat{b}_{n,k}\,\Theta(t-t_n)\, e^{-i2\pi k t/\Delta t} e^{i2\pi\ell t/\Delta t}\, dt. \quad \text{(C.3)}$$

Since $\Theta(t-t_n)$ *is nonzero only for* $t_n \leq t < t_n + \Delta t$, *and considering that the integration is specifically performed over the interval* t_m *to* $t_m + \Delta t$, $\Theta(t-t_n)$ *can be replaced with* 1 *within the integration interval* $[t_m, t_m + \Delta t)$ *and zero otherwise. This allows for the removal of the summation over n in the expression, as the only non-zero contribution comes from the term when* $n = m$, *i.e., when the interval picked by the Heaviside function coincides with the integration interval. Then, the rhs of Eq. (C.3) reduces to*

$$\int_{t_m}^{t_m+\Delta t} \hat{b}(t)e^{i2\pi\ell t/\Delta t}\, dt = \sum_{k=-\infty}^{\infty} \frac{1}{\sqrt{\Delta t}}\hat{b}_{m,k} \int_{t_m}^{t_m+\Delta t} e^{-i2\pi k t/\Delta t} e^{i2\pi\ell t/\Delta t}\, dt.$$

It can be further simplified by using the orthogonality of the exponential functions, i.e., meaning that for terms with $k \neq \ell$, *the integral over the interval* Δt *will be zero due to periodicity of the exponents over that time period, and for* $k = \ell$, *the integral will yield the interval* Δt:

$$\int_{t_m}^{t_m+\Delta t} e^{-i2\pi k t/\Delta t} e^{i2\pi\ell t/\Delta t}\, dt = \Delta t\, \delta_{k,\ell}, \quad \text{(C.4)}$$

where $\delta_{k,\ell}$ *denotes the Kronecker delta. Therefore,*

$$\int_{t_m}^{t_m+\Delta t} \hat{b}(t)e^{i2\pi\ell t/\Delta t}\, dt = \sum_{k=-\infty}^{\infty} \frac{1}{\sqrt{\Delta t}}\hat{b}_{m,k}\, \Delta t\, \delta_{k,\ell} = \hat{b}_{m,\ell}\sqrt{\Delta t},$$

such that we can easily isolate the discretized modes $\hat{b}_{n,k}$ *as:*

$$\hat{b}_{n,k} = \frac{1}{\sqrt{\Delta t}} \int_{t_n}^{t_n+\Delta t} \hat{b}(t)e^{i2\pi k t/\Delta t}\, dt. \quad \text{(C.5)}$$

D

Appendix for Chap. 4

D.1 WRITING A QUANTUM MAP AS A CONVEX MIXTURE OF UNITARIES

Theorem D.1 (Quantum map as a convex mixture of unitaries). *Given a unitary evolution governed by a Hamiltonian $\xi\hat{H}$*

$$\mathcal{U}_{\xi,\tau}[\,\cdot\,] = \mathrm{e}^{-\mathrm{i}\xi\hat{H}\tau} \cdot \mathrm{e}^{\mathrm{i}\xi\hat{H}\tau}, \tag{D.1}$$

whose scalar encoding $\xi \in \mathbb{R}$ (frequency) is randomly distributed according to a Gaussian probability density

$$\xi \sim p_{\mu,\sigma}(\xi) = \mathcal{N}(\mu(\tau), \sigma^2(\tau)) = \frac{1}{\sqrt{2\pi\sigma^2(\tau)}} \exp\left\{-\frac{(\xi - \mu(\tau))^2}{2\sigma^2(\tau)}\right\}, \tag{D.2}$$

then, the quantum map Ω can be written as a convex mixture of these unitaries as:

$$\rho(\tau) = \Omega[\rho(0)] = \mathbb{E}_{p(\xi)}[\mathcal{U}_{\xi,\tau}[\rho(0)]] = \int d\xi\, p_{\mu,\sigma}(\xi)\, \mathrm{e}^{-\mathrm{i}\xi\hat{H}\tau}\rho(0)\mathrm{e}^{\mathrm{i}\xi\hat{H}\tau}, \tag{D.3}$$

if $\rho(\tau)$ corresponds to the solution of the following master equation:

$$\frac{d\rho(\tau)}{d\tau} = -i\omega(\tau)[\hat{H}, \rho(\tau)] + \Gamma(\tau)\left(\hat{H}\rho(\tau)\hat{H} - \frac{1}{2}\{\hat{H}^2, \rho(\tau)\}\right) \tag{D.4}$$

$$= -i\omega(\tau)[\hat{H}, \rho(\tau)] - \frac{1}{2}\Gamma(\tau)\left[\hat{H}, [\hat{H}, \rho(\tau)]\right] \tag{D.5}$$

with the time-dependent frequency and decay parameters being

$$\omega(\tau) = \mu(\tau) + \tau\dot{\mu}(\tau) \quad \text{and} \quad \Gamma(\tau) = 2\sigma^2(\tau)\tau\left(1 + \frac{\dot{\sigma}(\tau)}{\sigma(\tau)}\tau\right). \tag{D.6}$$

Proof. *The first step is to differentiate $\rho(\tau)$, as defined in Eq. (D.3), with respect to τ:*

$$\frac{d\rho(\tau)}{d\tau} = \int d\xi \left(\frac{d}{d\tau} p_{\mu,\sigma}(\xi)\right) \mathcal{U}_{\xi,\tau}[\rho(0)] + \int d\xi p_{\mu,\sigma}(\xi)\frac{d}{d\tau}\mathcal{U}_{\xi,\tau}[\rho(0)], \tag{D.7}$$

which, in turn, requires the differentiation of the Gaussian probability distribution,

$$\frac{d}{d\tau} p_{\mu,\sigma}(\xi) = \frac{d}{d\tau}\left[\frac{1}{\sqrt{2\pi\sigma^2(\tau)}}\exp\left\{-\frac{(\xi-\mu(\tau))^2}{2\sigma^2(\tau)}\right\}\right]$$

$$= \frac{d}{d\tau}\left[\frac{1}{\sqrt{2\pi\sigma^2(\tau)}}\right]\exp\left\{-\frac{(\xi-\mu(\tau))^2}{2\sigma^2(\tau)}\right\} + \frac{1}{\sqrt{2\pi\sigma^2(\tau)}}\frac{d}{d\tau}\left[\exp\left\{-\frac{(\xi-\mu(\tau))^2}{2\sigma^2(\tau)}\right\}\right]$$

$$= -\frac{\dot{\sigma}(\tau)}{\sigma(\tau)}\frac{1}{\sqrt{2\pi\sigma^2(\tau)}}\exp\left\{-\frac{(\xi-\mu(\tau))^2}{2\sigma^2(\tau)}\right\} - \frac{1}{\sqrt{2\pi\sigma^2(\tau)}}\exp\left\{-\frac{(\xi-\mu(\tau))^2}{2\sigma^2(\tau)}\right\}\frac{d}{d\tau}\frac{(\xi-\mu(\tau))^2}{2\sigma^2(\tau)}$$

$$= p_{\mu,\sigma}(\xi)\left(-\frac{\dot{\sigma}(\tau)}{\sigma(\tau)} - \frac{2\sigma^2(\tau)\frac{d}{d\tau}(\xi-\mu(\tau))^2 - (\xi-\mu(\tau))^2\frac{d}{d\tau}2\sigma^2(\tau)}{4\sigma^4(\tau)}\right)$$

$$= p_{\mu,\sigma}(\xi)\left(-\frac{\dot{\sigma}(\tau)}{\sigma(\tau)} - \frac{-4\sigma^2(\tau)(\xi-\mu(\tau))\dot{\mu}(\tau) - (\xi-\mu(\tau))^2 4\sigma(\tau)\dot{\sigma}(\tau)}{4\sigma^4(\tau)}\right)$$

$$= p_{\mu,\sigma}(\xi)\left(\frac{(\xi-\mu(\tau))\sigma(\tau)\dot{\mu}(\tau) + ((\xi-\mu(\tau))^2 - \sigma^2(\tau))\dot{\sigma}(\tau)}{\sigma^3(\tau)}\right), \tag{D.8}$$

and the unitary map,

$$\frac{d}{d\tau}\mathcal{U}_{\xi,\tau}[\rho(0)] = \frac{d}{d\tau}\left[e^{-i\xi\hat{H}\tau}\rho(0)e^{i\xi\hat{H}\tau}\right] = \frac{d}{d\tau}\left[e^{-i\xi\hat{H}\tau}\right]\rho(0)e^{i\xi\hat{H}\tau} + e^{-i\xi\hat{H}\tau}\rho(0)\frac{d}{d\tau}\left[e^{i\xi\hat{H}\tau}\right]$$

$$= (-i\xi\hat{H})\mathcal{U}_{\xi,\tau}[\rho(0)] + \mathcal{U}_{\xi,\tau}[\rho(0)](i\xi\hat{H}) = -i\left[\hat{H},\xi\mathcal{U}_{\xi,\tau}[\rho(0)]\right]. \tag{D.9}$$

Therefore, the derivative of the density matrix with respect to τ becomes,

$$\frac{d\rho(\tau)}{d\tau} = \frac{1}{\sigma^3(\tau)}\int d\xi\, p_{\mu,\sigma}(\xi)\left((\xi-\mu(\tau))\sigma(\tau)\dot{\mu}(\tau) + \left((\xi-\mu(\tau))^2 - \sigma^2(\tau)\right)\dot{\sigma}(\tau)\right)\mathcal{U}_{\xi,\tau}[\rho(0)]$$

$$- i\int d\xi\, p_{\mu,\sigma}(\xi)\left[\hat{H},\xi\mathcal{U}_{\xi,\tau}[\rho(0)]\right]$$

$$= \frac{1}{\sigma^3(\tau)}\mathbb{E}_{p_{\mu,\sigma}(\xi)}\left[\left(\xi\sigma(\tau)\dot{\mu}(\tau) - \mu(\tau)\sigma(\tau)\dot{\mu}(\tau) + (\xi-\mu(\tau))^2\dot{\sigma}(\tau) - \sigma^2(\tau)\dot{\sigma}(\tau)\right)\mathcal{U}_{\xi,\tau}[\rho(0)]\right]$$

$$- i\left[\hat{H},\mathbb{E}_{p_{\mu,\sigma}(\xi)}[\xi\mathcal{U}_{\xi,\tau}[\rho(0)]]\right] = \frac{1}{\sigma^3(\tau)}\mathbb{E}_{p_{\mu,\sigma}(\xi)}[\xi\sigma(\tau)\dot{\mu}(\tau)\mathcal{U}_{\xi,\tau}[\rho(0)]]$$

$$- \frac{1}{\sigma^3(\tau)}\mathbb{E}_{p_{\mu,\sigma}(\xi)}[\mu(\tau)\sigma(\tau)\dot{\mu}(\tau)\mathcal{U}_{\xi,\tau}[\rho(0)]] + \frac{1}{\sigma^3(\tau)}\mathbb{E}_{p_{\mu,\sigma}(\xi)}\left[(\xi-\mu(\tau))^2\dot{\sigma}(\tau)\mathcal{U}_{\xi,\tau}[\rho(0)]\right]$$

$$- \frac{1}{\sigma^3(\tau)}\mathbb{E}_{p_{\mu,\sigma}(\xi)}\left[\sigma^2(\tau)\dot{\sigma}(\tau)\mathcal{U}_{\xi,\tau}[\rho(0)]\right] - i\left[\hat{H},\mathbb{E}_{p_{\mu,\sigma}(\xi)}[\xi\mathcal{U}_{\xi,\tau}[\rho(0)]]\right]$$

$$= \frac{\dot{\mu}(\tau)}{\sigma^2(\tau)}\mathbb{E}_{p_{\mu,\sigma}(\xi)}[\xi\mathcal{U}_{\xi,\tau}[\rho(0)]] - \frac{\mu(\tau)\dot{\mu}(\tau)}{\sigma^2(\tau)}\rho(\tau) + \frac{\dot{\sigma}(\tau)}{\sigma^3(\tau)}\mathbb{E}_{p_{\mu,\sigma}}\left[(\xi-\mu(\tau))^2\mathcal{U}_{\xi,\tau}[\rho(0)]\right]$$

$$- \frac{\dot{\sigma}(\tau)}{\sigma(\tau)}\rho(\tau) - i\left[\hat{H},\mathbb{E}_{p_{\mu,\sigma}(\xi)}[\xi\mathcal{U}_{\xi,\tau}[\rho(0)]]\right]. \tag{D.10}$$

To continue our proof, we must then explicitly evaluate the averaged expressions $\mathbb{E}_{p_{\mu,\sigma}(\xi)}[\xi\mathcal{U}_{\xi,\tau}[\rho(0)]]$ and $\mathbb{E}_{p_{\mu,\sigma}}[(\xi-\mu(\tau))^2\mathcal{U}_{\xi,\tau}[\rho(0)]]$. To do so we will employ the trick of taking the derivative of the Gaussian probability distribution with respect to $\mu(\tau)$. Namely,

$$\sigma^2(\tau)\frac{d}{d\mu(\tau)}p_{\mu,\sigma}(\xi) = \sigma^2(\tau)\frac{d}{d\mu(\tau)}\left[\frac{1}{\sqrt{2\pi\sigma^2(\tau)}}\exp\left\{-\frac{(\xi-\mu(\tau))^2}{2\sigma^2(\tau)}\right\}\right]$$

$$= \xi p_{\mu,\sigma}(\xi) - \mu(\tau)p_{\mu,\sigma}(\xi) \tag{D.11}$$

and

$$\sigma^4(\tau)\frac{\mathrm{d}^2}{\mathrm{d}\mu(\tau)^2}p_{\mu,\sigma}(\xi) = \sigma^2(\tau)\frac{\mathrm{d}}{\mathrm{d}\mu(\tau)}\left[(\xi-\mu)p_{\mu,\sigma}(\xi)\right]$$

$$= \sigma^2(\tau)\left(-p_{\mu,\sigma}(\xi) + (\xi-\mu(\tau))\frac{\mathrm{d}}{\mathrm{d}\mu(\tau)}p_{\mu,\sigma}(\xi)\right)$$

$$= -\sigma^2(\tau)\,p_{\mu,\sigma}(\xi) + (\xi-\mu(\tau))\sigma^2(\tau)\frac{\mathrm{d}}{\mathrm{d}\mu(\tau)}p_{\mu,\sigma}(\xi)$$

$$= -\sigma^2(\tau)\,p_{\mu,\sigma}(\xi) + (\xi-\mu(\tau))^2 p_{\mu,\sigma}(\xi). \tag{D.12}$$

Then, we can move to evaluating the averages, i.e.,

$$\mathbb{E}_{p_{\mu,\sigma}(\xi)}[\xi\,\mathcal{U}_{\xi,\tau}[\rho(0)]] = \int \mathrm{d}\xi\,\xi\,p_{\mu,\sigma}(\xi)\,\mathcal{U}_{\xi,\tau}[\rho(0)]$$

$$\overset{(\text{D.11})}{=} \int \mathrm{d}\xi\,\left(\sigma^2(\tau)\frac{\mathrm{d}}{\mathrm{d}\mu(\tau)}p_{\mu,\sigma}(\xi) + \mu(\tau)\,p_{\mu,\sigma}(\xi)\right)\mathcal{U}_{\xi,\tau}[\rho(0)]$$

$$= \sigma^2(\tau)\int \mathrm{d}\xi\,\frac{\mathrm{d}}{\mathrm{d}\mu(\tau)}p_{\mu,\sigma}(\xi)\,\mathcal{U}_{\xi,\tau}[\rho(0)] + \mu(\tau)\rho(\tau)$$

$$\overset{(\text{D.15})}{=} -\mathrm{i}\sigma^2(\tau)\tau\left[\hat{H},\rho(\tau)\right] + \mu(\tau)\rho(\tau), \tag{D.13}$$

and

$$\mathbb{E}_{p_{\mu,\sigma}}\left[(\xi-\mu(\tau))^2\,\mathcal{U}_{\xi,\tau}[\rho(0)]\right] = \int \mathrm{d}\xi\,(\xi-\mu(\tau))^2\,p_{\mu,\sigma}(\xi)\,\mathcal{U}_{\xi,\tau}[\rho(0)]$$

$$\overset{(\text{D.12})}{=} \int \mathrm{d}\xi\,\left(\sigma^4(\tau)\frac{\mathrm{d}^2}{\mathrm{d}\mu(\tau)^2}p_{\mu,\sigma}(\xi) + \sigma^2(\tau)\,p_{\mu,\sigma}(\xi)\right)\mathcal{U}_{\xi,\tau}[\rho(0)]$$

$$= \sigma^4(\tau)\int \mathrm{d}\xi\,\frac{\mathrm{d}^2}{\mathrm{d}\mu(\tau)^2}p_{\mu,\sigma}(\xi)\,\mathcal{U}_{\xi,\tau}[\rho(0)] + \sigma^2(\tau)\rho(\tau)$$

$$\overset{(\text{D.16})}{=} -\sigma^4(\tau)\tau^2\left[\hat{H},\left[\hat{H},\rho(\tau)\right]\right] + \sigma^2(\tau)\rho(\tau) \tag{D.14}$$

where in the last step of both expressions we have used the change of variable of $\xi \to \xi + \mu(\tau)$ to

calculate

$$\frac{\mathrm{d}}{\mathrm{d}\mu(\tau)} \int \mathrm{d}\xi \, p_{\mu,\sigma}(\xi) \, \mathcal{U}_{\xi,\tau}[\rho(0)] = \frac{\mathrm{d}}{\mathrm{d}\mu(\tau)} \int \mathrm{d}\xi \, p_{\mu,\sigma}(\xi + \mu(\tau)) \, \mathcal{U}_{\xi+\mu(\tau),\tau}[\rho(0)]$$

$$= \frac{\mathrm{d}}{\mathrm{d}\mu(\tau)} \int \mathrm{d}\xi \, p_{0,\sigma(\tau)}(\xi) \, \mathcal{U}_{\xi+\mu(\tau),\tau}[\rho(0)] = \int \mathrm{d}\xi \, p_{0,\sigma(\tau)}(\xi) \, \frac{\mathrm{d}}{\mathrm{d}\mu(\tau)} \mathcal{U}_{\xi+\mu(\tau),\tau}[\rho(0)]$$

$$\overset{(D.9)}{=} -\mathrm{i} \int \mathrm{d}\xi \, p_{0,\sigma(\tau)}(\xi) \left[\hat{H}, \tau \, \mathcal{U}_{\xi,\tau}[\rho(0)] \right] = -\mathrm{i}\tau \left[\hat{H}, \rho(\tau) \right], \tag{D.15}$$

which then implies that

$$\frac{\mathrm{d}^2}{\mathrm{d}\mu(\tau)^2} \int \mathrm{d}\xi \, p_{\mu,\sigma}(\xi) \, \mathcal{U}_{\xi,\tau}[\rho(0)] = \frac{\mathrm{d}}{\mathrm{d}\mu(\tau)} \left[\frac{\mathrm{d}}{\mathrm{d}\mu(\tau)} \int \mathrm{d}\xi \, p_{\mu,\sigma}(\xi) \, \mathcal{U}_{\xi,\tau}[\rho(0)] \right]$$

$$= -\mathrm{i}\tau \left[\hat{H}, \frac{\mathrm{d}}{\mathrm{d}\mu(\tau)} \rho(\tau) \right] = -\mathrm{i}\tau \left[\hat{H}, \frac{\mathrm{d}}{\mathrm{d}\mu(\tau)} \int \mathrm{d}\xi \, p_{\mu,\sigma}(\xi) \, \mathcal{U}_{\xi,\tau}[\rho(0)] \right]$$

$$= -\tau^2 \left[\hat{H}, \left[\hat{H}, \rho(\tau) \right] \right]. \tag{D.16}$$

