



Estimation d'état et de paramètres pour les systèmes quantiques ouverts

Pierre Six

► To cite this version:

Pierre Six. Estimation d'état et de paramètres pour les systèmes quantiques ouverts. Physique mathématique [math-ph]. Université Paris sciences et lettres, 2016. Français. NNT : 2016PSLEM019 . tel-01511698

HAL Id: tel-01511698

<https://pastel.hal.science/tel-01511698>

Submitted on 21 Apr 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THÈSE DE DOCTORAT

de l'Université de recherche Paris Sciences et Lettres
PSL Research University

Préparée à MINES ParisTech

Estimation d'état et de paramètres pour les systèmes quantiques ouverts

Estimation of state and parameters in open quantum systems

Ecole doctorale n°432

SCIENCES DES MÉTIERS DE L'INGÉNIEUR

Spécialité MATHÉMATIQUES ET AUTOMATIQUE

COMPOSITION DU JURY :

M. Yacine CHITOUR
Lab. des Signaux et Système, Université
Paris Sud, Rapporteur

M. Éric MOULINES
Centre de Mathématiques Appliquées,
École Polytechnique, Rapporteur

M. Madalin GUTA
School of Mathematical Sciences, Univ.
of Nottingham, Examineur

M. Igor DOTSENKO
Lab. Kastler Brossel, Collège de France,
Examineur

M. Pierre ROUCHON
Centre Automatiques et Systèmes, Mines
ParisTech, Examineur

M. Benjamin HUARD
Lab. Pierre Aigrain, École Normale
Supérieure, Président

Soutenue par PIERRE SIX
le 22 novembre 2016

Dirigée par **Pierre ROUCHON**



Résumé

La communauté scientifique a réussi ces dernières années à bâtir des systèmes quantiques simples sur lesquels des séries de mesures sont acquises successivement le long de trajectoires quantiques et sans réinitialisation de l'état (opérateur densité) par l'expérimentateur. L'objet de cette thèse est d'adapter les méthodes de tomographie quantique (estimation d'état et de paramètres) à ce cadre pour prendre en compte la rétroaction de la mesure sur l'état, la décohérence et les imperfections expérimentales.

Durant le processus de mesure, l'évolution de l'état quantique est alors gouvernée par un processus de Markov à états cachés (filtres quantiques de Belavkin). Pour des mesures en temps continu, nous commençons par montrer comment discrétiser l'équation maîtresse stochastique tout en préservant la positivité et la trace de l'état quantique, et ainsi se ramener aux filtres quantiques en temps discret. Ensuite, nous développons, à partir de trajectoires de mesures en temps discret, des techniques d'estimation par maximum de vraisemblance pour l'état initial et les paramètres. Cette estimation est accompagnée de son intervalle de confiance. Lorsqu'elle concerne des valeurs de paramètres (tomographie de processus quantique), nous donnons un résultat de robustesse grâce au formalisme des filtres particuliers et nous proposons une méthode de maximisation fondée sur le calcul du gradient par l'adjoint et bien adaptée au cas multiparamétrique. Lorsque l'estimation porte sur l'état initial (tomographie d'état quantique), nous donnons une formulation explicite de la fonction de vraisemblance grâce aux états adjoints, montrons que son logarithme est une fonction concave de l'état initial et élaborons une expression intrinsèque de la variance grâce à des développements asymptotiques de moyennes bayésiennes et reposant sur la géométrie de l'espace des opérateurs densité.

Ces méthodes d'estimation ont été appliquées et validées expérimentalement pour deux types de mesures quantiques : des mesures en temps discret non destructives de photons dans le groupe d'électrodynamique quantique en cavité du LKB au Collège de France, des mesures diffusives de la fluorescence d'un qubit supraconducteur dans le groupe d'électronique quantique du LPA à l'ENS Paris.

Mots-clés

Tomographie quantique ; Filtres quantiques ; Équation maîtresse stochastique ; Qubit supraconducteur ; Électrodynamique quantique en cavité ; Maximum de vraisemblance ; Estimation bayésienne ; Calcul du gradient par l'adjoint ; Développement asymptotique d'intégrales de Laplace

Abstract

In recent years, the scientific community has succeeded in experimentally building simple quantum systems on which series of measurements are successively acquired along quantum trajectories, without any reinitialization of their state (density operator) by the physicist. The subject of this thesis is to adapt the quantum tomography techniques (state and parameters estimation) to this frame, to take into account the feedback of the measurement on the state, the decoherence and experimental imperfections.

During the measurement process, the evolution of the quantum state is then governed by a hidden-state Markov process (Belavkin quantum filters). Concerning continuous-time measurements, we begin by showing how to discretize the stochastic master equation, while preserving the positivity and the trace of the quantum state, and so reducing to discrete-time quantum filters. Then, we develop, starting from trajectories of discrete-time measurements, some maximum-likelihood estimation techniques for initial state and parameters. This estimation is coupled with its confidence interval. When it concerns the value of parameters (quantum process tomography), we provide a result of robustness using the formalism of particular filters, and we propose maximization technique based on the calculus of gradient by adjoint method, well adapted to the multi-parametric case. When the estimation concerns the initial state (quantum state tomography), we give an explicit formulation of the likelihood function thanks to the adjoint states, show that its logarithm is a concave function of the initial state and build an intrinsic expression of the variance, obtained from asymptotic developments of bayesian means, lying on the geometry of the space of density operators.

These estimation techniques have been applied and experimentally validated for two types of quantum measurements : discrete-time non-destructive measurements of photons in the group of cavity quantum electro-dynamics of LKB at Collège de France, diffusive measurements of the fluorescence of a supra-conducting qubit in the quantum electronics group of LPA at ENS Paris.

Keywords

Quantum tomography ; Quantum filters ; Stochastic master equation ; Superconducting qubit ; Cavity quantum electrodynamics ; Maximum-likelihood ; Bayesian estimation ; Gradient calculus by adjoint method ; Asymptotic expansion of Laplace's integrals

Remerciements

J'écris ces remerciements plusieurs semaines après avoir soutenu le travail présenté ici ; c'est donc du haut de ma première expérience de docteur, comprenant a posteriori tout ce que m'a apporté cette thèse, que j'écris les quelques lignes qui suivent.

Mes premiers remerciements s'adressent, bien sûr, à Pierre Rouchon, avec qui j'ai eu un très grand plaisir de travailler durant trois ans. Nous avons eu, tout au long de cette thèse, un très grand nombre d'échanges scientifiques fructueux, qui m'ont habitué à comprendre la façon dont la recherche en mathématiques appliquées se déroule au quotidien.

Je tiens également à remercier mon jury : Yacine Chitour et Eric Moulines, qui ont accepté d'être mes rapporteurs, et Madalin Guta qui, bien qu'examineur, a tenu selon moi un rôle tout à fait analogue. Prisent part également au jury Benjamin Huard, du laboratoire Pierre Aigrain, et Igor Dotsenko, du laboratoire Kastler-Brossel, avec qui j'ai eu un grand plaisir à travailler durant ma thèse.

Je remercie également l'ensemble des chercheurs permanents et doctorants du Centre Automatique et Systèmes, au sein duquel a toujours régné une ambiance sympathique, quoique stimulante. Je remercie particulièrement Lionel Magnis et Pascal Combes dont le secours m'a parfois été précieux, et Nicolas Petit avec qui j'ai beaucoup aimé converser. Je tiens également à saluer toutes les personnes avec qui j'ai travaillé, particulièrement au sein des deux laboratoires expérimentaux que dirigent Benjamin Huard et Igor Dotsenko.

Enfin, je tiens à remercier mes parents, ma soeur Clémence, mes amis, et certaines personnes qui, si elles ont la pudeur de ne vouloir être citées ici, n'en furent pas moins importantes pour moi.

Table des matières

1	Introduction	13
2	Structure des modèles	17
2.1	Système isolé et mesures parfaites : fonctions d'onde	17
2.1.1	Propriétés des systèmes physiques classiques	17
2.1.2	Systèmes quantiques isolés	19
2.1.3	Systèmes ouverts à mesures parfaites	21
2.1.4	Ouverture	25
2.2	Systèmes ouverts à mesures imparfaites : matrices densité	26
2.2.1	Motivations	26
2.2.2	Notion d'opérateur densité	27
2.2.3	Adaptation des équations d'évolution aux matrices densité	30
2.2.4	Mesures non lues et mesures imparfaites	31
2.3	Trajectoire quantique et estimation de paramètres	37
2.3.1	Cadre général de la thèse	37
2.3.2	Filtre réel et filtres approchés	38
3	Tomographie d'état	41
3.1	Introduction	41
3.2	Préliminaires mathématiques	43
3.2.1	Notations utilisées	43
3.2.2	Probabilité des trajectoires de mesure	44
3.2.3	États adjoints	46
3.3	Estimation par Maximum de Vraisemblance	48
3.3.1	Contexte de l'optimisation	48
3.3.2	Comportement asymptotique et variance d'estimation	52
3.4	Passage à l'estimation bayésienne	55
3.4.1	Principe de l'estimation bayésienne	55
3.4.2	Comparaison avec le Maximum de Vraisemblance	56
3.5	Validations expérimentales	58
3.5.1	Cas d'un champ de photons : mesures discrètes	58
3.5.2	Cas d'un qubit : mesures diffusives	59
4	Estimation de paramètres et filtres particuliers	67
4.1	Introduction	67
4.2	Robustesse des filtres particuliers	68
4.2.1	Filtres particuliers	68
4.2.2	Description du système par filtre étendu	70
4.2.3	Robustesse de l'estimation par Maximum de Vraisemblance	73

4.3	Généralisation de la notion de robustesse	75
4.3.1	Extension à un espace de valeurs de paramètre possibles	75
4.3.2	Extension aux trajectoires de mesure successives	75
4.4	Filtre particulaire à temps continu et discrétisation	76
4.4.1	Filtre particulaire continu	76
4.4.2	Filtre particulaire discrétisé	78
4.5	Validations expérimentales	79
4.5.1	Estimation de l'efficacité de détection	80
4.5.2	Estimation du temps de relaxation	80
5	Tomographie de processus quantique	85
5.1	Introduction	85
5.1.1	Motivations	85
5.1.2	Cadre de la maximisation	85
5.2	Optimisation grâce à l'adjoint	86
5.2.1	Écriture du Lagrangien et optimalité	86
5.2.2	Lagrangien et conditions d'optimalité	87
5.2.3	Calcul du gradient et états adjoints	88
5.3	Validations expérimentales : champ de photons et mesures discrètes	92
5.3.1	Configuration	92
5.3.2	Cadre de l'estimation	92
5.3.3	Résultats	93
6	Conclusion	95
	Bibliographie	97
A	Systèmes expérimentaux utilisés	103
A.1	Boîte à photons	103
A.1.1	Dispositif expérimental	103
A.1.2	Cadre mathématique	103
A.2	Qubit supraconducteur	105
A.2.1	Dispositif expérimental	106
A.2.2	Cadre mathématique	106
B	Développements mathématiques	109
B.1	Hamiltonien de trace nulle	109
B.2	Mesure discrète par un système tiers	109
B.3	Processus de Wiener	111
B.4	Schéma de discrétisation des équations de mesures diffusives	112
B.5	Autour de la fidélité	115
B.5.1	Cas d'un état pur	115
B.5.2	Théorème d'Uhlmann	116
B.5.3	Propriétés de la fidélité	118
B.6	Maximisation de la fonction de log-vraisemblance pour la tomographie d'état	118
B.6.1	Projection sur l'ensemble des matrices densité admissibles	118
B.6.2	Descente par algorithme EM	119

C	Développements asymptotiques de l'estimation Bayésienne	121
C.1	Introduction	121
C.2	Développement asymptotique des intégrales de Laplace	123
C.3	Application à la tomographie d'état quantique	131
C.4	Conclusion	137

Chapitre 1

Introduction

Ce mémoire porte sur l'estimation d'états et de paramètres à partir de mesures en temps discret ou continu pour les systèmes quantiques ouverts. Plus précisément, il s'agit de systèmes physiques régis par les lois de la mécanique quantique, soumis à la fois à la décohérence et à l'action en retour du processus de mesure. Leur dynamiques sont bien décrites par des chaînes de Markov à états cachés combinant l'évolution cohérence selon l'équation de Schrödinger, la réduction du paquet d'ondes attachée à la mesure quantique et la loi de Bayes pour la prise en compte des imperfections et de la décohérence [HR06, WM09]. En temps continu, il s'agit du filtre quantique de Belavkin [Bel92], présenté de façon détaillée dans [BvHJ07] pour des signaux d'entrée quasi-classiques et dans [GJNC12] pour une classe de signaux d'entrée complètement quantiques (pour une version en temps discret voir par exemple [DMB⁺09, SDSR12], pour le passage du discret au continu voir par exemple [APR14]). Dans ce cadre, l'estimation d'état (resp. de paramètres) correspond en physique et information quantique à la tomographie d'état [PR04] (resp. de processus). Pour la tomographie d'état, il s'agit d'inférer sa valeur initiale connaissant le modèle (la chaîne de Markov à état caché) et un ensemble de mesures, toutes partant de cette valeur initiale à estimer. Pour la tomographie de processus, il s'agit d'inférer la valeur d'un ou de plusieurs paramètres connaissant la dépendance du modèle par rapport à ce ou ces paramètres et un ensemble de mesures partant d'états initiaux bien connus par ailleurs.

Dans ce mémoire nous nous sommes posés la question s'il était possible d'améliorer et d'enrichir les méthodes usuelles d'estimation quantique pour prendre en compte certaines spécificités des mesures actuellement disponibles expérimentalement. Nous ne nous sommes donc pas intéressés aux questions fondamentales de métrologie quantique où l'intrication, tant au niveau des entrées que des mesures en sortie, joue un rôle crucial [BD11, DDKG12]. Nous nous sommes donc appuyés sur le maximum de vraisemblance, un principe bien adapté aux chaînes de Markov à états cachés [DMOH11] et relié, via des développements asymptotiques d'intégrales de Laplace, à l'estimation bayésienne [CMR05].

Cette thèse a été réalisée à partir du mois d'octobre 2013 au sein du Centre Automatique et Systèmes de Mines ParisTech, sous la direction de Pierre Rouchon, qui a développé dans ce laboratoire une activité sur l'analyse mathématique et le contrôle des systèmes quantiques ouverts. Nous avons travaillé en collaboration avec deux laboratoires expérimentaux parisiens. Citons tout d'abord le laboratoire Kastler-Brossel et son expérience appelée "boîte à photons", déménagée de l'École Normale Supérieure au Collège de France au cours de cette thèse. Nos interlocuteurs principaux au sein de ce laboratoire furent Igor Dotsenko et Jean-Michel Raimond, que nous remercions particulièrement pour leur aide précieuse et pour les trajectoires de mesure qu'ils ont produites et que nous avons pu

exploiter. Le second laboratoire avec lequel nous avons collaboré est le laboratoire Pierre Aigrain, également rattaché à l'École Normale Supérieure, et plus particulièrement l'équipe de Benjamin Huard, qui a réalisé l'expérience de qubit supraconducteur que nous avons étudiée au cours de cette thèse. Nous les remercions également pour leur très étroite collaboration.

Durant la deuxième année de ma thèse, cette collaboration avec l'équipe de Benjamin Huard s'est concrétisée par la création du groupe de recherche Quantic, affilié à Inria, l'École Normale Supérieure, Mines ParisTech, le CNRS et l'Université Pierre et Marie Curie, avec comme membres permanents Benjamin Huard, Zaki Leghtas, François Mallet, Maziar Mirrahimi (chef de ce projet Inria), Pierre Rouchon et Alain Sarlette.

Durant cette thèse, j'ai participé à l'écriture de 5 publications, dont je liste ici le contenu et ma contribution personnelle :

- *Parameter estimation from measurements along quantum trajectories* [SCIB⁺15], déposé en mars 2015 sur arXiv, et accepté à la *Conference on Decision and Control 2015*, dont j'ai rédigé une grande partie ;
- *Quantum state tomography with non-instantaneous measurements, imperfections and decoherence* [SCID⁺16], déposé en octobre 2015 sur arXiv, et publié en 2016 dans le volume **93** de *Physical Review A*, dont j'ai également rédigé une grande partie ;
- *Observing quantum state diffusion by heterodyne detection of fluorescence* [CISB⁺16], déposé en novembre 2015 sur arXiv, et publié en 2016 dans le volume **6** de *Physical Review X*, dans lequel j'ai participé à l'élaboration du filtre quantique et à son implémentation numérique ;
- *Using Spontaneous Emission of a Qubit as a Resource for Feedback Control* [CIJC⁺16], déposé en février 2016 sur arXiv, et publié en 2016 dans le volume **117** de *Physical Review Letter*, dans lequel j'ai travaillé sur le passage à l'équation différentielle stochastique en boucle fermée, détaillé dans les suppléments ;
- *Asymptotic expansions of Laplace integrals for quantum state tomography* [SR16], déposé en juillet 2016 sur arXiv, qui est un travail nettement plus mathématique que les précédents et où ma contribution porte sur la moitié.

Ce manuscrit de thèse reprend principalement le contenu des papiers [SCIB⁺15, SCID⁺16, SR16]. Dans la première de ces trois publications, nous nous sommes intéressés à la tomographie de processus quantique par Maximum de Vraisemblance. Ici, il n'y avait qu'un paramètre réel à estimer. Ces données que nous exploitons sont des trajectoires de mesure, chaque trajectoire formant une série de résultats de mesure acquis successivement sur un même système, dans retour à l'état initial entre les mesures. Après avoir rappelé que les filtres quantiques sont robustes par rapport à l'état initial, nous avons utilisé le formalisme des filtres particuliers, qui permet de tester simultanément plusieurs valeurs de paramètre différentes, pour montrer la pertinence de l'estimation de paramètre par Maximum de Vraisemblance, même lorsque l'état initial n'est pas parfaitement connu. Nous avons ensuite explicité l'équation différentielle stochastique régissant un filtre particulier construit pour un système à mesures diffusives, puis avons présenté un premier schéma de discrétisation permettant de traiter les systèmes diffusifs de la même façon que les systèmes à mesures discrètes. Enfin, nous avons présenté une validation expérimentale, en estimant grâce aux filtres particuliers l'efficacité de détection du canal de mesure hétérodyne d'un qubit supraconducteur.

Ensuite, dans [SCID⁺16], nous nous sommes intéressés à la tomographie d'état quantique par Maximum de Vraisemblance. En nous plaçant d'abord dans le cadre des systèmes

à mesures discrètes, nous avons, en utilisant les applications de Kraus adjointes et la notion d'état adjoint, transformé la fonction de log-vraisemblance en une expression très facilement calculable, somme de logarithmes de produits scalaires entre les états adjoints de chaque trajectoire et la matrice densité testée. Après avoir rappelé le schéma de discrétisation utilisé pour les systèmes diffusifs, nous explicitons les conditions nécessaires et suffisantes que vérifie la matrice densité $\hat{\rho}_{ML}$ maximisant la fonction de log-vraisemblance. Nous donnons également la formule permettant de calculer, lorsque le nombre de trajectoires de mesure prises en compte est grand et lorsque la matrice $\hat{\rho}_{ML}$ est de rang plein ou de noyau uni-dimensionnel, une estimation de la variance de l'estimation de la valeur moyenne de tout observable A . Enfin, nous illustrons ces résultats par deux validations expérimentales : l'une montrant la détection de l'injection d'un photon dans un système à temps discret (boîte à photons), et l'autre montrant une tomographie d'état quantique validée par une estimation réalisée en moyennant les signaux de mesure, dans un système à mesures diffusives.

Enfin, dans [SR16], nous nous intéressons à la tomographie d'état quantique. L'objectif du papier est d'apporter un développement asymptotique, lorsque le nombre Ω de trajectoires de mesure prises en compte tend vers l'infini, de la moyenne et de la variance Bayésiennes de la valeur moyenne d'un observable A sur l'ensemble des matrices densité admissibles. Nous montrons que la moyenne Bayésienne d'un observable converge vers $\text{Tr}(\hat{\rho}_{ML}A)$ (où $\hat{\rho}_{ML}$ est l'estimation par Maximum de Vraisemblance), lorsque Ω tend vers l'infini. Puis, nous donnons la formule de la variance Bayésienne, liée à la variance de l'estimation par Maximum de Vraisemblance, de la valeur moyenne de tout observable, même lorsque $\hat{\rho}_{ML}$ a un noyau de rang supérieur ou égal à 2. Nous réalisons cela en utilisant le développement asymptotique d'intégrales de Laplace, et, lorsque $\hat{\rho}_{ML}$ est à la surface de cet ensemble, en paramétrant l'ensemble des matrices densité autour de $\hat{\rho}_{ML}$. Nous démontrons également les conditions d'optimalité que nous avons énoncées dans [SCID+16] pour $\hat{\rho}_{ML}$.

Nous avons conservé dans cette thèse les apports de ces trois publications. Après un premier chapitre où nous introduisons la structure des modèles d'évolution des systèmes quantiques ouverts, et où nous précisons les notations utilisées, nous détaillons dans le deuxième chapitre notre travail sur la tomographie d'état quantique, en reprenant les acquis de [SCID+16], complétés par les résultats récemment présentés dans [SR16] (la formule de la variance de l'estimation de la valeur moyenne d'un observable existe désormais pour toute matrice densité $\hat{\rho}_{ML}$ produite par le Maximum de Vraisemblance), et par l'exposé d'une méthode de maximisation de la log-vraisemblance plus rapide, lorsque les états adjoints de chaque trajectoire sont tous diagonaux dans la même base orthogonale. Le troisième chapitre reprend les résultats de [SCIB+15], en présentant une seconde validation expérimentale - l'estimation du temps de relaxation T_1 du qubit - qui met en relief la nécessité d'un raffinement du schéma de discrétisation des systèmes diffusifs initialement utilisé. Enfin, le dernier chapitre présente un développement qui n'a pas encore été publié, à savoir l'utilisation de la méthode de l'adjoint pour l'estimation par Maximum de Vraisemblance d'un paramètre évoluant dans un ensemble de grande dimension. Nous illustrons ces formules par l'estimation des coefficients apparaissant dans la mesure par un atome d'un champ électromagnétique, chaque paire de coefficients correspondant à un nombre n de photons, avec n entre 0 et 15. Nous donnons pour chaque coefficient son estimation et un intervalle d'erreur de cette estimation.

Enfin, nous donnons en annexe une présentation des systèmes expérimentaux qui nous ont servis à valider nos calculs, une série de développements mathématiques annexes, une traduction de [SR16], et, finalement, un ajout des papiers [CISB+16] et [CIJC+16] que

nous avons contribué à écrire, avec leurs suppléments.

