University of Pannonia
Doctoral School of Information Science and Technology

Parameter estimation and robustness analysis of quantum information systems

PhD thesis
by
Gábor Balló

DOI: 10.18136/PE.2014.546

Supervisor:
Dr. Katalin Hangos

March 2, 2014

Parameter estimation and robustness analysis of quantum information systems

Értekezés doktori (PhD) fokozat elnyerése érdekében

Írta:
Balló Gábor

Készült a Pannon Egyetem Informatikai Tudományok Doktori Iskolája keretében

Témavezető: Dr. Hangos Katalin

Elfogadásra javaslom (igen / nem) .....
(aláírás)

A jelölt a doktori szigorlaton 100%-ot ért el.

Az értekezést bírálóként elfogadásra javaslom:

Bíráló neve: ..... igen / nem .....
(aláírás)

Bíráló neve: ..... igen / nem .....
(aláírás)

A jelölt az értekezés nyilvános vitáján ..... %-ot ért el.

Veszprém, ...
a Bíráló Bizottság elnöke

A doktori (PhD) oklevél minősítése .....

.....
az EDT elnöke

Abstract

Quantum information systems have different determining factors of their performance. In particular, in the field of quantum error correction the accurate knowledge of the models of quantum physical processes, the so-called quantum channels is required. It follows that the ability to accurately identify the channels, together with the precise characterization of error correction robustness properties against the channel uncertainty is essential.

Therefore, the robustness of quantum error correction operations against completely unstructured type of channel uncertainties is first investigated in this thesis. We find that a channel-adapted optimal error correction operation does not only give the best possible channel fidelity but it is more robust against channel perturbations than any other error correction operation. Our results are valid for Pauli channels and stabilizer codes. Numerical results indicate that very similar conclusions can be drawn also in the general case.

In addition, a method of parameter estimation is proposed for quantum channels of which we have a priori structural information. In the case of channels with the Pauli channel structure, a special parametrization turns the parameter estimation problem into a convex optimization problem.

Furthermore, a novel experiment design method for the parameter estimation of Pauli channels is developed both for the cases of known and unknown channel structure. For qubit channels in the former case, this leads to an optimal setting that includes pure input states and projective measurements directed towards the channel directions. For qubit Pauli channels with unknown structure, an iterative method of estimating the channel directions is proposed, together with the study and comparison of an adaptive estimation procedure with simple non-adaptive methods.

Kivonat

A kvantum információs rendszerek teljesítményének különféle meghatározó tényezői vannak. Különönen a kvantum hibajavítás területén szükséges a kvantumfizikai folyamatok modelljeinek, az úgynevezett kvantumcsatornáknak a pontos ismerete. Ebből következik, hogy a csatornák pontos identifikációjának képessége, valamint a hibajavítás csatornabizonytalansággal szembeni robusztussági tulajdonságainak precíz jellemzése alapvető.

Tehát ebben a disszertációban elsőként a kvantum hibajavító operációk teljesen strukturálatlan típusú csatornabizonytalansággal szembeni robusztussága kerül vizsgálatra. Azt találjuk, hogy egy csatornaadaptív optimális hibajavító operáció nem csak a lehető legjobb csatornafidelity-t adja, de robusztusabb is a csatornaperturbációkkal szemben, mint bármely más hibajavító operáció. Eredményeink Pauli csatornákra és stabilizátor kódokra érvényesek. Numerikus eredmények alapján nagyon hasonló következtetések vonhatók le az általános esetben is.

Továbbá egy paraméterbecslő módszert javasolunk olyan kvantumcsatornákhoz, melyek struktúrájáról a priori információnk van. A Pauli struktúrával rendelkező csatornák esetében egy speciális paraméterezés konvex optimalizációs problémává alakítja a paraméterbecslési feladatot.

Ezenkívül egy újfajta kísérlettervezési módszert fejlesztünk ki a Pauli csatornák paraméterbecsléséhez, ismert és ismeretlen csatornastruktúra esetére is. Az előbbi esetben kubit csatornákra egy olyan optimális elrendezést kapunk, melyben tiszta bemenő állapotok és projektív mérések állnak a csatornairányok mentén. Ismeretlen struktúrájú kubit Pauli csatornákra egy iteratív módszert javasolunk a csatornairányok becslésére, valamint tanulmányozunk egy adaptív becslő eljárást, melyet összevetünk egyszerű nemadaptív módszerekkel.

Zusammenfassung

Quantum Informationssysteme haben unterschiedliche Faktoren in ihrer Leistung. Insbesondere im Bereich der Quanten-Fehlerkorrektur sind die genauen Kenntnisse von den Modellen des quantenphysikalischen Prozesses, die sogenannten Quantenkanäle erforderlich. Daraus folgt, dass die Fähigkeit die Kanäle genau zu identifizieren, zusammen mit der genauen Charakterisierung der Robustheitseigenschaften von der Fehlerkorrektur gegen den Kanal Unsicherheit wesentlich ist.

Daher wird die Robustheit der Quanten-Fehlerkorrektions Operationen gegen völlig unstrukturierte Kanal Unsicherheits Typen erstens in dieser Arbeit untersucht. Wir finden, dass eine optimale Kanal-adaptive Fehlerkorrektur Operation nicht nur den bestmöglichen Kanal Fidelity ergibt, aber es ist auch robuster gegen Kanal Störungen als jede andere Fehlerkorrektur Operation. Unsere Ergebnisse gelten für Pauli Kanäle und Stabilisator Codes. Numerische Ergebnisse zeigen, dass sehr ähnlichen Schlussfolgerungen auch im allgemeinen Fall gezogen werden können.

Darüber hinaus wird eine Methode zur Parameterschätzung für Quanten-Kanäle, mit a priori Strukturinformationen vorgeschlagen. Im Falle von Kanälen mit dem Pauli Kanalstruktur wird ein spezielles Parametrisierung das Parameterschätzproblem in eine konvexe Optimierungsproblem umwandeln.

Zudem wird ein neuartiges Versuchsplanungs-Methode für die Parameterschätzung von Pauli Kanäle sowohl für die Fälle von bekannten und unbekannten Kanal-Strukturen entwickelt. Für Qubit Kanäle führt dies zu einer optimalen Einstellung, die reine Eingangs-zustände und Projektiv-messungen nach Kanal Richtung gerichtet beinhaltet. Für Qubit Pauli Kanäle mit unbekannter Struktur wird eine iterative Methode zur Schätzung der Kanal Richtungen vorgeschlagen, zusammen mit der Untersuchung und Vergleich eines adaptiven Schätzverfahren mit einfachen, nicht-adaptiven Methoden.

Acknowledgement

I am grateful to Dr. Katalin Hangos, my supervisor, for her guidance and support during my PhD studies. Her advices helped me to learn the practice of scientific research.

I would also like to thank Dr. Péter Gurin, the coauthor of the first part of my thesis, for the joint work and inspiration during the early years of my research.

I must also thank Dr. Dénes Petz, László Ruppert, and Dr. Attila Magyar for their advices on my work.

I thank Róbert Galambos for the discussions about science, about being a PhD student, and for the help in running long simulations.

Big thanks goes also to Veronika Baráth, my girlfriend, for her love, and patience during times of hard work on research and publications.

Finally, I am also grateful to my parents for the strong and constant support they gave me during my student years.

Financial support

Financial support was provided by the Hungarian State and the European Union under the TAMOP-4.2.2.A-11/1/ KONV-2012-0072.

Contents

Chapter 1

Introduction

> *Atoms on a small scale behave like nothing on a large scale, for they satisfy the laws of quantum mechanics. So, as we go down and fiddle around with the atoms down there, we are working with different laws, and we can expect to do different things.*
> – R. P. Feynman

According to Feynman, quantum mechanics offers special possibilities, however, it also puts special obstacles in front of these possibilities. For instance, the building of quantum computers that can be used to solve realistic, large scale problems – including the breaking of cryptographic codes and the simulation of complex quantum systems – has two main difficulties from the theoretical point of view. The first is decoherence, in other words the unwanted but unavoidable coupling of quantum systems with their environment. This noise effect can alter or even completely destroy the information content of the system, causing errors in the calculation. The other main obstacle is the inability to fully manipulate and extract full information from physical systems, i.e., the control and estimation of quantum states and processes. This is a hard problem, because in quantum mechanics one has to face the difficulty of the treatment of measurements. Namely, that no measurement can be carried out on a quantum system without substantially disturbing the state of the system itself.

1.1 Background and motivation

A common element of the above introduced problems is the quantum channel, i.e., the general model of physical processes transforming quantum states to quantum states. Whether we would like to do quantum information processing or solve system- and control theoretic problems, the characterization of quantum channels, i.e., the modeling of quantum mechanical processes becomes essential.

Modeling is an important field in classical system- and control theory [1, 2]. Mathematical models are in general simplifictions of the physical reality, which

implies that all mathematical models of a physical system suffer from inaccuracies. These can result from non-exact measurements or from the inability to capture all involved physical phenomena or just from the requirement to obtain a simple model. The quality of a nominal model depends on how closely its behaviour matches that of the true system. In engineering context this mismatch between the nominal model and the real system is called uncertainty [3], referring not to the uncertain nature of the physical system, but to our incomplete knowledge. The degree of uncertainty can vary from incomplete knowledge about the model structure to uncertainty only in certain parameters.

The design of more accurate nominal models is not necessarily satisfactory in practice, mostly because the real system may not even be in the set of possible nominal models. Thus to have sufficient information about the usability and accuracy of the identified model, robustness analysis of some application specific property against the model uncertainty is also required [4, 3]. Robustness in general means the ability to resist change in certain conditions without adaptation. It means that robustness can only be analyzed after the precise characterization of these conditions.

The same problem arises in connection with the processing and control of quantum systems. The complexity of the environment does not allow us to take interactions between the system and its environment into account with perfect accuracy. Thus the quantum channel, too, will always remain just a model, which though describes reality more or less precisely, but we can not assume that it is fully accurate.

It is clear that each quantum information theoretic application has a specific threshold of performance. For instance, the error correction of quantum systems requires accurate models of the channels representing noise processes. This implies that accurate identification of quantum channels together with the study of the constraints on their uncertainty under which robustness properties hold are of essential importance.

1.2 Aims of the thesis

One of the central problems of quantum information processing is to find suitable error correcting procedures [5, 6]. Therefore, the development of the theory of quantum error correction (QEC) was a very important step, giving the possibility for quantum information theory to become a potentially applicable science from an only theoretically interesting field. Nowadays, there are two main approaches; the standard quantum error correcting method tries to adapt the techniques of classical information theory, while the optimization approach can be used – assuming some exact noise model – to find the best error correcting method.

In this thesis, one of our aims is to perform robustness analysis on the conventional and optimization based channel specific error correction procedures, assuming uncertainty in the channel model. At first sight, we could think that the conventional QEC – which makes very slight assumptions about the properties of the channel – is more insensitive to the uncertainties than an

error correction optimized for a specific channel. On the other hand, we can also think that if a solution is optimal, then in general, in case of a very small alteration of the conditions it decays only in second order, so it is robust in some measure. These questions apparently have greater significance when the channel is of general type, but because of the simplicity of the handling of Pauli channels, here we concentrate mainly on this specific case.

Another undoubtedly fundamental problem of quantum information theory is the task of the identification of quantum processes, commonly known as quantum process tomography (QPT) [7]. It has considerable relevance not only in quantum computers, but also in the field of quantum communication and cryptography. For example, quantum communication channels usually rely on a priori knowledge of the channel properties. This shows that the performance of quantum information theoretic applications can depend greatly on the accuracy of the estimated channel models.

Thus our second aim is to develop identification methods achieving high accuracy with favorable computational properties. Our convex optimization based approach uses the fact that often we have a priori knowledge of quantum processes, which can be used to reduce the level of uncertainty in the model and arrive at a parameter estimation problem significantly easier to solve with a predefined accuracy.

The problem of classical system identification is related to experiment design, the general aim of which is to determine experimental conditions that result in good or even optimal identification results. In the case of quantum channel parameter estimation, the design variables are the quantum input to the channel, and the measurements to be applied on the resulting quantum output system. These are called the experiment configuration, together with the number of measurements to be performed in the different experiment configurations if one has a few of them.

Therefore, our third aim is to extend our results on parameter estimation, and find optimal experiment configurations for the class of Pauli channels having various levels of uncertainty present in the model.

1.3 Structure of the thesis

The rest of the thesis is organized as follows. The introduction is followed by Chapter 2, which describes basic notions about quantum systems and quantum channels.

Part I treats the topic of quantum error correction. In Chapter 3 the basic theory of quantum error correcting methods is introduced. Then Chapter 4 presents our results on the robustness analysis of quantum error correcting procedures against uncertainties of Pauli and non-pauli channels.

Part II treats the topic of quantum process tomography. Chapter 5 gives an introduction on channel identification and the related experiment design problem. Then Chapter 6 presents our results on the convex optimization-based channel parameter estimation of Pauli channels with known structure. Chapter 7 presents our proposed experiment design method for the estimation

of Pauli channels with known channel structure. Finally, in Chapter 8 our two methods for the estimation of Pauli channels with unknown structure is presented.

Lastly, in Chapter 9 the conclusions are drawn in the form of thesis statements, the possible future research directions are discussed and the publication list of the author is presented.

Appendix A describes all necessary concepts needed by the main thesis chapters which are assumed beyond the level of an information technology master degree. Appendix B presents additional examples and simulation results complementing those of Chapter 6.

Throughout the thesis, our main results are phrased in the form of statements, while used known results are phrased as theorems.

1.4 Notations

Here in Table 1.1 we set a few notational conventions which we will use throughout the thesis. If the meaning of some of these symbols were different in exceptional cases, it will be apparent from the context.

Symbol	Meaning
$\mathcal{H}$	Hilbert space
$\mathcal{B}(\mathcal{H})$	set of (bounded) linear operators acting on $\mathcal{H}$
$\mathcal{B}(\mathcal{H}_1, \mathcal{H}_2)$	set of (bounded) linear operators mapping $\mathcal{H}_1$ to $\mathcal{H}_2$
$\mathcal{V}$	vector space
d	level of quantum system
$\lvert\,.\,\rangle, \langle\,.\,\rvert$	state vector, dual of state vector
ρ	density matrix
$\rho_{\mathrm{ms}}, \rho_{\mathrm{cs}}$	message, codeword
$\mathbf{V}$	matrix whose columns form a MUB
$\mathbf{S}$	qubit Pauli channel affine representation
λ	Pauli channel depolarizing parameters
$\sigma_{\mathrm{x}}, \sigma_{\mathrm{y}}, \sigma_{\mathrm{z}}$	Pauli matrices
X, Y, Z	Pauli matrices
X_i, Y_i, Z_i	Pauli matrices acting on the i^{th} qubit
$\bar{X}, \bar{Y}, \bar{Z}$	logical Pauli matrices
$\lvert\mathbf{e}_i\rangle, \lvert-\mathbf{e}_i\rangle$	eigenvectors of σ_i
$\lvert\mathbf{v}_i\rangle, \lvert-\mathbf{v}_i\rangle$	eigenvectors of $\mathbf{v}_i \cdot \vec{\sigma}$
$\mathbb{1}$	identity matrix
θ	Bloch vector
$\mathcal{C}, \mathcal{C}_S$	quantum code, stabilizer code
$\mathcal{E}$	quantum noise channel
$\mathcal{R}$	recovery operation
$X_{\mathcal{E}}$	choi matrix of $\mathcal{E}$
$\mathcal{I}$	identity map
*	Hermitian adjoint
*	denotes optimality
$\propto$	denotes proportionality
$\nabla_{\mathbf{p}}$	gradient with respect to $\mathbf{p}$
∂_{p_i}	partial derivative with respect to p_i
V_{LS}	least squares objective function
$\mathrm{diag}(\mathbf{a})$	matrix with vector $\mathbf{a}$ in its main diagonal
$\mathrm{diag}^{-1}(\mathbf{A})$	vector obtained by taking out the main diagonal of matrix $\mathbf{A}$

Table 1.1. Notations used in the thesis.

Chapter 2

Basic notions

Here we give a short description of the basic notions of finite dimensional quantum mechanical systems with emphasis on the concepts needed to understand the main chapters. The concepts discussed here can be found in [5, 8]. The mathematical preliminaries needed for the basic notions are briefly summarized in appendix A.

In section 2.1 the postulates of quantum mechanics are presented. Section 2.2 gives a detailed discussion on quantum channels in general. This is followed by section 2.3 on Pauli channels. Then, in section 2.4 several metrics used in quantum information are presented.

2.1 Postulates of quantum mechanics

Quantum mechanics is a mathematical framework used to develop physical theories. It does not give us any laws of physics or tell us any facts about specific physical systems, rather it can be used as a tool for formally describing these laws.

In the next few sections we give an introduction on the basic postulates of quantum mechanics. These postulates provide a connection between the physical world and the mathematical formalism of quantum mechanics. The derivation of the postulates was historically a long process of trial and error, understanding their motivation and meaning can lead to philosophical questions. In this chapter we restrict ourselves only to an application oriented discussion.

2.1.1 States of quantum systems

Two commonly used representations of finite dimensional quantum systems will be presented here, the latter being the most general.

Postulate 1. *The state space of any isolated physical system is a complex Hilbert space $\mathcal{H}$. The system can be completely described either by the state vector in $\mathcal{H}$ or by the density matrix acting on $\mathcal{H}$.*

For finite dimensional quantum systems the dimension d of $\mathcal{H}$ is finite and d is also the dimension (level number) of the quantum system.

Pure states

A quantum state providing maximal knowledge about the quantum system is said to be a *pure state*. A pure state is most commonly represented as a normalized *state vector* $|\psi\rangle \in \mathcal{H}$. The symbol $|\psi\rangle$ is the so-called Dirac "ket" notation of vectors. By the Dirac notation, the scalar product of two vectors is denoted by $\langle\varphi|\psi\rangle$, where the $\langle\varphi|$ "bra" vector is the natural pair of $|\varphi\rangle$ in the dual space.

The linear structure of $\mathcal{H}$ implies the *superposition principle*, namely that the normalized linear combination of two state vectors is also a possible quantum state. We can thus write any quantum state $|\psi\rangle \in \mathcal{H}$ as a normalized linear combination $\sum_i \alpha_i|i\rangle$ ($\alpha_i \in \mathbb{C}$) of an arbitrary orthonormal basis $\{|i\rangle\}$ of $\mathcal{H}$ with the normalization condition $\sum_i |\alpha_i|^2 = 1$.

Note that the description of the physical state by a state vector is unique only apart from global phase, i.e., $|\psi\rangle$ and $\mathrm{e}^{\mathrm{i}\theta}|\psi\rangle$ describe the same physical state. In case of two superposed states however, the phase difference between the corresponding state vectors is relevant.

Mixed states

If the state is not pure, then it is said to be a *mixed state*. A mixed state is represented as a positive semidefinite matrix ρ with unit trace, called *density matrix*:

$$\rho \geq 0, \quad \mathrm{Tr}(\rho) = 1$$

Such states are mixtures of pure states taken from the *ensemble* $\{p_i, |\psi_i\rangle\}_{i=1}^d$. This means that the system may be in the pure state $|\psi_i\rangle$ with probability p_i. Because of this uncertainty, a mixed state does not provide maximal knowledge about the system. The density matrix corresponding to the ensemble is $\rho = \sum_i p_i|\psi_i\rangle\langle\psi_i|$.[1] It follows that the density matrix of pure states are rank-1 orthogonal projections $\rho = |\psi\rangle\langle\psi|$.

The level of mixedness can be measured with the quantity $\mathrm{Tr}(\rho^2)$ called *purity*. The value goes from 1 indicating a pure state to $\frac{1}{d}$ indicating the completely mixed state $\rho = \frac{1}{d}\mathbb{1}$ corresponding to the ensemble $\{\frac{1}{d}, |\psi_i\rangle\}_{i=1}^d$ for a d level system.

Physical content of the density matrix

Regarding the physical content of a density matrix [9], let ρ be the pure state $|\psi\rangle\langle\psi|$. In a given basis $|i\rangle$, the diagonal elements $\langle i|\rho|i\rangle$ are referred to as *populations*, these give the weight $|\alpha_i|^2$ of the basis state $|i\rangle$ in the superposition $|\psi\rangle = \sum_i \alpha_i|i\rangle$.[2] The off-diagonal elements $\langle i|\rho|j\rangle$ (with $i \neq j$) are called *coherences*, they give information about the relative phase between components of the superposition $|\psi\rangle$. For an ensemble $\{p_k, |\psi_k\rangle\}_{k=1}^d$, these relative phases

[1] This correspondence is not one-to-one, there is a unitary freedom in the ensembles giving the same density matrix.

[2] Precisely, these are the probabilities $|\langle i|\psi\rangle|^2$ of jumping into the state $|i\rangle$ after measuring in the basis $\{|i\rangle\}$ (see later in section 2.1.3).

may be different for different pure states $|\psi_k\rangle$, thus in a mixed density matrix the coherences may get averaged out.

The qubit

The most basic data unit of quantum information theory is the two-state (two-level) quantum system. This is called "quantum bit", or *qubit* for short.[3]

The possible pure states of a qubit are the elements of the two dimensional Hilbert space. After choosing e.g. the orthonormal basis $\{|0\rangle, |1\rangle\}$ – the so-called *computational basis* – the states can be written as $|\psi\rangle = \alpha_0|0\rangle + \alpha_1|1\rangle$, where $|\alpha_0|^2 + |\alpha_1|^2 = 1$. Thus in contrast to the classical bit, which can only be 0 or 1, the qubit can also take the arbitrary complex superposition of the states $|0\rangle$ and $|1\rangle$.

Let $|\psi_0\rangle = \alpha_0|0\rangle + \alpha_1|1\rangle$ and $|\psi_1\rangle = \beta_0|0\rangle + \beta_1|1\rangle$ be pure states. Then all (including mixed) states of a qubit corresponding to the ensemble $\{p_i, |\psi_i\rangle\}_{i=1,2}$ can be written as the density matrix:

$$\rho = \sum_{k=0}^{1} p_k|\psi_k\rangle\langle\psi_k| = \sum_{i,j\in\{0,1\}} c_{i,j}|i\rangle\langle j|, \quad c_{i,j} = p_0\alpha_i\alpha_j^* + p_1\beta_i\beta_j^* \tag{2.1}$$

In effect, the rank-1 matrices $\{|i\rangle\langle j|\}$ $(i, j \in \{0, 1\})$ form the computational basis in the complex Hilbert space of all 2×2 matrices.

The Bloch picture

The density matrix of a qubit can be described in any basis similarly to (2.1), however, the most convenient basis is the Pauli basis. It consists of the identity matrix $\mathbb{1}$ (also denoted σ_0 for convenience) and the three Pauli matrices σ_x (σ_1), σ_y (σ_2), and σ_z (σ_3), or simply X, Y, and Z. Their matrices in the computational basis are:

$$\sigma_\mathrm{x} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_\mathrm{y} = \begin{pmatrix} 0 & -\mathrm{i} \\ \mathrm{i} & 0 \end{pmatrix}, \quad \sigma_\mathrm{z} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} . \tag{2.2}$$

The Pauli matrices are self-adjoint and they obey the commutation relations $\sigma_a\sigma_b = \delta_{ab}\mathbb{1} + \mathrm{i}\varepsilon_{abc}\sigma_c$ where ε_{abc} is the Levi-Civita symbol.[4] The symbol $\vec{\sigma}$ commonly denotes the formal vector $[\sigma_\mathrm{x}, \sigma_\mathrm{y}, \sigma_\mathrm{z}]$. A formal dot product with $\mathbf{v} \in \mathbb{R}^3$ is then $\mathbf{v} \cdot \vec{\sigma} \coloneqq \sum_{i=1}^{3} v_i\sigma_i$.

The convenience in using the basis $\{\mathbb{1}, \sigma_\mathrm{x}, \sigma_\mathrm{y}, \sigma_\mathrm{z}\}$ arises from the fact that it forms an orthogonal basis (with respect to the Hilbert–Schmidt inner product (A.1)) also in the real Hilbert space of 2×2 Hermitian matrices. The qubit

[3] Considering its specific physical realization, the qubit can be the spin of any half-integer spin particle (for example the spin of an electron), or the two different polarization states of a photon.

[4] If (a, b, c) is an even permutation of (x, y, z), then $\varepsilon_{abc} = 1$, in case of odd permutation $\varepsilon_{abc} = -1$, else $\varepsilon_{abc} = 0$.

density matrices can thus be written as

$$\rho = \frac{1}{2}\left(\mathbb{1} + \sum_{i=1}^{3} \theta_i \sigma_i\right) = \frac{1}{2}\begin{bmatrix} 1+\theta_3 & \theta_1 - \mathrm{i}\theta_2 \\ \theta_1 + \mathrm{i}\theta_2 & 1-\theta_3 \end{bmatrix} , \tag{2.3}$$

where the numbers $\theta_i = \mathrm{Tr}(\rho\sigma_i)$ form a three-dimensional vector θ, commonly called *Bloch vector*. The hermiticity and normalization of the density matrix are ensured by this formula. The remaining positivity constraint is expressed using the 2-norm of the Bloch vector as $\|\theta\|_2 \leq 1$. It follows that equation (2.3) gives a unique correspondence between all possible qubit states and points of the unit ball in $\mathbb{R}^3$, the so-called Bloch ball, which can be seen in Figure 2.1.

Let $|\mathbf{e}_i\rangle$ and $|-\mathbf{e}_i\rangle$ be the normalized eigenvectors of the Pauli matrix σ_i from the expansion

$$\sigma_i = |\mathbf{e}_i\rangle\langle\mathbf{e}_i| - |-\mathbf{e}_i\rangle\langle-\mathbf{e}_i| \ .$$

Through the correspondence (2.3), $|\mathbf{e}_i\rangle$ and $|-\mathbf{e}_i\rangle$ are identified with the two unit vectors $\pm\mathbf{e}_i$ along the Cartesian axis i.

The Bloch vector picture of qubits allows a convenient distinction of pure and mixed states. The pure (completely polarized) states take place on the surface of the Bloch ball, and the mixed (depolarized) states can be found in the interior, with decreasing purity towards the origin representing the completely mixed (completely depolarized) state.

Mutually unbiased bases

It is well-known that the three bases $\{|\mathbf{e}_i\rangle, |-\mathbf{e}_i\rangle\}_{i=1}^3$ form a set of mutually unbiased bases (MUB) (see appendix A.4). MUBs can be unitarily transformed

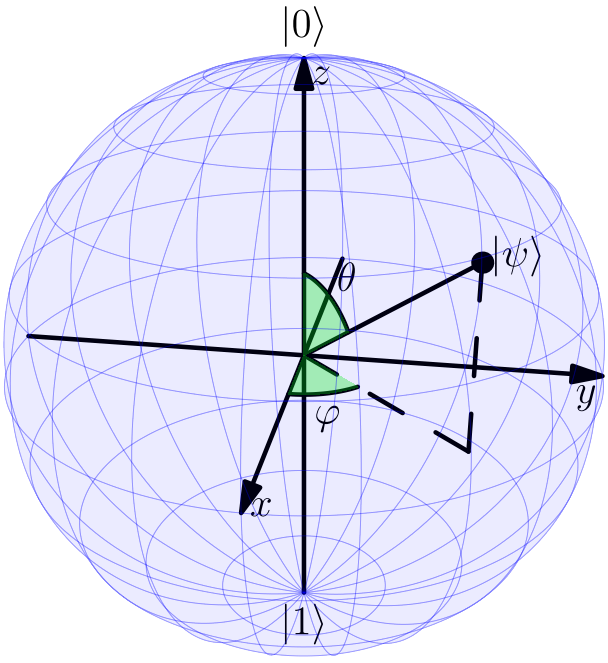


Figure 2.1. It can be seen on the Bloch sphere, that all pure states of a qubit can also be written in $|\psi(\theta, \varphi)\rangle$ form.

into each other. In the case of qubits, since any orthogonal transformation $\mathbf{V} = [\mathbf{v}_x, \mathbf{v}_y, \mathbf{v}_z]$ on $\mathbb{R}^3$ is induced by a unitary conjugation with $U_\mathbf{V}$ on 2×2 complex matrices, the basis vectors $\{U_\mathbf{V}|\pm\mathbf{e}_i\rangle\}$ will be the eigenvectors of $\mathbf{v}_i \cdot \vec{\sigma}$ corresponding to the Bloch vectors $\pm\mathbf{v}_i$. A set of MUB can then be identified with three orthogonal axes in the Bloch ball, and any two-dimensional MUB arises this way. Thus, the general notation $|\pm\mathbf{v}_i\rangle \coloneqq U_\mathbf{V}|\pm\mathbf{e}_i\rangle$ will be used in the following.

This symmetry of the Bloch ball shows that any MUB $\{|\pm\mathbf{v}_i\rangle\}_{i=1}^3$ on qubit systems can be used to describe qubit density matrices as Bloch vectors with respect to the MUB $\{|\pm\mathbf{v}_i\rangle\}$:

$$\begin{aligned}\rho = \frac{1}{2}\left(\mathbb{1} + \sum_{i=1}^{3} \vartheta_i \mathbf{v}_i \cdot \vec{\sigma}\right) &= \frac{1}{2}\left(\mathbb{1} + \sum_{i=1}^{3} \vartheta_i \left(|\mathbf{v}_i\rangle\langle\mathbf{v}_i| - |-\mathbf{v}_i\rangle\langle-\mathbf{v}_i|\right)\right) \\ &= \frac{1}{2}\left(1 - \sum_{i=1}^{3} \vartheta_i\right)\mathbb{1} + \sum_{i=1}^{3} \vartheta_i |\mathbf{v}_i\rangle\langle\mathbf{v}_i| \ ,\end{aligned} \tag{2.4}$$

where $\vartheta_i = \mathrm{Tr}(\rho \mathbf{v}_i \cdot \vec{\sigma}) = \mathbf{v}_i^{\mathrm{T}}\theta$ is now a vector component in the basis $\{\mathbf{v}_i\}_{i=1}^3$.

2.1.2 Time evolution

Postulate 2. *The continuous time evolution of a closed quantum system $|\psi\rangle$ is described by the Schrödinger equation*

$$\frac{\mathrm{d}}{\mathrm{d}t}|\psi(t)\rangle = -\frac{\mathrm{i}}{\hbar} H|\psi(t)\rangle \ , \tag{2.5}$$

where $\hbar$ is the Planck's constant, and H is a fixed hermitian operator known as the Hamiltonian of the closed system.

If we know the Hamiltonian of a quantum system, then (in principle) we understand its dynamics completely.[5] The value of $\hbar$ is not important for us, practically it can just be absorbed into H.

The solution of this equation is $|\psi(t)\rangle = \mathrm{e}^{-\mathrm{i}(t-t_0)H}|\psi(t_0)\rangle$, which indicates that the unitary matrix $U(t, t_0) = \mathrm{e}^{-\mathrm{i}(t-t_0)H}$ called *time-evolution operator* can be used to easily relate states at different time instants, i.e., to implement discrete time dynamics. It can be proven that any unitary arises in such form, thus the state dynamics governed by (2.5) is also known as *unitary evolution*.

The property of reversibly relating two quantum states allows us to regard unitaries also as *quantum logic gates*, and use them to implement analogues of classical reversible logic circuits.[6]

It must be noted that no such unitary U exists that could create an independent copy of an arbitrary quantum state $|\psi\rangle$, i.e., act as $U(|\psi\rangle \otimes |0\rangle) = |\psi\rangle \otimes |\psi\rangle$ for all $|\psi\rangle \in \mathcal{H}$. Such U exist only for sets of mutually orthogonal states, so

[5] In general, figuring out the Hamiltonian is a difficult problem in physics.

[6] Classical irreversible gates do not have a quantum analogue, because quantum computation is practically done on pure states, and irreversible quantum dynamics does not preserve purity, i.e., it acts as noise (see section 2.2).

copiers can only be constructed for the elements of a known basis of $\mathcal{H}$. This is the content of the so-called *no-cloning theorem*.

The equivalent of equation (2.5) for mixed states is the von Neumann equation $\frac{\mathrm{d}}{\mathrm{d}t}\rho(t) = -\frac{\mathrm{i}}{\hbar}[H, \rho(t)]$, and analogously $\rho(t) = U(t, t_0)\rho(t_0)U(t, t_0)^*$.

The simplest examples of quantum gates are the Pauli matrices (2.2). The matrix X is known as the quantum NOT gate or *bit-flip* suggested by the expression $X|0\rangle = |1\rangle$. Similarly, Z is called *phase-flip* because it flips the relative phase: $Z(\alpha_0|0\rangle+\alpha_1|1\rangle) = \alpha_0|0\rangle-\alpha_1|1\rangle$, $|\alpha_0|^2+|\alpha_1|^2 = 1$. Furthermore, as Y is proportional with XZ, we call it *bit-phase flip*.

2.1.3 Quantum measurements

Observing a quantum system, i.e., taking a measurement is a special example of non-unitary quantum dynamics.

Postulate 3. *Quantum measurement procedures are represented by a collection $\boldsymbol{M} \coloneqq \{M_m\}$ of self-adjoint positive operators acting on the state space $\mathcal{H}$ of the system being measured. The index m identifies each operator M_m with a corresponding outcome event that may occur as a result of the measurement. This implies that $\{M_m\}$ has to satisfy the completeness relation $\sum_m M_m = \mathbb{1}$. Such a set of operators are commonly called positive operator-valued measure (POVM).*

If the system is in state ρ then the probability distribution of possible outcomes is given by $p(m|\rho) = \mathrm{Tr}(\rho M_m)$. If we get the measurement result m and the measurements are further specified[7] *such that $M_m = F_m^* F_m$ uniquely, then the state ρ of the system collapses*[8] *into $\rho_m = \frac{F_m \rho F_m^*}{p(m|\rho)}$.*

Thus the main features of quantum measurement are the stochastic outcome and the unavoidable abrupt back-action on the measured system, making the state change discontinuously.

POVMs form a convex set [10]. $\mathbf{M} = p\mathbf{M}_1 + (1-p)\mathbf{M}_2$ with $0 \leq p \leq 1$ means that the distribution $p_{\mathbf{M}}(\,.\,|\rho)$ can be equivalently obtained as the convex combination of $p_{\mathbf{M}_1}(\,.\,|\rho)$ and $p_{\mathbf{M}_2}(\,.\,|\rho)$, i.e., measuring $\mathbf{M}$ is the same as randomly choosing between measurements $\mathbf{M}_1$ and $\mathbf{M}_2$. Formally, the convex combination of POVMs is calculated elementwise by treating POVMs as vectors of positive operators (with elements allowed to be the zero operator). The extremal points of the set of POVMs are called extremal POVMs. Let $\{|\psi_{\alpha,i}\rangle\}_{i=1}^{\mathrm{rank}(M_\alpha)}$ be the set of eigenvectors of POVM element M_α. Then the POVM $\mathbf{M} = \{M_\alpha\}$ is extremal if and only if the operators $|\psi_{\alpha,i}\rangle\langle\psi_{\alpha,j}|$ are linearly independent for all α, i and j.

[7] In spite of this inconvenience, POVMs are widely used, because in many cases only the measurement statistics are important.

[8] This means that measurement makes the state ρ jump into one of the possible outcome states ρ_m.

Projective measurement

If the positive operators are all orthogonal projections P_m onto pairwise orthogonal subspaces, then we have a special case of measurements, the so-called projective (von Neumann–Lüders) measurement.

This property allows projective measurements $\mathbf{M} = \{P_m\}$ to be represented as a single *observable* M, a Hermitian operator on the state space $\mathcal{H}$ of the system. The correspondence is made using the spectral decomposition $M = \sum_m mP_m$, i.e., each P_m will be a projector onto the eigenspace of M with eigenvalue m. If the projections are all of the form $P_m = |m\rangle\langle m|$ then we can speak about "measuring in the basis $|m\rangle$".

Furthermore, the hermiticity and idempotence of P_m allows us to derive the state after measurement directly; performing a projective measurement collapses the state into one of the eigenvectors (eigenstates) of the measured M observable. It follows that projective measurements – opposed to POVMs – are repeatable in the sense that repeating the same measurement will give the same result with certainty.

It can be proven that POVMs can be realized as projective measurements if we are allowed to compose the system with other quantum systems and to perform unitary dynamics.

Projective measurements are included in the set of extremal POVMs.

Observables for qubits are for example the Pauli matrices.[9] Let us have the state $|\psi\rangle = \alpha_0|0\rangle + \alpha_1|1\rangle$. Measuring in the computational basis corresponds to the observable $\sigma_z = |0\rangle\langle 0| - |1\rangle\langle 1|$ or to the POVM $\{|0\rangle\langle 0|, |1\rangle\langle 1|\}$. Then the state of the qubit after measurement will be $|0\rangle$ with probability $\mathrm{Tr}(|\psi\rangle\langle\psi|0\rangle\langle 0|) = |\alpha_0|^2$, and $|1\rangle$ with probability $\mathrm{Tr}(|\psi\rangle\langle\psi|1\rangle\langle 1|) = |\alpha_1|^2$. However, if we only know that a measurement has happened, but the result is unknown, then the state of the system can be given as the ensemble of the two possible outcome states, i.e., as the density matrix

$$\rho = |\alpha|^2|0\rangle\langle 0| + |\beta|^2|1\rangle\langle 1| \ .$$

In general, every projective measurement on a qubit can be written as a two element POVM. In the Bloch picture these two elements are $\frac{1}{2}(\mathbb{1} \pm \mathbf{v}\cdot\vec{\sigma})$ with $\|\mathbf{v}\|_2 = 1$, and they can be used for "measurement along the axis $\mathbf{v}$" in the Bloch ball. After measurement, the state changes either to the Bloch vector $\mathbf{v}$ or $-\mathbf{v}$.

2.1.4 Composite systems

Postulate 4. *The state space $\mathcal{H}$ of a bipartite physical system is the tensor product (see appendix A.1.3) $\mathcal{H}_1 \otimes \mathcal{H}_2$ of the state spaces $\mathcal{H}_1$ and $\mathcal{H}_2$ of the component systems. Moreover, if the first system is in state $|\psi_1\rangle$ and the second system is in state $|\psi_2\rangle$ then the joint state of the total system is $|\psi_1\rangle \otimes |\psi_2\rangle \in \mathcal{H}$.*

[9] Historically, observables correspond to physical quantities. The Pauli matrices can be used to measure spin components, while the Hamiltonian from section 2.1.2 is the observable of energy.

An arbitrary state of the space $\mathcal{H}$ can then be written in the following way:

$$|\psi\rangle = \sum_{i,j} \alpha_{i,j}|i\rangle \otimes |j\rangle, \quad \text{or in short} \quad |\psi\rangle = \sum_{i,j} \alpha_{i,j}|ij\rangle \ , \tag{2.6}$$

where $|i\rangle$ and $|j\rangle$ are orthonormal bases for the component Hilbert spaces $\mathcal{H}_1$ and $\mathcal{H}_2$.

If in (2.6) $|\psi\rangle = \sum_i \alpha_i|i\rangle \otimes \sum_j \alpha_j|j\rangle = |\psi_1\rangle \otimes |\psi_2\rangle$ with $|\psi_1\rangle \in \mathcal{H}_1$ and $|\psi_2\rangle \in \mathcal{H}_2$, then $|\psi\rangle$ is called a *separable state*. Otherwise, if the component systems can not be separated, then $|\psi\rangle$ is an *entangled state*. This means that there is a *quantum correlation* between its component systems.

However, even if $|\psi\rangle$ is entangled, we can still associate for example its first subsystem with a state, which provides the correct measurement statistics for measurements affecting only the first subsystem, i.e., measurements having POVM elements of the form $M_m \otimes \mathbb{1}$. It is calculated using the *partial trace* defined as the unique linear operator $\mathrm{Tr}_1 \colon \mathcal{B}(\mathcal{H}_1 \otimes \mathcal{H}_2) \to \mathcal{B}(\mathcal{H}_2)$ such that $\mathrm{Tr}_1(A_1 \otimes A_2) = A_2\mathrm{Tr}(A_1)$. Then the partial trace of $|\psi\rangle$ with respect to the second subsystem is

$$\rho_1 = \mathrm{Tr}_2(|\psi\rangle\langle\psi|) = \sum_{i,j,k,l} \alpha_{i,j}\alpha^*_{k,l}|i\rangle\langle k| \otimes \mathrm{Tr}(|j\rangle\langle l|) = \sum_{i,j,k} \alpha_{i,k}\alpha^*_{k,j}|i\rangle\langle k| \ . \tag{2.7}$$

The result ρ_1 is called *reduced density operator*. In terms of reduced densities, $|\psi\rangle$ is entangled if and only if ρ_1 (and ρ_2) is mixed, indicating that we can not get maximal knowledge of them independently of the other part of $|\psi\rangle$. Thus, any mixed state ρ_1 can be obtained as a correlated part of some larger system. Moreover, assuming that $\dim(\mathcal{H}_1) \leq \dim(\mathcal{H}_2)$, $|\psi\rangle$ is said to be maximally entangled if $\rho_1 = \mathrm{Tr}_2(|\psi\rangle\langle\psi|) = \frac{1}{\dim(\mathcal{H}_1)}\mathbb{1}$, i.e., ρ_1 is a maximally mixed state.

To find a pure state $|\psi\rangle$ for which (2.7) holds with a given mixed ρ_1 as subsystem is called *purification* of ρ_1. This procedure is of course not unique, the system and its environment as a whole can have many states whose subsystem is the same mixed state.

Consider now the two-qubit state $|\Phi^+\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$. This state is an example of entangled state; there are no single qubit states $|\phi_1\rangle$ and $|\phi_2\rangle$ such that $|\Phi^+\rangle = |\phi_1\rangle \otimes |\phi_2\rangle$. In fact, the partial traces $\rho_1 = \mathrm{Tr}_2(|\Phi^+\rangle\langle\Phi^+|) = \frac{1}{2}\mathbb{1} = \mathrm{Tr}_1(|\Phi^+\rangle\langle\Phi^+|) = \rho_2$ result in mixed states. The result $\rho_1 = \rho_2 = \frac{1}{2}\mathbb{1}$ also shows that $|\Phi^+\rangle$ is a two-qubit maximally entangled state.

Operators acting independently on each part of a composite system are also in tensor product form. For example, the tensor product of Pauli matrices can act on multiple qubit systems: $(X \otimes Z)(\alpha|0\rangle + \beta|1\rangle) = \alpha X|0\rangle + \beta Z|1\rangle = \alpha|1\rangle - \beta|1\rangle$. $X \otimes Z$ can also be denoted as X_1Z_2 meaning that X acts on the first qubit and Z acts on the second.

2.2 Quantum channels

In reality, a quantum system can never be perfectly closed. The interaction with the environment gives rise to more general physical processes causing

dynamic change in the state necessarily represented as a density matrix. These processes are thus modeled by *quantum channels* (or quantum operations) which map density matrices to density matrices.[10] In the following, we will present different mathematical representations of channels with their respective properties.

2.2.1 The Kraus representation

One way to arrive at a definition of the quantum channel $\mathcal{E}$ is to consider the quantum system together with its environment to be closed. Let ρ be the density matrix of the system with state space $\mathcal{H}_1$, and let the environment with state space $\mathcal{H}_2$ be in some pure state $|\varphi_0\rangle$.[11] Then the unitary evolution of the joint state $\rho \otimes |\varphi_0\rangle\langle\varphi_0|$ is $U(\rho \otimes |\varphi_0\rangle\langle\varphi_0|)U^*$. From this we get the dynamics of the $\mathcal{H}_1$ system alone by taking the partial trace with respect to the environment $\mathcal{H}_2$. This gives us the first definition of quantum channels:

$$\mathcal{E}(\rho) = \mathrm{Tr}_2\big(U(\rho \otimes |\varphi_0\rangle\langle\varphi_0|)U^*\big) = \sum_j \langle\varphi_j|U|\varphi_0\rangle\rho\langle\varphi_0|U^*|\varphi_j\rangle = \sum_j V_j \rho V_j^* \tag{2.8}$$

If $\{|i\rangle \otimes |\varphi_j\rangle\}$ is an orthonormal basis of $\mathcal{H}_1 \otimes \mathcal{H}_2$, then $V_j \coloneqq \langle\varphi_j|U|\varphi_0\rangle : \mathcal{H}_1 \to \mathcal{H}_1$ denotes the operator with matrix elements $(\langle i| \otimes \langle\varphi_j|)U(|i'\rangle \otimes |\varphi_0\rangle)$. This *operator element set* $\{V_j\}$ gives the so-called *Kraus representation* of $\mathcal{E}$. Because of the unitarity of U it also holds that

$$\sum_j V_j^* V_j = \mathbb{1} \ , \tag{2.9}$$

which is the *completeness relation* expressing the trace preserving property of $\mathcal{E}$.[12]

Knowledge of the operator elements $\{V_j\}$ makes it possible to describe the dynamics of the system $\mathcal{H}_1$ without needing the attributes of the environment $\mathcal{H}_2$ to be taken into account explicitly; all nesessary information is embedded into the operator elements, which only have effect on $\mathcal{H}_1$. It follows however, that the set of Kraus operators $\{V_j\}$ corresponding to a particular channel $\mathcal{E}$ is not unique. In fact the choice of basis $\{|\varphi_j\rangle\}$ is arbitrary in (2.8). This implies a unitary freedom in the operator element set for the channel $\mathcal{E}$: the operators $W_i = \sum_j u_{i,j} V_j$ (with $u_{i,j} \in \mathbb{C}$ being the elements of a unitary matrix) form an equivalent set of Kraus operators for the channel $\mathcal{E}$.[13]

Note that if we have a d level quantum system, then at most d^2 operator elements are enough to describe any possible quantum channel on the system.

[10] Channels are also called superoperators, because they map operators to operators.

[11] The environment could also be mixed, however, considering it pure is not a loss of generality. Moreover, if $\dim(\mathcal{H}_1) = d$ then it is sufficient to take $\dim(\mathcal{H}_2)$ to be at most d^2.

[12] Measurements can also be modeled using quantum channels, if we relax the completeness condition (2.9) to $\sum_j V_j^* V_j \leq \mathbb{1}$.

[13] If the two operator element sets do not contain the same number of operators, then we augment the smaller set with zero operators.

Note also that this channel definition can be generalized to the case of different input and output spaces. In this case, instead of making distinction between system and environment spaces, we speak about two systems; the first starts in state ρ, the second is in state $|\varphi_0\rangle$. We bring them into interaction, then by discarding the first system, i.e., by applying the partial trace on it, we obtain the remaining second system in the output state $\mathcal{E}(\rho)$.

Defining axioms

Quantum channels can also be defined as maps satisfying the following axioms:

- **linearity**: required by the ensemble interpretation of density matrices,
- **positivity** and **trace preservation**: these ensure that the channel outcome will also be a density matrix,
- **complete positivity**: this means that the composite channel $\mathcal{E} \otimes \mathcal{I}_n$ acting on the extended space $\mathcal{B}(\mathcal{H}_1) \otimes \mathcal{B}(\mathcal{H}_{\text{ext}})$ with $\dim(\mathcal{H}_{\text{ext}}) = n$ also has to be a positive map for all n.

Maps with these properties are commonly called *completely positive and trace preserving (CPTP)* maps.

The following theorem states that the above two definitions of quantum channels are equivalent:

Theorem 2.1. *A map $\mathcal{E}\colon \mathcal{B}(\mathcal{H}_1) \to \mathcal{B}(\mathcal{H}_2)$ taking density operators to density operators satisfies the above three axioms if and only if $\mathcal{E}(\rho) = \sum_j V_j \rho V_j^*$ for a set $\{V_j\colon \mathcal{H}_1 \to \mathcal{H}_2\}$ satisfying* (2.9).

2.2.2 The Choi matrix

The unitary freedom in the Kraus representation can be circumvented using another possible description of channels, the *Choi matrix*, which is in essence a matrix representation of the channel $\mathcal{E}$. In fact, $\mathcal{E}$ has an infinite number of equivalent Kraus representations, but has only one Choi matrix. It can be defined using the Choi–Jamiołkowski isomorphism [11, 12].

The Choi–Jamiołkowski isomorphism

The *Choi–Jamiołkowski isomorphism* associates an operator $C\colon \mathcal{H}_1 \to \mathcal{H}_2$ with a vector[14] $|C\rangle\!\rangle \in \mathcal{H}_1 \otimes \mathcal{H}_2$. Let $\{|i_1\rangle\}$ and $\{|j_2\rangle\}$ be some preferred basis in $\mathcal{H}_1$ and $\mathcal{H}_2$, and let $|\Phi_{\max}\rangle = \sum_i |i_1\rangle \otimes |i_1\rangle$ be the unnormalized maximally entangled state in $\mathcal{H}_1 \otimes \mathcal{H}_1$. Then the isomorphism is given by the following definition:

$$|C\rangle\!\rangle := (\mathbb{1} \otimes C)|\Phi_{\max}\rangle = \sum_i |i_1\rangle \otimes C|i_1\rangle = \sum_{i,j} |i_1\rangle \otimes \langle j_2|C|i_1\rangle|j_2\rangle$$

[14] The $|.\rangle\!\rangle$ notation tries to indicate that these vectors represent operators.

This is essentially the "stacking" of columns of the matrix C to form the vector $|C\rangle\!\rangle$. This assignment is obviously unique, and it also can be seen that $|\Phi_{\max}\rangle = |\mathbb{1}\rangle\!\rangle$. The inner product of these vectors in $\mathcal{H}_1 \otimes \mathcal{H}_2$ is defined in a natural way:

$$\langle\!\langle A|B\rangle\!\rangle = \left(\sum_i \langle i_1| \otimes \langle i_1|A^*\right)\left(\sum_j |j_1\rangle \otimes B|j_1\rangle\right) = \mathrm{Tr}(A^*B) \tag{2.10}$$

what is by definition the Hilbert–Schmidt inner product (A.1) of the two operators. Thus an isomorphism can be made between the operators in $\mathcal{B}(\mathcal{H}_1, \mathcal{H}_2)$ and the vectors of the space $\mathcal{H}_1 \otimes \mathcal{H}_2$.

Let A^{T} denote the transposition in the preferred basis $\{|i_1\rangle\}$. Then a useful relation follows directly from the above definition:

$$(A \otimes B)|C\rangle\!\rangle = |BCA^{\mathrm{T}}\rangle\!\rangle \ , \tag{2.11}$$

whenever the dimensions of A, B, and C indicate that BCA^{T} is a valid operator.[15]

Representing quantum channels

Using the fact that the space of maps $\mathcal{E}\colon \mathcal{B}(\mathcal{H}_1) \to \mathcal{B}(\mathcal{H}_2)$ is also a Hilbert space, we can use the Choi–Jamiołkowski isomorphism to associate $\mathcal{E}$ with an operator $X_{\mathcal{E}} \in \mathcal{B}(\mathcal{H}_1) \otimes \mathcal{B}(\mathcal{H}_2)$:

$$X_{\mathcal{E}} \coloneqq (\mathcal{I} \otimes \mathcal{E})\left(|\Phi_{\max}\rangle\langle\Phi_{\max}|\right) = \sum_{i,j,k} \Big(|i_1\rangle \otimes V_k|i_1\rangle\Big)\Big(\langle j_1| \otimes \langle j_1|V_k^*\Big)$$

where the operators V_k are the Kraus operator elements of $\mathcal{E}$. Continuing the derivation, two useful formulas can be obtained from this:

$$X_{\mathcal{E}} = \sum_k (\mathbb{1} \otimes V_k)|\mathbb{1}\rangle\!\rangle\langle\!\langle\mathbb{1}|(\mathbb{1} \otimes V_k^*) = \sum_k |V_k\rangle\!\rangle\langle\!\langle V_k| \tag{2.12}$$

and

$$X_{\mathcal{E}} = \sum_{i,j} |i_1\rangle\langle j_1| \otimes \sum_k V_k|i_1\rangle\langle j_1|V_k^* = \sum_{i,j} |i_1\rangle\langle j_1| \otimes \mathcal{E}(|i_1\rangle\langle j_1|) \ . \tag{2.13}$$

Actually, the above matrix is a block matrix, its $(i,j)^{\text{th}}$ element is $\mathcal{E}\big(|i_1\rangle\langle j_1|\big)$.