Armed now with the explicit form of the expected values given in Eq. (D.13) and Eq. (D.14), we can complete our derivation of Eq. (D.10). Namely,

$$\frac{\mathrm{d}\rho(\tau)}{\mathrm{d}\tau} = \frac{\dot{\mu}(\tau)}{\sigma^2(\tau)} \mathbb{E}_{p_{\mu,\sigma}(\xi)}[\xi \mathcal{U}_{\xi,\tau}[\rho(0)]] + \frac{\dot{\sigma}(\tau)}{\sigma^3(\tau)} \mathbb{E}_{p_{\mu,\sigma}} \left[(\xi - \mu(\tau))^2 \, \mathcal{U}_{\xi,\tau}[\rho(0)] \right]$$

$$- \left(\frac{\mu(\tau)\dot{\mu}(\tau)}{\sigma^2(\tau)} + \frac{\dot{\sigma}(\tau)}{\sigma(\tau)} \right) \rho(\tau) - \mathrm{i} \left[\hat{H}, \mathbb{E}_{p_{\mu,\sigma}(\xi)}[\xi \mathcal{U}_{\xi,\tau}[\rho(0)]] \right]$$

$$= \frac{\dot{\mu}(\tau)}{\sigma^2(\tau)} \left(-\mathrm{i}\sigma^2(\tau)\tau \left[\hat{H}, \rho(\tau) \right] + \mu(\tau)\rho(\tau) \right) - \left(\frac{\mu(\tau)\dot{\mu}(\tau)}{\sigma^2(\tau)} + \frac{\dot{\sigma}(\tau)}{\sigma(\tau)} \right) \rho(\tau)$$

$$+ \frac{\dot{\sigma}(\tau)}{\sigma^3(\tau)} \left(-\sigma^4(\tau)\tau^2 \left[\hat{H}, \left[\hat{H}, \rho(\tau) \right] \right] + \sigma^2(\tau)\rho(\tau) \right)$$

$$- \mathrm{i} \left[\hat{H}, \left(-\mathrm{i}\sigma^2(\tau)\tau \left[\hat{H}, \rho(\tau) \right] + \mu(\tau)\rho(\tau) \right) \right]$$

$$= -\mathrm{i}\dot{\mu}(\tau)\tau \left[\hat{H}, \rho(\tau) \right] - \sigma(\tau)\dot{\sigma}(\tau)\tau^2 \left[\hat{H}, \left[\hat{H}, \rho(\tau) \right] \right]$$

$$- \sigma^2(\tau)\tau \left[\hat{H}, \left[\hat{H}, \rho(\tau) \right] \right] - \mathrm{i}\mu(\tau) \left[\hat{H}, \rho(\tau) \right]$$

$$= -\mathrm{i}(\mu(\tau) + \dot{\mu}(\tau)\tau) \left[\hat{H}, \rho(\tau) \right] - \left(\frac{\dot{\sigma}(\tau)}{\sigma(\tau)}\tau + 1 \right) \sigma^2(\tau)\tau \left[\hat{H}, \left[\hat{H}, \rho(\tau) \right] \right] \tag{D.17}$$

which is then in the desired form of Eq. (D.5).

D.2 Gaussian properties II

Lemma D.2 (The integral of $N+1$ Gaussians). *The integral of $N+1$ Gaussians can be written as*

$$
\frac{1}{(2\pi V_c)^{1/2}} \frac{1}{(2\pi V_\ell)^{N/2}} \int d\xi \, e^{-\frac{\xi^2}{2V_c}} e^{-\sum_{i=1}^{N} \frac{(\zeta^{(i)} - \xi - \omega)^2}{2V_\ell}}
$$

$$
= \frac{1}{\sqrt{2\pi(V_c + V_\ell/N)}} f(\boldsymbol{\zeta}) \exp\left\{ -\frac{(\bar{\zeta} - \omega)^2}{2(V_c + V_\ell/N)} \right\} \tag{D.18}
$$

where $\bar{\zeta} := \frac{1}{N} \sum_{i=1}^{N} \zeta^{(i)}$ and

$$
f(\boldsymbol{\zeta}) = \sqrt{\frac{1}{N(2\pi V_\ell)^{N-1}}} \, \exp\left\{ -\frac{1}{2V_\ell} \left(\sum_{i=1}^{N} (\zeta^{(i)})^2 - N\bar{\zeta}^2 \right) \right\} \tag{D.19}
$$

Proof. Let us first start by expanding the exponent of the local term,

$$
(\zeta^{(i)} - (\xi + \omega))^2 = (\zeta^{(i)})^2 - 2(\xi + \omega)\zeta^{(i)} + (\xi + \omega)^2 \tag{D.20}
$$

such that when summing over $i = 1, \ldots, N$, we can it as

$$
\sum_{i=1}^{N} (\zeta^{(i)} - (\xi + \omega))^2 = \sum_{i=1}^{N} \left((\zeta^{(i)})^2 - 2(\xi + \omega)\zeta^{(i)} + (\xi + \omega)^2 \right)
$$

$$
= \sum_{i=1}^{N} (\zeta^{(i)})^2 - 2(\xi + \omega) \sum_{i=1}^{N} \zeta^{(i)} + N(\xi + \omega)^2 \tag{D.21}
$$

Next, we introduce the average of auxiliary frequencies experienced by N atoms,

$$
\bar{\zeta} := \frac{1}{N} \sum_{i=1}^{N} \zeta^{(i)} \tag{D.22}
$$

which will be employed to reformulate Eq. (D.21):

$$\sum_{i=1}^{N}(\zeta^{(i)} - (\xi + \omega))^2 = \sum_{i=1}^{N}(\zeta^{(i)})^2 - 2(\xi + \omega)N\bar{\zeta} + N(\xi + \omega)^2$$

$$= N\bar{\zeta}^2 - 2(\xi + \omega)N\bar{\zeta} + N(\xi + \omega)^2 + \sum_{i=1}^{N}(\zeta^{(i)})^2 - N\bar{\zeta}^2$$

$$= N(\bar{\zeta} - (\xi + \omega))^2 + \sum_{i=1}^{N}(\zeta^{(i)})^2 - N\bar{\zeta}^2, \tag{D.23}$$

where in the penultimate step we have added and subtracted $N\bar{\zeta}$. Crucially, the local exponential in Eq. (D.18) can now be divided into two terms

$$e^{-\sum_{i=1}^{N}\frac{(\zeta^{(i)} - \xi - \omega)^2}{2V_\ell}} = e^{-\frac{(\bar{\zeta} - (\xi + \omega))^2}{2V_\ell/N}} e^{-\frac{1}{2V_\ell}\left(\sum_{i=1}^{N}(\zeta^{(i)})^2 - N\bar{\zeta}^2\right)}, \tag{D.24}$$

leaving the first exponential as the only one depending on ξ and ω and the second term as a function of exclusively $\boldsymbol{\zeta}$.

Therefore, we have reduced the problem of integrating $N + 1$ Gaussian functions with respect to ξ, to integrating only two Gaussians, i.e.,

$$\int \mathrm{d}\xi\, e^{-\frac{\xi^2}{2V_c}} e^{-\sum_{i=1}^{N}\frac{(\zeta^{(i)} - \xi - \omega)^2}{2V_\ell}} = e^{-\frac{1}{2V_\ell}\left(\sum_{i=1}^{N}(\zeta^{(i)})^2 - N\bar{\zeta}^2\right)} \int \mathrm{d}\xi\, e^{-\frac{\xi^2}{2V_c}} e^{-\frac{(\bar{\zeta} - (\xi + \omega))^2}{2V_\ell/N}} \tag{D.25}$$

which has a known result:

$$\int \mathrm{d}\xi\, e^{-\frac{\xi^2}{2V_c}} e^{-\frac{(\bar{\zeta} - (\xi + \omega))^2}{2V_\ell/N}} = \sqrt{2\pi\frac{V_c V_\ell/N}{V_c + V_\ell/N}} \exp\left\{-\frac{(\bar{\zeta} - \omega)^2}{2(V_c + V_\ell/N)}\right\}. \tag{D.26}$$

Therefore,

$$\frac{1}{(2\pi V_c)^{1/2}}\frac{1}{(2\pi V_\ell)^{N/2}} \int \mathrm{d}\xi\, e^{-\frac{\xi^2}{2V_c}} e^{-\sum_{i=1}^{N}\frac{(\zeta^{(i)} - \xi - \omega)^2}{2V_\ell}}$$

$$= \frac{1}{\sqrt{2\pi(V_c + V_\ell/N)}} f(\boldsymbol{\zeta}) \exp\left\{-\frac{(\bar{\zeta} - \omega)^2}{2(V_c + V_\ell/N)}\right\} \tag{D.27}$$

where

$$f(\boldsymbol{\zeta}) = \sqrt{\frac{1}{N(2\pi V_\ell)^{N-1}}} \, \exp\left\{-\frac{1}{2V_\ell}\left(\sum_{i=1}^{N}(\zeta^{(i)})^2 - N\bar{\zeta}^2\right)\right\} \tag{D.28}$$

Lemma D.3 (Set of nested integrals of Gaussian functions). *Consider the following recurrence relation defining $\mathcal{P}_j(\omega_j)$ in terms of $\mathcal{P}_{j-1}(\omega_{j-1})$ for $j = 0, 1, 2, \ldots$:*

$$\mathcal{P}_j(\omega_j) = \int d\omega_{j-1} \frac{1}{\sqrt{2\pi V_P}} \exp\left\{-\frac{(\omega_j - \omega_{j-1})^2}{2V_P}\right\} \frac{1}{\sqrt{2\pi V_Q}} \exp\left\{-\frac{(\bar{\zeta}_{j-1} - \omega_{j-1})^2}{2V_Q}\right\} \mathcal{P}_{j-1}(\omega_{j-1}) \tag{D.29}$$

which involves a generalized convolution of Gaussian distributions with variances V_P and V_Q, respectively. The initial condition for the recursive relation is given by a Gaussian of the form

$$\mathcal{P}_0(\omega_0) = C_0 \exp\left\{-\frac{(\omega_0 - \mu_0)^2}{2V_0}\right\}, \tag{D.30}$$

where C_0 is a constant, $V_0 \geq 0$ and $\mu_0 \in \mathbb{R}$. Then, $\forall j \geq 1$, the solution to the recurrence relation is

$$\mathcal{P}_j(\omega_j) = C_j \exp\left\{-\frac{(\omega_j - \mu_j)^2}{2V_j}\right\}, \tag{D.31}$$

with parameters C_j, μ_j and V_j given in turn by the following coupled recurrence relations:

$$C_j = C_{j-1}\left(2\pi\left(V_P + V_Q + \frac{V_P V_Q}{V_{j-1}}\right)\right)^{-1/2} \exp\left\{-\frac{(\omega_{j-1} - \mu_{j-1})^2}{2(V_Q + V_{j-1})}\right\} \tag{D.32}$$

$$\mu_j = \frac{V_Q \mu_{j-1} + V_{j-1}\bar{\zeta}_{j-1}}{V_Q + V_{j-1}} \tag{D.33}$$

$$V_j = V_P + \frac{V_Q V_{j-1}}{V_Q + V_{j-1}}. \tag{D.34}$$

Proof. *As with many recursive problem, it is sufficient to use mathematical induction to prove that the proposed solution in Eq. (D.31) holds. This involves verifying the base case for $j = 0$, and then showing that the relation (D.29) is fulfilled by Eq. (D.31) for any $j \geq 1$. The base case is trivially satisfied by the definition. For the inductive step, we assume that $\mathcal{P}_{j-1}(\omega_{j-1})$*

takes the form given in Eq. (D.31), and we show that substituting this into the recurrence relation in Eq. (D.29) yields the correct form for $\mathcal{P}_j(\omega_j)$. This is done by performing the integral in Eq. (D.29) explicitly:

$$
\mathcal{P}_j(\omega_j) = \int d\omega_{j-1} \frac{1}{\sqrt{2\pi V_P}} \exp\left\{-\frac{(\omega_j - \omega_{j-1})^2}{2V_P}\right\} \frac{1}{\sqrt{2\pi V_Q}} \exp\left\{-\frac{(\bar{\zeta}_{j-1} - \omega_{j-1})^2}{2V_Q}\right\} C_{j-1} \exp\left\{-\frac{(\omega_{j-1} - \mu_{j-1})^2}{2V_{j-1}}\right\}
$$

$$
= \frac{C_{j-1}}{\sqrt{2\pi\left(V_Q + V_P + \frac{V_P V_Q}{V_{j-1}}\right)}} \exp\left\{-\frac{\omega_j^2 - 2\alpha\omega_j + \beta}{2\left(V_P + \frac{V_Q V_{j-1}}{V_Q + V_{j-1}}\right)}\right\} \tag{D.35}
$$

with α and β fulfilling:

$$
\alpha = \frac{V_Q \mu_{j-1} + V_{j-1}\bar{\zeta}_{j-1}}{V_Q + V_{j-1}} = \mu_j, \tag{D.36}
$$

$$
\beta = \frac{V_Q \mu_{j-1}^2 + V_{j-1}\bar{\zeta}_{j-1}^2 + V_P(\bar{\zeta}_{j-1} - \mu_{j-1})^2}{V_Q + V_{j-1}}, \tag{D.37}
$$

which are both independent of ω_j and ω_{j-1}. If now we 'complete the square', i.e., split the exponent of Eq. (D.31) into $\omega_j^2 - 2\alpha\omega_j + \beta = (\omega_j - \alpha)^2 - \alpha^2 + \beta$, and substitute the expressions for α and β, we get,

$$
\mathcal{P}_j(\omega_j) = \frac{C_{j-1} \exp\left\{-\frac{-\alpha^2 + \beta}{2\left(V_P + \frac{V_Q V_{j-1}}{V_Q + V_{j-1}}\right)}\right\}}{\sqrt{2\pi\left(V_Q + V_P + \frac{V_P V_Q}{V_{j-1}}\right)}} \exp\left\{-\frac{(\omega_j - \alpha)^2}{2\left(V_P + \frac{V_Q V_{j-1}}{V_Q + V_{j-1}}\right)}\right\} \tag{D.38}
$$

which, since $\alpha = \mu_j$ and

$$-\alpha^2 + \beta = -\mu_j^2 + \frac{V_Q \mu_{j-1}^2 + V_{j-1}\bar{\zeta}_{j-1}^2 + V_P(\bar{\zeta}_{j-1} - \mu_{j-1})^2}{V_Q + V_{j-1}}$$

$$= -\frac{(V_Q \mu_{j-1} + V_{j-1}\bar{\zeta}_{j-1})^2}{(V_Q + V_{j-1})^2} + \frac{V_Q \mu_{j-1}^2 + V_{j-1}\bar{\zeta}_{j-1}^2 + V_P(\bar{\zeta}_{j-1} - \mu_{j-1})^2}{V_Q + V_{j-1}}$$

$$= \frac{-V_Q^2 \mu_{j-1}^2 - 2V_Q \mu_{j-1}V_{j-1}\bar{\zeta}_{j-1} - V_{j-1}^2\bar{\zeta}_{j-1}^2 + V_Q^2 \mu_{j-1}^2 + V_{j-1}V_Q \mu_{j-1}^2}{(V_Q + V_{j-1})^2}$$

$$+ \frac{V_{j-1}^2\bar{\zeta}_{j-1}^2 + V_Q V_{j-1}\bar{\zeta}_{j-1}^2 + V_P(V_Q + V_{j-1})(\bar{\zeta}_{j-1}^2 - 2\bar{\zeta}_{j-1}\mu_{j-1} + \mu_{j-1}^2)}{(V_Q + V_{j-1})^2}$$

$$= \frac{-2V_Q \mu_{j-1}V_{j-1}\bar{\zeta}_{j-1} + V_{j-1}V_Q \mu_{j-1}^2 + V_Q V_{j-1}\bar{\zeta}_{j-1}^2 + V_P V_Q\bar{\zeta}_{j-1}^2 - 2V_P V_Q\bar{\zeta}_{j-1}\mu_{j-1}}{(V_Q + V_{j-1})^2}$$

$$+ \frac{V_P V_Q\mu_{j-1}^2 + V_P V_{j-1}\bar{\zeta}_{j-1}^2 - 2V_P V_{j-1}\bar{\zeta}_{j-1}\mu_{j-1} + V_P V_{j-1}\mu_{j-1}^2}{(V_Q + V_{j-1})^2}$$

$$= \frac{(V_Q V_{j-1} + V_P V_Q + V_P V_{j-1})\bar{\zeta}_{j-1}^2 - 2(V_Q V_{j-1} + V_P V_Q + V_P V_{j-1})\bar{\zeta}_{j-1}\mu_{j-1}}{(V_Q + V_{j-1})^2}$$

$$+ \frac{(V_Q V_{j-1} + V_P V_Q + V_P V_{j-1})\mu_{j-1}^2}{(V_Q + V_{j-1})^2} = \frac{(V_Q V_{j-1} + V_P V_Q + V_P V_{j-1})}{(V_Q + V_{j-1})^2}(\bar{\zeta}_{j-1} - \mu_{j-1})^2$$

$$= \left(V_P + \frac{V_Q V_{j-1}}{V_Q + V_{j-1}}\right)\frac{(\bar{\zeta}_{j-1} - \mu_{j-1})^2}{V_Q + V_{j-1}}, \tag{D.39}$$

yields the expected form of Eq. (D.31), with C_j, μ_j and V_j as specified by Eqs. (D.32-D.34).