Chapitre 2

Structure des modèles

2.1 Système isolé et mesures parfaites : fonctions d'onde

Nous parlons d'abord dans cette section d'un système quantique isolé, c'est-à-dire qui n'interagit pas avec l'extérieur¹ ; puis d'un système sur lequel toute mesure nouvellement appliquée est connue parfaitement par l'expérimentateur. Nous voyons notamment que ces deux types de systèmes peuvent être décrits par des fonctions d'onde.

2.1.1 Propriétés des systèmes physiques classiques

Lorsqu'il s'agit de modéliser des systèmes physiques à une échelle où l'on peut considérer que leur évolution est convenablement décrite par la physique classique, la représentation que nous choisissons pour décrire son état présente quelques caractéristiques générales, que nous détaillons ici pour mieux comprendre, par contraste, certaines spécificités des systèmes quantiques.

Nature d'un état classique

Un état classique est couramment décrit comme évoluant au sein d'un espace $\mathcal{S}$ isomorphe à $\mathbb{R}^q$, ou dans un de ses sous-ensembles : $\mathcal{S} \subset \mathbb{R}^q$. A un instant donné t , cet état est noté $\mathbf{x}(t) \in \mathcal{S}$. Cet état peut par exemple décrire la position du centre de masse d'un objet dans l'espace, auquel cas $\mathcal{S}$ sera un sous-ensemble de $\mathbb{R}^3$.

De plus, supposons que nous disposons de deux systèmes classiques d'états respectifs $\mathbf{x}_1(t) \in \mathcal{S}_1$ et $\mathbf{x}_2(t) \in \mathcal{S}_2$ à l'origine isolés, et que nous voulons considérer au sein d'un même système (par exemple pour étudier ce qu'un couplage entre ces deux systèmes induirait). L'état de ce système composite $\mathbf{z}(t)$ s'écrit simplement comme un vecteur agrandi :

$$\mathbf{z}(t) = (\mathbf{x}_1(t), \mathbf{x}_2(t)) \in \mathcal{S}_1 \times \mathcal{S}_2 \subset \mathbb{R}^{q_1+q_2},$$

où q_1 et q_2 sont les dimensions respectives des deux espaces d'état. On remarque qu'avec ces notations, le fait de vouloir attribuer à tout instant un état à l'un ou l'autre des systèmes *considérés isolément* garde un sens, puisqu'il suffira, pour l'obtenir, d'extraire soit les premières, soit les dernières dimensions de l'état $\mathbf{z}(t)$. Ainsi, deux systèmes classiques, même après une longue évolution commune où de nombreuses influences réciproques se sont entremêlées, restent toujours *séparables*.

1. Dont l'expérimentateur lui-même fait partie.

Déterminisme de l'évolution

Les équations d'évolution newtonienne ont une nature strictement déterministe. En effet, si on choisit deux instants successifs t_1 et t_2 , on peut définir une application Ξ_{t_1, t_2} telle que, si l'état du système au temps t_1 est $\mathbf{x}(t_1) = \mathbf{x}_1$, l'état du système au temps t_2 devient alors nécessairement :

$$\mathbf{x}(t_2) = \mathbf{x}_2 = \Xi_{t_1, t_2}(\mathbf{x}_1)$$

Mesures et altérations

A priori, un système classique peut être observé sans que cela n'altère en retour l'état de ce système. Bien sûr, cela ne veut pas dire que certains systèmes sont construits de telle sorte que leur observation soit intimement liée à l'existence d'une loi de commande, mais cela relève alors d'une contrainte extérieure qui n'est pas inhérente aux lois physiques classiques, mais qui résulte des possibilités matérielles restreintes de l'expérimentateur.

Incertitude sur l'état

Une éventuelle incertitude que l'expérimentateur aurait sur l'état réel $\mathbf{x}(t)$ du système à l'instant t peut se traduire par une certaine densité de probabilité $p_t(\mathbf{x})$ attribuée à chaque valeur possible $\mathbf{x} \in \mathcal{S}$ de l'état. Lorsque la valeur de l'état est connue, cette distribution est simplement un Dirac centré sur celle-ci.

Ce qui est remarquable dans les systèmes classiques, c'est que l'ensemble des valeurs possibles (les vecteurs $\mathbf{x}$ telles que $p_t(\mathbf{x}) \neq 0$) "n'interfèrent" pas entre elles. Toutes ces valeurs évoluent sans considération pour toutes les autres valeurs possibles de l'état. Formellement, si nous définissons un système que l'expérimentateur connaît avec la densité de probabilité $p_{t_1}(\mathbf{x})$ à l'instant t_1 , alors la valeur moyenne que renvoie un observateur $\mathbf{x} \mapsto A(\mathbf{x}) \in \mathbb{R}$ sur l'état passe de :

$$\langle A \rangle_{t_1} = \int_{\mathcal{S}} A(\mathbf{x}) p_{t_1}(\mathbf{x}) d\mathbf{x}$$

à l'instant t_1 à :

$$\langle A \rangle_{t_2} = \int_{\mathcal{S}} A(\Xi_{t_1, t_2}(\mathbf{x}_1)) p_{t_1}(\mathbf{x}) d\mathbf{x}$$

à l'instant t_2 .

Passage aux systèmes quantiques

Lorsque nous considérons des systèmes où les phénomènes de la mécanique quantique sont sensibles, les propriétés que nous venons d'énoncer ne s'appliquent plus. Les états quantiques ont une interprétation naturellement probabiliste. Ainsi, ils ne sont pas représentés de la même façon que les états de systèmes classiques. Leur évolution n'est pas toujours déterministe, puisque les mesures qu'on peut effectuer sur un système quantique renvoient la réalisation d'une variable aléatoire, et altèrent en retour l'état du système.

Dans ce chapitre, nous commençons par détailler la représentation par *fonction d'onde* permettant de décrire l'état d'un système quantique *isolé*, ce qui désigne ici un système qui n'interagit avec des systèmes extérieurs que lors de réductions du paquet d'ondes dont le résultat est parfaitement connu. L'étude des caractéristiques des systèmes quantiques ouverts, ou dont le résultat des mesures est imparfaitement connu est l'objet de la section suivante, avec l'*opérateur densité* comme représentation de l'état qui remplace et étend la notion de fonction d'onde.

2.1.2 Systèmes quantiques isolés

Pour une très bonne introduction à la notion d'état quantique, voir [CTDL77]. La nature d'un état quantique est intimement liée à celle de probabilité. La notion de probabilité est à la base de la représentation d'un état quantique, à cela près que nous nous attachons, dans ce cas, à des *amplitudes* de probabilité ; celles-ci font apparaître une phase dans le plan complexe qui permet de traduire dans l'état quantique le caractère ondulatoire des particules étudiées au sein des systèmes quantiques.

L'état d'un système quantique isolé, appelé fonction d'onde, est donc un vecteur d'un espace de Hilbert complexe noté $\mathcal{H}$, muni du produit scalaire Euclidien. Chaque dimension de cet espace correspond à un état physique possible du système quantique, et la composante qui lui est associée dans la fonction d'onde détermine l'amplitude de probabilité correspondant à cet état possible. L'espace de Hilbert $\mathcal{H}$ peut donc être, selon la nature des états physiques considérés comme possibles, de dimension finie ou infinie - nous travaillons ici avec un espace de Hilbert de dimension finie $N_{\mathcal{H}}$. Cette fonction d'onde est souvent notée $|\psi\rangle$, ou $|\psi(t)\rangle$ lorsque la dépendance en temps doit être soulignée : nous adoptons le point de vue de Schrödinger, et c'est l'état et non les observables qui dépendent du temps. La notation $|\psi\rangle$ représente un vecteur colonne, et $\langle\psi|$ le vecteur ligne qui en est l'adjoint. Le produit scalaire entre deux états $|\psi\rangle$ et $|\phi\rangle$ est donc $\langle\psi|\phi\rangle$. La norme d'une fonction d'onde est toujours égale à 1.

Nous prenons trois exemples pour illustrer la façon avec laquelle est construite cette fonction d'onde², puis nous donnons l'équation de Schrödinger qui permet de représenter l'évolution, déterministe, des systèmes isolés.

Fonction d'onde et système à deux niveaux

Ici, le système est une particule qui ne peut prendre que deux états physiques différents. Souvent, ces deux états correspondent à deux niveaux d'énergie, et on note ces deux états $|e\rangle = (1, 0)^T$ et $|g\rangle = (0, 1)^T$, l'état excité et l'état au repos respectivement. L'espace de Hilbert du système $\mathcal{H}$ est donc isomorphe à $\mathbb{C}^2$. L'ensemble des états possibles $|\psi\rangle$ s'écrit donc :

$$\left\{ |\psi\rangle = c_e|e\rangle + c_g|g\rangle / |c_e|^2 + |c_g|^2 = 1 \right\}$$

D'un point de vue informationnel, ces systèmes sont appelés *qubit*, par analogie au bit classique qui stocke une information binaire.

Fonction d'onde et population

Un autre type habituel de systèmes en physique classique consiste à modéliser une population de particules identiques. Dans ce cas, l'espace de Hilbert associé est alors $\mathbb{C}^{\mathbb{N}}$, chaque dimension représentant un nombre différent de particules.

La fonction d'onde associée au nombre $n > 0$ de particules est notée $|n\rangle$. L'ensemble de fonctions d'onde $(|n\rangle)_{n \geq 0}$ constitue une base orthonormale de l'espace de Hilbert associé, appelée base de Fock. On note $\mathbf{a}$ l'opérateur d'annihilation associé au système, vérifiant $\mathbf{a}|n\rangle = \sqrt{n}|n-1\rangle$:

$$\mathbf{a} = \sum_{n \geq 1} \sqrt{n} |n-1\rangle \langle n|$$

2. Le formalisme complet du qubit représenté par sphère de Bloch nécessitant l'introduction de la notion de matrice densité, nous y reviendrons en sous-section 2.2.2.

L'opérateur création est simplement $\mathbf{a}^\dagger$. Enfin, l'opérateur $\mathbf{N} = \mathbf{a}^\dagger \mathbf{a}$ est l'opérateur population, qui vérifie $\mathbf{N}|n\rangle = n|n\rangle$.

Pour analyser numériquement ces systèmes, l'espace de Hilbert est la plupart du temps tronqué à un nombre maximal $N_{\max}$ de particules, la base orthonormale en résultat étant $(|0\rangle, \dots, |N_{\max}\rangle)$.

Fonction d'onde et position dans l'espace

Supposons ici que le système quantique étudié est une particule isolé (au comportement ondulatoire), et que nous souhaitons décrire l'évolution de sa position. Ainsi, l'information que la fonction d'onde correspondante contient est l'amplitude de probabilité de telle ou telle position $\mathbf{x}$, vecteur à trois dimensions appartenant à $\mathcal{S}$, l'ensemble des positions autorisées.

Chaque position possible est ici représentée par un vecteur normé $|\mathbf{x}\rangle$ différent, chacun de ces vecteurs étant orthogonal à tous les autres. L'ensemble de ces vecteurs représente une sorte de base de l'espace de Hilbert $\mathcal{H}$, lequel est donc de dimension infinie. On peut donc écrire la fonction d'onde correspondant à l'état de la particule comme :

$$|\psi\rangle = \iiint_{\mathcal{S}} \bar{\psi}(\mathbf{x}) |\mathbf{x}\rangle d\mathbf{x}$$

où $\mathbf{x} \mapsto \bar{\psi}(\mathbf{x})$ est une fonction complexe sur $\mathcal{S}$.

La densité de probabilité d'observer la particule en position $\mathbf{x}$ est alors $|\bar{\psi}(\mathbf{x})|^2$. Cette notation impose donc à la fonction d'onde d'être toujours un vecteur normé, puisque :

$$\begin{aligned} \langle\psi|\psi\rangle &= \left(\iiint_{\mathcal{S}} \bar{\psi}^*(\mathbf{x}) \langle\mathbf{x}| d\mathbf{x} \right) \left(\iiint_{\mathcal{S}} \bar{\psi}(\mathbf{x}') |\mathbf{x}'\rangle d\mathbf{x}' \right) \\ &= \iiint_{\mathcal{S}} \iiint_{\mathcal{S}} \bar{\psi}^*(\mathbf{x}) \bar{\psi}(\mathbf{x}') \langle\mathbf{x}|\mathbf{x}'\rangle d\mathbf{x} d\mathbf{x}' \\ &= \iiint_{\mathcal{S}} |\bar{\psi}(\mathbf{x})|^2 d\mathbf{x} = 1 \end{aligned}$$

car $\langle\mathbf{x}|\mathbf{x}'\rangle = \delta(\mathbf{x} - \mathbf{x}')$.

Si nous voulons maintenant considérer, non pas la position de la particule, mais sa quantité de mouvement, il nous faut introduire une autre base orthonormale de l'espace de Hilbert, constituée des vecteurs $|\mathbf{p}\rangle$, pour tout vecteur $\mathbf{p}$ de $\mathbb{R}^3$. La reparamétrisation de la fonction d'onde se fait à l'aide du produit scalaire :

$$\langle\mathbf{x}|\mathbf{p}\rangle = \frac{e^{i\hbar^{-1}\mathbf{x}\cdot\mathbf{p}}}{(2\pi\hbar)^{3/2}}$$

Évolution déterministe : l'équation de Schrödinger

Au sein des systèmes quantiques isolés, l'évolution est fondamentalement déterministe. Cette évolution est décrite par l'équation de Schrödinger, et conserve certaines propriétés des évolutions classiques. Cette équation fait intervenir à tout instant t un opérateur Hermitien de $\mathcal{H}$, noté $\hat{H}(t)$, qui décrit l'évolution infinitésimale de la fonction d'onde à cet instant. Cet opérateur est appelé Hamiltonien du système, et l'évolution qui en découle est souvent appelée évolution Hamiltonienne.

L'équation de Schrödinger s'écrit alors :

$$d|\psi(t)\rangle = -i \frac{\hat{H}(t)}{\hbar} |\psi(t)\rangle dt \quad (2.1)$$

Il est facile de constater que cette équation préserve la norme de $|\psi(t)\rangle$, ce qui est requis par la nature même de la fonction d'onde, puisque :

$$\begin{aligned} d(\langle\psi(t)|\psi(t)\rangle) &= (d|\psi(t)\rangle)^\dagger |\psi(t)\rangle + \langle\psi(t)| (d|\psi(t)\rangle) \\ &= \langle\psi(t)| \left(i \frac{\hat{H}(t)}{\hbar} - i \frac{\hat{H}(t)}{\hbar} \right) |\psi(t)\rangle dt \\ &= 0 \end{aligned}$$

Pour un système quantique ne subissant que l'influence d'un Hamiltonien, décrite par l'équation (2.1), entre deux instants successifs t_0 et t_1 , il peut être intéressant d'introduire un opérateur de $\mathcal{H}$ appelé *opérateur d'évolution* du système entre ces deux instants, noté $U(t_0, t_1)$, qui permet d'obtenir :

$$|\psi(t_1)\rangle = U(t_0, t_1) |\psi(t_0)\rangle \quad (2.2)$$

Cet opérateur est obtenu en intégrant l'équation de Schrödinger entre ces deux instants, en remarquant que celle-ci impose :

$$U(t, t + dt) = I_{\mathcal{H}} - i \frac{\hat{H}(t)}{\hbar} dt$$

On obtient alors :

$$U(t_0, t_1) = I_{\mathcal{H}} - \frac{i}{\hbar} \int_{t_0}^{t_1} \hat{H}(s) U(t_0, s) ds \quad (2.3)$$

Entre autres propriétés de cet opérateur (décrites plus en détail dans [CTDL77, III, D-1-b]), nous insistons sur le fait que son existence implique la réversibilité de toute évolution Hamiltonienne : pour tout $U(t_0, t_1)$ généré par $t \in [t_0; t_1] \mapsto \hat{H}(t)$, il est possible de construire, pour un instant $t_2 > t_1$, une autre trajectoire Hamiltonienne $t \in [t_1; t_2] \mapsto \hat{H}'(t)$ construisant l'opérateur d'évolution $U(t_1, t_2)$ permettant d'obtenir :

$$|\psi(t_2)\rangle = U(t_1, t_2) U(t_0, t_1) |\psi(t_0)\rangle = |\psi(t_0)\rangle$$

L'évolution Hamiltonienne est donc, en tout cas théoriquement, toujours réversible, et aucune perte d'informations n'en découle.

Remarquons également que, puisque :

$$U^\dagger(t, t + dt) = U(t, t - dt) = U^{-1}(t, t + dt),$$

un opérateur d'évolution est unitaire.

2.1.3 Systèmes ouverts à mesures parfaites

L'évolution Hamiltonienne n'apporte en elle-même aucune information à l'expérimentateur. Lorsque celui-ci cherche à connaître l'état du système, il doit lui appliquer ce qu'on appelle une mesure quantique. Le système devient alors un système quantique ouvert. La valeur du résultat tiré de cette mesure est la réalisation d'une variable aléatoire dont la loi de probabilité dépend de l'état du système au moment de cette mesure. En retour, puisque l'état d'un système quantique a une nature probabiliste, ce processus de mesure l'altère de façon irréversible, et cette évolution dépend du résultat de la mesure obtenue.

Lorsque le résultat de ces mesures est parfaitement connu, le système peut être décrit par une fonction d'onde. C'est ce cas de figure que nous étudions dans cette sous-section, pour introduire ensuite le cadre plus général des mesures quantiques. Nous présentons d'abord le cas où la mesure est un processus discret, puis celui où le phénomène de mesure est continu et accompagné d'un bruit diffusif.

Mesure parfaite discrète

Notion d'observable Chaque mesure quantique est représentée mathématiquement par un opérateur de $\mathcal{H}$ appelé *observable*, noté $\hat{A} \in \mathcal{H}$. Notons que nous utilisons dans cette thèse le point de vue de Schrödinger ; les opérateurs de mesure $\hat{A}$ ne dépendent donc pas du temps. Une observable, en dimension finie, peut toujours être choisi Hermitien. Le troisième postulat de la physique quantique impose que la valeur d'une telle mesure soit nécessairement une valeur propre de l'opérateur $\hat{A}$ dans $\mathcal{H}$. Une première conséquence de ce postulat est que le nombre m de mesures possibles ne peut excéder $N_{\mathcal{H}}$, la dimension de l'espace de Hilbert.

Notons $(\lambda_j)_{j=1\dots m}$ ces valeurs possibles de mesure, formant le spectre de $\hat{A}$. Une propriété fondamentale des opérateurs hermitiens est que leurs sous-espaces propres respectifs sont orthogonaux deux à deux. Nous notons P_j le projecteur orthogonal associé à la valeur propre λ_j , ainsi :

$$\hat{A} = \sum_{j=1}^m \lambda_j P_j$$

Supposons que nous appliquions l'observable $\hat{A}$ ainsi définie sur l'état $|\psi\rangle$, et notons λ_μ le résultat de mesure ainsi produit. La probabilité d'obtenir ce résultat λ_μ est l'objet du quatrième postulat de la physique quantique ; cette probabilité, conditionnée au fait que l'état quantique juste avant la mesure est $|\psi\rangle$, vaut :

$$\mathbb{P}(\text{résultat } \lambda_\mu \text{ de la mesure } \hat{A} \mid \text{état } |\psi\rangle) = \langle \psi | P_\mu^\dagger P_\mu | \psi \rangle = \langle \psi | P_\mu | \psi \rangle \quad (2.4)$$

Là encore, il est facile de voir que la somme des probabilités de toutes les valeurs mesures est bien égale à un, puisque $\sum_{\mu=1}^m P_\mu^\dagger P_\mu = \sum_{\mu=1}^m P_\mu = I_{\mathcal{H}}$. Avant de montrer quel est l'effet d'une telle mesure sur l'évolution de l'état, nous donnons ici la formule de la valeur moyenne de l'observable sur l'état $|\psi\rangle$, notée $\langle \hat{A} \rangle_{|\psi\rangle}$:

$$\langle \hat{A} \rangle_{|\psi\rangle} = \sum_{j=1}^m \lambda_j \mathbb{P}(\text{résultat } \lambda_j \mid \text{état } |\psi\rangle) = \langle \psi | \hat{A} | \psi \rangle \quad (2.5)$$

Itération de l'état du système Supposons donc que la mesure effectuée par l'expérimentateur grâce à l'observable $\hat{A}$ a produit le résultat λ_μ associé au projecteur orthogonal P_μ , et que cette mesure a eu lieu entre les instants t et $t + \Delta t$, où Δt est un petit intervalle de temps bien inférieur aux temps caractéristiques du système. Ainsi, nous considérons qu'aucune évolution Hamiltonienne n'a eu lieu durant ce processus de mesure. Le cinquième postulat de la mécanique quantique nous indique que la mesure procède, entre ces deux instants, à une réduction du paquet d'ondes sur l'état originel $|\psi(t)\rangle$. En effet, on peut toujours décomposer l'état du système en $|\psi(t)\rangle = \sum_{j=1}^m c_j |\psi_j\rangle$, où $|\psi_j\rangle$, vecteur normé, appartient à l'espace image du projecteur P_j . $|\psi_j\rangle$ est donc associé au résultat de mesure λ_j , et à lui *seul*, puisque $|\psi_j\rangle = P_j |\psi_j\rangle$. Lorsque le résultat λ_μ est tiré lors de la mesure, l'état itéré ne peut plus avoir des composantes sur les sous-espaces images associés à des projecteurs P_j , tels que $j \neq \mu$. L'incertitude est "levée" sur toutes ces dimensions de $\mathcal{H}$. Ainsi, l'état est projeté sur le seul sous-espace image possible, celui de P_μ , et il devient :

$$|\psi(t + \Delta t)\rangle = \frac{P_\mu |\psi(t)\rangle}{\sqrt{\langle \psi | P_\mu | \psi \rangle}} = |\psi_\mu\rangle \quad (2.6)$$

On retrouve au dénominateur, sous une racine carrée, la probabilité qu'on avait a priori d'obtenir ce résultat de mesure, donnée par l'équation (2.4). Ici, cette probabilité vaut $|c_\mu|^2$.