The complete positivity of $\mathcal{E}$ is equivalent to the positivity of $X_{\mathcal{E}}$. Furthermore, $\mathcal{E}$ is trace preserving if and only if $\mathrm{Tr}\left[\mathcal{E}\big(|i_1\rangle\langle j_1|\big)\right] = \mathrm{Tr}\big(|i_1\rangle\langle j_1|\big) = \delta_{i,j}$ which means $\mathrm{Tr}_2(X_{\mathcal{E}}) = \mathbb{1}$.

The inverse of the map, i.e., the $X_{\mathcal{E}} \to \mathcal{E}$ relation is $\mathcal{E}(\rho) = \mathrm{Tr}_1\big((\rho^{\mathrm{T}} \otimes \mathbb{1})X_{\mathcal{E}}\big)$, which comes from (2.13).

Note that the definition of the Choi–Jamiołkowski isomorphism presented here is basis dependent, i.e., depends on the choice of $\{|i_1\rangle\}$.

The Kraus operator elements $|V_k\rangle\!\rangle$ can be obtained from the Choi matrix by applying a square root factorization. The unitary freedom of such a factorization corresponds to the unitary freedom of the Kraus representation.

Note also that the Choi matrix representation shows clearly that the set of quantum channels is convex.

[15] Another relation not used here follows for $C_1, C_2 \in \mathcal{B}(\mathcal{H}_1, \mathcal{H}_2)$: $\mathrm{Tr}_1(|C_1\rangle\!\rangle\langle\!\langle C_2|) = C_1C_2^* \in \mathcal{B}(\mathcal{H}_2)$

2.2.3 Channels on a qubit

In the two-level system case, as in section 2.1.1 we have the possibility to expand the density operator ρ with respect to any set of MUB $\{\pm|\mathbf{v}_i\rangle\}_{i=1}^3$. The most general channel on a qubit can then be associated with an affine map $T\colon \mathbb{R}^3 \to \mathbb{R}^3$ acting on the Bloch ball [8, 13]:

$$T(\theta) = \mathbf{S}\theta + \mathbf{t} \ ,$$

where θ is the Bloch vector of ρ, $\mathbf{t} \in \mathbb{R}^3$ and $\mathbf{S} = \mathbf{V}\Lambda\mathbf{Q}\mathbf{V}^{\mathrm{T}}$ is a real matrix decomposed into rotation matrices $\mathbf{Q}$ and $\mathbf{V} = [\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3]$ with the latter representing the MUB transformation. Finally, Λ is diagonal with elements λ_1, λ_2, and λ_3.

If $\mathbf{V}$ and $\mathbf{Q}$ are not important and we are free to set them then simply $\mathbf{V} = \mathbf{Q} = \mathbb{1}$, and then we have the map $T = \Lambda\theta + \mathbf{t}$ corresponding to the standard set of MUB $\{\pm|\mathbf{e}_i\rangle\}_{i=1}^3$. This form is generally sufficient for analyzing the properties of the channel.

Example qubit channels used throughout the thesis are described in Appendix A.5.1.

2.3 Pauli channels

A notable wide class of quantum channels are the *Pauli channels*.

2.3.1 Group theoretic definition on qubits

Pauli channels can be defined using the *Pauli group* defined in appendix A.3.1. If a channel has a set of operator elements, which contains scaled elements of the Pauli group, i.e., $\{V_i = \sqrt{a_i}g_i\}$ where $a_i \geq 0$ and $g_i \in \mathcal{P}_n$, then we call it an n-qubit Pauli channel.[16] Because of the trace preserving condition of the channel, $\sum_i a_i = 1$ has to hold. Note that (2.8) implies that two operator element sets differing only in factors of ± 1 or $\pm \mathrm{i}$ give rise to the same Pauli channel.

On a single qubit, the Pauli channel is thus the following:

$$V_0 = \sqrt{a_0}\mathbb{1},\ V_1 = \sqrt{a_1}X,\ V_2 = \sqrt{a_2}Y,\ V_3 = \sqrt{a_3}Z,\ a_i \geq 0,\ \sum_i a_i = 1 \quad (2.14)$$

2.3.2 Definition using matrix algebras

The most general Pauli channel definition for arbitrary level quantum systems is discussed in [14] in a matrix algebraic context (see appendix A.4 for basics).

Let the center of the full matrix algebra $M_d(\mathbb{C})$ be $Z_{M_d(\mathbb{C})} = \{c\mathbb{1} \,|\, c \in \mathbb{C}\}$, and let $A_i : M_d(\mathbb{C}) \to \mathcal{A}_i$ be a trace preserving projection (usually called conditional expectation). If the pairwise complementary subalgebras $\mathcal{A}_1, \mathcal{A}_2, \ldots, \mathcal{A}_u$

[16] In order to be Pauli, it is enough for the channel to have at least one such Kraus element set.

of $M_d(\mathbb{C})$ are given and they linearly span the whole algebra $M_d(\mathbb{C})$, then any matrix $D \in M_d(\mathbb{C})$ is the sum of the components $A_i(D) - \frac{\mathrm{Tr}(D)}{d}\mathbb{1}$ in the traceless subspaces $\mathcal{A}_i \cap Z^{\perp}_{M_d(\mathbb{C})}$ and the component $\frac{\mathrm{Tr}(D)}{d}\mathbb{1}$ in $Z_{M_d(\mathbb{C})}$:

$$D = \sum_{i=1}^{u} \left(A_i(D) - \frac{\mathrm{Tr}(D)}{d}\mathbb{1} \right) + \frac{\mathrm{Tr}(D)}{d}\mathbb{1} \ ,$$

By definition, the Pauli channel $\mathcal{E}$ is depolarizing on each $\mathcal{A}_i \cap Z^{\perp}_{M_d(\mathbb{C})}$, i.e., contracts each component $A_i(D) - \frac{\mathrm{Tr}(D)}{d}\mathbb{1}$ with parameter λ_i, resulting in

$$\mathcal{E}(D) = \left(1 - \sum_{i=1}^{u} \lambda_i \right) \frac{\mathrm{Tr}(D)}{d}\mathbb{1} + \sum_{i=1}^{u} \lambda_i A_i(D)$$

Trivially, if D is a density matrix and $\lambda_i = 0$ for all i then the result is the completely mixed state $\frac{1}{d}\mathbb{1}$. Conversely if $\lambda_i = 1$ for all i then D remains unchanged.

The channel is trace preserving by construction, and completely positive if and only if the parameters λ_i satisfy

$$1 + d\lambda_i \geq \sum_j \lambda_j \geq -\frac{1}{d-1} \ . \tag{2.15}$$

This constraint describes a polyhedron bounded by $\lambda_i \in [\frac{-1}{d-1}, 1]$.

In this thesis, we consider only the case when all of the complementary subalgebras are maximal Abelian, and the level d of quantum systems is prime number. In this case the subalgebras $\{\mathcal{A}_i\}_{i=1}^{d+1}$ can be obtained as the linear span of $d+1$ groups of unitaries (see appendix A.4), and the unitaries can be selected to be generalized Pauli matrices. Through these, the set of subalgebras correspond also to a set of MUB $\{|\phi_{i,j}\rangle\}_{i=1}^{d+1}$, and thus the orthogonal projection A_i can be constructed as

$$A_i(D) = \sum_{j=1}^{d} \langle\phi_{i,j}|D|\phi_{i,j}\rangle |\phi_{i,j}\rangle\langle\phi_{i,j}| \ ,$$

and the subalgebras can also be given as

$$\mathcal{A}_i = \left\{ \sum_{j=1}^{d} c_j |\phi_{i,j}\rangle\langle\phi_{i,j}| \,\middle|\, c_j \in \mathbb{C} \right\} .$$

2.3.3 Qubit Pauli channel

For qubit systems, a standard selection $\{\pm|\mathbf{e}_i\rangle\}_{i=1}^{3}$ of MUB can be obtained from the eigenvectors of the Pauli matrices. Then the

$$\mathcal{A}_i = \{a|\mathbf{e}_i\rangle\langle\mathbf{e}_i| + b|-\mathbf{e}_i\rangle\langle-\mathbf{e}_i| \,|\, a, b \in \mathbb{C}\} \quad (i = 1, 2, 3)$$

are commutative subalgebras and the projections onto them are for example

$$A_1\left(\begin{bmatrix} 1+\theta_3 & \theta_1 - i\theta_2 \\ \theta_1 + i\theta_2 & 1-\theta_3 \end{bmatrix}\right) = \begin{bmatrix} 1 & \theta_1 \\ \theta_1 & 1 \end{bmatrix}.$$

Using these, the Pauli channel on the input density matrix ρ of the form (2.3) is defined by

$$\mathcal{E}(\rho) = \frac{1}{2}\left(\mathbb{1} + \sum_{i=1}^{3} \lambda_i \theta_i \sigma_i\right) . \tag{2.16}$$

The Choi matrix of this channel is the following:

$$X_{\mathcal{E}} = \frac{1}{2}\begin{bmatrix} 1+\lambda_3 & 0 & 0 & \lambda_1+\lambda_2 \\ 0 & 1-\lambda_3 & \lambda_1-\lambda_2 & 0 \\ 0 & \lambda_1-\lambda_2 & 1-\lambda_3 & 0 \\ \lambda_1+\lambda_2 & 0 & 0 & 1+\lambda_3 \end{bmatrix} \tag{2.17}$$

This matrix satisfies trace preserving by construction. The conditions of positivity for $X_{\mathcal{E}}$ in terms of the parameters are

$$|1 \pm \lambda_3| \geq |\lambda_1 \pm \lambda_2| . \tag{2.18}$$

This constraint is symmetric in the parameters λ_i. It describes a tetrahedron in $\mathbb{R}^3$ bounded by $\|\lambda\|_\infty \leq 1$.

It can be shown that this channel is equivalent to the group theoretic definition in section 2.3.1 in the sense, that one is a reparametrization of the other. Using (2.14) and that by definition $\mathcal{E}(\sigma_i) = \lambda_i \sigma_i$, we get that the relationship between the parameters is

$$\begin{bmatrix} 1 \\ \lambda_1 \\ \lambda_2 \\ \lambda_3 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \end{bmatrix} = \begin{bmatrix} 1 \\ a_0+a_1-a_2-a_3 \\ a_0-a_1+a_2-a_3 \\ a_0-a_1-a_2+a_3 \end{bmatrix} \tag{2.19}$$

Denote the coefficient matrix in (2.19) with $\mathbf{Q}$. Then the first row of $\mathbf{Q}$ corresponds to $\sum_i a_i = 1$, and this additional row makes $\mathbf{Q}$ invertible. The columns of the remaining bottom part define four vertices of a tetrahedron in $\mathbb{R}^3$, and the right hand side of (2.19) describes the convex combination of these. It can be seen that (2.18) describes the same tetrahedron just in terms of the λ_i Cartesian coordinates. Thus the two definitions are truly equivalent on qubit systems.

Rotated Pauli channel

Similarly to general qubit channels, general two-level Pauli channels can also be defined in terms of any two-dimensional set of MUB $\{|\pm\mathbf{v}_i\rangle\}_{i=1}^3$. In other words, such a channel scales the components of a density matrix (2.4) in the subalgebras $\mathcal{A}_i = \{a|\mathbf{v}_i\rangle\langle\mathbf{v}_i| + b|-\mathbf{v}_i\rangle\langle-\mathbf{v}_i| \,|\, a, b \in \mathbb{C}\}$ as

$$\mathcal{E}(\rho) = \frac{1}{2}\left(1 - \sum_{i=1}^{3} \lambda_i \vartheta_i\right)\mathbb{1} + \sum_{i=1}^{3} \lambda_i \vartheta_i |\mathbf{v}_i\rangle\langle\mathbf{v}_i|.$$

Then – beside the λ_i – the vectors $|\mathbf{v}_1\rangle$,$|\mathbf{v}_2\rangle$ and $|\mathbf{v}_3\rangle$ are further unknown parameters of the channel that are called *channel directions*.

Of course, the channel model first has to transform the state in the MUB $\{|\pm\mathbf{v}_i\rangle\}_{i=1}^3$ and – after scaling – it has to transform it back. This can be described most simply in the Bloch picture as $\mathcal{E}(\theta) = \mathbf{S}\theta + \mathbf{t} = \mathbf{V}\Lambda\mathbf{V}^{\mathrm{T}}\theta$. We see that for Pauli channels $\mathbf{S}$ is a real symmetric matrix and $\mathbf{t} = 0$.

Note that in the group theoretical approach the rotation of the channel directions with an orthogonal $\mathbf{V}$ corresponds to the conjugation of the Kraus operator elements with the corresponding unitary $U_{\mathbf{V}}$.

2.4 Measures in quantum information

2.4.1 Distance of states

The distance of quantum states has several common measures, for example the trace distance, the relative entropy, and the *fidelity*. The latter is used in this thesis, so it is introduced here. The general definition of the fidelity is

$$F(\rho_1, \rho_2) = \mathrm{Tr}\left(\sqrt{\rho_1^{\frac{1}{2}}\rho_2\rho_1^{\frac{1}{2}}}\right) .$$

The fidelity takes values between 0 and 1, i.e., $0 \leq F(\rho_1, \rho_2) \leq 1$. If $\rho_1 = \rho_2$ then $F(\rho_1, \rho_2) = 1$, and if ρ_1 and ρ_2 have orthogonal support then $F(\rho_1, \rho_2) = 0$. It follows that the fidelity is not a distance in the mathematical sense. It also does not satisfy the triangle inequality. It is important that the fidelity is invariant under unitary transformations, i.e., $F(U\rho_1U^*, U\rho_2U^*) = F(\rho_1, \rho_2)$. We must also note that the fidelity is symmetric in its variables (see page 411 of [5]). In some special cases, the definition of the fidelity can be given more simply. If one of the states is pure then $F(|\psi\rangle, \rho) = \sqrt{\langle\psi|\rho|\psi\rangle}$, and for two pure states $F(|\psi\rangle, |\varphi\rangle) = |\langle\psi|\varphi\rangle|$.

In some applications, the Hilbert–Schmidt norm arising from (A.1) is also used as a distance measure between density matrices or Choi matrices of quantum channels: $\|\rho_1 - \rho_2\|_{\mathrm{HS}} = \sqrt{\mathrm{Tr}\left[(\rho_1 - \rho_2)^2\right]}$.

2.4.2 Channel performance

We also need to measure how much does a channel preserve quantum information. It must be emphasized that this is in general not the same as preserving the input state of the channel, because in many applications this input system can be in a mixed state, i.e., correlated (entangled) with its environment. In such cases the exact correlations – the one specific purification among the many possible – must be preserved!

Formally, let the system $\mathcal{H}$ be in state ρ and let $|\psi\rangle \in \mathcal{H} \otimes \mathcal{H}_{\mathrm{ref}}$ be a purification of ρ with a reference system $\mathcal{H}_{\mathrm{ref}}$, i.e., $\rho = \mathrm{Tr}_2(|\psi\rangle\langle\psi|)$. Then the *entanglement fidelity* [15] measures how well the channel $\mathcal{E}$ preserves the

state $|\psi\rangle$, in other words, how well $\mathcal{E}$ preserves the entanglement of ρ with the reference system:

$$F_{\text{ent}}(\rho, \mathcal{E}) \coloneqq F(|\psi\rangle, (\mathcal{E} \otimes \mathcal{I})(|\psi\rangle\langle\psi|))^2 = \langle\psi|\big[(\mathcal{E} \otimes \mathcal{I})(|\psi\rangle\langle\psi|)\big]|\psi\rangle$$

Thus the entanglement fidelity is derived from the fidelity. If the entanglement is preserved then $F_{\text{ent}}(\rho, \mathcal{E}) = 1$. The entanglement fidelity has the following properties derived from the properties of the fidelity:

- Linear in the channel,
- Its value is independent of the purifying system $\mathcal{H}_{\text{ref}}$,
- For pure states $F_{\text{ent}}(|\phi\rangle, \mathcal{E}) = F\big(|\phi\rangle, \mathcal{E}(|\phi\rangle\langle\phi|)\big)^2$. This shows that the entanglement fidelity measures the preservation of the state as well as entanglement.[17]

If the Kraus elements V_k of the channel are known then we get the useful formula

$$F_{\text{ent}}(\rho, \mathcal{E}) = \sum_k |\text{Tr}(\rho V_k)|^2 \ . \tag{2.20}$$

Notice that the entanglement fidelity depends not only on the channel, but also on the input state. To avoid this, in this thesis we always use the completely depolarized state as an input, which is the reduced density matrix of the maximally entangled state. If the channel can preserve the maximally entangled state, then it can preserve other entangled states, too. This special case of the entanglement fidelity is sometimes called *channel fidelity* F_{ch} [16], because it depends only on the properties of the channel.

[17] This implies that the quantity $F_{\text{ent}}(\rho, \mathcal{E})$ may be lowered by local unitary operations on $\mathcal{H}$. If needed, this can be taken into account using a modified definition which allows for evolution of the channel input that does not decrease entanglement (see [15]).

Part I

Quantum error correction

Chapter 3

Theory of quantum error corretion

This chapter explains the basic principles of quantum error correction (QEC) building on the notions on quantum channels and summarizes the specialized results available in the literature, that were used in developing the results of Chapter 4. The general theory of quantum error correcting codes was developed in [17]. The notions discussed here are also based on [5, 18, 19].

Section 3.1 discusses the standard theory of QEC, and section 3.2 presents the optimization approach of QEC.

3.1 Standard QEC

The general aim of quantum information theory is to generalize the concepts and goals known from classical information theory (discussed in e.g. [20]) to a setting where the underlying physical laws are the laws of quantum mechanics, i.e., where the basic physical systems holding information are quantum mechanical systems and the physical processes acting on these information containers are also processes of quantum mechanical nature [5].

One of the leading subfields of quantum information theory is quantum error correction. It studies the possibilities of restoring quantum systems corrupted by the effect of noise-like quantum processes, thus enabling reliable information processing. As stated in the introduction and in section 2.2, quantum noise is essentially the development of unwanted and irreversible correlation between the system and its environment, leading to the loss of information stored in the system [5]. This process is also called *decoherence*[18] [21]. Despite the fundamental differences between classical and quantum systems discussed in section 2.1, the basic approaches of quantum error correction follow the principles of classical methods.

3.1.1 Basic theory

The key idea of quantum error correction is analogous to that of classical error correction; we must complement the original message with enough re-

[18]The name refers to the nature of this process, namely that in a special basis it suppresses the off-diagonal (coherence) elements of the density matrix, i.e., generates a mixed state.

dundancy, i.e., *encode* the message in such a way that the information content is recoverable after the noise process has acted on the encoded message.

The encoding

Making redundancy by producing independent copies of the message quantum state is forbidden in quantum mechanics (see the no-cloning theorem in section 2.1.2). To circumvent this limitation, the encoding is rather done by *distributing* the message state – the logical information – over a bigger system of correlated states.

Formally, a *quantum error correcting code* is a vector space $\mathcal{C}$ defined by the unitary embedding[19] $U_{\mathcal{C}}\colon \mathcal{H}_{\mathrm{ms}} \to \mathcal{C} \subset \mathcal{H}_{\mathrm{cs}}$ of the *message space* $\mathcal{H}_{\mathrm{ms}}$ – the information source – into the *code space* $\mathcal{H}_{\mathrm{cs}}$ containing the *codewords*.

The messages are usually carried by two-level systems thus codewords are also called *logical states* of the *logical qubits*. If m logical qubits are encoded into a block of c physical ones then we have a so-called $[[c, m]]$-*quantum code*, and $\dim(\mathcal{H}_{\mathrm{ms}}) = \dim(\mathcal{C}) = 2^m < 2^c = \dim(\mathcal{H}_{\mathrm{cs}})$.

From here on, we omit the coding/decoding and assume that the error correction procedure starts and ends in $\mathcal{C}$. Thus the error operators are of type $E_i\colon \mathcal{H}_{\mathrm{cs}} \to \mathcal{H}_{\mathrm{cs}}$ and the state $|\psi_{\mathrm{cs}}\rangle$ or ρ_{cs} will denote the encoded message $U_{\mathcal{C}}|\psi_{\mathrm{ms}}\rangle$ or $U_{\mathcal{C}}\rho_{\mathrm{ms}}U_{\mathcal{C}}^*$.

The error correctability conditions

Physical noise processes are represented by quantum channels in general. Each such noise channel has a set of Kraus operator elements, with the different elements corresponding to different error types. Thus it is reasonable to derive correctability conditions for sets of individual error operators, which can then be applied to any noise channel.

The set of n_{err} error operators $\{E_i\}$ is said to be *correctable on the code* $\mathcal{C}$, if a *recovery operation* $\mathcal{R}$ exists such that $(\mathcal{R} \circ \mathcal{E})(\rho) \propto \rho$ with $\mathrm{supp}(\rho) \in \mathcal{C}$.[20] The conditions on the existence of such an $\mathcal{R}$ is given by the following theorem:

Theorem 3.1 (Knill–Laflamme)**.** *The set* $\{E_i\}$ *is correctable perfectly on the code* $\mathcal{C}$ *with some* $\mathcal{R}$ *recovery operation if and only if*

$$P_{\mathcal{C}} E_i^* E_j P_{\mathcal{C}} = h_{i,j} P_{\mathcal{C}} \ , \tag{3.1}$$

where $P_{\mathcal{C}}$ *is the orthogonal projection onto* $\mathcal{C}$ *and the entries* $h_{i,j} \in \mathbb{C}$ *form a* $n_{err} \times n_{err}$ *Hermitian matrix* h*.*

An important result allowing effective error correction states that if the set of operators $\{E_i\}$ is correctable with $\mathcal{R}$ on $\mathcal{C}$, then any set of linear combinations $\{F_j \,|\, F_j = \sum_i c_{j,i} E_i,\ c_{j,i} \in \mathbb{C}\}$ is also correctable with $\mathcal{R}$ on $\mathcal{C}$. This

[19] Our discussion of quantum error-correction assumes that encoding and decoding of quantum states can be done perfectly, without error. In reality, the theory of fault-tolerant quantum computation has also to be considered.

[20] The symbol $\propto$ means proportionality here.

shows that the error operators correctable with $\mathcal{R}$ on $\mathcal{C}$ form an at most n_{err} dimensional complex linear space $\mathcal{V}_{\mathcal{R},\mathcal{C},\{E_i\}} = \text{span}(\{E_i\})$. Moreover, there always exist a set of $n_{\text{s}} = \dim(\mathcal{V}_{\mathcal{R},\mathcal{C},\{E_i\}})$ independent linear combinations $\{D_\mu\}$ in $\mathcal{V}_{\mathcal{R},\mathcal{C},\{E_i\}}$ such that the matrix h in (3.1) is diagonal.

As an example, the Pauli basis forms the linear space of all single-qubit operators. Thus if the error operator set $\{\mathbb{1}^{\otimes c}, X_k, Y_k, Z_k\}$ is correctable with some $\mathcal{R}$ then any operator causing error only on a single qubit will become correctable using the same recovery operation $\mathcal{R}$.

The meaning of (3.1) is most intuitively understood through the set of errors $\{D_\mu\}$ defined above. Each D_μ in this special set isometrically rotates $\mathcal{C}$ into n_{s} mutually orthogonal subspaces $\mathcal{S}_\mu \subset \mathcal{H}_{\text{cs}}$ called *syndrome subspaces* with $\mathcal{S}_0 = \mathcal{C}$. More formally, $D_\mu P_\mathcal{C} = \sqrt{c_\mu} A_\mu W_\mu P_\mathcal{C}$, where W_μ is the isometry rotating $\mathcal{C}$ to $\mathcal{S}_\mu$ and A_μ is the unitary specifying the effect of the error. This error can then be identified by measuring the observable $\sum_\mu \mu P_\mu$ built from the projections P_μ onto each of the subspaces $\mathcal{S}_\mu$. The resulting μ specifies the *error syndrome*. This effectively tells us which error D_μ has occurred without disturbing the state. In fact the (projective) *syndrome measurement* identifies the subspace $\mathcal{S}_\mu$ by collapsing the corrupted codeword $\sum_\mu D_\mu|\psi_{\text{cs}}\rangle$ into $\sqrt{c_\mu} A_\mu W_\mu|\psi_{\text{cs}}\rangle \in \mathcal{S}_\mu$. This can then be recovered easily by rotating it back into the code subspace $\mathcal{C}$ with the operator $R_\mu = W_\mu^* A_\mu^*$. This procedure thus ensures that the message is perfectly preserved from the error operator set $\mathcal{V}_{\mathcal{R},\mathcal{C},\{D_\mu\}}$.

Of course a recovery $\mathcal{R}$ perfectly correcting the errors in $\mathcal{V}_{\mathcal{R},\mathcal{C},\{E_i\}}$ on the code $\mathcal{C}$ is not unique. Moreover, it may correct some errors outside $\mathcal{V}_{\mathcal{R},\mathcal{C},\{E_i\}}$ too. It follows that in practice $\mathcal{R}$ not only corrects $\mathcal{V}_{\mathcal{R},\mathcal{C},\{E_i\}}$, but the whole space $\mathcal{V}_{\mathcal{R},\mathcal{C}}$ of errors it is able correct on $\mathcal{C}$.[21] This can result in slight performance differences for otherwise equivalent codes (see appendix A.5.4 for an example).

Efficiency of QEC

The quantum Hamming bound (see appendix A.5.3) implies that we have limited space in $\mathcal{H}_{\text{cs}}$ for the $\mathcal{S}_\mu$ syndrome subspaces, i.e., a limited number of independent errors to correct. Thus the best strategy is evidently to correct the most probable errors caused by a certain noise channel.

It is a common approximation to assume that the error operators are tensor products of single-qubit operators, in other words, the noise channel is such that the action of its error operators on different qubits is uncorrelated. In this context we can speak about the *weight* of an error operator. The weight of an uncorrelated error operator E is the number of non-identity operators in the tensor product expansion of E. For instance, suppose we have a two-level

[21] Operators acting differently on the whole $\mathcal{H}_{\text{cs}}$ can still act the same way on $\mathcal{C}$, i.e., two independent correctable errors can map the codewords into dependent states. If there are such errors in $\mathcal{V}_{\mathcal{R},\mathcal{C},\{E_i\}}$, then the code $\mathcal{C}$ is said to be *degenerate* with respect to $\mathcal{V}_{\mathcal{R},\mathcal{C},\{E_i\}}$. This is equivalent to $\text{rank}(h)$ from (3.1) being less than $\dim(\mathcal{V}_{\mathcal{R},\mathcal{C},\{E_i\}})$, meaning that we only need to actively correct a subspace of $\mathcal{V}_{\mathcal{R},\mathcal{C},\{E_i\}}$. Degeneracy is in fact a non-classical feature that only quantum codes can have.

noise channel $\mathcal{E}$ with operator elements $\{\sqrt{1-p_\mathrm{n}}\mathbb{1}, \sqrt{p_\mathrm{n}}F\}$, where p_n is the probability of error operator F acting on a qubit. Then the total effect of the noise on the codeword ρ_cs can be expanded as

$$\mathcal{E}^{\otimes c}(\rho_\mathrm{cs}) \approx (1-p_\mathrm{n})^c \rho_\mathrm{cs} + \sum_{i=1}^{c} (1-p_\mathrm{n})^{c-i} p_\mathrm{n}^i \sum_{E \in \left\{ \substack{\text{weight-}i \\ \text{errors}} \right\}} E \rho_\mathrm{cs} E^* \ .$$

In such an expansion, the errors acting on fewer qubits, i.e., the errors with smaller weight are more probable. It follows that the correction of these can improve resistance against such uncorrelated noise models. This statement is similar for qubit channels with multiple non-identity error operators, and most generally holds also for channels without an operator element proportional to $\mathbb{1}$, provided that the noise is sufficiently weak.[22] Based on this, constructing a quantum code to perfectly correct the minimum weight errors, i.e., errors corrupting the least number of qubits is called the *standard strategy of QEC*. This strategy can achieve an adequate level of noise reduction independently of the actual type of the noise channel.

3.1.2 Pauli errors on stabilizer codes

We know from section 2.3.1 that scaled elements of the Pauli group $\mathcal{P}_c$ are a special class of error operators and noise channels having such operator elements are the Pauli channels. The specialty of Pauli error operators is most apparent using a certain class of quantum codes, the so-called stabilizer codes (see appendix A.5.2). Recall from section 3.1.1 that for general errors the corrupted codeword was sent into one of the syndrome subspaces through measurement. In contrast, because Pauli errors either commute or anticommute with the generators g_i of the stabilizer code, they directly rotate the codeword into a syndrome subspace.

This fact can be used to classify Pauli errors and study their correction differently than in appendix A.5.2. Let S be the stabilizer of the code $\mathcal{C}_S$, and $N(S)$ be its normalizer in $\mathcal{P}_c$. Then it is known that the group theoretic relations $S < N(S) < \mathcal{P}_c$ hold, which means that $\mathcal{P}_c$ can be partitioned to left cosets $\mathcal{P}_c/N(S)$, and $N(S)$ can be partitioned to left cosets $N(S)/S$. By choosing fixed representatives $\mathrm{i}^r A_p$ in the cosets of S and W_μ in the cosets of $N(S)$ we can uniquely associate any Pauli operator g with $s \in S$ and indices (p, μ, r) as $g = \mathrm{i}^r A_p W_\mu s$, such that

- the $A_p \in N(S)$ normalizer operators are logical Pauli operators, which act unitarily within $\mathcal{C}_S$, i.e., $A_p S A_p = \langle A_p g_1 A_p, \ldots, A_p g_{c-m} A_p \rangle = S$. r is actually unimportant, because $\mathrm{i}^r A_p S$ is in the same coset as $A_p S$ and (2.8) shows that $\mathrm{i}^r g$ acts the same way on a density operator as g. Thus practically only the factor group $\mathcal{P}_c/\langle \mathrm{i} \rangle$ is important.

[22] This assumed local nature of the noise is compatible with the expectation that the distribution of information into nonlocal correlations (entanglement) by coding will help us to preserve it.

- the W_μ operators are isometries for which $W_\mu S W_\mu^*$ stabilizes the syndrome subspace $\mathcal{S}_\mu$. The orthogonal projection onto $\mathcal{S}_\mu$ will be $P_\mu = \prod_{i=1}^{c-m} \frac{1}{2}\big(\mathbb{1} + W_\mu g_i W_\mu^*\big)$, and $W_\mu g_i W_\mu^* = \pm g_i$ proves that the syndrome subspaces are orthogonal.

This way all Pauli operators are in one of the left cosets of the stabilizer $A_p W_\mu S$, where μ indexes the isometric mapping of $\mathcal{C}_S$ onto $\mathcal{S}_\mu$ and p indexes a further unitary operation. Note that by definition W_μ and A_p commute.

It is now apparent that the operators from the same coset $A_p W_\mu S$ transform the code $\mathcal{C}_S$ equivalently, that is, they cause the same error which can be corrected in the same way.[23] Therefore, given a noise channel with operator elements $\{E_i = \sqrt{a_i}\, g_i\}$ with $g_i \in \mathcal{P}_c$ it is convenient to introduce the equivalence classes $[A_p W_\mu]$ and identify them with the error operator

$$E_{p,\mu} = \sqrt{a_{p,\mu}}\, g_{p,\mu} \coloneqq \sqrt{a_{p,\mu}} A_p W_\mu \ , \tag{3.2}$$

where $a_{p,\mu} = \sum_{\{i \,|\, E_i \in [A_p W_\mu]\}} a_i$ is the total probability of the error $g_{p,\mu}$ in the channel.

Suppose $\mu \neq 0$. Then the error class $E_{p,\mu}$ can be corrected obviously by $E_{p,\mu}^* = W_\mu^* A_p$. However, the syndrome measurement can only identify μ, but not p. This implies that two error operators E_{p_1,μ_1} and E_{p_2,μ_2} can be corrected at the same time if and only if $\mu_1 \neq \mu_2$, i.e., they map the codeword into different syndrome subspaces. This means that we can perfectly correct only one error class from each coset $W_\mu N(S)$.

In the $\mu = 0$ case $E_{p,0} = \sqrt{a_{p,0}} A_p \in N(S)$, which means that either $p = 0$ and the codeword remained valid or $p \neq 0$ and the error is undetectable thus uncorrectable.

As an example on quantum stabilizer codes, the so-called five-qubit code is discussed in appendix A.5.4. This type of code was used in the thesis, too (see later in Chapter 4).

3.2 Optimal QEC

As we have seen, standard QEC is a completely noise independent procedure, i.e., it makes no assumptions on the type of the noise process, rather its aim is to perfectly correct a fixed group of commonly probable errors. In contrast, if the type of the noise process is known, it is more reasonable to try protecting the information against that specific process type. This approach is called the *optimization approach of QEC*. Its goal is to find the optimal recovery operation $\mathcal{R}^\star$ given a code $\mathcal{C}$ and a noise channel $\mathcal{E}$. In the literature authors mostly follow this approach [22, 23, 24] although there are also alternate attempts [25, 26].

A recovery operation $\mathcal{R}^\star$ is said to be optimal, if it maximizes the channel fidelity F_{ch}. The reason for this choice of objective function is apparent

[23] When there are more than one error operator in the same coset of S which is correctable by some $\mathcal{R}$ recovery operator, then it is the case of degenerate QEC.

based on section 2.4.2; error correction has to take into account the possibility that it is only applied to a part of a larger quantum system, i.e., it not only has to preserve the state on which it is applied, but also has to preserve the entanglement between the state and parts of its environment.

More formally, the following optimization problem has to be solved:

$$\mathcal{R}^\star = \arg\max_{\{\mathcal{R}\}} F_{\text{ch}}(\mathcal{R} \circ \mathcal{E}) \ . \tag{3.3}$$

In contrast to a Kraus operator element set the Choi matrix is a unique channel representation in the sense that it has no unitary freedom (see section 2.2.2), which makes it suitable for optimization purposes. Using this, (2.20), (2.10) and (2.11), the objective of (3.3) turns into

$$F_{\text{ch}}(\mathcal{R} \circ \mathcal{E}) = \sum_{i,j} \langle\!\langle \rho_{\text{cs}} | R_i E_j \rangle\!\rangle \langle\!\langle R_i E_j | \rho_{\text{cs}} \rangle\!\rangle = \langle\!\langle \rho_{\text{cs}} | X_{\mathcal{R}\circ\mathcal{E}} | \rho_{\text{cs}} \rangle\!\rangle = \text{Tr}(X_{\mathcal{R}} C_{\rho_{\text{cs}},\mathcal{E}}) \ .$$

where the R_i are the Kraus elements of $\mathcal{R}$, $C_{\rho_{\text{cs}},\mathcal{E}} \coloneqq \sum_i |\rho_{\text{cs}} E_i^* \rangle\!\rangle \langle\!\langle \rho_{\text{cs}} E_i^*|$, and $\rho_{\text{cs}} = U_{\mathcal{C}} \frac{\mathbb{1}}{2^m} U_{\mathcal{C}}^* = \frac{P_{\mathcal{C}}}{2^m} \in \mathcal{B}(\mathcal{H}_{\text{cs}})$ is the codeword of the maximally entangled message state $\frac{\mathbb{1}}{2^m}$.

Taking into account that $\mathcal{R}$ must be CPTP, the final form of (3.3):

$$\begin{gathered} X_{\mathcal{R}}^\star = \arg\max_X \text{Tr}(X C_{\frac{P_{\mathcal{C}}}{2^m},\mathcal{E}}) \ , \\ \text{so that } X \geq 0, \text{ and } \text{Tr}_2(X) = \mathbb{1} \ . \end{gathered} \tag{3.4}$$

This problem is a semidefinite programming problem, a class of convex optimization problems for which efficient solvers exist (see section A.6). The properties of such a problem can be seen in appendix A.2.3.

Note that for a $[[c, m]]$ code, (3.4) is a $2^{4c} - 2^{2c}$ (real) dimensional optimization problem. In practice the number of dimensions can be reduced to $2^{2(m+c)} - 2^{m+c}$ by merging the noise and recovery operator elements with the encoder as $E_i' \coloneqq E_i U_{\mathcal{C}}$, $R_i' \coloneqq U_{\mathcal{C}}^* R_i$ and operating directly on the message space $\mathcal{H}_{\text{ms}}$.

3.2.1 Pauli case

In the case of Pauli channels and stabilizer codes the optimal correction operation $\mathcal{R}^\star$ can be generated analytically [19, 25]. This is a very essential result, as it applies to many important channels; furthermore, it may also help understand more complex cases where only numerical methods are available.

As stated in section 3.1.2, only one Pauli error class $E_{p,\mu} = \sqrt{a_{p,\mu}} A_p W_\mu$ can be corrected for each μ. It is then apparent that the correction will be the most effective if for each μ we choose and correct the most probable error. In contrast, standard QEC chooses to correct the minimum weight (more often only the single-qubit) errors. This means that for stabilizer codes and Pauli channels, standard QEC will be optimal except when the error probabilities are sufficiently unequal. Further comparison of standard and optimal QEC with examples can be seen in appendix A.5.5.

Formalizing the above, let the index p of the most probable error operator A_p for each $W_\mu N(S)$ coset be denoted by $p^\star_\mu$. Now the operator element set of the optimal $\mathcal{R}^\star$ recovery operation can be obtained as $\{W^*_\mu A_{p^\star_\mu}\}$, $\mu = 0, \ldots, 2^{c-m} - 1$. Using this, the channel fidelity turns into

$$F_{\mathrm{ch}}(\mathcal{R}^\star \circ \mathcal{E}) = \sum_{p,\mu} \sum_{\mu'} \left| \mathrm{Tr}\left(W^*_{\mu'} A_{p^\star_{\mu'}} \sqrt{a_{p,\mu}} A_p W_\mu \frac{P_\mathcal{C}}{2^m} \right) \right|^2 = \sum_\mu a_{p^\star_\mu,\mu} \ .$$

We used that $[W_\mu, A_p] = 0$, $W^*_{\mu'} W_\mu = \delta_{\mu,\mu'} P_\mathcal{C}$ and the logical Pauli operators are mutually Hilbert–Schmidt orthogonal, i.e., $\mathrm{Tr}\big(A_p A_{p'}\big) = \delta_{p,p'} 2^m$.

By this formula, the channel fidelity for Pauli channels is thus the sum of the $a_{p_\mu,\mu}$ probabilities of the errors selected for correction in each syndrome subspace. It is evident that we get the maximal channel fidelity if we design $\mathcal{R}$ to correct the error with the greatest probability in each syndrome subspace. Strictly speaking, we have proved only that among the $(2^{2m})^{2^{c-m}}$ different QEC operations of the form

$$\mathcal{R} = \{W^*_\mu A_{p_\mu}\} \tag{3.5}$$

there does not exist any better than the one with $p_\mu = p^\star_\mu$. However, there is not any other kind of QEC operation which is better (see [19] for a complete proof).

Chapter 4

Robustness of quantum error correction

This chapter summarizes a set of new results. The robustness of QEC procedures against channel perturbations is first defined. Using this definition, the robustness of recovery operations for the class of Pauli channels is analyzed in detail. The robustness domains splitting up the set of Pauli channels are explicitly characterized. Furthermore, several case studies on the robustness of recovery operations against general non-Pauli perturbating channels is also studied using the same definitions, with the aim of finding possible generalization of the results for Pauli channels, in particular the general robustness domains and domain borders.

In section 4.1 the general definitions of robustness and related notions is given. Section 4.2 discusses robustness for the case of Pauli channels, while section 4.3 deals with the general case. Finally, 4.4 summarizes the results.

4.1 Robustness in the context of QEC

As we have seen in section 3.2 the optimal QEC is clearly the best in terms of efficiency. In real world situations, however, robustness should also be considered as an indicator of usefulness.

As we mentioned in the introduction 1.1, any precise definition of robustness against a change must also include the accurate description of the changing conditions together with the bounds of their change. In system-and control theory we can speak about robustness of a system model property against uncertainties arising from the differences between the model and reality. Representations of these uncertainties vary primarily in terms of the amount of structure they contain [3]. Unmodeled dynamics, nonlinearities, effects of linearization and system randomness are usually called *unstructured uncertainties* (also nonparametric uncertainties). These uncertainties are usually defined as set membership statements for the model. In contrast, *structured uncertainty* (or parametric uncertainty) assumes a known system model, only its actual parameters are uncertain. Therefore, structured uncertainty is defined as set membership statements (intervals) for the parameters. Robustness of some

property can then be defined as resistance against the variations in the given set of uncertainties.

In the context of quantum error correction robustness of recovery operations can also be analyzed using a similar classification of uncertainties in the noise channel. One possibility is to study robustness against variations in the channel parameters, i.e, assume structured uncertainties. This approach is shortly discussed by Fletcher in [19]. Another approach of [19] is to try taking a – less structured – uncertainty into account in the formulation of the optimization problem. They define the channel as an integral weighted by a probability distribution then maximize the average entanglement fidelity. A similar problem statement can be found in the work [26], which tries to solve the problem with a reformulated, "indirect" version of the optimization problem. Of course, to use these approaches we must first consider by what assumptions the distribution will be physically realistic.

The drawback of the above-mentioned approaches is that they assume a very special type of uncertainty structure. These can be very useful in certain cases, but here we want to study the question of robustness from a more general point of view, by assuming the least possible structure in the uncertainties of the noise channel. By knowing less about the uncertainty we can be less prepared against it.

We seek answers for this question in the following way. Let $\mathcal{R}_0$ be a recovery operation originally used to correct the nominal noise channel $\mathcal{E}_0$. First, we characterize the robustness of $\mathcal{R}_0$ against the structured channel set obtained by moving in the set of channels from $\mathcal{E}_0$ in the direction given by a perturbation channel $\mathcal{E}_1$, and checking how effective the recovery operation $\mathcal{R}_0$ remains on the new altered channel. Then by removing the knowledge of $\mathcal{E}_1$, we generalize our results to obtain an exact description of the neighborhood of $\mathcal{E}_0$ in the set of channels on which $\mathcal{R}_0$ is a robust recovery operation. This way we get information on the robustness of $\mathcal{R}_0$ against completely unstructured type of uncertainties.

4.1.1 Perturbation of quantum channels

To describe moving from the nominal noise channel $\mathcal{E}_0$ towards the perturbation channel $\mathcal{E}_1$, i.e., the *mixing* of the channels, we can utilize the convexity of the set of channels and define channel mixing as the convex combination of two channels. The operator elements of the mixed channel $\mathcal{E}_\gamma = (1-\gamma)\mathcal{E}_0 + \gamma\mathcal{E}_1$ will be the following:

$$\{E_{\gamma,i}\} = \left\{\sqrt{1-\gamma}E_{0,j}\right\} \uplus \left\{\sqrt{\gamma}E_{1,k}\right\} , \tag{4.1}$$

where $\{E_{0,j}\}$ and $\{E_{1,k}\}$ are the sets of operator elements of $\mathcal{E}_0$ and $\mathcal{E}_1$; furthermore, $\gamma \in [0,1]$ is the *mixing parameter*. Notation $\uplus$ means here that if $\{E_{0,j}\}$ and $\{E_{1,k}\}$ contains an element which is the same in these two sets then this common element will appear two times in $\{E_{\gamma,i}\}$. It is easy to see that the obtained operator element set still represents a quantum channel as channels form a convex set, so the trace preserving constraint (2.9) remains valid.

Thus a parameterized set $\mathcal{E}_\gamma$ of channels is obtained as a model of structured uncertainty with structure imposed on it by $\mathcal{E}_0$ and $\mathcal{E}_1$.

Note that mixing could also be defined using the corresponding Choi matrices: $X_{\mathcal{E}_\gamma} = (1-\gamma)X_{\mathcal{E}_0} + \gamma X_{\mathcal{E}_1}$.

4.1.2 Efficiency of correction on the mixed channel

To get closer to the notion of robustness of a recovery operation, we must measure the effectiveness of $\mathcal{R}_0$ on the mixed channel $\mathcal{E}_\gamma$. The effectiveness of a recovery operation $\mathcal{R}$ on the channel $\mathcal{E}$ is given by the channel fidelity $F_{\text{ch}}(\mathcal{R} \circ \mathcal{E})$, however, instead of just working with the pure value, it is worth comparing it with the best possibility $\mathcal{R}^\star_\mathcal{E}$ on the channel $\mathcal{E}$ with some given encoding. Thus we arrive at the definition:

Definition 4.1 (Relative efficiency)**.** *The relative efficiency of the recovery operation $\mathcal{R}$ for a channel $\mathcal{E}$ is*

$$\epsilon_\mathcal{R}(\mathcal{E}) \coloneqq F_{\text{ch}}(\mathcal{R}^\star_\mathcal{E} \circ \mathcal{E}) - F_{\text{ch}}(\mathcal{R} \circ \mathcal{E}) \ .$$

Using this quantity we can analyze whether the efficiency of $\mathcal{R}_0$ remains close to the best possible value or not for $\mathcal{E}_\gamma$, i.e., while we alter the nominal channel $\mathcal{E}_0$ by mixing it with $\mathcal{E}_1$. This way it turns out against which type of noise $\mathcal{E}_1$ is the recovery operation sensitive and against which type it remains efficient (i.e., closely optimal). Thus the relative efficiency of $\mathcal{R}_0$ for $\mathcal{E}_\gamma$ will be

$$\epsilon_{\mathcal{R}_0}(\mathcal{E}_\gamma) = F_{\text{ch}}(\mathcal{R}^\star_\gamma \circ \mathcal{E}_\gamma) - F_{\text{ch}}(\mathcal{R}_0 \circ \mathcal{E}_\gamma) \ . \tag{4.2}$$

The first term is the channel fidelity for the case when the mixed noise channel $\mathcal{E}_\gamma$ is corrected by the best possible $\mathcal{R}^\star_\gamma$ recovery operation – which gives the biggest possible channel fidelity – and the second term tells us how much smaller is the channel fidelity we get compared to this optimal value by using the original $\mathcal{R}_0$ recovery operation.

This difference can be greater than zero because of two reasons. First, it can be greater than zero if $\mathcal{R}_0$ is not robust. Second, it is possible that the examined $\mathcal{R}_0$ recovery operation is robust, but not optimal for the $\mathcal{E}_0$ channel. The $\epsilon_{\mathcal{R}_0}(\mathcal{E}_0)$ difference, which is independent of the mixing parameter γ characterizes optimality, i.e., how effective the $\mathcal{R}_0$ recovery for the $\mathcal{E}_0$ channel is.

4.1.3 Robustness domains

We want to use the expression "robustness" to characterize the feature of the recovery operation, which shows how the efficiency of $\mathcal{R}_0$ changes with the mixing parameter γ. Thus it turns out that the robustness against the perturbation of the channel can be measured by the inspection of the γ dependence of the $\epsilon_{\mathcal{R}_0}(\mathcal{E}_\gamma)$ function. If $\epsilon_{\mathcal{R}_0}(\mathcal{E}_\gamma)$ grows fast with γ then the $\mathcal{R}_0$ correction is sensitive to the perturbation with $\mathcal{E}_1$ or else when $\mathcal{R}_0$ remains closely as effective as it was for the $\mathcal{E}_0$ channel, it is robust agaist $\mathcal{E}_1$. The robustness could be

characterized as the derivative $\frac{\mathrm{d}}{\mathrm{d}\gamma}\epsilon_{\mathcal{R}_0}(\mathcal{E}_\gamma)$; however, it is more useful to define the robustness more generally with an $\mathcal{E}_1$ independent *domain* in which the efficiency of $\mathcal{R}_0$ is not worse than it is for the original channel plus a given δ value:

Definition 4.2 (δ-robustness domain)**.** *The boundary point of the δ-robustness domain in the direction given by $\mathcal{E}_1$ is the channel $\mathcal{E}_{\gamma_\delta}$, where*

$$\gamma_\delta = \inf\{\gamma \,|\, \epsilon_{\mathcal{R}_0}(\mathcal{E}_\gamma) - \epsilon_{\mathcal{R}_0}(\mathcal{E}_0) > \delta\}$$

The whole δ-robustness domain of a recovery operation $\mathcal{R}_0$ around the channel $\mathcal{E}_0$ consists of all the channels which are inside the boundary.

To determine the boundary of the δ-robustness domain we need to know (4.2). The second term gives us how good channel fidelity we get for the channel $\mathcal{E}_\gamma$ using the $\mathcal{R}_0$ recovery operation. It can be seen that we do not need optimization to determine this term. Using the operator elements of the mixed channel $\mathcal{E}_\gamma$ as in (4.1),

$$F_{\mathrm{ch}}(\mathcal{R}_0 \circ \mathcal{E}_\gamma) = (1-\gamma)F_{\mathrm{ch}}(\mathcal{R}_0 \circ \mathcal{E}_0) + \gamma F_{\mathrm{ch}}(\mathcal{R}_0 \circ \mathcal{E}_1)\ . \tag{4.3}$$

That is, this term varies linearly in function of γ as the distance from the channel $\mathcal{E}_0$ grows. This follows from our definition of channel mixing and the linearity of the trace.

It is more difficult to handle the first term of (4.2). In the most general case, this calculation requires the solution of the semidefinite programming problem (3.4), as it was discussed in section 3.2.

4.2 Robustness in the Pauli channel case

This section draws conclusions about the robustness of Pauli channel error correction on stabilizer codes using the notions defined in section 4.1.

Assume that $\mathcal{E}_0$, $\mathcal{E}_1$ and so $\mathcal{E}_\gamma$ are Pauli channels. Then we can directly apply the analytical results of section 3.2.1 to determine the first term of the relative efficiency (4.2). Therefore, the operator elements from (4.1) can be written equivalently in terms of Pauli error classes (3.2) as

$$\left\{\sqrt{(1-\gamma)a^0_{p,\mu} + \gamma a^1_{p,\mu}}A_p W_\mu\right\}_{p,\mu}\ ,$$

where $\sqrt{a^0_{p,\mu}}$ and $\sqrt{a^1_{p,\mu}}$ are the coefficients of the operator element classes of the channels $\mathcal{E}_0$ and $\mathcal{E}_1$, respectively, and $\sqrt{a^\gamma_{p,\mu}} \coloneqq \sqrt{(1-\gamma)a^0_{p,\mu} + \gamma a^1_{p,\mu}}$ are the coefficients of the operator element class of the channel $\mathcal{E}_\gamma$.

To obtain the optimal correction for the mixed channel $\mathcal{E}_\gamma$, we apply the expression (2.20). For this, we need to determine the most probable one among the errors which move the code subspace $\mathcal{C}$ into a given syndrome subspace $\mathcal{S}_\mu$, i.e., the greatest $a^\gamma_{p,\mu}$ for a given μ.