E
Appendix for Chap. 5

Consider the most general dephasing evolution possible:

$$d\rho_{(c)}(t) = -i\,\omega(t)\left[\hat{J}_z, \rho_{(c)}(t)\right]dt + \sum_{\alpha=x,y,z} \kappa_\alpha \mathcal{D}[\hat{J}_\alpha]\rho_{(c)}(t)dt + M\mathcal{D}[\hat{J}_y]\rho_{(c)}(t)dt, \qquad \text{(E.1)}$$

where collective dephasing occurs along the three directions x, y and z at rates κ_x, κ_y and κ_z, respectively. From this master equation, a set of differential equations describing the evolution of

the corresponding observables $\hat{J}_x$, $\hat{J}_y$, and $\hat{J}_z$ can be derived:

$$d\omega = -\chi\omega(t)dt + \sqrt{q_\omega}dW_\omega, \tag{E.2}$$

$$d\langle\hat{J}_x\rangle = -\omega(t)\langle\hat{J}_y\rangle dt - \frac{1}{2}(M + \kappa_y + \kappa_z)\langle\hat{J}_x\rangle dt, \tag{E.3}$$

$$d\langle\hat{J}_y\rangle = \omega(t)\langle\hat{J}_x\rangle dt - \frac{1}{2}(\kappa_x + \kappa_z)\langle\hat{J}_y\rangle dt, \tag{E.4}$$

$$d\langle\hat{J}_z\rangle = -\frac{1}{2}\left(M + \kappa_y + \kappa_x\right)\langle\hat{J}_z\rangle dt, \tag{E.5}$$

Given that Eqs. (E.2-E.4) form a closed set of coupled differential equations, they can be numerically solved in order to yield the unconditional evolution of $\langle\hat{J}_x(t)\rangle$. However, since we would rather find an approximate analytical expression, we focus on short timescales such that $\bar{\omega}(t)\,t \ll 1$. Then, by substituting the random variable $\omega(t)$ by its time average $\bar{\omega}(t)$ into Eqs. (E.3-E.5):

$$\bar{\omega}(t) = \frac{1}{t}\int_0^t d\tau\,\omega(\tau), \tag{E.6}$$

the system of differential equations yields,

$$\langle\hat{J}_x(t)\rangle = \frac{J}{2\Theta}\,e^{-(M+\kappa_x+\kappa_y+2\kappa_z+\Theta)\,t/4}\left(M-\kappa_x+\kappa_y+\Theta-e^{t\Theta/2}(M-\kappa_x+\kappa_y-\Theta)\right), \tag{E.7}$$

where $\Theta = \sqrt{(M - \kappa_x + \kappa_y)^2 - 16\bar{\omega}^2(t)}$ and the initial conditions set were to the CSS state: $\langle\hat{J}_x(0)\rangle = J$ and $\langle\hat{J}_y(0)\rangle = 0$. By expanding the unconditional evolution of $\hat{J}_x$ to first order in $\bar{\omega}(t)$, we obtain

$$\langle\hat{J}_x(t)\rangle \approx J e^{-(M+\kappa_z+\kappa_y)\,t/2}$$
$$+ 2J e^{-(M+\kappa_z+\kappa_y)\,t/2}\bar{\omega}^2(t)\frac{2 - 2e^{(M-\kappa_x+\kappa_y)t/2} + t(M - \kappa_x + \kappa_y)}{(M - \kappa_x + \kappa_y)^2}. \tag{E.8}$$

Hence, he unconditional evolution of $\hat{J}_x$ can be approximated as

$$\langle\hat{J}_x(t)\rangle \approx J e^{-(M+\kappa_z+\kappa_y)\,t/2}, \tag{E.9}$$

when the first term of E.8 dominates over the second. In other words, when the time average of $\omega(t)$ fulfills

$$|\bar{\omega}(t)| \leq \frac{\sqrt{2}}{t}, \tag{E.10}$$

where we have approximated the exponent $e^{(M-\kappa_x+\kappa_y)t/2}$ in (E.8) as its Taylor expansion up to second order. Namely, $e^{(M-\kappa_x+\kappa_y)t/2} \approx 1 + \frac{1}{2}(M - \kappa_x + \kappa_y)t + \frac{1}{8}(M - \kappa_x + \kappa_y)^2 t^2$. Additionally, note that within the linear-Gaussian regime, the approximation Eq. (E.9) is independent of the decoherence rate κ_x. Therefore, the decoherence rate κ_x is redundant and disappears from Eqs. (5.16-5.19) when performing the Holstein-Primakoff transformation of the SME according to the $\hat{X}$ and $\hat{P}$ quadratures of Eq. (5.15). Mathematically, one can check that the dissipative terms disappear, since $\text{Tr}\{\hat{P}\mathcal{D}[\hat{X}]\rho_{(c)}\} = \text{Tr}\{\hat{X}\mathcal{D}[\hat{P}]\rho_{(c)}\} = 0$. This is can be intuitively explained—deviations of $\hat{J}(t)$ from the x-direction are then too small for the collective noise manifested via $\mathcal{D}[\hat{J}_x]$ in equation Eq. (E.1) to have any effect on the quadratures Eq. (5.15).

Next, we would like to find the mean and variance of the time average of $\omega(t)$,

$$\mathbb{E}[\bar{\omega}(t)] = \frac{1}{t} \int_0^t \mathbb{E}[\omega(\tau)] \, d\tau, \tag{E.11}$$

$$V[\bar{\omega}(t)] = \mathbb{E}\left[(\bar{\omega}(t) - \mathbb{E}[\bar{\omega}(t)])^2\right]. \tag{E.12}$$

As discussed in Sec. 1.3.6, the mean of a random variable $\omega(t)$ driven by an OU process of the form in Eq. (E.2) reads

$$\mathbb{E}[\omega(t)] = \omega(0) e^{-\chi t} = 0, \tag{E.13}$$

where we have taken $\omega(0) = 0$, while its covariance (as shown in Eq. (1.134)) is:

$$C[\omega(s), \omega(t)] = \mathbb{E}[\omega(s)\omega(t)] = \frac{q_\omega}{2\chi}\left(e^{-\chi|t-s|} - e^{-\chi(t+s)}\right). \tag{E.14}$$
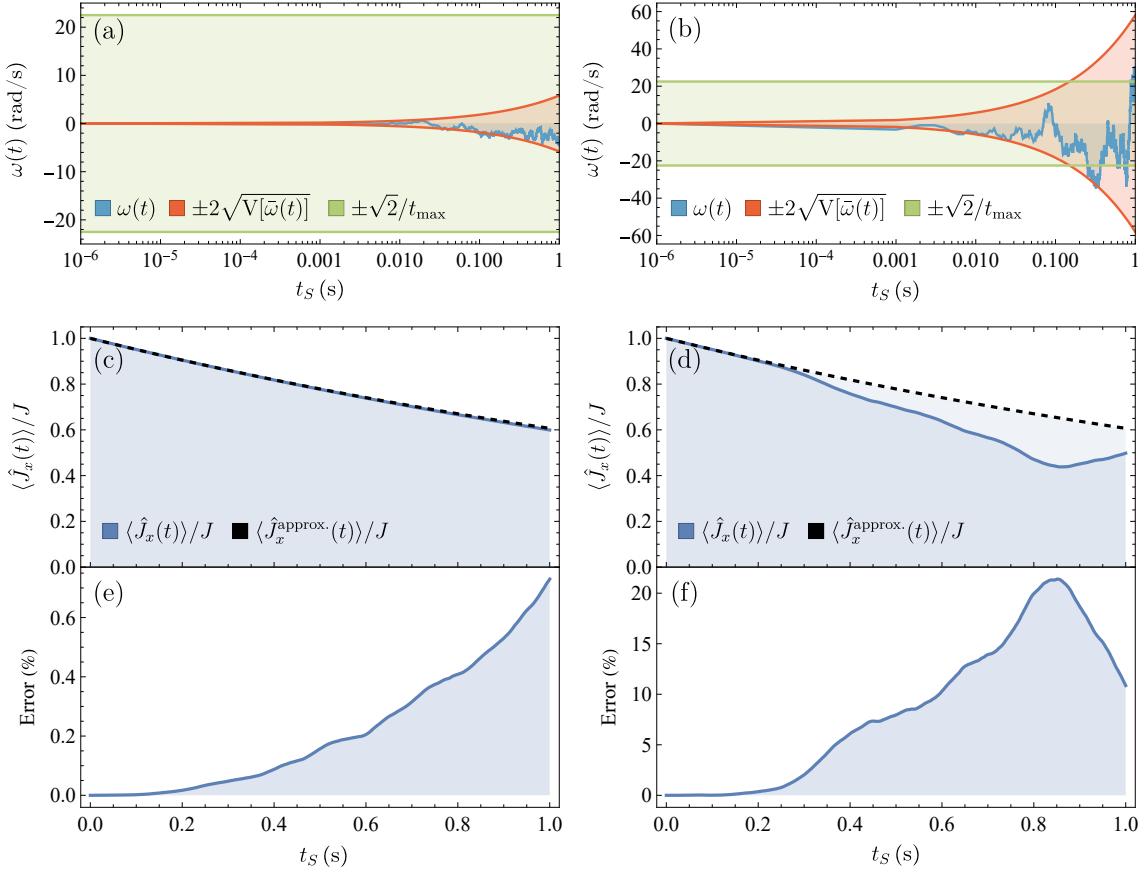
Figure E.1: Validity of the LG approximation for $\langle \hat{J}_x(t) \rangle$ at different field fluctuation strengths The parameters used to generate the plots are $M = 100\,\text{kHz}$, $\kappa_z = \kappa_c = 1\,\text{Hz}$, $J = 10^7$, $\eta = 1$, and $\chi = 0$, with $t_S = (M + \kappa_c)t$ being the scaled time such that $t_S = 1$ when $t = (M + \kappa_c)^{-1}$. Plots (a), (c), and (d) (left column) have been generated with a field fluctuation strength of $q_\omega = 10^{14}\,\text{rad}\,\text{s}^{-3}$, and plots (b), (d), (f) (right column) with $q_\omega = 10^{16}\,\text{rad}\,\text{s}^{-3}$. The first row (plots (a) and (b)) show the fluctuating field in solid blue juxtaposed with the confidence interval of $\bar{\omega}(t)$ (∂) as well as the upper bound for $|\bar{\omega}(t)|$ which stems from the Taylor expansion of $\langle \hat{J}_x(t) \rangle$. The plots in the second row (subfigures (c) and (d)), compare the exact solution of $\langle \hat{J}_x(t) \rangle$ with its approximation $Je^{-(M+\kappa_c)t/2}$. Finally, in the bottom row, plots (e) and (f) show the error percentage of the approximation of $\langle \hat{J}_x(t) \rangle$.

Hence, $\mathbb{E}[\bar{\omega}(t)] = 0$ and its variance:

$$\mathrm{V}[\bar{\omega}(t)] = \mathbb{E}\left[(\bar{\omega}(t))^2\right] = \mathbb{E}\left[\frac{1}{t}\int_0^t \mathrm{d}\tau_1\,\omega(\tau_1)\,\frac{1}{t}\int_0^t \mathrm{d}\tau_2\,\omega(\tau_2)\right] \tag{E.15}$$

$$= \frac{1}{t^2}\int_0^t \mathrm{d}\tau_1 \int_0^t \mathrm{d}\tau_2\,\mathbb{E}[\omega(\tau_1)\omega(\tau_2)] = \frac{1}{t^2}\int_0^t \mathrm{d}\tau_1\,\mathrm{d}\tau_2\,\frac{q_\omega}{2\chi}\left(e^{-\chi|\tau_1-\tau_2|} - e^{-\chi(\tau_1+\tau_2)}\right) \tag{E.16}$$

$$= \frac{q_\omega}{2\chi t^2}\left[\int_0^t \mathrm{d}\tau_1 \left(\int_0^{\tau_1}\mathrm{d}\tau_2\,e^{-\chi(\tau_1-\tau_2)} + \int_{\tau_1}^t \mathrm{d}\tau_2\,e^{\chi(\tau_1-\tau_2)}\right) - \left(\frac{1-e^{-\chi t}}{\chi}\right)^2\right] \tag{E.17}$$

$$= \frac{q_\omega}{2\chi t^2}\left[\frac{1}{\chi}\int_0^t \mathrm{d}\tau_1\,\left(1 - e^{-\chi\tau_1} - e^{\chi(\tau_1-t)} {}_2 {}_{33} 1\right) - \left(\frac{1-e^{-\chi t}}{\chi}\right)^2\right] \tag{E.18}$$

$$= \frac{q_\omega}{2\chi^3 t^2}\left(4e^{-\chi t} + 2\chi t - e^{-2\chi t} - 3\right). \tag{E.19}$$

If $\chi \ll \frac{4}{3t}$, we can correctly approximate the variance as:

$$\mathrm{V}[\bar{\omega}(t)] = \frac{q_\omega t}{3}. \tag{E.20}$$

So, using the 68-95-99.7 rule and taking the confidence interval for $\bar{\omega}(t)$ to be

$$\left| \mathbb{E}[\bar{\omega}(t)] \pm 2\sqrt{\mathrm{V}[\bar{\omega}(t)]} \right| = 2\sqrt{\frac{q_\omega t}{3}}, \tag{E.21}$$

we can define an inequality that ensures (E.10) to hold with high (95%) probability:

$$2\sqrt{\frac{q_\omega t}{3}} \lesssim \frac{\sqrt{2}}{t}. \tag{E.22}$$

As shown in figure Fig. E.1, the inequality in (E.22) correctly assures the approximation $\langle \hat{J}_x(t) \rangle \approx Je^{-(M+\kappa_c)t/2}$ to hold. Moreover, it gives an upper bound to the value of q_ω,

$$q_\omega \lesssim \frac{3}{2t^3}. \tag{E.23}$$

In summary, the unconditional evolution of $\hat{J}_x$ can be taken to be

$$\langle \hat{J}_x(t) \rangle \approx Je^{-(M+\kappa_c)t/2}, \tag{E.24}$$

if $\chi \leq \frac{4}{3t}$ and $q_\omega \lesssim \frac{3}{2t^3}$. Note that we have renamed $\kappa_z =: \kappa_c$ and dropped κ_y, since it is unnecessary. Namely, by transforming the parameters of the continuous measurement in Eq. (5.16) as follows: $M \to M - \kappa_y$, $\eta \to \eta M/(M - \kappa_y)$ and $y(t) \to y(t)\sqrt{M/(M - \kappa_y)}$; we retrieve the conditional dynamics Eqs. (5.16-5.18) with $\kappa_y = 0$. Hence, the impact of the collective noise introduced via $\mathcal{D}[\hat{J}_y]$ in Eq. (E.1) can always be interpreted and incorporated into a modified form of the continuous measurement Eq. (5.16).

E.2 Conditional dynamics of the variance $\mathrm{V}_y(t)$ in the LG regime

When taking into account possible decoherence mechanisms along the same direction of the magnetic field, the differential equation for the conditional variance of $\mathrm{V}_y(t)$ is shown to be,

$$d\mathrm{V}_y(t) = -4M\eta \mathrm{V}_y(t)dt + \kappa_c J^2 e^{-(M+\kappa_c)t}dt, \tag{E.25}$$

whose solution exists and can be given in terms of modified Bessel functions of first and second kind ($\mathcal{I}_\beta[\cdot]$ and $\mathcal{K}_\beta[\cdot]$), and regularized confluent hypergeometric functions ($_0F_1[\cdot]$). Namely,

$$
V_y(t) = V_e(t) = Je^{-(M+\kappa_c)t/2}\left(\mathcal{I}_1\Big[2\beta\Big]\left(\sqrt{\eta\,\kappa_c\mathcal{M}}\mathcal{K}_0[2\alpha] - \kappa_c\mathcal{K}_1[2\alpha]\right) + \right.
$$

$$
\left. + \mathcal{K}_1\Big[2\beta\Big]\left(\kappa_c\,\mathcal{I}_1[2\alpha] + \sqrt{\eta\,\kappa_c\mathcal{M}}\,\,_0F_1[1,\alpha^2]\right) \right)
$$

$$
\Bigg/ \left(2\,_0F_1[1,\beta^2]\left(\sqrt{\eta\,\kappa_c\mathcal{M}}\mathcal{K}_1[2\alpha] - M\eta\,\mathcal{K}_0[2\alpha]\right) \right.
$$

$$
\left. + \frac{2\eta M}{M+\kappa_c}\mathcal{K}_0[2\beta]\Big((M+\kappa_c)\,_0F_1[1,\alpha^2] + 2\kappa_c J\,_0F_1[2,\alpha^2]\Big) \right), \qquad \text{(E.26)}
$$

where $\alpha = 2J\sqrt{\eta\,\kappa_c\mathcal{M}}/(M+\kappa_c)$ and $\beta = \alpha e^{-(M+\kappa_c)t/2}$. The behavior of the solution in E.26 can be better understood when broken down into different regimes.

$\mathcal{I}_1[2\beta] \approx \dfrac{e^{2\beta}}{2\sqrt{\pi\beta}}$	$\mathcal{K}_0[2\alpha] \approx \dfrac{1}{2}\sqrt{\dfrac{\pi}{\alpha}}e^{-2\alpha}$	$\mathcal{K}_1[2\alpha] \approx \dfrac{1}{2}\sqrt{\dfrac{\pi}{\alpha}}e^{-2\alpha}$
$\mathcal{K}_1[2\beta] \approx \dfrac{1}{2}\sqrt{\dfrac{\pi}{\beta}}e^{-2\beta}$	$\mathcal{I}_1[2\alpha] \approx \dfrac{e^{2\alpha}}{2\sqrt{\pi\alpha}}$	$_0F_1[1,\alpha^2] \approx \dfrac{e^{2\alpha}}{2\sqrt{\pi\alpha}}$
$_0F_1[1,\beta^2] \approx \dfrac{e^{2\beta}}{2\sqrt{\pi\beta}}$	$\mathcal{K}_0[2\beta] \approx \dfrac{1}{2}\sqrt{\dfrac{\pi}{\beta}}e^{-2\beta}$	$_0F_1[2,\alpha^2] \approx \dfrac{e^{2\alpha}}{2\alpha\sqrt{\pi\alpha}}$

Table E.1: **Series expansions of the Bessel functions** for $1/\alpha$ and $1/\beta$ around $1/\alpha_0 = 0$ and $1/\beta_0 = 0$, to leading order.

In order to do so, the first step is to expand the modified Bessel functions and the regularized confluent hypergeometric functions around infinity and only keep the first order, since $\alpha \gg 1$ and $\beta \gg 1$. The relevant expansions are shown in Table E.1.