Généralisation de la mesure discrète La mesure discrète parfaite a été présentée ci-dessus comme un phénomène quasi-instantané (avec Δt très petit), et en l'absence d'évolution Hamiltonienne. C'est ce qui permet aux matrices de mesures P_μ d'être des projecteurs orthogonaux, et donc d'être Hermitiens. Cependant, il se peut que, entre les deux instants t et $t + \Delta t$ pendant lesquels on veut connaître l'état du système, une évolution Hamiltonienne ait lieu avant, et une autre après le processus de mesure. Ces évolutions sont respectivement représentées par les opérateurs d'évolution U et V . Il est alors simple de remarquer que, en notant $M_\mu = VP_\mu U$, le processus de mesure discret est alors défini par

$$\begin{aligned} \mathbb{P}(\text{résultat } \lambda_\mu \text{ de la mesure } \hat{A} \mid \text{état } |\psi\rangle) &= \langle \psi | M_\mu^\dagger M_\mu | \psi \rangle \\ |\psi(t + \Delta t)\rangle &= \frac{M_\mu |\psi(t)\rangle}{\sqrt{\langle \psi | M_\mu^\dagger M_\mu | \psi \rangle}} \end{aligned} \quad (2.7)$$

Nous retrouvons des équations de forme identique à celles du paragraphe précédent, à ceci près que les matrices de mesure M_μ ne sont plus nécessairement Hermitiennes. Nous adopterons à partir de maintenant ce schéma de mesure lorsqu'il s'agira de mesures discrètes parfaites. Notons que cet ensemble de matrices de mesure $(M_j)_{j \in \{1, \dots, m\}}$ conserve la relation :

$$\sum_{j=1}^m M_j^\dagger M_j = I_{\mathcal{H}} \quad (2.8)$$

Équation différentielle stochastique : mesure parfaite inachevée

Dans la sous-section précédente, nous avons en réalité présenté un processus de mesure achevé, à la fin duquel une seule valeur de mesure émerge. Si nous notons T_{mes} le temps caractéristique de cette mesure, nous avons plus haut $T_{\text{mes}} \ll \Delta t$. Dans ce paragraphe, nous travaillons dans le régime inverse, et présenterons les résultats sous forme différentielle, entre deux instants t et $t + dt$ de sorte que $dt \ll T_{\text{mes}}$.

Le résultat d'une telle mesure n'est plus une des valeurs propres de l'observable, mais la valeur moyenne d'une observable à tout instant t , à laquelle est ajoutée un bruit quantique. Ce résultat est donc remplacé par un processus stochastique différentiel, définissable à tout instant t . Nous nous limitons au cas des systèmes diffusifs, au sein desquels ce bruit est un bruit Gaussien, représenté par le processus de Wiener W_t , dont les propriétés sont rappelées en annexe B.3. Les raisonnements physiques permettant de parvenir aux formules suivantes peuvent être consultés dans [HR06, 4.3]. Pour un traitement plus complet des systèmes diffusifs, nous renvoyons le lecteur à [BG09].

Canal de mesure unique Nous commençons par aborder le cas où une seule mesure diffusive est effectuée. L'opérateur qui lui est associé - qui n'est pas nécessairement Hermitien - est noté L , le processus de mesure dy_t est généré par l'équation différentielle stochastique suivante :

$$dy_t = dW_t + \langle \psi(t) | (L + L^\dagger) | \psi(t) \rangle dt \quad (2.9)$$

A chaque instant, la valeur moyenne de ce processus de mesure est donc donnée par :

$$\langle dy_t \rangle = \langle \psi(t) | (L + L^\dagger) | \psi(t) \rangle dt \quad (2.10)$$

Le système est en retour affecté par cette mesure, et on peut également définir cette évolution par une équation différentielle stochastique. Si on prend en compte un éventuel Hamiltonien $\hat{H}(t)$ agissant simultanément sur le système, cette équation s'écrit, sous une forme analogue à celle de l'équation (2.7) concernant les mesures discrètes, mais avec une matrice M_{t,dy_t} proche de l'identité :

$$|\psi(t + dt)\rangle = \frac{M_{t,dy_t} |\psi(t)\rangle}{\sqrt{\langle \psi(t) | M_{t,dy_t}^\dagger M_{t,dy_t} | \psi(t) \rangle}} \quad (2.11)$$

où l'opérateur de mesure différentiel M_{t,dy_t} vaut :

$$M_{t,dy_t} = I_{\mathcal{H}} - \left(\frac{i}{\hbar} \hat{H}(t) + \frac{L^\dagger L}{2} \right) dt + L dy_t \quad (2.12)$$

On peut également définir la loi de probabilité des mesures diffusives dy_t à partir de cette matrice M_{t,dy_t} . L'équation (2.9) nous enseigne que pour un intervalle de temps dt infinitésimal, la mesure dy_t est de l'ordre $\sqrt{dt}$. Définissons donc le processus de mesure entre t et $t + dt$ en écrivant que $dy_t = s\sqrt{dt}$. Nous montrerons en annexe B.4 (en utilisant le formalisme des opérateurs densité) que, si ds est un élément infinitésimal de s , nous pouvons écrire :

$$\mathbb{P}(dy_t \in [s; s + ds] \sqrt{dt} \text{ à l'instant } t \mid \text{état } |\psi(t)\rangle) = \langle \psi | M_{t,s\sqrt{dt}}^\dagger M_{t,s\sqrt{dt}} | \psi \rangle \frac{e^{-s^2}}{\sqrt{2\pi}} ds \quad (2.13)$$

Le terme en e^{-s^2} ne dépendant pas de l'état du système, il peut être incorporé dans l'expression de $M_{t,s\sqrt{dt}}$. On peut ainsi écrire :

$$\tilde{M}_{t,dy_t} = \frac{e^{-\frac{(dy_t)^2}{2dt}}}{(2\pi)^{\frac{1}{4}}} \left(I_{\mathcal{H}} - \left(\frac{i}{\hbar} \hat{H}(t) + \frac{L^\dagger L}{2} \right) dt + L dy_t \right)$$

et ainsi pouvoir écrire :

$$|\psi(t + dt)\rangle = \frac{\tilde{M}_{t,dy_t} |\psi(t)\rangle}{\sqrt{\langle \psi(t) | \tilde{M}_{t,dy_t}^\dagger \tilde{M}_{t,dy_t} | \psi(t) \rangle}}$$

$$\mathbb{P}(dy_t \in [s; s + ds] \sqrt{dt} \text{ à l'instant } t \mid \text{état } |\psi(t)\rangle) = \langle \psi | \tilde{M}_{t,s\sqrt{dt}}^\dagger \tilde{M}_{t,s\sqrt{dt}} | \psi \rangle ds$$

ce qui est de la même forme que celle des équations (2.7) régissant les systèmes à mesures discrètes.

Canaux de mesures multiples Plusieurs mesures diffusives peuvent s'effectuer simultanément ; on appelle alors chacune des trajectoires produites des *canaux de mesure*. Notons M le nombre de ces mesures, indexées par $\nu = 1 \dots M$. Chaque mesure est représentée

par l'opérateur L_ν , et produit la trajectoire dy_t^ν . La valeur moyenne d'une telle mesure est indépendante des autres canaux de mesure, on retrouve donc :

$$\begin{aligned} dy_t^\nu &= dW_t^\nu + \langle \psi(t) | (L_\nu + L_\nu^\dagger) | \psi(t) \rangle dt \\ \langle dy_t^\nu \rangle &= \langle \psi(t) | (L_\nu + L_\nu^\dagger) | \psi(t) \rangle dt \end{aligned} \quad (2.14)$$

Dans ces expressions, dW_t^ν est un processus de Wiener propre au canal de mesure ν , qui vérifie les propriétés habituelles des processus de Wiener rappelées plus haut. Ces processus sont indépendants entre eux, de sorte que leur covariance est donnée par $\mathbb{E}(dW_t^\nu dW_t^{\nu'}) = \delta_{\nu,\nu'} dt$. Notons $\mathbf{dy}_t$ le vecteur comportant toutes ces mesures : $\mathbf{dy}_t = (dy_t^1, \dots, dy_t^M)$. L'évolution que la fonction d'onde décrit sous tous ces canaux de mesure est donnée par l'équation différentielle stochastique :

$$\begin{aligned} |\psi(t+dt)\rangle &= \frac{M_{t,\mathbf{dy}_t} |\psi(t)\rangle}{\sqrt{\langle \psi(t) | M_{t,\mathbf{dy}_t}^\dagger M_{t,\mathbf{dy}_t} | \psi(t) \rangle}} \\ M_{t,\mathbf{dy}_t} &= I_{\mathcal{H}} - \left(\frac{i}{\hbar} \hat{H}(t) + \sum_{\nu=1}^M \frac{L_\nu^\dagger L_\nu}{2} \right) dt + \sum_{\nu=1}^M L_\nu dy_t^\nu \end{aligned} \quad (2.15)$$

Remarquons que ces équations d'évolution ont une forme analogue à celles qui traduisent la réduction du paquet d'onde lors d'une mesure complètement achevée (appelée désormais mesure *discrète*) ; par conséquent, ces équations préservent la norme de la fonction d'onde à chaque instant.

Donnons enfin, de manière analogue à celle de l'équation (2.13), une loi de probabilité sur la valeur des mesures diffusives $\mathbf{dy}_t$ entre les instants t et $t+dt$. Nous procédons à la même paramétrisation, à savoir que $dy_t^\nu = s^\nu \sqrt{dt}$ pour tout canal de mesure $\nu = 1 \dots M$, et que le vecteur $\mathbf{s} = (s^1, \dots, s^M)$. L'expression de cette probabilité est donnée par :

$$\begin{aligned} \mathbb{P}(\mathbf{dy}_t \in [\mathbf{s}; \mathbf{s} + d\mathbf{s}] \sqrt{dt} \mid |\psi(t)\rangle) &= \prod_{\nu=1}^M \mathbb{P}(dy_t^\nu \in [s^\nu; s^\nu + ds^\nu] \sqrt{dt} \mid |\psi(t)\rangle) = \\ &= \frac{\langle \psi | M_{t,\mathbf{s}\sqrt{dt}}^\dagger M_{t,\mathbf{s}\sqrt{dt}} | \psi \rangle}{(2\pi)^{\frac{M}{2}}} \exp\left(-\frac{\|\mathbf{s}\|^2}{2}\right) \prod_{\nu=1}^M ds^\nu \end{aligned} \quad (2.16)$$

où $\|\mathbf{s}\|$ est la norme euclidienne de $\mathbf{s}$.

2.1.4 Ouverture

Nous avons détaillé ici les équations permettant de prendre en compte les évolutions déterministe et stochastique de l'état d'un système quantique, lorsque le résultat des mesures qui lui sont appliquées est connu parfaitement par l'expérimentateur. Dans ce cas, l'état du système est représenté par une fonction d'onde.

Comparons l'évolution d'un système dont l'état est $|\psi\rangle \in \mathcal{H}$, et d'un système évoluant dans le même espace de Hilbert, et dont l'état est, au même instant initial t_0 , représenté par $|\phi\rangle = \lambda|\psi\rangle$, avec λ nombre complexe unitaire. Voyons comment ces deux états évoluent lors des mêmes processus physiques :

- une évolution Hamiltonienne jusqu'à l'instant t_1 , représentée par l'opérateur d'évolution $U(t_0, t_1)$ transforme ces états respectivement en :

$$|\psi(t_1)\rangle = U(t_0, t_1) |\psi\rangle \text{ et } |\phi(t_1)\rangle = U(t_0, t_1) (\lambda|\psi\rangle)$$

L'évolution Hamiltonienne préserve donc la relation de proportionnalité $|\phi\rangle = \lambda|\psi\rangle$.

- une mesure discrète pure, réalisée avec l'observable $\hat{A}$, présente tout d'abord les mêmes statistiques de mesure a priori, puisque pour tout projecteur orthogonal M_μ , on a $\langle \phi | M_\mu^\dagger M_\mu | \phi \rangle = \langle \psi | M_\mu^\dagger M_\mu | \psi \rangle$. En outre, il est facile de montrer, à partir de l'équation (2.7), qu'une mesure discrète parfaite conserve la relation de proportionnalité entre deux états.
- enfin, une mesure diffusive stochastique décrite par les équations (2.14) et (2.15) génère sur ces deux systèmes la même statistique de mesure dy_t' sur chacun de ses canaux, et la mise à jour de l'état qui en découle préserve également la relation de proportionnalité.

Ainsi, ces deux états $|\psi\rangle$ et $\lambda|\psi\rangle$ représentent en réalité la même situation physique. Ce qui est important dans l'état $|\psi\rangle$, c'est le $\mathbb{C}$ -espace vectoriel que ce vecteur engendre. Toute l'information sur cet état est donc contenue l'expression de son projecteur orthogonal associé $|\psi\rangle\langle\psi|$. Cette considération permet de comprendre l'introduction des matrices densité, faite dans la section suivante.

Nous montrons en annexe B.1 que ces considérations ont pour conséquence de pouvoir imposer au Hamiltonien $\hat{H}(t)$ d'être de trace nulle, ce que nous aurons par la suite.

2.2 Systèmes ouverts à mesures imparfaites : matrices densité

2.2.1 Motivations

Jusqu'à maintenant, nous avons uniquement traité le cas de système dont l'état quantique était parfaitement connu. Les mesures qui lui ont été appliquées étaient supposées parfaites, et renvoyaient au physicien qui en tenait compte des résultats qui ne pouvaient être partiellement ou complètement erronés. A l'inverse, le physicien peut rencontrer des imperfections de mesure, imperfections classiques introduisant une incertitude statistique dans notre connaissance du résultat de la mesure, et donc de l'état du système. De plus, nous n'avons pas non plus abordé la possibilité de mesures non lues, à la fin desquelles le résultat de la mesure est considéré comme "omis" par l'expérimentateur (ce qui est un cas critique d'imperfections de mesure).

Nous allons montrer par un exemple en quoi la représentation des états quantiques à l'aide du formalisme des fonctions d'onde n'est pas apte à prendre en compte ces mesures appelées respectivement imparfaites et non lues. Prenons un espace de Hilbert $\mathcal{H}$ isomorphe à $\mathbb{C}^2$, qui est celui, par exemple, d'un état de spin à deux niveaux. Admettons que son état initial, que l'expérimentateur connaît, est $|\psi\rangle_0 = (1, 0)$. Appliquons à cet état une première mesure, dont l'observable associée est $\hat{M}_x$ décrit par :

$$\hat{M}_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \hat{M}_x = |+\rangle\langle+| - |-\rangle\langle-|$$

où $|+\rangle = (1, 1)/\sqrt{2}$ et $|-\rangle = (1, -1)/\sqrt{2}$ sont deux vecteurs propres de norme 1 de l'opérateur $\hat{M}_x$ pour les valeurs propres 1 et -1 respectivement. Cette mesure projette donc l'état soit sur $|+\rangle$, soit sur $|-\rangle$, avec la même probabilité $1/2$. Appliquons ensuite au système une seconde mesure dont l'observable est noté $\hat{M}_z$. Si l'expérimentateur n'a pas lu le résultat de la première mesure, il doit, pour modéliser la seconde mesure sur le système, suivre les deux hypothèses *séparément* et garder à l'esprit qu'elles ont la même probabilité d'être advenues.

En l'occurrence, définissons le second observable $\hat{M}_z$ comme :

$$\hat{M}_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

Quelque soit le résultat de la première mesure, et donc quelque soit l'état $|+\rangle$ ou $|-\rangle$ du système, cette observable donne comme résultat de mesure 1 et -1 avec la même probabilité. Or, si la première mesure n'avait simplement pas eu lieu, l'état serait resté $|\psi\rangle_0 = (1, 0)$, et la mesure effectuée avec $\hat{M}_z$ n'aurait pu produire que le résultat 1.

Cette expérience de pensée montre que la mesure non lue, dans le monde quantique, a une réalité physique qui la différencie de l'absence simple de mesure, et cela parce qu'un état quantique représente une loi de probabilité. Il est malaisé de tenir compte de ce phénomène en représentant l'état par une fonction d'onde, car cela nous conduit à des séparations statistiques qui peuvent croître exponentiellement au fur et à mesure des mesures non lues. Il est à noter que nous aurions trouvé le même type de raisonnement dans le cas d'imperfections de mesure - c'est-à-dire si on avait procédé à une mesure qui, certes lue, aurait eu une certaine probabilité de renvoyer un résultat erroné.

On appelle état pur les états quantiques susceptibles d'être décrits par une fonction d'onde. La notion d'opérateur densité détaillée plus bas permet de représenter des états qui, suite à une mesure imparfaite ou non lue, ne sont plus purs.

2.2.2 Notion d'opérateur densité

Nous avons écrit que la valeur moyenne d'une observable A sur un système quantique décrit par la fonction d'onde $|\psi\rangle$ est donnée par :

$$\langle A \rangle_{|\psi\rangle} = \langle \psi | A | \psi \rangle$$

On peut transformer cette écriture en :

$$\langle A \rangle_{|\psi\rangle} = \text{Tr}(|\psi\rangle\langle\psi|A)$$

Ici, $|\psi\rangle\langle\psi|$ est un opérateur de $\mathcal{H}$ - le projecteur orthogonal sur le sous-espace vectoriel généré par $|\psi\rangle$. Notons $\rho_{|\psi\rangle}$ ce projecteur, qui est donc de trace un. Réciproquement, pour un projecteur de rang un et de trace un noté ρ de l'espace $\mathcal{H}$, il existe un vecteur $|\psi\rangle$ de $\mathcal{H}$ tel que $\rho = \rho_{|\psi\rangle}$. Alors, pour tout complexe λ du cercle unitaire ($|\lambda|^2 = 1$), ρ est également le projecteur sur le sous-espace propre généré par $\lambda|\psi\rangle$. Nous avons montré en 2.1.4 que deux états quantiques représentés par les fonctions d'onde $|\psi\rangle$ et $\lambda|\psi\rangle$ décrivent exactement la même situation physique, et sont donc interchangeables. Il est donc tout à fait justifié de décrire désormais un état quantique pur par $\rho_{|\psi\rangle}$ plutôt que par $|\psi\rangle$, ce que nous ferons désormais.

Matrice densité et espérance

La notion de matrice densité a été introduite pour prendre en compte des événements qui altèrent l'état du système d'une façon qui reste incertaine pour le physicien : des mesures dont le résultat produit n'est pas connu parfaitement.

Supposons qu'à un instant donné, une mesure transforme un état $|\psi\rangle$ en un état $|\psi_j\rangle$ avec la probabilité p_j , pour $j = 1 \dots m$. Les statistiques des prochaines mesures sur le système devront être formulées en considérant chacune de ces évolutions possibles. La

probabilité d'obtenir le résultat μ au cours d'une seconde mesure se calcule selon la formule :

$$\mathbb{P}(\mu) = \sum_{j=1}^m p_j \mathbb{P}(\mu | j) = \sum_{j=1}^m p_j \mathbb{P}(\mu | |\psi_j\rangle)$$

où $\mathbb{P}(\mu | |\psi_j\rangle)$ représente la probabilité d'obtenir le résultat μ à partir de l'état $|\psi_j\rangle$. Cette probabilité est tirée de l'équation (2.4), et est linéaire en $|\psi_j\rangle\langle\psi_j|$. Ainsi, réécrivons cette probabilité conditionnelle : $\mathbb{P}(\mu | |\psi\rangle) = g_\mu(|\psi\rangle\langle\psi|)$. La probabilité de produire le résultat μ en seconde mesure est donc :

$$\mathbb{P}(\mu) = \sum_{j=1}^m p_j g_\mu(|\psi_j\rangle\langle\psi_j|) = \sum_{j=1}^m g_\mu(\rho), \text{ où } \rho = \sum_{j=1}^m p_j |\psi_j\rangle\langle\psi_j| \quad (2.17)$$

Il est donc possible d'adapter simplement les équations de mesure discrète, et de remplacer la fonction d'onde par un opérateur linéaire ρ de $\mathcal{H}$ - qu'on appellera désormais matrice densité - pour prendre en compte les mesures non lues qui ont pu être réalisées auparavant. Le même travail aurait pu être fait pour les mesures avec imperfections de mesure, et pour des mesures diffusives.

Il faut donc interpréter la matrice densité comme l'espérance du projecteur orthogonal $|\psi\rangle\langle\psi|$ de l'état sur tous les processus de mesure que le physicien ne connaît pas parfaitement. Nous pourrions constater, en donnant les schémas de mesure pour des systèmes à matrice densité, que celle-ci, à chaque instant, représente toute l'information disponible sur le système à cet instant : les systèmes quantiques ont donc une structure naturellement Markovienne.

Propriétés des matrices densité

L'interprétation de l'opérateur densité comme somme de projecteurs orthogonaux sur l'espace de Hilbert $\mathcal{H}$, telle qu'exposée par l'équation (2.17), impose certaines propriétés que les matrices densité doivent respecter :

- il est facile de constater que leur trace doit valoir un ;
- la matrice densité doit être un opérateur non-négatif de $\mathcal{H}$.

A l'inverse, chaque opérateur de $\mathcal{H}$ vérifiant ces deux propriétés peut s'écrire sous la forme de l'équation (2.17), et peut donc être une matrice densité, car toute matrice non-négative admet une telle décomposition spectrale.

L'ensemble des matrices densité de $\mathcal{H}$ s'écrit donc :

$$\{\rho \in \mathcal{L}(\mathcal{H}) / \text{Tr}(\rho) = 1, \rho \geq 0\} \quad (2.18)$$

La positivité de ρ implique son *hermiticité* : $\rho = \rho^\dagger$.

Remarquons que chaque opérateur densité peut donc se décomposer comme :

$$\rho = \sum_{j=1}^J p_j |\psi_j\rangle\langle\psi_j|$$

pourvu que la somme des p_j , tous positifs ou nuls, vaille un, et que les $|\psi_j\rangle\langle\psi_j|$ soient des projecteurs orthogonaux entre eux. L'intérêt est que $J \leq N_{\mathcal{H}}$ la dimension de $\mathcal{H}$. Ainsi, à chaque instant, la matrice densité d'un système peut être représentée comme une superposition statistique d'un nombre fini de fonctions d'onde.

Enfin, on appelle *état pur* un état ρ qui peut s'écrire comme une projection orthogonale sur un vecteur $|\psi\rangle$: $\rho = |\psi\rangle\langle\psi|$. Par conséquent, un état est pur si et seulement si son état vérifie $\text{Tr}(\rho^2) = 1$. Par ailleurs, on a, pour toute matrice densité :

$$0 < \text{Tr}(\rho^2) \leq 1$$

Lorsque $\text{Tr}(\rho^2) < 1$, l'état est dit *mixte*.

Qubit, matrices de Pauli et sphère de Bloch

Comme nous l'avons dit dans la section précédente, les systèmes à deux niveaux, ou qubit, évoluent dans un espace de Hilbert à deux dimensions : $\mathcal{H} = \mathbb{C}^2$. Nous nous intéressons ici à la façon dont on représente habituellement les états quantiques d'un tel système. Un état correspond à une matrice densité, qui est donc Hermitienne et de trace un. Ces deux conditions définissent un sous-espace linéaire de $\mathcal{M}(\mathbb{C}^2)$ - l'ensemble des opérateurs linéaires de $\mathbb{C}^2$ -, de dimension 3.