4.2.1 Results for the case of single syndrome subspace

To see what exactly happens in each syndrome subspace with the optimal recovery operation during mixing, let us first take a simple case $a^1_{p,\mu} = \delta_{p,\bar{p}}\delta_{\mu,\bar{\mu}}$, where $\bar{p}$ and $\bar{\mu}$ are fixed, i.e., $\mathcal{E}_1$ has the single operator element $A_{\bar{p}}W_{\bar{\mu}}$. In this case, only in the syndrome subspace $\mathcal{S}_{\bar{\mu}}$ will there be any change in the ratios of the $a^0_{p,\mu}$ probabilities, so there is no need to examine any other syndrome subspace, as in those the most probable error remains unchanged during the mixing. With the change in the γ parameter in the mixed channel, only the probability $a^\gamma_{\bar{p},\bar{\mu}}$ grows; the other probabilities are decreasing. There are three cases:

1. $a^0_{\bar{p},\bar{\mu}}$ is smaller than some other $a^0_{p^\star_{\bar{\mu}},\bar{\mu}}$, i.e., in the channel $\mathcal{E}_0$ it is not the $A_{\bar{p}}W_{\bar{\mu}}$ error operator belonging to the channel $\mathcal{E}_1$, which is corrected by the current optimal recovery $\mathcal{R}^\star_0$. In this case, however, the probability of the error $A_{\bar{p}}W_{\bar{\mu}}$ is increasing as $a^\gamma_{\bar{p},\bar{\mu}} = (1-\gamma)a^0_{\bar{p},\bar{\mu}} + \gamma$ while we mix the channels; but the optimal recovery remains *unchanged* until the decreasing probability $a^\gamma_{p^\star_{\bar{\mu}},\bar{\mu}} = (1-\gamma)a^0_{p^\star_{\bar{\mu}},\bar{\mu}}$ remains greater.
2. $a^0_{\bar{p},\bar{\mu}}$ is uniquely the greatest error probability in the syndrome subspace $\mathcal{S}_{\bar{\mu}}$. In this case, it will increase further during the mixing while the other error probabilities will decrease, so $A_{\bar{p}}W_{\bar{\mu}}$ will be the most probable error and the optimal correction will be unchanged during the whole mixing.
3. Lastly, it can occur that $a^0_{\bar{p},\bar{\mu}}$ is one of the greatest error probabilities of $\mathcal{E}_0$ in the syndrome subspace $\mathcal{S}_{\bar{\mu}}$, i.e., there is more than one error operator with the same probability. (In other words, at least two different indices $p^\star_{\bar{\mu}} \neq p^{\star\star}_{\bar{\mu}} \neq \ldots$ exist, for which $a^0_{p^\star_{\bar{\mu}},\bar{\mu}} = a^0_{p^{\star\star}_{\bar{\mu}},\bar{\mu}} = \ldots$, and $p^\star_{\bar{\mu}} = \bar{p}$.) In this case, the optimal correction of the channel $\mathcal{E}_0$ is not unique. We can choose to correct any of the most probable errors and we get the same channel fidelity by these completely different recovery operations. But when the channel is altered by $\mathcal{E}_1$, for an arbitrary small value of γ the optimal recovery operation will uniquely be the one which corrects the error $A_{\bar{p}}W_{\bar{\mu}}$.

We can summarize the three cases in the following statement:

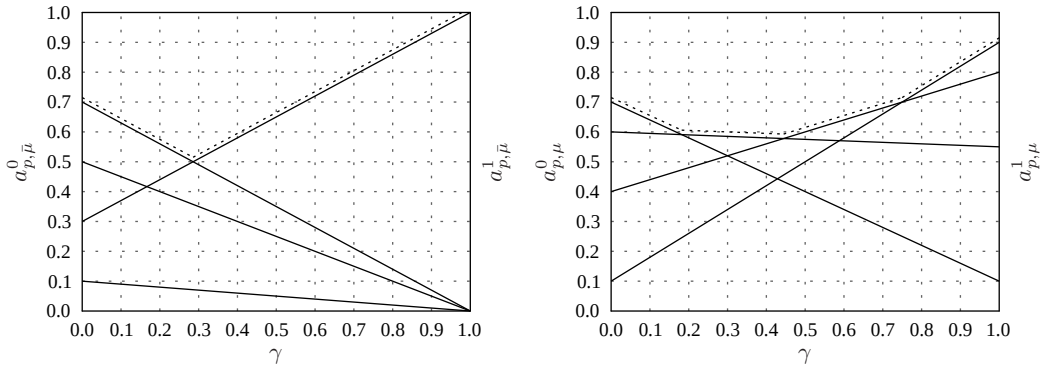
Statement 4.1 (Optimality in unitary case)**.** *Assume that $\mathcal{R}^\star_0$ – which is designed to correct the unitary error $\mathcal{E}_1 = A_{p^\star_{\bar{\mu}}}W_{\bar{\mu}}$ – is an unique optimal recovery operation for the channel $\mathcal{E}_0$. Then it is also optimal recovery for the channel $\mathcal{E}_\gamma$ arising as a mix of $\mathcal{E}_0$ with $\mathcal{E}_1$ if*

- *$\bar{p} \neq p^\star_{\bar{\mu}}$ and the inequality*

$$(1-\gamma)a^0_{p^\star_{\bar{\mu}},\bar{\mu}} \geq (1-\gamma)a^0_{\bar{p},\bar{\mu}} + \gamma \tag{4.4}$$

  holds,
- *$\bar{p} = p^\star_{\bar{\mu}}$ and γ is arbitrary.*

The above described behaviour of the $a^\gamma_{p,\bar{\mu}}$ probabilities is illustrated in Figure 4.1(a).



(a) The location of the breakpoint in case of unitary perturbation channel in the $\mathcal{S}_{\bar{\mu}}$ syndrome subspace. It can be seen that the probability of the perturbing error operator grows to 1, while the others decrease to 0.

(b) In the general case more breakpoints can appear.

Figure 4.1. These figures are only illustrations, there is no specific channel associated with them. Optimal channel fidelity is denoted by the dotted line.

The robustness domains

To obtain the δ-robustness domain of a recovery operation $\mathcal{R}_0$, we need to study the relative efficiency function (4.2). First, we assume that $\mathcal{R}_0 = \mathcal{R}_0^\star$ is the unique optimal error correction of the nominal channel $\mathcal{E}_0$. Then it remains optimal, under the conditions stated in **Statement 4.1**. This means that there exists a zero-robustness domain of nonzero size for $\mathcal{R}_0$ around $\mathcal{E}_0$. Thus we can state the following:

Statement 4.2 (Zero-robustness domain in unitary case)**.** *The boundary point of the zero-robustness domain for the $\mathcal{R}_0^\star$ unique optimal recovery of the channel $\mathcal{E}_0$ with perturbation $A_{\bar{p}}W_{\bar{\mu}}$ is that $\mathcal{E}_{\gamma_0}$ channel, for which γ_0 is the greatest possible solution of inequality* (4.4)*,*

$$\gamma_0 = \frac{a^0_{p^\star_{\bar{\mu}},\bar{\mu}} - a^0_{\bar{p},\bar{\mu}}}{a^0_{p^\star_{\bar{\mu}},\bar{\mu}} - a^0_{\bar{p},\bar{\mu}} + 1} \ ,$$

or $\gamma_0 = 1$ if $\bar{p} = p^\star_{\bar{\mu}}$. For the channel $\mathcal{E}_{\gamma_0}$, even the recovery operation $\mathcal{R}_0^\star$ is optimal, but not uniquely. Furthermore, exceeding this point, the optimal recovery operation changes abruptly and a completely different error correction will become optimal.

This can be seen also by looking at the channel fidelities in (4.2). Before the γ_0 boundary point, the optimal fidelity goes as

$$F_{\mathrm{ch}}(\mathcal{R}^\star_\gamma \circ \mathcal{E}_\gamma) = \sum_\mu (1-\gamma) a^0_{p^\star_\mu,\mu} = (1-\gamma) F_{\mathrm{ch}}(\mathcal{R}_0^\star \circ \mathcal{E}_0) \ ,$$

which is the same as the linear curve in (4.3), so $\epsilon_{\mathcal{R}_0}(\mathcal{E}_\gamma)$ is identically zero. (We used that $F_{\mathrm{ch}}(\mathcal{R}_0^\star \circ \mathcal{E}_1) = 0$.) But when $a^\gamma_{\bar{p},\bar{\mu}}$ exceeds $a^\gamma_{p^\star_{\bar{\mu}},\bar{\mu}}$, i.e., instead of

the correction of the error $A_{p^\star_{\bar\mu}}W_{\bar\mu}$, the correction of $A_{\bar p}W_{\bar\mu}$ will be the optimal solution, the optimal fidelity will go as

$$F_{\mathrm{ch}}(\mathcal{R}^\star_\gamma \circ \mathcal{E}_\gamma) = \sum_{\mu\neq\bar\mu}(1-\gamma)a^0_{p^\star_\mu,\mu} + (1-\gamma)a^0_{\bar p,\bar\mu} + \gamma \ .$$

As can be seen by comparing this equation and (4.3), the boundary of the δ-robustness domain for a $\delta > 0$ is that $\mathcal{E}_{\gamma_\delta}$ channel, for which

$$\gamma_\delta = \gamma_0 + \frac{\delta}{a^0_{p^\star_{\bar\mu},\bar\mu} - a^0_{\bar p,\bar\mu} + 1} \ .$$

In contrast, for every other $\mathcal{R}_0$ recovery operation of type (3.5), which is not optimal for the $\mathcal{E}_0$ channel, $\gamma_\delta \propto \delta$ without constant term.

4.2.2 Results for the general case

In general, when we mix the nominal Pauli channel $\mathcal{E}_0$ with an arbitrary other Pauli channel $\mathcal{E}_1$ which has more operator elements belonging to different p, μ index pairs, then in a given $\mathcal{S}_\mu$ syndrome subspace all error probabilities can vary differently, and this happens in more than one syndrome subspace at the same time. In this case the following generalization of **Statement 4.1** holds:

Statement 4.3 (Optimality in general case)**.** *The unique optimal recovery operation, which is designed to correct the errors $A_{p^\star_\mu}W_\mu$ in each syndrome subspace $\mathcal{S}_\mu$, remains optimal also for the mixed channel $\mathcal{E}_\gamma$ arising as a mix of $\mathcal{E}_0$ with $\mathcal{E}_1$ as long as the inequalities*

$$(1-\gamma)a^0_{p^\star_\mu,\mu} + \gamma a^1_{p^\star_\mu,\mu} \geq (1-\gamma)a^0_{p,\mu} + \gamma a^1_{p,\mu} \tag{4.5}$$

hold for all p and μ.

The above described behaviour of the $a^\gamma_{p,\mu}$ probabilities is illustrated in Figure 4.1(b). The general version of **Statement 4.2** can also be given:

Statement 4.4 (Zero-robustness domain in general case)**.** *The greatest value of γ for which the inequalities* (4.5) *are all true defines a γ_0, for which $\mathcal{E}_{\gamma_0}$ is the boundary point of the zero-robustness domain of $\mathcal{R}^\star_0$ in the direction given by $\mathcal{E}_1$. Going through this point, the optimal recovery operation changes abruptly and $\epsilon_{\mathcal{R}_0}(\mathcal{E}_\gamma)$ goes linearly with γ.*

It follows that the boundary point of the δ-robustness domain for $\delta > 0$ is that $\mathcal{E}_{\gamma_\delta}$ channel, for which $\gamma_\delta = \gamma_0 + \mathrm{const}\cdot\delta$ (which is 0 only if $\gamma_0 = 0$). In contrast, for an arbitrary non-optimal recovery operation $\mathcal{R}$ of type (3.5) used on a Pauli channel $\mathcal{E}_0$, the γ_δ value starts to increase linearly with δ in general, i.e., $\gamma_\delta = \mathrm{const}\cdot\delta$. Because of the global optimality of $\mathcal{R}^\star_0$ this is also true for any other type of QEC operation (including standard QEC recovery operation), having a growth with lower powers of δ at most in special cases. This brings us to one of our main conclusions:

Statement 4.5. *For the case of Pauli channels, the optimal recovery operation $\mathcal{R}^\star$ does not only give better entanglement fidelity but it remains robust in general, except for the case when the greatest common solution γ_0 of inequalities* (4.5) *is zero, i.e., the channel is on the boundary surface of a zero-robustness domain.*

Therefore, we should be careful when the standard QEC differs from the optimal. The latter one takes slight assumptions about the channel, but it does not guarantee the robustness at all. Surprisingly, the channel specific optimal solution can give a better guarantee for the robustness.

4.2.3 Geometric picture of Pauli robustness domains

Using the following geometric picture, we can easily determine and visualize the boundaries of the zero-robustness regions, in which the optimal $\mathcal{R}^\star$ recovery operation is the same. Since all the Pauli channels which have the same effect on the code subspace can be given by the numbers $0 \leq a_{p,\mu} \leq 1$ and because of the trace preserving condition $\sum_{p,\mu} a_{p,\mu} = 1$, all of these channels can be represented by a point of a $(2^{c+m}-1)$-dimensional simplex, which has 2^{c+m} extremal points.[24] The extremal points of the simplex will correspond to the unitary channels given by the equivalence classes $[A_pW_\mu]$ with $[A_0W_0]$ at the origin. A general Pauli channel with operator elements $\{\sqrt{a_{p,\mu}}A_pW_\mu\}$ then corresponds to the point of the simplex with coordinates $a_{p,\mu}$ $(p,\mu \neq 0)$. Any parameterized Pauli channel can then be associated with a curve in the simplex. In particular, for channels assuming that noise acts on each qubit independently the curve starts from the origin and ends on the face opposite to the origin. Moreover, channels arising as convex combination, i.e., which are parameterized by the mixing parameter γ have straight curves between the two endpoints.

Using this simplex model, it is easy to visualize the zero-robustness domains, i.e., those in which the optimal recovery operation is the same:

Statement 4.6 (Uniqueness of optimal recovery)**.** *A channel $\mathcal{E}_0$ has unique optimal recovery operation $\mathcal{R}_0^\star$ if and only if the $p_\mu^\star$ indices are uniquely determined for all μ by the strict inequalities $a^0_{p_\mu^\star,\mu} > a^0_{p,\mu}$ for all μ and $p \neq p_\mu^\star$.*

Mixing $\mathcal{E}_0$ with another channel $\mathcal{E}_1$, the optimal error recovery operation will remain $\mathcal{R}_0^\star$ until (4.5) holds, i.e., $a^\gamma_{p_\mu^\star,\mu}$ is maximal among $a^\gamma_{p,\mu}$. The point given by the $a^\gamma_{p,\mu}$ numbers is on the boundary between two (or more) domains if there exist at least one index μ and at least two indices $p^\star \neq p^{\star\star}$ for which $a^\gamma_{p^\star,\mu} = a^\gamma_{p^{\star\star},\mu} \geq a^\gamma_{p,\mu}$, where p can be arbitrary. These points are on the $(2^{n+k}-2)$-dimensional plane which is determined by the following $2^{n+k}-1$ points: all the extremal points except $A_{p^\star}W_\mu$ and $A_{p^{\star\star}}W_\mu$ and the midpoint between these two latter extremal points.

As a simple example, let us see a 4^m-1 dimensional projection of the simplex which we get as the convex hull of A_pW_μ vertices for a fixed μ. If we

[24] Given a $[[c,m]]$-stabilizer code, the factor group $G_c/\langle \mathrm{i} \rangle$ has 2^{2c} elements and the stabilizer S of the code has 2^{c-m} elements. Thus, $G_c/\langle \mathrm{i} \rangle$ factorized by S will have 2^{c+m} elements.

have one logical qubit, i.e., $m = 1$ then this simplex is three-dimensional. The zero-robustness domain borders can be obtained in the following way. Select any two vertices A_p, for instance $A_0 = \bar{\mathbb{1}}$ and $A_1 = \bar{X}$. Then the midpoint channel $\frac{1}{2}\bar{\mathbb{1}} + \frac{1}{2}\bar{X}$ of the line connecting these is a border channel, because $a_{0,\mu} = a_{1,\mu} = \frac{1}{2}$. If we go from this point towards the $A_2 = \bar{Z}$ vertex, then the weights $a_{0,\mu}$ and $a_{1,\mu}$ will decrease, but remain equal. This implies that we are still on a border as long as $a_{2,\mu} \leq a_{0,\mu} = a_{1,\mu} \leq \frac{1}{3}$. Proceeding similarly for all p we get the zero-robustness domains. This example can be seen in Figure 4.2.

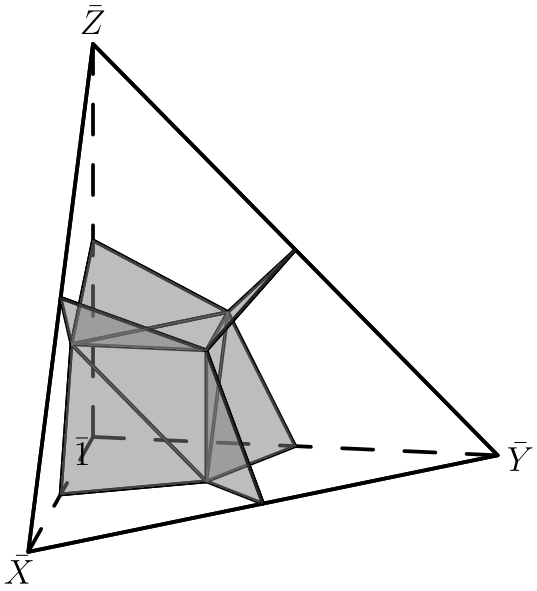


Figure 4.2. A 3 dimensional simplex with the zero-robustness domains inside.

Thus, the Pauli channels with the same optimal QEC operation form a zero-robustness domain in the space of Pauli channels. There are finite numbers of such domains and they cover the whole space of Pauli channels. If a channel is an interior point of such a domain then the optimal QEC operation is robust against *any* Pauli-type alteration of the channel, i.e., mixing the channel with any other Pauli channel $\mathcal{E}_1$, the boundary point $\mathcal{E}_{\gamma_0}$ of the zero-robustness domain is in non-zero γ_0 distance. Moreover, the boundary point of an arbitrary δ-robustness domain with $\delta > 0$ is outside the zero-robustness domain; therefore, $\gamma_\delta > \gamma_0$ and $\lim_{\delta\to 0} \gamma_\delta > 0$.)

As Pauli channels are examples of uncorrelated noise models, i.e., noise with operator elements given as tensor products of qubit operators, the following fact follows from the simplex picture:

Corollary 4.1. *Pauli channels with parameterized qubit operator elements are always inside the interior of a zero-robustness domain for small parameter values.*

For small parameter values, the minimal weight errors will be the most probable. Then these are associated with different syndrome subspaces and can be corrected at the same time. Thus there will never be two equal $a_{p,\mu}$

probabilities with the same μ, which means that the recovery operation for this channel will remain optimal for growing parameter values until an error with bigger weight gets more probable than one of the small weight errors. The more symmetric is the channel, i.e., the closer are the probabilities of the errors X, Y, and Z, the later will this happen. The most symmetric Pauli channel assuming uncorrelated noise is the depolarizing channel (A.5). For this, the optimal recovery is the same on the full domain of the noise parameter.

In Figure 4.3 an example can be seen for a strongly asymmetric channel. It has non-analytic breakpoints in the optimal channel fidelity curve in function of the noise parameter p, where the curve of the channel crosses robustness domain borders in the simplex, indicating an abrupt change of the optimal recovery operation. The most asymmetric case is the phase damping channel. It has a breakpoint at the 0 value of the noise parameter, so it is not visible in Figure A.1, that is, the optimal recovery is the same again on the full domain of the noise parameter.

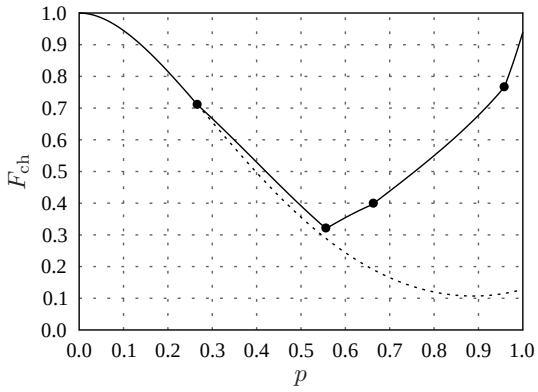


Figure 4.3. A strongly asymmetric Pauli channel crossing several zero-robustness domains with changing noise parameter p. The single-qubit operator elements of the channel: $\left\{\sqrt{(1-p)}\mathbb{1}, \sqrt{\frac{8}{10}p}X, \sqrt{\frac{2}{10}p}Y\right\}$. The dashed line shows that after the first breakpoint the optimal recovery changes.

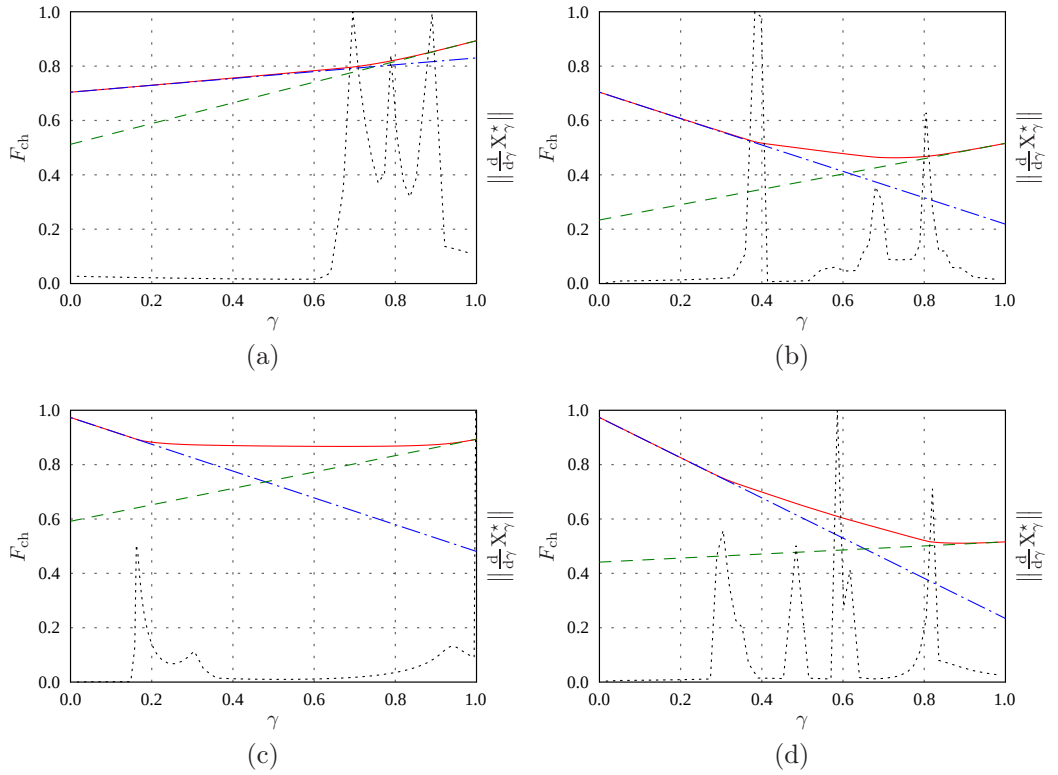


Figure 4.4. The two terms of $\epsilon_{\mathcal{R}_0}(\mathcal{E}_\gamma)$ in function of γ for the mixing of Pauli and non-Pauli channels. $\gamma = 0$ denotes the initial Pauli channel on each figure. The mixtures are (a) the depolarizing channel ($p = 0.3$) and the amplitude damping channel ($g = 0.3$); (b) the depolarizing channel ($p = 0.3$) and the pure states rotation channel ($\theta = \frac{5\pi}{12}, \phi = \frac{5\pi}{36}$); (c) the phase damping channel ($p = 0.3$) and the amplitude damping channel ($g = 0.3$); and (d) the phase damping channel ($p = 0.3$) and the pure states rotation channel ($\theta = \frac{5\pi}{12}, \phi = \frac{5\pi}{36}$). The solid line denotes the optimal channel fidelity term of $\epsilon_{\mathcal{R}_0}(\mathcal{E}_\gamma)$, the dashed and dotted-dashed lines denote the linear terms of $\epsilon_{\mathcal{R}_0}(\mathcal{E}_\gamma)$ for the two initial channels, and the dotted line denotes the approximate derivative $\left\|\frac{\mathrm{d}}{\mathrm{d}\gamma}X^\star_\gamma\right\|$, which is normalized to its maximum.

4.3 Case studies for non-Pauli channels

The results in section 4.2 are valid for Pauli channels and stabilizer codes, but based on some numerical results, we believe that the main conclusions are similar also in the case of arbitrary channels. To demonstrate the general situation, we study in detail a few examples using the definitions from section 4.1.

For this purpose, we choose two non-Pauli channels as perturbation channels; the amplitude damping channel, and the pure states rotation channel (see appendix A.5.1 for the definitions). For simplicity, in the latter the value of the θ parameter is chosen to be $\theta = \frac{5\pi}{12}$. In the first two case studies, Pauli channels were also used, these were the phase damping channel (A.6) and the depolarization channel (A.5).

As we discussed previously, for general non-Pauli channels, we have no analytical method to determine the optimal recovery operation. Therefore, hereafter we follow the numerical method discussed in section 3.2. In our

examples, as encoding we use the standard five-qubit code defined by the generators of Table A.1(a).

Note that in all cases, we studied systematically the dependence of our results on the noise strength of the two channels that were mixed. We found that the picture is similar in each case, i.e, there are no qualitative differences in the results while we modify the noise strength of the depolarizing, phase damping and amplitude damping channels between 0 and 0.7 and the parameter ϕ of the pure states rotation channel between 0 and $\frac{5\pi}{24}$. Thus in the following examples the noise strength dependence of the results is assumed negligible.

4.3.1 Pauli channel with non-Pauli perturbation

As a first step, we study the robustness of the optimal QEC operation $\mathcal{R}_0^\star$ of a nominal Pauli channel $\mathcal{E}_0$ against a non-Pauli perturbation $\mathcal{E}_1$. We have studied four different cases by mixing two types of Pauli channels with the above two types of non-Pauli ones in all possible combinations.

In Figure 4.4 we present typical results for all of the four cases. In the figures, the first and the second terms of the expression $\epsilon_{\mathcal{R}_0}(\mathcal{E}_\gamma)$ from (4.2) can be seen, that is, the optimal channel fidelity $F_{\text{ch}}(\mathcal{R}_\gamma^\star \circ \mathcal{E}_\gamma)$ (solid line) and the linear term $F_{\text{ch}}(\mathcal{R}_0 \circ \mathcal{E}_\gamma)$ (dotted-dashed line), respectively. It can be seen that the two curves are almost identical at a neighborhood of the initial Pauli channel. (Really, close to the Pauli channel, there is no difference in the two curves in the order of numerical error.) Nevertheless, we can not determine exactly the border of the zero-robustness domain and, in principle, γ_0 could be zero; but $\epsilon_{\mathcal{R}_0}(\mathcal{E}_\gamma)$ starts to increase very slowly with γ in higher than linear order. This region could be the generalization of the zero-robustness domain of the pure Pauli case. Increasing the distance from the original Pauli channel, we reach another domain where $\epsilon_{\mathcal{R}_0}(\mathcal{E}_\gamma)$ clearly has a linear term as a function of γ. Between these two domains, there is a transitional region which could be the generalization of the boundary surfaces of the zero-robustness domains. To a first approximation, we could summarize our observations based on the analysis of these fidelity curves:

Statement 4.7. *The longer the period on which the $\epsilon_{\mathcal{R}_0}(\mathcal{E}_\gamma)$ has only higher than linear order term, the more we can consider the corresponding recovery operation $\mathcal{R}_0^\star$ robust against the non-Pauli perturbation $\mathcal{E}_1$.*

Informally, general robustness domains would be those domains, in which the optimal recovery operation changes slowly with γ, and the borders between the regions would be those parts, where it changes faster. The necessary condition for the possibility of robust error correction would be that the channel is inside a robustness region.

To analyze in more detail the above observations and to clarify the differences and similarities with the pure Pauli case discussed in section 4.2, we study the change in the optimal recovery operation $\mathcal{R}_\gamma^\star$ with the mixing parameter γ. In the Pauli case, $\mathcal{R}_0^\star$ was completely unchanged in the zero-robustness domain, but it changed abruptly at the boundary of the domain. In the general

case, to measure how fast is the change in $\mathcal{R}^\star_\gamma$ we compute the $\left\|\frac{\mathrm{d}}{\mathrm{d}\gamma}X^\star_\gamma\right\|$ derivative of the Choi matrix $X^\star_\gamma$, where the norm is the Hilbert–Schmidt norm. We compute this derivative approximately; it is plotted by the dotted line in Figures 4.4(a)-(d).

We observe that $\mathcal{R}^\star_\gamma$ varies very slowly near the nominal Pauli channel $\mathcal{E}_0$, i.e., if γ is small. Even more interestingly, we can see that there are abrupt changes in $\mathcal{R}^\star_\gamma$ denoted by the peaks of the dotted curves. The better the numerical precision, the higher these peaks are, indicating discontinuous jumps in the $\mathcal{R}^\star_\gamma$ optimal recovery operation. These jumps correspond to breakpoints in the optimal channel fidelity curve. These breakpoints can not always be seen, because of the scaling of the figure, however, changing the scale and enlarging the vicinity of a point where $\mathcal{R}^\star_\gamma$ is discontinuous, the breakpoint can always be found. Between these non-continuous changes indicated by breakpoints, $\mathcal{R}^\star_\gamma$ is not constant but changes relatively slowly as a smooth function of γ. These observations indicate the following statement:

Statement 4.8. *Mixing a nominal Pauli-channel with a non-Pauli perturbation, the optimal recovery operation $\mathcal{R}^\star_\gamma$ changes in two different ways:*

- *In the non-analytic breakpoints – similarly to the domain boundary of the Pauli case given by **Statement 4.4** – the optimal $\mathcal{R}^\star_\gamma$ is not uniquely defined and there is no good choice out of the possible $\mathcal{R}^\star_\gamma$ operations against all possible perturbation $\mathcal{E}_1$. These breakpoints, however, form a discrete set in the $\gamma \in [0,1]$ interval and they correspond to a subset of zero measure in the set of all possible channels.*
- *Between these non-continuous changes the optimal QEC operation varies analytically, i.e., a small change in γ causes a change in the at least second order in $\epsilon_{\mathcal{R}_0}(\mathcal{E}_\gamma)$.*

In light of the above results, we can assert also the following:

Statement 4.9. *The optimal recovery operation $\mathcal{R}_0$ of a Pauli channel $\mathcal{E}_0$ is also robust against non-Pauli channel perturbations.*

4.3.2 Non-Pauli channel with Pauli perturbation

On the other hand, the curves of Figure 4.4 can also be viewed from the other side, i.e., from the $\gamma = 1$ endpoint. This point describes the amplitude damping or pure states rotation channels, respectively. Going backward from this point in γ, we can study the robustness of the recovery operation $\mathcal{R}^\star_1$ which is optimal for a nominal non-Pauli channel $\mathcal{E}_1$, against a Pauli-type perturbation $\mathcal{E}_0$. Thus the role of $\mathcal{E}_0$ and $\mathcal{E}_1$ is switched in this case study.

As can be seen from (4.3), the second term of the relative efficiency $\epsilon_{\mathcal{R}_1}(\mathcal{E}_\gamma)$ varies linearly in the function of γ for an arbitrary nominal channel $\mathcal{E}_1$ and an arbitrary perturbation $\mathcal{E}_0$. This is indicated by the dashed line in Figure 4.4. Similarly to the previous case study, we can see from the dotted lines that the derivative of the Choi matrix is not zero. This means that the $\mathcal{R}^\star_\gamma$ optimal

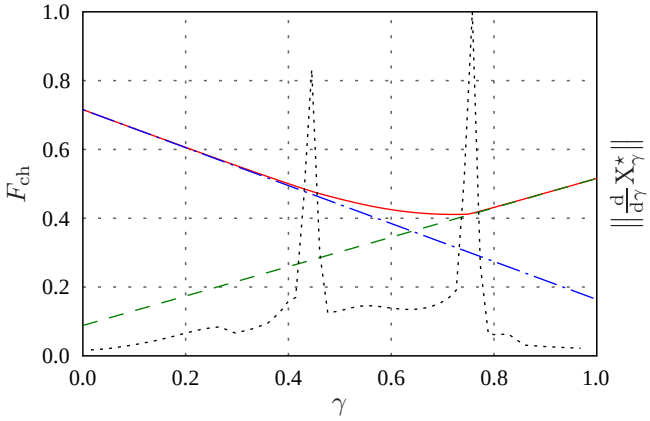


Figure 4.5. The two terms of $\epsilon_{\mathcal{R}_0}(\mathcal{E}_\gamma)$ in function of γ for the mixing of the amplitude damping channel ($g = 0.5$) and the pure states rotation channel ($\theta = \frac{5\pi}{12}, \phi = \frac{5\pi}{36}$), and the approximate derivative $\left\|\frac{\mathrm{d}}{\mathrm{d}\gamma}X^\star_\gamma\right\|$. Notations are the same as in Figure 4.4.

recovery operation changes by arbitrary small perturbation of the channel $\mathcal{E}_1$. As a consequence, $\gamma_0 = 0$, i.e., there is no finite zero-robustness domain around a non-Pauli channel. However, from the channel fidelity curves we can see that $\epsilon_{\mathcal{R}_1^\star}(\mathcal{E}_\gamma)$ has only higher than linear order term and γ_δ starts to increase only with lower powers of δ and not linearly. Therefore,

Statement 4.10. *The optimal recovery operation proved to be more robust in general than any other QEC operation, for which the γ_δ value of the δ-robustness domain starts to increase linearly with δ.*

Only when we are in a non-analytic point where the optimal recovery operation changes abruptly – at the peaks of $\left\|\frac{\mathrm{d}}{\mathrm{d}\gamma}X^\star_\gamma\right\|$ – is the case worse, similarly to the boundaries of zero-robustness domains in the case of Pauli channels.

We can see an interesting example in Figure 4.4(c). Very close to the amplitude damping channel $\mathcal{E}_1$, there is a very narrow peak in the derivative $\left\|\frac{\mathrm{d}}{\mathrm{d}\gamma}X^\star_\gamma\right\|$ indicating an abrupt change in the optimal QEC operation; therefore, the two terms of $\epsilon_{\mathcal{R}_1^\star}(\mathcal{E}_\gamma)$, i.e., the solid and dotted-dashed lines split at that point. Enlarging the vicinity of $\gamma = 1$ (see Figure 4.6) we observe that the solid and the dotted-dashed lines are almost identical when $0.9982 \lessapprox \gamma$ and the difference between the two terms grows linearly as γ decreases, i.e., as the distance from the initial amplitude damping channel $\mathcal{E}_1$ is increasing. The picture is very similar to the Pauli case; nevertheless, the robustness domain is very small.

4.3.3 Mixing of non-Pauli channels

To make sure that our statements are general, we also studied a case where two non-Pauli channels are mixed. For this purpose, we mix the amplitude damping ($\mathcal{E}_0$) and the pure states rotation channel ($\mathcal{E}_1$). As a typical example, we plot the $g = 0.5$ and $\phi = \frac{5\pi}{36}$ cases in Figure 4.5. Our observations are essentially the same as in the previously discussed case study. Actually, we can also see such non-Pauli–non-Pauli channel-mixing cases anywhere in the

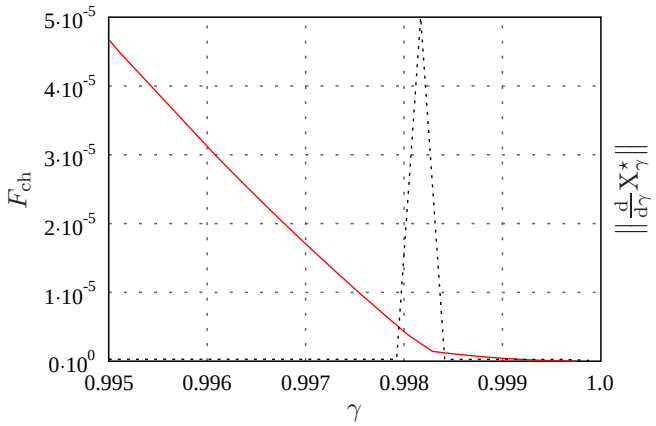


Figure 4.6. The enlarged version in Figure 4.4c in the small vicinity of $\gamma = 1$. The solid line shows the $\epsilon_{\mathcal{R}_1^\star}(\mathcal{E}_\gamma)$ function. The dotted line is again the normalized approximate derivative $\left\|\frac{\mathrm{d}}{\mathrm{d}\gamma}X_\gamma^\star\right\|$.

figures of Pauli–non-Pauli cases, when we take two channels corresponding to nonzero γ points and mix them.

4.4 Summary

In this chapter, the robustness of error correction operations based on stabilizer codes for the case of Pauli channels was investigated. It was found that the channels which have the same optimal QEC operation form a zero-robustness domain in the space of all Pauli channels. There are finite numbers of such domains and they form a partition of the whole set of Pauli channels represented as a simplex. If a channel is an interior point of such a domain then the optimal QEC operation is robust against any Pauli-type alteration of the channel. Only when we are on the boundary between two zero-robustness regions, there are more (at least two) completely different optimal QEC operations giving the same maximal channel fidelity. In this case, there is no such QEC operation, which is resistant against any type of channel perturbation, i.e., if we do not know something about the uncertainty of the channel then we can not protect against this uncertainty. However, this case is exceptional; most of the channels are interior points of a zero-robustness domain.

Thus, for Pauli channels the optimal recovery operation is better not only in entanglement fidelity, but also in robustness than any other recovery operation, assuming the channel is not on the boundary surface of a zero-robustness domain. If we have a more or less reliable channel model to describe the interactions between the system and its environment then by choosing the optimal recovery operation we can protect against the uncertainty of the channel model without any knowledge about it.

Moreover, the robustness of recovery operations was further investigated for the case of non-Pauli channels and perturbations. The obtained results strongly indicate – however, not prove – the following consequences. In contrast to the pure Pauli case, zero-robustness domains in which the optimal

recovery operation does not change at all, does not cover the whole space. Instead, the optimal recovery operation can change in two different ways. The first is similar to the case of boundaries of Pauli zero-robustness domains. In these special points the optimal recovery operation is not unique, it changes abruptly and the channel fidelity is non-analytic function of the mixing parameter. Between these points, it changes analytically with the mixing parameter like the channel fidelity, and therefore, the optimal QEC operation can said to be robust.

Part II

Quantum process tomography and experiment design

Chapter 5

Theory of quantum process tomography

This chapter gives an introduction to the theory of quantum process tomography and the related experiment design problem, building on the notions on quantum channels from section 2.2 and concepts from classical estimation theory in [1]. The presentation is mostly based on [27].

Section 5.1 introduces quantum tomography in general, while section 5.2 concentrates on the problem of process tomography. Finally, in section 5.3 the related experiment design problem is discussed.

5.1 Quantum tomography

Continuing the line of thought of the introduction, the differences between models and reality can have significant effect on the performance of the models. This means that accurate identification methods are necessary. The case is similar in the field of quantum information, where it can be of great importance to obtain precise estimates of quantum states or processes.

5.1.1 Quantum tomography as an identification problem

The problem of quantum estimation or *quantum tomography*[25] can be approached using the tools of classical identification theory (described e.g. in [1]). However, quantum systems are quite special stochastic nonlinear systems, where the stochasticity and nonlinearity are caused by the back-action of the measurements on the measured system (see section 2.1.3). Therefore, even in the simplest static case, when a non-dynamic quantum system is to be estimated, one needs special estimation methods [28]. Thus quantum tomography is a widely investigated problem in mathematical physics, with the goal to develop methods that can produce accurate estimators of unknown quantum states and processes.

[25]The general term "tomography" refers to the estimation techniques of infinite dimensional quantum systems, which are similar to the technique used in medical imaging.

5.1.2 Quantum state tomography

The simplest, but essential example of quantum tomography tasks is the identification of a quantum system, i.e., *quantum state tomography*. In fact, basically any other quantum tomography problem can be formulated indirectly as state tomography.

The disturbing nature of quantum measurement implies a fundamental requirement in state tomography, namely that sufficiently many identical copies of the state must be available. This way, we can perform a single measurement independently on each of the identical quantum systems, making the state after the measurement irrelevant in tomography tasks [7]. Thus in state tomography we are only interested in the measurement outcome probabilities, i.e., the statistical behaviour of the state with respect to the measurements. It follows that to have a unique state estimator, the set of used measurements must be *tomographically complete*. In other words, it must provide maximal information about the unknown quantum system.[26] For instance, in the most general case a set of observables which form an operator basis on the Hilbert space of the system is tomographically complete.

5.2 Quantum process tomography

Quantum process tomography deals with the estimation of quantum channels representing quantum physical processes. The paper [29] can be used as a general review and comparison of the available methods for process tomography. In its simplest version the problem can be formulated indirectly, i.e., by tracing it back to quantum state tomography.

There are mainly two methods for this. The first method (standard method) uses fixed d-dimensional *input states*, which are exposed to the effect of the channel, and performs state tomography on the resulting *output states*. The second method (ancilla-assisted method) is based on the fact that the Choi matrix of the channel can be interpreted also as a d^2-dimensional scaled density matrix, obtained by letting the channel act on one half of a suitable combined quantum system. This allows us to simply use state tomography for the channel estimation. Thus the ancilla-assisted method uses only one input state (principal system combined with ancilla system), however this must be d^2-dimensional, and preferably also maximally entangled to achieve best performance. Entanglement is harder to maintain, and if two-body interactions are not naturally available (e.g., photons), then the ancilla-assisted method can not be implemented efficiently. In contrast, the standard method does not use entanglement as a resource, and works with lower dimensional quantum systems and measurements, thus it is easier to apply in practice. Therefore, in this thesis, the standard method is used, however, the same questions could also be studied for the ancilla-assisted method. A formal presentation of the

[26] This approach of state tomography fits well to the interpretation of the quantum state as the set of outcome probability distributions of all possible measurements.

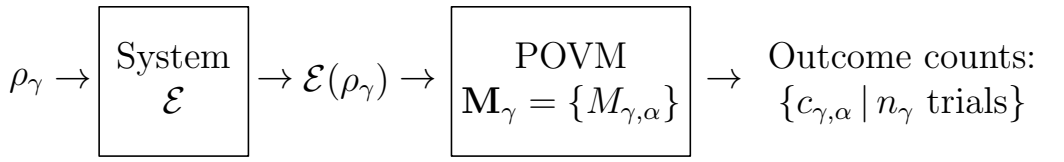


Figure 5.1. The scheme of data collection for channel tomography.

standard method is given in the following, in analogy with classical system identification.

5.2.1 Statistical model

The unknown quantum channel $\mathcal{E}\colon \mathcal{B}(\mathcal{H}) \to \mathcal{B}(\mathcal{H})$ taken from a set of all possible CPTP maps $\mathcal{M}_{\text{ch}}$ can be associated with a statistical model $\text{Tr}[\mathcal{E}(\,.\,)\,.\,]$ which captures the statistical behaviour of $\mathcal{E}$ with respect to so-called *experiment configurations* $(\rho, \mathbf{M})$ having the following elements:

- A density matrix ρ on the Hilbert space $\mathcal{H}$, used as probe (input) state of the channel $\mathcal{E}$.
- A POVM $\mathbf{M} = \{M_\alpha\}$ with which the measurements are performed on $\mathcal{E}(\rho)$ to obtain experimental data.

Thus the input of the model is a configuration $(\rho, \mathbf{M})$ and the output is the probability distribution $p(\alpha|\mathcal{E}) = \text{Tr}[\mathcal{E}(\rho)M_\alpha]$.

Note that – similarly to state tomography – the unique identification of $\mathcal{E}$ requires a tomographically complete set $\{(\rho_\gamma, \mathbf{M}_\gamma)\}$ of experiment configurations. This will be defined more exactly later in section 5.2.3.

5.2.2 Experimental data collection

The experiments[27] are performed n_γ times independently with each configuration $(\rho_\gamma, \mathbf{M}_\gamma)$ taken from a tomographically complete configuration set. The different α outcomes of the experiments are counted in the variable $c_{\gamma,\alpha}$ and put in the measurement record $\mathbf{D}$. This scheme can be seen in Figure 5.1. Then obviously $\sum_\alpha c_{\gamma,\alpha} = n_\gamma$, so we performed a total number of $n_{\text{tot}} = \sum_\gamma n_\gamma$ independent experiments. The estimate $\hat{\mathcal{E}}$ of the channel $\mathcal{E}$ will be calculated from these outcome counts, based on some estimation procedure.

5.2.3 Estimation procedure

The next step of the tomography problem is to choose a suitable estimation procedure. Two commonly used optimization based procedures will be presented in the following. Similarly to section 3.2, the optimization variable is choosen to be the Choi matrix $X_\mathcal{E}$ instead of a Kraus operator element set, because of its uniqueness. This allows us using (2.8), (2.10), (2.11), and (2.12)

[27] We assume here that the experiments are performed ideally, i.e., without measurement noise.

to write the statistical model of the channel with the more convenient notation $\mathrm{Tr}(\,.\,X_{\mathcal{E}})$, because

$$p_\gamma(\alpha|X_{\mathcal{E}}) = \sum_i \langle\!\langle M_{\gamma,\alpha} E_i | E_i \rho_\gamma \rangle\!\rangle = \sum_i \mathrm{Tr}\left[\rho_\gamma^{\mathrm{T}} \otimes M_{\gamma,\alpha} | E_i \rangle\!\rangle \langle\!\langle E_i | \right] = \mathrm{Tr}(C_{\gamma,\alpha} X_{\mathcal{E}}) \ , \tag{5.1}$$

where the configuration (block) matrix $C_{\gamma,\alpha} = \rho_\gamma^{\mathrm{T}} \otimes M_{\gamma,\alpha}$ depends on the channel input ρ and on the measured POVM elements in configuration γ. It is now apparent that in general a configuration set $\mathbf{C} = \{C_k\}$ is tomographically complete if and only if a linear span of operators C_k contains the whole set of possible Choi matrices $\mathcal{M}_{\mathrm{ch}}$ [30]. This can be achieved by using d^2 linearly independent input density matrices (a basis for the space of $d \times d$ matrices) and performing a tomographocally complete measurement on the channel output corresponding to each of them [29].

Maximum likelihood estimation

The independence of the experiments implies that the model probability of obtaining the data set $\mathbf{D}$ is $p(\mathbf{D}|\mathcal{E}) = \prod_{\gamma,\alpha} p_\gamma(\alpha|\mathcal{E})^{c_{\gamma,\alpha}}$, which is the likelihood function of $\mathcal{E}$. Taking its logarithm and substituting (5.1) we arrive at the optimization problem:

$$\begin{aligned} &\arg\max_{X_{\mathcal{E}}} \sum_{\gamma,\alpha} c_{\gamma,\alpha} \log\left(\mathrm{Tr}[C_{\gamma,\alpha} X_{\mathcal{E}}]\right) \\ &\text{so that} \quad X_{\mathcal{E}} \geq 0, \quad \mathrm{Tr}_2(X_{\mathcal{E}}) = \mathbb{1} \end{aligned} \tag{5.2}$$

The solution of this is the maximum likelihood (ML) estimate $\hat{\mathcal{E}}^{\mathrm{ML}}$ of $\mathcal{E}$. It is important that the objective function is concave and the constraints are convex in $X_{\mathcal{E}}$, hence, (5.2) is a convex optimization problem (see appendix A.2 for details).

Least squares estimation

The least squares (LS) is a popular method in classical estimation theory, because of its simplicity and good statistical properties. It is easy to implement, and it results in an estimation procedure which can be used in higher dimensions and for multiple parameters.[28]

We can arrive at the more convenient LS form of problem (5.2) too, if the number of experiments n_γ in configuration γ is sufficiently high. The relative frequency $\hat{p}_\gamma(\alpha|\mathcal{E}) = \frac{c_{\gamma,\alpha}}{n_\gamma}$ is an unbiased estimator of the model probability $p_\gamma(\alpha|\mathcal{E})$. If $n_\gamma \to \infty$ then $\mathrm{Var}\big(\hat{p}_\gamma(\alpha|\mathcal{E})\big) \to 0$, because $\hat{p}_\gamma(\alpha|\mathcal{E})$ has a binomial distribution. It follows that $\hat{p}_\gamma(\alpha|\mathcal{E}) \approx p_\gamma(\alpha|\mathcal{E})$ for large n_γ. Furthermore, assuming that $\mathcal{E}^{\mathrm{true}} \in \mathcal{M}_{\mathrm{ch}}$, i.e., $p_\gamma(\alpha|X_{\mathcal{E}}^{\mathrm{true}}) = \mathrm{Tr}(C_{\gamma,\alpha} X_{\mathcal{E}}^{\mathrm{true}})$, we arrive at the

[28] Another property of the (unconstrained) LS estimator is that if the system model is a linear function of the estimation variable then the estimator can be given analytically.

following least squares problem:

$$\arg\min_{X_\mathcal{E}} \sum_{\gamma,\alpha} \left[\hat{p}_\gamma(\alpha|X_\mathcal{E}) - \mathrm{Tr}(C_{\gamma,\alpha}X_\mathcal{E})\right]^2 , \quad \text{so that} \quad X_\mathcal{E} \geq 0, \quad \mathrm{Tr}_2(X_\mathcal{E}) = \mathbb{1} \tag{5.3}$$

The solution is thus the least squares (LS) estimate $\hat{\mathcal{E}}^{\mathrm{LS}}$ of $\mathcal{E}$. This problem is also a convex optimization problem in the Choi matrix $X_\mathcal{E}$, in particular a semidefinite programming problem, thus can be solved relatively easily (see in appendix A.2.3).

5.3 Experiment design

Recall from the introduction that the goal of experiment design is to search for experimental conditions that result in better or even optimal identification results [1]. The nature and properties of the system to be identified have a major influence on identification and experiment design. In general, the aim is to choose the experiment configuration (set of design variables) such that the data set of experimental results contains the most information with respect to the set of possible system models.

In connection with the estimation of a quantum channel $\mathcal{E}$, we would like to get the most information on the set of possible channels $\mathcal{M}_{\mathrm{ch}}$. This means – similarly to classical systems – that if the estimate $\hat{\mathcal{E}}$ converges in the number of experiments to $\tilde{\mathcal{E}}$, then $\tilde{\mathcal{E}}$ should be a unique approximation of the true channel $\mathcal{E}$, furthermore, if $\mathcal{E} \in \mathcal{M}_{\mathrm{ch}}$ then $\tilde{\mathcal{E}} = \mathcal{E}$ should hold.

Once the design variables are chosen such that the limiting estimate $\tilde{\mathcal{E}}$ is acceptable then it may become interesting to further select them so that the covariance matrix of $\tilde{\mathcal{E}}$ is minimized. This is the problem of optimal experiment design. As we have seen in section 5.2, the design variables of quantum process tomography were the input state ρ, the measurement POVM $\mathbf{M}$, the number of measurements per configuration n_γ, and the type of the estimator. The inconvenience that the optimal design depends on the estimator can be eliminated – assuming some parametrization of $\mathcal{E}$ – using the *Fisher information matrix*.