By then substituting the leading order expansions of Table E.1 to solution E.26 and approxi-
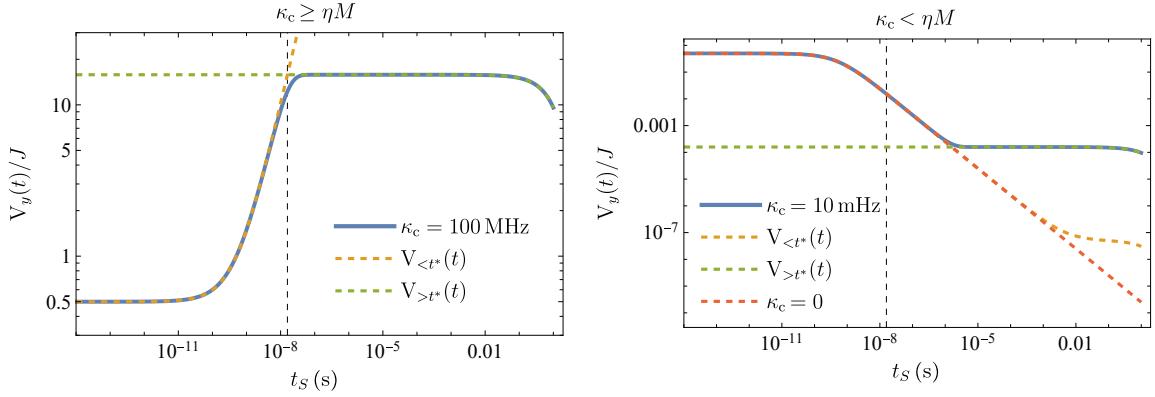
Figure E.2: Strong and weak dephasing regimes for the variance. In the subfigure on the left ($\kappa_{\rm c} = 10\,\mathrm{mHz} < M$), the exact variance solution $\mathrm{V}_y(t)$ is compared to the approximated functions $\mathrm{V}_{<t^*}(t)$ and $\mathrm{V}_{>t^*}(t)$ (dashed green and yellow, respectively). The transition time t^* between these two regimes is marked with a dotted black vertical line. In the second graph (on the right, with $\kappa_{\rm c} = 100\,\mathrm{MHz} > M$), the two different regimes $\mathrm{V}_{<t^*}(t)$ (dashed green) and $\mathrm{V}_{>t^*}(t)$ (dashed yellow) are superimposed with the exact solution of $\mathrm{V}_y(t)$ (in solid blue). Notation t_s refers to a scaled time, $t_s = t(M + \kappa_{\rm c})$. All plots have been generated with $M = 100\,\mathrm{kHz}$, $\eta = 1$, and $J = 10^9$.

mating $\mathrm{e}^{4\beta}$ as $\mathrm{e}^{-2\alpha(-2+t(M+\kappa_{\rm c}))}$, the variance of $\hat{J}_z(t)$ simplifies to

$$\mathrm{V}_y(t) \approx \frac{1}{2} J \mathrm{e}^{-(M+\kappa_{\rm c})t/2} \frac{\sqrt{M\kappa_{\rm c}\eta}\cosh\left(2Jt\sqrt{M\kappa_{\rm c}\eta}\right) + \kappa_{\rm c}\sinh\left(2Jt\sqrt{M\kappa_{\rm c}\eta}\right)}{\sqrt{M\kappa_{\rm c}\eta}\cosh\left(2Jt\sqrt{M\kappa_{\rm c}\eta}\right) + M\eta\sinh\left(2Jt\sqrt{M\kappa_{\rm c}\eta}\right)}. \tag{E.27}$$

Note that if $2Jt\sqrt{M\kappa_{\rm c}\eta} \gg 1$, then, $\cosh\left(2Jt\sqrt{M\kappa_{\rm c}\eta}\right) \approx \frac{1}{2}\mathrm{e}^{2Jt\sqrt{M\kappa_{\rm c}\eta}}$ and $\sinh\left(2Jt\sqrt{M\kappa_{\rm c}\eta}\right) \approx \frac{1}{2}\mathrm{e}^{2Jt\sqrt{M\kappa_{\rm c}\eta}}$, such that,

$$\mathrm{V}_y(t) \approx \mathrm{V}_{>t^*}(t) = \frac{1}{2} J \mathrm{e}^{-(M+\kappa_{\rm c})t/2}\sqrt{\frac{\kappa_{\rm c}}{\eta M}}. \tag{E.28}$$

If $2Jt\sqrt{M\kappa_{\rm c}\eta} \ll 1$, then, $\cosh\left(2Jt\sqrt{M\kappa_{\rm c}\eta}\right) \approx 1$ and $\sinh\left(2Jt\sqrt{M\kappa_{\rm c}\eta}\right) \approx 2Jt\sqrt{M\kappa_{\rm c}\eta}$ such that,

$$\mathrm{V}_y(t) \approx \mathrm{V}_{<t^*}(t) = J\mathrm{e}^{-(M+\kappa_{\rm c})t/2}\frac{(1 + 2Jt\kappa_{\rm c})}{2 + 4JtM\eta}. \tag{E.29}$$

Moreover, note that since $2Jt\sqrt{M\kappa_{\rm c}\eta} \ll 1$, we can then derive the condition

$$t \ll t^* = \frac{1}{2J\sqrt{M\kappa_{\rm c}\eta}} \Rightarrow 2Jt\kappa_{\rm c} \ll \sqrt{\frac{\kappa_{\rm c}}{M\eta}}, \tag{E.30}$$

236

for which E.29 holds. From it, it also follows that if $\kappa_c < M$, then $(1 + 2Jt\kappa_c) \approx 1$ and,

$$V_y(t) \approx \frac{J}{2 + 4JtM\eta}e^{-(M+\kappa_c)t/2} \approx \frac{J}{2 + 4JtM\eta}, \tag{E.31}$$

where in the last step we used the short-time condition $t \ll t^* < (M + \kappa_c)^{-1}$.

Next, by showing that $V_y(t)$ is a non-decreasing function at $t \approx 0$ if $\kappa_c \geq \eta M$, we can prove that (E.31) will hold *if and only if* $\kappa_c < \eta M$ for $J \gg 1$. Namely, that we can consider the global decoherence κ_c to be insignificant for small times $t \ll t^*$ only when $\kappa_c < \eta M$. In order to do so, we take the derivative with respect to time of the function E.29 and then compute the limit for time approaching zero:

$$\lim_{t \to 0} \frac{\mathrm{d}}{\mathrm{d}t} [V_{<t^*}(t)] = -\frac{1}{4}J(M + \kappa_c - 4J\kappa_c + 4JM\eta). \tag{E.32}$$

By then setting the solution above equal to zero, we find the value of κ_c for which the derivative changes signs,

$$\kappa_c = \frac{M + 4JM\eta}{4J - 1} \approx M\eta + \frac{M(\eta + 1)}{4J} + \frac{M(\eta + 1)}{16J^2} + \mathcal{O}\left[\frac{1}{J}\right]^3, \tag{E.33}$$

which can be correctly approximated as $\kappa_c = \eta M$ when $J \gg 1$. Hence,

$$\lim_{t \to 0} \frac{\mathrm{d}}{\mathrm{d}t} [V_{<t^*}(t)] \geq 0 \text{ if } \kappa_c \geq \eta M, \tag{E.34}$$

and

$$\lim_{t \to 0} \frac{\mathrm{d}}{\mathrm{d}t} [V_{<t^*}(t)] < 0 \text{ if } \kappa_c < \eta M. \tag{E.35}$$

Thus proving that the variance of $\hat{J}_y$ can be approximated as (E.31) only when $\kappa_c < \eta M$.

E.3   Derivation of the CoG dynamical model of Eq. (5.28)

The set of stochastic differential equations (5.28) can be derived by carefully applying the rules of Itô calculus, e.g., by noting that the differential of any two functions of time and a stochastic process, f and g, reads $\mathrm{d}(fg) = f\mathrm{d}g + g\mathrm{d}f + \mathrm{d}f\mathrm{d}g$. In our case, these functions are the means, variances and covariances of some quantum observable $\hat{\mathcal{O}}$, whose dynamical evolution can then be computed by substituting the conditional dynamics (5.13) of $\mathrm{d}\rho_{(c)}$ into $\mathrm{d}\langle\hat{\mathcal{O}}\rangle = \mathrm{Tr}\{\hat{\mathcal{O}}\,\mathrm{d}\rho_{(c)}\}$.

In particular, considering $\hat{J}_\alpha$, $V_\alpha^{(c)}$ and $C_{\alpha\beta}^{(c)}$ with $\alpha, \beta = x, y, z$ appearing in Eq. (5.28), which satisfy

$$d\langle \hat{J}_\alpha \rangle_{(c)} = \text{Tr}[\hat{J}_\alpha \, d\rho_{(c)}], \tag{E.36a}$$

$$dV_\alpha^{(c)} = d\langle \hat{J}_\alpha^2 \rangle_{(c)} - d\left(\langle \hat{J}_\alpha \rangle_{(c)}^2 \right) \tag{E.36b}$$

$$= d\langle \hat{J}_\alpha^2 \rangle_{(c)} - 2\langle \hat{J}_\alpha \rangle_{(c)} d\left(\langle \hat{J}_\alpha \rangle_{(c)} \right) - d\langle \hat{J}_\alpha \rangle_{(c)} d\langle \hat{J}_\alpha \rangle_{(c)},$$

$$dC_{\alpha\beta}^{(c)} = \frac{1}{2}d\langle \hat{J}_\alpha \hat{J}_\beta \rangle_{(c)} + \frac{1}{2}d\langle \hat{J}_\beta \hat{J}_\alpha \rangle_{(c)} - d\left(\langle \hat{J}_\alpha \rangle_{(c)} \langle \hat{J}_\beta \rangle_{(c)} \right)$$

$$= \frac{1}{2}d\langle \hat{J}_\alpha \hat{J}_\beta \rangle_{(c)} + \frac{1}{2}d\langle \hat{J}_\beta \hat{J}_\alpha \rangle_{(c)} - \langle \hat{J}_\beta \rangle_{(c)} d\langle \hat{J}_\alpha \rangle_{(c)}$$

$$- \langle \hat{J}_\alpha \rangle_{(c)} d\langle \hat{J}_\beta \rangle_{(c)} - d\langle \hat{J}_\alpha \rangle_{(c)} d\langle \hat{J}_\beta \rangle_{(c)}, \tag{E.36c}$$

with the initial conditions set to the mean, variances and co-variances of a CSS along x (see Sec. 1.4.4):

$$\langle \hat{\boldsymbol{J}}(0) \rangle_{(c)} = \left(\langle \hat{J}_x(0) \rangle_{(c)}, \ \langle \hat{J}_y(0) \rangle_{(c)}, \ \langle \hat{J}_z(0) \rangle_{(c)} \right)^{\text{T}} = \left(\frac{N}{2}, \ 0, \ 0 \right)^{\text{T}}, \tag{E.37}$$

$$\langle \mathbf{C}_J^{(c)}(0) \rangle_{(c)} = \begin{pmatrix} V_x^{(c)}(0) & C_{xy}^{(c)}(0) & C_{xz}^{(c)}(0) \\ C_{xy}^{(c)*}(0) & V_y^{(c)}(0) & C_{zy}^{(c)*}(0) \\ C_{xz}^{(c)*}(0) & C_{zy}^{(c)}(0) & V_z^{(c)}(0) \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & N/4 & 0 \\ 0 & 0 & N/4 \end{pmatrix}. \tag{E.38}$$

By carefully working to the relevant order $O(\mathrm{d}t^{3/2})$:

$$\mathrm{d}\langle\hat{J}_x\rangle_{(c)} = -(\omega(t)+u(t))\langle\hat{J}_y\rangle_{(c)}\mathrm{d}t - \frac{1}{2}(\kappa_c+2\kappa_\ell+M)\langle\hat{J}_x\rangle_{(c)}\mathrm{d}t + 2\sqrt{\eta M}\mathrm{C}^{(c)}_{xy}\,\mathrm{d}W \tag{E.39a}$$

$$\mathrm{d}\langle\hat{J}_y\rangle_{(c)} = (\omega(t)+u(t))\langle\hat{J}_x\rangle_{(c)}\mathrm{d}t - \frac{1}{2}(\kappa_c+2\kappa_\ell)\langle\hat{J}_y\rangle_{(c)}\mathrm{d}t + 2\sqrt{\eta M}\mathrm{V}^{(c)}_y\,\mathrm{d}W \tag{E.39b}$$

$$\mathrm{d}\langle\hat{J}_z\rangle_{(c)} = -\frac{1}{2}M\langle\hat{J}_z\rangle_{(c)}\mathrm{d}t + 2\sqrt{\eta M}\mathrm{C}^{(c)}_{zy}\,\mathrm{d}W \tag{E.39c}$$

$$\mathrm{d}\mathrm{V}^{(c)}_x = -2(\omega(t)+u(t))\mathrm{C}^{(c)}_{xy}\mathrm{d}t + \kappa_c\left(\mathrm{V}^{(c)}_y+\langle\hat{J}_y\rangle^2_{(c)}-\mathrm{V}^{(c)}_x\right)\mathrm{d}t + \kappa_\ell\left(\frac{N}{2}-2\mathrm{V}^{(c)}_x\right)\mathrm{d}t$$
$$+M\left(\mathrm{V}^{(c)}_z-\mathrm{V}^{(c)}_x-4\eta\mathrm{C}^{(c)2}_{xy}\right)\mathrm{d}t + 2\sqrt{\eta M}\left(\frac{1}{2}\mathrm{cov}_{(c)}(\hat{J}^2_x\hat{J}_y)+\frac{1}{2}\mathrm{cov}_{(c)}(\hat{J}_y\hat{J}^2_x)\right)\mathrm{d}W \tag{E.39d}$$

$$\mathrm{d}\mathrm{V}^{(c)}_y = 2(\omega(t)+u(t))\mathrm{C}^{(c)}_{xy}\mathrm{d}t + \kappa_c\left(\mathrm{V}^{(c)}_x+\langle\hat{J}_x\rangle^2_{(c)}-\mathrm{V}^{(c)}_y\right)\mathrm{d}t + \kappa_\ell\left(\frac{N}{2}-2\mathrm{V}^{(c)}_y\right)\mathrm{d}t$$
$$-4\eta M\mathrm{V}^{(c)2}_y\mathrm{d}t + 2\sqrt{\eta M}\,\mathrm{cov}_{(c)}(\hat{J}^3_y)\,\mathrm{d}W \tag{E.39e}$$

$$\mathrm{d}\mathrm{V}^{(c)}_z = M\left(\mathrm{V}^{(c)}_x+\langle\hat{J}_x\rangle^2_{(c)}-\mathrm{V}^{(c)}_z\right)\mathrm{d}t$$
$$+2\sqrt{\eta M}\left(\frac{1}{2}\mathrm{cov}_{(c)}(\hat{J}^2_z\hat{J}_y)+\frac{1}{2}\mathrm{cov}_{(c)}(\hat{J}_y\hat{J}^2_z)\right)\mathrm{d}W \tag{E.39f}$$

$$\mathrm{d}\mathrm{C}^{(c)}_{xy} = (\omega(t)+u(t))\left(\mathrm{V}^{(c)}_x-\mathrm{V}^{(c)}_y\right)\mathrm{d}t - \kappa_c\left(2\mathrm{C}^{(c)}_{xy}+\langle\hat{J}_x\rangle_{(c)}\langle\hat{J}_y\rangle_{(c)}\right)\mathrm{d}t$$
$$-2\kappa_\ell\mathrm{C}^{(c)}_{xy}\mathrm{d}t - \frac{1}{2}M\mathrm{C}^{(c)}_{xy}\left(1+8\eta\mathrm{V}^{(c)}_y\right)\mathrm{d}t$$
$$+2\sqrt{\eta M}\left(\frac{1}{4}\mathrm{cov}_{(c)}(\hat{J}_x\hat{J}^2_y)+\frac{1}{2}\mathrm{cov}_{(c)}(\hat{J}_y\hat{J}_x\hat{J}_y)+\frac{1}{4}\mathrm{cov}_{(c)}(\hat{J}^2_y\hat{J}_x)\right)\mathrm{d}W \tag{E.39g}$$

$$\mathrm{d}\mathrm{C}^{(c)}_{zy} = (\omega(t)+u(t))\mathrm{C}^{(c)}_{xz}\mathrm{d}t - \frac{1}{2}\left(\kappa_c+2\kappa_\ell+M\left(1+8\eta\mathrm{V}^{(c)}_y\right)\right)\mathrm{C}^{(c)}_{zy}\mathrm{d}t$$
$$+2\sqrt{\eta M}\left(\frac{1}{4}\mathrm{cov}_{(c)}(\hat{J}_z\hat{J}^2_y)+\frac{1}{2}\mathrm{cov}_{(c)}(\hat{J}_y\hat{J}_z\hat{J}_y)+\frac{1}{4}\mathrm{cov}_{(c)}(\hat{J}^2_y\hat{J}_z)\right)\mathrm{d}W \tag{E.39h}$$

$$\mathrm{d}\mathrm{C}^{(c)}_{xz} = -(\omega(t)+u(t))\mathrm{C}^{(c)}_{zy}\mathrm{d}t - \frac{1}{2}\left(\kappa_c+2\kappa_\ell+4M\right)\mathrm{C}^{(c)}_{xz}\mathrm{d}t - M\langle\hat{J}_z\rangle_{(c)}\langle\hat{J}_x\rangle_{(c)}\mathrm{d}t$$
$$-4\eta M\mathrm{C}^{(c)}_{xy}\mathrm{C}^{(c)}_{zy}\mathrm{d}t + 2\sqrt{\eta M}\left(\frac{1}{4}\mathrm{cov}_{(c)}(\hat{J}_x\hat{J}_z\hat{J}_y)+\frac{1}{4}\mathrm{cov}_{(c)}(\hat{J}_y\hat{J}_x\hat{J}_z)\right.$$
$$\left.+\frac{1}{4}\mathrm{cov}_{(c)}(\hat{J}_z\hat{J}_x\hat{J}_y)+\frac{1}{4}\mathrm{cov}_{(c)}(\hat{J}_y\hat{J}_z\hat{J}_x)\right)\mathrm{d}W \tag{E.39i}$$

$$\mathrm{d}\omega = -\chi\omega(t)\mathrm{d}t + \sqrt{q_\omega}\mathrm{d}W_\omega \tag{E.39j}$$

where for any three operators $\hat{A}$, $\hat{B}$ and $\hat{C}$, we define

$$\mathrm{cov}_{(c)}(\hat{A}\hat{B}\hat{C}) := \langle\hat{A}\hat{B}\hat{C}\rangle_{(c)} - \langle\hat{A}\rangle_{(c)}\langle\hat{B}\hat{C}\rangle_{(c)} - \langle\hat{B}\rangle_{(c)}\langle\hat{A}\hat{C}\rangle_{(c)}$$
$$- \langle\hat{C}\rangle_{(c)}\langle\hat{A}\hat{B}\rangle_{(c)} + 2\langle\hat{A}\rangle_{(c)}\langle\hat{B}\rangle_{(c)}\langle\hat{C}\rangle_{(c)}. \tag{E.40}$$

We simplify the full dynamical model given in Eq. (E.39) by applying a cut-off approximation that discards the third-order moments and higher. This step is crucial for two reasons: it allows us to construct an EKF and provides a self-contained set of stochastic differential equations describing our sensor. Importantly, the impact of neglecting third-order moments on the CoG model is limited, as these moments appear only within the stochastic terms of the second-order moment dynamics.