Or, en définissant le produit de Frobenius comme produit scalaire sur l'espace des matrices densité admissibles (ce que nous ferons tout au long de cette thèse), les matrices de Pauli, définies comme :

$$\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix},$$

définissent une base orthonormée de l'espace tangent à l'intérieur de l'espace des matrices admissibles. En effet, on a les relations suivantes :

$$\begin{aligned} \forall \xi = x, y, z, \quad \text{Tr}(\sigma_\xi) &= 0, \quad \sigma_\xi^2 = I_{\mathcal{H}} \\ \sigma_x \sigma_y &= i \sigma_z, \quad \sigma_y \sigma_z = i \sigma_x, \quad \sigma_z \sigma_x = i \sigma_y. \end{aligned}$$

A chaque matrice densité ρ , on peut donc associer un jeu de coordonnées (x, y, z) tel que :

$$\rho = \frac{I + x\sigma_x + y\sigma_y + z\sigma_z}{2} \quad (2.19)$$

Il reste à déterminer quel est l'ensemble des triplets (x, y, z) tels que l'opérateur ρ que ce triplet construit est positif, et est donc bien une matrice densité. Il s'agit des triplets tels que $x^2 + y^2 + z^2 \leq 1$. L'ensemble des matrices densité sur $\mathbb{C}^2$ est donc isomorphe à la boule unité de $\mathbb{C}^3$, appelée *sphère de Bloch*.

Nous représenterons donc désormais l'état d'un qubit grâce à ces coordonnées (x, y, z) . La pureté d'une telle matrice, définie par $\text{Tr}(\rho^2)$ est égale à :

$$\text{Tr}(\rho^2) = \frac{1 + x^2 + y^2 + z^2}{2}$$

Une matrice densité représente donc un état pur si et seulement si ses coordonnées sont sur la surface de la sphère de Bloch.

Lorsque toutes les coordonnées sont nulles, la matrice est "au centre" de la sphère, et aucune information n'est disponible sur le système.

Lorsqu'un qubit décrit un système à deux niveaux d'énergie, l'état de plus haute énergie, noté $|e\rangle\langle e|$, correspond à $z = 1$, et l'état de plus faible énergie, noté $|g\rangle\langle g|$, à $z = -1$:

$$|e\rangle\langle e| = \frac{I + \sigma_z}{2}, \quad |g\rangle\langle g| = \frac{I - \sigma_z}{2}$$

L'opérateur σ_z permet donc de mesurer le niveau d'énergie du qubit, puisque $\sigma_z = |e\rangle\langle e| - |g\rangle\langle g|$. Enfin, nous notons $\sigma_- = |g\rangle\langle e|$ l'opérateur d'annihilation, qui permet de descendre de e vers g , et $\sigma_+ = (\sigma_-)^\dagger = |e\rangle\langle g|$ l'opérateur de création, jouant le rôle inverse.

2.2.3 Adaptation des équations d'évolution aux matrices densité

Nous adaptons ici les équations d'évolution que nous avons écrites plus haut dans le formalisme des matrices densité. Pour cela, nous écrivons la matrice densité $\rho(t)$ du système sous la forme d'une somme de projecteurs unitaires :

$$\rho(t) = \sum_{j=1}^m p_j |\psi_j(t)\rangle\langle\psi_j(t)| \quad (2.20)$$

Équation de Schrödinger

Commençons par l'équation de Schrödinger. Insérer l'équation originelle (2.1) dans cette décomposition nous permet de montrer que :

$$\begin{aligned} d\rho(t) &= d \left(\sum_{j=1}^m p_j |\psi_j(t)\rangle\langle\psi_j(t)| \right) \\ &= \sum_{j=1}^m p_j \left(|\psi_j(t)\rangle (d|\psi_j(t)\rangle)^\dagger + d|\psi_j(t)\rangle\langle\psi_j(t)| \right) \\ &= \frac{i}{\hbar} \sum_{j=1}^m p_j \left(|\psi_j(t)\rangle\langle\psi_j(t)| \hat{H}(t) + \hat{H}(t) |\psi_j(t)\rangle\langle\psi_j(t)| \right) dt \\ &= -\frac{i}{\hbar} [\hat{H}(t), \rho(t)] dt \end{aligned} \quad (2.21)$$

De plus, une évolution Hamiltonienne entre les instants t_0 et t_1 , décrite par l'opérateur d'évolution $U(t_0, t_1)$, transforme l'état $\rho(t_0)$ en :

$$\rho(t_1) = U(t_0, t_1) \rho(t_0) U(t_0, t_1)^\dagger \quad (2.22)$$

Mesure parfaite en temps discret

Montrons maintenant comment on calcule les statistiques d'une mesure discrète parfaite, et l'évolution qui en découle, à partir de l'état $\rho(t)$. Cette mesure est représentée par les opérateurs de mesure M_μ définis au paragraphe 2.1.3, et correspondant respectivement au résultat de mesure λ_μ . La probabilité d'obtenir la mesure λ_μ à partir de l'opérateur densité $\rho(t)$ est obtenue en considérant que la matrice densité représente un système qui a la probabilité p_j d'être dans l'état $|\psi_j(t)\rangle$, d'une façon analogue à ce qui a été présenté au paragraphe 2.2.2, puis en injectant l'équation (2.4) pour chacune des fonctions d'onde. Ainsi, on obtient la formule suivante :

$$\begin{aligned} \mathbb{P}(\text{résultat } \lambda_\mu \mid \text{état } \rho) &= \sum_{j=1}^m p_j \langle\psi_j(t)| M_\mu^\dagger M_\mu |\psi_j(t)\rangle \\ &= \sum_{j=1}^m p_j \text{Tr} \left(M_\mu |\psi_j(t)\rangle\langle\psi_j(t)| M_\mu^\dagger \right) \end{aligned} \quad (2.23)$$

$$= \text{Tr} \left(M_\mu \rho(t) M_\mu^\dagger \right) \quad (2.24)$$

Lorsque la mesure est réalisée entre les instants t et $t + \Delta t$, et renvoie le résultat λ_μ , la matrice densité de l'état devient alors, par réduction du paquet d'onde :

$$\rho(t + \Delta t) = \frac{M_\mu \rho(t) M_\mu^\dagger}{\text{Tr}(M_\mu \rho(t) M_\mu^\dagger)} \quad (2.25)$$

Mesure parfaite diffusive en temps continu

Enfin, voyons comment évolue le même état quantique non nécessairement pur $\rho(t)$ soumis à M canaux de mesure $(L_\nu)_{\nu=1\dots M}$. Tout d'abord, il est facile de vérifier, de la même façon que pour les mesures discrètes, que ces canaux de mesure vérifient l'équation différentielle stochastique suivante :

$$\begin{aligned} dy_t^\nu &= dW_t^\nu + \text{Tr}\left((L_\nu + L_\nu^\dagger) \rho(t)\right) dt \\ < dy_t^\nu > &= \text{Tr}\left((L_\nu + L_\nu^\dagger) \rho(t)\right) dt \end{aligned} \quad (2.26)$$

Après un tel résultat de mesure sur chacun des canaux, l'état devient $\rho(t + dt)$ où :

$$\begin{aligned} \rho(t + dt) &= \frac{M_{t,dy_t} \rho(t) M_{t,dy_t}^\dagger}{\text{Tr}(M_{t,dy_t} \rho(t) M_{t,dy_t}^\dagger)} \\ M_{t,dy_t} &= I_{\mathcal{H}} - \left(\frac{i}{\hbar} \hat{H}(t) + \sum_{\nu=1}^M \frac{L_\nu^\dagger L_\nu}{2} \right) dt + \sum_{\nu=1}^M L_\nu dy_t^\nu \end{aligned} \quad (2.27)$$

Cette équation est facile à discrétiser, et prend rigoureusement la même forme que celle de l'équation (2.16) décrivant la réduction du paquet d'ondes à la suite d'une mesure discrète. On peut néanmoins la transformer en une équation différentielle stochastique, en tenant compte des règles de calcul de Itô, ce qui nous donne, en écrivant que $\rho(t + dt) = \rho(t) + d\rho(t)$:

$$\begin{aligned} d\rho(t) &= \left(-\frac{i}{\hbar} [\hat{H}(t), \rho(t)] + \sum_{\nu=1}^M \mathcal{L}_\nu(\rho(t)) \right) dt \\ &\quad + \sum_{\nu=1}^M \left(\mathcal{M}_\nu(\rho(t)) - \overline{\mathcal{M}}_\nu(\rho(t)) \rho(t) \right) dW_t^\nu \end{aligned} \quad (2.28)$$

où $\mathcal{L}_\nu(\rho) = L_\nu \rho L_\nu^\dagger - (L_\nu^\dagger L_\nu \rho + \rho L_\nu^\dagger L_\nu)/2$ est appelé *Lindbladien* pour le canal de mesure ν , $\mathcal{M}_\nu(\rho) = L_\nu \rho + \rho L_\nu^\dagger$ et $\overline{\mathcal{M}}_\nu(\rho) = \text{Tr}\left((L_\nu + L_\nu^\dagger) \rho\right)$ sa moyenne intervenant dans l'évolution de la mesure dy_t^ν . Cette équation est appelée *équation maîtresse* du système (*master equation*). Lorsqu'une discrétisation est nécessaire, pour simuler une trajectoire quantique numériquement par exemple, il est préférable de se référer à l'équation discrète (2.27), car l'équation maîtresse, dans sa forme discrétisée, peut ne pas préserver la positivité de $\rho(t)$.

2.2.4 Mesures non lues et mesures imparfaites

Cas des mesures discrètes

Nous décrivons ici le processus d'une mesure discrète, dont le résultat est imparfaitement ou non lu, entre les instants t et $t + \Delta t$.

Mesures non lues Nous nous replaçons ici dans le cadre des mesures discrètes, décrites par les équations (2.24) et (2.25). Or, nous considérons désormais que le résultat de la mesure μ n'est pas lu. La mesure a bien eu lieu, mais l'expérimentateur n'a pas pris connaissance du résultat produit. Pour déterminer en quoi l'état est altéré par cette mesure, il convient d'utiliser le fait qu'une matrice densité est une espérance sur les résultats de mesure non lus. Ainsi :

$$\begin{aligned}\rho(t + \Delta t) &= \mathbb{E}_\mu(\rho(t + \Delta t) \mid \rho(t)) \\ &= \sum_{\mu=1}^m \rho_\mu(t + \Delta t) \mathbb{P}(\text{résultat } \lambda_\mu \mid \text{état } \rho(t))\end{aligned}$$

où $\mathbb{E}_\mu(\cdot)$ désigne une espérance sur le résultat de la mesure μ et $\rho_\mu(t + \Delta t)$ la valeur de l'état final si la mesure avait été lue en λ_μ . Son expression est donnée par la formule (2.25), et celle de la probabilité par (2.24). On peut ainsi écrire que, lorsque la mesure n'est pas lue, l'évolution du système est donnée par :

$$\rho(t + \Delta t) = \sum_{j=1}^m M_j \rho(t) M_j^\dagger \quad (2.29)$$

Mesures imparfaites Dans ce paragraphe, nous considérons un cas intermédiaire : celui où le résultat de la mesure nous parvient, mais faussé par un bruit de mesure classique. En d'autres termes, nous séparons ici la mesure μ effectivement produite par le système quantique, et la mesure y qui est le résultat que l'expérimentateur parvient à lire. y est générée à partir de μ comme une réalisation d'une variable aléatoire discrète dont la loi de probabilité est donnée par :

$$\mathbb{P}(y = \tilde{y} \mid \mu = \tilde{\mu}) = \eta_{\tilde{\mu}\tilde{y}} \quad (2.30)$$

On considère que le nombre Y de mesures qui peuvent être lues par l'expérimentateur peut être différent de m , le nombre de mesures que le système peut produire (le nombre de valeurs propres de l'observable associée à la mesure). Ces mesures possibles forment, pour simplifier, l'ensemble $\{1; \dots; Y\}$. On peut remarquer que la matrice $\tilde{\eta} = (\eta_{\mu y})_{\mu \leq m, y \leq Y}$ est une matrice stochastique. Les coefficients de cette matrice sont indépendants de l'état du système $\rho(t)$. En reprenant l'interprétation probabiliste de la matrice densité, nous écrivons, après une mesure $\tilde{y}$ lue par l'expérimentateur, que :

$$\begin{aligned}\rho(t + \Delta t) &= \mathbb{E}_\mu(\rho(t + \Delta t) \mid \rho(t), \tilde{y}) \\ &= \sum_{\mu=1}^m \rho_\mu(t + \Delta t) \mathbb{P}(\lambda_\mu \mid \rho(t), \tilde{y})\end{aligned}$$

Un élément nous manque ici : la valeur de $\mathbb{P}(\lambda_\mu \mid \rho(t), \tilde{y})$. En utilisant la loi de Bayes, on obtient :

$$\begin{aligned}\mathbb{P}(\lambda_\mu \mid \rho(t), \tilde{y}) &= \frac{\mathbb{P}(\tilde{y} \mid \rho(t), \lambda_\mu) \mathbb{P}(\lambda_\mu \mid \rho(t))}{\mathbb{P}(\tilde{y} \mid \rho(t))} \\ &= \frac{\eta_{\mu\tilde{y}} \text{Tr}(M_\mu \rho(t) M_\mu^\dagger)}{\mathbb{P}(\tilde{y} \mid \rho(t))}\end{aligned}$$

Le terme apparaissant au dénominateur est un terme de normalisation (puisque'on doit garder $\text{Tr}(\rho(t + \Delta t)) = 1$), qui nous donnera la loi de probabilité de y selon $\rho(t)$. On écrit

donc que la mesure $\tilde{y}$ est donnée avec la probabilité :

$$\mathbb{P}(y = \tilde{y} \mid \rho(t)) = \text{Tr} \left(\sum_{\mu=1}^m \eta_{\mu\tilde{y}} M_{\mu} \rho(t) M_{\mu}^{\dagger} \right) \quad (2.31)$$

Après cette mesure, l'état du système devient :

$$\rho(t + \Delta t) = \frac{\sum_{\mu=1}^m \eta_{\mu\tilde{y}} M_{\mu} \rho(t) M_{\mu}^{\dagger}}{\text{Tr} \left(\sum_{\mu=1}^m \eta_{\mu\tilde{y}} M_{\mu} \rho(t) M_{\mu}^{\dagger} \right)} \quad (2.32)$$

Systèmes ouverts et mesures par couplage avec l'environnement Jusqu'à présent, les processus de mesure que nous avons analysés consistaient à réduire le paquet d'onde sur le système lui-même. Dans un système quantique ouvert, la mesure peut également être indirecte. En effet, après avoir couplé un système avec un système tiers, une réduction du paquet d'onde sur ce système tiers a une influence et produit une information sur le premier système.

Appelons "expérience" le système dont on cherche à connaître l'état, évoluant dans $\mathcal{H}$, et "environnement" ce système tiers dont l'état est un opérateur de l'espace de Hilbert noté $\mathcal{E}$. L'état du système composite est un opérateur de l'espace formé par le produit tensoriel $\mathcal{H} \otimes \mathcal{E}$. Supposons d'abord qu'initialement, les deux systèmes étaient initialement séparables : l'état initial du système s'écrit donc :

$$\rho_{\mathcal{H} \otimes \mathcal{E}} = \rho_{\mathcal{H}} \otimes \rho_{\mathcal{E}} \quad (2.33)$$

Par la suite, une évolution Hamiltonienne sur $\mathcal{H} \otimes \mathcal{E}$ a intriqué les deux systèmes. Ainsi, il existe un opérateur d'évolution, donc auto-adjoint, de $\mathcal{H} \otimes \mathcal{E}$, noté U , tel que l'état du système composite juste avant le processus de mesure est :

$$\rho_{\mathcal{H} \otimes \mathcal{E}}^{-} = U (\rho_{\mathcal{H}} \otimes \rho_{\mathcal{E}}) U^{\dagger} \quad (2.34)$$

Quel est l'effet d'une mesure par réduction du paquet d'ondes du système "environnement" sur l'état du système "expérience" ? Les détails mathématiques conduisant aux résultats suivants sont donnés en annexe [B.2](#).

Nous notons les résultats possibles de cette mesure $y = 1 \dots Y$. Pour chacun de ces résultats possibles, il existe un ensemble d'opérateurs $(M_l)_{l \in I_y}$, où I_y est un ensemble fini d'indices, tel que le processus de mesure, lorsque $\rho_{\mathcal{E}}$ est connu de l'expérimentateur, peut être exprimé comme dépendant de la matrice $\rho_{\mathcal{H}}$:

$$\mathbb{P}(\text{résultat } y \mid \text{état } \rho_{\mathcal{H}}) = \text{Tr} \left(\sum_{l \in I_y} M_l \rho_{\mathcal{H}} M_l^{\dagger} \right) \quad (2.35)$$

$$\rho_{\mathcal{H}}^+ = \frac{\sum_{l \in I_y} M_l \rho_{\mathcal{H}} M_l^{\dagger}}{\text{Tr} \left(\sum_{l \in I_y} M_l \rho_{\mathcal{H}} M_l^{\dagger} \right)} \quad (2.36)$$

où $\rho_{\mathcal{H}}^+$ désigne l'état du système "expérience" après la mesure faite sur l'environnement.

Cadre général des mesures quantiques discrètes Nous avons donc présenté ici le cadre le plus général des mesures discrètes qui peuvent être effectuées sur un système quantique à dimension finie. La mesure y est un entier entre 1 et Y . A chacun de ces

résultats possibles, nous associons un super-opérateur de $\mathcal{H}$ ³ linéaire lui correspondant. Ce super-opérateur, noté $\mathbf{K}_y$ pour le résultat de mesure y , est appelé *application de Kraus partielle*, a une structure particulière, puisqu'il existe un ensemble fini d'opérateurs $(M_l)_{l \in I_y}$ tel que, pour tout opérateur densité :

$$\mathbf{K}_y(\rho) = \sum_{l \in I_y} M_l \rho M_l^\dagger$$

L'ensemble des applications de Kraus forme un ensemble complet d'applications de Kraus. Lorsque la mesure est non lue, l'état est itéré avec :

$$\rho(t + \Delta t) = \mathbf{K}(\rho) = \sum_{y=1}^Y \mathbf{K}_y(\rho)$$

Cet ensemble est dit complet, car la préservation de la trace de l'opérateur densité impose l'équation suivante (*completeness equation* dans [NC00, 2.4.2]) :

$$\sum_{y=1}^Y \sum_{l \in I_y} M_l^\dagger M_l = I_{\mathcal{H}}$$

Cas des mesures diffusives

Canaux de mesure non lus Commençons par donner ici l'évolution d'un système diffusif dont aucun des canaux de mesure n'est lu. Puisque l'opérateur densité représente une espérance par rapport aux résultats des mesures effectuées, on peut retrouver l'évolution d'un système diffusif due à des canaux de mesure non lus en moyennant l'évolution normale par rapport au bruit de mesure résiduel, décrite par l'équation (2.28). On obtient alors rapidement, en indexant par λ ces canaux de mesure non lus :

$$d\rho(t) = \left(-\frac{i}{\hbar} [\hat{H}(t), \rho(t)] + \sum_{\lambda=1}^{\Lambda} \mathcal{L}_{\lambda}(\rho(t)) \right) dt \quad (2.37)$$

où $\rho \mapsto \mathcal{L}_{\lambda}(\rho)$ est le Linbladien du canal λ . Cette équation diffusive est également discrétisable. Néanmoins, il est impossible de vouloir utiliser tel quel le schéma de discrétisation de l'équation (2.27), puisque la mesure $d\mathbf{y}_t$ n'existe plus. On utilise seulement la matrice M_t avec des termes de complément, ce qui donne :

$$\begin{aligned} \rho(t + dt) &= \frac{M_t \rho(t) M_t^\dagger + \sum_{\lambda=1}^{\Lambda} L_{\lambda} \rho(t) L_{\lambda}^\dagger dt}{\text{Tr} \left(M_t \rho(t) M_t^\dagger + \sum_{\lambda=1}^{\Lambda} L_{\lambda} \rho(t) L_{\lambda}^\dagger dt \right)} \\ M_t &= I_{\mathcal{H}} - \left(\frac{i}{\hbar} \hat{H}(t) + \sum_{\lambda=1}^{\Lambda} \frac{L_{\lambda}^\dagger L_{\lambda}}{2} \right) dt \end{aligned} \quad (2.38)$$

Il est à noter que cette discrétisation reprend le cadre général des mesures en temps discret, avec pour opérateurs de Kraus M_t et les matrices $L_{\lambda} \sqrt{dt}$. Cette équivalence dans la façon de prendre en compte l'influence des mesures discrètes et celle des mesures diffusives nous sera très utile dans cette thèse.

3. Une application de $\mathcal{L}(\mathcal{H})$ dans lui-même.

Mesures imparfaites et efficacité de détection Nous traitons désormais le cas où les canaux de mesure, indexés par ν , sont à nouveau détectés par l'utilisateur, mais avec un bruit gaussien additionnel. Ce bruit, s'il est normalisé, conduit à ce que la part déterministe de la mesure - qui est le terme $\text{Tr} \left((L_\nu + L_\nu^\dagger) \rho(t) \right) dt$ - dans la mesure totale dy_t^ν du canal ν perçue par le physicien décroisse. Pour prendre cela en compte, on affecte à chaque canal de mesure une efficacité de détection notée $\eta_\nu \in [0; 1]$ telle que :

$$\begin{aligned} dy_t^\nu &= dW_t^\nu + \sqrt{\eta_\nu} \text{Tr} \left((L_\nu + L_\nu^\dagger) \rho(t) \right) dt \\ < dy_t^\nu > &= \sqrt{\eta_\nu} \text{Tr} \left((L_\nu + L_\nu^\dagger) \rho(t) \right) dt \end{aligned} \quad (2.39)$$

L'équation maîtresse décrivant l'effet de ces mesures imparfaites prend la forme suivante :

$$\begin{aligned} d\rho(t) &= \left(-\frac{i}{\hbar} [\hat{H}(t), \rho(t)] + \sum_{\nu=1}^M \mathcal{L}_\nu(\rho(t)) \right) dt \\ &\quad + \sum_{\nu=1}^M \sqrt{\eta_\nu} \left(\mathcal{M}_\nu(\rho(t)) - \overline{\mathcal{M}}_\nu(\rho(t)) \rho(t) \right) dW_t^\nu \end{aligned} \quad (2.40)$$

Nous rappelons que le Linbladien $\mathcal{L}_\nu$ vérifie $\mathcal{L}_\nu(\rho) = L_\nu \rho L_\nu^\dagger - (L_\nu^\dagger L_\nu \rho + \rho L_\nu^\dagger L_\nu)/2$, que $\mathcal{M}_\nu(\rho) = L_\nu \rho + \rho L_\nu^\dagger$, et que $\overline{\mathcal{M}}_\nu(\rho) = \text{Tr}(\mathcal{M}_\nu(\rho))$.