5.3.1 Fisher information

The Fisher information reflects the amount of information that a measured random variable can carry about the parameter $\mathbf{p}$ of interest. In other words, it measures the accuracy of the unbiased estimator $\hat{\mathbf{p}}$ of $\mathbf{p}$. Fisher information is a classical concept in statistics [31] and in system identification [1].

The Fisher information matrix in the quantum setting for a parametrized quantum state $\rho_\mathbf{p}$ measured using POVM $\mathbf{M}$ is by [32]

$$F(\mathbf{p}|\mathbf{M}) = \sum_\alpha \frac{1}{p(\alpha|\rho_\mathbf{p})} \nabla_\mathbf{p} p(\alpha|\rho_\mathbf{p}) \nabla_\mathbf{p}^{\mathrm{T}} p(\alpha|\rho_\mathbf{p}) , \tag{5.4}$$

where $p(\alpha|\rho_{\mathbf{p}}) = \mathrm{Tr}(\rho_{\mathbf{p}} M_\alpha)$ from section 2.1.3. Note that this definition requires $p(\alpha|\rho_{\mathbf{p}}) > 0$ for all α and $\mathbf{p}$.

The Fisher information is related to the covariance matrix through the so-called *Cramér–Rao matrix inequality*:

$$\mathrm{Var}_{\mathbf{p}}(\hat{\mathbf{p}}) \geq F(\mathbf{p}|\mathbf{M})^{-1} \tag{5.5}$$

This bound shows that the higher the Fisher information, the better estimation we can have. It is also apparent that in the quantum case, F depends on the actual measurement POVM $\mathbf{M}$ used in the experiments.

Based on the above, the case of a parameterized quantum channel $\mathcal{E}_{\mathbf{p}}$ can also be handled through the expression $\rho_{\mathbf{p}} = \mathcal{E}_{\mathbf{p}}(\rho)$. Then obviously $p(\alpha|\rho_{\mathbf{p}}) = \mathrm{Tr}(C_\alpha X_{\mathbf{p}})$ from (5.1), where $C_\alpha = \rho^{\mathrm{T}} \otimes M_\alpha$ is the configuration matrix. This implies that in this case F depends on the experiment configuration set $\mathbf{C} = \{C_\alpha\}$, which is $F(\mathbf{p}|\mathbf{C})$ in our notation.

Thus the inverse of the Fisher information gives us a global lower bound on the efficiency of any unbiased estimator[29] for a given $\mathbf{p}$. This indicates that the experiment design problem can also be an optimization problem; we have to choose the remaining three design variables – the input state ρ, the measurement POVM $\mathbf{M}$, and the number of measurements per configuration n_γ – such that they minimize the lower bound, or equivalently, maximize the Fisher information. Of course, the optimal design will depend on the value of $\mathbf{p}$ in general, which suggests that it may only be computable approximately in practice.

[29] There is a generalization for the case of biased estimators too.

Chapter 6

Parameter estimation for Pauli channels

In this chapter the identification of Pauli channels is formulated as a parameter estimation problem. By refining the general least squares based solution method from the convex optimization viewpoint, we arrive at statements regarding the optimization problem class of the parameter estimation of Pauli channels.

Section 6.1 proposes a method for convex optimization based channel parameter estimation. Section 6.2 applies this method to the case of Pauli channels, which is demonstrated using case studies in section 6.3. Lastly, section 6.4 summarizes the results.

6.1 Parameter estimation of quantum channels

It was discussed in section 1.1 that system models can have uncertainty in their structure and in their parameters in the general case. The same can be stated also for quantum channels. There are several attempts for the identification of a completely, i.e., structurally unknown channel. For instance, [33] considers channel estimation by randomizing the experiment configurations in a maximum-likelihood formulation. The goal of [30] is to extend the maximum entropy principle to the case of incomplete quantum channel estimation. The work [27] formulates the task of process tomography as a convex optimization problem.

In practice, however, it is reasonable to assume that we have a priori information about the structure of the channel, and only the unknown values of its parameters have to be estimated [34]. In other words, we can put constraints on the set $\mathcal{M}_{\text{ch}}$ of all channels. Some authors proposed approaches using such prior information on the channel, obtaining a well conditioned parameter estimation problem. This prior knowledge was mainly derived from physical interactions involved in the dynamics [35, 36].

As we have seen in section 5.2.3, the convex optimization based least squares method is very useful for structure estimation. However, in the case when the structure of the optimization variable (the Choi matrix of the chan-

nel) is known, i.e., only several specific parameters have to be estimated, then in general the LS method assuming unstructured variables can suffer from overparametrization.

A possible solution is to study the internal structure of the Choi matrix, and use this information to select more appropriate structure specific parameters as optimization variables, such that the constrained set $\mathcal{M}_{\text{ch}}$ remains convex, i.e., the LS method remains a convex optimization problem. Effectively, this should constrain the set of optimal solutions of problem (5.3) to solutions, which are consistent with the desired channel structure.

6.1.1 Affine approximation

The natural choice would be to select the unknown channel parameters as optimization variables. However, it can be easily seen that this choice could ruin convexity, as in the most general case the Choi matrix can depend on these parameters in a nonlinear way.

Thus, we propose the following decomposition of the Choi matrix structure. Let $H_0, H_1, \ldots, H_v$ denote constant Hermitian matrices such that we can expand the Choi matrix as an affine function for all value of $\mathbf{p}$:

$$X_{\mathcal{E}_\mathbf{p}} = \sum_{k=1}^{v} H_k h_k(\mathbf{p}) + H_0 \ , \tag{6.1}$$

where the $h_1(\mathbf{p}), \ldots, h_v(\mathbf{p})$ denote functions of the channel parameters $\mathbf{p}$. These functions $h_k(\mathbf{p})$ can now be used as the components of the new optimization variable $\mathbf{h}$, resulting in an approximation of the original problem.

Note that using this approximation may allow us to omit the trace preserving constraint from (5.3). Specific parameterized channel models are usually constructed to be trace preserving, and in such cases this property is independent of the parameter value $\mathbf{p}$. If trace preservation can be made independent of $\mathbf{h}$ too, then we can omit the trace preserving constraint. This holds, when the possible values of $\mathbf{h}$ are in one-to-one correspondence with the values of $\mathbf{p}$, i.e., when optimizing $\mathbf{h}$ instead of $\mathbf{p}$ is not an approximation.

6.1.2 Convex constraints

It may occur that some of the functions $h_k(\mathbf{p})$ depend on one or more other h_i $(i = 1, \ldots, v)$ functions, i.e, $h_k(\mathbf{p}) = (f \circ [h_1, \ldots, h_v])(\mathbf{p})$. If this dependence f is convex then we can use it to define additional constraints on the structure to make the estimation more accurate while still preserving convexity. We can determine a minimal set $\mathcal{P} = \{h_j(\mathbf{p})\}$, where $j \in I_\mathcal{P}$ and $|\mathcal{P}| = m$. The set $I_\mathcal{P}$ contains the indices j of those functions $h_j(\mathbf{p})$ among which there are no convex dependence. Putting these in an m dimensional vector $\mathbf{h}$, the remaining $v - m$ functions can be written as $h_k(\mathbf{p}) = f^{\text{cvx}}(\mathbf{h})$. Thus, if we have convex dependence between the new optimization variables $h_k(\mathbf{p})$, then the explicit statement of these dependences in the optimization problem can help to further

reduce the number of variables, which can be helpful in the subsequent steps of the estimation (see section 6.1.3).

However, the convex dependences can not be simply added to the optimization problem (5.3), because a strictly convex function with an equality statement is not a convex constraint. In these cases we can relax the equality, and add the convex constraint to the problem as inequality. If $v - m$ such independent constraints are found, then we obtain the following optimization problem (omitting the trace preserving constraint):

$$\begin{aligned} &\arg\min_{\mathbf{h}} V_{\mathrm{LS}}\big(X_{\mathcal{E}}(\mathbf{h})\big) \ , \\ &\text{so that} \quad X_{\mathcal{E}}(\mathbf{h}) \geq 0 \ , \\ \text{and} \quad &f_k^{\mathrm{cvx}}(\mathbf{h}) \leq h_k(\mathbf{p}), \quad k = 1, \ldots, v, \quad k \notin I_{\mathcal{P}}, \end{aligned} \tag{6.2}$$

where V_{LS} denotes the objective function of (5.3). After solving this problem we get an optimal value $v_0^\star = V_{\mathrm{LS}}(X_{\mathcal{E}}^\star)$ for the objective, where the * denotes the optimality. Then we solve the following auxiliary problem:

$$\begin{aligned} &\arg\min_{\mathbf{h}} \sum_{k=1,\ k\notin I_{\mathcal{P}}}^{v} h_k(\mathbf{p}) \ , \\ &\text{so that} \quad X_{\mathcal{E}}(\mathbf{h}) \geq 0 \ , \\ \text{and} \quad &f_k^{\mathrm{cvx}}(\mathbf{h}) \leq h_k(\mathbf{p}), \quad k = 1, \ldots, v, \quad k \notin I_{\mathcal{P}}, \\ &\text{and} \quad V_{\mathrm{LS}}\big(X_{\mathcal{E}}(\mathbf{h})\big) \leq v_0^\star \end{aligned} \tag{6.3}$$

By solving this, we make sure that the additional convex constraints in (6.2) get as close to equality as possible, thus making the parameter estimation procedure more accurate. In practice, the effectiveness of this heuristics is based only on the fact, that in case of the optimal solution (the model corresponding to the real channel), all of the convex constraints are saturated.

It is interesting to see whether this method always guarantees a global optimum of V_{LS}. If all dependences between the new optimization variables are convex, and thus can be taken into account, then the global optimum should be reachable, because the solution of the problem uniquely corresponds to a channel parameter value. It follows that the trace preserving constraint is also unnecessary in this case. If, however, some dependences remain which can not be taken into account as a convex constraint, then the estimation method will only find an approximate solution. A small example on convex constraints is presented in section 6.1.4.

6.1.3 Determining the model parameters

After we performed all the needed optimization steps and obtained an optimal Choi matrix $X_{\mathcal{E}}^\star$ together with the optimal variables $\mathbf{h}^\star$, the next step of parameter estimation is to obtain the original channel parameter vector $\mathbf{p}$ from the $\mathbf{h}^\star$ values. In essence, this problem can be written as

$$\arg\min_{\mathbf{p}} \sum_{l\in I_{\mathcal{P}}} \big(h_l^\star - h_l(\mathbf{p})\big)^2 \ . \tag{6.4}$$

This is a nonlinear least squares problem, which is not even convex in the general case, because of the nonlinearities in the $h_l(\mathbf{p})$ functions.

6.1.4 General example to parameter estimation

Here we present a simple but representative example on the affine approximation based parameter estimation of quantum channels. The example channel is the generalized amplitude damping described in section A.5.1. The Choi matrix of this channel is

$$X_{\mathcal{E}} = \begin{bmatrix} 1-g+pg & 0 & 0 & \sqrt{1-g} \\ 0 & g-pg & 0 & 0 \\ 0 & 0 & pg & 0 \\ \sqrt{1-g} & 0 & 0 & -pg+1 \end{bmatrix} .$$

Using (6.1), the above Choi matrix can be decomposed in the following way:

$$H_0 = \frac{1}{2}(\mathbb{1}_{4\times 4} + \sigma_{\mathrm{z}} \otimes \sigma_{\mathrm{z}}), \qquad H_1 = -\frac{1}{2}(\sigma_{\mathrm{z}} \otimes \sigma_{\mathrm{z}} + \mathbb{1}_{2\times 2} \otimes \sigma_{\mathrm{z}}),$$
$$H_2 = \mathbb{1}_{2\times 2} \otimes \sigma_{\mathrm{z}}, \qquad H_3 = \frac{1}{2}(\sigma_{\mathrm{x}} \otimes \sigma_{\mathrm{x}} - \sigma_{\mathrm{y}} \otimes \sigma_{\mathrm{y}}) ,$$

and the optimization variables will be

$$h_1(g,p) = g, \quad h_2(g,p) = pg, \quad h_3(g,p) = \sqrt{1-g} .$$

Notice that there is a convex dependence between h_3 and h_1, so we can further simplify the problem: $h_3^2 - 1 = -h_1$.

Using this, we arrive at the optimization problem

$$\arg\min_{h_2,h_3} V_{\mathrm{LS}}\big(X_{\mathcal{E}}(h_2,h_3)\big) ,$$
$$\text{so that} \quad X_{\mathcal{E}}(h_2,h_3) \geq 0, \quad \text{and} \quad h_3^2 - 1 \leq -h_1$$

Note that the TP constraint is not necessary here, because the above Choi matrix is TP by construction, and any value of the new optimization variables together with the convex constraint uniquely corresponds to the value of the channel parameters. The type (6.3) auxiliary problem with $v_0^\star = V_{\mathrm{LS}}(X_{\mathcal{E}}^\star)$ will be:

$$\arg\min_{h_2,h_3} -h_1 ,$$
$$\text{so that} \quad X_{\mathcal{E}}(h_2,h_3) \geq 0,\ h_3^2 - 1 \leq -h_1 \text{ and } V_{\mathrm{LS}}\big(X_{\mathcal{E}}(h_2,h_3)\big) \leq v_0^\star$$

The final step is the parameter extraction using (6.4). In this case it is clearly nonconvex, but the simplicity of this two dimensional example allows us to analytically derive the solution of (6.4). It will be $g = h_1^\star$ and $p = \frac{h_2^\star}{h_1^\star}$.

6.2 Estimation of Pauli channels

The parameter estimation problem (6.2) becomes really simple in the case of Pauli channels.

6.2.1 Qubit Pauli channel

In this case, knowing the channel structure effectively means the knowledge of the channel directions $|\mathbf{v}_i\rangle$. This knowledge allows us to solve the estimation problem for channel directions $|\mathbf{e}_i\rangle$ without loss of generality, because we can always align our reference frame with the known vectors $|\mathbf{v}_i\rangle$ (see section 2.3.3).

We will now show that in this case, matrices H_k $(k = 0, \ldots, 3)$ can be found such that $h_k(\lambda) = \lambda_k$ $(k = 1, \ldots, 3)$, i.e., the Choi matrix of the qubit Pauli channel is affine in the original channel parameters. By relating (2.12) with (6.1) using (2.14) and (2.19), we get the following Hermitian matrices:

$$H_k = \sum_{i=0}^{3} q_{k,i}^{-1} |\sigma_i\rangle\rangle\langle\langle\sigma_i|,$$

where the number $q_{k,i}^{-1}$ is the $(k+1, i+1)^{\text{th}}$ element of $\mathbf{Q}^{-1}$, the inverse of the coefficient matrix $\mathbf{Q}$ from (2.19).

We see that in this case there are no convex dependences, so the parameter extraction step (6.4) is unnecessary, and the trace preserving constraint can be omitted from the resulting convex problem, which is

$$\begin{aligned} &\arg\min_{\lambda} V_{\text{LS}}\big(X_{\mathcal{E}}(\lambda)\big) \ , \\ &\text{so that} \quad X_{\mathcal{E}}(\lambda) \geq 0 \ . \end{aligned} \tag{6.5}$$

To summarize, we can formulate the statement:

Statement 6.1. *The least squares based parameter estimation of any two dimensional Pauli channel is a convex problem of the form* (6.5)*, as the optimization variables are exactly the channel parameters to be estimated.*

6.2.2 Pauli channels for prime-level systems

The qubit case can be generalized to higher level Pauli channel cases, where the level d of the system is prime. In such cases, we can obtain $d+1$ cyclic groups of order d from the d-level Pauli matrices, each group spanning a complementary subalgebra. The notations X and Z for this case follow appendix A.4. Let $U_\mu = XZ^{\mu-1}$ $(\mu = 1, \ldots, d)$ and $U_{d+1} = Z$. Then each U_μ generates a cyclic group, and j indexes the group elements U_μ^j. Thus the channel structure is assumed again to be known.

The Kraus representation of a d-level Pauli channel is then by [37]:

$$\mathcal{E}(\rho) = a_0\rho + \frac{1}{d-1}\sum_{\mu=1}^{d+1} a_\mu \sum_{j=1}^{d-1} U_\mu^j \rho U_\mu^{-j}, \quad a_\mu \geq 0, \quad \sum_\mu a_\mu = 1 \ . \tag{6.6}$$

The $\mathbb{1}$ operators from each group are treated separately, because these have a different parameter a_0.

Using now that the channel acts by definition as $\mathcal{E}(U_i) = \lambda_i U_i$, we get

$$\begin{aligned}\lambda_i U_i &= a_0 U_i + \frac{1}{d-1}\sum_{\mu=1}^{d+1} a_\mu \sum_{j=1}^{d-1} U_\mu^j U_i U_\mu^{-j} \\ &= a_0 U_i + \frac{1}{d-1}\sum_{\mu=1}^{d+1} a_\mu \sum_{j=1}^{d-1} \omega^{j[\mu(1-\delta_{i,d+1})-i(1-\delta_{\mu,d+1})]} U_i \ ,\end{aligned}$$

where $\omega = \mathrm{e}^{\mathrm{i}\frac{2\pi}{d}}$. Noticing that

$$\sum_{j=1}^{d-1} \omega^{j[\mu(1-\delta_{i,d+1})-i(1-\delta_{\mu,d+1})]} = \begin{cases} d-1 & \mu = i \\ -1 & \mu \neq i \end{cases} ,$$

we arrive at the correspondence between λ and $\mathbf{a} = [a_0, \ldots, a_{d+1}]^{\mathrm{T}}$ parameter vectors:

$$\begin{bmatrix} 1 \\ \lambda_1 \\ \vdots \\ \lambda_{d+1} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & \cdots & \cdots & 1 \\ 1 & 1 & \frac{-1}{d-1} & \cdots & \cdots & \frac{-1}{d-1} \\ 1 & \frac{-1}{d-1} & 1 & \ddots & & \vdots \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \vdots & & \ddots & 1 & \frac{-1}{d-1} \\ 1 & \frac{-1}{d-1} & \cdots & \cdots & \frac{-1}{d-1} & 1 \end{bmatrix} \begin{bmatrix} a_0 \\ \vdots \\ a_{d+1} \end{bmatrix} , \tag{6.7}$$

where the coefficient matrix will be denoted by $\mathbf{Q}$, similarly to the qubit case.

Now the matrices H_k $(k = 0, \ldots, d+1)$ can be given such that $h_k(\lambda) = \lambda_k$ $(k = 1, \ldots, d+1)$. By relating (2.12) with (6.1) using (6.6) and (6.7), we get the following Hermitian matrices:

$$H_k = \frac{1}{d-1}\sum_{\mu=0}^{d+1} q_{k,\mu}^{-1} \sum_{j=1}^{d-1} |U_\mu^j\rangle\!\rangle\langle\!\langle U_\mu^j|, \tag{6.8}$$

where $U_0 = \mathbb{1}$ and the number $q_{k,\mu}^{-1}$ is the $(k+1, \mu+1)^{\text{th}}$ element of $\mathbf{Q}^{-1}$ from (6.7).

Similarly to the qubit case, in the prime-level case there are no convex dependences, so the parameter extraction step (6.4) is unnecessary. The trace preserving constraint can also be omitted from the resulting convex problem, which is of the same form as (6.5).

To summarize, we can formulate the statement:

Statement 6.2. *The least squares based formulation of the Pauli channel parameter estimation problem is a convex optimization problem, and has the form of* (6.5) *in any prime dimension.*

Note that in principle, this result could be further generalized to the case when d is prime power using tensor products of Pauli operators (see appendix A.4).

Example with $d = 3$

For the sake of simplicity we choose the simplest 3-level (qutrit) case, i.e., when $d = 3$ in section 2.3.2 to present an example. The 3-level unitary Pauli matrices and the identity form the following 4 cyclic groups:

$$\begin{aligned}\mathcal{U}_1 &= \{\mathbb{1}, X, X^2\},\\ \mathcal{U}_2 &= \{\mathbb{1}, XZ, \omega X^2 Z^2\},\\ \mathcal{U}_3 &= \{\mathbb{1}, XZ^2, \omega^2 X^2 Z\},\\ \mathcal{U}_4 &= \{\mathbb{1}, Z, Z^2\},\end{aligned}$$

where $\omega = \mathrm{e}^{\mathrm{i}\frac{2\pi}{3}}$, and the complementary subalgebra $\mathcal{A}_\mu$ of the channel is obtained as the linear span of the group $\mathcal{U}_\mu$. The Choi matrix of this channel will be the following:

$$X_{\mathcal{E}} = \frac{1}{3}\begin{bmatrix} f_1 & 0 & 0 & 0 & f_3 & 0 & 0 & 0 & f_3 \\ 0 & f_2 & 0 & 0 & 0 & f_4 & f_4^* & 0 & 0 \\ 0 & 0 & f_2 & f_4^* & 0 & 0 & 0 & f_4 & 0 \\ 0 & 0 & f_4 & f_2 & 0 & 0 & 0 & f_4^* & 0 \\ f_3 & 0 & 0 & 0 & f_1 & 0 & 0 & 0 & f_3 \\ 0 & f_4^* & 0 & 0 & 0 & f_2 & f_4 & 0 & 0 \\ 0 & f_4 & 0 & 0 & 0 & f_4^* & f_2 & 0 & 0 \\ 0 & 0 & f_4^* & f_4 & 0 & 0 & 0 & f_2 & 0 \\ f_3 & 0 & 0 & 0 & f_3 & 0 & 0 & 0 & f_1 \end{bmatrix},$$

where

$$\begin{aligned} f_1 = 1 + 2\lambda_4, \quad f_2 = 1 - \lambda_4, \quad f_3 = \lambda_1 + \lambda_2 + \lambda_3,\\ f_4 = \lambda_1 - \lambda_2 \mathrm{e}^{\mathrm{i}\frac{\pi}{3}} - \lambda_3 \mathrm{e}^{-\mathrm{i}\frac{\pi}{3}}. \end{aligned}$$

6.3 Case studies

The aim of the following simulation experiments was to provide examples on the proposed parameter estimation method for Pauli channels, and compare them to results obtained using unstructured Choi matrix, i.e., the general process tomography method discussed in Chapter 5. The results were generated with the simulation tools discussed in section A.6.

6.3.1 Tomography settings

A *single experiment in the configuration* γ consisted of acting with the unknown channel $\mathcal{E}$ on the prepared state ρ_γ, then measuring $\mathcal{E}(\rho_\gamma)$ using the POVM $\mathbf{M}_\gamma$. Each configuration was set up as follows.

- The used input states were all pure states, because this way the outputs provide the most information. Recall from Chapter 5 that for the estimation of d-level channels, at least d^2 linearly independent input density

matrices are necessary. The corresponding pure states are commonly selected to be the elements of the orthonormal basis $\{|n\rangle\}_{n=0}^{d-1}$ and the set of states $\{\frac{|n\rangle+|m\rangle}{\sqrt{2}}, \frac{|n\rangle+\mathrm{i}|m\rangle}{\sqrt{2}}\}$ with $m, n = 0 \dots d-1$ and $m > n$.
- Appropriate POVMs were selected to obtain a tomographically complete experiment configuration sets.

We say that a *complete experiment* was performed if a single experiment was done in all configurations.

As the total number of available pure quantum states were assumed to be n_{tot}, a full estimation procedure consisted of total number of n_{tot} measurements. These were distributed among all n_{cfg} configurations equally. Thus, $n_{\mathrm{exp}} = \frac{n_{\mathrm{tot}}}{n_{\mathrm{cfg}}}$ single experiments were done in each configuration, i.e., n_{exp} complete experiments were done. Each estimation procedure was repeated 10 times, and the resulting Choi matrices and channel parameter vectors in the k^{th} procedure were $\hat{X}_{\mathcal{E}}^{(k)}$ and $\hat{\lambda}^{(k)}$. Using these, the following three estimation performance measuring quantities were calculated:

- Using the empirical mean $\bar{\lambda} = \frac{1}{10}\sum_{k=1}^{10} \hat{\lambda}^{(k)}$ of the estimated parameter vectors $\hat{\lambda}^{(k)}$ from each run, the distance

$$\left\|\bar{\lambda} - \lambda\right\|_1 = \sum_i |\bar{\lambda}_i - \lambda_i|$$

  of $\bar{\lambda}$ and the real parameter vector λ was used as performance indicator.
- The empirical variance of the estimated parameter vectors $\hat{\lambda}^{(k)}$ from each run. The full covariance matrix is not useful here, because the computing of the off-diagonal elements corresponding to correlated parameters would require joint measurements of possibly incompatible observables. On the other hand, if we can select observables such that the parameters can be estimated independently, then the resulting covariance matrix is diagonal. Thus, only these diagonal elements were used, i.e., the average variance of the parameters was computed using the unbiased variance estimator:

$$\mathrm{Avg}\big(\mathrm{Var}(\hat{\lambda})\big) = \frac{1}{d+1}\sum_i \frac{1}{9}\sum_{k=1}^{10}(\hat{\lambda}_i^{(k)} - \bar{\lambda}_i)^2$$

- Average Hilbert–Schmidt norm of the estimation error:

$$\mathrm{Avg}(\|\hat{X}_{\mathcal{E}} - X_{\mathcal{E}}\|) = \frac{1}{10}\sum_{k=1}^{10} \|\hat{X}_{\mathcal{E}}^{(k)} - X_{\mathcal{E}}\|$$

6.3.2 Experiment configurations

Here we describe the experiment configurations used in the simulation examples. These configurations were selected to be completely general in the sense that they are not adjusted to the structure of the example channels.

Qubit channels

For qubit channels, $n_{\text{tot}} = 18000$ was used. The experiment configurations were made up from the following elements:

- To obtain a tomographically complete configuration set, minimum 4 independent input densities are necessary. Based on section 6.3.1, the inputs were the states with Bloch vectors $\pm\frac{1}{\sqrt{3}}[1,1,1]^{\mathrm{T}}$, $\frac{1}{\sqrt{2}}[-1,1,0]^{\mathrm{T}}$, and $\frac{1}{\sqrt{6}}[1,1,-2]^{\mathrm{T}}$.
- Each channel output state must be measured using a tomographically complete set of POVMs. The following two cases were tested:
  (a) Minimal POVM described by [38]: This is a tomographically complete extremal POVM. It allows us to do channel estimation using only 4 experiment configurations, however it can not be used to estimate each λ_i parameter independently, which results in a nondiagonal empirical covariance matrix. The POVM elements are given by $M_j = \frac{1}{4}(\mathbb{1} + \mathbf{m}_j \cdot \vec{\sigma})$, where $\mathbf{m}_1 = \frac{1}{\sqrt{3}}[1,1,1]^{\mathrm{T}}$, $\mathbf{m}_2 = \frac{1}{\sqrt{3}}[1,-1,-1]^{\mathrm{T}}$, $\mathbf{m}_3 = \frac{1}{\sqrt{3}}[-1,1,-1]^{\mathrm{T}}$, and $\mathbf{m}_4 = \frac{1}{\sqrt{3}}[-1,-1,1]^{\mathrm{T}}$.
  (b) Standard qubit tomography observable set: This consist of three complementary observables of the general form $\mathbf{v}_i \cdot \vec{\sigma}$, where the vectors $\mathbf{v}_i$ are the columns of a rotation matrix $\mathbf{V}$. These form a tomographically complete set of observables for the case of two dimensional states. Each can be used as a two-element POVM in a different configuration, resulting in a total number of 12 experiment configurations. Such set of observables allow the independent estimation of each Pauli channel parameter λ_i only if they correspond to measurements aligned with the channel directions, which is not true in this general case. $\mathbf{V}$ can be obtained from the unit vector $\mathbf{u}$ (axis of rotation) and the angle of rotation θ as $\mathbf{V} = \mathrm{e}^{[\mathbf{u}]_\times \theta}$, where $[\mathbf{u}]_\times$ is the cross product matrix

$$\begin{bmatrix} 0 & -u_3 & u_2 \\ u_3 & 0 & -u_1 \\ -u_2 & u_1 & 0 \end{bmatrix} .$$

  To have a general setting, $\mathbf{u} = \frac{1}{\sqrt{14}}[1,2,3]^{\mathrm{T}}$ and $\theta = \frac{\pi}{4}$ were used in the experiments.

3-level channels

For qutrit channels, $n_{\text{tot}} = 108000$ was used. The experiment configurations were made up from the following elements:

- To obtain a tomographically complete configuration set, minimum 9 independent input densities are necessary. Recall from appendix A.4 that together with $\mathbb{1}$, the 3-level unitary Pauli matrices form an orthonormal

basis in the space of 3×3 complex matrices. Then, using the notations of section 6.2.2, any qutrit density matrix ρ can be expanded as

$$\rho = \frac{1}{3}\left(\mathbb{1} + \sum_{\mu=1}^{4}\sum_{j=1}^{2} b_{\mu,j} U_\mu^j\right) .$$

In particular, a pure state in each complementary subalgebra $\mathcal{A}_\mu$ of the Pauli channel is $\rho_\mu = \frac{1}{3}\left(\mathbb{1} + \sum_{j=1}^{2} U_\mu^j\right)$. Using these, a sufficiently general set of input states can be obtained in a way similar to the 2-level case. The eigenvectors of the convex combination $\frac{1}{4}\sum_\mu \rho_\mu$ give the basis

$$|0\rangle = \begin{bmatrix} \sqrt{\frac{2}{5-\sqrt{5}}} \\ \sqrt{\frac{2}{5+\sqrt{5}}} \\ 0 \end{bmatrix}, |1\rangle = \begin{bmatrix} -\sqrt{\frac{2}{5+\sqrt{5}}} \\ \sqrt{\frac{2}{5-\sqrt{5}}} \\ 0 \end{bmatrix}, |2\rangle = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} ,$$

which can be used to define the rest of the necessary 9 input states.

- The eight Gell-Mann matrices were selected as a tomographically complete set of observables to measure the channel outputs:

$$\Sigma_1 = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \Sigma_2 = \begin{bmatrix} 0 & -\mathrm{i} & 0 \\ \mathrm{i} & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \Sigma_3 = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}, \Sigma_4 = \begin{bmatrix} 0 & 0 & -\mathrm{i} \\ 0 & 0 & 0 \\ \mathrm{i} & 0 & 0 \end{bmatrix},$$
$$\Sigma_5 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}, \Sigma_6 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -\mathrm{i} \\ 0 & \mathrm{i} & 0 \end{bmatrix}, \Sigma_7 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \Sigma_8 = \frac{1}{\sqrt{3}} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{bmatrix}$$

These are the Hermitian generalization of the Pauli matrices for 3-level systems. Each can be used as a three-element POVM in a different configuration, resulting in a total number of 72 experiment configurations. This set of observables can not be used to estimate all channel parameters independently, thus it results in a nondiagonal empirical covariance matrix.

6.3.3 Qubit Pauli channel examples

The estimation of the qubit Pauli channel defined in (2.16) was simulated using the configurations in section 6.3.2. The parameter vector was selected to be $\lambda = (-0.4, -0.6, 0.2)$. This was taken from the interior of the parameter space.

The performance indicator quantities of the estimations were calculated in function of the number of complete experiments n_{exp} both for the case when the Choi matrix is unstructured, and when the structure is taken into account in the LS estimation problem. The results are plotted in Figure 6.1 for the case of the minimal POVM, and in Figure 6.2 for the case of the standard POVM set.

More simulation examples with other channel parameter values taken both from the interior and the border of the parameter space can be seen in appendix B.1.1.

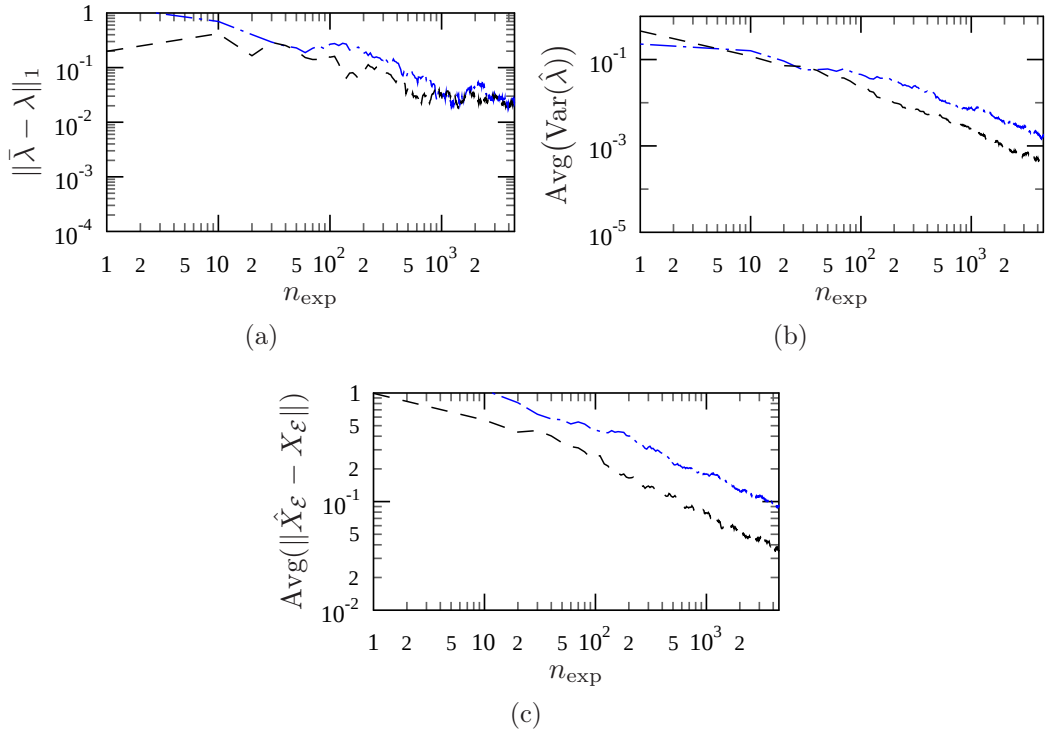


Figure 6.1. Minimal tomography with $\lambda = (-0.4, -0.6, 0.2)$. The three Figures show the performance indicators defined in section 6.3.1 respectively, in function of the number of complete experiments. The dotted-dashed line corresponds to tomography with unstructured Choi matrix, and the dashed line shows results from parameter estimation using Choi matrix structure.

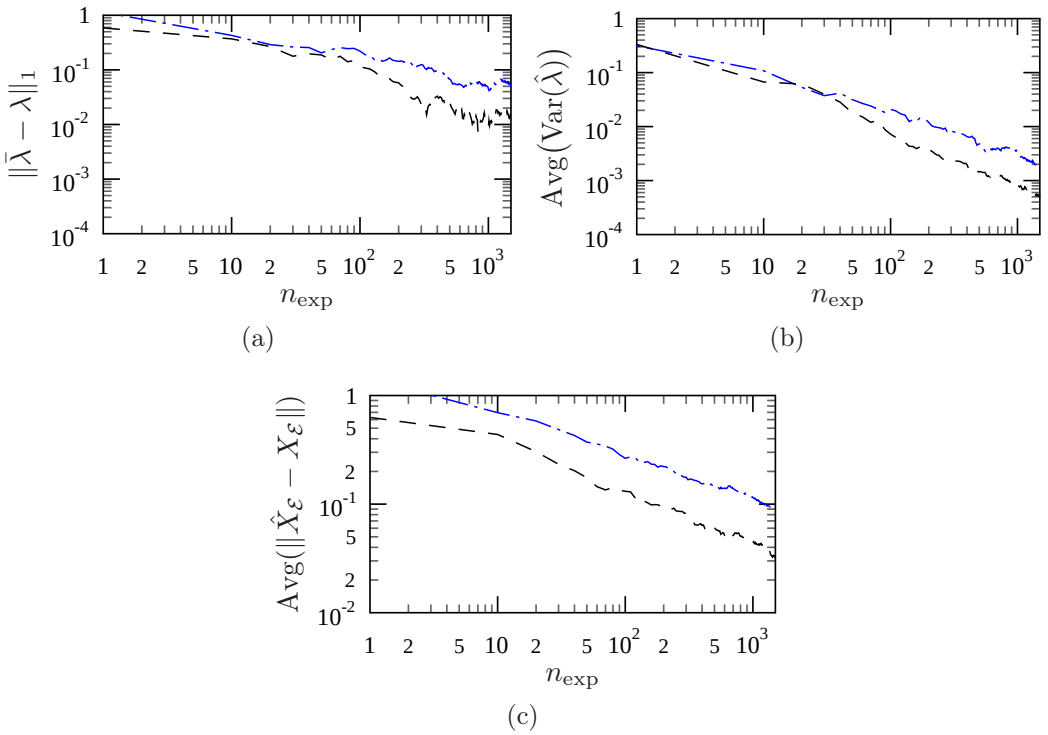


Figure 6.2. Standard tomography with $\lambda = (-0.4, -0.6, 0.2)$. The three Figures show the performance indicators defined in section 6.3.1 respectively, in function of the number of complete experiments. The dotted-dashed line corresponds to tomography with unstructured Choi matrix, and the dashed line shows results from parameter estimation using Choi matrix structure.

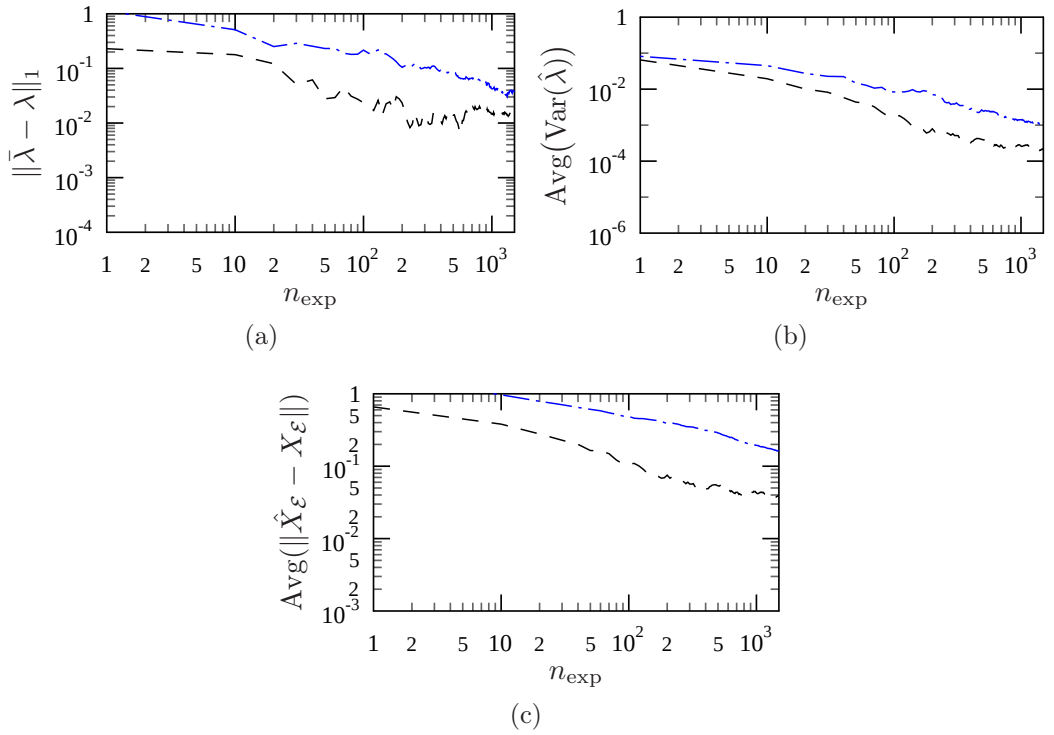


Figure 6.3. 3-level tomography with $\lambda = (0.4, 0.475, 0.325, 0.55)$. The three Figures show the performance indicators defined in section 6.3.1 respectively, in function of the number of complete experiments. The dotted-dashed line corresponds to tomography with unstructured Choi matrix, and the dashed line shows results from parameter estimation using Choi matrix structure.

It can be seen on the Figures that for qubit Pauli channels, using the same fixed amount of resources, taking the channel structure into account results in better estimation performance.

6.3.4 3-level Pauli channel example

The estimation of the qutrit Pauli channel defined in section 6.2.2 was simulated using the configurations in section 6.3.2. The parameter vector was selected to be $\lambda = (0.4, 0.475, 0.325, 0.55)$. This was taken from the interior of the parameter space.

The performance indicator quantities of the estimations were calculated in function of the number of complete experiments n_{exp} both for the case when the Choi matrix is unstructured, and when the structure is taken into account in the LS estimation problem. The results are plotted in Figure 6.3.

More simulation examples with other channel parameter values taken both from the interior and the border of the parameter space can be seen in appendix B.1.2.

We can see that even in the case of a Pauli channel with a higher number of parameters, the proposed parameter estimation method (6.5) outperforms general channel tomography using the same fixed amount of resources.

6.4 Summary

In this chapter a parameter estimation method based on convex optimization was proposed for Pauli channels. It was found that after the affine decomposition of the Choi matrix, the problem of least squares based Pauli channel tomography turns into a convex parameter estimation problem, which is solvable in any prime dimension. Moreover, using simulation case studies it has been shown, that for a fixed amount of resources, taking the known channel structure into account using the proposed method of affine decomposition of the Choi matrix can significantly increase the accuracy of the estimation, compared to the case when no information on the channel structure is used.

In addition, with performing auxiliary optimization problems, convex relations between optimization variables can also be exploited to further improve the estimation. In the case of non-Pauli channels, however, beside the convex part the method may need a nonconvex optimization step as well.

Chapter 7

Experiment design for Pauli channels with known structure

This chapter formulates the experiment design problem corresponding to parameter estimation discussed in Chapter 6 and derives the optimal experiment configuration for Pauli channels in the case of known channel structure using a convex maximization approach.

Section 7.1 formulates the problem of experiment design as an optimization problem. Section 7.2 derives the optimal experiment configuration for the case of Pauli channels with known structure. The performance of the optimal design is demonstrated on examples in section 7.3. Finally, section 7.4 summarizes the results.

7.1 Problem statement of experiment design

The field of experiment design for quantum channel identification has not matured yet. Even the main problems have not been formalized completely. A recent paper of [35] gives a good overview of the state of the art in this field. Only a few papers exist that aim at determining elements of the experiment configuration, i.e., the input quantum system and the measurement POVM (see e.g. [39] and [40]). These papers, however, fix one of the elements – the input quantum system, for example – and determine the other (say the POVM) according to some optimality criteria. For example, in [41] the problem of optimal input design with fixed measurements is examined. The only papers that use a convex optimization approach to experiment design solve only a restricted problem; the determination of the number of measurements to be performed in the different experiment configurations [27, 36]. The same goal is set in [39], where the authors seek to optimize experiment design for general one parameter quantum channels using analytical methods.

Our aim is to consider the full experiment configuration, i.e., the input state and measurement POVM together as design variable, and solve the problem of experiment design for quantum channel parameter estimation in an optimization context.

Formally, suppose we have a quantum channel $\mathcal{E}_{\mathbf{p}}$ with some fixed channel

parameter vector $\mathbf{p}$. We would like to find the optimal experiment configuration, i.e., that input state ρ and a measurement POVM $\mathbf{M}$ for which the channel output $\mathcal{E}_{\mathbf{p}}(\rho)$ measured by the POVM $\mathbf{M}$ gives the most information on the channel parameters. Optimality in terms of information is understood here in the sense that we seek the pair $(\rho, \mathbf{M})$ for which a suitable scalar function of the Fisher information matrix $F(\mathbf{p}|\rho, \mathbf{M})$ of the channel parameters $\mathbf{p}$ is maximal.

It is important to note that the optimal configuration alone does not necessarily constitute an optimal, tomographically complete experiment setup. Instead, it only tells us what properties the configurations of an effective setup should have. Note also that – as mentioned in section 5.3.1 – the optimal design $(\rho^\star, \mathbf{M}^\star)$ can in general depend on the value of $\mathbf{p}$.

7.1.1 The optimization problem

A mathematical representation of the configuration $(\rho, \mathbf{M})$ is obtained through (5.1); the probability $p(\alpha|\mathbf{p})$ of the measurement outcome α can be written as $\mathrm{Tr}(C_\alpha X_{\mathbf{p}})$, where $C_\alpha = \rho^{\mathrm{T}} \otimes M_\alpha$ is the configuration matrix. Using this and $\mathbf{C} = \{C_\alpha\}$, the Fisher information matrix (5.4) will be

$$F(\mathbf{p}|\mathbf{C}) = \sum_\alpha \frac{1}{\mathrm{Tr}(C_\alpha X_{\mathbf{p}})} \nabla_{\mathbf{p}} \mathrm{Tr}\left(C_\alpha X_{\mathbf{p}}\right) \nabla_{\mathbf{p}}^{\mathrm{T}} \mathrm{Tr}\left(C_\alpha X_{\mathbf{p}}\right) \tag{7.1}$$

assuming $\mathrm{Tr}(C_\alpha X_{\mathbf{p}}) > 0$.

To be able to do maximization we need to obtain a scalar value from the matrix $F(\mathbf{p}|\mathbf{C})$. Out of many possible choices, the trace was selected for this purpose because of its mathematical simplicity, and the property $\mathrm{Tr}(A) \leq \mathrm{Tr}(B)$ whenever $A \leq B$ for Hermitian matrices A and B. Thus, the objective function will be

$$\tilde{F}(\mathbf{p}|\mathbf{C}) = \sum_{i,\alpha} \frac{\mathrm{Tr}(C_\alpha[\partial_{p_i} X_{\mathbf{p}}])^2}{\mathrm{Tr}(C_\alpha X_{\mathbf{p}})} \tag{7.2}$$

It can be shown that the function $\tilde{F}$ is convex in the set of configuration matrices $\{C_\alpha\}$. This can be proven using the composition rule in **Theorem A.4** (see appendix A.2). Suppose that $\mathbb{S}^d$ is the set of $d \times d$ Hermitian matrices, then

- $h\colon \mathbb{R}^2 \to \mathbb{R}$, $h(\mathbf{x}) = \frac{x_1^2}{x_2}$ is known to be convex if $x_2 > 0$ (see [42]),
- $\mathbf{g}_i\colon \mathbb{S}^d \to \mathbb{R}^2$, $\mathbf{g}_i(C_\alpha) = \left[\mathrm{Tr}(C_\alpha[\partial_{p_i} X_{\mathbf{p}}]), \mathrm{Tr}(C_\alpha X_{\mathbf{p}})\right]^{\mathrm{T}}$ is affine,

thus

$$(h \circ \mathbf{g}_i)(C_\alpha) = \frac{\mathrm{Tr}(C_\alpha[\partial_{p_i} X_{\mathbf{p}}])^2}{\mathrm{Tr}(C_\alpha X_{\mathbf{p}})}$$

is convex, if the probability $\mathrm{Tr}(C_\alpha X_{\mathbf{p}}) > 0$, which is assumed. We know that the sum of convex functions is also convex, which means $\tilde{F}(\mathbf{p}|\mathbf{C})$ is convex. Note that then $\tilde{F}$ is also convex both in the input ρ and in the used measurement POVM $\mathbf{M}$ if we fix the other to be a constant.

It follows that $\tilde{F}$ takes its maximum at an extremal point of the feasible region containing the possible experiment configurations $\{C_\alpha\}$. The set of extremal points of the quantum state space is the set of pure states, i.e., the set of rank one projections. Similarly, the extremal points of the set of POVMs are the extremal POVMs. These imply the following statement:

Statement 7.1. *In terms of the objective* (7.2)*, the optimal input state* $\rho^\star$ *will be pure, and the optimal measurement (POVM)* $\boldsymbol{M}^\star$ *will be an extremal POVM in the optimal experiment configuration.*

7.2 Experiment design for Pauli channels

In this section the optimal experiment configuration is found for the subclass of Pauli channels. The channel structure, i.e., the channel directions in the qubit case and the set of pairwise complementary subalgebras in the general case are assumed to be known.

7.2.1 Optimal configuration for qubit Pauli channels

In the special case of qubit Pauli channels $\mathcal{E}$ with known channel directions, **Statement 7.1** can be made stronger. Because of the rotational symmetry of the Bloch ball, the channel directions can be assumed to be $\{|\mathbf{e}_i\rangle\}$ and the obtained results apply to any other channel direction setting.

For simplicity, a further assumption is that the used measurements are projective, i.e., can be represented with two-element extremal POVMs $\{|\psi\rangle\langle\psi|, I - |\psi\rangle\langle\psi|\}$ [10]. Recall from section 2.1.3 that these POVM elements have the corresponding Bloch vectors $\mathbf{m}$ and $-\mathbf{m}$. Furthermore, the pure input state is also identified with the Bloch vector $\mathbf{b}$.

Based on section 2.3.3 the channel output with channel affine map $\Lambda = \mathrm{diag}([\lambda_1, \lambda_2, \lambda_3]^{\mathrm{T}})$ will be $\mathcal{E}(\mathbf{b}) = \Lambda\mathbf{b} = \sum_{i=1}^{3} \lambda_i b_i \mathbf{e}_i$, and the Fisher information matrix takes the form

$$F(\lambda|\mathbf{b}, \mathbf{m}) = \frac{\nabla_\lambda(\mathbf{m}^{\mathrm{T}}\Lambda\mathbf{b})\nabla_\lambda^{\mathrm{T}}(\mathbf{m}^{\mathrm{T}}\Lambda\mathbf{b})}{1 - (\mathbf{m}^{\mathrm{T}}\Lambda\mathbf{b})^2} \ .$$

Taking the trace, we arrive at

$$\tilde{F}(\lambda|\mathbf{b}, \mathbf{m}) = \sum_i \frac{(\mathbf{m}^{\mathrm{T}}[\partial_{\lambda_i}\Lambda]\mathbf{b})^2}{1 - (\mathbf{m}^{\mathrm{T}}\Lambda\mathbf{b})^2} = \frac{m_1^2 b_1^2 + m_2^2 b_2^2 + m_3^2 b_3^2}{1 - (m_1 b_1 \lambda_1 + m_2 b_2 \lambda_2 + m_3 b_3 \lambda_3)^2} \ . \quad (7.3)$$

Note that in the qubit Pauli channel case (apart from $\|\lambda\|_\infty = 1$), the requirement $\mathrm{Tr}(C_\alpha X_\lambda) > 0$ always holds, because the channel output states $\mathcal{E}_\lambda(\rho)$ are always mixed, which means that $\mathrm{Tr}(\mathcal{E}_\lambda(\rho)M_\alpha) = 0$ if and only if $M_\alpha = \mathbf{0}$.

The task is then to determine the maximal value of $\tilde{F}(\lambda|\mathbf{b}, \mathbf{m})$ by choosing $\mathbf{m}$ and $\mathbf{b}$. Recall that the unit length requirement on the optimal vectors $\mathbf{b}^\star$ and $\mathbf{m}^\star$ follows from the convexity of $\tilde{F}$ which we want to maximize. Note also that the above formula is a special case of (7.2) using projective measurements.