Additionally, we omit the differential equations for $\langle\hat{J}_z\rangle_{(c)}$, $C_{xz}^{(c)}$, and $C_{zy}^{(c)}$ in the main text because these quantities remain consistently zero throughout the time evolution. This is due to their initial values being zero (CSS-state conditions: $\langle\hat{J}_z(0)\rangle_{(c)} = C_{xz}^{(c)}(0) = C_{zy}^{(c)}(0) = 0$) and their exclusively decaying dynamics. By disregarding these irrelevant terms, we arrive at the dynamical equations presented in the main text as Eq. (5.28).

Finally, in order to solve the system of SDE numerically, is also convenient to normalize the state and hence the system of SDEs w.r.t. $\sqrt{N}$. Namely,

$$\hat{X} := \hat{J}_x/\sqrt{N} \qquad\qquad \hat{Y} := \hat{J}_y/\sqrt{N}, \tag{E.41}$$

with new variances and covariances being:

$$\langle\Delta^2\hat{X}\rangle_{(c)} = V_x/N \quad \langle\Delta^2\hat{Y}\rangle_{(c)} = V_y/N \quad \langle\Delta^2\hat{Z}\rangle_{(c)} = V_z/N \quad \langle\Delta(\widehat{XY})\rangle_{(c)} = C_{xy}/N. \tag{E.42}$$

Then, the system of SDE simulating the system in the co-moving approximation is:

$$d\langle\hat{X}\rangle_{(c)} = -(\omega(t)+u(t))\langle\hat{Y}\rangle_{(c)}dt - \frac{1}{2}(\kappa_c+2\kappa_\ell+M)\langle\hat{X}\rangle_{(c)}dt$$
$$+ 2\sqrt{\eta MN}\langle\Delta(\widehat{XY})\rangle_{(c)}dW \tag{E.43}$$

$$d\langle\hat{Y}\rangle_{(c)} = (\omega(t)+u(t))\langle\hat{X}\rangle_{(c)}dt - \frac{1}{2}(\kappa_c+2\kappa_\ell)\langle\hat{Y}\rangle_{(c)}dt + 2\sqrt{\eta MN}\langle\Delta^2\hat{Y}\rangle_{(c)}dW \tag{E.44}$$

$$d\langle\Delta^2\hat{X}\rangle_{(c)} = -2(\omega(t)+u(t))\langle\Delta(\widehat{XY})\rangle_{(c)}dt + \kappa_c\left(\langle\Delta^2\hat{Y}\rangle_{(c)}+\langle\hat{Y}\rangle_{(c)}^2-\langle\Delta^2\hat{X}\rangle_{(c)}\right)dt$$
$$+ \kappa_\ell\left(\frac{1}{2}-2\langle\Delta^2\hat{X}\rangle_{(c)}\right)dt + M\left(\langle\Delta^2\hat{Z}\rangle_{(c)}-\langle\Delta^2\hat{X}\rangle_{(c)}-4\eta N\langle\Delta(\widehat{XY})\rangle_{(c)}^2\right)dt \tag{E.45}$$

$$d\langle\Delta^2\hat{Y}\rangle_{(c)} = 2(\omega(t)+u(t))\langle\Delta(\widehat{XY})\rangle_{(c)}dt + \kappa_c\left(\langle\Delta^2\hat{X}\rangle_{(c)}+\langle\hat{X}\rangle_{(c)}^2-\langle\Delta^2\hat{Y}\rangle_{(c)}\right)dt$$
$$+ \kappa_\ell\left(\frac{1}{2}-2\langle\Delta^2\hat{Y}\rangle_{(c)}\right)dt - 4\eta MN\langle\Delta^2\hat{Y}\rangle_{(c)}^2dt \tag{E.46}$$

$$d\langle\Delta^2\hat{Z}\rangle_{(c)} = M\left(\langle\Delta^2\hat{X}\rangle_{(c)}+\langle\hat{X}\rangle_{(c)}^2-\langle\Delta^2\hat{Z}\rangle_{(c)}\right)dt \tag{E.47}$$

$$d\langle\Delta(\widehat{XY})\rangle_{(c)} = (\omega(t)+u(t))\left(\langle\Delta^2\hat{X}\rangle_{(c)}-\langle\Delta^2\hat{Y}\rangle_{(c)}\right)dt - \kappa_c\left(2\langle\Delta(\widehat{XY})\rangle_{(c)}+\langle\hat{X}\rangle_{(c)}\langle\hat{Y}\rangle_{(c)}\right)dt$$
$$- 2\kappa_\ell\langle\Delta(\widehat{XY})\rangle_{(c)}dt - \frac{1}{2}M\langle\Delta(\widehat{XY})\rangle_{(c)}\left(1+8\eta N\langle\Delta^2\hat{Y}\rangle_{(c)}\right)dt \tag{E.48}$$

$$d\omega = -\chi\omega(t)\,dt + \sqrt{q_\omega}\,dW_\omega, \tag{E.49}$$

with the initial values for the means: $\left(\langle\hat{X}(0)\rangle_{(c)},\langle\hat{Y}(0)\rangle_{(c)}\right) = (\sqrt{N}/2, 0)$, the variances and co-variances: $\left(\langle\Delta^2\hat{X}(0)\rangle_{(c)},\langle\Delta^2\hat{Y}(0)\rangle_{(c)},\langle\Delta^2\hat{Z}(0)\rangle_{(c)},\langle\Delta(\widehat{XY})(0)\rangle_{(c)}\right) = (0, 1/4, 1/4, 0)$ and the Larmor frequency $\omega(0) = \mu$. $\langle\Delta(\widetilde{\hat{X}}\hat{Y})\rangle_{(c)}$

E.4   Verification of the CoG approximation

In this section, we simulate the exact dynamics of the density matrix for low values of N using the SME of Eq. (5.13) in order to verify that the approximate evolution of the lowest moments given by Eq. (5.28) correctly captures the system behavior for moderate dephasing and measurement-strength parameters. We observe that the agreement between the full model and the approximate equations improves with increasing atomic number at short timescales, and because the experimental regimes involve large $N \approx 10^5$–10^{13} [13, 14, 15, 16, 17, 63], we subsequently use Eq. (5.28) to simulate the dynamics of the atomic sensor with sufficient accuracy.

Fig. 5.3 shows the architecture of the feedback loop employed in our atomic magnetometry
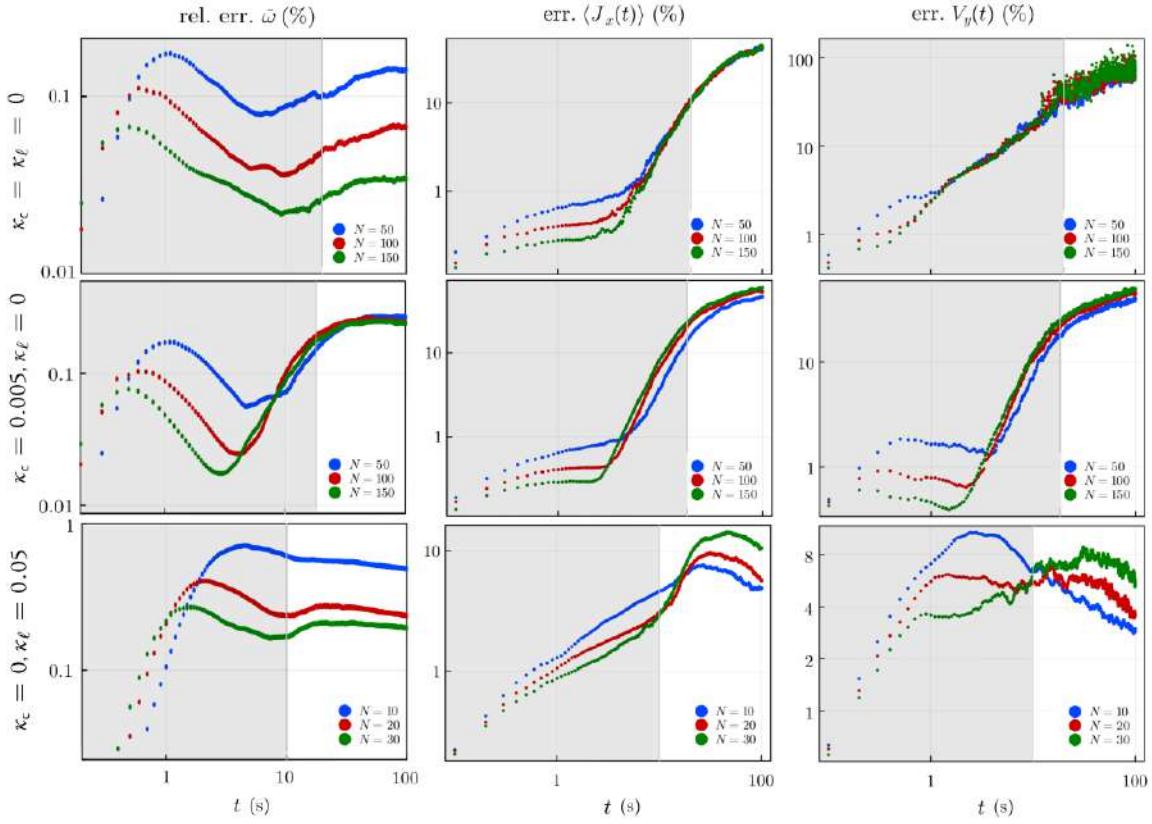
Figure E.3: Performance of the CoG approximation: Larmor frequency estimation and moment error analysis. (Left column) Evolution of the relative error in estimating the Larmor frequency when comparing the exact model (SME, Eq. (5.13)) with the approximate CoG model of Eq. (5.28). Each graph shows the relative error (in %) for three different noise scenarios (from top to bottom): dephasing induced solely by continuous measurement; combined measurement-induced and collective dephasing; and combined measurement-induced, collective, and local dephasing. For the first two cases, system sizes $N = 50, 100, 150$ (blue, red, and green, respectively) are considered, while for the case including local decoherence, $N = 10, 20, 30$ are used. In all cases, the error remains below 1% and decreases with increasing N. **(Center and right columns)** Comparative error analysis of the moments $\left\langle \hat{J}_x(t) \right\rangle_{(c)}$ and $V_y^{(c)}(t)$ between the exact SME solution and the CoG model. Here, the relative error (in %) is defined in Eq. (E.50). The analysis is performed for the same three decoherence scenarios: (top row) continuous measurement only ($M = 0.05, \kappa_c = \kappa_\ell = 0$); (middle row) measurement-induced and collective decoherence ($M = 0.05, \kappa_c = 0.005, \kappa_\ell = 0$); and (bottom row) measurement-induced and local decoherence ($M = 0.05, \kappa_c = 0, \kappa_\ell = 0.05$). In each plot, increasing system sizes (either $N = 50, 100, 150$ or $N = 10, 20, 30$ for the local case) demonstrate that the CoG approximation becomes more accurate at short times as N increases. All error values are obtained by averaging over $\nu = 1000$ measurement trajectories. Figure adapted from Ref. [65].

scheme. In each round, the measurement data $\boldsymbol{y}(t)$ is generated by simulating the "*Atomic ensemble*" either exactly – evolving its full conditional density matrix $\rho_{(c)}$ via the SME (5.13) — or approximately, through the dynamics of its relevant first and second moments, $\left\langle \hat{J}_x \right\rangle_{(c)}$, $\left\langle \hat{J}_y \right\rangle_{(c)}$ and $V_y^{(c)}$, according to the CoG model (5.28). The measurement record generated by the system is then processed by the "*Estimator*" (i.e., the EKF), which provides in real time not only an esti-

mate of the Larmor frequency, $\tilde{\omega}(t)$, but also estimates of the dynamical parameters, $\tilde{\boldsymbol{x}}(t)$. These estimates are used by the "*Controller*" (i.e., the LQR) to adjust the system dynamics on the fly by modifying $u(t)$.

To assess the accuracy of the CoG approximation of Eq. (5.28) when simulating the system dynamics, we benchmark it against the exact SME solution for moderate atomic numbers, where the exact simulation is computationally feasible. This comparison is carried out at two levels: (1) we focus solely on the estimation task by computing the average error in the real-time estimate $\tilde{\omega}(t)$ (large box in Fig. 5.3), and (2) we adopt a stricter criterion by requiring that the relevant moments, $\langle \hat{J}_x \rangle_{(c)}$, $\langle \hat{J}_y \rangle_{(c)}$, and $V_y^{(c)}$, are accurately reproduced when compared to their exact values obtained from evolving $\rho_{(c)}(t)$ with the SME(smaller box in Fig. 5.3).

In the left column of Fig. E.3, we present in percentage the *average relative error* between the real-time estimate $\tilde{\omega}(t)$ of ω obtained using the exact model (full SME solution) and the approximate model (CoG), i.e.

$$
\begin{aligned}
\mathbb{E}[\delta_{\tilde{\omega}}]\,(\%) &= 100 \times \mathbb{E}\left[\left|\frac{\tilde{\omega}_{\mathrm{SME}} - \tilde{\omega}_{\mathrm{CoG}}}{\tilde{\omega}_{\mathrm{SME}}}\right|\right] \\
&= 100 \times \int \mathrm{d}\omega\, p(\omega) \int \mathcal{D}\boldsymbol{y}_{\leq t}\, p(\boldsymbol{y}_{\leq t}|\omega) \left|\frac{\tilde{\omega}_{\mathrm{SME}} - \tilde{\omega}_{\mathrm{CoG}}}{\tilde{\omega}_{\mathrm{SME}}}\right|
\end{aligned}
\tag{E.50}
$$

where the expectation is taken over the realizations of the experiment. In the left column of Fig. E.3, three plots corresponding to a different noise scenario are showcased. From top to bottom: only measurement decoherence ($\kappa_c = \kappa_\ell = 0$), combined measurement and collective decoherence ($M = 0.05$, $\kappa_c = 0.005$ and $\kappa_\ell = 0$), and combined measurement and local decoherence ($M = \kappa_\ell = 0.05$ and $\kappa_c = 0.005$). Each plot shows the averaged relative error for increasing system sizes — specifically, $N = 50, 100, 150$ for the first two scenarios, and $N = 10, 20, 30$ for the local case. In each case the average relative error decreases as the system size increases, reaching below 1% error, which indicates that for large ensembles ($N \sim 10^5 - 10^{13}$) the CoG approximation is sufficiently accurate for generating the measurement data used to estimate the Larmor frequency.

We further assess the CoG model by comparing the evolution of key dynamical moments ($\langle \hat{J}_x \rangle_{(c)}$, $\langle \hat{J}_y \rangle_{(c)}$, and $V_y^{(c)}$) to their exact values computed from the full density matrix $\rho_{(c)}$ (via the SME, Eq. (1)). In this comparison, the error is quantified as

$$
\mathbb{E}[\delta_x] = 100 \times \frac{\mathbb{E}[|x_{\mathrm{SME}} - x_{\mathrm{CoG}}|]}{\mathbb{E}[|x_{\mathrm{SME}}|]} \quad \text{for} \quad x \in \{\langle \hat{J}_x \rangle_{(c)}, \langle \hat{J}_y \rangle_{(c)}, V_y^{(c)}\},
\tag{E.51}
$$

where x stands for any of the moments. As illustrated in the center and right columns of Fig. E.3, for short times the error in simulating these moments decreases with increasing system size, and remains below approximately 10% for times $t \lesssim 1/(M + \kappa_c + 2\kappa_\ell)$. This supports the validity of using the CoG approximation to predict quantities such as the spin-squeezing parameter in large atomic ensembles.

E.5 Steady-state solution of the Kalman Filter for $\chi \neq 0$ in the LG regime

Let us consider the covariance differential equation introduced in (5.36), with matrices $\boldsymbol{F}(t)$, $\boldsymbol{G}(t)$, and $\boldsymbol{H}(t)$ defined in (5.23) and noise covariance matrices being $\boldsymbol{Q} = \mathrm{Diagonal}[1, q_\omega]$, $\boldsymbol{R} = \eta$, and $\boldsymbol{S} = (\sqrt{\eta} \quad 0)^{\mathrm{T}}$.

Next, for simplicity in the upcoming analysis, we will rename the elements of the covariance matrix $\boldsymbol{\Sigma}$ as $\Delta^2(\tilde{\hat{J}}_y)_t \equiv x(t)$, $\Delta(\tilde{\hat{J}}_y \tilde{\omega})_t = \Delta(\tilde{\omega}\tilde{\hat{J}}_y)_t \equiv y(t)$, and $\Delta^2\tilde{\omega}_t \equiv z(t)$. Then, the system of equations in the steady state ($d\boldsymbol{\Sigma}_t = 0$) can be written as,

$$-8M\eta V_{>t^*}(t)\,x(t) - 4M\eta\,x^2(t) - 2Je^{-(M+\kappa_c)t/2}\,y(t) = 0, \qquad (\mathrm{E.52})$$

$$-\chi y(t) - 4M\eta V_{>t^*}(t)\,y(t) - 4M\eta x(t)y(t) - Je^{-(M+\kappa_c)t/2}z(t) = 0, \qquad (\mathrm{E.53})$$

$$q_\omega - 4M\eta y^2(t) - 2\chi z(t) = 0, \qquad (\mathrm{E.54})$$

where the variance $\left\langle \Delta^2 \hat{J}_y(t) \right\rangle_{(c)}$ in the steady state ($t \gg t^*$) is $\left\langle \Delta^2 \hat{J}_y(t) \right\rangle_{(c)} = V_{>t^*}(t)$ (5.24b). It follows that the solution for $z(t)$ is

$$
\begin{aligned}
z(t) = {} & -\kappa_c\chi - \frac{\chi^3}{4J^2 M\eta}e^{(M+\kappa_c)t} - \frac{\chi}{J\sqrt{M\eta}}\sqrt{q_\omega + \kappa_c\chi^2}\,e^{(M+\kappa_c)t/2} \\
& + \frac{1}{2JM\eta}\sqrt{q_\omega + \kappa_c\chi^2}\left(\sqrt{M\eta} + \frac{\chi^2 e^{(M+\kappa_c)t/2}}{2J\sqrt{q_\omega + \kappa_c\chi^2}}\right) \times \\
& \times \sqrt{\chi^2 e^{(M+\kappa_c)t} + 4J\left(\kappa_c JM\eta + e^{(M+\kappa_c)t/2}\sqrt{M\eta(q_\omega + \kappa_c\chi^2)}\right)}, \qquad (\mathrm{E.55})
\end{aligned}
$$

which, when expanded in powers of $1/J$ around zero and truncated at first order, can be approximated as

$$\Delta^2\tilde{\omega}^{\mathrm{SS}} \equiv z(t) \approx -\kappa_c\chi + \sqrt{\kappa_c q_\omega + \kappa_c^2\chi^2}. \qquad (\mathrm{E.56})$$

Moreover, when $\chi = 0$, the term (E.55) becomes,

$$\Delta^2 \tilde{\omega}^{\mathrm{SS}}(t)|_{\chi=0} = \left(q_\omega \, \kappa_c + \frac{1}{J} \sqrt{\frac{q_\omega^3}{M\eta}} \, e^{(M+\kappa_c)t/2} \right)^{1/2}, \tag{E.57}$$

matching the solution introduced in (5.46).