Ainsi, pour donner une formule générale à l'évolution de tous les systèmes quantiques diffusifs, nous prendrons en compte les canaux de mesure non lus, en les indexant par $\lambda = 1, \dots, \Lambda$, et les canaux de mesure lus, indexés par $\nu = 1, \dots, M$, caractérisés par leur efficacité de détection η_ν . L'équation maîtresse stochastique prend alors la forme :

$$\begin{aligned} d\rho(t) &= \rho(t+dt) - \rho(t) = \left(-\frac{i}{\hbar} [\hat{H}(t), \rho(t)] + \sum_{\nu=1}^M \mathcal{L}_\nu(\rho(t)) + \sum_{\lambda=1}^\Lambda \mathcal{L}_\lambda(\rho(t)) \right) dt \\ &\quad + \sum_{\nu=1}^M \sqrt{\eta_\nu} \left(\mathcal{M}_\nu(\rho(t)) - \overline{\mathcal{M}}_\nu(\rho(t)) \rho(t) \right) dW_t^\nu, \end{aligned} \quad (2.41)$$

avec $dy_t^\nu = dW_t^\nu + \sqrt{\eta_\nu} \text{Tr} \left((L_\nu + L_\nu^\dagger) \rho(t) \right) dt$.

Schémas de discrétisation Il est donc nécessaire d'adjoindre à l'équation maîtresse stochastique que nous venons de présenter un schéma de discrétisation, adapté de celui de l'équation (2.38), pour pouvoir travailler numériquement sur ces systèmes. Les mesures discrétisées sont notées :

$$\Delta y_t^\nu = \int_t^{t+\Delta t} dy_t^\nu$$

et $\Delta \mathbf{y}_t$ est le vecteur $(\Delta y_t^1, \dots, \Delta y_t^M)$.

Les schémas de discrétisation que nous allons présenter doivent avoir une forme analogue aux équations modélisant les mesures discrètes, c'est-à-dire :

$$\rho(t + \Delta t) = \frac{\mathbf{K}_{t, \Delta \mathbf{y}_t}(\rho(t))}{\text{Tr}(\mathbf{K}_{t, \Delta \mathbf{y}_t}(\rho(t)))} \quad (2.42)$$

où $\rho \mapsto \mathbf{K}_{t, \Delta \mathbf{y}_t}(\rho)$ est un super-opérateur linéaire, ayant la structure d'une application de Kraus partielle. Ces schémas apportent une loi de probabilité sur la valeur des sorties

discrétisées Δy_t^ν sur $[t; t + \Delta t]$. Une première formulation d'un tel filtre avait déjà été donné dans [RR15], sans y adjoindre de justification mathématique de validité. Puisque ces sorties sont de l'ordre de $\sqrt{\Delta t}$, on les paramètre avec $\Delta y_t^\nu = s^\nu \sqrt{\Delta t}$, en écrivant $\mathbf{s} = (s^1, \dots, s^M)$. On a ensuite la loi de probabilité approchée :

$$\mathbb{P}(\Delta \mathbf{y}_t \in [\mathbf{s}; \mathbf{s} + d\mathbf{s}] \sqrt{\Delta t} \mid \rho(t)) = \prod_{\nu=1}^M \mathbb{P}(\Delta y_t^\nu \in [s^\nu; s^\nu + ds^\nu] \sqrt{\Delta t} \mid \rho(t)) = \quad (2.43)$$

$$\frac{\text{Tr}(\mathbf{K}_{t, \mathbf{s} \sqrt{\Delta t}}(\rho(t)))}{(\sqrt{2\pi})^M} \exp\left(-\frac{\|\mathbf{s}\|^2}{2}\right) \prod_{\nu=1}^M ds^\nu$$

où $\|\mathbf{s}\|$ est la norme euclidienne de $\mathbf{s}$.

Remarquons que cette loi de probabilité fait intervenir deux éléments : un terme exponentiel qui ne dépend absolument pas de l'état du système $\rho(t)$, et la trace de l'application de Kraus qui, appliquée à $\rho(t)$, permet de déterminer l'évolution du système. Ce second terme est tout à fait analogue à l'expression de la probabilité d'une mesure discrète. Les schémas de discrétisation présentés ci-dessous prennent donc la forme d'un processus de Markov, analogue à celui régissant les systèmes à temps discret.

Un premier schéma de discrétisation, appelé schéma de discrétisation simple, est construit avec l'application de Kraus $\mathbf{K}_{t, \Delta \mathbf{y}_t}^{\text{sim}}$ suivante :

$$\text{avec } \mathbf{K}_{t, \Delta \mathbf{y}_t}^{\text{sim}}(\rho) = M_{t, \Delta \mathbf{y}_t}^{\text{sim}} \rho (M_{t, \Delta \mathbf{y}_t}^{\text{sim}})^\dagger + \sum_{\nu=1}^M (1 - \eta_\nu) L_\nu \rho L_\nu^\dagger \Delta t + \sum_{\lambda=1}^\Lambda L_\lambda \rho L_\lambda^\dagger \Delta t, \quad (2.44)$$

$$\text{et } M_{t, \Delta \mathbf{y}_t}^{\text{sim}} = I_{\mathcal{H}} + \sum_{\nu=1}^M \sqrt{\eta_\nu} L_\nu \Delta y_t^\nu - \left(\frac{i}{\hbar} \hat{H}(t) + \sum_{\nu=1}^M \frac{L_\nu^\dagger L_\nu}{2} + \sum_{\lambda=1}^\Lambda \frac{L_\lambda^\dagger L_\lambda}{2} \right) \Delta t$$

Nous donnons en annexe, en nous basant sur notre papier [SCIB⁺15], et sur les preuves exposées dans [JCRH15] et [APR14], les arguments mathématiques montrant que ce schéma de discrétisation converge, lorsque Δt tend vers dt , vers l'équation différentielle stochastique maîtresse.

Néanmoins, ce schéma de discrétisation peut ne pas satisfaire les attentes du physicien lorsque celui-ci cherche à estimer la valeur d'un paramètre intervenant dans l'évolution du système. Puisque nous avons donné une expression de la densité de probabilité des mesures à un instant donné, vérifions que l'intégrale de cette densité sur l'ensemble des mesures possibles vaut 1 jusqu'à un certain ordre de Δt , et caractérisons le terme d'erreur qui en résultera. On écrit :

$$\int_{\mathbb{R}^M} \mathbb{P}(\Delta \mathbf{y}_t \in [\mathbf{s}; \mathbf{s} + d\mathbf{s}] \sqrt{\Delta t} \mid \rho(t)) d\mathbf{s} = \int_{\mathbb{R}^M} \frac{\text{Tr}(\mathbf{K}_{t, \mathbf{s} \sqrt{\Delta t}}(\rho(t)))}{(\sqrt{2\pi})^M} \exp\left(-\frac{\|\mathbf{s}\|^2}{2}\right) d\mathbf{s} \quad (2.45)$$

en utilisant la pseudo-application de Kraus partielle de l'équation (2.44).

Les calculs qui suivent cette réécriture nous permettent de montrer que :

$$\int_{\mathbb{R}^M} \mathbb{P}(\Delta \mathbf{y}_t \in [\mathbf{s}; \mathbf{s} + d\mathbf{s}] \sqrt{\Delta t} \mid \rho(t)) d\mathbf{s} = 1 + \text{Tr}(C^\dagger(t) C(t) \rho(t)) (\Delta t)^2 \quad (2.46)$$

où $C(t)$ est l'opérateur défini par $C(t) = \frac{i}{\hbar} \hat{H}(t) + \sum_{\nu=1}^M \frac{L_\nu^\dagger L_\nu}{2} + \sum_{\lambda=1}^\Lambda \frac{L_\lambda^\dagger L_\lambda}{2}$. Lorsqu'aucun Hamiltonien n'est exercé sur le système, cet opérateur est Hermitien. Ce schéma est donc valable jusqu'à l'ordre Δt seulement.

Notons que l'expression de $C(t)$ peut dépendre de la valeur de paramètres physiques, intervenant dans les opérateurs de mesure L_ν ou dans le Hamiltonien, paramètres qu'on peut vouloir estimer. Cette estimation étant de nature bayésienne, les valeurs pour lesquelles $\text{Tr}(C^\dagger(t)C(t)\rho(t))$ est grand seront donc considérées d'emblée comme étant, à l'ordre $(\Delta t)^2$, dotés d'une plus grande probabilité initiale, ce qui faussera l'estimation. Il y a plusieurs façons de créer une application de Kraus modifiée qui prend en compte cette dépendance en $(\Delta t)^2$, pour l'annuler. Nous présentons donc ici un premier schéma alternatif, qui consiste à construire de nouveau l'application de Kraus grâce à l'équation (2.44), en utilisant, à la place de $M_{t,\Delta\mathbf{y}_t}^{\text{sim}}$, la nouvelle matrice $M_{t,\Delta\mathbf{y}_t}$:

$$\begin{aligned} M_{t,\Delta\mathbf{y}_t} &= I_{\mathcal{H}} + \sum_{\nu=1}^M \sqrt{\eta_\nu} L_\nu \Delta y_t^\nu - \left(\frac{i}{\hbar} \hat{H}(t) + \sum_{\nu=1}^M \frac{L_\nu^\dagger L_\nu}{2} + \sum_{\lambda=1}^\Lambda \frac{L_\lambda^\dagger L_\lambda}{2} \right) \Delta t - \frac{C^\dagger(t)C(t)}{2} (\Delta t)^2 \\ &= M_{t,\Delta\mathbf{y}_t}^{\text{sim}} - \frac{C^\dagger(t)C(t)}{2} (\Delta t)^2 \end{aligned} \quad (2.47)$$

Ce schéma a l'avantage d'être facile à implémenter numériquement. Nous en donnons ensuite un second, qui reprend la pseudo-application de Kraus $\mathbf{K}_{t,\Delta\mathbf{y}_t}^{\text{sim}}$ du schéma de discrétisation simple (voir l'équation (2.44)), utilisée non plus sur $\rho(t)$ directement, mais sur $\sigma(t)\rho(t)\sigma(t)$, avec $\sigma(t)$ la matrice Hermitienne :

$$\sigma(t) = \sqrt{(I_{\mathcal{H}} + C^\dagger(t)C(t)(\Delta t)^2)^{-1}},$$

de sorte que ce schéma de discrétisation a lui aussi la structure donnée par les équations (2.42) et (2.43), avec comme application de Kraus :

$$\rho \mapsto \mathbf{K}_{t,\Delta\mathbf{y}_t}(\rho) = \mathbf{K}_{t,\Delta\mathbf{y}_t}^{\text{sim}}(\sigma(t)\rho(t)\sigma(t)), \quad \text{avec } \sigma(t) = \sqrt{(I_{\mathcal{H}} + C^\dagger(t)C(t)(\Delta t)^2)^{-1}} \quad (2.48)$$

Ce schéma a l'avantage de vérifier, comme le schéma de discrétisation simple, la convergence en loi vers l'équation différentielle stochastique maîtresse, et de permettre également que, pour toute matrice densité ρ :

$$\begin{aligned} \int_{\mathbb{R}^M} \mathbb{P}(\Delta\mathbf{y}_t \in [\mathbf{s}; \mathbf{s} + d\mathbf{s}] \sqrt{\Delta t} \mid \rho) d\mathbf{s} &= \int_{\mathbb{R}^M} \frac{\text{Tr}(\mathbf{K}_{t,\mathbf{s}\sqrt{\Delta t}}^{\text{sim}}(\sigma(t)\rho\sigma(t)))}{(\sqrt{2\pi})^M} \exp\left(-\frac{\|\mathbf{s}\|^2}{2}\right) d\mathbf{s} \\ &= \int_{\mathbb{R}^M} \frac{\text{Tr}((\mathbf{K}_{t,\mathbf{s}\sqrt{\Delta t}}^{\text{sim}})^*(I_{\mathcal{H}})\sigma(t)\rho\sigma(t))}{(\sqrt{2\pi})^M} \exp\left(-\frac{\|\mathbf{s}\|^2}{2}\right) d\mathbf{s} \\ &= \text{Tr}(\sigma(t)(I_{\mathcal{H}} + C^\dagger(t)C(t)(\Delta t)^2)\sigma(t)\rho) = 1 \end{aligned} \quad (2.49)$$

2.3 Trajectoire quantique et estimation de paramètres

2.3.1 Cadre général de la thèse

Nous traitons dans cette thèse des systèmes quantiques sur lesquels agissent l'un ou l'autre des deux types de mesures quantiques que nous avons présentés plus haut :

- les systèmes sur lesquels les mesures sont en temps discret ;
- les systèmes sur lesquels les mesures en temps continu sont diffusives.

Puisque nous analysons numériquement ces systèmes, le temps sera, sauf mention contraire, considéré comme discret. L'instant initial est noté t_0 ; puis l'état est calculé aux instants successifs t_k , tels que $t_0 < \dots < t_k < t_{k+1}$. Le dernier de ces instants est noté t_T . Pour alléger les formules, on écrit l'état initial du système :

$$\rho(t_0) = \rho_0$$

et l'état du système à l'instant t_k , de façon analogue⁴ :

$$\rho(t_k) = \rho_k$$

- Au sein des systèmes "discrets", chaque intervalle de temps $[t_{k-1}; t_k]$ est le moment réservé à une et une seule mesure, dont le résultat est noté y_k . Nous distinguerons, lorsqu'il y aura lieu, cette mesure y_k , qui est celle que le physicien peut lire en dernière instance du processus, et μ_k , qui est le résultat que le système quantique produit, avant que les imperfections de mesure classique n'interviennent. La mesure y_k est donc la réalisation d'une variable aléatoire dont la loi de probabilité dépend de μ_k .
- Au sein des systèmes "diffusifs", tous les intervalles de temps $[t_{k-1}; t_k]$ ont la même durée $\Delta t = t_k - t_{k-1}$. Nous disposons, après chacun d'entre eux, et grâce à d'éventuels raffinements expérimentaux en amont, les signaux discrétisés de chaque canal de mesure :

$$\Delta y_k^\nu = \int_{t_{k-1}}^{t_k} dy_t^\nu$$

et nous notons $\Delta \mathbf{y}_k$ le vecteur contenant toutes ces mesures : $\Delta \mathbf{y}_k = (\Delta y_k^1, \dots, \Delta y_k^M)$. Le canal de mesure ν a pour efficacité de détection η_ν . Pour incrémenter l'état du système en prenant en compte les mesures successives $\Delta \mathbf{y}_k$, nous utilisons le schéma de discrétisation, d'ordre 1 ou 2 selon les objectifs visés, décrit par les équations (2.42) et (2.43).

2.3.2 Filtre réel et filtres approchés

Unification des deux types de systèmes

Grâce au schéma de discrétisation que nous avons présenté, il nous est toujours possible, que le système étudié soit à mesures discrètes ou à mesures diffusives, de décrire le processus de mesure intervenant entre les instants t_{k-1} et t_k par une application de Kraus partielle. Dans les systèmes à temps discret, l'application de Kraus partielle correspondant à une mesure y_k entre ces deux instants est notée $\mathbf{K}_{k,y_k}$ (la dépendance en k indique le fait que les opérateurs de mesure composant l'application de Kraus peuvent dépendre du temps), et l'ensemble des applications de Kraus partielles correspondant à chaque valeur de mesure

$$\sum_{y=1}^Y \mathbf{K}_{k,y} = \mathbf{K}_k, \quad \text{avec } \text{Tr}(\mathbf{K}_k(\rho)) = \text{Tr}(\rho) \quad (2.50)$$

pour tout opérateur hermitien ρ de $\mathcal{H}$. L'application de Kraus complète $\mathbf{K}_k$ permet alors de décrire l'évolution du système sans tenir compte du résultat de la mesure.

4. Cette représentation de l'instant t_k par k se retrouvera également dans la représentation des opérateurs de mesure et des résultats qu'ils produisent.

Il est possible d'écrire, pour les systèmes à mesures diffusives, des notations semblables, puisque que les schémas de discrétisation permettent de décrire l'évolution de l'état entre les instants t_{k-1} et t_k avec une application de Kraus notée $\mathbf{K}_{k,\Delta\mathbf{y}_k}$. Ici, une relation analogue à l'équation (2.50) est donnée par :

$$\mathbf{K}_k(\rho) = \int_{\mathbb{R}^M} \frac{\text{Tr}(\mathbf{K}_{k,s\sqrt{\Delta t}}(\rho(t)))}{\sqrt{2\pi}} \exp\left(-\frac{\|\mathbf{s}\|^2}{2}\right) d\mathbf{s}, \quad (2.51)$$

avec $\text{Tr}(\mathbf{K}_k(\rho)) = \text{Tr}(\rho) + \mathcal{O}((\Delta t)^2)$ pour le schéma de discrétisation d'ordre 1 et $\text{Tr}(\mathbf{K}_k(\rho)) = \text{Tr}(\rho) + \mathcal{O}((\Delta t)^3)$ pour celui d'ordre 2.

Ainsi, à partir d'un état initial ρ_0 , nous procédons à ce qui peut être considéré comme un ensemble de T mesures successives, donnant lieu aux états $\rho_1, \dots, \rho_T$. Ces états sont générés par la formule (voir, par exemple, [WM09]) :

$$\rho_k = \frac{\mathbf{K}_{k,y_k}(\rho_{k-1})}{\text{Tr}(\mathbf{K}_{k,y_k}(\rho_{k-1}))} \quad (2.52)$$

où, pour simplifier, on note, dans le cas des mesures diffusives, $y_k = \Delta\mathbf{y}_k$. Cette mesure y_k est la réalisation d'une variable aléatoire, discrète pour les mesures discrètes et continue pour les mesures diffusives, dont les lois sont données par :

$$\begin{aligned} \mathbb{P}(y_k = y \mid \rho_0, y_1, \dots, y_{k-1}) &= \mathbb{P}(y_k = y \mid \rho_{k-1}) \\ &= \text{Tr}(\mathbf{K}_{k,y}(\rho_{k-1})) \end{aligned}$$

pour les systèmes à temps discret et :

$$\begin{aligned} \mathbb{P}(\Delta\mathbf{y}_k \in [\Delta\mathbf{y}; \Delta\mathbf{y} + d\mathbf{y}] \mid \rho_0, y_1, \dots, y_{k-1}) &= \mathbb{P}(\Delta\mathbf{y}_k \in [\Delta\mathbf{y}; \Delta\mathbf{y} + d\mathbf{y}] \mid \rho_{k-1}) \\ &\propto \text{Tr}(\mathbf{K}_{k,\Delta\mathbf{y}}(\rho_{k-1})) d\mathbf{y}_1 \dots d\mathbf{y}_M \end{aligned}$$

où le signe "proportionnel" montre que la seule dépendance par rapport aux états précédents du système dans cette distribution de probabilité est $\text{Tr}(\mathbf{K}_{k,\Delta\mathbf{y}}(\rho_{k-1}))$. L'ensemble $[\Delta\mathbf{y}; \Delta\mathbf{y} + d\mathbf{y}]$ représente le pavé, dont $\Delta\mathbf{y}_k \in [\Delta\mathbf{y}; \Delta\mathbf{y} + d\mathbf{y}]$ impose que, pour tout $\nu = 1 \dots M$, $\Delta y_k^\nu \in [\Delta y^\nu; \Delta y^\nu + dy^\nu]$.

Puisque cette loi de probabilité ne dépend que de l'état ρ_{k-1} précédent, nous avons affaire, pour ces deux types de systèmes, à un processus de Markov en temps discret dont l'état est ρ_k et la mesure y_k ou $\Delta\mathbf{y}_k$.

Filtres approchés

Nous venons de donner ici l'expression des deux filtres Markoviens permettant d'obtenir l'évolution de l'état réel du système. L'utilisation de ces filtres suppose de connaître :

- l'état initial ρ_0 du système,
- l'expression exacte des applications de Kraus partielles à tout instant. Or, des ces expressions sont obtenues à partir de paramètres physiques précis, dont la valeur exacte de certains d'entre eux peut ne pas être connue a priori par l'expérimentateur. Nous notons $p \in \mathbb{R}^q$ les valeurs de ces paramètres. Pour prendre en compte la dépendance des applications de Kraus partielles à ces valeurs de paramètres, nous notons désormais $\mathbf{K}_{k,y}^{\tilde{p}}$ l'application correspondant à l'instant t_k et la mesure y générée par la valeur de paramètres $\tilde{p}$.

Il nous faut préciser, que ce soit pour l'état initial comme pour les paramètres, les notations que nous allons utiliser pour distinguer valeurs exacte, estimée et courante de ces grandeurs physiques. Les valeurs qui ont réellement servi à produire les trajectoires de mesure sont notées respectivement $\bar{\rho}$ et $\bar{p}$. Les estimations qui peuvent être faites de ces grandeurs physiques sont notées respectivement $\hat{\rho}$ et $\hat{p}$ (le circonflexe dénotant toujours la notion d'estimation au cours de cette thèse). Enfin, lorsqu'on voudra parler d'un état initial ou d'une valeur de paramètres quelconques, nous les représenterons simplement ρ et p .

L'expérimentateur peut être amené à se baser sur des valeurs approchées de l'état initial et des paramètres, notées respectivement $\hat{\rho}$ et $\hat{p}$. Dans ce contexte, celui-ci doit construire un filtre *approché*, utilisant les résultats de mesure obtenus par le filtre réel, mais initialisé et itéré avec des valeurs $\hat{\rho}$ et $\hat{p}$ approchées. L'état de ce filtre approché à l'instant t_k est noté $\hat{\rho}_k$, et est calculé avec :

$$\hat{\rho}_0 = \hat{\rho}, \quad \hat{\rho}_k = \frac{\mathbf{K}_{k,y_k}^{\hat{p}}(\hat{\rho}_{k-1})}{\text{Tr}(\mathbf{K}_{k,y_k}^{\hat{p}}(\hat{\rho}_{k-1}))} \quad (2.53)$$

Chapitre 3

Tomographie d'état

Nota Bene. Ce chapitre est inspiré du papier [SCID⁺16] paru dans *Physical Review A*, écrit par Pierre Six, Philippe Campagne-Ibarcq, Igor Dotsenko^{*}, Alain Sarlette[†], Benjamin Huard et Pierre Rouchon.