Let us now define the vector $\mathbf{c} = [m_1 b_1, m_2 b_2, m_3 b_3]^{\mathrm{T}}$, which can be seen as the configuration vector of the channel estimation problem, including both the input state and measurement information together with the assumptions on the channel structure. The objective (7.3) will then be

$$\tilde{F}(\lambda|\mathbf{c}) = \frac{\mathbf{c}^{\mathrm{T}}\mathbf{c}}{1 - (\mathbf{c}^{\mathrm{T}}\lambda)^2} \ .$$

It is easy to see that the set of all possible $\mathbf{c}$ vectors is inside an octahedron in the Bloch sphere, whose vertices are the unit vectors pointing to the three axes $\mathbf{e}_i$ corresponding to the directions $|\mathbf{e}_i\rangle$ of the channel. This can be proven using the Cauchy–Schwarz–Bunyakovsky (CSB) inequality:

$$\|\mathbf{c}\|_1 = \sum_{i=1}^{3} |m_i b_i| \leq \|\mathbf{m}\|_2 \|\mathbf{b}\|_2 = 1 \tag{7.4}$$

To be precise, we could try proving that the possible $\mathbf{c}$ vectors fill the whole octahedron. However, as we will exploit the convexity of the octahedron, we would like to know only whether any of its extremal points correspond to possible $\mathbf{c}$ vectors. More than this can be shown; any $\mathbf{c}$ with $\|\mathbf{c}\|_1 = 1$ can be obtained as an elementwise product $c_i = b_i m_i$, where $\|\mathbf{b}\|_2 = \|\mathbf{m}\|_2 = 1$. In general, we have equality in (7.4) if and only if $|b_i|^2 = |m_i|^2$. Thus we can choose for example $b_i = \sqrt{|c_i|}$ and $m_i = \mathrm{sgn}(c_i)\sqrt{|c_i|}$.

Now, if $\tilde{F}$ is convex in the $\mathbf{c}$ variable then we can characterize the optimal $\mathbf{b}$ and $\mathbf{m}$ vectors. Convexity is proven using **Theorem A.3** from appendix A.2. Suppose $f(\mathbf{c}) = \mathbf{c}^{\mathrm{T}}\mathbf{c}$ and $g(\mathbf{c}) = \frac{1}{1-(\mathbf{c}^{\mathrm{T}}\lambda)^2}$. Then f is convex and nonnegative on $\mathbb{R}^3$, while g is convex and nonnegative on the domain where $|\mathbf{c}^{\mathrm{T}}\lambda| \leq 1$ holds. The octahedral set of $\mathbf{c}$ vectors is inside this domain, i.e., $\|\mathbf{c}\|_1 \leq 1 \Rightarrow |\mathbf{c}^{\mathrm{T}}\lambda| \leq 1$, because $\|\lambda\|_\infty \leq 1$ and Hölder's inequality implies

$$|\mathbf{c}^{\mathrm{T}}\lambda| \leq \sum_i |c_i\lambda_i| \leq \|\mathbf{c}\|_1 \|\lambda\|_\infty \leq 1 \ ,$$

with equality possible only if $|\lambda_i| = 1$ for all i where $c_i \neq 0$, if such λ corresponds to a CPTP map. Thus the convexity of $\tilde{F}(\lambda|\mathbf{c})$ in $\mathbf{c}$ can be verified by checking

$$\nabla f(\mathbf{c}) \nabla^{\mathrm{T}} g(\mathbf{c}) = \frac{2\mathbf{c}^{\mathrm{T}}\lambda}{\left[1 - (\mathbf{c}^{\mathrm{T}}\lambda)^2\right]^2} \mathbf{c}\lambda^{\mathrm{T}} \geq 0 \ .$$

This in turn holds because its only nonzero eigenvalue is $\frac{2(\mathbf{c}^{\mathrm{T}}\lambda)^2}{[1-(\mathbf{c}^{\mathrm{T}}\lambda)^2]^2} \geq 0$.

It is now easy to see that $\tilde{F}$ takes its maximum at a vertex of the octahedral feasible set. It follows that $\|\mathbf{c}^\star\|_1 = \|\mathbf{c}^\star\|_2 = 1$, i.e., $\sum_i |c_i^\star|^2 = \left(\sum_i |c_i^\star|\right)^2$, which is the same as $\sum_{i<j} |c_i^\star c_j^\star| = 0$. This can only happen if only one component of $\mathbf{c}^\star$ is nonzero, which means that both the input $\mathbf{b}^\star$ and the measurement $\mathbf{m}^\star$ has to be in the same Bloch axis corresponding to a channel direction. This implies that the objective $\tilde{F}$ is maximized clearly if the axis of $\mathbf{c}^\star$ is that axis,

for which the corresponding $|\lambda_i|$ is maximal. Let this be for example λ_j, then the optimal objective will be

$$\tilde{F}(\lambda|\mathbf{c}^\star) = \frac{1}{1-\lambda_j^2} \ .$$

This result on the parallelity of the optimal vectors $\mathbf{b}^\star$ and $\mathbf{m}^\star$ motivates the following definition for qubit Pauli channels:

Definition 7.1. *An experimental configuration of the form* $(|\psi\rangle\langle\psi|, \{|\psi\rangle\langle\psi|, \mathbb{1}-|\psi\rangle\langle\psi|\})$ *– or* $(\boldsymbol{b}, \{\pm\boldsymbol{b}\})$ *for qubits if* $\boldsymbol{b}$ *is the Bloch vector of* $|\psi\rangle$ *– will be called parallel configuration.*

We see that the optimum is independent of the parameter vector λ in this case. However, the Bloch axes of the channel directions are orthogonal, thus performing experiments in one axis does not give any information on the other channel parameters. To have a tomographically complete setting, we have to search for additional experimental configurations. Let the axis of the optimal configuration found first be the axis of $\mathbf{e}_1$ with the corresponding channel parameter λ_1. If we now constrain the objective (7.3) to the plane orthogonal to $\mathbf{e}_1$, then we get the constraints $m_1 = 0$ and $b_1 = 0$. Using the same derivation as in the general three dimensional case we get that the next optimal configuration will be the $\mathbf{e}_2$ or $\mathbf{e}_3$ axis, and so on. To summarize, in the case of the qubit Pauli channel, a tomographically complete set consisting of the above optimal experiment configurations gives an optimal way of estimating each channel parameter independently. It follows, that searching for an optimal experiment configuration that gives information simultaneously on all parameters could lead to different results. Therefore, we can state the following:

Statement 7.2. *In terms of the objective* (7.3)*, the optimal experiment configurations for the qubit Pauli channel with known channel directions* $\{|\boldsymbol{v}_i\rangle\}$ *are the parallel configurations* $(\boldsymbol{v}_i, \{\pm\boldsymbol{v}_i\})$*,* $i = 1, 2, 3$*. These make up an optimal set of configurations for estimating each channel parameter independently.*

7.2.2 Generalization to higher level Pauli channels

The trace of the Fisher information for d-level Pauli channels

More generally, the objective function (7.2) can be derived also for the case of d-level Pauli channels, where d is prime. Using the notations of section 6.2.2, a d-dimensional operator D can be expanded as

$$D = \frac{1}{d}\left(b_0\mathbb{1} + \mathbf{b}\cdot\vec{U}\right) = \frac{1}{d}\left(b_0\mathbb{1} + \sum_{\mu=1}^{d+1}\sum_{j=1}^{d-1} b_{\mu,j}U_\mu^j\right) \ , \tag{7.5}$$

where $b_0 = \mathrm{Tr}(D)$, the coefficients $b_{\mu,j} = \mathrm{Tr}(U_\mu^{j*}D)$ make up the generalized Bloch vector $\mathbf{b}$, and $\vec{U}$ is the formal vector of the Pauli operators U_μ^j. The numbers $b_{\mu,j}$ may be complex even if D is Hermitian, because the operators

U_μ^j are not Hermitian for $d > 2$. However, for a Hermitian D, $b^*_{\mu,j} = b_{\mu,d-j}$. Furthermore, if D is a density matrix, then $b_0 = 1$. Then it is known that $|b_{\mu,j}| \leq 1$, and $\|\mathbf{b}\|_2 \leq \sqrt{d-1}$ with equality in the latter if and only if D is pure (see [37]). It follows that every pure state has at least $d-1$ non-zero coefficients. Note that for $d > 2$ these conditions are necessary but not sufficient for (7.5) to define a positive operator.

Now, to derive (7.2), we first express $\mathrm{Tr}(C_\alpha[\partial_{\lambda_k} X_\lambda]) = \mathrm{Tr}(C_\alpha H_k)$ ($k = 0, \ldots, d+1$) using (6.8), (7.5), (2.11), and that $C_\alpha = \rho^{\mathrm{T}} \otimes M_\alpha$, where M_α can be a positive operator in general:

$$\begin{aligned}
&\mathrm{Tr}(C_\alpha H_k) = \\
&= \mathrm{Tr}\left(\frac{1}{d}\big(\mathbb{1} + \mathbf{b}_\alpha \cdot \vec{U}\big)^{\mathrm{T}} \otimes \frac{1}{d}\big(m_{\alpha,0}\mathbb{1} + \mathbf{m}_\alpha \cdot \vec{U}\big) \frac{1}{d-1} \sum_{\mu=0}^{d+1} q^{-1}_{k,\mu} \sum_{j=1}^{d-1} |U_\mu^j\rangle\!\rangle\langle\!\langle U_\mu^j| \right) \\
&= \frac{1}{d^2(d-1)} \sum_{\mu=0}^{d+1} q^{-1}_{k,\mu} \sum_{j=1}^{d-1} \Big(m_{\alpha,0} d + \langle\!\langle U_\mu^j | (\mathbf{m}_\alpha \cdot \vec{U}) U_\mu^j (\mathbf{b}_\alpha \cdot \vec{U}) \rangle\!\rangle \Big) \\
&= \frac{1}{d^2(d-1)} \sum_{\mu=0}^{d+1} q^{-1}_{k,\mu} \sum_{j=1}^{d-1} \left(m_{\alpha,0} d + \sum_{\nu,\nu'=1}^{d+1} \sum_{x,x'=1}^{d-1} m_{\alpha,\nu,x} b_{\alpha,\nu',x'} \mathrm{Tr}\big(U_\mu^{j*} U_\nu^x U_\mu^j U_{\nu'}^{x'}\big) \right)
\end{aligned}$$

Let $\omega = \mathrm{e}^{\mathrm{i}\frac{2\pi}{d}}$. Then by the commutation relations,

$$\mathrm{Tr}\big(U_\mu^{j*} U_\nu^x U_\mu^j U_{\nu'}^{x'}\big) = d\omega^{jx[\nu(1-\delta_{\mu,d+1}-\delta_{\mu,0})-\mu(1-\delta_{\nu,d+1})]} \delta_{\nu',\nu} \delta_{x',d-x} \ .$$

Continuing the previous derivation, we get

$$\begin{aligned}
&\mathrm{Tr}(C_\alpha H_k) = \\
&= \frac{1}{d^2-d} \sum_{\mu=0}^{d+1} q^{-1}_{k,\mu} \sum_{j=1}^{d-1} \left(m_{\alpha,0} + \sum_{\nu=1}^{d+1} \sum_{x=1}^{d-1} m_{\alpha,\nu,x} b^*_{\alpha,\nu,x} \omega^{jx[\nu(1-\delta_{\mu,d+1}-\delta_{\mu,0})-\mu(1-\delta_{\nu,d+1})]} \right) \\
&= \frac{1}{d} m_{\alpha,0} \underbrace{\sum_{\mu=0}^{d+1} q^{-1}_{k,\mu}}_{\delta_{k,0}} + \frac{1}{d} \sum_{\mu=0}^{d+1} q^{-1}_{k,\mu} \sum_{\nu=1}^{d+1} \underbrace{\frac{d(\delta_{\mu,\nu}+\delta_{\mu,0})-1}{d-1}}_{q_{\mu,\nu}} \sum_{x=1}^{d-1} m_{\alpha,\nu,x} b^*_{\alpha,\nu,x} \ ,
\end{aligned}$$

where $q_{\mu,\nu}$ is the $(\mu+1, \nu+1)^{\mathrm{th}}$ element of the coefficient matrix $\mathbf{Q}$ from (6.7), and we used that $\sum_{j=1}^{d-1} \omega^{jx[\nu(1-\delta_{\mu,d+1}-\delta_{\mu,0})-\mu(1-\delta_{\nu,d+1})]} = d(\delta_{\mu,\nu}+\delta_{\mu,0}) - 1$. Simplifying the formula, we arrive at

$$\mathrm{Tr}(C_\alpha H_k) = \frac{1}{d} \mathbf{q}_k^{-\mathrm{T}} \mathbf{Q} \begin{bmatrix} m_{\alpha,0} \\ \mathbf{c}_\alpha \end{bmatrix} = \frac{1}{d}\big(\delta_{k,0} m_{\alpha,0} + (1-\delta_{k,0}) c_{\alpha,k}\big) \ ,$$

where $\mathbf{q}_k^{-\mathrm{T}}$ is the k^{th} row of $\mathbf{Q}^{-1}$, and $\mathbf{c}_\alpha$ is a $d+1$ dimensional vector with $c_{\alpha,\nu} = \sum_{x=1}^{d-1} m_{\alpha,\nu,x} b^*_{\alpha,\nu,x}$. Note that the properties of Bloch vectors imply that $\mathbf{c}_\alpha$ is a real vector.

We can now write the objective function (7.2) in a simpler form using (6.1):

$$\tilde{F}(\lambda|\{\mathbf{c}_\alpha\}) = \frac{1}{d} \sum_{\alpha,k} \frac{c^2_{\alpha,k}}{\mathbf{c}_\alpha^{\mathrm{T}} \lambda + m_{\alpha,0}} = \frac{1}{d} \sum_{\alpha} \frac{\mathbf{c}_\alpha^{\mathrm{T}} \mathbf{c}_\alpha}{\mathbf{c}_\alpha^{\mathrm{T}} \lambda + m_{\alpha,0}} \tag{7.6}$$

Optimal configuration for d-level Pauli channels

Numerical maximization of (7.2) was done in the $d = 3$ case, starting from multiple random input states and 3-element (extremal) projective measurements. The results suggested that the optimal input state and one element of the extremal POVM are the same, and are inside one of the complementary subalgebras $\mathcal{A}_i$ of the Pauli channel, while the remaining two POVM elements seemed to be arbitrary.

Based on this, in the following we will derive the optimal configuration for d-level Pauli channels (when d is prime) for the simple case of two-element extremal POVMs. Similarly to section 7.2.1, the pure input state and the extremal POVM element $|\psi\rangle\langle\psi|$ will be represented with Bloch vectors $\mathbf{b}$ and $\mathbf{m}$, while the POVM element $\mathbb{1} - |\psi\rangle\langle\psi|$ will be represented with Bloch vector $-\mathbf{m}$ together with $m_0 = d - 1$.

The objective (7.6) can then be written as

$$\tilde{F}(\lambda|\mathbf{c}) = \frac{1}{d}\left(\frac{\mathbf{c}^{\mathrm{T}}\mathbf{c}}{\mathbf{c}^{\mathrm{T}}\lambda + 1} + \frac{\mathbf{c}^{\mathrm{T}}\mathbf{c}}{-\mathbf{c}^{\mathrm{T}}\lambda + d - 1}\right) = \frac{\mathbf{c}^{\mathrm{T}}\mathbf{c}}{(d-2)\mathbf{c}^{\mathrm{T}}\lambda - (\mathbf{c}^{\mathrm{T}}\lambda)^2 + d - 1} \,. \tag{7.7}$$

To maximize this function, we first need to characterize the set of $\mathbf{c}$ vectors and prove the convexity of $\tilde{F}$. The set of $\mathbf{c}$ vectors is inside a so-called hyperoctahedron in $\mathbb{R}^{d+1}$. This can be seen from the inequalities

$$\|\mathbf{c}\|_1 = \sum_{\mu=1}^{d+1}\left|\sum_{j=1}^{d-1} m_{\mu,j} b^*_{\mu,j}\right| \leq \sum_{\mu=1}^{d+1}\sum_{j=1}^{d-1} |m_{\mu,j} b^*_{\mu,j}| \leq \|\mathbf{m}\|_2 \|\mathbf{b}\|_2 = d - 1 \,. \tag{7.8}$$

Assume that $d > 2$. Then in the first inequality (triangle inequality) we have equality if and only if all $m_{\mu,j}b^*_{\mu,j}$ have the same phase. Then, because $\mathbf{c}$ is real, all $m_{\mu,j}b^*_{\mu,j}$ must be real with the same sign, too. The second inequality (CSB) will be equality if and only if $|b_{\mu,j}|^2 = |m_{\mu,j}|^2$, which implies $m_{\mu,j}b^*_{\mu,j} = |b_{\mu,j}|^2 \mathrm{e}^{\mathrm{i}\varphi_j}$ for arbitrary φ_j. Finally, the last equality is implied by the purity of $\mathbf{b}$ and $\mathbf{m}$. Thus, if both inequalities are saturated, then $m_{\mu,j}b^*_{\mu,j} = \pm|b_{\mu,j}|^2$ with the same sign for all j. It follows that either $\mathbf{b} = \mathbf{m}$ or $\mathbf{b} = -\mathbf{m}$. However, it is easy to check that if $\mathbf{m}$ is a pure Bloch vector then $-\mathbf{m}$ does not represent a positive operator, so only the $\mathbf{b} = \mathbf{m}$ case is allowed. From this we can conclude that each $\mathbf{c}$ with $\|\mathbf{c}\|_1 = d - 1$ can be related to a pure Bloch vector, and no $\mathbf{c}$ vectors with negative component exist such that $\|\mathbf{c}\|_1 = d - 1$. For this reason, in the following, we will be interested in the positive orthant of the hyperoctahedron, which thus contains all extremal configuration vectors $\mathbf{c}$ with $\|\mathbf{c}\|_1 = d - 1$. The $d = 2$ case is special; then the triangle inequality is always saturated, thus the constraints on the $\mathbf{c}$ vectors with unit 1-norm are different, e.g. negative $\mathbf{c}$ vectors are allowed, too (see section 7.2.1 for details).

We see that not all points of the hyperoctahedron are valid $\mathbf{c}$ configuration vectors, but in its positive orthant, the vector having $\|\mathbf{c}\|_1 = d - 1$ in each coordinate axis ν is valid, i.e., can be related to a pure Bloch vector $\mathbf{b}_\nu$. Moreover, we can select $\mathbf{b}_\nu$ to represent the pure state ρ_ν from the corresponding complementary subalgebra $\mathcal{A}_\nu$. There are d pure states $\{\rho_\nu^{(k)}\}_{k=0}^{d-1}$ in $\mathcal{A}_\nu$. The

nonzero components $b^{(k)}_{\nu,j}$ of their Bloch vectors $\mathbf{b}^{(k)}_\nu$ is derived in the following. We diagonalize the U^j_μ in (7.5) for the case of pure density matrices in $\mathcal{A}_\nu$: $\rho^{(k)}_\nu = \frac{1}{d} V_\nu \left(\mathbb{1} + \sum_{j=1}^{d-1} b^{(k)}_{\nu,j} Z^j \right) V^*_\nu$, where V_ν is unitary and Z was defined in section A.4. Then $\rho^{(k)}_\nu$ will be a pure state for all k if and only if the matrix B_ν whose $(j,k)^{\text{th}}$ element is $b^{(k)}_{\nu,j}$ satisfies

$$\left[\text{diag}^{-1}(Z), \ldots, \text{diag}^{-1}(Z^{d-1})\right] \cdot B_\nu = \begin{bmatrix} d-1 & -1 & -1 & \cdots & -1 \\ -1 & d-1 & -1 & \cdots & -1 \\ -1 & -1 & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & -1 \\ -1 & -1 & \cdots & -1 & d-1 \end{bmatrix} .$$

The solution is $B_\nu = \left[\text{diag}^{-1}(Z), \ldots, \text{diag}^{-1}(Z^{d-1})\right]^*$, thus the vectors $\mathbf{b}^{(k)}_\nu$ are determined.

Convexity of $\tilde{F}$ can be proven in almost exactly the same way as in the qubit case, using **Theorem A.3** from appendix A.2. Let $f(\mathbf{c}) = \mathbf{c}^{\mathrm{T}}\mathbf{c}$ and $g(\mathbf{c}) = \frac{1}{(d-2)\mathbf{c}^{\mathrm{T}}\lambda - (\mathbf{c}^{\mathrm{T}}\lambda)^2 + d - 1}$. Then f is convex and nonnegative on $\mathbb{R}^{d+1}$, while g is convex and nonnegative on the domain where $-1 \leq \mathbf{c}^{\mathrm{T}}\lambda \leq d-1$ holds. The positive orthant of the hyperoctahedron is inside this domain, because $\frac{-1}{d-1} \leq \lambda_i \leq 1$ $(i = 1, \ldots, d+1)$ implies for any positive $\mathbf{c}$ with $\|\mathbf{c}\|_1 \leq d-1$:

$$-1 \leq \frac{-1}{d-1} \|\mathbf{c}\|_1 = \frac{-1}{d-1} \sum_i c_i \overset{\lambda_i = \frac{-1}{d-1}}{\leq} \mathbf{c}^{\mathrm{T}}\lambda \overset{\lambda_i = 1}{\leq} \sum_i c_i = \|\mathbf{c}\|_1 \leq d-1 ,$$

with equality possible only in the case of $\lambda_i = 1$ for all i where $c_i \neq 0$, or $\lambda_i = \frac{-1}{d-1}$ for all i where $c_i \neq 0$, if such λ corresponds to a CPTP map. Therefore, the convexity of $\tilde{F}(\lambda|\mathbf{c})$ in $\mathbf{c}$ can be verified by checking

$$\nabla f(\mathbf{c}) \nabla^{\mathrm{T}} g(\mathbf{c}) = \frac{2\mathbf{c}^{\mathrm{T}}\lambda}{\left[(d-2)\mathbf{c}^{\mathrm{T}}\lambda - (\mathbf{c}^{\mathrm{T}}\lambda)^2 + d - 1\right]^2} \mathbf{c}\lambda^{\mathrm{T}} \geq 0 ,$$

which holds because its only nonzero eigenvalue is $\frac{2(\mathbf{c}^{\mathrm{T}}\lambda)^2}{\left[(d-2)\mathbf{c}^{\mathrm{T}}\lambda - (\mathbf{c}^{\mathrm{T}}\lambda)^2 + d - 1\right]^2} \geq 0$.

Now we know that $\tilde{F}$ takes its maximum at a vertex of the positive orthant of the hyperoctahedron. Recall that the positive vertex on coordinate axis ν corresponds to the pure states in subalgebra $\mathcal{A}_\nu$. This means that the optimal experiment configuration $\mathbf{c}^\star$ is made up using a pure state $\mathbf{b}_\nu$ both as input and POVM element, taken from that subalgebra $\mathcal{A}_\nu$ for which $|\lambda_\nu|$ is maximal. Then the optimal objective will be

$$\tilde{F}(\lambda|\mathbf{c}^\star) = \frac{d-1}{(d-2)\lambda_\nu - (d-1)\lambda_\nu^2 + 1} .$$

Note that the optimum is independent of the parameter vector λ also in this general case. However, as the optimal configuration only gives information on one channel parameter, we have to search for additional experimental

configurations to have a tomographically complete experiment setting. If the optimal configuration $\mathbf{c}^\star$ is nonzero at the ν^{th} coordinate, then a second optimal configuration can be found by optimizing with the constraint $c_\nu = 0$. Then the feasible region will become the positive orthant of a one less dimensional hyperoctahedron, which implies that the second optimal configuration will also be a positive vertex. Continuing this procedure, we arrive at a tomographically complete set of optimal configurations, which gives an optimal way of estimating each channel parameter independently. Moreover, this method implies that the empirical covariance matrix of the estimated parameters will be diagonal. Of course, searching for an optimal experiment configuration that gives information simultaneously on all parameters could lead to different results, as in the qubit case.

A case study on the optimal d-level Pauli channel estimation can be seen for the qutrit case in subsection 7.3.2.

7.2.3 Analytical solution of parameter estimation

For Pauli channels, we can express the outcome probability (5.1) with respect to the rank-1 POVM element $|\psi_\gamma\rangle\langle\psi_\gamma|$ in the configuration γ using the $\mathbf{c}_\gamma$ configuration vector as $p_\gamma = \frac{1+\mathbf{c}_\gamma^{\mathrm{T}}\lambda}{d}$. Then the probability with respect to $\mathbb{1} - |\psi_\gamma\rangle\langle\psi_\gamma|$ will be $\frac{d-1-\mathbf{c}_\gamma^{\mathrm{T}}\lambda}{d} = 1 - p_\gamma$. Using these, the least squares objective function (6.5) used for parameter estimation can be written in a simpler form, without the Choi matrix:

$$\arg\min_\lambda \sum_\gamma 2\left(\hat{p}_\gamma - \frac{1+\mathbf{c}_\gamma^{\mathrm{T}}\lambda}{d}\right)^2 ,$$
$$\text{so that} \quad 1 + d\lambda_i \geq \sum_j \lambda_j \geq -\frac{1}{d-1} .$$

If we know the channel structure, i.e., that the channel is truly Pauli, then the real λ satisfies the constraints and the global minimum of the objective function will be inside the feasible region. Thus the problem will be unconstrained:

$$\arg\min_\lambda \frac{2}{d}\left(\frac{1}{d}\lambda^{\mathrm{T}}\mathbf{C}^{\mathrm{T}}\mathbf{C}\lambda + \frac{2}{d}(1 - d\hat{\mathbf{p}})^{\mathrm{T}}\mathbf{C}\lambda + (d\hat{\mathbf{p}} - 2)^{\mathrm{T}}\hat{\mathbf{p}} + \frac{n_{\text{cfg}}}{d}\right) , \qquad (7.9)$$

where the rows of the $n_{\text{cfg}} \times (d+1)$ sized matrix $\mathbf{C} = [\mathbf{c}_1, \ldots, \mathbf{c}_{n_{\text{cfg}}}]^{\mathrm{T}}$ contain the n_{cfg} number of used configuration vectors, and the $\hat{\mathbf{p}}$ vector contains the measured outcome probabilities – the relative frequencies – obtained with respect to the rank-1 POVM element in each configuration. The problem (7.9) can now be easily solved analytically by setting the gradient equal to zero:

$$\hat{\lambda} = (\mathbf{C}^{\mathrm{T}}\mathbf{C})^{-1}\mathbf{C}^{\mathrm{T}}(d\hat{\mathbf{p}} - 1)$$

This shows that the solution of the parameter estimation problem (7.9) does not require the tomographically complete reconstruction of the Choi matrix.

To estimate the parameters, it is sufficient to have only $d+1$ linearly independent configuration vectors $\mathbf{c}_\gamma$. If we select these to be the optimal configuration vectors, then each of them corresponds to a state aligned with a complementary subalgebra, giving $\mathbf{C} = (d-1)\mathbb{1}$, and we arrive at the optimal estimator for the channel parameters:

$$\hat{\lambda} = \frac{d\hat{\mathbf{p}} - 1}{d-1}$$

Note that if needed, the Choi matrix can also be computed after the parameter estimation using (6.1) and the known H_k structure matrices. This method was used also in the case studies.

Thus we can summarize:

Statement 7.3. *For prime numbers d, the least squares based parameter estimation problem of d-level Pauli channels with known channel structure can be solved analytically, using only $d+1$ linearly independent configuration vectors. Moreover, if the optimal configurations are used, which are directed along the complementary subalgebras, then the problem can be solved in a very simple and efficient way.*

7.3 Case studies

The aim of the following simulation experiments was to demonstrate the differences between nonoptimal and optimal experiment configurations for the parameter estimation method of Pauli channels proposed in Chapter 6. The results were generated with the simulation tools discussed in section A.6. To facilitate comparison with the nonoptimal estimation cases, the tomography settings were the same here as in section 6.3.1, and the used channel parameter values were also the same as in sections 6.3.3 and 6.3.4.

7.3.1 Optimal estimation of qubit Pauli channel

The estimation of the qubit Pauli channel defined in (2.16) was simulated using the optimal configuration set described in section 7.2.1. The parameter vector was selected to be $\lambda = (-0.4, -0.6, 0.2)$, taken from the interior of the parameter space.

Assuming $n_{\text{tot}} = 18000$, the performance indicator quantities of the estimations were calculated in function of the number of complete experiments $n_{\text{exp}} = 6000$ for the case of the optimal configuration ($n_{\text{cfg}} = 3$), and $n_{\text{exp}} = 1500$ for the case of the nonoptimal configuration using the standard POVM set ($n_{\text{cfg}} = 12$) described in section 6.3.2. Thus the optimal method uses more complete experiments, but the reason for this is that the fixed amount of resources was distributed over a smaller number of configurations. The results are plotted in Figure 7.1.

More simulation examples with other channel parameter values taken both from the interior and the border of the parameter space can be seen in appendix B.1.1.

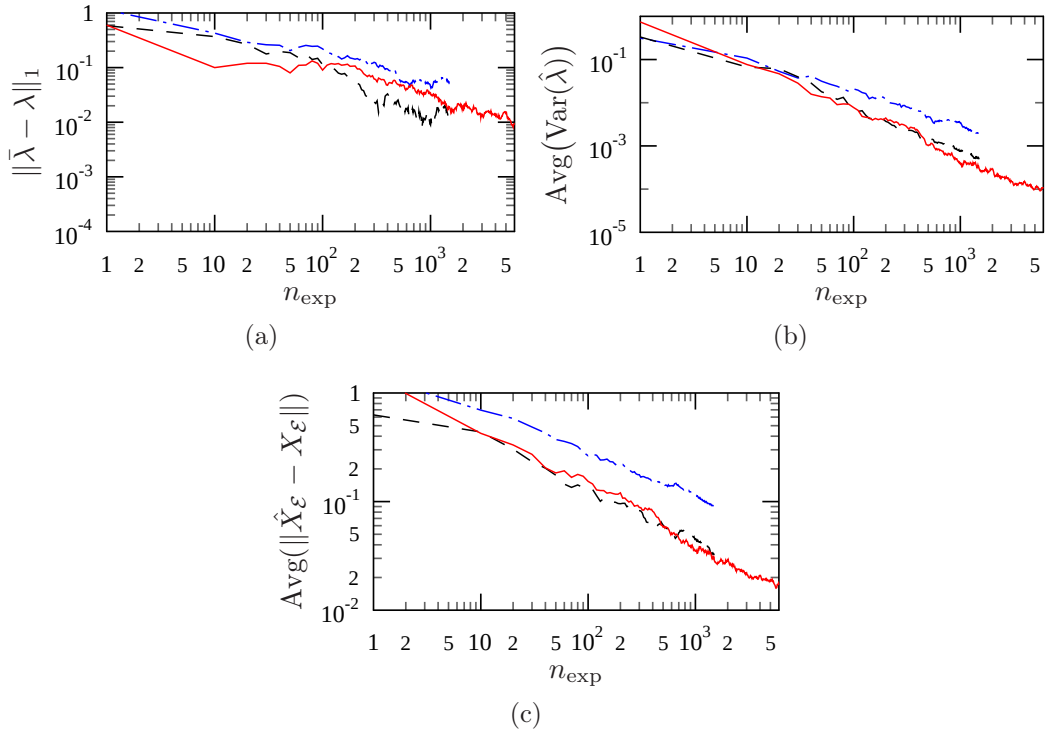


Figure 7.1. Optimal tomography with $\lambda = (-0.4, -0.6, 0.2)$. The three Figures show the performance indicators defined in section 6.3.1 respectively, in function of the number of complete experiments. The dotted-dashed line corresponds to tomography with unstructured Choi matrix, the dashed line shows results from parameter estimation using Choi matrix structure, and the solid line is obtained using the optimal configuration set.

The results indicate, that in terms of the performance indicators, the efficiency of the optimal experiment configuration highly outperforms the nonoptimal ones.

Robustness of the optimal experiment design

It is often the case that the optimal experimental configurations designed to probe a set of parameters are sensitive to the assumed parts of the model used to derive the optimal settings. This can be an issue mainly because even if the channel structure is known, that knowledge may be approximate. Thus, the aim of this subsection is to present a small example on the performance of the optimal experimental configuration for the case when the actual channel directions are slightly perturbed from the assumed channel directions.

We know from section 2.3.3, that in the qubit case, the channel directions $|\mathbf{v}_1\rangle$, $|\mathbf{v}_2\rangle$ and $|\mathbf{v}_3\rangle$ must be elements of bases that form a set of MUB. This tells us that the found (possibly inaccurate) channel directions $\{|\mathbf{v}'_i\rangle\}$ must be unitarily transformed versions of the real channel directions. This transformation corresponds to a rotation of the Bloch vectors $\mathbf{v}_1$, $\mathbf{v}_2$, and $\mathbf{v}_3$, around a given axis $\mathbf{a}$ with a given angle α.

The following example is a modified version of case study 7.3.1. The parameter estimation was done assuming that the $\{\mathbf{v}_i\}$ basis represents the channel directions, but the real channel was simulated using a perturbed basis $\{\mathbf{v}'_i\}$, where $\mathbf{v}'_i = R_{\mathbf{a}}(\alpha)\mathbf{v}_i$, the matrix $R_{\mathbf{a}}(\alpha)$ being a rotation matrix. The axis of rotation $\mathbf{a}$ was given by the Bloch vector $\frac{1}{\sqrt{3}}[1,1,1]^{\mathrm{T}}$. In Figure 7.2(a) the Hilbert–Schmidt norm, and in Figure 7.2(b) the estimated parameter means for each parameter were depicted in function of α after $n_\gamma = 1500$ measurements in each tomography configuration.

The Hilbert–Schmidt norm is clearly periodic. This is because in this example, after rotating the $\{\mathbf{v}_i\}$ basis by the angle $\alpha = \frac{2\pi}{3}$, we get the same basis $\{\mathbf{v}_i\}$, but with the order of the basis vectors permuted. This can be also seen in Figure 7.2(b), where we get the valid parameter values again after rotating by $\alpha = \frac{2\pi}{3}$ just in a different order. Of course, from the aspect of robustness, we are only interested in small perturbations, i.e., small values of α.

From this example, it can be seen that considering the Hilbert–Schmidt norm, the optimal experimental configurations are indeed sensitive to the accuracy of the assumed channel directions, as the norm changes linearly for small perturbations with the perturbation parameter α (see Figure 7.2(a)). However, the results in Figure 7.2(b) suggest that the estimated mean values of the parameters change only quadratically for small α values. This means that the optimal parameter estimation method is robust in this sense.

7.3.2 Optimal estimation of 3-level Pauli channel

The estimation of the qutrit Pauli channel defined in section 6.2.2 was simulated using the optimal configuration. The parameter vector was selected to be $\lambda = (0.4, 0.475, 0.325, 0.55)$. This was taken from the interior of the parameter space.

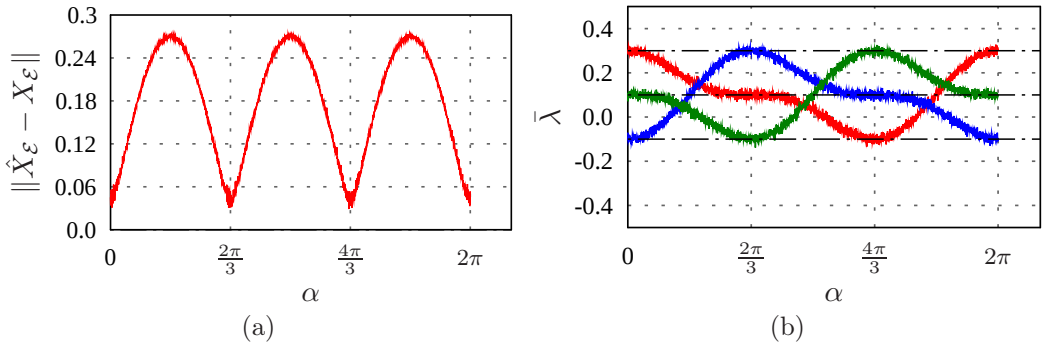


Figure 7.2. Example results on the robustness of the optimal experiment design for a qubit Pauli channel with parameters $\lambda = (0.3, -0.1, 0.1)$ against perturbations in the channel structure. In Figure (a) the Hilbert–Schmidt norm and in Figure (b) the estimated parameter means were depicted in function of the perturbation parameter α after $n_{\text{exp}} = 1500$ measurements in each tomography configuration.

Assuming $n_{\text{tot}} = 108000$, the performance indicator quantities of the estimations were calculated in function of the number of complete experiments $n_{\text{exp}} = 27000$ for the case of the optimal configuration ($n_{\text{cfg}} = 4$), and $n_{\text{exp}} = 1500$ for the case of the nonoptimal configuration ($n_{\text{cfg}} = 72$) described in section 6.3.2. The optimal method now uses considerably more complete experiments, but the reason for this is again that the fixed amount of resources was distributed over a considerably smaller number of configurations. The results are plotted in Figure 7.3.

More examples with other channel parameter values taken both from the interior and the border of the parameter space can be seen in appendix B.1.2.

The results indicate, that in terms of the performance indicators, the efficiency of the optimal experiment configuration highly outperforms the nonoptimal one also in the 3-level case.

7.4 Summary

In this Chapter, an experiment design procedure based on maximizing the trace of the Fisher information of the quantum channel output state was presented. It is shown that the objective function is convex in the configuration parameters. This way we have proven that the optimal input state should be pure and the measurement POVM should be extremal.

For Pauli channels in prime dimensions, this formulation leads to an optimal setting that includes pure input states and projective measurements directed towards the complementary subalgebras defining the channel. A simple way of estimating the channel parameters in the optimal configuration is also given, and the robustness of the optimal configuration was considered in the qubit case, too. The effect of the optimal configuration on the parameter estimation performance, compared to nonoptimal settings is demonstrated using case studies.

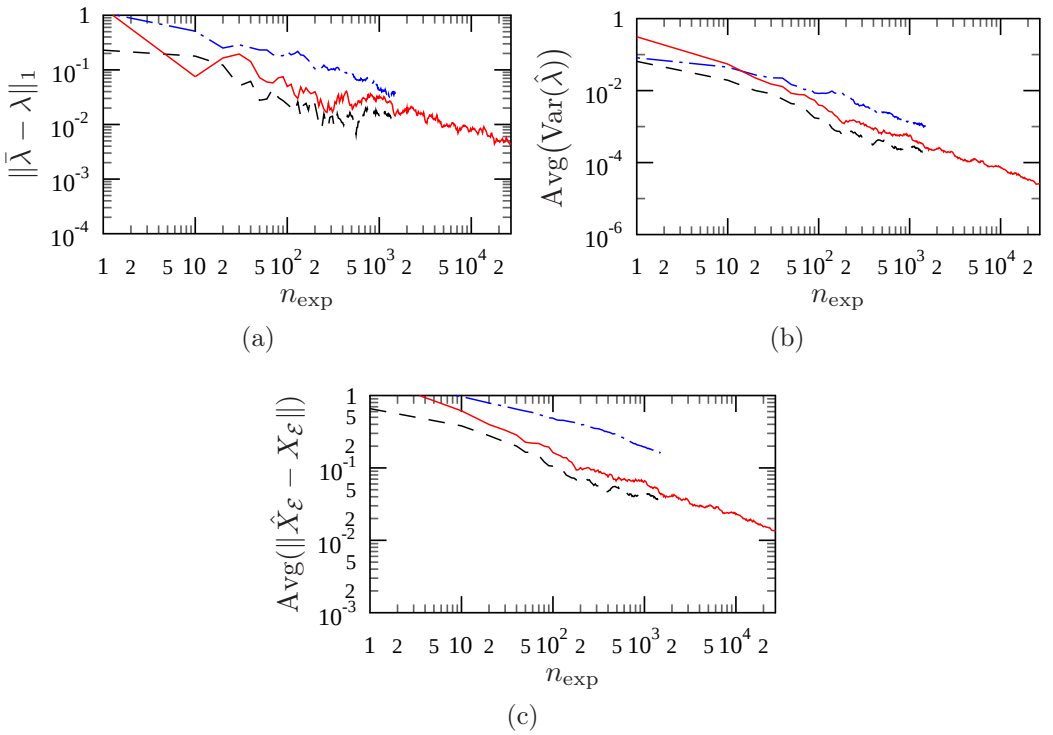


Figure 7.3. Optimal tomography with $\lambda = (0.4, 0.475, 0.325, 0.55)$. The three Figures show the performance indicators defined in section 6.3.1 respectively, in function of the number of complete experiments. The dotted-dashed line corresponds to tomography with unstructured Choi matrix, the dashed line shows results from parameter estimation using Choi matrix structure, and the solid line is obtained using the optimal configuration set.

Chapter 8

Identification of a Pauli channel with unknown structure

In this chapter, the identification of a Pauli channel with unknown structure is investigated. Two approaches are proposed, which can be used to estimate a completely undetermined two-level Pauli channel. The first one aims at estimating the channel directions in a one by one manner, while the second tries to find the optimal experiment design in an adaptive way, thus approaching the truly optimal channel estimate. Both methods are presented with numerical examples, their resource and computational requirements are discussed and compared to simpler methods.

In section 8.1 the problem of Pauli channel estimation with unknown structure is discussed. In section 8.2 an algorithm for qubit Pauli channel direction estimation is proposed, while in section 8.3 an adaptive estimation algorithm is discussed and compared to non-adaptive methods for the same problem. Finally, section 8.4 summarizes the results.

8.1 Problem statement

In section 5.3.1 it was mentioned that the optimal experiment design – in terms of Fisher information – is in general dependent on the parameters to be estimated. It follows that either the optimal design depends only on a subset of the parameters, which are assumed known, and it is independent of the rest of the parameters, or we can compute the optimal design only approximately using a first guess on the value of the parameters. The former case was discussed in chapters 6 and 7 about our results on the parameter estimation of Pauli channels and the related experiment design problem. In these chapters it was generally assumed that the Pauli channel directions are known. This fact was essential in the process of obtaining and applying our results. Recall from **Statement 7.2** that in the qubit scenario the optimal experiment design was to use parallel configurations, i.e., to prepare input states and perform measurements in the channel directions. Thus, the optimal configuration was independent of the unknown λ vector, and it was dependent of the known channel directions. The solution was similar in the 3-level case,

too. This shows that the knowledge of the Pauli channel structure is necessary for the application of the optimal experiment design. It follows that if – as it is the case in general – this knowledge is not available, i.e., we have a Pauli channel with

- unknown depolarizing parameters λ_i, and
- an unknown set of complementary subalgebras in which the channel is depolarizing (see section 2.3.2),

then for such Pauli channels we need fundamentally different approaches to solve the optimal identification problem.

For qubits, this unknown channel can be modeled with the matrix $\mathbf{S}$ of the rotated Pauli channel in section 2.3.3. This model contains six independent matrix elements, these can form the parameter vector $\mathbf{s} = [s_1, \ldots, s_6]^{\mathrm{T}}$. After channel estimation, the eigenvectors of the estimate $\hat{\mathbf{S}}$ will be the estimated Bloch vectors $\hat{\mathbf{v}}_i$ of the channel directions, and its eigenvalues will be the corresponding estimated scaling values $\hat{\lambda}_i$. The Choi matrix of such a generalized channel is

$$X_{\mathbf{S}} = \frac{1}{2}\begin{bmatrix} 1+s_3 & s_5 - \mathrm{i}s_6 & s_5 + \mathrm{i}s_6 & s_1 + s_2 \\ s_5 + \mathrm{i}s_6 & 1 - s_3 & s_1 - s_2 + 2\mathrm{i}s_4 & -s_5 - \mathrm{i}s_6 \\ s_5 - \mathrm{i}s_6 & s_1 - s_2 - 2\mathrm{i}s_4 & 1 - s_3 & -s_5 + \mathrm{i}s_6 \\ s_1 + s_2 & -s_5 + \mathrm{i}s_6 & -s_5 - \mathrm{i}s_6 & 1 + s_3 \end{bmatrix}$$

This matrix is linear in its parameters s_i, thus an exact affine decomposition based parametrization can be made. This implies

Statement 8.1. *The parameter estimation of the qubit Pauli channel with unknown channel directions remains a convex optimization problem.*

Of course, we could try to simply find the optimal experiment configuration $(\mathbf{b}, \{\pm\mathbf{m}\})$ for this channel similarly as in the known channel structure case in section 7.2.1, namely, by maximizing the Fisher information matrix $F(\mathbf{s}|\mathbf{b}, \mathbf{m})$ of a single experiment:

$$F(\mathbf{s}|\mathbf{b}, \mathbf{m}) = \frac{\nabla_{\mathbf{s}}(\mathbf{m}^{\mathrm{T}}\mathbf{S}\mathbf{b})\nabla_{\mathbf{s}}^{\mathrm{T}}(\mathbf{m}^{\mathrm{T}}\mathbf{S}\mathbf{b})}{1 - (\mathbf{m}^{\mathrm{T}}\mathbf{S}\mathbf{b})^2} \tag{8.1}$$

However, we see that this matrix depends on the full channel matrix $\mathbf{S}$. This clearly shows for qubits that – as we mentioned above based on **Statement 7.2** – the optimal configuration surely depends on the channel structure, i.e., the channel directions; thus it can not be determined without knowing $\mathbf{S}$ itself!

The aim of this chapter is to overcome this problem. In the following two sections, methods for the estimation of such general qubit Pauli channel will be studied using two fundamentally different approaches.

8.2 Channel direction estimation

The method described in this section estimates the unknown channel directions of a rotated Pauli channel from section 2.3.3, while resulting in a first estimate on the λ parameter values too.

8.2.1 Estimation algorithm for channel directions

Let the three directions in which the qubit Pauli channel is depolarizing be $|\mathbf{v}_1\rangle$, $|\mathbf{v}_2\rangle$ and $|\mathbf{v}_3\rangle$. Then quantum state estimation steps can be used to determine these. The proposed method is essentially an adaptation of the power iterations algorithm from linear algebra.

The effect of the channel $\mathcal{E}$ for the input pure Bloch vector $\tilde{\mathbf{b}}$ can be written as $\mathcal{E}(\tilde{\mathbf{b}}) = \mathbf{S}\tilde{\mathbf{b}} = \sum_{i=1}^{3} \lambda_i \mathbf{v}_i^{\mathrm{T}} \tilde{\mathbf{b}} \mathbf{v}_i$. In the rest of this section, the words "vector" and "state" are used as synonyms, both referring to Bloch vectors.

The task is to estimate the three depolarizing directions $\{|\mathbf{v}_i\rangle\}$ of $\mathcal{E}$ by estimating the corresponding Bloch vectors $\mathbf{v}_i$. Let the set of found channel direction Bloch vectors be $\mathbf{D}$. Let $\mathbf{D} = \{\}$ and $n = 0$, this is the initialization step. The following algorithm describes the direction estimation procedure.

Algorithm 1 Direction estimation

1: **repeat**
2: Prepare a pure state $\tilde{\mathbf{b}}^{(n)} \in \mathbf{D}^{\perp}$.
3: **repeat**
4: Put $\tilde{\mathbf{b}}^{(n)}$ into the composite channel $\mathcal{E}^k$ formed by cascading k instances of the channel $\mathcal{E}$, then get the output $\mathbf{b}^{(n+1)}$.
5: Perform quantum state tomography on $\mathbf{b}^{(n+1)}$ to get the estimate $\hat{\mathbf{b}}^{(n+1)}$.
6: Project $\hat{\mathbf{b}}^{(n+1)}$ to $\mathbf{D}^{\perp}$ to get $\hat{\mathbf{b}}^{(n+1)}_{\mathrm{proj}}$.
7: Normalize $\hat{\mathbf{b}}^{(n+1)}_{\mathrm{proj}}$ to get the pure state $\tilde{\mathbf{b}}^{(n+1)}$.
8: Increase n by 1.
9: **until** The distance $\left\|\tilde{\mathbf{b}}^{(n)} - \tilde{\mathbf{b}}^{(n+1)}\right\|$ is smaller than some prescribed value.
10: Put $\tilde{\mathbf{b}}^{(n+1)}$ into $\mathbf{D}$, set n to 0.
11: **until** Dimension of $\mathbf{D}^{\perp}$ is 0.

We now give the *mathematical arguments* that support the steps of the above algorithm.

Case of different channel parameter values

Assume that all of the λ_i channel parameters have different absolute values, and recall that $|\lambda_i| < 1,\ i = 1, 2, 3$.

- Step 4: Assume we use a pure state $\tilde{\mathbf{b}}^{(n)}$ as input to the channel and obtain the output $\mathbf{b}^{(n_1)} = \mathbf{S}\tilde{\mathbf{b}}^{(n)}$. If we repeat this procedure, i.e., we put the channel output $\mathbf{b}^{(n_1)}$ back into the channel as input to get the output $\mathbf{b}^{(n_2)}$, then by the power iterations method, the vector $\frac{\mathbf{b}^{(n_\ell)}}{\|\mathbf{b}^{(n_\ell)}\|}$ will converge to the axis of $\mathbf{v}_m$ corresponding to the dominant channel parameter λ_m, i.e., the parameter with the largest absolute value. The normalization in the above sequence is inevitable, as the output states do not remain pure during the iterated channel effect, i.e., the length $\left\|\mathbf{b}^{(n_\ell)}\right\|$ of the sequence will not remain 1, it will converge to zero instead.

- Step 5-7: Thus, to avoid the vector sequence from converging to the maximally mixed state, we have to do the normalization of the output Bloch vector $\mathbf{b}^{(n+1)} = \mathbf{b}^{(n_k)} = \mathbf{S}^k \tilde{\mathbf{b}}^{(n)}$ manually after each step. This means that we have to exchange the output state with the pure state which points in the same direction. In order to do this we need to perform quantum state tomography (Step 5) to get an estimate $\hat{\mathbf{b}}^{(n+1)}$ and normalize it in Step 7 to obtain the pure state $\tilde{\mathbf{b}}^{(n+1)} = \frac{\hat{\mathbf{b}}^{(n+1)}}{\|\hat{\mathbf{b}}^{(n+1)}\|}$ which can be put again into the channel. This way, the sequence of vectors will indeed converge to $\mathbf{v}_m$.
- Step 2,6,10,11: After the first channel direction $\mathbf{v}_m$ was found using this procedure, we can continue the search in the plane $\mathbf{D}^\perp$ orthogonal to $\mathbf{v}_m$ (Step 10 and 2). However, due to the inaccuracies in state tomography, the direction we will find will not be exactly $\mathbf{v}_m$, rather some vector $\tilde{\mathbf{b}}^\star \approx \mathbf{v}_m$. Thus convergence is more robust against these inaccuracies if we apply a projection to the output vector in Step 6 onto the subspace $\mathbf{D}^\perp$. When the second direction is found, then the third can be easily obtained, as it will be the one orthogonal to both the first and the second direction. Thus the direction estimation procedure is finished using only two iterations (Step 11).

Special cases

In the degenerate cases when some of the channel parameters λ_i have equal absolute values, then the channel is equally depolarizing in the linear span of those directions, i.e., there are no exact channel directions defined in that subspace. This means that we can use any state inside this subspace as channel direction, so the sequence $\left\{\tilde{\mathbf{b}}^{(n)}\right\}_{n=0}^{\infty}$ of states is only required to converge to an arbitrary state inside this subspace, which is guaranteed by the above procedure.

It follows, that if all the channel parameters have equal absolute values, then the channel is the depolarizing channel, which means that any three orthogonal Bloch vectors can be used to represent channel directions.

In the case of $\lambda_i = 1$ for some i, the above argument differs only in the fact that in theory, the normalization step is not necessary. The iterated channel effect does not make the length of the input Bloch vector tend to zero.