E.6 Derivation of the LQR control law

In this section we present a detailed derivation, along the lines of Ref. [55], of the optimal control law used in our LQR design. The final control law is given by

$$u(t) = -\tilde{\omega}(t) - \lambda \left\langle \hat{\tilde{J}}_y(t) \right\rangle_{(c)}, \tag{E.58}$$

where the constant parameter $\lambda = \sqrt{\dfrac{p_J}{\nu}}$ is determined by the weights in the quadratic cost function. Our starting point is the linearized system dynamics in the LG regime. In this regime the state of interest is the reduced vector

$$\boldsymbol{z}(t) = \begin{pmatrix} \left\langle \hat{J}_y(t) \right\rangle_{(c)} \\ \omega(t) \end{pmatrix}, \tag{E.59}$$

which evolves according to

$$\dot{\boldsymbol{z}}(t) = \boldsymbol{A}\, \boldsymbol{z}(t) + \boldsymbol{B}\, u(t) + \boldsymbol{G}(t)\, \boldsymbol{q}(t), \tag{E.60}$$

with

$$\boldsymbol{A} = \begin{pmatrix} 0 & J \\ 0 & -\chi \end{pmatrix}, \quad \boldsymbol{B} = \begin{pmatrix} J \\ 0 \end{pmatrix}. \tag{E.61}$$

The performance of the controller is measured by the quadratic cost function

$$I(u) = \int_0^\infty \left[\boldsymbol{z}^{\mathrm{T}}(t)\, \boldsymbol{P}\boldsymbol{z}(t) + \nu\, u^2(t) \right] \mathrm{d}t, \tag{E.62}$$

where

$$\boldsymbol{P} = \begin{pmatrix} p_J & 0 \\ 0 & p_\omega \end{pmatrix} \geq 0, \quad \nu > 0.$$

For such a linear system with cost (E.62), the optimal state feedback control law takes the form

$$u(t) = -\boldsymbol{K}_C \tilde{\boldsymbol{z}}(t), \tag{E.63}$$

with the gain matrix computed as

$$\boldsymbol{K}_C = \nu^{-1} \boldsymbol{B}^{\mathrm{T}} \boldsymbol{\Lambda}, \tag{E.64}$$

where $\boldsymbol{\Lambda}$ is the unique positive semidefinite solution of the algebraic Riccati equation (ARE)

$$\boldsymbol{A}^{\mathrm{T}} \boldsymbol{\Lambda} + \boldsymbol{\Lambda} \boldsymbol{A} + \boldsymbol{P} - \boldsymbol{\Lambda} \boldsymbol{B} \nu^{-1} \boldsymbol{B}^{\mathrm{T}} \boldsymbol{\Lambda} = 0. \tag{E.65}$$

We now parameterize $\boldsymbol{\Lambda}$ as a symmetric 2×2 matrix:

$$\boldsymbol{\Lambda} = \begin{pmatrix} \Lambda_{11} & \Lambda_{12} \\ \Lambda_{12} & \Lambda_{22} \end{pmatrix} \tag{E.66}$$

in order to solve the ARE Eq. (E.65) component by component. Focusing first on the $(1, 1)$ entry, we obtain

$$0 + p_J - \frac{J^2}{\nu} \Lambda_{11}^2 = 0 \implies \Lambda_{11} = \frac{\sqrt{p_J \nu}}{J}. \tag{E.67}$$

Next, the $(1, 2)$ entry of Eq. (E.65) yields

$$J \Lambda_{11} - \chi \Lambda_{12} - \frac{J^2}{\nu} \Lambda_{11} \Lambda_{12} = 0 \implies \Lambda_{12} = \frac{J \Lambda_{11}}{\chi + \dfrac{J^2}{\nu} \Lambda_{11}} = \frac{\sqrt{p_J \nu}}{\chi + J \sqrt{\dfrac{p_J}{\nu}}}, \tag{E.68}$$

where in the last step we substituted the expression from Eq. (E.67) into Eq. (E.68). The $(2, 2)$ entry of Eq. (E.65) is

$$2 \left(J \Lambda_{12} - \chi \Lambda_{22} \right) + p_\omega - \frac{J^2}{\nu} \Lambda_{12}^2 = 0, \tag{E.69}$$

which can be rearranged to solve for Λ_{22}:

$$\Lambda_{22} = \frac{1}{2\chi} \left(2J\Lambda_{12} + p_\omega - \frac{J^2}{\nu} \Lambda_{12}^2 \right), \tag{E.70}$$

which fully determines the explicit form of Λ_{22}, even though it is not required for the final control law. Having determined the components of $\boldsymbol{\Lambda}$ in the steady state, we now compute the gain matrix $\boldsymbol{K}_C$ from Eq. (E.64). Given the form of $\boldsymbol{B}$ of Eq. (E.61), we have

$$\boldsymbol{K}_C = \frac{1}{\nu} \boldsymbol{B}^{\mathrm{T}} \boldsymbol{\Lambda} = \frac{J}{\nu} \begin{pmatrix} \Lambda_{11} & \Lambda_{12} \end{pmatrix}. \tag{E.71}$$

Thus, the state feedback control law (E.63) can be written as

$$\begin{aligned} u(t) &= -\boldsymbol{K}_C \tilde{\boldsymbol{z}}(t) \\ &= -\frac{J\Lambda_{11}}{\nu} \left\langle \hat{\tilde{J}}_y(t) \right\rangle_{(c)} - \frac{J\Lambda_{12}}{\nu} \tilde{\omega}(t). \end{aligned} \tag{E.72}$$

It is straightforward to verify that using Eq. (E.67) and Eq. (E.68) we obtain

$$\frac{J\Lambda_{11}}{\nu} = \frac{J}{\nu} \cdot \frac{\sqrt{p_J \nu}}{J} = \sqrt{\frac{p_J}{\nu}} = \lambda, \tag{E.73}$$

$$\frac{J\Lambda_{12}}{\nu} = \frac{1}{1 + \frac{\chi}{J}\sqrt{\frac{\nu}{p_J}}} = \frac{1}{1 + \frac{\chi}{J\lambda}} \approx 1, \tag{E.74}$$

where the second term has been approximated to 1 since under typical experimental conditions, χ is small compared to J. Thus, the control law Eq. (E.72) reduces to the desired form:

$$u(t) = -\tilde{\omega}(t) - \lambda \left\langle \hat{\tilde{J}}_y(t) \right\rangle_{(c)}. \tag{E.75}$$

E.7 Experimentally realistic parameters

Sec. 5.4.2.2 and Sec. 5.4.2.3 analyze the estimation of a fluctuating and time-varying magnetic field, respectively, in experimentally realistic conditions, using parameters inspired by Ref. [63]. Most parameters are straightforward to define: we take the atomic ensemble size as $N = 10^{13}$ and set the coherence time to $T_2 = 10$ ms, which implies a local dephasing rate of $\kappa_\ell = 1/T_2 = 100$ Hz. We further consider frequencies on the order of kHz, specifically $\omega_0 = 10^4 \mathrm{rad\, s}^{-1}$. The
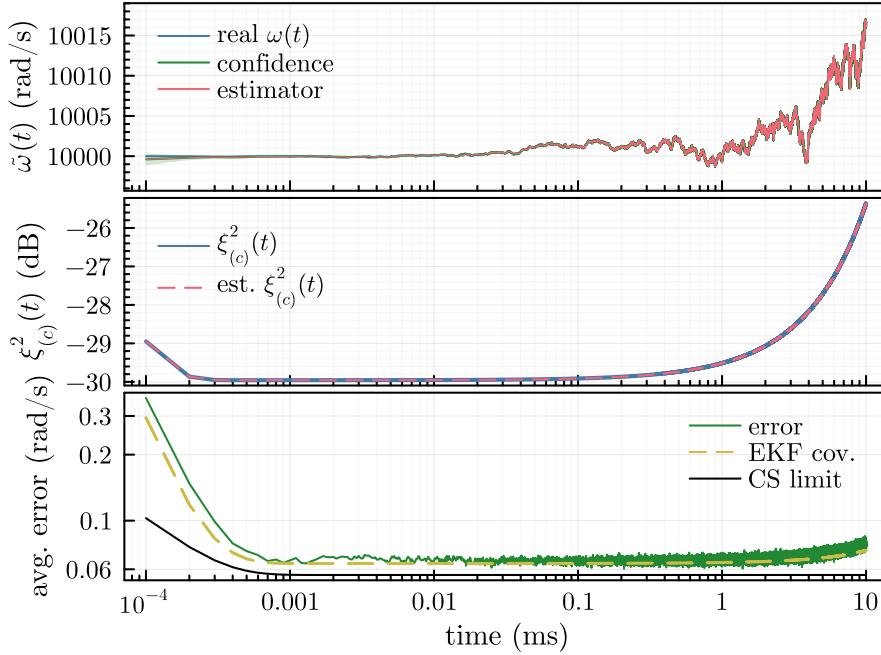
Figure E.4: Tracking a fluctuating magnetic field with a strong continuous measurement. The *top plot* illustrates how the EKF estimate (solid, red) closely follows the true field OUP dynamics (solid, blue), staying well within the error bounds of $\pm 2\sqrt{\mathbb{E}\left[\Delta^2 \tilde{\omega}(t)\right]}$, which are so small compared to the fluctuating field that they are nearly imperceptible. The *middle plot* shows the conditional spin squeezing (solid, blue) induced by the strong continuous measurement of strength $M = 1\,\text{mHz} \gg \kappa_c = 1\,\text{nHz}$ and the LQR feedback, along with its real-time estimation by the EKF (dashed red). In the *bottom plot*, the estimation error of $\omega(t)$ (solid, green) reaches a sensitivity of $\sim 0.066\,\text{rad s}^{-1}$ that matches the square-root of the EKF covariance (dashed, yellow). While the stronger measurement significantly enhances precision, the quantum limit dictated by dephasing (solid, black) at roughly $0.056\,\text{rad s}^{-1}$ is not perfectly saturated. A further increase in M could bring the error closer to this optimal limit [65]. The results in the bottom two plots are obtained by averaging over 1000 stochastic field-atom trajectories.

one parameter that is trickier to determine is the measurement strength parameter M. To establish M, we compare the equation for the photocurrent of Eq. (5.14) with Equation 18 from Ref. [63]. Namely, since the Wiener differential in Eq. (5.14) has a variance of $\mathrm{d}t$, then Equation 18 from Ref. [63] should be normalized by $\sqrt{\eta q_c^2 \dot{N}}$. Expressing it in our notation, the measurement equation of Ref. [63] can be then rewritten as:

$$I(t)\mathrm{d}t = \eta \sqrt{g^2 \dot{N}} \left\langle \hat{J}_y(t) \right\rangle_{(c)} \mathrm{d}t + \sqrt{\eta}\,\mathrm{d}\mathrm{W}, \qquad (\text{E.76})$$

where $\mathbb{E}\left[\mathrm{d}\mathrm{W}^2\right] = \mathrm{d}t$, and we write $\hat{J}_y$ instead of $\hat{J}_z$ to account for the different experimental geometry. By directly comparing this to Eq. (5.14), we identify the measurement strength pa-

rameter as:

$$M = \frac{g^2 \dot{N}}{4}. \tag{E.77}$$

Here, $\dot{N}$ represents the photon flux, given by:

$$\dot{N} = \frac{P}{2\pi\hbar\nu} \tag{E.78}$$

where P is the probe power, varying between $0.5\,\mathrm{mW}$ and $2\,\mathrm{mW}$, and ν is the frequency of the probe light, detuned by $\Delta\nu = \nu_{D_1 Rb} - \nu$ from the Rb D_1 transition at $\nu_{D_1 Rb} = c/\lambda_{D_1 Rb}$ of $794.8\,\mathrm{nm}$. The coupling constant g in Eq. (E.77) is defined in Ref. [63] as:

$$g \approx \frac{c\, r_e f_{osc}}{A_{eff}} \frac{1}{\Delta\nu} \tag{E.79}$$

where c is the speed of light, $r_e = 2.82 \times 10^{-13}\,\mathrm{cm}$ is the classical electron radius, $f_{osc} = 0.34$ is the oscillator strength for the Rb D_1 transition, and $A_{eff} = 0.0503\,\mathrm{cm}^2$ is the effective beam area. As a result, M is expected to lie within the range of $1 \times 10^{-10}\,\mathrm{Hz}$ and $1 \times 10^{-8}\,\mathrm{Hz}$, depending on the probe power P and optical detuning $\Delta\nu$, which can vary from $\Delta\nu \approx 24\,\mathrm{GHz}$ to $64\,\mathrm{GHz}$ when off-resonance.

Physically, the measurement strength M characterizes the balance between the light-atom interaction to the photon shot-noise in the detection process of Eq. (E.76). However, quantum backaction from continuous measurement unfolds on a timescale dictated by $M' = MN$, rather than M alone. This arises because in the quantum model of Eq. (5.13), the variance of the relevant spin operators decays with an effective rate $1/M'$. This behavior has been rigorously established in both decoherence-free cases [54] and scenarios with collective noise [69, 65], and is further corroborated here (see Fig. 5.10), as well as experimentally [63].

E.8 Exact forms of gradient matrices $F(t)$, $G(t)$ and $H(t)$ for an OUP

$$\tilde{x} = \begin{pmatrix}\tilde{x}_1 & \tilde{x}_2 & \tilde{x}_3 & \tilde{x}_4 & \tilde{x}_5 & \tilde{x}_6 & \tilde{x}_7\end{pmatrix} = \begin{pmatrix}\langle\tilde{\hat{X}}\rangle_{(c)} & \langle\tilde{\hat{Y}}\rangle_{(c)} & \langle\Delta^2\hat{X}\rangle_{(c)} & \langle\Delta^2\hat{Y}\rangle_{(c)} & \langle\Delta^2\hat{Z}\rangle_{(c)} & \langle\Delta(\widetilde{XY})\rangle_{(c)} & \tilde{\omega}\end{pmatrix},$$

(E.80)

$$F(t) = \nabla_x f\big|_{(\tilde{x},u,0,t)} =$$

$$=\begin{pmatrix}
-(\kappa_c+2\kappa_\ell+M)/2 & -(\tilde{x}_7+u) & 0 & 0 & 0 & 0 & -\tilde{x}_2\\[4pt]
(\tilde{x}_7+u) & -(\kappa_c+2\kappa_\ell)/2 & 0 & 0 & 0 & 0 & \tilde{x}_1\\[4pt]
0 & 2\kappa_c\tilde{x}_2 & -(\kappa_c+2\kappa_\ell+M) & \kappa_c & M & -2(\tilde{x}_7+u)-8\eta MN\tilde{x}_6 & -2\tilde{x}_6\\[4pt]
2\kappa_c\tilde{x}_1 & 0 & \kappa_c & -\kappa_c-2\kappa_\ell-8\eta MN\tilde{x}_4 & 0 & 2(\tilde{x}_7+u) & 2\tilde{x}_6\\[4pt]
2M\tilde{x}_1 & 0 & M & 0 & -M & 0 & 0\\[4pt]
-\kappa_c\tilde{x}_2 & -\kappa_c\tilde{x}_1 & \tilde{x}_7+u & -(\tilde{x}_7+u)-4\eta MN\tilde{x}_6 & -M & -\left(2\kappa_c+2\kappa_\ell+\tfrac{M}{2}\right)-4\eta MN\tilde{x}_4 & \tilde{x}_3-\tilde{x}_4\\[4pt]
0 & 0 & 0 & 0 & 0 & 0 & -\chi
\end{pmatrix},$$

$$G(t) = \nabla_\xi f\big|_{\tilde{x}} = \begin{pmatrix}
2\sqrt{\eta MN}\,\tilde{x}_6 & 0\\
2\sqrt{\eta MN}\,\tilde{x}_4 & 0\\
0 & 0\\
0 & 0\\
0 & 0\\
0 & 0\\
0 & \sqrt{q_K}
\end{pmatrix},\qquad
Q = \begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix},\qquad
S = \begin{pmatrix}\sqrt{\eta}\\ 0\end{pmatrix},\qquad R = \eta,$$

(E.81)

$$H = \nabla_x b = 2\eta\sqrt{MN}\begin{pmatrix}0 & 1 & 0 & 0 & 0 & 0 & 0\end{pmatrix}.$$