3.1 Introduction

Dans le cadre des expériences de physique quantique, il est crucial de pouvoir déterminer précisément l'état du système, lors que le processus de sa préparation n'est pas assez bien connu. Contrairement à ce qu'on peut connaître en physique classique, la détermination de cet état $\bar{\rho}$, appelée *tomographie*, nécessite un grand nombre Ω de mesures indépendantes [PR04][Hares]. En effet, l'état d'un système quantique a, comme nous l'avons montré dans le chapitre d'introduction, une nature purement statistique, puisqu'il permet de calculer la loi de probabilité de toute mesure effectuée sur le système qu'il représente. Ces mesures sont habituellement représentées par un jeu de matrices positives ou nulles π_n , appelées POVM (*positive operator-valued measurement*), et telles que $\sum_n \pi_n = I$. La probabilité d'obtenir le résultat n est alors $\text{Tr}(\bar{\rho}\pi_n)$. Lorsque Ω est suffisamment grand, la tomographie de $\bar{\rho}$ utilise le fait que $\text{Tr}(\bar{\rho}\pi_n)$ doit être proche de Ω_n/Ω , où Ω_n est le nombre de mesures ayant donné le résultat n : $\sum_n \Omega_n = \Omega$.

Plusieurs méthodes de reconstruction existent déjà. Les méthodes d'entropie maximale [Buz04] et d'acquisition comprimée [GLF⁺10] (*compressed sensing*) sont utilisées dans le cas d'ensemble de mesures où l'information est incomplète. Lorsque, en revanche, l'information apportée par les mesures parcourt toutes les dimensions du système, on utilise le plus souvent la reconstruction par Maximum de Vraisemblance [LR09], qui consiste à choisir comme estimation de $\bar{\rho}$ l'état $\hat{\rho}_{ML}$ qui maximise la fonction de vraisemblance $\rho \mapsto \mathbb{P}(Y | \rho)$. Cette fonction parcourt l'ensemble des matrices densité ρ pour l'espace de Hilbert du système, et prend pour valeur la probabilité des résultats de mesure $Y = (N_n)_n$, en supposant que l'état initial du système vaut ρ . Pour l'ensemble de POVM $(\pi_n)_n$, la

^{*}. Laboratoire Kastler-Brossel, Ecole Normale Supérieure-PSL Research University, Université Pierre et Marie Curie-Sorbonne Universités, CNRS, Collège de France, 11 place Marcelin Berthelot, 75005 Paris, France

[†]. INRIA Paris, 2, rue Simone Iff, CS 42112, 75589 Paris Cedex 12, France

fonction de vraisemblance prend simplement la forme :

$$\mathbb{P}(Y | \rho) = \prod_n (\text{Tr}(\rho \pi_n))^{N_n} \quad (3.1)$$

Dans ce cas, le processus de mesure est supposé instantané, et exempt d'imperfections et de décohérence (rentrant dans le cadre présenté en sous-section 2.2.3). Dans certaines situations expérimentales spécifiques (par exemple, [OJTBG07]), il n'est pas très difficile de prendre en compte certaines imperfections de mesure. Le problème du calcul de la fonction de vraisemblance, en présence d'imperfections de mesure et de décohérence pendant la mesure, n'avait en revanche pas encore été résolu dans un cadre général : c'est ce que nous avons fait dans cette thèse. Les travaux de Belavkin ont permis de prendre en compte les imperfections de mesure et la décohérence dans le filtrage quantique [Bel92], ce que nous avons présenté en sous-section 2.2.4. Nous montrons ici comment nous utilisons les équations d'évolution présentées dans le chapitre précédent, qui permettent d'obtenir l'état conditionnel ρ_k ou ρ_t après un ensemble de mesures successives, pour donner l'expression générale de la fonction de vraisemblance. Son expression, donnée par l'équation (3.19), est une simple généralisation de l'équation donnée plus haut, dans laquelle les π_n sont remplacées par les états adjoints aux temps initiaux des Ω trajectoires de mesures. Ces états adjoints évoluent selon une équation maîtresse apparaissant dans le *quantum smoothing* [Tsa09, GW15] et correspondent à l'opérateur E définissant le *past quantum state* (ρ, E) introduit dans [GJM13].

Quand le support de la fonction de vraisemblance est principalement concentré autour de son maximum $\hat{\rho}_{ML}$, il est connu que $\hat{\rho}_{ML}$ est une bonne approximation de l'estimation Bayésienne $\hat{\rho}_{BM}$ définie par (voir, par exemple, [BK10]) :

$$\hat{\rho}_{BM} = \frac{\int_{\mathcal{D}_{\mathcal{H}}} \rho \mathbb{P}(Y | \bar{\rho} = \rho) p(\bar{\rho} = \rho) d\rho}{\int_{\mathcal{D}_{\mathcal{H}}} \mathbb{P}(Y | \bar{\rho} = \rho) p(\bar{\rho} = \rho) d\rho} \quad (3.2)$$

où $\mathcal{D}_{\mathcal{H}}$ est l'espace convexe des opérateurs densité sur l'espace de Hilbert $\mathcal{H}$ de dimension finie, et $p(\bar{\rho} = \rho)$ est une densité de probabilité *a priori*, quelconque, de ρ (voir, par exemple, l'ensemble Gaussien unitaire [Meh04]). L'approximation de $\hat{\rho}_{BM}$ par $\hat{\rho}_{ML}$ ainsi faite s'appuie sur les premiers termes du développement asymptotique d'intégrales de Laplace multi-dimensionnelles [BH86], en supposant certaines propriétés de régularité. Ces développements sont donnés en annexe C.

Nous montrons ensuite que ces développements asymptotiques permettent également d'obtenir un intervalle de confiance pour $\text{Tr}(\hat{\rho}_{ML} A)$, pour n'importe quel opérateur Hermiteen A . Pour ce faire, nous calculons de la même façon le premier terme du développement de la variance Bayésienne de $\text{Tr}(\hat{\rho}_{ML} A)$, que nous notons $\sigma_{BM}^2(A)$. Les expressions obtenues ne dépendent alors que des dérivées première et seconde de la fonction de log-vraisemblance à son maximum $\hat{\rho}_{ML}$. Lorsque ce maximum est de rang plein, on retrouve les équivalences asymptotiques classiques intervenant dans le Maximum de Vraisemblance, à savoir la matrice d'information de Fisher et la borne de Cramér-Rao [CMR05]. En revanche, si $\hat{\rho}_{ML}$ est un opérateur de rang faible, il se situe sur la frontière de l'ensemble des matrices densité admissibles, et une étude spécifique doit avoir lieu.

Après avoir exposé la théorie générale de la tomographie quantique d'état, nous appliquons celle-ci à deux différents ensembles de données expérimentales.

Le premier ensemble de mesures provient d'une expérience de *quantum non-demolition photon counting* dans une cavité, où la mesure se fait grâce aux atomes de Rydberg, présentée dans [RPP⁺15]. Dans ce cas, l'équation stochastique gouvernant le système est

une chaîne de Markov en temps discret (comme présenté dans le premier chapitre), dont l'état exprime le nombre de photons dans la cavité.

Le second jeu de données que nous utilisons est formé de mesures de fluorescence hétérodynes d'un qubit supraconducteur. Cette expérience est présentée dans [CIBF⁺14], et est décrite succinctement dans A.2. Ce système est observé de façon continue : l'évolution de son état est alors décrit par une équation différentielle stochastique, dans laquelle interviennent deux canaux de mesures (et donc deux processus de Wiener).

Dans les deux cas, nous commençons par réaliser une estimation $\hat{\rho}_{ML}$ par Maximum de Vraisemblance de l'état initial $\bar{\rho}$ du système. Ensuite, nous réalisons le même travail pour tous les instants t entre 0 et T (la durée des trajectoires), de la même façon, en omettant alors de prendre en compte les résultats de mesure acquis entre les instants 0 et t . On peut alors réaliser de manière "artificielle" la tomographie d'état du système à chaque instant.

Nous donnons aussi, pour chaque observable A ayant une pertinence physique, l'intervalle de confiance habituel, fixé à 95%, qu'on peut approximer à $\text{Tr}(\hat{\rho}_{ML}A) \pm 2\sigma_{ML}(A)$. Précisons la signification de cet intervalle de confiance. La formule que nous venons de donner signifie que, si l'expérimentateur devait produire à nouveau un nombre équivalent de trajectoires de mesure, dans les mêmes conditions physiques, et calculer grâce à ceux-ci la valeur $\hat{\rho}_{ML}$ correspondante, il y aurait environ 95% de chance que l'estimation de $\text{Tr}(\bar{\rho}A)$ de la valeur moyenne de A à l'instant initial, réalisée grâce à la seconde estimation de $\bar{\rho}$, soit comprise entre $\text{Tr}(\hat{\rho}_{ML}A) - 2\sigma_{ML}(A)$ et $\text{Tr}(\hat{\rho}_{ML}A) + 2\sigma_{ML}(A)$. Cela n'a rien à voir avec le caractère probabiliste de l'état quantique $\bar{\rho}$. Ici, $\bar{\rho}$ est considéré comme un paramètre, intervenant dans un tirage probabiliste dont le produit est la succession de trajectoires de mesure acquises expérimentalement. Nous verrons dans le prochain chapitre que tout paramètre constant intervenant dans l'équation d'évolution du système peut être estimé de façon similaire.

3.2 Préliminaires mathématiques

3.2.1 Notations utilisées

Considérons que nous avons à notre disposition un nombre Ω de trajectoires de mesure, démarrant du même état initial $\bar{\rho}$ que nous cherchons à estimer. Ces trajectoires de mesure sont indexées par $n \in \{1; \dots; \Omega\}$, et la trajectoire n est écrite $(y_k^{(n)})_{k=1\dots T_n}$. Le temps est ici discret, et est donc noté comme un entier k entre 0 et T_n , pour la trajectoire n . Pour chaque réalisation, le filtre quantique est capable de donner l'état à chaque instant du système, $\rho_k^{(n)}$, qui est calculé conditionnellement à l'état initial du filtre, et à toutes les mesures antérieures $(y_1^{(n)}, \dots, y_k^{(n)})$.

Cet état du système est un opérateur densité de l'espace de Hilbert $\mathcal{H}$, supposé ici de dimension finie. Nous notons $\mathcal{L}(\mathcal{H})$ l'ensemble des applications linéaires de $\mathcal{H}$ (représentés par la matrice carrée qui leur correspond), et $\mathcal{H}_{\text{herm}}$ le sous-ensemble de $\mathcal{L}(\mathcal{H})$ composé des applications hermitiennes (donc représentés par des matrices hermitiennes).

Pour alléger les notations, nous notons désormais $k^{(n)} = (k, n)$ l'instant k de la n -ième trajectoire, et $\tilde{y}_k^{(n)} = (k^{(n)}, y_k^{(n)})$, qui contient toute l'information nécessaire pour définir le processus de mesure au temps k de la trajectoire n .

Le filtre quantique, une version discrète du filtre continu de Belavkin, est défini à tout instant $k^{(n)}$ par l'ensemble d'applications complètement positives $(\mathbf{K}_{k^{(n)}, y})_{y \in \{1; \dots; J\}}$, appelées ensemble complet d'applications de Kraus. On suppose, dans ce chapitre, que la

formule permettant de calculer cet ensemble d'applications est connue à chaque instant. L'estimation ne porte donc que sur l'état initial $\bar{\rho}$ du système physique. Nous rappelons que le filtre réel est défini à chaque instant par :

$$\rho_0^{(n)} = \bar{\rho}, \quad \rho_k^{(n)} = \frac{\mathbf{K}_{\tilde{y}_k^{(n)}}(\rho_{k-1}^{(n)})}{\text{Tr}\left(\mathbf{K}_{\tilde{y}_k^{(n)}}(\rho_{k-1}^{(n)})\right)} \quad (3.3)$$

$$\mathbb{P}\left(y_k^{(n)} = y \mid \bar{\rho}, y_1^{(n)}, \dots, y_{k-1}^{(n)}\right) = \mathbb{P}\left(y_k^{(n)} = y \mid \rho_{k-1}^{(n)}\right) = \text{Tr}\left(\mathbf{K}_{\tilde{y}_k^{(n)}}(\rho_{k-1}^{(n)})\right) \quad (3.4)$$

Nous rappelons que ce filtre a une structure Markovienne, puisque toute l'information que l'état initial et les mesures réalisées avant l'instant k peuvent apporter est contenue dans l'état du filtre au temps précédent ; et que la formule des applications de Kraus ne dépend que de caractères physiques qui ne dépendent pas des états antérieurs du filtre : ces applications ne sont générées que par la valeur de $\tilde{y}_k^{(n)} = (k^{(n)}, y_k^{(n)})$.

3.2.2 Probabilité des trajectoires de mesure

Puisque nous nous plaçons du point de vue du physicien, qui ne connaît pas la valeur exacte $\bar{\rho}$ de l'état initial de chacune des réalisations du filtre, il nous faut utiliser un état initial estimé noté $\hat{\rho}$. C'est un opérateur densité sur l'espace de Hilbert $\mathcal{H}$, vérifiant donc les conditions que nous avons détaillées dans le paragraphe 2.2.2. A priori, le physicien, n'ayant à disposition que cet état initial approché $\bar{\rho}$, peut seulement définir un état approché du filtre, noté $\hat{\rho}_k^{(n)}$ à l'instant k de la trajectoire n , initialisé par $\hat{\rho}$ et itéré grâce à :

$$\hat{\rho}_0^{(n)} = \hat{\rho}, \quad \hat{\rho}_k^{(n)} = \frac{\mathbf{K}_{\tilde{y}_k^{(n)}}(\hat{\rho}_{k-1}^{(n)})}{\text{Tr}\left(\mathbf{K}_{\tilde{y}_k^{(n)}}(\hat{\rho}_{k-1}^{(n)})\right)} \quad (3.5)$$

Cherchons maintenant à déterminer la probabilité des sorties du filtre - les mesures $y_k^{(n)}$ -, en fonction donc de l'opérateur densité $\hat{\rho}$ choisi pour itérer les filtres de chaque trajectoire. Notons l'ensemble $Y^{(n)} = (y_k^{(n)})_{k=1\dots T_n}$ des mesures de la trajectoire n , et $Y = (Y^{(n)})_{n=1\dots\Omega}$ l'ensemble des mesures que le système a produites. La probabilité d'une trajectoire n est calculée avec :

$$\begin{aligned} \mathbb{P}\left(Y^{(n)} \mid \bar{\rho} = \hat{\rho}\right) &= \prod_{k=1}^{T_n} \mathbb{P}\left(y_k^{(n)} \mid \bar{\rho} = \hat{\rho}, y_1^{(n)}, \dots, y_{k-1}^{(n)}\right) \\ &= \prod_{k=1}^{T_n} \mathbb{P}\left(y_k^{(n)} \mid \hat{\rho}_{k-1}^{(n)}\right) \end{aligned} \quad (3.6)$$

où $\hat{\rho}_{k-1}^{(n)}$ est généré par le filtre de l'équation (3.5). Ce filtre est considéré alors comme un filtre approché, puisque son état initial n'est pas nécessairement celui du système physique : on ne peut donc pas affirmer que c'est ce filtre qui a généré les mesures $(y_k^{(n)})_{k=1\dots T_n}$ de la n -ième trajectoire. En revanche, la probabilité conditionnelle $\mathbb{P}\left(y_k^{(n)} \mid \hat{\rho}_{k-1}^{(n)}\right)$ suppose que la mesure $y_k^{(n)}$ a été réellement générée par ce filtre. Le passage de la première à la seconde ligne de l'équation (3.6) provient de la nature Markovienne des filtres quantiques. L'expression de la probabilité conditionnelle $\mathbb{P}\left(y_k^{(n)} \mid \hat{\rho}_{k-1}^{(n)}\right)$ est alors donnée par l'équation (3.4), en

remplaçant l'état $\rho_{k-1}^{(n)}$ par l'état $\hat{\rho}_{k-1}^{(n)}$ obtenu grâce au filtre approché. Nous obtenons alors :

$$\mathbb{P}\left(Y^{(n)} \mid \bar{\rho} = \hat{\rho}\right) = \prod_{k=1}^{T_n} \text{Tr}\left(\mathbf{K}_{\hat{y}_k^{(n)}}\left(\hat{\rho}_{k-1}^{(n)}\right)\right) \quad (3.7)$$

En injectant dans cette équation la formule (3.5) d'itération du filtre, il est facile d'en déduire, par récurrence, que la probabilité notée $\mathbb{P}_n(\hat{\rho})$ d'obtenir la trajectoire de mesure $(y_1^{(n)}, \dots, y_{T_n}^{(n)})$ en supposant que l'état initial du filtre soit $\bar{\rho}$ est :

$$\mathbb{P}_n(\hat{\rho}) = \mathbb{P}\left(Y^{(n)} \mid \bar{\rho} = \hat{\rho}\right) = \text{Tr}\left(\mathbf{K}_{\hat{y}_{T_n}^{(n)}} \circ \dots \circ \mathbf{K}_{\hat{y}_1^{(n)}}(\hat{\rho})\right) \quad (3.8)$$

Enfin, puisque toutes les trajectoires de mesure sont supposées indépendantes, et de même état initial, la probabilité d'obtenir *toutes* les trajectoires de mesure, information exprimée par la notation $Y = (Y^{(n)})_{n=1\dots\Omega} = (y_k^{(n)})_{\substack{n=1\dots\Omega, \\ k=1\dots T_n}}$, en supposant un état initial $\hat{\rho}$, a un sens. Cette probabilité, notée $\mathbb{P}(\hat{\rho})$, est obtenue par la formule :

$$\mathbb{P}(\hat{\rho}) = \mathbb{P}(Y \mid \bar{\rho} = \hat{\rho}) = \prod_{n=1}^{\Omega} \mathbb{P}\left(Y^{(n)} \mid \bar{\rho} = \hat{\rho}\right) = \prod_{n=1}^{\Omega} \text{Tr}\left(\mathbf{K}_{\hat{y}_{T_n}^{(n)}} \circ \dots \circ \mathbf{K}_{\hat{y}_1^{(n)}}(\hat{\rho})\right) \quad (3.9)$$

Enfin, remarquons que, puisque nous cherchons à estimer l'état initial réel $\bar{\rho}$ du système physique, ce que nous cherchons est plutôt la densité de probabilité de tel ou tel état initial "candidat", conditionnellement à toutes les trajectoires de mesure que le système a produites. Cette densité de probabilité s'obtient facilement grâce à la loi de Bayes. Cependant, puisqu'on doit définir ici des densités de probabilité, il convient d'apporter quelques précisions préalables.

Rappelons que $\mathcal{D}_{\mathcal{H}}$ est l'ensemble des opérateurs densité admissibles sur l'espace de Hilbert $\mathcal{H}$; l'équation (2.18) nous enseigne que :

$$\mathcal{D}_{\mathcal{H}} = \{\rho \in \mathcal{L}(\mathcal{H}) \mid \text{Tr}(\rho) = 1, \rho \geq 0\}$$

$\mathcal{D}_{\mathcal{H}}$ est donc un sous-ensemble de l'espace vectoriel $\mathcal{H}_{\text{herm}}$ des opérateurs Hermitiens sur $\mathcal{H}$. Notons $\mathcal{U}_{\rho}$ un sous-ensemble de $\mathcal{D}_{\mathcal{H}}$, de volume infinitésimal dv , défini comme la restriction à $\mathcal{D}_{\mathcal{H}}$ d'un voisinage de ρ dans $\mathcal{H}_{\text{herm}}$. Pour toute information conditionnelle Y provenant des résultats de mesure effectués sur le système initialisé en $\bar{\rho}$, la loi de Bayes peut s'écrire :

$$\mathbb{P}(\bar{\rho} \in \mathcal{U}_{\rho} \mid Y) = \frac{1}{\mathbb{P}(Y)} \int_{\sigma \in \mathcal{U}_{\rho}} \mathbb{P}(Y \mid \bar{\rho} = \sigma) d\mathbb{P}_0(\bar{\rho} = \sigma) \quad (3.10)$$

où $\sigma \mapsto d\mathbb{P}_0(\bar{\rho} = \sigma)$ est la densité de probabilité de $\bar{\rho} = \sigma$, appelée initiale, car elle ne prend en compte aucune information sur σ . Nous imposons que la mesure qui décrit cette probabilité soit absolument continue³, ce qui nous permet de noter $d\mathbb{P}_0(\sigma) = p_0(\bar{\rho} = \sigma)d\sigma$, où $d\sigma$ est l'élément de volume infinitésimal autour de σ , défini par le produit scalaire de Frobenius, et $\sigma \mapsto p_0(\bar{\rho} = \sigma)$ la densité de probabilité initiale de l'état σ . Nous choisirons dans nos validations expérimentales que cette densité soit constante.

3. Cela signifie que, en notant μ_p la mesure telle que $\mu_p(A) = \int_{\sigma \in A} d\mathbb{P}_0(\bar{\rho} = \sigma)$ et μ la mesure de Lebesgue, alors pour tout mesurable A tel que $\mu(A) = 0$, on a $\mu_p(A) = 0$.

Lorsque l'élément de volume $\mathcal{U}_\rho$ est infinitésimal, on peut écrire, puisque la fonction $\mathbb{P}(Y | \bar{\rho} = \sigma)$, calculée grâce à l'équation (3.9), est analytique :

$$\mathbb{P}(\bar{\rho} \in \mathcal{U}_\rho | Y) = \frac{\mathbb{P}(Y | \bar{\rho} = \rho) p_0(\bar{\rho} = \rho)}{\mathbb{P}(Y)} dv + \mathcal{O}(|dv|^2)$$

Ainsi, on peut définir une densité de probabilité concernant la valeur de $\bar{\rho}$ conditionnellement à Y , vérifiant :

$$p(\bar{\rho} = \rho | Y) = \frac{\mathbb{P}(Y | \bar{\rho} = \sigma) p_0(\bar{\rho} = \sigma)}{\mathbb{P}(Y)} \quad (3.11)$$

C'est cette fonction que nous avons à maximiser selon ρ . Or, puisque nous considérons que $\rho \mapsto p_0(\bar{\rho} = \rho)$ est constante (ou d'amplitude devenant peu significatives devant celle de $\rho \mapsto \mathbb{P}(Y | \bar{\rho} = \rho)$), l'estimation par Maximum de Vraisemblance (*maximum-likelihood estimation*) de l'état initial $\bar{\rho}$ revient donc à trouver la matrice densité sur $\mathcal{H}$, notée $\hat{\rho}_{ML}$, qui maximise la fonction de vraisemblance :

$$\rho \in \mathcal{D}_{\mathcal{H}} \mapsto F(\rho) = \mathbb{P}(Y | \bar{\rho} = \rho) = \prod_{n=1}^{\Omega} \text{Tr} \left(\mathbf{K}_{\tilde{y}_{T_n}^{(n)}} \circ \dots \circ \mathbf{K}_{\tilde{y}_1^{(n)}}(\rho) \right) \quad (3.12)$$

Pour que le calcul d'une telle fonction soit possible, nous étudierons désormais son logarithme. Puisque le logarithme est une fonction strictement croissante, l'estimation par maximum de vraisemblance est donc possible en choisissant la matrice densité $\hat{\rho}_{ML}$ qui maximise :

$$\begin{aligned} \rho \in \mathcal{D}_{\mathcal{H}} \mapsto F_{\log}(\rho) &= \sum_{n=1}^{\Omega} \log(\mathbb{P}_n(\rho)) \\ &= \sum_{n=1}^{\Omega} \log \left(\text{Tr} \left(\mathbf{K}_{\tilde{y}_{T_n}^{(n)}} \circ \dots \circ \mathbf{K}_{\tilde{y}_1^{(n)}}(\rho) \right) \right) \end{aligned} \quad (3.13)$$

3.2.3 États adjoints

A première vue, il peut paraître difficile d'effectuer la maximisation de la fonction $F_{\log}$. En effet, calculer la valeur de cette fonction pour une matrice densité donnée nécessite d'appliquer des applications de Kraus partielles un nombre de fois égal au nombre de mesures que le système a produites.