Moreover, in the special case when the starting vector $\tilde{\mathbf{b}}^{(n)}$ has zero component along the axis corresponding to the dominant channel parameter λ_m, then in theory, the algorithm will find the channel direction with the second dominant parameter first.

Accuracy and efficiency

The accuracy of this procedure has of course a limit, set by the accuracy of quantum state tomography. Convergence to a channel direction is guaranteed only until the difference in the input and output state is not comparable with

the uncertainty of the state estimation procedure. Thus, when the sequence reaches this limit, the searching procedure should stop. It can also occur that we give a good initial guess, and start with an input state which is very close to a channel direction. Then the procedure can finish almost immediately, because slow convergence can only occur close to channel directions.

This algorithm – though rather resource intensive – thus estimates the directions of a qubit Pauli channel. By using the algorithm we can get information also on the channel parameters which can then be made more accurate using the optimal tomography configurations described in section 7.2.1, thus making a two step Pauli channel estimation procedure.

Finally, an interesting question can be raised. If the channel directions are unknown then how practical is it to try obtaining them, possibly by using the method presented in this section? To answer this question, a comparison would be necessary from the aspect of resource requirement between our method of direction estimation combined with optimal experiment design and a general channel estimation method that uses no a priori knowledge about the channel structure. This study is not in the scope of this thesis, but the papers [33] and [27] suggest, that in order to achieve an estimation accuracy of order comparable with the results given in section 7.3 for qubit channels without making assumptions on the channel structure, may require a number of measurements of order 10^4–10^5. This is at least about the same order as the approximate measurement requirement of our two step procedure.

8.2.2 A simple numerical example

In order to illustrate the operation and properties of the above proposed channel direction estimation algorithm, a simple illustrative numerical example is presented here for a qubit Pauli channel with different parameters $\lambda_1 = 0.6$, $\lambda_2 = 0.3$, and $\lambda_3 = 0.1$.

The three unknown channel directions were chosen to be the eigenvectors of the Pauli matrices. The uncertainty in the estimated channel output state arising from quantum state tomography was simulated using random perturbations in the output state. The perturbation of the i^{th} Bloch vector component b_i is a random term of the form

$$\xi\sqrt{\frac{1-b_i}{N}}\ ,$$

where ξ is a random number taken form the standard normal distribution, N is the number of measurements in the state tomography step, and $\frac{1-b_i}{N}$ is the variance of the estimator $\hat{b_i}$.

The result of the numerical test can be seen in Figure 8.1. The three unknown channel directions are shown by the black axes in the Bloch sphere. The prime labeled vectors indicate the perturbed and normalized input states in each step, and the number labeled vectors indicate channel outputs. The starting input vector was chosen randomly at the beginning of the search. The

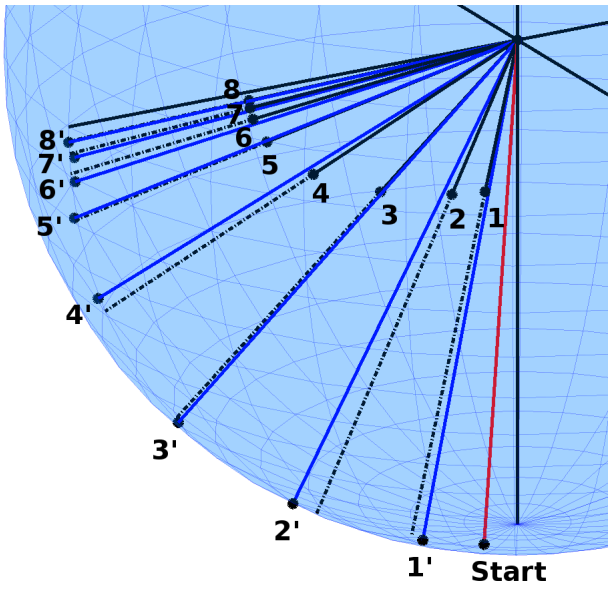


Figure 8.1. Channel direction estimation example for the qubit Pauli channel with parameters $\lambda_1 = 0.6$, $\lambda_2 = 0.3$, and $\lambda_3 = 0.1$, estimating the direction $m = 1$. The unknown channel directions are shown by the black axes in the Bloch sphere. The numbered vectors indicate the channel output (n) and its perturbed and normalized form (n') in the n^{th} step. The starting input vector was chosen randomly at the beginning of the search.

perturbation in the output states assumed $N = 5000$ measurements in the state tomography steps.

It can be seen from the figure that the sequence of input states converges to the dominant channel direction in the subspace of searching in a few iteration steps.

8.3 Adaptive channel estimation

A possible alternative to the direction estimation procedure in solving the problem of unknown Pauli channel directions is to make an adaptive estimation procedure. Such an approach is studied in the papers [43] ans [44] for the case of quantum state estimation. The main idea in these is that the optimal design is not decided in advance of the experiments, but the measurements are adaptively re-optimised depending on the collected data.

In other words, if the experimental configurations used in each experiment are not selected independently and/or identically, but are related, then we can speak about adaptive experiment design. The relation between the subsequent experiment configurations is based on an *update criterion* selected such that the obtained configuration sequence approaches optimality in some sense. Such an adaptive approach was not yet studied for quantum channels.

8.3.1 Adaptive estimation algorithm for quantum channels

For simplicity, assume that the set of possible input states and the set of possible measurement POVMs is fixed during the whole procedure. Let the configuration used in the n^{th} experiment be $(\rho_n, \mathbf{M}_n)$ and the update criterion be u_n. The fact that each configuration is used only once will be reflected in the notation: A_n will refer to the quantity A in the n^{th} experiment, and $\mathbf{A}_{(n)}$ will denote the sequence $\{A_1, \ldots, A_n\}$. Then the algorithm for adaptive process tomography can be described as follows:

Algorithm 2 Adaptive channel estimation

1: Prepare an initial configuration $(\rho_0, \mathbf{M}_0)$.
2: Set n to 0.
3: **repeat**
4: Perform the n^{th} experiment: α_n is the measurement outcome of the channel output $\mathcal{E}(\rho_n)$ using $\mathbf{M}_n$.
5: Build the data set $D_n = (\rho_n, \mathbf{M}_n, \alpha_n)$, and the data record $\mathbf{D}_{(n)} = \{D_1, \ldots, D_n\}$.
6: Calculate the channel estimate $\hat{\mathcal{E}}_n \coloneqq \hat{\mathcal{E}}\big(\mathbf{D}_{(n)}\big)$.
7: Calculate the next configuration $(\rho_{n+1}, \mathbf{M}_{n+1}) = u_n\big(\mathbf{D}_{(n)}\big)$.
8: Increase n by 1.
9: **until** Stop condition is satisfied.

The update criterion u_n can be selected arbitrarily in each experiment. In this thesis, we will use the same criterion for all n, i.e., the sequence u_n will be constant. This criterion will be the A-optimality criterion described in [44]. It is based on the Cramér–Rao bound (5.5). Taking the trace of both sides, we arrive at $\mathrm{Var}(\hat{\mathcal{E}}_n) \geq \mathrm{Tr}[F(\mathcal{E}|u_{(n)})^{-1}]$, where $F(\mathcal{E}|u_{(n)})$ is the Fisher information of the probability distribution $p(\mathbf{D}_{(n)}|\mathcal{E})$, i.e., the average information we can gain about the true channel from an experiment sequence of length n, using the update criterion sequence u_n. Based on this, each estimation step is followed by an experiment design step, which tries to search for that configuration, which maximizes the information gain from the current sequence of experiments. Thus, for each n the next configuration will be that $(\rho_{n+1}, \mathbf{M}_{n+1})$, which minimizes $F(\mathcal{E}|u_{(n+1)})^{-1}$, so we need to solve the optimization problem

$$(\rho_{n+1}, \mathbf{M}_{n+1}) = \arg\min_{(\rho,\mathbf{M})} \mathrm{Tr}[F(\mathcal{E}|u_{(n+1)})^{-1}]$$

Obtaining a solvable form

This problem, however, is not solvable practically [44], because the true value of $\mathcal{E}$ is unknown, and the calculation of the Fisher information would require taking the expectation over many experiment sequences, which is computationally intensive. To overcome these problems in the n^{th} step,

- we replace the true $\mathcal{E}$ with the currently available approximation $\hat{\mathcal{E}}_n$,
- we replace the exact $F(\hat{\mathcal{E}}_n|u_{(n)})$ with the *observed information* matrix $J(\hat{\mathcal{E}}_n|u_{(n)}) \coloneqq \sum_{k=1}^{n} F(\hat{\mathcal{E}}_n|\mathbf{C}_k)$. This is the sum of the information gain $F(\hat{\mathcal{E}}_n|\mathbf{C}_k)$ from individual experiments using the sets of configuration matrices $\mathbf{C}_k = \{C_{k,m} = \rho_k^{\mathrm{T}} \otimes M_{k,m}\}$. Recall from (7.1) that the Fisher information of the distribution $p(\alpha|\mathcal{E})$, i.e., a single experiment using configuration $\mathbf{C}$ is

$$F(\mathcal{E}_{\mathbf{p}}|\mathbf{C}) = \sum_m \frac{1}{\mathrm{Tr}(C_m X_{\mathbf{p}})} \nabla_{\mathbf{p}} \mathrm{Tr}\left(C_m X_{\mathbf{p}}\right) \nabla_{\mathbf{p}}^{\mathrm{T}} \mathrm{Tr}\left(C_m X_{\mathbf{p}}\right) \ ,$$

  where $\mathbf{p}$ is some parametrization of the channel $\mathcal{E}$.

Thus, we arrive at a solvable form of the optimal design problem of the $(n+1)^{\mathrm{th}}$ set of configuration matrices:

$$\mathbf{C}_{n+1} = \arg\min_{\mathbf{C}} \mathrm{Tr}\big([J(\hat{\mathcal{E}}_n|u_{(n)}) + F(\hat{\mathcal{E}}_n|\mathbf{C})]^{-1}\big)$$

As $F(\hat{\mathcal{E}}_n|\mathbf{C})$ is rank-1, the inversion of the matrix $[J(\hat{\mathcal{E}}_n|u_{(n)}) + F(\hat{\mathcal{E}}_n|\mathbf{C})]$ can be performed using the formula $(A+B)^{-1} = A^{-1} - \frac{A^{-1}BA^{-1}}{1+\mathrm{Tr}(A^{-1}B)}$ where B is rank-1. Thus only the inversion of $J(\hat{\mathcal{E}}_n|u_{(n)})$ is required, which is independent of the optimization variable, so it needs to be computed only once during optimization. From the optimization viewpoint, the constant term $J(\hat{\mathcal{E}}_n|u_{(n)})^{-1}$ can be omitted by the same argument. Thus the resulting final problem is

$$\mathbf{C}_{n+1} = \arg\max_{\mathbf{C}} \frac{\mathrm{Tr}[J(\hat{\mathcal{E}}_n|u_{(n)})^{-1} F(\hat{\mathcal{E}}_n|\mathbf{C}) J(\hat{\mathcal{E}}_n|u_{(n)})^{-1}]}{1 + \mathrm{Tr}[J(\hat{\mathcal{E}}_n|u_{(n)})^{-1} F(\hat{\mathcal{E}}_n|\mathbf{C})]}. \tag{8.2}$$

Note that the experiment sequences obtained this way must be tomographycally complete, because the update criterion has the freedom to select that configuration which gives the most information on all parameters.

The above suggested adaptive method will be called the A-optimal adaptive method below.

8.3.2 Qubit Pauli case

Algorithm 2 can now be applied on the Pauli channel estimation problem described in section 8.1. Let us restrict ourselves to the case of two-element POVMs, i.e., von Neumann measurements. Then the design variable, i.e., the experiment configuration can be written as $(\mathbf{b}, \{\pm\mathbf{m}\})$. The Fisher information of a single experiment will then be (8.1). Using this, the A-optimality update criterion will need the solution of (8.2) in the form

$$(\mathbf{b},\mathbf{m})_{n+1} = \arg\max_{(\mathbf{b},\mathbf{m})} \frac{\mathrm{Tr}\big[J(\hat{\mathbf{s}}_n|u_{(n)})^{-1} F(\hat{\mathbf{s}}_n|\mathbf{b},\mathbf{m}) J(\hat{\mathbf{s}}_n|u_{(n)})^{-1}\big]}{1 + \mathrm{Tr}\big[J(\hat{\mathbf{s}}_n|u_{(n)})^{-1} F(\hat{\mathbf{s}}_n|\mathbf{b},\mathbf{m})\big]}. \tag{8.3}$$

Unfortunately, the solution of this problem requires global optimization, because this objective is nonconvex. This is a disadvantage of this method, because of the amount of computational resource needed.

In the first few experiments, the matrix $J(\hat{\mathbf{s}}_n|u_{(n)})$ may be singular. In these cases, we can calculate the Moore–Penrose pseudoinverse. This introduces error only in the first few steps of the algorithm, the sequence of optimal configurations $(\mathbf{b}, \{\pm\mathbf{m}\})_n$ will eventually span the full configuration space.

It is important to note that in the numerical case studies in section 8.3.4 the solution of (8.3) was always a parallel configuration (see **Definition 7.1**), i.e., the input part $\mathbf{b}$ of the optimal solution was always a pure state, and the measurement $\mathbf{m}$ was always parallel to $\mathbf{b}$.

Finally, as we discussed in section 8.1, the channel estimation step can be performed with the ML estimation method, because similarly to the case with known channel directions, the problem is convex in the channel parameters. The LS method can not be used in this case, because we perform only one experiment with each configuration (see section 5.2.3).

8.3.3 Non-adaptive methods

Here we shortly describe three other non-adaptive experiment configuration selection methods. The performance of these were tested and compared to the A-optimal adaptive method.

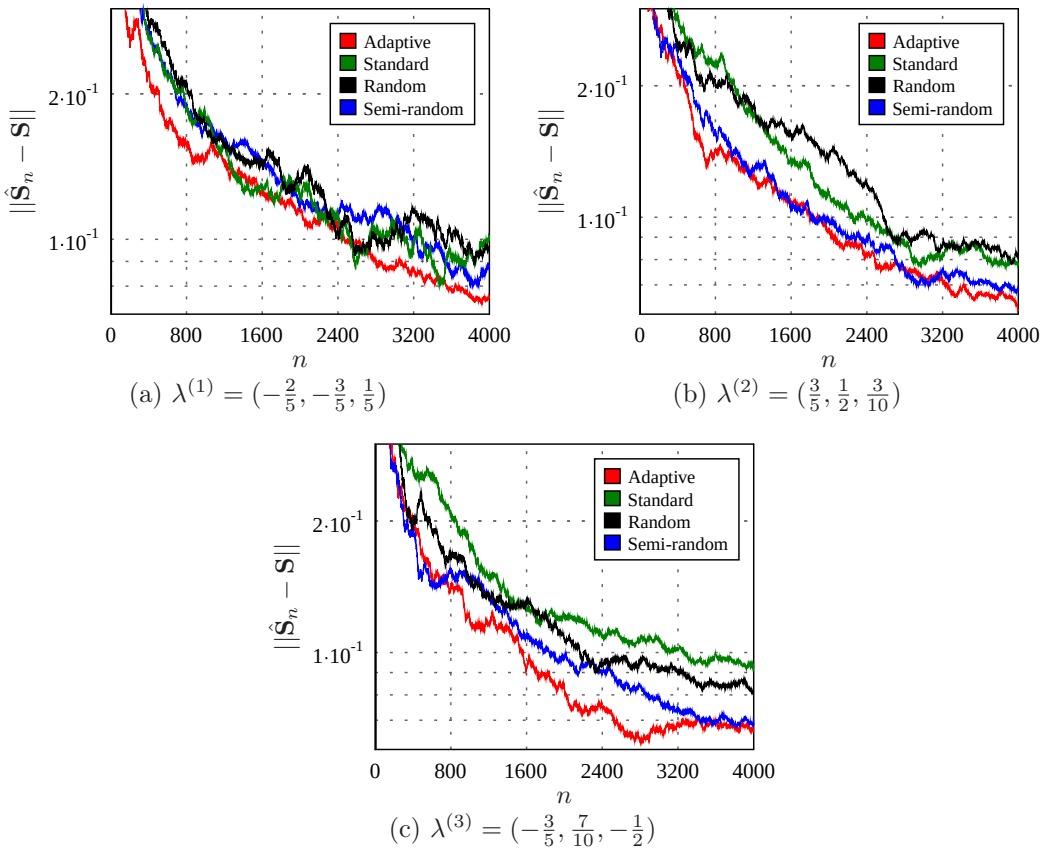
- Standard method: The experiment configuration variables $(\mathbf{b}, \mathbf{m})$ were selected to be Bloch vectors of states and measurements along the fixed set of channel directions $\{|\mathbf{e}_i\rangle\}$, i.e., $(\mathbf{b}, \mathbf{m}) = (\mathbf{e}_i, \mathbf{e}_j)$ with $i, j = 1, 2, 3$. This is a tomographically complete experiment setting for qubit Pauli channels.
- Random method: This method uses random input Bloch vectors $\mathbf{b}$ and random measurement vectors $\mathbf{m}$ independently of $\mathbf{b}$.
- Semi-random method: This method uses configuration variables $(\mathbf{b}, \mathbf{m})$ which make parallel configurations, i.e., are of the form $(\mathbf{b}, \mathbf{b})$ using random unit length Bloch vectors $\mathbf{b}$. This method was proposed based on the following observations suggesting that parallel configurations are optimal:
  - According to **Statement 7.2**, the optimal configuration for the given channel direction case was a parallel configuration.
  - In the numerical tests in section 8.3.4, the solutions of the adaptive method were also parallel configurations.

8.3.4 Case studies

Simulation experiments were used to compare the performance of the adaptive and non-adaptive experiment design methods for qubit Pauli channels. In the experiments, pure states and projective measurements were used.

To have a general setup, the Pauli channel directions used in the tests were represented with the matrix

$$[\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3] = \begin{bmatrix} 0.08908 & -0.96225 & 0.25717 \\ 0.44543 & -0.19245 & -0.87438 \\ 0.89087 & 0.19245 & 0.41147 \end{bmatrix}.$$



(a) $\lambda^{(1)} = (-\frac{2}{5}, -\frac{3}{5}, \frac{1}{5})$

(b) $\lambda^{(2)} = (\frac{3}{5}, \frac{1}{2}, \frac{3}{10})$

(c) $\lambda^{(3)} = (-\frac{3}{5}, \frac{7}{10}, -\frac{1}{2})$

Figure 8.2. The Hilbert–Schmidt distance between affine representation matrices of the real and estimated channels in function of the number of experiments. Colors red, green, black and blue correspond to the adaptive, fixed, random and semi-random strategies respectively.

Two sets of λ_i depolarizing parameters were taken from the convex tetrahedron parametrized by the λ_i (see section 2.3.3): $\lambda^{(1)} = (-\frac{2}{5}, -\frac{3}{5}, \frac{1}{5})$, $\lambda^{(2)} = (\frac{3}{5}, \frac{1}{2}, \frac{3}{10})$ and $\lambda^{(3)} = (-\frac{3}{5}, \frac{7}{10}, -\frac{1}{2})$. These values were taken from the interior of the tetrahedron of channel parameters.

Ten independent simulations of each setup was averaged in order to obtain the average estimation efficiency of each method. The estimation efficiency was measured by the Hilbert–Smith distance $\|\hat{\mathbf{S}}_n - \mathbf{S}\|$ of the estimated and real affine representation matrices and was plotted in function of the number of experiments in Figures 8.2(a)-(c).

As a conclusion, we can state that

Statement 8.2. *Efficiency of the A-optimal adaptive experiment design outperforms the non-adaptive ones, hovewer, its high computational resource requirement may not make it worth using, compared to the proposed non-adaptive semi-random strategy, which has the second best efficiency.*

8.4 Summary

In this chapter, two approaches were proposed to handle the case where the Pauli channel directions are not known. First, an efficient iterative method of estimating the channel directions was proposed for the qubit Pauli channel case, based on the power iterations method for matrices. The algorithm is resource intensive, but offers a two-step estimation procedure for completely unknown channels.

The second approach studied the A-optimal adaptive procedure for quantum state estimation, and proposed a similar method for the channel tomography case. Finally, the efficiency of the obtained adaptive procedure was tested against simple, non-adaptive methods. It was found that for the qubit Pauli channel, the A-optimal design has a better performance than other methods, however, its high resource requirement suggests that the proposed non-adaptive semi-random strategy can be a reasonable alternative in practice. This latter strategy is based on the optimality conditions suggested by the case of experiment design with known channel directions, and the A-optimal design results.

Chapter 9

Conclusions

This chapter summarizes the new scientific contribution of this thesis. Section 9.1 presents the new results. Section 9.2 gives possible directions of further research and section 9.3 lists the publications of the author.

9.1 New results

The new scientific results presented in this thesis are summarized here. They are arranged in four thesis statements as follows.

Thesis 1. *The notion of robustness of quantum error correcting (QEC) operations against noisy channel uncertainties was defined. Using the definition, the robustness of QEC operations was characterized.* (See in Chapter 4 and in [**P1**].)

(i) Channel uncertainty was represented as the convex combination of a nominal channel and a perturbation channel. By assuming arbitrary perturbations (unstructured uncertainty), robustness domains were defined as regions around the nominal noise channel. Pauli channels having the same optimal QEC operation with respect to a stabilizer code form a zero-robustness domain in the set of all Pauli channels. There are finite numbers of such domains and they form a partition of the whole set of Pauli channels represented as a simplex. If a channel is an interior point of such a domain then the optimal QEC operation is robust against any Pauli-type perturbation of the channel. On the boundary between two zero-robustness regions, however, there are more (at least two) completely different optimal QEC operations giving the same maximal channel fidelity.

(ii) For Pauli channels, the optimal QEC operation is better in robustness than any other recovery operation, assuming the channel is not on the boundary surface of a zero-robustness domain.

(iii) Case studies with non-Pauli channels suggest that the optimal QEC operation can in general change in two different ways. The first is similar to the case of boundary points of Pauli zero-robustness domains. In these special points the optimal recovery operation is not unique, it changes

abruptly and the channel fidelity is a non-analytic function of the mixing parameter. Between these points it changes analytically with the mixing parameter, indicating that the optimal QEC operation is robust in general.

Thesis 2. *A parameter estimation method of Pauli channels with known channel structure based on convex optimization was proposed.* (See in Chapter 6 and in [**P2**, **P3**, **P4**].)

(i) Utilizing the fact that the Choi matrix has an affine dependence on functions of the channel parameters, an approximation method was given, that can take into account the channel structure in quantum process tomography, yielding a parameter estimation problem. It was found that for Pauli channels, this parameter estimation problem is always purely convex, and solvable in any prime dimension. Simulation case studies performed for the case of qubit and qutrit Pauli channels show, that compared to the case when no channel structure is assumed, the proposed method of affine decomposition of the Choi matrix can significantly increase the accuracy of the parameter estimation.

(ii) For non-Pauli channels, convex relations between optimization variables can also be exploited to improve the estimation using auxiliary optimization steps. In this case, however, beside the convex part one may need a nonconvex optimization step, as well.

Thesis 3. *A convex maximization-based experiment design procedure was proposed for prime-level Pauli channels with known channel structure.* (See in Chapter 7 and in [**P2**, **P5**].)

(i) It was shown that the trace of the Fisher information matrix is a convex function in the configuration parameters. Using it as objective, it is proven that in the optimal experiment configuration the optimal input state must be pure and the measurement POVM must be extremal.

(ii) Based on maximizing the trace of the Fisher information matrix of a qubit Pauli channel output, this formulation leads to optimal configurations containing pure input states and projective measurements directed towards the channel directions. These constitute an optimal set of configurations for the independent estimation of the channel parameters.

(iii) A simple way of estimating the channel parameters with the optimal configuration is also given. Furthermore, the robustness of the optimal configuration against channel uncertainties was studied in the qubit case. It was found that the estimated parameters are robust against small perturbations in the channel directions. Finally, the superior parameter estimation performance of the optimal configuration set compared to other widely used ones was demonstrated using case studies for qubit and 3-level Pauli channels.

Thesis 4. *Two approaches were proposed to handle the case of qubit Pauli channel estimation where the channel structure is unknown.* (See in Chapter 8 and in [**P2**, **P5**, **P6**].)

(i) An efficient iterative method of estimating the channel directions was proposed for the qubit Pauli channel case, based on the power iterations method for matrices. The algorithm is resource intensive, but offers an optimal two-step estimation procedure for completely unknown channels.

(ii) An adaptive procedure was designed for the estimation of Pauli channels with unknown directions, based on the A-optimality update criterion. The efficiency of this procedure was tested against simple, non-adaptive methods. One of these was the semi-random strategy, which was proposed based on the optimality conditions of **Thesis 3** and numerical test results of the adaptive method. It was found, that for the qubit Pauli channel, the A-optimal design has a better performance than other methods, however, its high resource requirement suggests that the proposed non-adaptive semi-random strategy can be a reasonable alternative in practice.

9.2 Further work

The results of Chapter 4 could be directly continued with further studies of the robustness in the non-Pauli channel case, in particular by studying the distribution of the non-analytic points in the space of all quantum channels. Other interesting topics could be the analytic derivation of optimal QEC operations for more general classes of channels.

The work in Chapter 6 could be developed further by studying the case of completely general complementary subalgebras.

The results of Chapter 7 could be extended by deriving and proving analytic results also for arbitrary dimensional Pauli channels.

The results of Chapter 8 could be strengthened by finding a direction estimation algorithm for higher level Pauli channels, and by finding a computationally less intensive adaptive procedure that can also be generalized to higher level channels.

9.3 Publications

The results of this thesis were published in journals or presented in conferences as enlisted below. The relevant theses are indicated in parentheses.

Thesis related publications

[**P1**] **G. Balló**, **P. Gurin**, Robustness of channel-adapted quantum error correction, *Phys. Rev. A*, **80 (1)**, 012326 (Jul 2009), URL `arXiv:0905.3838v2` (**Thesis 1**)

[**P2**] **G. Balló**, **K. M. Hangos**, **D. Petz**, Convex Optimization-Based Parameter Estimation and Experiment Design for Pauli Channels, *IEEE Transactions on Automatic Control*, **57 (8)**, 2056–2061 (Aug 2012) (**Thesis 2, 3, 4**)

[**P3**] **G. Balló**, **A. Magyar**, **K. M. Hangos**, Quantum process tomography using optimization methods, *6th Central European Quantum Information Processing Workshop (CEQIP)* (Jun 2009, Jindřichův Hradec, Czech Republic) (**Thesis 2**)

[**P4**] **G. Balló**, **A. Magyar**, **K. M. Hangos**, Parameter estimation of quantum processes using convex optimization, in *Proceedings of the International Symposium on Mathematical Theory of Networks and Systems (MTNS)* (Jul 2010, Budapest, Hungary), URL `arXiv:1004.5209v1` (**Thesis 2**)

[**P5**] **G. Balló**, **K. M. Hangos**, Experiment Design for Pauli Channel Estimation, *18th IFAC World Congress* (Aug 2011, Milan, Italy) (**Thesis 3, 4**)

[**P6**] **G. Balló**, **K. M. Hangos**, Comparison of experiment design approaches for Pauli channel tomography, *10th Central European Quantum Information Processing Workshop (CEQIP)* (Jun 2013, Valtice, Czech Republic) (**Thesis 4**)

Other publications

[**O1**] **S. Varga**, **G. Balló**, **P. Gurin**, Structural properties of hard disks in a narrow tube, *Journal of Statistical Mechanics: Theory and Experiment*, **2011 (11)**, P11006 (Nov 2011)

Citations in journals

[**C1**] **S. Taghavi**, **T. A. Brun**, **D. A. Lidar**, Optimized entanglement-assisted quantum error correction, *Phys. Rev. A*, **82**, 042321 (Oct 2010) (cites [**P1**])

[**C2**] **Y. Ouyang**, **W. H. Ng**, Truncated quantum channel representations for coupled harmonic oscillators, *Journal of Physics A-Mathematical and Theoretical*, **46 (20)** (May 2013) (cites [**P1**])

Appendix

Appendix A

Basic notions in mathematics and information theory

The purpose of this appendix is a short overview, not the full and mathematically precise exposition of several topics related to the thesis. Basic knowledge in mathematical analysis and functional analysis is assumed.

A.1 Basics of Hilbert spaces

This section introduces the mathematical basics on vector spaces, that are needed in the thesis. We are concerned only in finite-dimensional vector spaces, thus finite-dimensionality is a general assumption throughout the section. The presentation follows [45].

A.1.1 The Hilbert space

Definition A.1. *Let $\mathcal{H}$ be a vector space over the complex numbers. Then the function $\langle\,.\,,\,.\,\rangle\colon \mathcal{H}\times\mathcal{H}\to\mathbb{C}$ is an inner product if it satisfies the following for all $x, y, z \in \mathcal{H}$ and $c \in$:*

- *Linearity in the second argument: $\langle x, cy + z\rangle = c\langle x, y\rangle + \langle x, z\rangle$,*
- *Conjugate symmetry: $\langle x, y\rangle^* = \langle y, x\rangle$,*
- *Positive definitness: $\langle x, x\rangle \geq 0$ and $\langle x, x\rangle = 0$ if and only if $x = 0$.*

A complex vector space $\mathcal{H}$ with an inner product defined on it is a *Hilbert space*. Furthermore, any such Hilbert space is isomorphic to $\mathbb{C}^d$, where d is the dimension of the space.[30]

A.1.2 Linear operators on Hilbert spaces

In the thesis *linear operators on Hilbert spaces* are of great importance. Through the above isomorphism a linear operator T on the d dimensional Hilbert space $\mathcal{H}$ can be represented as a complex $d \times d$ matrix.

[30]Note that infinite dimensional Hilbert spaces have a much more complicated definition.

For each linear operator T, we can define the *adjoint operator* T^* by the formula $\langle Tx, y\rangle = \langle x, T^*y\rangle$ for all $x, y \in \mathcal{H}$. The matrix isomorphic to T^* is the transposed and complex conjugated matrix of T.

The set $\mathcal{B}(\mathcal{H})$ of all (bounded) linear operators on $\mathcal{H}$ is also a complex Hilbert space with respect to the *Hilbert–Schmidt inner product* defined on linear operators A and B as

$$\langle A, B\rangle = \mathrm{Tr}(A^*B) \ . \tag{A.1}$$

The composition of linear operators corresponds to the product of their matrices. It follows that $\mathcal{B}(\mathcal{H})$ is also an algebra, which is isomorphic to the full matrix algebra $M_d(\mathbb{C})$, i.e., the algebra of all $d \times d$ complex matrices.

Self-adjoint operators

A linear operator T is called *self-adjoint* (Hermitian) if $T = T^*$. Since the real linear combinations of self-adjoint operators is self-adjoint, the set of self-adjoint operators $\mathcal{B}_{\mathrm{sa}}(\mathcal{H}) \subset \mathcal{B}(\mathcal{H})$ is a real Hilbert space. Note however, that $\mathcal{B}_{\mathrm{sa}}(\mathcal{H})$ is not an algebra.

A self-adjoint operator S always satisfies $\langle x, Sx\rangle \in \mathbb{R}$ for every $x \in \mathcal{H}$. If $\langle x, Sx\rangle \geq 0$ for all $x \in \mathcal{H}$ then S is called *positive* (positive semidefinite), $S \geq 0$ in notation. It follows that the matrix of a positive S has nonnegative eigenvalues.

Given $T \in \mathcal{B}(\mathcal{H})$ the operator T^*T is always positive, and any positive operator S can be written as $S = T^*T$ for some $T \in \mathcal{B}(\mathcal{H})$. Moreover, there is always a unique positive $T_{\mathrm{sa}} \in \mathcal{B}_{\mathrm{sa}}(\mathcal{H})$ for which $S = T_{\mathrm{sa}}^2$. Then T_{sa} is denoted by $S^{\frac{1}{2}}$.

Projections

A self-adjoint operator P is a *projection* if $P = P^2$. Projections are always positive operators. If P is a projection then $I - P$ is called the *complement* of P.

If the range of P is one-dimensional then P is a *one-dimensional* projection. Such projections can be defined using unit vectors. If $x \in \mathcal{H}$ is a unit vector then the operator P_x given by $P_x y = \langle x, y\rangle x$ is a one-dimensional projection. It follows that the matrix of P is rank-1. Two one-dimensional projections P_x and P_y are orthogonal, i.e., $P_xP_y = 0$ if and only if the unit vectors x and y defining them are orthogonal.

Unitary operators

A linear operator U is called *unitary* if $UU^* = U^*U = \mathbb{1}$. It follows from the definition that U is invertible and $U^{-1} = U^*$. Thus the matrix of U has eigenvalues with 1 absolute value.

Unitaries preserve the inner product, i.e, $\langle x, y\rangle = \langle Ux, Uy\rangle$. This implies that unitaries transform orthonormal bases to orthonormal bases. Unitary operators form a group with respect to operator composition.

A.1.3 Dual space and tensor product

Let $\mathcal{V}$ be a complex vector space. A linear mapping $f\colon \mathcal{V} \to \mathbb{C}$ is called a *linear functional*. The set of all linear functionals on $\mathcal{V}$ is called the *dual vector space* $\mathcal{V}^*$ of $\mathcal{V}$.

If $\mathcal{H}$ is a Hilbert space then each vector $x \in \mathcal{H}$ defines a continuous linear functional f_x on $\mathcal{H}$ by the formula $f_x(y) = \langle x, y\rangle$. It is known that all continuous linear functionals on $\mathcal{H}$ have this form, which implies that the space $\mathcal{H}$ and its dual $\mathcal{H}^*$ can be identified naturally through the scalar product.

Definition A.2. *Let $\mathcal{H}_1$ and $\mathcal{H}_2$ be two Hilbert spaces. For each $x_1 \in \mathcal{H}_1$ and $x_2 \in \mathcal{H}_2$ we define the mapping $x_1 \otimes x_2\colon \mathcal{H}_1 \times \mathcal{H}_2 \to \mathbb{C}$ as*

$$(x_1 \otimes x_2)(z_1, z_2) = \langle z_1, x_1\rangle\langle z_2, x_2\rangle \ . \tag{A.2}$$

Then the set of all such mappings span the Hilbert space $\mathcal{H}_1 \otimes \mathcal{H}_2$ called tensor product space of $\mathcal{H}_1$ and $\mathcal{H}_2$ with the inner product

$$\langle x_1 \otimes x_2, y_1 \otimes y_2\rangle = \langle x_1, x_2\rangle\langle y_1, y_2\rangle \ .$$

Note that (A.2) implicitly enforces the bilinearity of the tensor product operator $\otimes\colon \mathcal{H}_1 \times \mathcal{H}_2 \to \mathcal{H}_1 \otimes \mathcal{H}_2$ assigning $x \otimes y$ to (x, y).

Similarly to vectors, the tensor product of two linear operators $T_1 \in \mathcal{B}(\mathcal{H}_1)$ $T_2 \in \mathcal{B}(\mathcal{H}_2)$ can be defined. The product $T_1 \otimes T_2$ is an element of $\mathcal{B}(\mathcal{H}_1 \otimes \mathcal{H}_2)$ with action for all $x \in \mathcal{H}_1$ and $y \in \mathcal{H}_2$ given as

$$(T_1 \otimes T_2)(x \otimes y) = T_1 x \otimes T_2 y \ .$$

The tensor product of operators corresponds to the Kronecker product of their matrices. The Kronecker product of an $n \times m$ matrix A with any other matrix B is defined to be

$$A \otimes B = \begin{bmatrix} a_{1,1}B & \cdots & a_{1,m}B \\ \vdots & \ddots & \vdots \\ a_{n,1}B & \cdots & a_{n,m}B \end{bmatrix} \ .$$

A.2 Convex optimization

This section is based on the work [42], which also contains the proofs of the theorems presented here.

A.2.1 Convex functions

Here we give the definition of convex functions then the first and second order conditions of convexity.

Definition A.3. *A function $f\colon \mathbb{R}^n \to \mathbb{R}$ is convex if $dom(f)$ is a convex set, and for all $x_1, x_2 \in dom(f)$ and $t \in [0, 1]$, we have $f\big(tx_1 + (1-t)x_2\big) \leq tf(x_1) + (1-t)f(x_2)$.*

Theorem A.1. *If $f\colon \mathbb{R}^n \to \mathbb{R}$ is differentiable then f is convex if and only if $dom(f)$ is a convex set and for all $x, y \in dom(f)$*

$$f(y) \geq f(x) + \nabla^{\mathrm{T}} f(x)(y - x) \ .$$

Theorem A.2. *If $f\colon \mathbb{R}^n \to \mathbb{R}$ is twice differentiable then f is convex if and only if $dom(f)$ is a convex set and for all $x \in dom(f)$ the Hessian $\nabla^2 f(x)$ is positive semidefinite, i.e., $\nabla^2 f(x) \geq 0$.*

Note that if we change each of the inequalities above to strict inequalities then the funtion f will be *strictly convex*.

Many operations exist that preserve convexity of a function. The following two theorems are used also in the thesis.

Theorem A.3 (Product of convex functions)**.** *If $f, g\colon \mathbb{R}^n \to \mathbb{R}$ are differentiable, nonnegative and convex functions on a convex domain satisfying $\nabla f(x)\nabla^{\mathrm{T}} g(x) \geq 0$ for all x in that domain, then fg is also convex.*

The semidefinitness condition $\nabla f(x)\nabla^{\mathrm{T}} g(x) \geq 0$ corresponds to the requirement that f and g are simultaneously nondecreasing (or nonincreasing) at every x.[31]

Theorem A.4 (Composition of convex functions)**.** *If $h\colon \mathbb{R}^m \to \mathbb{R}$ is nondecreasing (nonincreasing) in each argument and each component g_k of $\boldsymbol{g}\colon \mathbb{R}^n \to \mathbb{R}^m$ is convex (concave), then $h \circ \boldsymbol{g}$ is convex.*

A.2.2 Convex optimization

In many engineering problems, such as control systems, estimation and signal processing, communications and networks, electronic circuit design, data analysis and modeling, convex optimization has a wide range of applications, therefore, it is an important subfield of mathematical optimization. Some very popular methods also belong to the class of convex optimization problems, such as least squares, and linear programming.

The general form of a mathematical optimization problem with functions $f, g_i, h_i\colon \mathbb{R}^n \to \mathbb{R}$ is

$$\begin{aligned} &(\arg)\min f(x) \text{ so that} \\ &g_i(x) \leq 0, \quad i = 1, \ldots, m \\ &h_i(x) = 0, \quad i = 1, \ldots, p \ . \end{aligned} \tag{A.3}$$

We would like to find an x that minimizes the *objective function* $f(x)$ while being a *feasible point*, i.e., satisfying the *constraints* $g_i(x) \leq 0$ and $h_i(x) = 0$. The set of all x points satisfying the constraints is the *feasible region* inside the domain $\mathcal{D} \coloneqq [\bigcap_{i=1}^{m} \mathrm{dom}(g_i)] \cap [\bigcap_{i=1}^{p} \mathrm{dom}(h_i)] \cap \mathrm{dom}(f)$ of the optimization

[31] This theorem can be proven for nondifferentiable f and g too. Then the condition is $\big(f(x_1) - f(x_2)\big)\big(g(x_1) - g(x_2)\big) \geq 0$ for all x_1, x_2.

problem. Note that speaking only about minimization is not a loss of generality. By multiplying f with -1 we arrive at an equivalent maximization problem.

That feasible $x^\star \in \mathcal{D}$ which minimizes $f(x)$ is called *optimal point* and the corresponding function value is the *optimal value* $f(x^\star)$. The feasible point $x^\star$ is *locally optimal* if it minimizes $f(x)$ only for a subset of feasible points $\{y \,|\, y \in \mathcal{D}, \|x^\star - y\| \leq r\}$ for some $r > 0$. Otherwise, $x^\star$ is *globally optimal*. The optional „arg" in the problem statement (A.3) indicates that we are interested in $x^\star$ instead of $f(x^\star)$ as a result.

In order (A.3) to be a convex optimization problem, f and each g_i must be convex and each h_i must be affine. This means that the feasible region of the problem is convex too. The most important properties of a convex optimization problem are the following:

- If there exists a local minimum point $x^\star$, then it is a global minimum point.
- The set of all global minimum points is convex.
- If the function f is strictly convex, then there exists at most one global minimum point.
- The solution can be obtained to within any desired accuracy with very efficient algorithms.

Convex optimization is exceptionally useful in problems related to quantum mechanical systems, because many objects in quantum mechanics form convex sets, e.g., measurement outcome probabilities, density operators, POVMs, and Choi matrices.

A.2.3 The semidefinite programming problem

The semidefinite programming (SDP) problem is a widely used subclass of convex optimization problems, which generalizes many important problems like linear or quadratic programming. Its standard form is the following:

$$\begin{gathered} \max \operatorname{Tr}(A_0 X) \\ \text{so that } \operatorname{Tr}(A_k X) = b_k, \quad k = 1, \ldots, n \\ X \geq 0 \end{gathered}$$

Here the matrices A_i and X are Hermitian matrices. Thus we seek to minimize a linear objective function subject to linear and matrix semidefinitness constraints (linear matrix inequalities), which are convex constraints.

SDP problems can be solved very efficiently both from theoretical and practical point of view. The most modern solvers use the so-called interior point methods which have the following favourable properties:

- Practical efficiency: Performs similar to other methods on smaller problems, but performs substantially faster on bigger problems.
- Theoretical efficiency: For a fixed accuracy, the resource requirements grow only polinomially with the problem size.

- Exploiting problem structure: The used interior point methods are based on the solution of least squares problems, which can take the problem structure into account.

There are also many solver programs available, a few of them – the ones in particular that are free and can be used with MATLAB – are SeDuMi, SDPT3, MOSEK, and SDPA. A short description and a comprehensive benchmark of these and other solvers can be found in [46].

A.3 Group theory

This section follows the appendix of [5], concentrating on the most needed concepts.

A.3.1 General concepts

Definition A.4. *A group $(G, \cdot)$ is a non-empty set G with a binary operation "$\cdot$" having the following properties:*

- *Closure: $g_1 \cdot g_2 \in G$ for all $g_1, g_2 \in G$,*
- *Associativity: $(g_1 \cdot g_2) \cdot g_3 = g_1 \cdot (g_2 \cdot g_3)$ for all $g_1, g_2, g_3 \in G$,*
- *Identity: there exists $e \in G$ such that $\forall g \in G$, $g \cdot e = e \cdot g = g$,*
- *Inverse: for all $g \in G$, there exists $g^{-1} \in G$ such that $g \cdot g^{-1} = g^{-1} \cdot g = e$.*

We often leave out the operation $\cdot$ in $g_1 \cdot g_2$ and write simply $g_1 g_2$.

A group G is finite if the number of elements, i.e., the *order* of G denoted by $|G|$ is finite. A group G is *Abelian* if the operation $\cdot$ is commutative, i.e, $g_1 g_2 = g_2 g_1$ for all $g \in G$.

A *subgroup* H of G is a subset of G which forms a group under the same operation $\cdot$ as G. In notation we write this as $H < G$. If $g_1, g_2 \in G$ then the *conjugate* of g_2 with respect to g_1 is $g_1 g_2 g_1^{-1} \in G$. Now we can present two important notions.

Definition A.5. *The centralizer $Z(S)$ of a subset $S \subset G$ (not necessarily a subgroup) is the set $\{g \in G \,|\, gsg^{-1} = s, \forall s \in S\}$.*

Definition A.6. *The normalizer $N(S)$ of a subset $S \subset G$ (not necessarily a subgroup) is the set $\{g \in G \,|\, gsg^{-1} \in S, \forall s \in S\}$.*

The study of a group can be greatly simplified by the use of a special subset of the group as a compact description:

Definition A.7. *The elements $g_1, \ldots, g_l$ in a group G are said to be the generators of G if every element of G can be written as a product of (possibly repeated) elements from the set $\{g_1, \ldots, g_l\}$. In notation we write $G = \langle g_1, \ldots, g_l \rangle$.*

It can be shown that a group G can always be generated with a set of at most $\log_2(|G|)$ independent generators.

The following definition is of essential importance in the thesis:

Definition A.8. *For a subgroup $H < G$, the left coset of H in G determined by $g \in G$ is the set $gH \coloneqq \{gh \,|\, h \in H\}$. The right coset Hg is defined similarly.*

Elements of a particular coset gH are known as *coset representatives* of that coset. Cosets gH define an equivalence relation $\sim$ on G given by $g_1 \sim g_2$ if and only if $g_1 h = g_2$ for some $h \in H$. It follows that any two left cosets gH in G are either identical or disjoint. In other words the set of left cosets G/H form a partition of G.

An example of groups is the group of $n \times n$ unitary matrices $U(n)$ with the matrix multiplication as binary operation. Another example important in the thesis is the Pauli group on n qubits $\mathcal{P}_n < U(2^n)$. It consists of all n-fold tensor products of the Pauli matrices defined in (2.2) with a possible ± 1 or $\pm \mathrm{i}$ factor. Thus it is generated as $\mathcal{P}_n = \langle \bigcup_{i=1}^{n} \{X_i, Z_i\}, \mathrm{i}\mathbb{1} \rangle$, that is, all Pauli operators X_i, Y_i, and Z_i acting only on the i^{th} qubit together with the possible ± 1 or $\pm \mathrm{i}$ factor. The order of the Pauli group is $|\mathcal{P}_n| = 2^{2n+2}$. For example, the Pauli group for one qubit is $\mathcal{P}_1 = \langle X, Z, \mathrm{i}\mathbb{1} \rangle$. The most important properties of $\mathcal{P}_n$ are the following:

1. All $g \in \mathcal{P}_n$ are either Hermitian or antihermitian.
2. Any two elements $g, h \in \mathcal{P}_n$ either commute or anticommute, i.e., $gh = \pm hg$. Note that this property implies that for any subset $S \subset \mathcal{P}_n$ for which $-\mathbb{1} \notin S$, the centralizer $Z(S)$ and the normalizer $N(S)$ are the same.

In many cases group elements $g \in G$ can be tought of as transformations on some other set $\mathcal{V}$. In this context we can speak about stabilizers.

Definition A.9. *The group element $g \in G$ fixes (stabilizes) $x \in \mathcal{V}$ if $gx = x$. For any $x \in \mathcal{V}$ the stabilizer subgroup $S < G$ of x is the set of all $g \in G$ that fix x, i.e., $S = \{g \in G \,|\, gx = x\}$.*

In particular, if $\mathcal{V}$ is a vector space then it is easy to see that the elements of G stabilize their whole $+1$ eigenspace. The applications of this property in quantum information processing is discussed in the followig using the Pauli group $\mathcal{P}_n$.

A.3.2 The stabilizer formalism

The stabilizer formalism is an advantageous group theoretic approach to a wide class of actions in quantum mechanics with main application in quantum error correction (see in appendix A.5.2). The presentation follows [5].

Stabilizers of quantum states

Let S be a subgroup of the Pauli group $\mathcal{P}_n$ on n qubits. Then the space $\mathcal{V}_S$ is the space of all n-qubit states fixed by the elements of S, i.e., the space obtained as the intersection of $+1$ eigenspaces of all $g \in S$. In other words $\mathcal{V}_S$ is the *vector space stabilized by the stabilizer* S: $g|\psi\rangle = |\psi\rangle$, $\forall |\psi\rangle \in \mathcal{V}_S$, $\forall g \in S$.

A big advantage of the stabilizer formalism comes from the possibility of describing a group by its generators. A vector is stabilized by S if and only if it is stabilized by the generators of S. The fact that any group G has at most $\log_2(|G|)$ generators allows a very compact representation and easy handling of stabilizers. For example an n-qubit state is stabilized by a subgroup of $\mathcal{P}_n$ having n Pauli group generators and a generator is a product of the at most $2n+2$ number of original Pauli group generators. This means that all states of $\mathcal{V}_S$ can be described by at most $\mathcal{O}(n^2)$ amount of information.

In practice we seek the stabilizer of nontrivial vector spaces. It can be seen that the vector space $\mathcal{V}_S$ is nontrivial if and only if all element of S commute and $-\mathbb{1} \notin S$. Thus in the following we always assume independent generators having these properties. For nontrivial $\mathcal{V}_S$ the following statements hold:

1. If the subgroup $S < \mathcal{P}_n$ has $n-k$ generators then $\mathcal{V}_S$ has 2^k dimensions.
2. For the stabilizer generators of every $S = \langle g_1, \ldots, g_l \rangle$ there exist $g \in \mathcal{P}_n$ such that $g g_i g^\dagger = -g_i$ for some i, but $g g_j g^\dagger = g_j$, $\forall j \neq i$.

Unitary transformations using stabilizers

The stabilizer formalism can also be used to describe unitary dynamics in vector spaces. Suppose we act with a unitary U on the vector space $\mathcal{V}_S$ stabilized by $S < \mathcal{P}_n$. Let $|\psi\rangle \in \mathcal{V}_S$. Then for all $g \in S$

$$U|\psi\rangle = Ug|\psi\rangle = UgU^*U|\psi\rangle \ , \tag{A.4}$$

thus the state $U|\psi\rangle$ is stabilized by UgU^*. It follows that the stabilizer of $U\mathcal{V}_S$ is the group $USU^* \equiv \{UgU^* \,|\, g \in S\}$. Furthermore, if S is generated by $g_1, \ldots, g_l$ then USU^* is generated by $Ug_1U^*, \ldots, Ug_lU^*$. Thus it is enough to compute the effect of U on the generators.

Note that using these principles, the stabilizer formalism allows an efficient classical simulation of many quantum behaviour. However, it can not efficiently simulate all of quantum mechanics. The exact limitations are stated by the famous Gottesman–Knill theorem [5]. In general those unitaries (gates) U can be efficiently simulated, for which $U\mathcal{P}_nU^* = \mathcal{P}_n$, i.e., elements of the normalizer $N(\mathcal{P}_n)$. Fortunately, encoding, decoding, error-detection and recovery for stabilizer based quantum error correcting codes (see Chapter 3) are all such *normalizer gates*.

Quantum measurement using stabilizers

Using the stabilizer formalism, measurements made in the computational basis can also be efficiently described. Apart from a possible $-1, \pm\mathrm{i}$ factor, $g \in \mathcal{P}_n$ is self-adjoint and can be seen as an observable. Suppose then g is an observable and the system is in state $|\psi\rangle$ with stabilizer $S = \langle g_1, \ldots, g_l \rangle$. Then there are two possibilities:

1. g commutes with all generators of S.

2. g anticommutes with g_1 and commutes with all other g_i, $i > 1$. This is not a loss of generality, because if g anticommutes also with g_i then it commutes with $g_1 g_i$, thus we can just swap g_i with $g_1 g_i$.

In the first case $g_j g|\psi\rangle = g g_j|\psi\rangle = g|\psi\rangle$ implies $g|\psi\rangle \in \mathcal{V}_S$, i.e., $g|\psi\rangle$ is proportional to $|\psi\rangle$, because $\mathcal{V}_S$ one dimenisonal. The hermiticity of g implies $g^2 = \mathbb{1}$, thus $g|\psi\rangle = \pm|\psi\rangle$. This means that either g or $-g$ is element of S. If $g \in S$ then $g|\psi\rangle = |\psi\rangle$ and the measurement result is 1 with 1 probability. If in turn $-g \in S$ then $g|\psi\rangle = -|\psi\rangle$ and the measurement result is -1 with 1 probability. Furthermore, the measurement does not ruin the state, because the stabilizer does not change.