(E.82)

$$\tilde{\boldsymbol{x}} = \begin{pmatrix} \tilde{x}_1 & \tilde{x}_2 & \tilde{x}_3 & \tilde{x}_4 & \tilde{x}_5 & \tilde{x}_6 & \tilde{x}_7 & \tilde{x}_8 & \tilde{x}_9 \end{pmatrix} = \begin{pmatrix} \langle\hat{X}\rangle_{(c)} & \langle\hat{Y}\rangle_{(c)} & \langle\Delta^2\hat{X}\rangle_{(c)} & \langle\Delta^2\hat{Y}\rangle_{(c)} & \langle\Delta^2\tilde{Z}\rangle_{(c)} & \langle\Delta(\widehat{XY})\rangle_{(c)} & \tilde{\nu} & \tilde{\omega} & \tilde{v} \end{pmatrix},$$ (E.83)

$$\boldsymbol{F}(t) = \nabla_{\boldsymbol{x}}\boldsymbol{f}\big|_{(\tilde{\boldsymbol{x}},u,0,t)} =$$

with columns labelled $\big(\langle\hat{X}\rangle_{(c)}\;\; \langle\hat{Y}\rangle_{(c)}\;\; \langle\Delta^2\hat{X}\rangle_{(c)}\;\; \langle\Delta^2\hat{Y}\rangle_{(c)}\;\; \langle\Delta^2\tilde{Z}\rangle_{(c)}\;\; \langle\Delta(\widehat{XY})\rangle_{(c)}\;\; \tilde{\nu}\;\; \tilde{\omega}\;\; \tilde{v}\big)$

$$=$$

$$\begin{pmatrix}
-\tfrac{(\kappa_c+2\kappa_\ell+M)}{2} & -(\tilde{x}_7+u) & 0 & 0 & 0 & 0 & 0 & -\tilde{x}_2 & 0 \\[4pt]
(\tilde{x}_7+u) & -\tfrac{(\kappa_c+2\kappa_\ell)}{2} & 0 & 0 & 0 & 0 & 0 & \tilde{x}_1 & 0 \\[4pt]
0 & 2\kappa_c\tilde{x}_2 & -(\kappa_c+2\kappa_\ell+M) & \kappa_c & M & -2(\tilde{x}_7+u)-8\eta MN\tilde{x}_6 & 0 & -2\tilde{x}_6 & 0 \\[4pt]
2\kappa_c\tilde{x}_1 & 0 & \kappa_c & -\kappa_c-2\kappa_\ell-8\eta MN\tilde{x}_4 & 0 & 2(\tilde{x}_7+u) & 0 & 2\tilde{x}_6 & 0 \\[4pt]
2M\tilde{x}_1 & 0 & M & 0 & -M & 0 & 0 & 0 & 0 \\[4pt]
-\kappa_c\tilde{x}_2 & -\kappa_c\tilde{x}_1 & \tilde{x}_7+u & -(\tilde{x}_7+u)-4\eta MN\tilde{x}_6 & 0 & -\left(2\kappa_c+2\kappa_\ell+\tfrac{M}{2}\right)-4\eta MN\tilde{x}_4 & 0 & \tilde{x}_3-\tilde{x}_4 & 0 \\[4pt]
0 & 0 & 0 & 0 & 0 & 0 & k_K/m_K & \dfrac{-p_K}{m_K} & -2c_K\tilde{x}_4/m_K \\[6pt]
0 & 0 & 0 & 0 & 0 & 0 & -\dfrac{1+\tilde{x}_3/|\tilde{x}_3|}{2T_K} & \dfrac{2c_K(1-\tilde{x}_5)}{2T_K} & 0 \\[6pt]
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -\dfrac{1}{T_K}
\end{pmatrix},$$

$$\boldsymbol{G}(t) = \nabla_{\boldsymbol{\xi}}\boldsymbol{f}\big|_{\tilde{\boldsymbol{x}}} = \begin{pmatrix}
2\sqrt{\eta MN}\,\tilde{x}_6 & 0 \\
2\sqrt{\eta MN}\,\tilde{x}_4 & 0 \\
0 & 0 \\
0 & 0 \\
0 & 0 \\
0 & 0 \\
0 & 0 \\
0 & \sqrt{q_K} \\
0 & 0
\end{pmatrix}, \qquad \boldsymbol{Q} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \qquad \boldsymbol{S} = \begin{pmatrix} \sqrt{\eta} \\ 0 \end{pmatrix}, \qquad \boldsymbol{R} = \eta,$$ (E.84)

$$\boldsymbol{H} = \nabla_{\boldsymbol{x}}\boldsymbol{b} = 2\eta\sqrt{MN}\begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$ (E.85)

References

[1] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory* (Prentice Hall, 1993).

[2] H. L. V. Trees and K. L. Bell, *Bayesian Bounds for Parameter Estimation and Nonlinear Filtering/Tracking* (Wiley-IEEE Press, 2007).

[3] M. Born, *"Zur quantenmechanik der stoßvorgänge,"* Zeitschrift für Physik **37**, 863 (1926).

[4] C. M. Caves, *"Quantum-mechanical noise in an interferometer,"* Phys. Rev. D **23**, 1693 (1981).

[5] D. J. Wineland, J. J. Bollinger, W. M. Itano, F. L. Moore, and D. J. Heinzen, *"Spin squeezing and reduced quantum noise in spectroscopy,"* Phys. Rev. A **46**, R6797 (1992).

[6] D. J. Wineland, J. J. Bollinger, W. M. Itano, and D. J. Heinzen, *"Squeezed atomic states and projection noise in spectroscopy,"* Phys. Rev. A **50**, 67 (1994).

[7] J. J. . Bollinger, W. M. Itano, D. J. Wineland, and D. J. Heinzen, *"Optimal frequency measurements with maximally correlated states,"* Phys. Rev. A **54**, R4649 (1996).

[8] G. M. D'Ariano, P. Lo Presti, and M. G. A. Paris, *"Using entanglement improves the precision of quantum measurements,"* Phys. Rev. Lett. **87**, 270404 (2001).

[9] V. Giovannetti, S. Lloyd, and L. Maccone, *"Quantum-enhanced positioning and clock synchronization,"* Nature **412**, 417 (2001).

[10] V. Giovannetti, S. Lloyd, and L. Maccone, *"Quantum-enhanced measurements: Beating the Standard Quantum Limit,"* Science **306**, 1330 (2004).

[11] V. Giovannetti, S. Lloyd, and L. Maccone, *"Quantum metrology,"* Phys. Rev. Lett. **96**, 010401 (2006).

[12] L. Pezzè, A. Smerzi, M. K. Oberthaler, R. Schmied, and P. Treutlein, *"Quantum metrology with nonclassical states of atomic ensembles,"* Rev. Mod. Phys. **90**, 035005 (2018).

[13] A. Kuzmich, L. Mandel, and N. P. Bigelow, *"Generation of spin squeezing via continuous quantum nondemolition measurement,"* Phys. Rev. Lett. **85**, 1594 (2000).

[14] W. Wasilewski, K. Jensen, H. Krauter, J. J. Renema, M. V. Balabas, and E. S. Polzik, *"Quantum noise limited and entanglement-assisted magnetometry,"* Phys. Rev. Lett. **104**, 133601 (2010).

[15] V. Shah, G. Vasilakis, and M. V. Romalis, *"High bandwidth atomic magnetometry with continuous quantum nondemolition measurements,"* Phys. Rev. Lett. **104**, 013601 (2010).

[16] M. Koschorreck, M. Napolitano, B. Dubost, and M. W. Mitchell, *"Sub-projection-noise sensitivity in broadband atomic magnetometry,"* Phys. Rev. Lett. **104**, 093602 (2010).

[17] R. J. Sewell, M. Koschorreck, M. Napolitano, B. Dubost, N. Behbood, and M. W. Mitchell, *"Magnetic sensitivity beyond the projection noise limit by spin squeezing,"* Phys. Rev. Lett. **109**, 253605 (2012).

[18] B. M. Escher, R. L. de Matos Filho, and L. Davidovich, *"General framework for estimating the ultimate precision limit in noisy quantum-enhanced metrology,"* Nature Phys. **7**, 406 (2011).

[19] R. Demkowicz-Dobrzański, J. Kołodyński, and M. Guţă, *"The elusive Heisenberg limit in quantum-enhanced metrology,"* Nat. Commun. **3**, 1063 (2012).

[20] M. W. Mitchell, J. S. Lundeen, and A. M. Steinberg, *"Super-resolving phase measurements with a multiphoton entangled state,"* Nature **429**, 161 (2004).

[21] T. Nagata, R. Okamoto, J. L. O'Brien, K. Sasaki, and S. Takeuchi, *"Beating the standard quantum limit with four-entangled photons,"* Science **316**, 726 (2007).

[22] M. Pannetier-Lecoeur, L. Parkkonen, N. Sergeeva-Chollet, H. Polovy, C. Fermon, and C. Fowley, *"Magnetocardiography with sensors based on giant magnetoresistance,"* Appl. Phys. Lett. **98**, 153705 (2011).

[23] G. Bison, N. Castagna, A. Hofer, P. Knowles, J.-L. Schenker, M. Kasprzak, H. Saudan, and A. Weis, *"A room temperature 19-channel magnetic field mapping device for cardiac signals,"* Appl. Phys. Lett. **95**, 173701 (2009).

[24] K. Jensen, M. A. Skarsfeldt, H. Stærkind, J. Arnbak, M. V. Balabas, S.-P. Olesen, B. H. Bentzen, and E. S. Polzik, *"Magnetocardiography on an isolated animal heart with a room-temperature optically pumped magnetometer,"* Sci. Rep. **8**, 16218 (2018).

[25] Y. J. Kim, I. Savukov, and S. Newman, *"Magnetocardiography with a 16-channel fiber-coupled single-cell rb optically pumped magnetometer,"* Applied Physics Letters **114**, 143702 (2019).

[26] Y. Yang, M. Xu, A. Liang, Y. Yin, X. Ma, Y. Gao, and X. Ning, *"A new wearable multichannel magnetocardiogram system with a serf atomic magnetometer array,"* Scientific Reports **11**, 5564 (2021).

[27] R. Hari and R. Salmelin, *"Magnetoencephalography: From squids to neuroscience: Neuroimage 20th anniversary special edition,"* NeuroImage **61**, 386 (2012).

[28] J. Jung, J. Park, J. Choi, and H.-T. Choi, *"Autonomous mapping of underwater magnetic fields using a surface vehicle,"* IEEE Access **6**, 62552 (2018).

[29] J. L. Crassidis and J. L. Junkins, *Optimal Estimation of Dynamic Systems*, Chapman & Hall/CRC Applied Mathematics & Nonlinear Science (CRC Press, 2011).

[30] S. Särkkä, *Bayesian Filtering and Smoothing* (Cambridge University Press, 2013).

[31] R. E. Kalman, *"A new approach to linear filtering and prediction problems,"* J Basic Eng-t Asme **82**, 35 (1960).

[32] R. E. Kalman and R. S. Bucy, *"New results in linear filtering and prediction theory,"* J Basic Eng-t Asme **83**, 95 (1961).

[33] D. Simon, *Optimal State Estimation: Kalman, H Infinity, and Nonlinear Approaches* (Wiley, 2006).

[34] C. M. Caves, K. S. Thorne, R. W. P. Drever, V. D. Sandberg, and M. Zimmermann, *"On the measurement of a weak classical force coupled to a quantum-mechanical oscillator. I. Issues of principle,"* Rev. Mod. Phys. **52**, 341 (1980).

[35] V. B. Braginsky and F. Y. Khalili, *"Quantum nondemolition measurements: the route from toys to tools,"* Rev. Mod. Phys. **68**, 1 (1996).

[36] E. Davies, *Quantum Theory of Open Systems* (Academic Press, 1976).

[37] M. Srinivas and E. Davies, *"Photon counting probabilities in quantum optics,"* Optica Acta: International Journal of Optics **28**, 981 (1981).

[38] H. Carmichael, *An Open Systems Approach to Quantum Optics*, Lectures Presented at the Université Libre de Bruxelles, October 28 to November 4, 1991 No. v. 18 (Springer Berlin Heidelberg, 1993).

[39] M. D. Srinivas, *"Quantum theory of continuous measurements and its applications in quantum optics,"* Pramana **47**, 1–23 (1996).

[40] H. M. Wiseman and G. J. Milburn, *"Quantum theory of field-quadrature measurements,"* Phys. Rev. A **47**, 642 (1993).

[41] H. M. Wiseman, *Quantum Trajectories and Feedback*, Phd thesis, University of Queensland (1994), available at https://howardwiseman.me/PhDThesis.pdf.

[42] H. Wiseman and G. Milburn, *Quantum Measurement and Control* (Cambridge University Press, 2010).

[43] K. Iwasawa, K. Makino, H. Yonezawa, M. Tsang, A. Davidovic, E. Huntington, and A. Furusawa, *"Quantum-limited mirror-motion estimation,"* Phys. Rev. Lett. **111**, 163602 (2013).

[44] W. Wieczorek, S. G. Hofer, J. Hoelscher-Obermaier, R. Riedinger, K. Hammerer, and M. Aspelmeyer, *"Optimal state estimation for cavity optomechanical systems,"* Phys. Rev. Lett. **114**, 223601 (2015).

[45] D. J. Wilson, V. Sudhir, N. Piro, R. Schilling, A. Ghadimi, and T. J. Kippenberg, *"Measurement-based control of a mechanical oscillator at its thermal decoherence rate,"* Nature **524**, 325 (2015).

[46] M. Rossi, D. Mason, J. Chen, Y. Tsaturyan, and A. Schliesser, *"Measurement-based quantum control of mechanical motion,"* Nature **563**, 53 (2018).

[47] M. Rossi, D. Mason, J. Chen, and A. Schliesser, *"Observing and verifying the quantum trajectory of a mechanical resonator,"* Phys. Rev. Lett. **123**, 163601 (2019).

[48] V. Sudhir, D. J. Wilson, R. Schilling, H. Schütz, S. A. Fedorov, A. H. Ghadimi, A. Nunnenkamp, and T. J. Kippenberg, *"Appearance and disappearance of quantum correlations in measurement-based feedback control of a mechanical oscillator,"* Phys. Rev. X **7**, 011001 (2017).

[49] A. Setter, M. Toroš, J. F. Ralph, and H. Ulbricht, *"Real-time Kalman filter: Cooling of an optically levitated nanoparticle,"* Phys. Rev. A **97**, 033822 (2018).

[50] L. Magrini, P. Rosenzweig, C. Bach, A. Deutschmann-Olek, S. G. Hofer, S. Hong, N. Kiesel, A. Kugi, and M. Aspelmeyer, *"Real-time optimal quantum control of mechanical motion at room temperature,"* Nature **595**, 373 (2021).

[51] I. K. Kominis, T. W. Kornack, J. C. Allred, and M. V. Romalis, *"A subfemtotesla multichannel atomic magnetometer,"* Nature **422**, 596 (2003).

[52] J. Clarke and A. I. Braginski, *The Squid Handbook*, Vol. 1 (Wiley Online Library, 2004).

[53] J. Kitching, *"Chip-scale atomic devices,"* Appl. Phys. Rev. **5**, 031302 (2018).

[54] J. M. Geremia, J. K. Stockton, A. C. Doherty, and H. Mabuchi, *"Quantum Kalman filtering and the Heisenberg limit in atomic magnetometry,"* Phys. Rev. Lett. **91**, 250801 (2003).

[55] J. K. Stockton, J. M. Geremia, A. C. Doherty, and H. Mabuchi, *"Robust quantum parameter estimation: Coherent magnetometry with feedback,"* Phys. Rev. A **69**, 032109 (2004).

[56] F. Albarelli, M. A. C. Rossi, M. G. A. Paris, and M. G. Genoni, *"Ultimate limits for quantum magnetometry via time-continuous measurements,"* New J. Phys. **19**, 123011 (2017).

[57] R. Jiménez-Martínez, J. Kołodyński, C. Troullinou, V. G. Lucivero, J. Kong, and M. W. Mitchell, *"Signal tracking beyond the time resolution of an atomic sensor by Kalman filtering,"* Phys. Rev. Lett. **120**, 040503 (2018).

[58] C. Troullinou, R. Jiménez-Martínez, J. Kong, V. G. Lucivero, and M. W. Mitchell, *"Squeezed-light enhancement and backaction evasion in a high sensitivity optically pumped magnetometer,"* Phys. Rev. Lett. **127**, 193601 (2021).

[59] C. Troullinou, V. G. Lucivero, and M. W. Mitchell, *"Quantum-enhanced magnetometry at optimal number density,"* Phys. Rev. Lett. **131**, 133602 (2023).

[60] L. B. Madsen and K. Mølmer, *"Spin squeezing and precision probing with light and samples of atoms in the Gaussian description,"* Phys. Rev. A **70**, 052324 (2004).

[61] K. Mølmer and L. B. Madsen, *"Estimation of a classical parameter with Gaussian probes: Magnetometry with collective atomic spins,"* Phys. Rev. A **70**, 052102 (2004).

[62] M. A. C. Rossi, F. Albarelli, D. Tamascelli, and M. G. Genoni, *"Noisy quantum metrology enhanced by continuous nondemolition measurement,"* Phys. Rev. Lett. **125**, 200505 (2020).

[63] J. Kong, R. Jiménez-Martínez, C. Troullinou, V. G. Lucivero, G. Tóth, and M. W. Mitchell, *"Measurement-induced, spatially-extended entanglement in a hot, strongly-interacting atomic system,"* Nat. Commun. **11**, 1 (2020).

[64] D. Budker and M. Romalis, *"Optical magnetometry,"* Nat. Phys. **3**, 227 (2007).

[65] J. Amorós-Binefa and J. Kołodyński, *"Noisy atomic magnetometry with Kalman filtering and measurement-based feedback,"* arXiv:2403.14764 (2024).

[66] J. Amorós-Binefa, M. W. Mitchell, and J. Kołodyński, *"Tracking time-varying signals with quantum-enhanced atomic magnetometers,"* arXiv:2503.14793 (2025).

[67] C. Fritsche, E. Özkan, L. Svensson, and F. Gustafsson, *"A fresh look at bayesian cramér-rao bounds for discrete-time nonlinear filtering,"* 17th International Conference on Information Fusion (FUSION) , 1 (2014).

[68] K. Matsumoto, *"On metric of quantum channel spaces,"* arXiv:1006.0300 (2010).

[69] J. Amorós-Binefa and J. Kołodyński, *"Noisy atomic magnetometry in real time,"* New J. Phys. **23**, 123030 (2021).

[70] C. W. Gardiner *et al.*, *Handbook of Stochastic Methods*, Vol. 3 (Springer Berlin, 1985).

[71] M. Hayes, *Statistical Digital Signal Processing and Modeling* (Wiley, 1996).

[72] A. Papoulis and S. Pillai, *Probability, Random Variables, and Stochastic Processes*, McGraw-Hill series in electrical and computer engineering (McGraw-Hill, 2002).

[73] V. Gorini, A. Kossakowski, and E. C. G. Sudarshan, *"Completely positive dynamical semigroups of n-level systems,"* Journal of Mathematical Physics **17**, 821 (1976), https://pubs.aip.org/aip/jmp/article-pdf/17/5/821/19090720/821_1_online.pdf .

[74] G. Lindblad, *"On the generators of quantum dynamical semigroups,"* Communications in Mathematical Physics **48**, 119 (1976).

[75] L. C. Biedenharn and J. D. Louck, *Angular Momentum in Quantum Physics: Theory and Application*, Encyclopedia of Mathematics and its Applications (Cambridge University Press, 1984).