Or, nous présentons une réécriture de cette fonction, rendant beaucoup plus rapide le calcul de la fonction $F_{\log}$ en une succession de matrices densité testées.

Applications de Kraus adjointes

Rappelons la structure particulière d'une application de Kraus partielle. Il s'agit d'un super-opérateur de $\mathcal{H}$ (voir le paragraphe 2.2.4), noté $\mathbf{K}$, défini par un ensemble fini I de $\mathbb{N}$ et $(M_l)_{l \in I}$ un ensemble de matrices sur $\mathcal{H}$, selon la formule :

$$\forall \rho \in \mathcal{L}(\mathcal{H}), \mathbf{K}(\rho) = \sum_{l \in I} M_l \rho M_l^\dagger \quad (3.14)$$

Nous appelons ensuite un ensemble complet d'applications de Kraus, un ensemble $(\mathbf{K}_y)_{y=1\dots J}$ d'applications de Kraus partielles, associées respectivement aux ensembles de matrices $(M_l)_{l \in I_y}$, qui vérifient l'équation suivante (*completeness equation*) :

$$\sum_{y=1}^J \sum_{l \in I_y} M_l^\dagger M_l = I_{\mathcal{H}}$$

Nous introduisons, pour chaque application de Kraus partielle $\mathbf{K}$ dont l'expression est donnée par l'équation (3.14), son application de Kraus adjointe notée $\mathbf{K}^*$, telle que, pour tous opérateurs A et B sur $\mathcal{H}$, on ait la relation suivante :

$$\text{Tr} (A \mathbf{K} (B)) = \text{Tr} (\mathbf{K}^* (A) B) \quad (3.15)$$

L'existence, pour chaque application de Kraus, de son application adjointe découle de sa structure, que nous venons de rappeler. En effet :

$$\begin{aligned} \text{Tr} (A \mathbf{K} (B)) &= \text{Tr} \left(A \left(\sum_{l \in I} M_l B M_l^\dagger \right) \right) \\ &= \sum_{l \in I} \text{Tr} (A M_l B M_l^\dagger) \\ &= \text{Tr} \left(\left(\sum_{l \in I} M_l^\dagger A M_l \right) B \right) \end{aligned} \quad (3.16)$$

Cette réécriture montre de manière évidente que l'expression de ces applications adjointes est :

$$\mathbf{K} (\rho) = \sum_{l \in I} M_l \rho M_l^\dagger \implies \mathbf{K}^* (\rho) = \sum_{l \in I} M_l^\dagger \rho M_l \quad (3.17)$$

Filtre quantique adjoint

A l'aide de ces applications de Kraus adjoints, nous pouvons générer, à partir d'une trajectoire de mesure $(y_k^{(n)})_{k=1\dots T_n}$, un filtre quantique adjoint normalisé (voir aussi [Tsa09] et [GJM13]) dont l'état est $E_k^{(n)}$, et qui est défini par la relation suivante :

$$E_{T_n}^{(n)} = \frac{I_{\mathcal{H}}}{\text{Tr} (I_{\mathcal{H}})}, \quad E_{k-1}^{(n)} = \frac{\mathbf{K}_{\tilde{y}_k^{(n)}}^* (E_k^{(n)})}{\text{Tr} (\mathbf{K}_{\tilde{y}_k^{(n)}}^* (E_k^{(n)}))} \quad \forall k \geq 1 \quad (3.18)$$

L'avantage de ce filtre adjoint est qu'il est construit à partir des résultats de mesure $y_k^{(n)}$, sans dépendre de l'état initial $\bar{\rho}$ dont on ignore la valeur. Voyons comment on peut se servir de ce filtre adjoint pour calculer la densité de probabilité $\rho \mapsto F_{\log}(\rho)$ définie par l'équation (3.13), en introduisant l'état adjoint $E_k^{(n)}$ dans l'expression de $\mathbb{P}(Y^{(n)} | \bar{\rho} = \rho)$. Remarquons d'abord que :

$$\begin{aligned} \mathbb{P}(Y^{(n)} | \bar{\rho} = \rho) &= \text{Tr} \left(I_{\mathcal{H}} \mathbf{K}_{\tilde{y}_{T_n}^{(n)}} \circ \dots \circ \mathbf{K}_{\tilde{y}_1^{(n)}} (\rho) \right) \\ &= \text{Tr} \left(\mathbf{K}_{\tilde{y}_{T_n}^{(n)}}^* (I_{\mathcal{H}}) \mathbf{K}_{\tilde{y}_{T_n-1}^{(n)}} \circ \dots \circ \mathbf{K}_{\tilde{y}_1^{(n)}} (\rho) \right) \\ &= \text{Tr} \left(\mathbf{K}_{\tilde{y}_1^{(n)}}^* \circ \dots \circ \mathbf{K}_{\tilde{y}_{T_n}^{(n)}}^* (I_{\mathcal{H}}) \rho \right) \end{aligned}$$

Or, on reconnaît dans cette expression l'itération d'applications de Kraus adjointes qui génère le filtre adjoint (3.18). Il est alors facile de montrer par récurrence que la densité de probabilité de la trajectoire n en supposant la valeur initiale ρ vaut :

$$\mathbb{P}\left(Y^{(n)} \mid \bar{\rho} = \rho\right) = \text{Tr}\left(E_0^{(n)}\rho\right) \prod_{k=1}^{T_n} \text{Tr}\left(\mathbf{K}_{\tilde{y}_k^{(n)}}^*\left(E_k^{(n)}\right)\right) \quad (3.19)$$

et la fonction $F_{\log}$ que nous devons maximiser s'écrit alors :

$$F_{\log}(\rho) = \sum_{n=1}^{\Omega} \left[\log\left(\text{Tr}\left(E_0^{(n)}\rho\right)\right) + \sum_{k=1}^{T_n} \log\left(\text{Tr}\left(\mathbf{K}_{\tilde{y}_k^{(n)}}^*\left(E_k^{(n)}\right)\right)\right) \right] \quad (3.20)$$

L'avantage de cette notation est qu'elle ne fait intervenir, en ce qui concerne sa dépendance en ρ , qu'une somme de n produits scalaires, entre ρ lui-même et les états adjoints "finaux" $E_0^{(n)}$ (pour alléger les formules, nous notons désormais $E_0^{(n)} = E^{(n)}$). Or, le calcul de ceux-ci ne dépend précisément pas de l'état initial ρ testé. Ainsi, l'estimation par maximum de vraisemblance de l'état initial consiste à choisir la matrice densité $\hat{\rho}_{ML}$ qui maximise la fonction :

$$\rho \in \mathcal{D}_{\mathcal{H}} \mapsto f(\rho) = \sum_{n=1}^{\Omega} \log\left(\text{Tr}\left(E^{(n)}\rho\right)\right) \quad (3.21)$$

3.3 Estimation par Maximum de Vraisemblance

L'estimation par Maximum de Vraisemblance consiste donc à choisir comme valeur estimée $\hat{\rho}_{ML}$ de $\bar{\rho}$ la valeur qui, choisie comme état initial du système physique, maximise la log-probabilité des résultats de mesure $y_k^{(n)}$. Son expression est formulée dans l'équation (3.21). Nous cherchons également à donner la variance d'une telle estimation. Cette variance sera notée σ_{ML}^2 . C'est un tenseur d'ordre 4 dans $\mathcal{H}$; aussi, pour des raisons pratiques, nous ne donnerons pas par la suite sa valeur explicite, mais pour chaque observable A sur $\mathcal{H}$, la valeur $\sigma_{ML}^2(A)$, qui correspond à la variance de l'estimation $\text{Tr}(\hat{\rho}_{ML}A)$ de la valeur moyenne $\text{Tr}(\bar{\rho}A)$ de A à l'instant initial. Sa définition exacte sera rappelée dans le paragraphe 3.3.2 de cette section.

On peut trouver dans ⁴ plus de précisions générales sur l'estimation par Maximum de Vraisemblance.

3.3.1 Contexte de l'optimisation

Rappelons d'abord que la fonction à maximiser est la fonction f définie par :

$$\rho \in \mathcal{D}_{\mathcal{H}} \mapsto f(\rho) = \sum_{n=1}^{\Omega} \log\left(\text{Tr}\left(E^{(n)}\rho\right)\right)$$

Cette fonction est une somme de logarithmes de produits scalaires sur l'argument ρ . Or, le logarithme préserve la concavité des fonctions : la fonction f est donc une fonction concave. Nous pouvons également remarquer que l'ensemble $\mathcal{D}_{\mathcal{H}}$ est convexe, puisque la somme de deux matrices positives forme une matrice positive. L'ensemble des valeurs de ρ maximisant cette fonction sur $\mathcal{D}_{\mathcal{H}}$ est donc un ensemble convexe, à l'intérieur duquel le

4. ajouter les deux references de E. Moulines.

gradient de la fonction s'annule (voir, par exemple, [BV09]). On suppose néanmoins que cet ensemble se réduit à un point, de sorte que l'on peut noter $\hat{\rho}_{ML}$ cette matrice densité de $\mathcal{H}$ qui vérifie :

$$\forall \rho \in \mathcal{D}_{\mathcal{H}}, \rho \neq \hat{\rho}_{ML} \implies f(\rho) < f(\hat{\rho}_{ML})$$

Les conditions d'optimalité seront donc différentes, que $\hat{\rho}_{ML}$ soit à l'intérieur ou sur la frontière de $\mathcal{D}_{\mathcal{H}}$. Il faut également tenir compte de la contrainte affine $\text{Tr}(\rho) = 1$ que vérifie tout élément de $\mathcal{D}_{\mathcal{H}}$, de sorte que, même si $\hat{\rho}_{ML}$ est un point intérieur de $\mathcal{D}_{\mathcal{H}}$, le gradient en ce point a seulement une composante nulle sur l'espace tangent associé en $\hat{\rho}_{ML}$, à savoir l'ensemble des matrices Hermitiennes de trace nulle.

Topologie de l'espace de maximisation

Nous munissons tout d'abord l'ensemble $\mathcal{H}_{\text{herm}}$ des opérateurs Hermitiens de $\mathcal{H}$ d'un produit scalaire : il s'agit du produit scalaire de Frobenius, et sa norme associée. Pour deux opérateurs A et B de $\mathcal{H}_{\text{herm}}$, leur produit scalaire est simplement défini par :

$$\langle A, B \rangle = \text{Tr}(AB) \quad (3.22)$$

et la norme associée est notée $\|A\| = \sqrt{\langle A, A \rangle}$.

Comme nous l'avons dit, l'espace $\mathcal{D}_{\mathcal{H}}$ des opérateurs densité sur $\mathcal{H}$ est un espace convexe, et un sous-espace de $\mathcal{H}_{\text{herm}}$. Notre ensemble de maximisation, $\mathcal{D}_{\mathcal{H}}$, hérite donc du produit scalaire défini ci-dessus. En outre, c'est un ensemble fermé et borné (puisque la norme au carré d'une matrice positive est la somme des carrés de ses valeurs propres, et que les valeurs propres d'une matrice densité sont toutes de carré inférieur à 1). L'espace $\mathcal{D}_{\mathcal{H}}$ est donc un compact. Cet espace est en partie défini par la contrainte $\text{Tr}(\rho) = 1$: l'espace tangent en chaque point de $\mathcal{D}_{\mathcal{H}}$ est donc contenu dans l'ensemble des opérateurs Hermitiens de $\mathcal{H}$ de trace nulle, noté $\mathcal{H}_{\text{herm}}^0$.

Déterminons enfin la nature du plan tangent à chaque point de $\mathcal{D}_{\mathcal{H}}$ (à chaque matrice densité admissible de $\mathcal{H}$) : cela aura une certaine importance par la suite. En un point ρ de l'intérieur de $\mathcal{D}_{\mathcal{H}}$ (une matrice ρ de rang plein), le plan tangent est simplement $\mathcal{H}_{\text{herm}}^0$. Lorsque le point ρ appartient à la frontière de $\mathcal{D}_{\mathcal{H}}$, deux cas de figure peuvent se présenter.

- Lorsque le noyau de la matrice ρ est seulement de dimension 1, on note celui-ci $\text{Ker}(\rho) = \text{Vect}(|\psi\rangle)$ avec $|\psi\rangle$ un vecteur normé de $\mathcal{H}$. La surface de $\mathcal{D}_{\mathcal{H}}$ au point ρ est alors localement une variété affine $\mathcal{C}^\infty$ de $\mathcal{H}_{\text{herm}}^0$, avec pour normale sortante la matrice $M_\perp = -|\psi\rangle\langle\psi|$, opposé du projecteur sur $|\psi\rangle$. Le plan tangent de $\mathcal{D}_{\mathcal{H}}$ en ρ est donc composé des matrices A de $\mathcal{H}_{\text{herm}}^0$ vérifiant $\langle\psi|A|\psi\rangle = 0$.
- Lorsque le noyau de la matrice ρ est de dimension supérieure, la surface n'est plus régulière, et les calculs de variance d'estimation en un tel point deviennent très difficiles à réaliser. Le plan tangent de $\mathcal{D}_{\mathcal{H}}$ en un tel point ρ est composé des matrices A de $\mathcal{H}_{\text{herm}}^0$ vérifiant $\langle\psi|A|\psi\rangle = 0$ pour tout vecteur $|\psi\rangle$ du noyau de ρ . Par conséquent, cet espace tangent n'est pas de dimension pleine et la surface de $\mathcal{D}_{\mathcal{H}}$ en ρ n'est pas régulière.

Pour toute matrice ρ de $\mathcal{D}_{\mathcal{H}}$, le plan tangent à $\mathcal{D}_{\mathcal{H}}$ en ρ est donc constitué des matrices de $\mathcal{H}_{\text{herm}}^0$ orthogonales à tout projecteur sur un vecteur du noyau de ρ . Seulement, lorsque ρ a un noyau d'une dimension strictement supérieure à 1, cette condition implique que la surface de $\mathcal{D}_{\mathcal{H}}$ ne soit pas régulière au voisinage de ρ , et cela complique grandement tout calcul de variance d'estimation lorsque la valeur estimée $\hat{\rho}_{ML}$ se trouve être une telle matrice.

Gradients et Hessiens de la fonction de vraisemblance

Le gradient de la fonction de maximisation f est assez facile à calculer. En effet, si l'on note $\nabla f(\rho)$ le gradient de f en ρ tel que, pour toute matrice Hermitienne infinitésimale $\delta\rho$, ce gradient vérifie : $f(\rho + \delta\rho) = f(\rho) + \text{Tr}(\nabla f(\rho) \cdot \delta\rho) + \mathcal{O}(\|\delta\rho\|^2)$, la formule permettant d'obtenir ce gradient est :

$$\nabla f(\rho) = \sum_{n=1}^{\Omega} \frac{E^{(n)}}{\text{Tr}(\rho E^{(n)})} \quad (3.23)$$

De même, il est possible de définir le Hessian de cette fonction. Notons ce Hessian $\nabla^2 f(\rho)$ au point ρ : il s'agit d'un tenseur d'ordre 2 dans l'espace $\mathcal{H}_{\text{herm}}$ des matrices hermitiennes de $\mathcal{H}$. Ainsi, pour deux matrices A et B hermitiennes, le produit $A \cdot (\nabla^2 f(\rho)) \cdot B$ est un scalaire, de telle sorte que, pour toute matrice $\delta\rho$ hermitienne et infinitésimale, et tout élément ρ de $\mathcal{H}$, on ait :

$$f(\rho + \delta\rho) = f(\rho) + \text{Tr}(\nabla f(\rho) \cdot \delta\rho) + \frac{\delta\rho \cdot (\nabla^2 f(\rho)) \cdot \delta\rho}{2} + \mathcal{O}(\|\delta\rho\|^3)$$

Le Hessian de la fonction f en un opérateur ρ de $\mathcal{D}_{\mathcal{H}}$ vérifie, pour tous opérateurs Hermitiens A et B sur $\mathcal{H}$:

$$A \cdot (\nabla^2 f(\rho)) \cdot B = - \sum_{n=1}^{\Omega} \frac{\text{Tr}(A E^{(n)}) \text{Tr}(B E^{(n)})}{\text{Tr}^2(\rho E^{(n)})} \quad (3.24)$$

Conditions d'optimalité

De plus, nous donnons, adaptée du critère d'optimalité classique (voir encore [BV09]) intervenant dans l'optimisation des fonctions convexes, la condition nécessaire et suffisante vérifiée par $\hat{\rho}_{ML}$. Notons d'abord ∇f_{ML} la valeur $\nabla f(\hat{\rho}_{ML})$ du gradient en ce point. Séparons deux cas : le cas où $\hat{\rho}_{ML}$ est un point intérieur de $\mathcal{D}_{\mathcal{H}}$ (donc de rang plein), et le cas où il est sur la frontière de $\mathcal{D}_{\mathcal{H}}$ ($\hat{\rho}_{ML}$ non inversible). Lorsque $\hat{\rho}_{ML}$ est de rang plein, la condition d'optimalité est donc simplement qu'il existe un scalaire λ_{ML} positif tel que :

$$\nabla f_{ML} = \lambda_{ML} I_{\mathcal{H}}, \quad (3.25)$$

car le gradient ∇f_{ML} doit avoir une projection orthogonale nulle sur $\mathcal{H}_{\text{herm}}^0$, et parce que la trace du gradient $\nabla f(\rho)$ est positive en tout point ρ de $\mathcal{D}_{\mathcal{H}}$.

Lorsque $\hat{\rho}_{ML}$ est de rang faible, la condition d'optimalité est un peu plus difficile à construire. Il nous faut tout d'abord noter P_{ML} le projecteur orthogonal sur l'image de $\hat{\rho}_{ML}$. Là encore, il existe un scalaire λ_{ML} positif tel que :

$$[\hat{\rho}_{ML}, \nabla f_{ML}] = 0, \quad \text{et} \quad \lambda_{ML} P_{ML} \leq \nabla f_{ML} \leq \lambda_{ML} I_{\mathcal{H}} \quad (3.26)$$

Ces propriétés s'appuient sur le théorème suivant, dont la démonstration est donnée en annexe, au lemme 6.

Théorème 1. Supposons que la matrice Hermitienne ρ^* , de dimension $N_{\mathcal{H}}$, maximise sur $\mathcal{D}_{\mathcal{H}}$ l'ensemble des matrices densité de dimension $N_{\mathcal{H}}$ la fonction $\rho \in \mathcal{D}_{\mathcal{H}} \mapsto f(\rho)$ définie par l'équation (3.21). Alors ρ^* vérifie les conditions suivantes :

- $\text{Tr}(\rho^* E^{(n)}) > 0$ pour chaque $n = 1, \dots, \Omega$;
- $[\rho^*, \nabla f(\rho^*)] = \rho^* \nabla f(\rho^*) - \nabla f(\rho^*) \rho^* = 0$ ($\rho \mapsto \nabla f(\rho) = \sum_{n=1}^{\Omega} \frac{E^{(n)}}{\text{Tr}(\rho E^{(n)})}$ est le gradient de f pour le produit de Frobenius);

- $\exists \lambda^* > 0$ tel que $\lambda^* P^* \leq \nabla f(\rho^*) \leq \lambda^* I_{\mathcal{H}}$, où P^* est le projecteur orthogonal sur l'image de ρ^* , et $I_{\mathcal{H}}$ l'opérateur identité.

Ces conditions sont suffisantes et caractérisent un unique maximum lorsque, de plus, l'espace vectoriel généré par les matrices $E^{(n)}$ correspond à l'ensemble des matrices Hermitiennes de dimension $N_{\mathcal{H}}$.

La condition (3.26) est en fait équivalente à ce que, pour tout opérateur densité ρ sur l'espace de Hilbert $\mathcal{H}$, on conserve $\text{Tr}((\rho - \hat{\rho}_{ML}) \nabla f_{ML}) \leq 0$. Lorsque $\hat{\rho}_{ML}$ est de rang plein, le projecteur orthogonal P_{ML} vaut $I_{\mathcal{H}}$: la condition (3.26) est donc générale, et s'applique à toute matrice $\hat{\rho}_{ML}$.

Lorsque $\hat{\rho}_{ML}$ est de rang faible, le point $\hat{\rho}_{ML}$ est sur la frontière $\partial \mathcal{D}_{\mathcal{H}}$ de $\mathcal{D}_{\mathcal{H}}$. Le projecteur P_{ML} est alors inférieur à l'identité (ses valeurs propres valent 0 ou 1). Notons $\nabla f_{ML}^{\parallel}$ la projection du gradient ∇f_{ML} sur $\mathcal{H}_{\text{herm}}^0$. Dans ce cas, la première condition d'optimalité de l'équation (3.26) signifient alors que $\nabla f_{ML}^{\parallel}$ pointe vers l'extérieur de $\mathcal{D}_{\mathcal{H}}$, et est orthogonal en $\hat{\rho}_{ML}$ à la sous-variété constituée des matrices densité de même rang que $\hat{\rho}_{ML}$: ∇f_{ML} est la somme d'une matrice proportionnelle à $I_{\mathcal{H}}$ et d'une matrice dont l'image est contenue dans le noyau de $\hat{\rho}_{ML}$.

Méthodes de descente envisagées

Nous avons vu plus haut que la fonction de maximisation est concave, et que l'ensemble de maximisation est convexe. L'optimisation est donc très simple à réaliser.