In the second case, as the eigenvalues of g are ± 1 the spectral decomposition of g is $g = P_+ - P_-$, where $P_+ = \frac{\mathbb{1}+g}{2}$ is the projection onto the $+1$ eigenspace and $P_- = \frac{\mathbb{1}-g}{2}$ is the projection onto the -1 eigenspace. Then using the measurement postulate together with $g g_1 = -g_1 g$ the p_+ and p_- probabilities of the $+1$ and -1 results:

$$p_+ = \langle\psi|P_+ g_1|\psi\rangle = \langle\psi|g_1 P_-|\psi\rangle = p_- = \frac{1}{2} \ .$$

Then the state after measurement will be

$$|\psi_+\rangle = \sqrt{2}P_+|\psi\rangle, \text{ with stabilizer } \langle g, g_2, \ldots, g_l\rangle \ ,$$
$$|\psi_-\rangle = \sqrt{2}P_-|\psi\rangle, \text{ with stabilizer } \langle -g, g_2, \ldots, g_l\rangle \ .$$

A.4 Matrix algebras

In this section a few notions on matrix algebras is presented, based on [47, 14].

Let $M_d(\mathbb{C})$ denote the full matrix algebra, i.e., the algebra of all $d \times d$ complex matrices.

Definition A.10 (Center of an algebra)**.** *The center of an algebra $\mathcal{A}$ is the set $Z_\mathcal{A} \coloneqq \{x \in \mathcal{A} \,|\, xa = ax, \forall a \in \mathcal{A}\}$, i.e., the set of elements in $\mathcal{A}$ that commute with all other elements of $\mathcal{A}$.*

The center of $M_d(\mathbb{C})$ is then $Z_{M_d(\mathbb{C})} = \{c\mathbb{1} \,|\, c \in \mathbb{C}\}$. This leads to the following important notions.

Definition A.11 (Complementary subalgebras)**.** *Two subalgebras $\mathcal{A}_i$ and $\mathcal{A}_j$ of $M_d(\mathbb{C})$ are called complementary if their traceless subspaces $\mathcal{A}_i \cap Z^\perp_{M_d(\mathbb{C})}$ and $\mathcal{A}_j \cap Z^\perp_{M_d(\mathbb{C})}$ are orthogonal with respect to the Hilbert–Schmidt inner product.*[32]*.*

Definition A.12 (Mutually unbiased bases)**.** *Two orthonormal bases $\{|\psi_{1,j}\rangle\}$ and $\{|\psi_{2,j}\rangle\}$ are said to be mutually unbiased if for all $k, l = 1, \ldots, d$ they satisfy $|\langle\psi_{1,k}|\psi_{2,l}\rangle|^2 = \frac{1}{d}$.*

[32] This condition is equivalent to $\mathrm{Tr}(A_i) = \mathrm{Tr}(A_j) = 0 \Rightarrow \mathrm{Tr}(A_i A_j) = 0$, $A_i \in \mathcal{A}_i$, $A_j \in \mathcal{A}_j$, $i \neq j$

In d dimensions, the maximal number of pairwise mutually unbiased bases (MUB) is $d+1$. Currently, this maximal number of MUB has been found only in dimensions which are prime power [48].

A general theorem on the number of MUB in d dimensions can be stated [49]:

Theorem A.5. *There exist u MUB in d dimensions if and only if there are u classes $\mathcal{U}_1, \ldots, \mathcal{U}_u$ each consisting of d commuting unitaries such that $\mathcal{U}_i \cap \mathcal{U}_j = \{\mathbb{1}\}$ and all operators in $\bigcup_i \mathcal{U}_i$ are pairwise orthogonal.*

It follows that if $u = d+1$ then $\bigcup_i \mathcal{U}_i$ is a basis of $M_d(\mathbb{C})$.

If the complementary subalgebras $\mathcal{A}_i$ are maximal Abelian, i.e., commutative and are not properly contained in any other commutative subalgebra of $M_d(\mathbb{C})$ then $\mathcal{A}_i = \mathrm{span}(\mathcal{U}_i)$, and thus they can be related to mutually unbiased bases.

Theorem A.6. *Two orthonormal bases $\{|\psi_{1,j}\rangle\}$ and $\{|\psi_{2,j}\rangle\}$ are mutually unbiased if and only if the maximal Abelian subalgebras $\mathcal{A}_1$ and $\mathcal{A}_2$ containing operators diagonal in the bases $\{|\psi_{1,j}\rangle\}$ and $\{|\psi_{2,j}\rangle\}$ are complementary.*

In the following, a unitary generalization of Pauli matrices is given for d dimensions. Let $|\varphi_0\rangle, \ldots, |\varphi_{d-1}\rangle$ be a d dimensional basis. Let X and Z be unitary operators such that

$$X|\varphi_k\rangle = |\varphi_{(k+1) \bmod d}\rangle \ ,$$
$$Z|\varphi_k\rangle = \omega^k |\varphi_k\rangle \ ,$$

with $\omega = \mathrm{e}^{\mathrm{i}\frac{2\pi}{d}}$. Then $XZ = \omega ZX$ and $X^d = Z^d = \mathbb{1}$. Using these, the general Pauli matrices are the traceless unitaries $U_{j,k} \coloneqq X^j Z^k$ $(j,k = 0, \ldots, d-1)$. They are pairwise orthogonal with respect to the Hilbert–Schmidt inner product, and satisfy the commutation relation $U_{j,k}U_{l,m} = \omega^{kl-jm} U_{l,m}U_{j,k}$. Together with $\mathbb{1}$, these matrices form an orthonormal basis in $M_d(\mathbb{C})$.

If d is prime, then the set of $U_{j,k}$ can be partitioned into cyclic groups, each of which can be used as $\mathcal{U}_i$ in **Theorem A.5**. Then by taking a unitary other than $\mathbb{1}$ from each group, for example Z and XZ^k for $k = 0, \ldots, d-1$, their eigenvectors give us a set of $d+1$ MUB.

When d is not prime, then we can still form an orthogonal basis of unitaries from tensor products of generalized Pauli matrices. However, only when d is prime power can we make $d+1$ (not necessarily cyclic) groups of unitaries in a similar way. Let $d = p^m$ for a prime p. Then the elements of the operator basis are taken from the Pauli group $\mathcal{P}_m^p$:

$$\omega^j U_{[\mathbf{x},\mathbf{z}]} = \omega^j X(\mathbf{x}) Z(\mathbf{z}) = \omega^j \bigotimes_{i=1}^{m} X^{x_i} Z^{z_i}, \quad 0 \le j, x_i, z_i \le p-1$$

We are only interested in operators with $j = 0$. These are uniquely determined by the vectors $\mathbf{x}$ and $\mathbf{z}$ over the finite field $\mathbb{F}_p$. Their concatenation $[\mathbf{x}, \mathbf{z}]$ is then an element of $\mathbb{F}_p^{2m}$. The commutation relation of such

operators is $U_{[\mathbf{x},\mathbf{z}]}U_{[\mathbf{x}',\mathbf{z}']} = \omega^{\langle[\mathbf{x},\mathbf{z}],[\mathbf{x}',\mathbf{z}']\rangle_{\mathrm{spl}}}U_{[\mathbf{x}',\mathbf{z}']}U_{[\mathbf{x},\mathbf{z}]}$, where $\langle[\mathbf{x},\mathbf{z}],[\mathbf{x}',\mathbf{z}']\rangle_{\mathrm{spl}} = \langle\mathbf{x},\mathbf{z}'\rangle - \langle\mathbf{x}',\mathbf{z}\rangle = \sum_{j=1}^{m} x_j z'_j - x'_j z_j$ is the symplectic inner product.

A method for constructing such an operator basis satisfying the conditions of **Theorem A.5** can be found in [49].

A.5 Some more notions and results from QEC

This section provides some examples to quantum channels and error correcting codes. Moreover, some additional material is given on the subject of Chapter 3.

A.5.1 Example quantum channels

In the following, four quantum noise models are presented as examples.

The depolarizing channel

Under the effect of the depolarizing channel an input qubit will be depolarized with probability p, i.e., it transforms into the completely mixed state $\rho = \frac{1}{2}\mathbb{1}$, and with probability $1-p$ it remains intact. The parameter p of the channel thus indicates the strength of depolarization.

The Kraus operator elements of this channel are the following:

$$E_0 = \sqrt{1-\frac{3}{4}p}\mathbb{1},\ E_1 = \frac{1}{2}\sqrt{p}X,\ E_2 = \frac{1}{2}\sqrt{p}Y,\ E_3 = \frac{1}{2}\sqrt{p}Z \tag{A.5}$$

Because each operator element is proportional to a Pauli matrix or to the identity, this channel belongs to the class of Pauli channels. Another common set of Kraus operators for the depolarizing channel is obtained using the reparametrization $q = \frac{3}{4}p$:

$$F_0 = \sqrt{1-q}\mathbb{1},\ F_1 = \sqrt{\frac{q}{3}}X,\ F_2 = \sqrt{\frac{q}{3}}Y,\ F_3 = \sqrt{\frac{q}{3}}Z$$

The affine map T of this qubit channel on the Bloch vector θ can also be derived, it is $T(\theta) = (1-p)\mathbb{1}\theta$. This shows that the channel contracts θ equally in any direction.

The phase damping channel

The phase damping channel describes an exclusively quantum mechanical noise process, the loss of quantum information without loss of energy.[33]

The Kraus operator elements of this channel are

$$E_0 = \sqrt{1-p}\mathbb{1},\ E_1 = \sqrt{p}\frac{\mathbb{1}+Z}{2},\ E_2 = \sqrt{p}\frac{\mathbb{1}-Z}{2}\ .$$

[33]Such process is for example the random scattering of a photon in an optical fiber, or the disturbation of electron states originating from interactions with distant charges.

This is also a Pauli channel. To make this apparent from the operator elements, we must construct another Kraus element set:

$$F_0 = \sqrt{1-q}\mathbb{1}, \quad F_1 = \sqrt{q}Z \tag{A.6}$$

Using the Choi matrix it can be derived that $q = \frac{p}{2}$. Substituting this, we can find the unitary U relating the two operator element sets:

$$U = \begin{bmatrix} \sqrt{\frac{2(1-p)}{2-p}} & \sqrt{\frac{p}{2(2-p)}} & \sqrt{\frac{p}{2(2-p)}} \\ 0 & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ -\sqrt{\frac{p}{2-p}} & \sqrt{\frac{1-p}{2-p}} & \sqrt{\frac{1-p}{2-p}} \end{bmatrix}$$

Here we appended the $F_2 = 0$ operator to the set $\{F_0, F_1\}$ to get a 3×3 matrix. Thus this operator element set also serves as an example on the unitary freedom of Kraus representation.

The affine map T of this qubit channel on the Bloch vector θ is

$$T(\theta) = \begin{bmatrix} 1-p & 0 & 0 \\ 0 & 1-p & 0 \\ 0 & 0 & 1 \end{bmatrix} \theta \ .$$

The amplitude damping channel

This channel is physically well motivated. It describes the spontaneous emission of an atom with excited state $|1\rangle$ and ground state $|0\rangle$. Thus the effect of the channel is that with probability g the state transforms into the state $|0\rangle$. The Kraus operator elements of this channel can be written as:

$$E_0 = \begin{bmatrix} 1 & 0 \\ 0 & \sqrt{1-g} \end{bmatrix}, \quad E_1 = \begin{bmatrix} 0 & \sqrt{g} \\ 0 & 0 \end{bmatrix}$$

In contrast to the previous two example channels, this channel is probably the most well-known non-Pauli channel. A generalized version of this channel can also be defined:

$$E_0 = \sqrt{p}\begin{bmatrix} 1 & 0 \\ 0 & \sqrt{1-g} \end{bmatrix}, \qquad E_1 = \sqrt{p}\begin{bmatrix} 0 & \sqrt{g} \\ 0 & 0 \end{bmatrix},$$
$$E_2 = \sqrt{1-p}\begin{bmatrix} \sqrt{1-g} & 0 \\ 0 & 1 \end{bmatrix}, \qquad E_3 = \sqrt{1-p}\begin{bmatrix} 0 & 0 \\ \sqrt{g} & 0 \end{bmatrix}$$

Here p parameterizes the output state for $g = 1$, it takes values from the interval $[0, 1]$. This is $|0\rangle$ when $p = 1$, and $|1\rangle$ when $p = 0$ in the computational basis, thus it can be seen that $p = 1$ gives back the first simpler definition.

The affine map T of this qubit channel on the Bloch vector θ is

$$T(\theta) = \begin{bmatrix} \sqrt{1-g} & 0 & 0 \\ 0 & \sqrt{1-g} & 0 \\ 0 & 0 & 1-g \end{bmatrix} \theta + \begin{bmatrix} 0 \\ 0 \\ 2pg-g \end{bmatrix} \ .$$

The pure state rotation channel

This channel is physically much less motivated than the other examples. It serves for us only as a second example for a non-Pauli channel from [24] besides the amplitude damping, broadening the scope of our results. The single-qubit operator elements are the following:

$$E_0 = \alpha \begin{pmatrix} \frac{\cos\left(\frac{\theta-\phi}{2}\right)}{\cos\left(\frac{\theta}{2}\right)} & 0 \\ 0 & \frac{\sin\left(\frac{\theta-\phi}{2}\right)}{\sin\left(\frac{\theta}{2}\right)} \end{pmatrix} ,$$

and

$$E_\pm = \beta \begin{pmatrix} \cos\left(\frac{\theta-\phi}{2}\right)\sin\left(\frac{\theta}{2}\right) & \pm\cos\left(\frac{\theta-\phi}{2}\right)\cos\left(\frac{\theta}{2}\right) \\ \pm\sin\left(\frac{\theta-\phi}{2}\right)\sin\left(\frac{\theta}{2}\right) & \sin\left(\frac{\theta-\phi}{2}\right)\cos\left(\frac{\theta}{2}\right) \end{pmatrix} .$$

Here α and β are determined by the trace preserving condition. In this channel, the parameter ϕ characterizes the strength of the noise.

A.5.2 Stabilizer codes

Continuing appendix A.3.2, we use the stabilizer formalism to compactly and efficiently describe a wide class of quantum codes, the *stabilizer codes* or additive quantum codes. These codes are analogous to classical linear codes [18]. Stabilizer codes are exceptionally useful against errors that are elements of the Pauli group defined in appendix A.3.1. Channels defined using Pauli group operators are a special case of uncorrelated noise. We will assume such errors in the rest of this section.

Code construction

By definition the $[[c, m]]$-stabilizer code is the $+1$ eigenspace $\mathcal{C}_S \subset \mathcal{H}_{\text{cs}}$ of the Abelian subgroup $S < \mathcal{P}_c$ given by $c - m$ independent generators $g_1, \ldots, g_{c-m}$ with $-\mathbb{1} \notin S$. These generators are of course not unique. The codewords in $\mathcal{C}_S$ can be described and transformed into each other by the *logical Pauli operators* $\bar{Z}_i$ and $\bar{X}_i$. These act on the i^{th} logical qubit the same way as Z or X act on a physical qubit (see in section 2.1.2). These must be elements of the normalizer $N(S)$ but must be outside S itself so that they act (nontrivially) within the code subspace $\mathcal{C}_S$.[34] Then the codeword $|b_1, \ldots, b_m\rangle_{\text{L}}$ is stabilized by $(-1)^{b_1}\bar{Z}_1, \ldots, (-1)^{b_m}\bar{Z}_m$ inside the code $\mathcal{C}_S$. For example if $m = 1$ then the stabilizer of the logical basis vector $|0\rangle_{\text{L}}$ is $\bar{Z}$ and the stabilizer of $|1\rangle_{\text{L}}$ is $-\bar{Z}$ in $\mathcal{C}_S$.

Error correction conditions

In case of a Pauli error E using a stabilizer code, the syndrome measurement is done by measuring each generator g_i. The error either commutes or

[34] This means that the logical Z and X operators must form an independent and commuting set together with the generators. The standard way of selecting them can be found in [18].

anticommutes with the generators g_i (i.e., $Eg_iE^* = (-1)^{\mu_i} g_i$) thus the result $\mu_i = 0$ or $\mu_i = 1$ tells us that the corrupted codeword is either in the $+1$ or in the -1 eigenspace of g_i. Therefore, the syndrome μ is obtained in the form $\mu = 2^{c-m-1}\mu_{c-m} + \cdots + 2^0\mu_1$, identifying the $\mathcal{S}_\mu$ syndrome subspace in which the corrupted codeword $E|\psi_{\text{cs}}\rangle$ lies (specially $\mu = 0$ for $E \in S$, i.e., $E|\psi_{\text{cs}}\rangle \in \mathcal{S}_0$).

However, in the case when $E \in Z(S)$, i.e., E commutes with all generators but $E \notin S$, the error is undetectable thus uncorrectable. We know that for the Pauli group $\mathcal{P}_c$ the centralizer $Z(S)$ is the same as the normalizer $N(S)$, thus we arrive at the error correcting conditions of stabilizer codes:

Theorem A.7 (EC conditions for stabilizer codes and Pauli errors)**.** *Let S be the stabilizer of the code $\mathcal{C}_S$. The set $\{E_i\} \subset \mathcal{P}_c$ is correctable perfectly on $\mathcal{C}_S$ with some $\mathcal{R}$ recovery operation if and only if $E_i^* E_j \notin N(S) - S$ for all j, k.*

To recover an error E taken from a correctable set $\{E_i\}$ after syndrome measurement, we only need to select an $E_\mu \in \{E_i\}$ error operator known to have the obtained syndrome μ, and apply E_μ^* on $E|\psi_{\text{cs}}\rangle$. Because $E_\mu^* E \in S$ (they are correctable errors), we get back $|\psi_{\text{cs}}\rangle$.

Comparison to non-stabilizer codes

We can observe the following differences between general quantum codes and stabilizer codes. For a general code $\mathcal{C}$ the set of all errors $\mathcal{V}_{\mathcal{R},\mathcal{C}}$ correctable with some $\mathcal{R}$ is characterized through the Knill–Laflamme conditions (3.1). An exact sufficient $\mathcal{R}$ is determined by the set $\{E_i\}$ of errors selected for correction through the determination of both the syndrome subspace structure – the W_μ operators – and the unitary rotations A_μ corrected inside each syndrome subspace, as we described in section 3.1.1.

In contrast, for a stabilizer code $\mathcal{C}_S$, the generators g_i determine not only the code, but also fix the syndrome subspaces through the intersections of the ± 1 eigenspaces of each generator. Thus we have freedom only in the selection of the exact A_μ unitaries – or A_{p_μ} for Pauli errors – which we would like to correct as described in section 3.1.2.

A.5.3 The quantum Hamming-bound

Recall from section 3.1.1 that the Knill–Laflamme conditions can tell us whether a code corrects a set of error operators or not, but we get no information on the effectiveness of the code. In fact, there might be better codes, tipically ones having the same set of correctable errors while requiring only a smaller amount of physical qubits.

To calculate at least how many physical qubits are needed, i.e., how many dimensions the code space $\mathcal{H}_{\text{cs}}$ must have to correct the desired errors, the *quantum Hamming bound* can be used. To correct N independent errors, i.e., to have N different 2^m dimensional syndrome subspaces we must have $\dim(\mathcal{H}_{\text{cs}}) = 2^c \geq 2^m N$. On each qubit three independent errors – X, Y and Z – can happen, thus the number of error operators corrupting exactly l qubits

is $3^l\binom{c}{l}$. If we would like to correct all errors affecting at most t qubits then we have $N = \sum_{l=0}^{t} 3^l\binom{c}{l}$. To summarize, the quantum Hamming bound is

$$2^m \sum_{l=0}^{t} 3^l \binom{c}{l} \leq 2^c \ . \tag{A.7}$$

For instance, let us encode one logical qubit with the requirement of correcting all single-qubit errors on the codeword, i.e., $m = 1$ and $t = 1$. What is the minimum number c of physical qubits for which such a coder exist? From (A.7) we get the solution $c \geq 5$ and in the $c = 5$ case (A.7) is saturated, meaning that such a code with 5 physical qubits – a $[[5, 1]]$ code – is a *perfect code*. The problem of finding a code with $m = 1$ and $t = 1$ was first solved by the pioneering work of [50] using a $[[9, 1]]$ code. The first $[[5, 1]]$ code – commonly called *five-qubit code* – was published not much later by [51]. Such code was used throughout the thesis, thus let us take a more detailed look at it in section A.5.4.

A.5.4 The five-qubit code

Many commonly known quantum codes are stabilizer codes. This is true also for five-qubit codes. Here we present three examples on this code. The first two are both standard $[[5, 1]]$ codes, the third is special and will be discussed later. Their generators together with the logical $\bar{X}$ and $\bar{Z}$ operators can be seen in Table A.1. The code in Table A.1(b) is the code of [51] which stabilizes the subspace given by the following codeword basis vectors

$$\begin{aligned}
|0\rangle_{\mathrm{L}} = \frac{1}{\sqrt{2^8}}\big(&|00000\rangle - |01111\rangle - |10011\rangle + |11100\rangle \\
&+ |00110\rangle + |01001\rangle + |10101\rangle + |11010\rangle\big) \ , \\
|1\rangle_{\mathrm{L}} = \frac{1}{\sqrt{2^8}}\big(&|11111\rangle - |10000\rangle + |01100\rangle - |00011\rangle \\
&+ |11001\rangle + |10110\rangle - |01010\rangle - |00101\rangle\big) \ .
\end{aligned}$$

Table A.1(a) contains the code presented in [5, 52]. This code was used throughout the thesis. Its codeword basis is given by the logical qubits

$$\begin{aligned}
|0\rangle_{\mathrm{L}} = \frac{1}{4}\big(&|00000\rangle + |10010\rangle + |01001\rangle + |10100\rangle \\
&+ |01010\rangle - |11011\rangle - |00110\rangle - |11000\rangle \\
&- |11101\rangle - |00011\rangle - |11110\rangle - |01111\rangle \\
&- |10001\rangle - |01100\rangle - |10111\rangle + |00101\rangle\big) \ , \\
|1\rangle_{\mathrm{L}} = \frac{1}{4}\big(&|11111\rangle + |01101\rangle + |10110\rangle + |01011\rangle \\
&+ |10101\rangle - |00100\rangle - |11001\rangle - |00111\rangle \\
&- |00010\rangle - |11100\rangle - |00001\rangle - |10000\rangle \\
&- |01110\rangle - |10011\rangle - |01000\rangle + |11010\rangle\big) \ ,
\end{aligned}$$

Name	Operator
g_1	$ZXXZ\mathbb{1}$
g_2	$\mathbb{1}ZXXZ$
g_3	$Z\mathbb{1}ZXX$
g_4	$XZ\mathbb{1}ZX$
$\bar{Z}$	$ZZZZZ$
$\bar{X}$	$XXXXX$

(a)

Name	Operator
g_1	$X\mathbb{1}XZX$
g_2	$ZXZ\mathbb{1}X$
g_3	$XYZY\mathbb{1}$
g_4	$XX\mathbb{1}XZ$
$\bar{Z}$	$\mathbb{1}X\mathbb{1}ZX$
$\bar{X}$	$XXYYX$

(b)

Name	Operator
g_1	$X\mathbb{1}\mathbb{1}\mathbb{1}X$
g_2	$\mathbb{1}X\mathbb{1}\mathbb{1}X$
g_3	$\mathbb{1}\mathbb{1}X\mathbb{1}X$
g_4	$\mathbb{1}\mathbb{1}\mathbb{1}XX$
$\bar{Z}$	$ZZZZZ$
$\bar{X}$	$\mathbb{1}\mathbb{1}\mathbb{1}\mathbb{1}X$

(c)

Table A.1. Generators and logical Pauli operators for five-qubit codes.

It is interesting that this code from Table A.1(a) the stabilizer generators are cyclic permutations of each other. This means that this code is a *cyclic code*.

Are there any difference between the error correcting properties of the codes in Tables A.1(a) and A.1(b)? Assuming uncorrelated noise models the answer is clearly yes, however, in the context of standard QEC these differences are of little importance; as the two codes have different syndrome subspaces, they correct different higher weight (two or more qubit) errors beside the standard set of single weight errors.

In contrast for the case of optimal QEC the exact form of the code can have a significant impact on the performance of error correction. This – together with the code in Table A.1(c) – is discussed in more detail in section A.5.5.

A.5.5 Comparison of standard and optimal QEC

We now mention three cases in which the correction operation based on optimization could give significantly better result than the standard strategy. In these cases, one should study carefully which are the most probable errors. When

1. the noise does not act independently on each qubit, i.e., an error operator is not the tensor product of single-qubit operators,
2. the noise is not weak,
3. even though the channel is tensor product and the noise is weak, it can still occur that a single-qubit error has smaller probability than a several-qubit error.

A good example on this third case is the phase damping channel, in which X- and Y-type errors do not occur at all. Let us take the representation of this channel given by the Kraus operators (A.6) substituting $q = \frac{p}{2}$. If we use against this noise the five-qubit code from Table A.1(a) and, according to the standard strategy, we use the 15 syndrome subspaces to correct all the fifteen different single-qubit errors then we get significantly worse channel fidelity than in the optimal case, when we use the 2×5 syndrome subspaces originally designed to correct the single-qubit X and Y errors to correct the two-qubit Z errors instead. Actually, all the two-qubit Z errors can be corrected in this

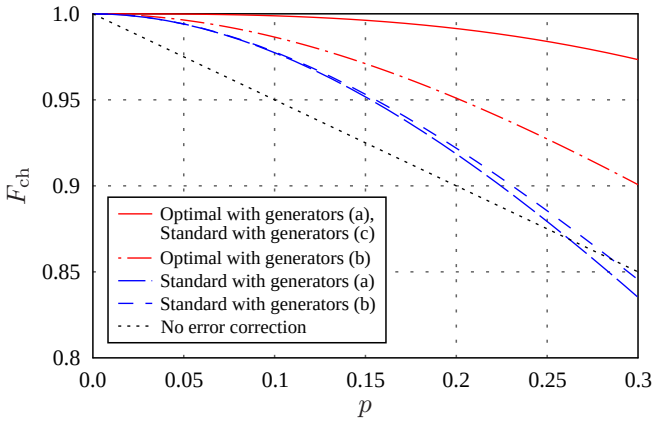


Figure A.1. The channel fidelity for the phase damping channel using the code in Table A.1(a) and (i) the optimal error correction operation: solid curve, third order in p; (ii) the standard QEC operation: long-dashed line, second order in p; using the code in Table A.1(b) and (iii) the optimal error correction operation: dotted-dashed line, second order in p; and (iv) the standard QEC operation: short-dashed line, second order in p; finally, the case of no error correction (dotted line) is also included for comparison.

way by this five-qubit code. The result can be seen in Figure A.1, where the solid line shows the channel fidelity computed by the optimized QEC operation $\mathcal{R}^\star$; this curve is cubic in the noise parameter p in contrast to the long-dashed curve which is only quadratic, showing the channel fidelity computed by the standard QEC operation. We also see that the optimal correction is better on the full domain of the noise parameter $0 \leq p \leq 1$ not only than the standard curves but also than the no error correction curve.

As we have mentioned in section A.5.4, there is a difference between the standard and optimal QEC also from the point of view of the coder. The difference between the two codes for the standard QEC case can be seen from the short- and long-dashed lines in Figure A.1. In the optimal case we used the coder from Table A.1(a) previously, but we can also use the five-qubit code defined by the generators of Table A.1(b) to correct the errors of the phase damping channel. In fact the two codes defined in Table A.1(a) and (b) follow essentially the same standard strategy, i.e., both are $[[5, 1]]$ codes which correct perfectly all one-qubit errors. However, the optimal QEC operations based on these codes are very different. As can be seen from the dotted-dashed line in Figure A.1, using the code from Table A.1(b) the correction of all two-qubit errors is not possible; therefore, the curve is only quadratic!

This observation raises the question whether it is possible to reach the optimal performance against the phase damping channel shown by the solid curve in Figure A.1 using the standard strategy. It turns out that this can be done. The extreme asymmetry of the phase damping channel allows us to construct a code which corrects all two-qubit errors without optimization. According to section A.5.2 we can select the generators of the code in multiple ways. A five-qubit code has 15 syndrome subspaces (plus the code subspace), which is just enough to correct all 5 possible one-qubit Z errors and all 10 possible two-qubit Z errors. Using the method described in [18] the generators

and logical Pauli operators of such a code can be constructed, it can be seen in Table A.1(c). Note that this code construction can also be viewed as an example of channel adapted code optimization, which is a similar to recovery optimization (see [19]).

Contrary to the above-mentioned three cases, the standard QEC strategy is acceptable in most cases, when the channel is not very asymmetric, i.e., the probability of each single-qubit error is about the same. In this case, it is easy to see that the most likely errors will be the single-qubit errors and the traditional QEC correction operation is just designed to correct these; thus, it is clearly optimal. A good example on this is the depolarizing channel, for which the optimal entanglement fidelity just equals with the value obtained from the conventional QEC error correction. This result – considered surprising by the authors of [16] – is obvious in light of the above.

A.6 Numerical simulation tools

Most of the main results of this thesis involved numerical computations, primarily convex optimization. The used software tools were the following:

- All simulations were done in MATLAB environment.
- The semidefinite programming problems were solved using YALMIP modeling language [53] and the SDPT3 solver [54].
- Global optimization was done using MATLAB built-in algorithms.

The simulation of quantum measurements is also a fundamental part of the numerical experiments. The measurement of a quantum system in state ρ using an apparatus represented by the POVM $\mathbf{M} = \{M_\alpha\}$ was simulated by generating a random number from the probability distribution $p(\,.\,)$ of the measurement outcomes $p(\alpha) = \mathrm{Tr}(\rho M_\alpha)$. This number α then determines the corresponding state after measurement.

Appendix B

Supplementary material on channel parameter estimation and experiment design

This appendix shows some additional results to the ones given in Chapters 6 and 7.

B.1 Further examples for Pauli channel estimation

In this section, we provide further experimental results for the case studies in section 6.3 with different channel parameters, which were taken from the interior and the border of the parameter space. These examples also use the configurations defined in sections 6.3.2 and 6.3.2 together with the two and three-level optimal configuration sets.

B.1.1 Qubit Pauli channel

Here are some additional examples to the qubit case studies in Chapters 6 and 7 depicted in Figures B.1-B.16.

The three subfigures on each Figure show the performance indicators defined in section 6.3.1 respectively, in function of the number of complete experiments. The dotted-dashed line corresponds to tomography with unstructured Choi matrix, the dashed line shows results from parameter estimation using Choi matrix structure, and the solid line is obtained using the optimal configuration set. The chosen λ vectors are written in the Figure captions.

B.1.2 Pauli channel for a 3-level quantum system

Here are some additional examples to the qutrit case studies in Chapters 6 and 7 depicted in Figures B.17-B.24. The chosen λ vectors together with other information are written in the Figure captions.

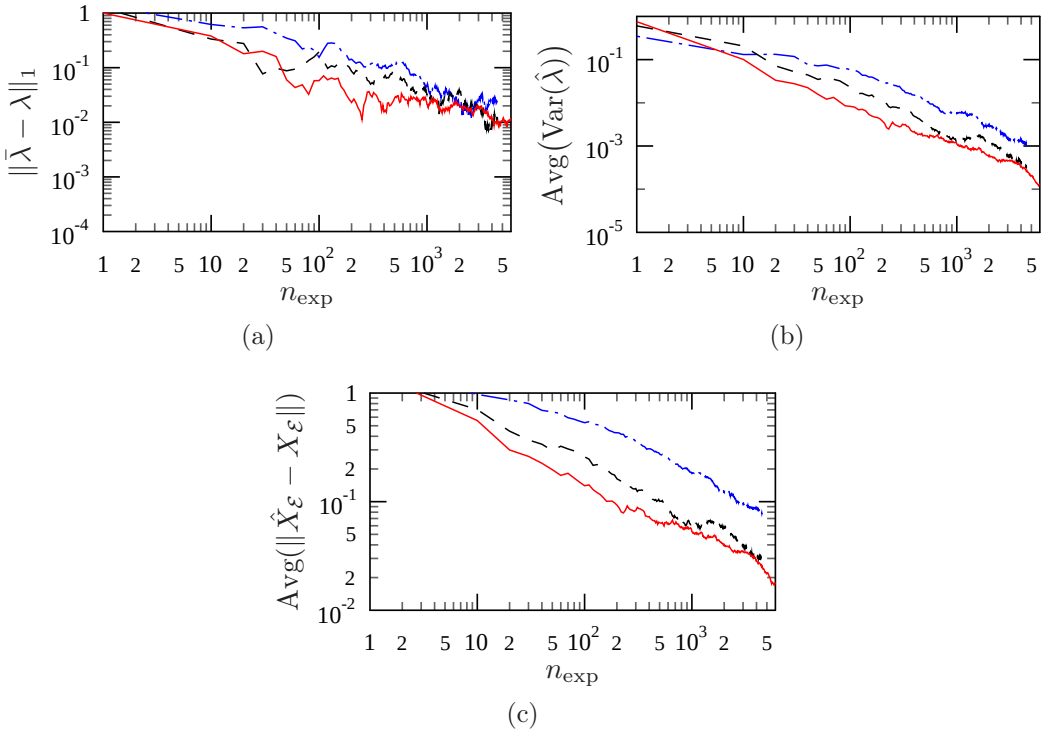


Figure B.1. Minimal tomography with $\lambda = (0.4, -0.2, -0.4)$ taken from the interior of the parameter space.

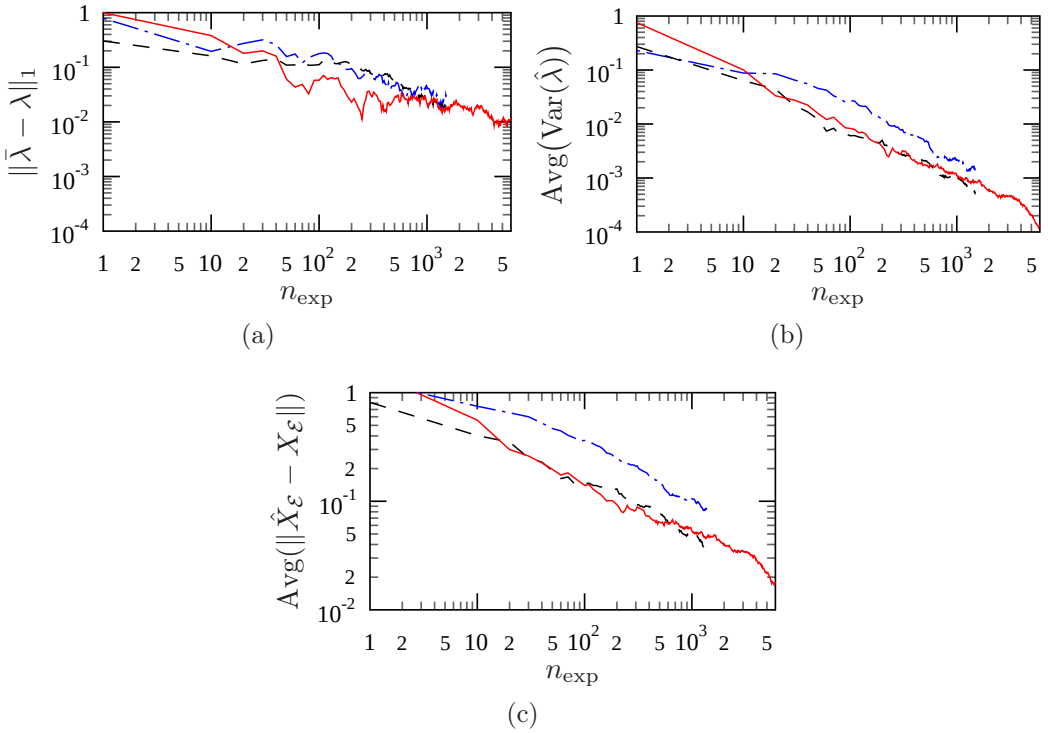


Figure B.2. Standard tomography with $\lambda = (0.4, -0.2, -0.4)$ taken from the interior of the parameter space.

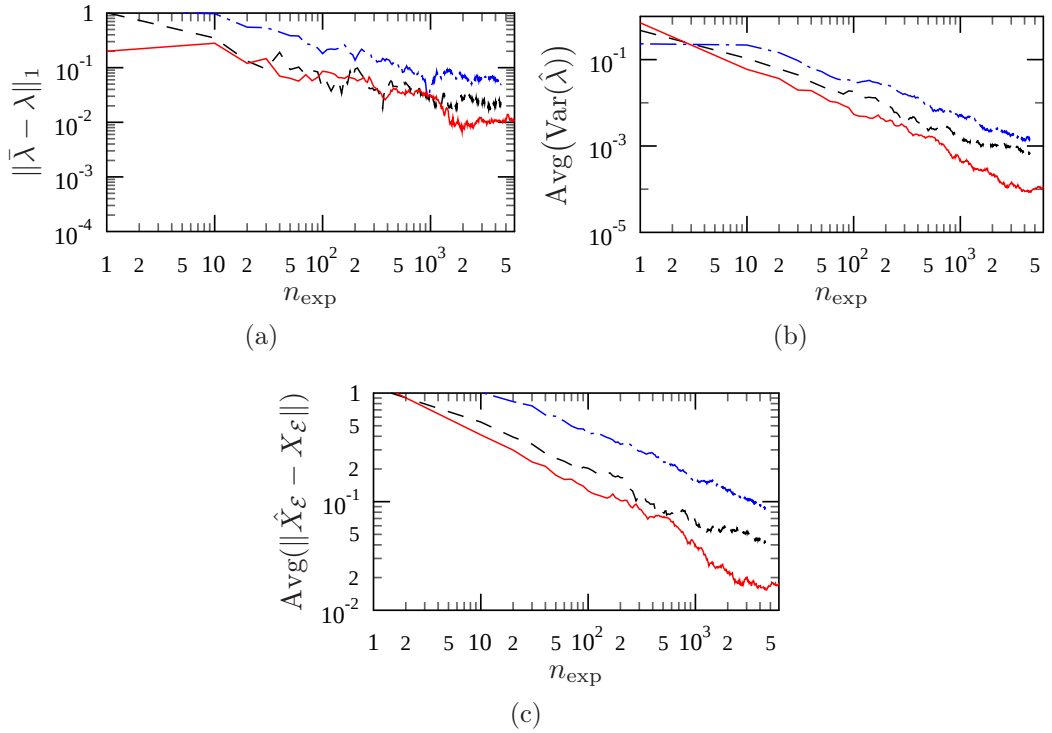


Figure B.3. Minimal tomography with $\lambda = (-0.6, 0.7, -0.5)$ taken from the interior of the parameter space.

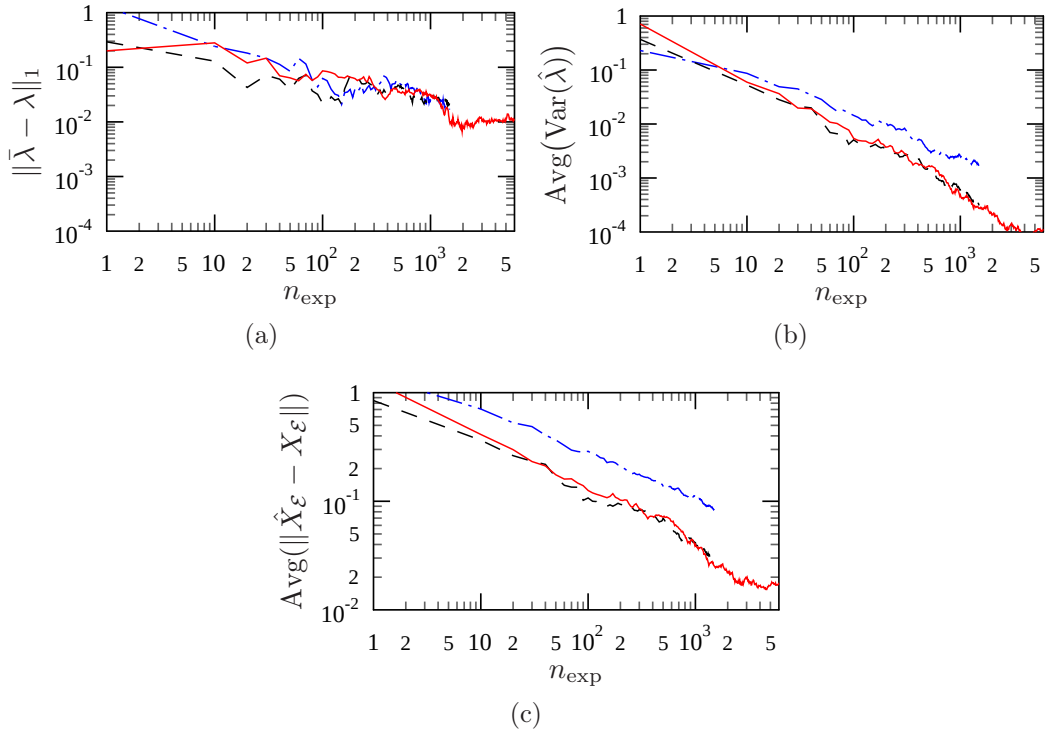


Figure B.4. Standard tomography with $\lambda = (-0.6, 0.7, -0.5)$ taken from the interior of the parameter space.

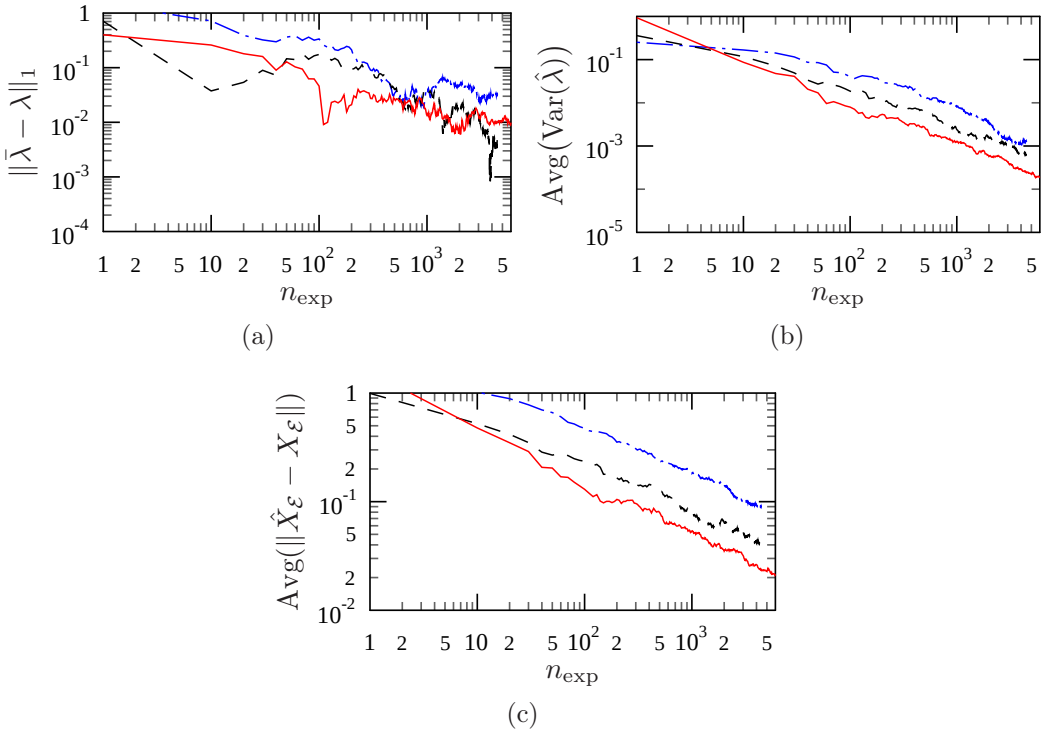


Figure B.5. Minimal tomography with $\lambda = (0.6, 0.5, 0.3)$ taken from the interior of the parameter space.

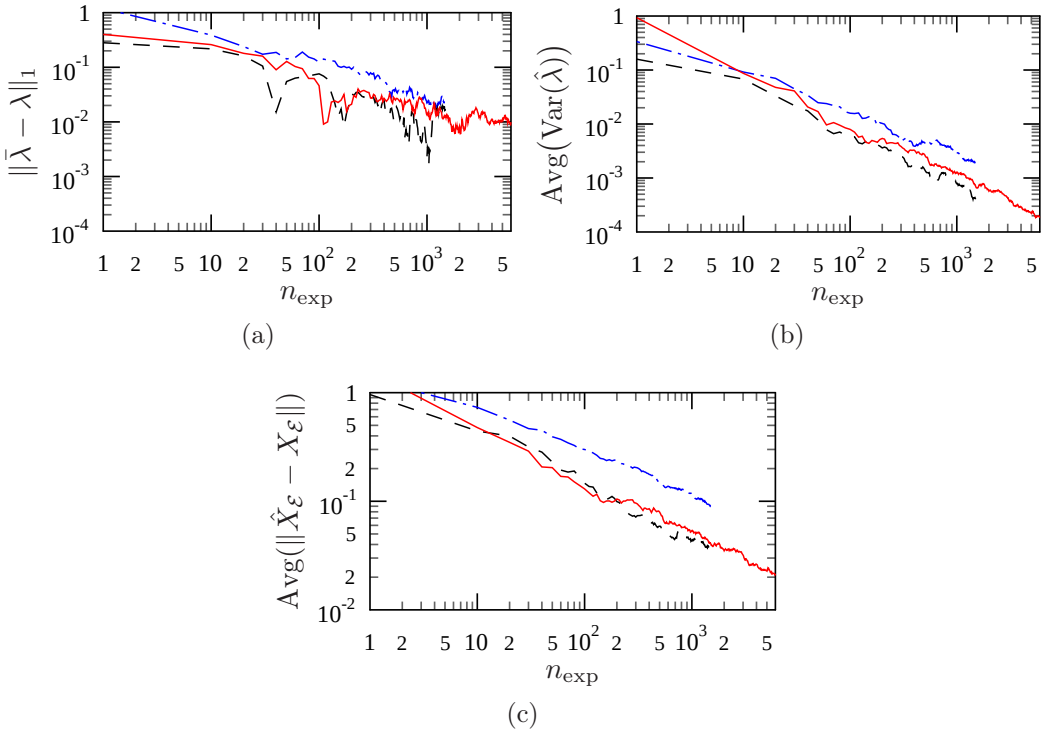


Figure B.6. Standard tomography with $\lambda = (0.6, 0.5, 0.3)$ taken from the interior of the parameter space.

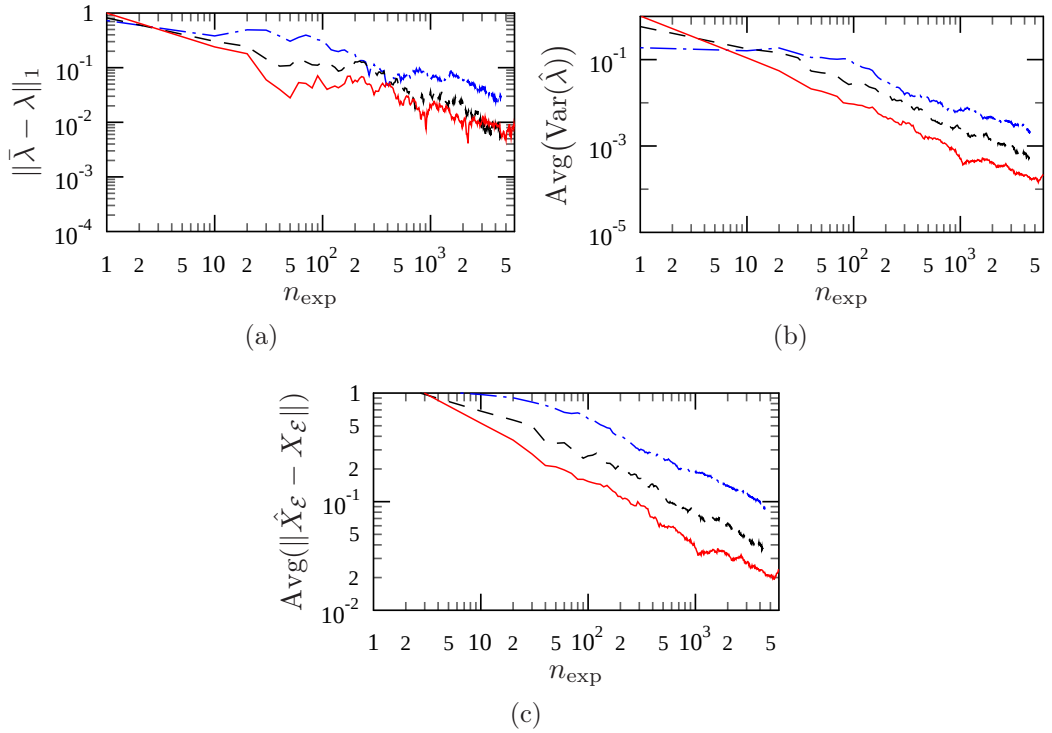


Figure B.7. Minimal tomography with $\lambda = (0, -0.3, 0.1)$ taken from the interior of the parameter space.

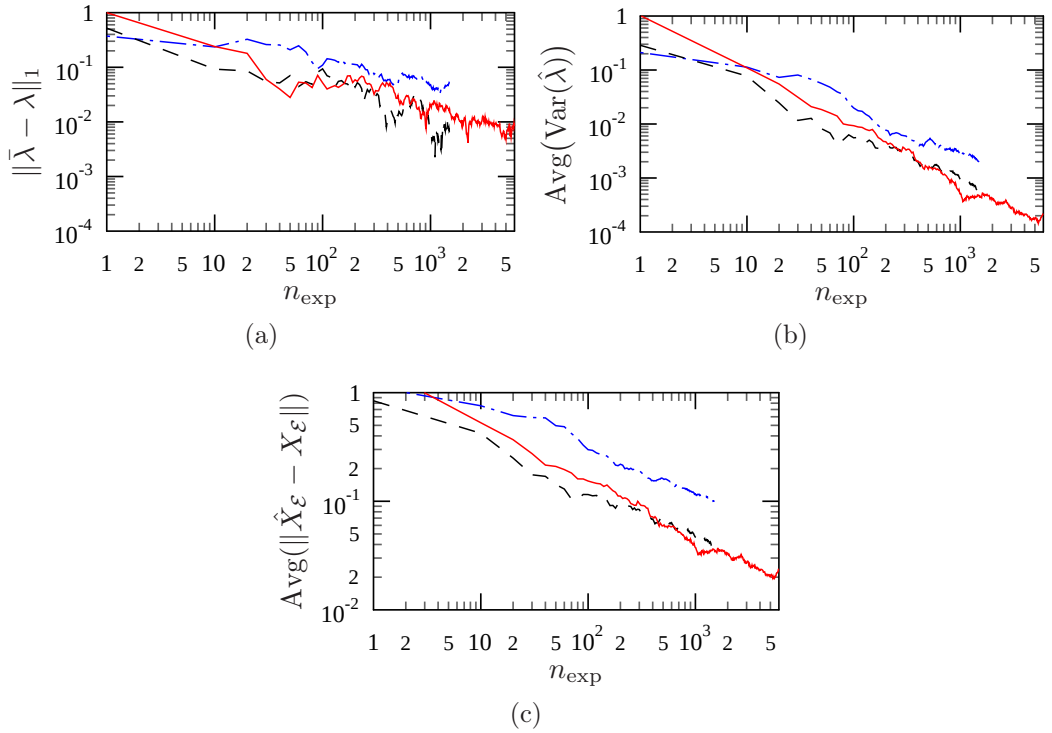


Figure B.8. Standard tomography with $\lambda = (0, -0.3, 0.1)$ taken from the interior of the parameter space.