[76] N. Shammah, S. Ahmed, N. Lambert, S. De Liberato, and F. Nori, *"Open quantum systems with local and collective incoherent processes: Efficient numerical simulations using permutational invariance,"* Phys. Rev. A **98**, 063815 (2018).

[77] J. M. Radcliffe, *"Some properties of coherent spin states,"* Journal of Physics A: General Physics **4**, 313 (1971).

[78] F. T. Arecchi, E. Courtens, R. Gilmore, and H. Thomas, *"Atomic coherent states in quantum optics,"* Phys. Rev. A **6**, 2211 (1972).

[79] J. P. Dowling, G. S. Agarwal, and W. P. Schleich, *"Wigner distribution of a general angular-momentum state: Applications to a collection of two-level atoms,"* Phys. Rev. A **49**, 4101 (1994).

[80] R. Schmied and P. Treutlein, *"Tomographic reconstruction of the Wigner function on the bloch sphere,"* New Journal of Physics **13**, 065019 (2011).

[81] V. K. Khersonskii, A. N. Moskalev, and D. A. Varshalovich, *Quantum theory of angular momemtum* (World Scientific, 1988).

[82] G. S. Agarwal, *"Relation between atomic coherent-state representation, state multipoles, and generalized phase-space distributions,"* Phys. Rev. A **24**, 2889 (1981).

[83] M. Kitagawa and M. Ueda, *"Squeezed spin states,"* Phys. Rev. A **47**, 5138 (1993).

[84] T. Bayes, *An Essay Towards Solving a Problem in the Doctrine of Chances*, Vol. 53 (Royal Society, 1763) pp. 370–418.

[85] G. Kolosov, *Optimal Design of Control Systems: Stochastic and Deterministic Problems*, Pure and Applied Mathematics: A Series of Monographs and Textbooks/221 (CRC Press, 2020).

[86] B. Z. Bobrovsky, E. Mayer-Wolf, and M. Zakai, *"Some classes of global cramér-rao bounds,"* Ann. Statist. **15**, 1421 (1987).

[87] M. Tsang, *"Optimal waveform estimation for classical and quantum systems via time-symmetric smoothing,"* Phys. Rev. A **80**, 033840 (2009).

[88] S. C. Edwards and V. P. Belavkin, *"Optimal quantum filtering and quantum feedback control,"* arXiv:quant-ph/0506018 (2005).

[89] W. Heisenberg, *Physical Principles of the Quantum Theory*, Dover Books on Physics (Dover Publications, 1949).

[90] P. Dirac, *The Principles of Quantum Mechanics* (Oxford University Press, 1930).

[91] J. Von Neumann, *Mathematical foundations of quantum mechanics*, edited by N. A. Wheeler (Princeton University Press, 2018).

[92] K. Kraus, *States, Effects, and Operations*, 1983rd ed., edited by A. Bohm, J. D. Dollard, and W. H. Wootters, Lecture Notes in Physics (Springer, 1983).

[93] D. J. Wineland, *"Superposition, entanglement, and raising Schrödinger's cat,"* Annalen der Physik **525**, 739 (2013).

[94] S. Haroche, *"Nobel lecture: Controlling photons in a box and exploring the quantum to classical boundary,"* Rev. Mod. Phys. **85**, 1083 (2013).

[95] W. Nagourney, J. Sandberg, and H. Dehmelt, *"Shelved optical electron amplifier: Observation of quantum jumps,"* Phys. Rev. Lett. **56**, 2797 (1986).

[96] J. C. Bergquist, R. G. Hulet, W. M. Itano, and D. J. Wineland, *"Observation of quantum jumps in a single atom,"* Phys. Rev. Lett. **57**, 1699 (1986).

[97] S. Gleyzes, S. Kuhr, C. Guerlin, J. Bernu, S. Deléglise, U. Busk Hoff, M. Brune, J.-M. Raimond, and S. Haroche, *"Quantum jumps of light recording the birth and death of a photon in a cavity,"* Nature **446**, 297–300 (2007).

[98] R. J. Cook and H. J. Kimble, *"Possibility of direct observation of quantum jumps,"* Phys. Rev. Lett. **54**, 1023 (1985).

[99] J. Javanainen, *"Possibility of quantum jumps in a three-level system,"* Phys. Rev. A **33**, 2121 (1986).

[100] P. Zoller, M. Marte, and D. F. Walls, *"Quantum jumps in atomic systems,"* Phys. Rev. A **35**, 198 (1987).

[101] R. Blatt and P. Zoller, *"Quantum jumps in atomic systems,"* European Journal of Physics **9**, 250 (1988).

[102] C. Sayrin, I. Dotsenko, X. Zhou, B. Peaudecerf, T. Rybarczyk, S. Gleyzes, P. Rouchon, M. Mirrahimi, H. Amini, M. Brune, J.-M. Raimond, and S. Haroche, *"Real-time quantum feedback prepares and stabilizes photon number states,"* Nature **477**, 73–77 (2011).

[103] M. Aspelmeyer, T. J. Kippenberg, and F. Marquardt, *"Cavity optomechanics,"* Rev. Mod. Phys. **86**, 1391 (2014).

[104] H. M. Wiseman, *"Quantum theory of continuous feedback,"* Phys. Rev. A **49**, 2133 (1994).

[105] H. M. Wiseman and G. J. Milburn, *"Squeezing via feedback,"* Phys. Rev. A **49**, 1350 (1994).

[106] E. B. Davies, *"Quantum stochastic processes,"* Communications in Mathematical Physics **15**, 277–304 (1969).

[107] E. B. Davies, *"Quantum stochastic processes ii,"* Communications in Mathematical Physics **19**, 83–105 (1970).

[108] E. B. Davies, *"Quantum stochastic processes iii,"* Communications in Mathematical Physics **22**, 51–70 (1971).

[109] R. L. Hudson and K. R. Parthasarathy, *"Quantum ito's formula and stochastic evolutions,"* Communications in Mathematical Physics **93**, 301–323 (1984).

[110] A. Barchielli, *"Measurement theory and stochastic differential equations in quantum mechanics,"* Phys. Rev. A **34**, 1642 (1986).

[111] V. Belavkin, in *Information Complexity and Control in Quantum Physics*, edited by A. Blaquiere, S. Diner, and G. Lochak (Springer Vienna, 1987) pp. 311–329.

[112] V. P. Belavkin, in *Modeling and Control of Systems*, edited by A. Blaquiére (Springer Berlin Heidelberg, 1989) pp. 245–265.

[113] A. Barchielli and V. P. Belavkin, *"Measurements continuous in time and a posteriori states in quantum mechanics,"* Journal of Physics A: Mathematical and General **24**, 1495 (1991).

[114] V. Belavkin, *"Measurement, filtering and control in quantum open dynamical systems,"* Reports on Mathematical Physics **43**, A405 (1999).

[115] C. W. Gardiner and M. J. Collett, *"Input and output in damped quantum systems: Quantum stochastic differential equations and the master equation,"* Phys. Rev. A **31**, 3761 (1985).

[116] C. W. Gardiner, A. S. Parkins, and P. Zoller, *"Wave-function quantum stochastic differential equations and quantum-jump simulation methods,"* Phys. Rev. A **46**, 4363 (1992).

[117] R. van Handel, J. K. Stockton, and H. Mabuchi, *"Modelling and feedback control design for quantum state preparation,"* Journal of Optics B: Quantum and Semiclassical Optics **7**, S179 (2005).

[118] J. A. Gross, C. M. Caves, G. J. Milburn, and J. Combes, *"Qubit models of weak continuous measurements: markovian conditional and open-system dynamics,"* Quantum Science and Technology **3**, 024005 (2018).

[119] F. Albarelli and M. G. Genoni, *"A pedagogical introduction to continuously monitored quantum systems and measurement-based feedback,"* Physics Letters A **494**, 129260 (2024).

[120] I. H. Deutsch and P. S. Jessen, *"Quantum control and measurement of atomic spins in polarization spectroscopy,"* Opt. Commun. **283**, 681 (2010).

[121] C. Gardiner and P. Zoller, *Quantum Noise: A Handbook of Markovian and Non-Markovian Quantum Stochastic Methods with Applications to Quantum Optics*, Springer Series in Synergetics (Springer, 2004).

[122] H.-P. Breuer and F. Petruccione, *The Theory of Open Quantum Systems* (Oxford University Press, 2002).

[123] K. Jacobs, *Quantum Measurement Theory and its Applications* (Cambridge University Press, 2014).

[124] D. V. Kupriyanov, O. S. Mishina, I. M. Sokolov, B. Julsgaard, and E. S. Polzik, *"Multimode entanglement of light and atomic ensembles via off-resonant coherent forward scattering,"* Phys. Rev. A **71**, 032348 (2005).

[125] J. M. Geremia, J. K. Stockton, and H. Mabuchi, *"Tensor polarizability and dispersive quantum measurement of multilevel atoms,"* Phys. Rev. A **73**, 042112 (2006).

[126] G. A. Smith, S. Chaudhury, and P. S. Jessen, *"Faraday spectroscopy in an optical lattice: a continuous probe of atom dynamics,"* Journal of Optics B: Quantum and Semiclassical Optics **5**, 323 (2003).

[127] G. A. Smith, S. Chaudhury, A. Silberfarb, I. H. Deutsch, and P. S. Jessen, *"Continuous weak measurement and nonlinear dynamics in a cold spin ensemble,"* Phys. Rev. Lett. **93**, 163602 (2004).

[128] A. Kuzmich, L. Mandel, J. Janis, Y. E. Young, R. Ejnisman, and N. P. Bigelow, *"Quantum nondemolition measurements of collective atomic spin,"* Phys. Rev. A **60**, 2346 (1999).

[129] M. Born, E. Wolf, A. B. Bhatia, P. C. Clemmow, D. Gabor, A. R. Stokes, A. M. Taylor, P. A. Wayman, and W. L. Wilcock, *Principles of Optics: Electromagnetic Theory of Propagation, Interference and Diffraction of Light*, 7th ed. (Cambridge University Press, 1999).

[130] S. R. de Echaniz, M. W. Mitchell, M. Kubasik, M. Koschorreck, H. Crepaz, J. Eschner, and E. S. Polzik, *"Conditions for spin squeezing in a cold 87rb ensemble,"* Journal of Optics B: Quantum and Semiclassical Optics **7**, S548 (2005).

[131] S. R. de Echaniz, M. Koschorreck, M. Napolitano, M. Kubasik, and M. W. Mitchell, *"Hamiltonian design in atom-light interactions with rubidium ensembles: A quantum-information toolbox,"* Phys. Rev. A **77**, 032316 (2008).

[132] G. Colangelo, R. J. Sewell, N. Behbood, F. M. Ciurana, G. Triginer, and M. W. Mitchell, *"Quantum atom–light interfaces in the Gaussian description for spin-1 systems,"* New Journal of Physics **15**, 103007 (2013).

[133] Y. Takahashi, K. Honda, N. Tanaka, K. Toyoda, K. Ishikawa, and T. Yabuzaki, *"Quantum nondemolition measurement of spin via the paramagnetic faraday rotation,"* Phys. Rev. A **60**, 4974 (1999).

[134] D. Budker, W. Gawlik, D. F. Kimball, S. M. Rochester, V. V. Yashchuk, and A. Weis, *"Resonant nonlinear magneto-optical effects in atoms,"* Rev. Mod. Phys. **74**, 1153 (2002).

[135] E. Boto, N. Holmes, J. Leggett, G. Roberts, V. Shah, S. S. Meyer, L. D. Muñoz, K. J. Mullinger, T. M. Tierney, S. Bestmann, G. R. Barnes, R. Bowtell, and M. J. Brookes, *"Moving magnetoencephalography towards real-world applications with a wearable system,"* Nature **555**, 657 (2018).

[136] M. Tsang, H. M. Wiseman, and C. M. Caves, *"Fundamental quantum limit to waveform estimation,"* Phys. Rev. Lett. **106**, 090401 (2011).

[137] M. Ernzer, M. Bosch Aguilera, M. Brunelli, G.-L. Schmid, T. M. Karg, C. Bruder, P. P. Potts, and P. Treutlein, *"Optical coherent feedback control of a mechanical oscillator,"* Phys. Rev. X **13**, 021023 (2023).

[138] D. Budker and D. F. Jackson Kimball, eds., *Optical Magnetometry* (Cambridge University Press, 2013).

[139] D. Budker and M. G. Kozlov, *"Sensing: Equation one,"* arXiv:2011.11043 (2020).

[140] W. Happer and A. C. Tam, *"Effect of rapid spin exchange on the magnetic-resonance spectrum of alkali vapors,"* Phys. Rev. A **16**, 1877 (1977).

[141] I. M. Savukov and M. V. Romalis, *"Effects of spin-exchange collisions in a high-density alkali-metal vapor in low magnetic fields,"* Phys. Rev. A **71**, 023405 (2005).

[142] W. Happer, *"Optical pumping,"* Rev. Mod. Phys. **44**, 169 (1972).

[143] J. Ma, X. Wang, C. P. Sun, and F. Nori, *"Quantum spin squeezing,"* Phys. Rep. **509**, 89 (2011).

[144] S. Chatterjee, R. S. Thakur, R. N. Yadav, L. Gupta, and D. K. Raghuvanshi, *"Review of noise removal techniques in ECG signals,"* IET Signal Proc. **14**, 569 (2020).

[145] S. F. Huelga, C. Macchiavello, T. Pellizzari, A. K. Ekert, M. B. Plenio, and J. I. Cirac, *"Improvement of frequency standards with quantum entanglement,"* Phys. Rev. Lett. **79**, 3865 (1997).

[146] U. Dorner, *"Quantum frequency estimation with trapped ions and atoms,"* New Journal of Physics **14**, 043011 (2012).

[147] D. Plankensteiner, J. Schachenmayer, H. Ritsch, and C. Genes, *"Laser noise imposed limitations of ensemble quantum metrology,"* Journal of Physics B: Atomic, Molecular and Optical Physics **49**, 245501 (2016).

[148] M. A. C. Rossi, F. Albarelli, D. Tamascelli, and M. G. Genoni, "Qcontinuousmeasurement.jl (2021)," https://github.com/matteoacrossi/QContinuousMeasurement.jl (2021).

[149] B. A. Chase and J. M. Geremia, *"Collective processes of an ensemble of spin-1/2 particles,"* Phys. Rev. A **78**, 052101 (2008).

[150] B. Q. Baragiola, B. A. Chase, and J. Geremia, *"Collective uncertainty in partially polarized and partially decohered spin-$\frac{1}{2}$ systems,"* Phys. Rev. A **81**, 032104 (2010).

[151] P. Rouchon and J. F. Ralph, *"Efficient quantum filtering for quantum feedback control,"* Phys. Rev. A **91**, 012118 (2015).

[152] F. Albarelli, M. A. C. Rossi, D. Tamascelli, and M. G. Genoni, *"Restoring Heisenberg scaling in noisy quantum metrology by monitoring the environment,"* Quantum **2**, 110 (2018).

[153] R. Kubo, *"Generalized cumulant expansion method,"* Journal of the Physical Society of Japan **17**, 1100 (1962).

[154] D. Plankensteiner, C. Hotter, and H. Ritsch, *"QuantumCumulants.jl: A Julia framework for generalized mean-field equations in open quantum systems,"* Quantum **6**, 617 (2022).

[155] K. Jacobs and D. A. Steck, *"A straightforward introduction to continuous quantum measurement,"* Contemp. Phys. **47**, 279 (2006).

[156] M. H. Muñoz Arias, P. M. Poggi, P. S. Jessen, and I. H. Deutsch, *"Simulating nonlinear dynamics of collective spins via quantum measurement and feedback,"* Phys. Rev. Lett. **124**, 110503 (2020).

[157] J. G. Bohnet, K. C. Cox, M. A. Norcia, J. M. Weiner, Z. Chen, and J. K. Thompson, *"Reduced spin measurement back-action for a phase sensitivity ten times beyond the standard quantum limit,"* Nat. Photonics **8**, 731 (2014).

[158] O. Hosten, N. J. Engelsen, R. Krishnakumar, and M. A. Kasevich, *"Measurement noise 100 times lower than the quantum-projection limit using entangled atoms,"* Nature **529**, 505 (2016).

[159] V. Petersen and K. Mølmer, *"Estimation of fluctuating magnetic fields by an atomic magnetometer,"* Phys. Rev. A **74**, 043802 (2006).

[160] M. Saarinen, P. Siltanen, P. Karp, and T. Katila, *"The normal magnetocardiogram: I morphology,"* Ann. Clin. Res. **10**, 1 (1978).

[161] B. Kaplan, I. Gabay, G. Sarafian, and D. Sarafian, *"Biological applications of the "filtered" Van der Pol oscillator,"* J. Franklin Inst. **345**, 226 (2008).

[162] S. Das and K. Maharatna, *"Fractional dynamical model for the generation of ECG like signals from filtered coupled van-der pol oscillators,"* Computer Methods and Programs in Biomedicine **112**, 490 (2013).

[163] H. I. Nurdin and N. Yamamoto, *Linear Dynamical Quantum Systems* (Springer, 2017).

[164] L. K. Thomsen, S. Mancini, and H. M. Wiseman, *"Continuous quantum nondemolition feedback and unconditional atomic spin squeezing,"* J. Phys. B: At., Mol. Opt. Phys. **35**, 4937 (2002).

[165] W. E. Bell and A. L. Bloom, *"Optically driven spin precession,"* Phys. Rev. Lett. **6**, 280 (1961).

[166] M. P. Ledbetter, V. M. Acosta, S. M. Rochester, D. Budker, S. Pustelny, and V. V. Yashchuk, *"Detection of radio-frequency magnetic fields using nonlinear magneto-optical rotation,"* Phys. Rev. A **75**, 023405 (2007).

[167] A. Weis, G. Bison, and A. S. Pazgalev, *"Theory of double resonance magnetometers based on atomic alignment,"* Phys. Rev. A **74**, 033401 (2006).

[168] M. Koźbiał, L. Elson, L. M. Rushton, A. Akbar, A. Meraki, K. Jensen, and J. Kołodyński, *"Spin noise spectroscopy of an alignment-based atomic magnetometer,"* Phys. Rev. A **110**, 013125 (2024).

[169] W. Happer, Y.-Y. Jau, and T. Walker, *Optically Pumped Atoms* (John Wiley & Sons, 2010).

[170] W. Górecki, F. Albarelli, S. Felicetti, R. D. Candia, and L. Maccone, *"Interplay between time and energy in bosonic noisy quantum metrology,"* arXiv:2409.18791 (2025)