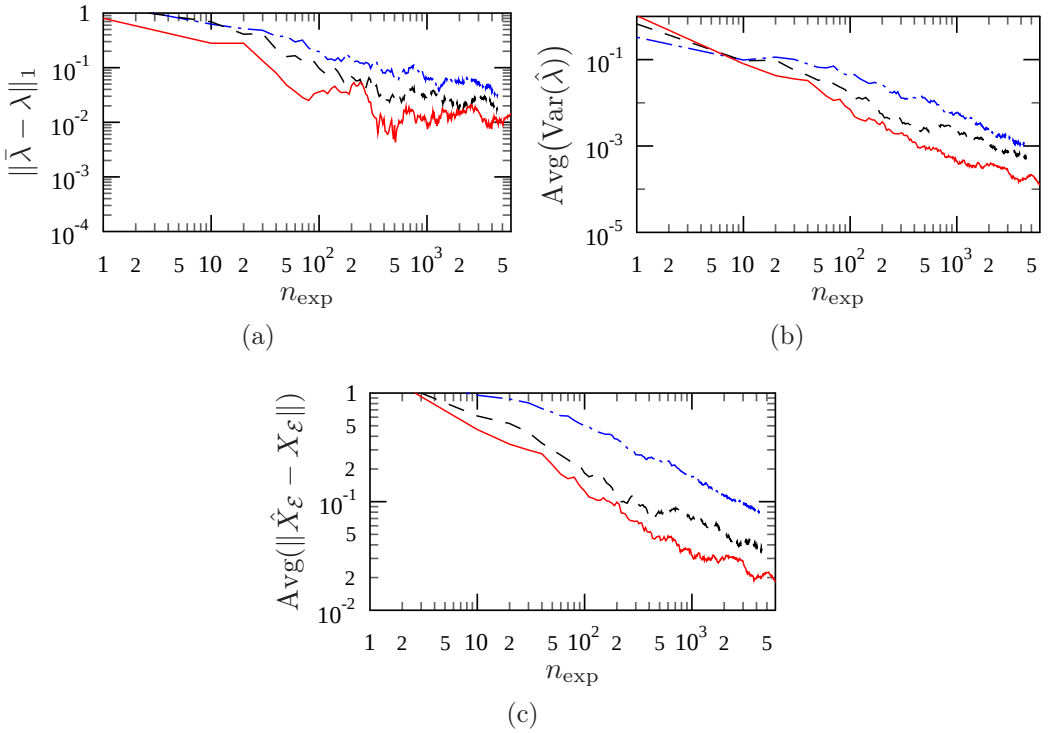


Figure B.9. Minimal tomography with $\lambda = (0, -0.4, 0.6)$ taken from a face of the parameter space.

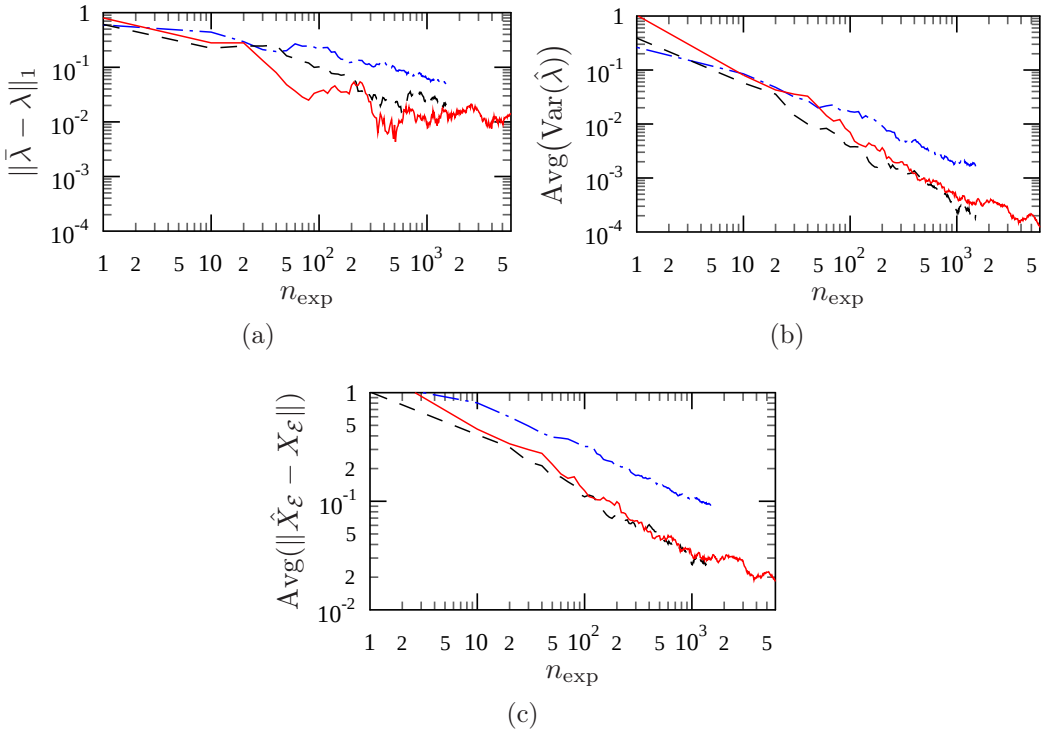


Figure B.10. Standard tomography with $\lambda = (0, -0.4, 0.6)$ taken from a face of the parameter space.

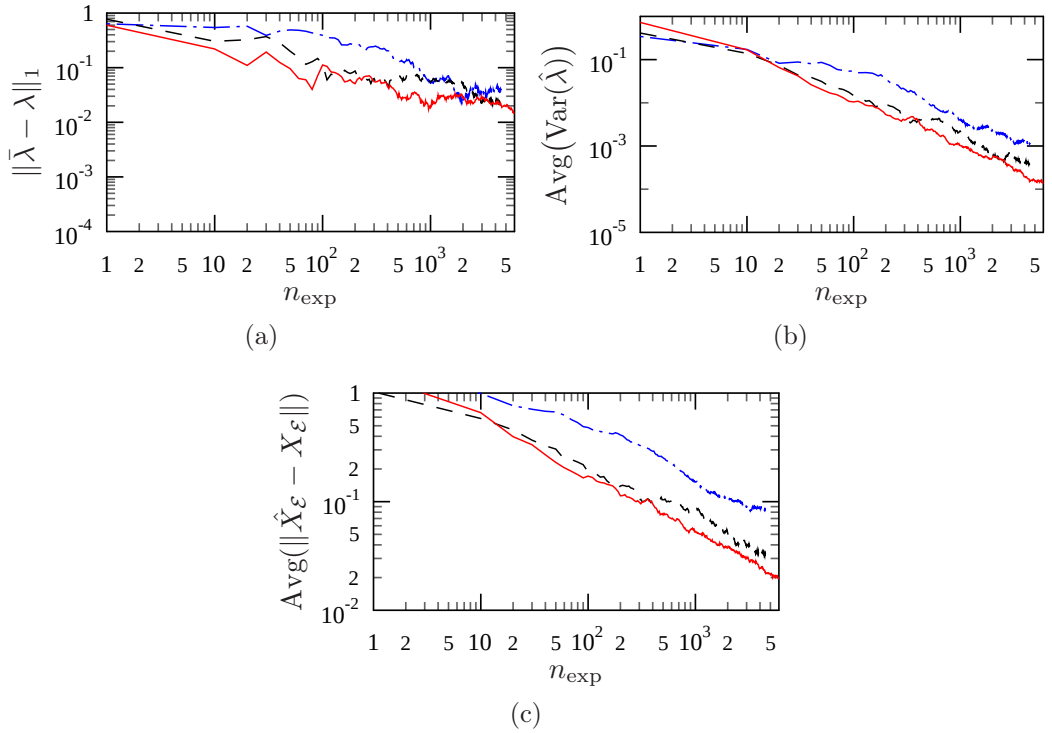


Figure B.11. Minimal tomography with $\lambda = (-0.4, -0.4, -0.2)$ taken from a face of the parameter space.

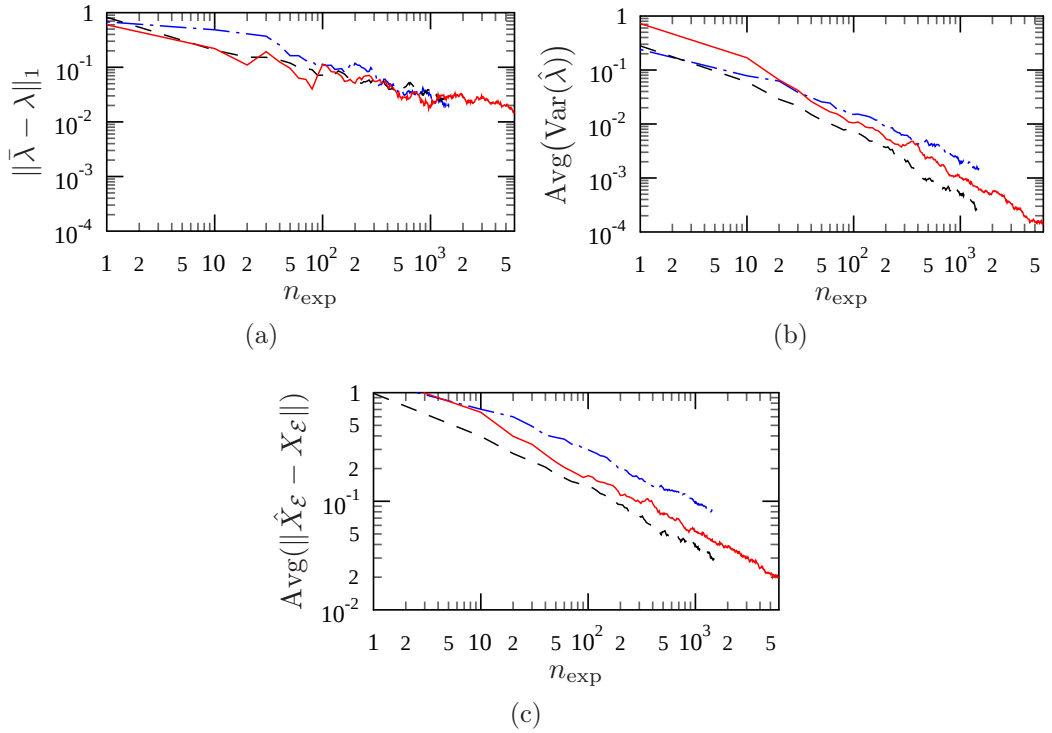


Figure B.12. Standard tomography with $\lambda = (-0.4, -0.4, -0.2)$ taken from a face of the parameter space.

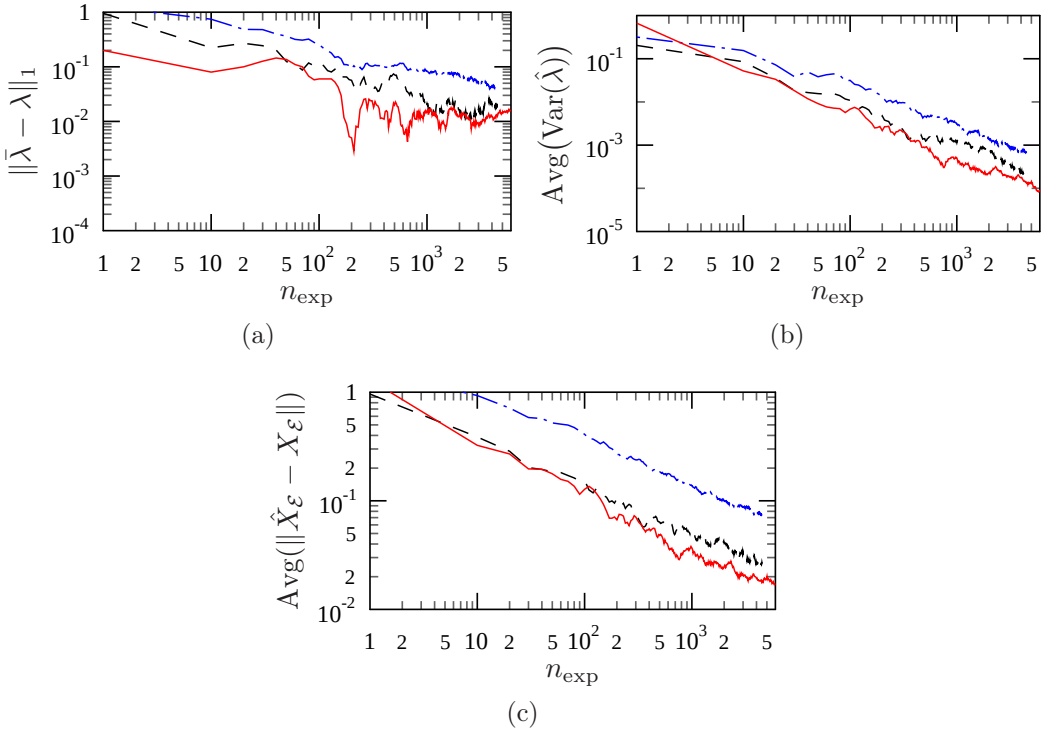


Figure B.13. Minimal tomography with $\lambda = (-0.4, 0.4, -1)$ taken from an edge of the parameter space.

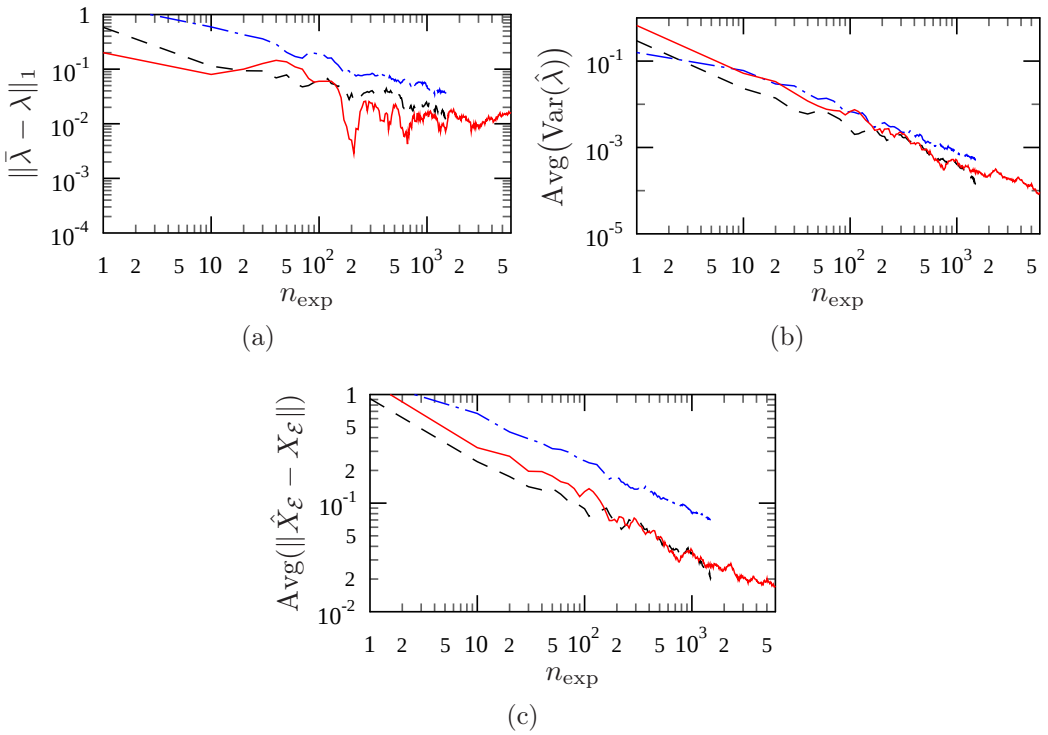


Figure B.14. Standard tomography with $\lambda = (-0.4, 0.4, -1)$ taken from an edge of the parameter space.

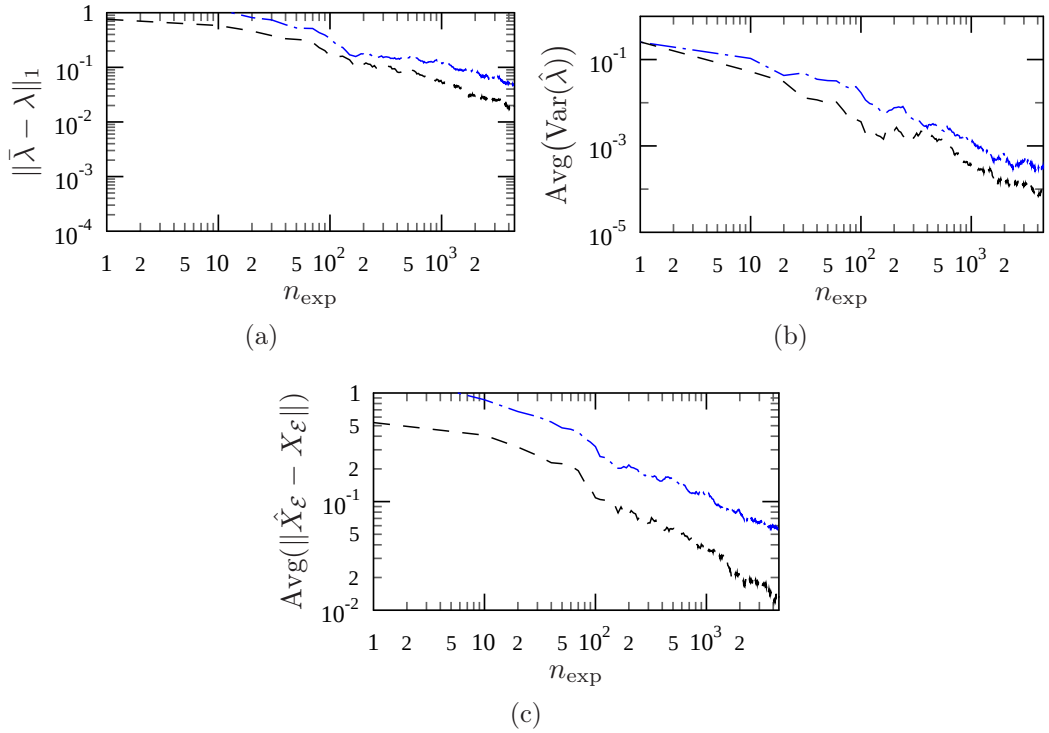


Figure B.15. Minimal tomography with $\lambda = (1, -1, -1)$, a vertex of the parameter space. The solid line is zero here, because extremal Pauli channels can be estimated with perfect accuracy using the optimal method.

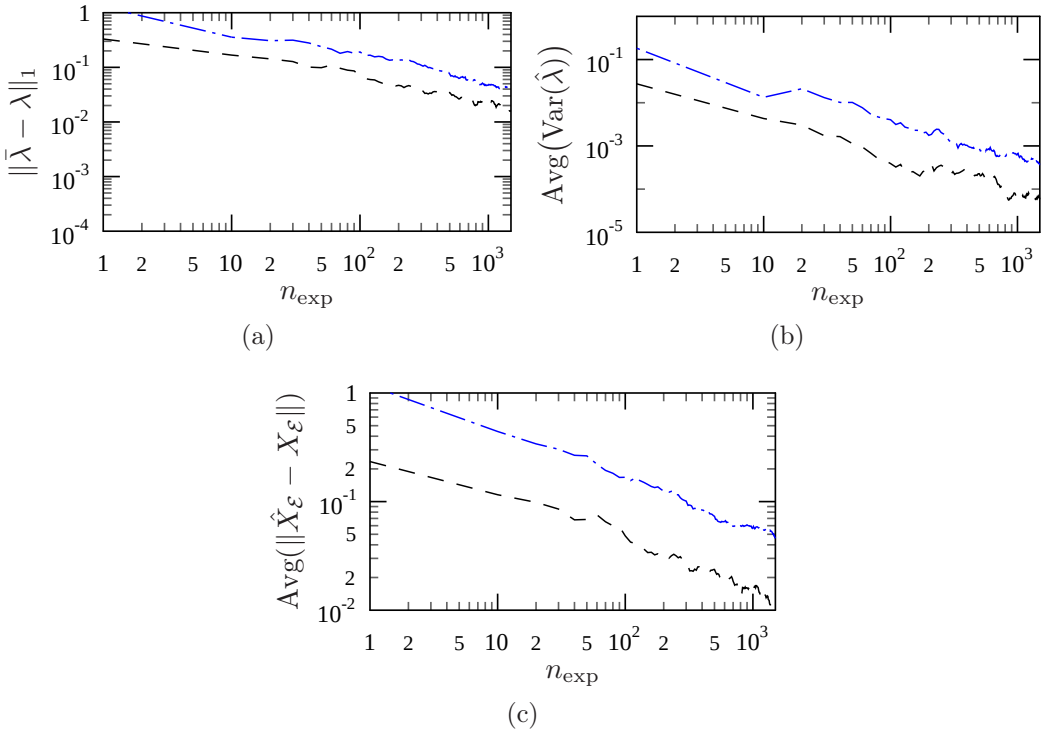


Figure B.16. Standard tomography with $\lambda = (1, -1, -1)$, a vertex of the parameter space. The solid line is zero here, because extremal Pauli channels can be estimated with perfect accuracy using the optimal method.

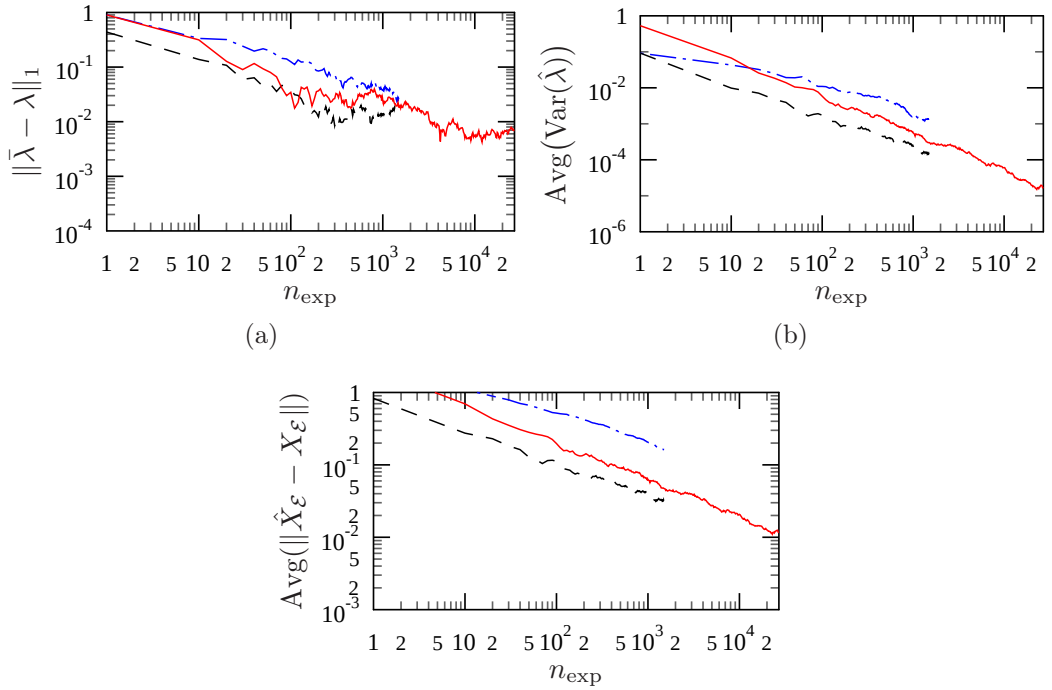


Figure B.17. 3-level tomography with $\lambda = (0.55, 0.1, 0.1, 0.1)$ taken from the interior of the parameter space.

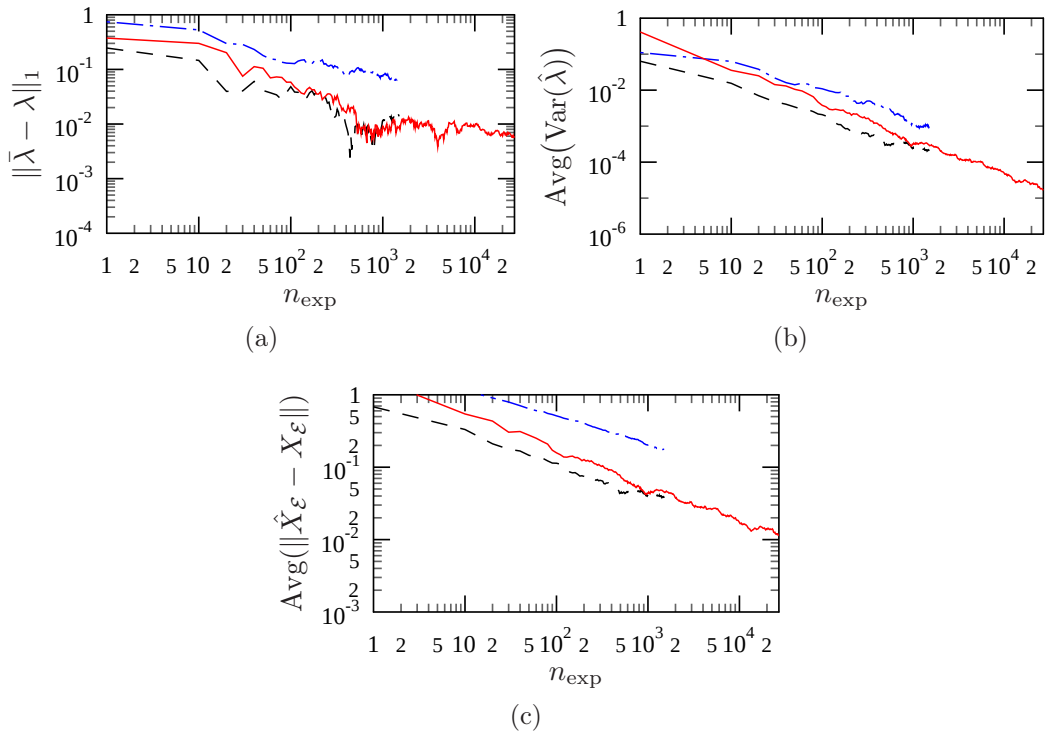


Figure B.18. 3-level tomography with $\lambda = (-0.275, 0.175, -0.35, 0.175)$ taken from the interior of the parameter space.

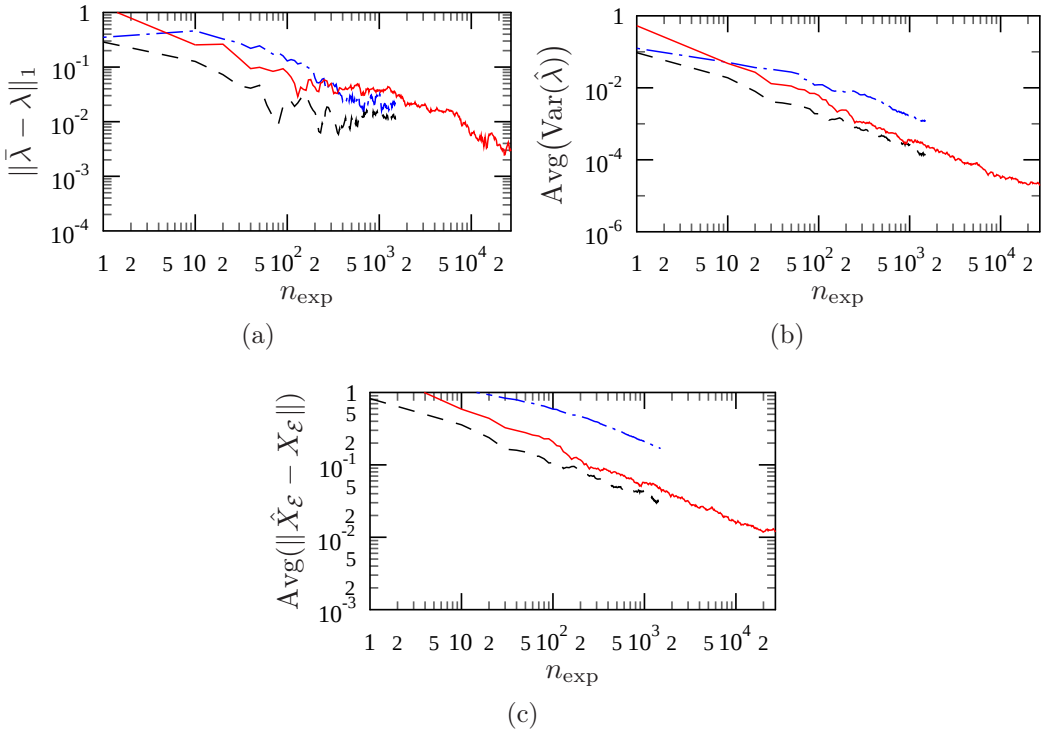


Figure B.19. 3-level tomography with $\lambda = (0.1, 0.025, 0.175, -0.125)$ taken from the interior of the parameter space.

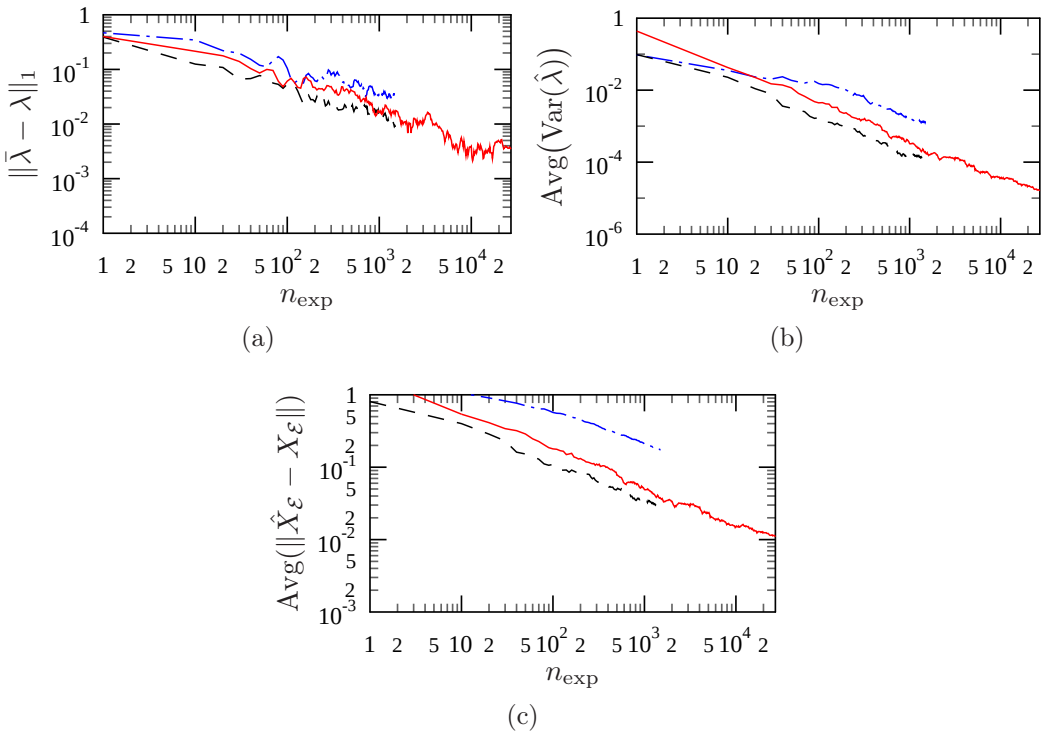


Figure B.20. 3-level tomography with $\lambda = (-0.2, -0.125, -0.025, 0.25)$ taken from the interior of the parameter space.

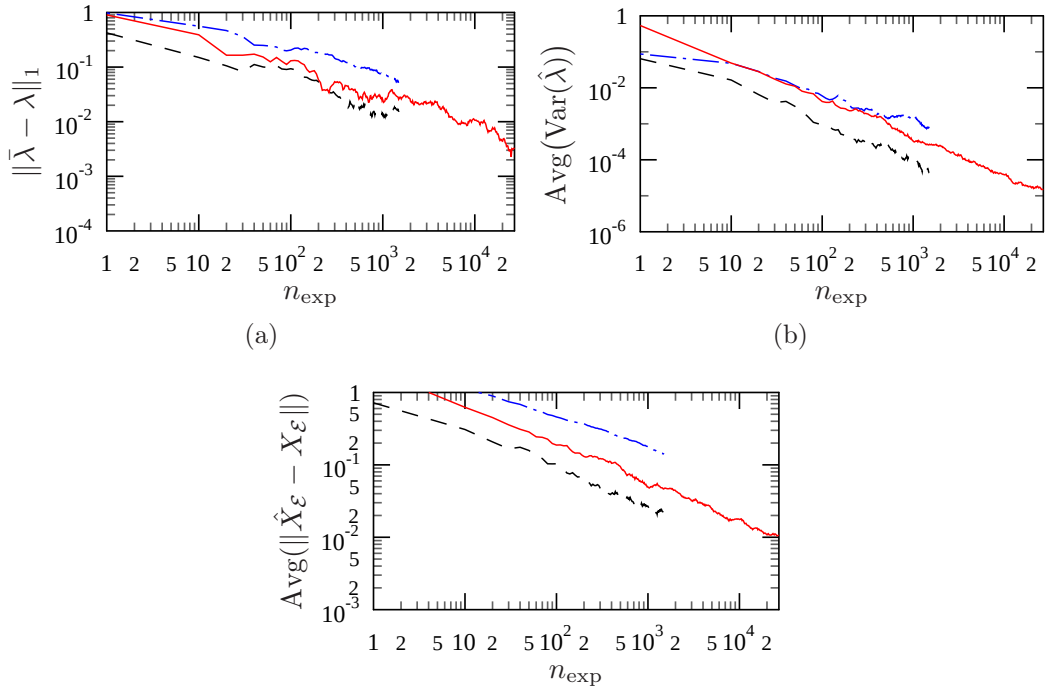


Figure B.21. 3-level tomography with $\lambda = (-0.35, -0.2, 0.1, -0.05)$ taken from a face of the parameter space.

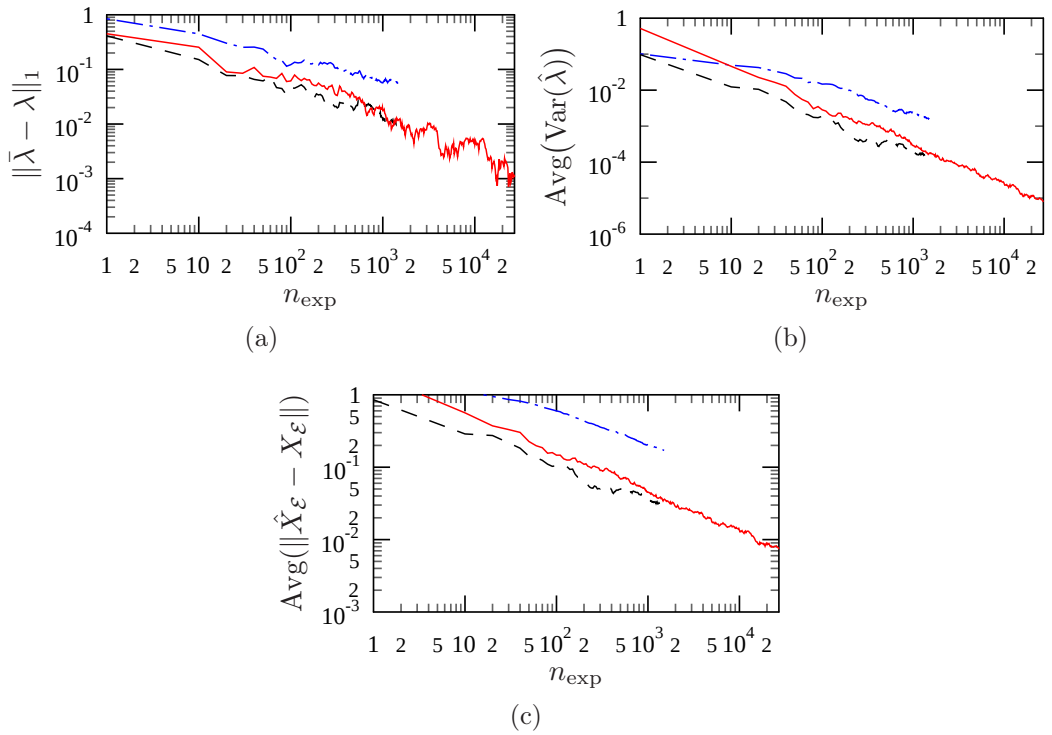


Figure B.22. 3-level tomography with $\lambda = (0.25, -0.2, -0.2, 0.55)$ taken from a face of the parameter space.

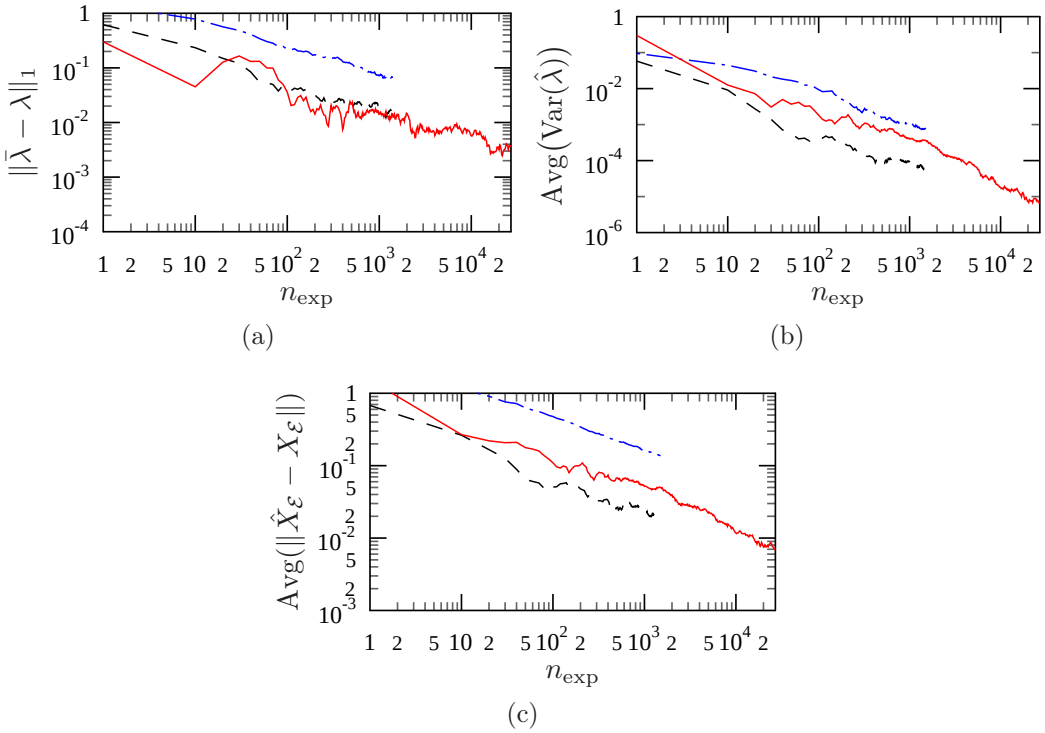


Figure B.23. 3-level tomography with $\lambda = (-0.5, -0.5, 0.1, 0.4)$ taken from an edge of the parameter space.

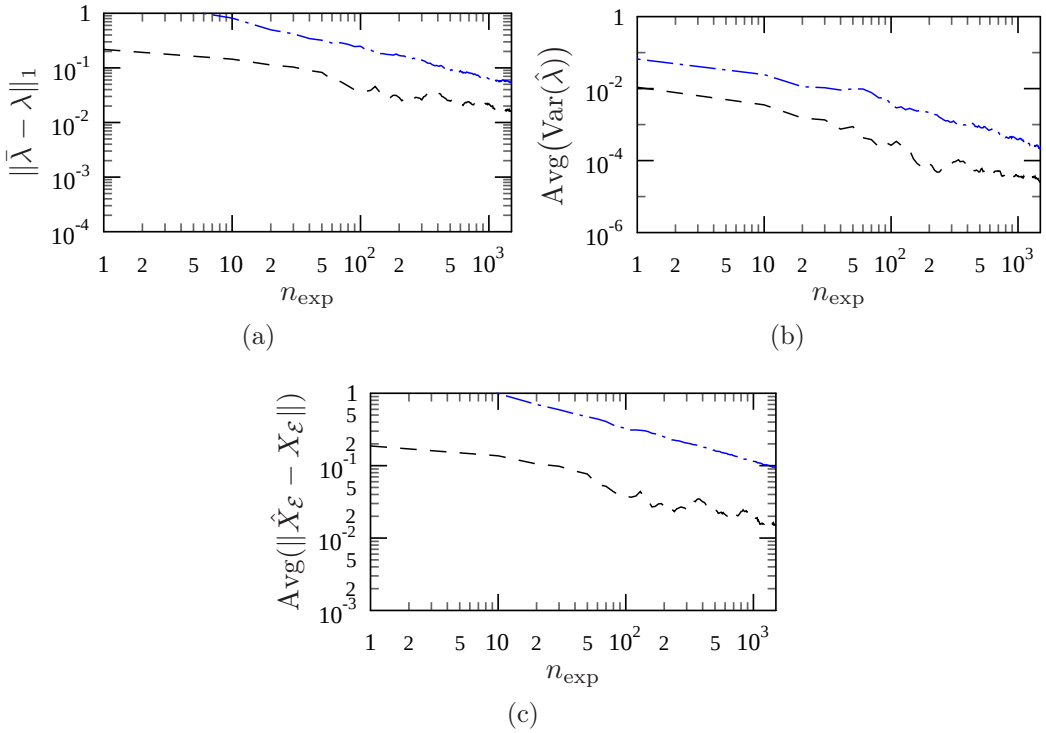


Figure B.24. 3-level tomography with $\lambda = (-0.5, 1, -0.5, -0.5)$, a vertex of the parameter space. The solid line is zero here, because extremal Pauli channels can be estimated with perfect accuracy using the optimal method.

Bibliography

[1] L. Ljung, *System Identification: Theory for the User*, Prentice Hall, Upper Saddle River, New Jersey (1999)

[2] L. Ljung, G. T., *Modeling of Dynamic Systems*, Prentice Hall (1994)

[3] K. Zhou, J. C. Doyle, *Essentials of Robust Control*, Prentice-Hall (1998)

[4] J. Chen, G. Gu, *Control Oriented System Identification: An $\mathcal{H}_\infty$ Approach*, Wiley-Interscience (2000)

[5] M. A. Nielsen, I. L. Chuang, *Quantum Computation and Quantum Information*, Cambridge University Press (2000)

[6] D. Lidar, T. Brun, *Quantum Error Correction*, Cambridge University Press (2013)

[7] M. G. A. Paris, J. Rehácek (eds.), *Quantum State Estimation*, *Lecture Notes in Physics, Berlin Springer Verlag*, vol. 649 (2004)

[8] D. Petz, *Quantum Information Theory and Quantum Statistics*, Theoretical and Mathematical Physics, Springer-Verlag (2008)

[9] K. Jacobs, D. Steck, A straightforward introduction to continuous quantum measurement, *Contemporary Physics*, **47, 5**: 279–303 (2006)

[10] G. D'Ariano, P. Lo Presti, P. Perinotti, Classical randomness in quantum measurements, *Journal of Physics A Mathematical General*, **38**: 5979–5991 (2005), `arXiv:quant-ph/0408115`

[11] M. Horodecki, P. Horodecki, R. Horodecki, General teleportation channel, singlet fraction, and quasidistillation, *Physical Review A*, **60, 3**: 1888 (1999)

[12] A. Jamiolkowski, Linear transformations which preserve trace and positive semidefiniteness of operators, *Reports on Mathematical Physics*, **3**: 275–278 (1972)

[13] A. Fujiwara, P. Algoet, One-to-one parametrization of quantum channels, *Phys. Rev. A*, **59**: 3290–3294 (1999)

[14] D. Petz, H. Ohno, Generalizations of Pauli channels, *Acta Math. Hungar*, **124**: 165–177 (2009), 0812.2668

[15] B. Schumacher, Sending entanglement through noisy quantum channels, *Phys. Rev. A*, **54, 4**: 2614–2628 (1996)

[16] M. Reimpell, R. F. Werner, Iterative Optimization of Quantum Error Correcting Codes, *Phys. Rev. Lett.*, **94, 8**: 080501 (2005)

[17] E. Knill, R. Laflamme, Theory of quantum error-correcting codes, *Phys. Rev. A*, **55, 2**: 900–911 (1997)

[18] D. Gottesman, Stabilizer codes and quantum error correction, Ph.D. thesis, California Institute of Technology (1997)

[19] A. S. Fletcher, Channel-Adapted Quantum Error Correction, Ph.D. thesis, MIT (2007), 0706.3400

[20] I. Csiszar, J. Körner, *Information theory : coding theorems for discrete memoryless systems*, Cambridge University Press, Cambridge, New York (NY) (2011)

[21] K. Hornberger, Introduction to decoherence theory (2008), quant-ph/0612118

[22] N. Yamamoto, S. Hara, K. Tsumura, Suboptimal quantum-error-correcting procedure based on semidefinite programming, *Phys. Rev. A*, **71, 2**: 022322 (2005)

[23] K. Audenaert, B. De Moor, Optimizing completely positive maps using semidefinite programming, *Phys. Rev. A*, **65, 3**: 030302 (2002)

[24] A. S. Fletcher, P. W. Shor, M. Z. Win, Optimum quantum error recovery using semidefinite programming, *Phys. Rev. A*, **75, 1**: 012338 (2007), arXiv:quant-ph/0606035

[25] A. S. Fletcher, P. W. Shor, M. Z. Win, Structured near-optimal channel-adapted quantum error correction, *Phys. Rev. A*, **77, 1**: 012320 (2008), 0708.3658

[26] R. L. Kosut, A. Shabani, D. A. Lidar, Robust Quantum Error Correction via Convex Optimization, *Phys. Rev. Lett.*, **100, 2**: 020502 (2008), arXiv:quant-ph/0703274

[27] R. Kosut, I. A. Walmsley, H. Rabitz, Optimal Experiment Design for Quantum State and Process Tomography and Hamiltonian Parameter Estimation, *arXiv:quant-ph*, **0411093**: 1–51 (2004)

[28] M. Paris, J. Rehácek, *Quantum state estimation, Lect. Notes Phys.* **649**, Springer, Berlin (2004)

[29] M. Mohseni, A. T. Rezakhani, D. A. Lidar, Quantum Process Tomography: Resource Analysis of Different Strategies, *Physical Review A*, **77**: 032322 (2008)

[30] M. Ziman, Process positive-operator-valued measure: A mathematical framework for the description of process tomography experiments, *Phys. Rev. A*, **77**: 062112 (2008)

[31] F. Pukelsheim, *Optimal Design of Experiments*, Classics in Applied Mathematics 50, SIAM (2006)

[32] R. D. Gill, S. Massar, State estimation for large ensembles (2002), URL `arXiv:quant-ph/9902063v2`

[33] M. F. Sacchi, Maximum-likelihood reconstruction of completely positive maps, *Phys. Rev. A*, **63, 5**: 054104 (2001)

[34] M. Sasaki, M. Ban, S. M. Barnett, Optimal parameter estimation of a depolarizing channel, *Phys. Rev. A*, **66, 2**: 022308 (2002)

[35] M. Branderhorst, J. Nunn, I. Walmsley, R. Kosut, Simplified quantum process tomography, *New Journal of Physics*, **11**: 115010 (2009)

[36] K. C. Young, M. Sarovar, R. Kosut, K. B. Whaley, Optimal quantum multiparameter estimation and application to dipole- and exchange-coupled qubits, *Phys. Rev. A*, **79, 6**: 062301 (2009)

[37] M. Nathanson, M. B. Ruskai, Pauli diagonal channels constant on axes, *Journal of Physics A Mathematical General*, **40**: 8171–8204 (2007), `arXiv:quant-ph/0611106`

[38] J. Rehácek, B.-G. Englert, D. Kaszlikowski, Minimal qubit tomography, *Phys. Rev. A*, **70, 5**: 052321 (2004)

[39] M. Sarovar, G. Milburn, Optimal estimation of one-parameter quantum channels, *Journal of Physics A: Mathematical and General*, **39**: 8487 (2006)

[40] A. Fujiwara, H. Imai, Quantum parameter estimation of a generalized Pauli channel, *Jornal of Physics A: Mathematical and General*, **36**: 8093–8103 (2003)

[41] A. Fujiwara, Quantum channel identification problem, *Phys. Rev. A*, **63, 4**: 042304 (2001)

[42] S. Boyd, L. Vandenberghe, *Convex Optimization*, Cambridge University Press (2004)

[43] F. Huszár, N. M. T. Houlsby, Adaptive Bayesian quantum tomography, *Phys. Rev. A*, **85**: 052120 (2012)

[44] T. Sugiyama, P. S. Turner, M. Murao, Adaptive experimental design for one-qubit state estimation with finite data based on a statistical update criterion, *Phys. Rev. A*, **85**: 052107 (2012), URL `http://link.aps.org/doi/10.1103/PhysRevA.85.052107`

[45] T. Heinosaari, M. Ziman, *The Mathematical Language of Quantum Theory: From Uncertainty to Entanglement*, Cambridge University Press (2012)

[46] H. D. Mittelmann, An independent benchmarking of SDP and SOCP solvers., *Math. Program.*, **95, 2**: 407–430 (2003)

[47] D. Petz, Complementarity in quantum systems, *Reports on Mathematical Physics*, **59**: 209–224 (2007), `arXiv:quant-ph/0610189`

[48] S. Bandyopadhyay, P. O. Boykin, V. Roychowdhury, F. Vatan, A new proof for the existence of mutually unbiased bases, *ArXiv Quantum Physics e-prints* (2001), `arXiv:quant-ph/0103162`

[49] S. Bandyopadhyay, P. O. Boykin, V. P. Roychowdhury, F. Vatan, A New Proof for the Existence of Mutually Unbiased Bases., *Algorithmica*, **34, 4**: 512–528 (2002)

[50] P. W. Shor, Scheme for reducing decoherence in quantum computer memory, *Phys. Rev. A*, **52, 4**: R2493–R2496 (1995)

[51] R. Laflamme, C. Miquel, J. P. Paz, W. H. Zurek, Perfect Quantum Error Correcting Code, *Phys. Rev. Lett.*, **77, 1**: 198–201 (1996)

[52] A. R. Calderbank, E. M. Rains, P. W. Shor, N. J. A. Sloane, Quantum Error Correction and Orthogonal Geometry, *Phys. Rev. Lett.*, **78, 3**: 405–408 (1997)

[53] J. Löfberg, YALMIP : A Toolbox for Modeling and Optimization in MATLAB, in *Proceedings of the CACSD Conference*, p. 284–289, Taipei, Taiwan (2004)

[54] K. C. Toh, M. Todd, R. Tütüncü, SDPT3 - a MATLAB software package for semidefinite programming, *Optimization Methods and Software*, **11**: 545–581 (1998)