Direction du gradient On peut utiliser dans ce cas l'optimisation la plus basique qui soit, c'est-à-dire celle qui choisit comme direction de descente la valeur du gradient. En notant $\hat{\rho}_j$ le j -ième état estimé par l'optimisation, l'état suivant est alors donné par :

$$\hat{\rho}_{j+1} = \mathcal{P}_{\mathcal{D}_{\mathcal{H}}}(\hat{\rho}_j + \alpha_j \nabla f(\hat{\rho}_j))$$

Ici, α_j est un scalaire positif, qui peut être choisi par optimisation linéaire, ou simplement être fixé à une valeur constante ϵ . $\mathcal{P}_{\mathcal{D}_{\mathcal{H}}}$ représente la projection orthogonale sur l'ensemble des matrices admissibles $\mathcal{D}_{\mathcal{H}}$. Le gradient étant aussi une matrice Hermitienne, elle se réalise en projetant d'abord l'état sur l'ensemble des matrices de trace un, puis grâce à une autre projection sur l'ensemble des matrices positives. La façon dont cette projection se réalise est détaillée en annexe B.6.1.

Descente par algorithme EM Un algorithme alternatif peut être utilisé dans le cas où les matrices d'état ρ , et les états adjoints $E^{(n)}$ restent diagonaux dans une base orthonormale commune (c'est ce que nous verrons dans le cas des trajectoires générées par la boîte à photons). Cet algorithme se construit grâce au théorème suivant :

Théorème 2. Définissons un opérateur $\hat{\rho}_j$ et un ensemble d'opérateurs $(E^{(n)})_{n=1, \dots, \Omega}$ du même espace de Hilbert fini $\mathcal{H}$, tous ces opérateurs étant diagonaux. Définissons également la fonction f comme :

$$\rho \in \mathcal{D}_{\mathcal{H}} = \mathcal{L}(\mathcal{H}) \mapsto f(\rho) = \sum_{n=1}^{\Omega} \log(\text{Tr}(\rho E^{(n)}))$$

Alors la matrice $\hat{\rho}_{j+1}$ définie comme :

$$\hat{\rho}_{j+1} = \frac{\hat{\rho}_j}{\Omega} \nabla f(\hat{\rho}_j) = \frac{1}{\Omega} \sum_{n=1}^{\Omega} \frac{\hat{\rho}_j E^{(n)}}{\text{Tr}(\hat{\rho}_j E^{(n)})} \quad (3.27)$$

vérifie $f(\hat{\rho}_{j+1}) \geq f(\hat{\rho}_j)$.

La preuve de ce théorème est donnée en annexe [B.6.2](#). Cette preuve est basée sur le principe de l'algorithme *Expectation-Maximization*, ce qui donne son nom à cette méthode de descente.

Ce théorème n'est pas valable lorsque les matrices en question ne sont pas diagonales.

3.3.2 Comportement asymptotique et variance d'estimation

Dans cette sous-section, nous nous basons sur les ouvrages de van der Vaart [[vdV98](#)] et de Lehmann et Castella [[LC98](#)], auxquels nous renvoyons les lecteurs pour plus de précisions sur les arguments mathématiques présentés ici.

Biais d'estimation

Nous nous replaçons ici dans le cadre général de l'estimation de paramètre par mesures répétées indépendantes. Soit θ un paramètre à estimer, évoluant dans Θ l'ensemble des paramètres admissibles - sous-ensemble d'un $\mathbb{R}$ -espace vectoriel de dimension finie -, et $\bar{\theta} \in \Theta$ la valeur réelle de θ . Le système génère des mesures aléatoires indépendantes et de même loi notées X_i . Ces X_i sont des réalisations de la densité de probabilité :

$$X \mapsto p(X|\bar{\theta})$$

De même, on connaît la loi conditionnelle $p(\cdot|\theta)$ pour toutes les valeurs possibles $\theta \in \Theta$.

Soit Ω un nombre de réalisations donné ; on note $\mathbf{X}^\Omega = (X_1, \dots, X_\Omega)$. Soit $C(\mathbf{X}^\Omega)$ un estimateur quelconque de θ prenant en argument Ω réalisations. On dit que cet estimateur est *non biaisé* si et seulement si :

$$\mathbb{E}_{\mathbf{X}^\Omega} (C(\mathbf{X}^\Omega) | \bar{\theta}) = \bar{\theta}$$

Le Maximum de Vraisemblance, qui consiste à rechercher comme valeur estimée une solution de $\frac{\partial}{\partial \theta} (p(\mathbf{X}^\Omega | \theta))$ - où $p(\mathbf{X}^\Omega | \bar{\theta}) = \prod_{n=1}^{\Omega} p(X_n | \bar{\theta})$ -, n'est pas une estimation non biaisée, en règle générale. Néanmoins, certains résultats existent, qui tendent à montrer que cette estimation tend à être non biaisée lorsque le nombre de réalisations prises en compte devient asymptotiquement grand. Lorsque l'espace de maximisation Θ est fini, la Maximum de Vraisemblance est alors un estimateur consistant. Lorsqu'à l'inverse Θ est un espace non fini, on peut donner, sous certaines conditions de régularité que nous détaillerons ensuite, le résultat suivant.

Lemme 1 ([[LC98](#)]). L'équation :

$$\frac{\partial}{\partial \theta} (p(\mathbf{X}^\Omega | \theta)) = 0$$

a une racine notée $\hat{\theta}_\Omega(\mathbf{X}^\Omega)$ qui tend en probabilité vers la valeur réelle $\bar{\theta}$ (celle qui génère les valeurs $\mathbf{X}^\Omega$) lorsque Ω tend vers $+\infty$.

Les conditions nécessaires pour obtenir ce résultat sont :

1. les densités de probabilité $X \mapsto p(X|\theta)$ sont distinctes pour chaque θ
2. les densités de probabilité $X \mapsto p(X|\theta)$ ont le même support
3. il existe un voisinage ouvert $\mathcal{U}$ de $\bar{\theta}$ dans Ω

4. pour presque tout X , la quantité $p(X|\theta)$ est différentiable en θ sur Θ

En rappelant que, dans le cas de la tomographie d'état quantique, le paramètre est $\rho \in \mathcal{D}_{\mathcal{H}}$ de valeur réelle $\bar{\rho}$ et la réalisation est une trajectoire de mesures, on s'aperçoit que ces propriétés sont vérifiées à deux conditions : que pour toute application de Kraus $\mathbf{K}_y$ intervenant dans la génération et pour toute matrice densité ρ , $\mathbf{K}_y(\rho)$ ne soit pas nul ; et, surtout, que $\bar{\rho}$ soit une matrice densité dans l'intérieur de $\mathcal{D}_{\mathcal{H}}$, donc de rang plein. La première hypothèse est vérifiée dans les systèmes considérés ici, et on peut supposer que la seconde l'est aussi, car une matrice strictement de rang faible n'a pas de sens physique.

Ainsi, le résultat de non-biais asymptotique présenté plus haut est vérifié par notre estimation par Maximum de Vraisemblance.

Variance d'estimation

Dans ce paragraphe, nous nous intéressons au comportement de la variance d'estimation du Maximum de Vraisemblance lorsque le nombre de réalisations prises en compte augmente asymptotiquement. Pour tout estimateur C , cette variance, notée Var_C , est définie par :

$$\text{Var}_C = \mathbb{E}_{\mathbf{X}^\Omega} \left(\left(C(\mathbf{X}^\Omega) - \bar{C}(\Omega) \right) \left(C(\mathbf{X}^\Omega) - \bar{C}(\Omega) \right)^T \middle| \bar{\theta} \right)$$

où $\bar{C}(\Omega) = \mathbb{E}_{\mathbf{X}^\Omega} \left(C(\mathbf{X}^\Omega) \middle| \bar{\theta} \right)$. Cette grandeur représente les fluctuations moyennes observées sur la valeur de l'estimation, lorsqu'un grand nombre d'estimations sont effectuées sur des groupes successifs de Ω réalisations.

Dans le cas du Maximum de Vraisemblance appliquée à la tomographie d'état (puis, plus tard, de processus), cette variance est notée σ_{ML}^2 . Puisque le nombre de réalisations prises en compte est, dans nos validations expérimentales, très grand, et en supposant que $\bar{C}(\Omega) = \mathbb{E}_{Y^{(1)}, \dots, Y^{(\Omega)}} \left(\hat{\rho}_{ML}(Y^{(1)}, \dots, Y^{(\Omega)}) \middle| \bar{\rho} \right) = \bar{\rho}$, on peut assimiler σ_{ML}^2 à une variance permettant d'obtenir l'erreur moyenne séparant une estimation particulière $\hat{\rho}_{ML}$ de $\bar{\rho}$ (celle qui est faite à partir de toutes les trajectoires expérimentales à notre disposition) de la valeur réelle $\bar{\rho}$.

Enfin, nous donnerons par la suite, non pas une expression directe de σ_{ML}^2 , qui est un tenseur d'ordre 4 dans $\mathcal{H}$, mais l'expression de la variance de l'estimation de $\text{Tr}(\bar{\rho}A)$, la valeur initiale moyenne d'un observable A quelconque. Cette variance est notée $\sigma_{ML}^2(A)$.

Information de Fisher et borne de Cramér-Rao

Nous présentons ici des résultats mathématiques permettant de comprendre d'où proviennent les formules de la variance de l'estimation par Maximum de Vraisemblance.

Reprenons le cadre général de l'estimation de $\bar{\theta}$ grâce à la mesure X . L'information que cette réalisation extraite de $\bar{\theta}$ contient sur $\bar{\theta}$ peut être quantifiée par l'information de Fisher, laquelle est exprimée comme une fonction :

$$\theta \in \Theta \mapsto \mathcal{I}(\theta) = \mathbb{E}_X \left(\left(\frac{\partial}{\partial \theta} \log(p(X|\theta)) \right)^2 \middle| \theta \right) \quad (3.28)$$

où l'espérance est calculée comme précédemment sur la valeur de X généré par un paramètre θ quelconque.

Cette formule peut être modifiée. En effet, en notant $\mathcal{D}_X$ l'ensemble des valeurs possibles de X , on peut écrire :

$$\mathcal{I}(\theta) = \int_{\mathcal{D}_X} p(X|\theta) \left(\frac{\partial}{\partial \theta} \log(p(X|\theta)) \right)^2 dX$$

Or, d'un autre côté, on peut montrer que :

$$\begin{aligned}
& - \int_{\mathcal{D}_X} p(X | \theta) \frac{\partial^2}{\partial \theta^2} \log(p(X | \theta)) dX \\
&= \int_{\mathcal{D}_X} p(X | \theta) \left(\left(\frac{\frac{\partial}{\partial \theta} p(X | \theta)}{p(X | \theta)} \right)^2 - \frac{\frac{\partial^2}{\partial \theta^2} p(X | \theta)}{p(X | \theta)} \right) dX \\
&= \mathcal{I}(\theta) - \int_{\mathcal{D}_X} \frac{\partial^2}{\partial \theta^2} p(X | \theta) dX \\
&= \mathcal{I}(\theta) - \frac{\partial^2}{\partial \theta^2} \int_{\mathcal{D}_X} p(X | \theta) dX = \mathcal{I}(\theta) - \frac{\partial^2}{\partial \theta^2} (1) \\
&= \mathcal{I}(\theta)
\end{aligned} \tag{3.29}$$

Ainsi, dans notre cas, nous pouvons écrire que l'information de Fisher est simplement donnée par :

$$\mathcal{I}(\theta) = -\mathbb{E} \left(\frac{\partial^2}{\partial \theta^2} \log(p(X | \theta)) \right) \tag{3.30}$$

Or, lorsqu'on cherche à donner une estimation $\hat{\theta}$ de $\bar{\theta}$ à partir des réalisations X en utilisant un estimateur non biaisé C , cette information de Fisher intervient pour déterminer un minorant de la variance de cette estimation :

$$\text{Var}_C \geq \mathcal{I}^{-1}(\hat{\theta}) \tag{3.31}$$

L'inverse de l'information, appelée borne de Cramér-Rao, est donc un minorant de la variance d'estimation des estimateurs non biaisés. Dans le cas du Maximum de Vraisemblance, il existe un résultat montrant que, sous certaines conditions, la variance d'estimation du Maximum de Vraisemblance atteint la borne de Cramér-Rao lorsque Ω tend vers l'infini. Dans ce cas, le Maximum de Vraisemblance est dit asymptotiquement efficace. Ce résultat est le suivant :

Lemme 2. Sous certaines conditions (voir [LC98][1]), et en reprenant les notations du lemme 1, on a, pour toute valeur $\bar{\theta} \in \Theta$ et toute suite $\hat{\theta}_\Omega(\mathbf{X}^\Omega)$ tendant en probabilité vers $\bar{\theta}$:

$$\sqrt{\Omega}(\hat{\theta}_\Omega(\mathbf{X}^\Omega) - \bar{\theta}) \rightarrow N(0, (I(\bar{\theta}))^{-1})$$

où la convergence est en probabilité et $N(\mu, \sigma^2)$ une loi normale de moyenne μ et de variance σ^2 . Nous insistons sur le fait que l'information de Fisher n'est calculée qu'à partir d'une seule réalisation X .

Désormais, nous déciderons, lorsque ces lemmes seront applicables, donc lorsque $\hat{\rho}_{ML}$ sera un point intérieur de $\mathcal{D}_\mathcal{H}$, d'assimiler la variance d'estimation du Maximum de Vraisemblance à l'inverse de l'opposé du Hessien de la log-vraisemblance en $\hat{\rho}_{ML}$, remplaçant l'intégrale de l'équation (3.30) par une simple identification en ce point.

Formules de la variance

Les formules permettant d'obtenir la variance de $\text{Tr}(A\hat{\rho}_{ML})$ dépendent de la position de $\hat{\rho}_{ML}$ dans l'espace $\mathcal{D}_\mathcal{H}$, puisqu'a priori, l'expression de la variance obtenue par la borne de Cramér-Rao n'est valable que lorsque le résultat $\hat{\rho}_{ML}$ de l'estimation par Maximum de Vraisemblance est un point intérieur de l'ensemble des valeurs que peut prendre ce

paramètre, ici $\mathcal{D}_{\mathcal{H}}$. Nous devons donc donner une première formule valable lorsque $\hat{\rho}_{ML}$ est de rang plein, puis passerons par l'estimation bayésienne pour obtenir cette variance lorsque le noyau de $\hat{\rho}_{ML}$ est de dimension 1.

Lorsque $\hat{\rho}_{ML}$ est de rang plein, nous devons tout de même tenir compte de la topologie de $\mathcal{D}_{\mathcal{H}}$ autour de ce point. Or, une matrice densité proche de $\hat{\rho}_{ML}$ ne peut évoluer qu'à trace constante. Ainsi, pour tout observable Hermitien A sur $\mathcal{H}$, aucune incertitude n'est due à la composante identité de cet observable. En d'autres termes, si l'on définit $A_{\parallel}$ la projection orthogonale de A sur $\mathcal{H}_{\text{herm}}^0$:

$$A_{\parallel} = A - \frac{\text{Tr}(A)}{\text{Tr}(I_{\mathcal{H}})} I_{\mathcal{H}} \quad (3.32)$$

(où, nous le rappelons, $I_{\mathcal{H}}$ est l'opérateur identité sur $\mathcal{H}$), la variance de $\langle A \rangle_{ML}$ et celle de $\langle A_{\parallel} \rangle_{ML}$ sont égales. Nous rappelons que nous nommons cette variance $\sigma_{ML}^2(A)$. L'expression de cette variance fait intervenir l'inverse de la matrice Hessienne de f en $\hat{\rho}_{ML}$ projetée orthogonalement sur les dimensions de $\mathcal{H}_{\text{herm}}^0$. Cette variance est donc équivalente, lorsque le nombre de mesures prises en compte tend vers l'infini, à :

$$\sigma_{ML}^2(A) \equiv \text{Tr} \left(A_{\parallel} \mathbf{R}^{-1} (A_{\parallel}) \right) \quad (3.33)$$

où $\mathbf{R}^{-1}(A_{\parallel})$ provient de l'inverse de l'application :

$$X \mapsto \mathbf{R}(X) = \sum_{n=1}^{\Omega} \frac{\text{Tr} \left(A E_{\parallel}^{(n)} \right) E_{\parallel}^{(n)}}{\text{Tr}^2(\hat{\rho}_{ML} E^{(n)})} \quad (3.34)$$

Pour simplifier, nous avons écrit $E^{(n)} = E_0^{(n)}$, et $E_{\parallel}^{(n)}$ est la projection orthogonale de $E^{(n)}$ sur $\mathcal{H}_{\text{herm}}^0$, construite comme $A_{\parallel}$.

En revanche, lorsque $\hat{\rho}_{ML}$ appartient à la frontière de $\mathcal{D}_{\mathcal{H}}$, la courbure de cette frontière est à prendre en compte dans l'expression du Hessian. L'analyse de l'estimation Bayésienne de $\bar{\rho}$ nous permettra de donner une formule de la variance de l'estimation lorsque $\hat{\rho}_{ML}$ est sur une surface régulière de $\mathcal{D}_{\mathcal{H}}$, c'est-à-dire lorsque son noyau est de dimension 1, puis d'étendre la formule obtenue au cas où ce noyau est à dimensions multiples.

3.4 Passage à l'estimation bayésienne

3.4.1 Principe de l'estimation bayésienne

Nous venons de présenter la façon dont le principe du Maximum de Vraisemblance résout le problème de la tomographie d'état quantique. Cependant, le choix sous-jacent de sélectionner comme estimation de l'état initial la valeur maximisant la probabilité inverse $\mathbb{P}(Y | \bar{\rho} = \rho)$ des mesures peut paraître arbitraire.

Nous avons présenté en sous-section 2.2.2 l'interprétation qui peut être faite de la matrice densité comme espérance sur les mesures effectuées *antérieurement*. Ici, la situation est légèrement différente ; les mesures ont été réalisées après l'instant où l'état est estimé. Ce sont les mesures qui sont influencées par l'état, et non plus l'inverse. Néanmoins, on peut toujours exprimer la probabilité de l'état conditionnellement à toute l'information que nous disposons sur le système : c'est ce que nous avons fait dans l'équation (3.10). L'interprétation de la matrice densité comme espérance, que nous avons évoquée plus

haut, nous amène à définir une autre façon d'estimer l'état initial. Il s'agit de l'estimation Bayésienne, dont le résultat $\hat{\rho}_{BM}$ est donné par la formule :

$$\hat{\rho}_{BM} = \frac{\int_{\mathcal{D}_{\mathcal{H}}} \rho p(\bar{\rho} = \rho | Y) d\rho}{\int_{\mathcal{D}_{\mathcal{H}}} p(\bar{\rho} = \rho | Y) d\rho} \quad (3.35)$$

$$= \frac{\int_{\mathcal{D}_{\mathcal{H}}} \rho \exp(f(\rho)) d\rho}{\int_{\mathcal{D}_{\mathcal{H}}} \exp(f(\rho)) d\rho} \quad (3.36)$$

où $p(\bar{\rho} = \rho | Y)$ est la densité de probabilité définie plus haut par l'équation (3.11), et où l'intégration se fait avec la métrique dérivée de la norme de Frobenius.

De la même manière, cette estimation Bayésienne peut également être faite pour le tenseur d'erreur d'estimation σ^2 . Nous notons de même ce nouveau tenseur σ_{BM}^2 . L'erreur est définie, bien sûr, depuis l'estimation bayésienne $\hat{\rho}_{BM}$. Il prend, pour l'observable A , la valeur :

$$\sigma_{BM}^2(A) = \frac{\int_{\mathcal{D}_{\mathcal{H}}} p(\bar{\rho} = \rho | Y) \text{Tr}^2((\rho - \hat{\rho}_{BM}) A) d\rho}{\int_{\mathcal{D}_{\mathcal{H}}} p(\bar{\rho} = \rho | Y) d\rho} \quad (3.37)$$

$$= \frac{\int_{\mathcal{D}_{\mathcal{H}}} \exp(f(\rho)) \text{Tr}^2((\rho - \hat{\rho}_{BM}) A) d\rho}{\int_{\mathcal{D}_{\mathcal{H}}} \exp(f(\rho)) d\rho} \quad (3.38)$$

Remarquons que la variance doit être calculée ici à partir de $\hat{\rho}_{BM}$, et non pas à partir de $\hat{\rho}_{ML}$. Cela dit, nous allons montrer que cela n'a pas d'influence, puisque ces deux valeurs sont liées par une équivalence asymptotique lorsque le nombre de trajectoires prises en compte tend vers l'infini.

3.4.2 Comparaison avec le Maximum de Vraisemblance

Nous souhaitons étudier les relations qui lient ces estimations bayésiennes de celles qui sont tirées du principe du Maximum de Vraisemblance, et ce lorsque le nombre de mesures prises en compte tend vers l'infini. Ainsi, nous considérons que la longueur de chaque trajectoire (le nombre de mesures T_n) reste dans certaines bornes, et que le nombre de trajectoires Ω tend vers l'infini. La fonction f et ses variations étant proportionnelles à Ω , il peut être judicieux de définir une fonction moyenne de f , notée $\bar{f}$, qui vérifie :

$$\forall \rho \in \mathcal{D}_{\mathcal{H}}, \quad \bar{f}(\rho) = \frac{1}{\Omega} f(\rho) \quad (3.39)$$

Les formules permettant d'obtenir les estimations bayésiennes deviennent finalement :

$$\hat{\rho}_{BM} = \frac{\int_{\mathcal{D}_{\mathcal{H}}} \rho \exp(\Omega \bar{f}(\rho)) d\rho}{\int_{\mathcal{D}_{\mathcal{H}}} \exp(\Omega \bar{f}(\rho)) d\rho} \quad (3.40)$$

$$\sigma_{BM}^2(A) = \frac{\int_{\mathcal{D}_{\mathcal{H}}} \exp(\Omega \bar{f}(\rho)) \text{Tr}^2((\rho - \hat{\rho}_{BM}) A) d\rho}{\int_{\mathcal{D}_{\mathcal{H}}} \exp(\Omega \bar{f}(\rho)) d\rho} \quad (3.41)$$

Les intégrales qui apparaissent ici respectent le formalisme des intégrales de Laplace et de leur développement asymptotique. Le détail des calculs qui suivent est donné en annexe C. Le premier résultat que l'on peut donner est que :

$$\hat{\rho}_{BM} = \hat{\rho}_{ML} + \mathcal{O}\left(\frac{1}{\Omega}\right) \quad (3.42)$$

Un résultat similaire peut être obtenu à propos des estimations d'erreur. En effet, pour tout observable A :

$$\sigma_{BM}^2(A) = \sigma_{ML}^2(A) + \mathcal{O}\left(\frac{1}{\Omega}\right) \quad (3.43)$$

Il est important de noter que ces deux résultats sont valables, que $\hat{\rho}_{ML}$ soit de rang plein ou de rang faible. Or, les développements asymptotiques réalisés en annexe C permettent d'obtenir la variance $\sigma_{BM}^2(A)$ même lorsque $\hat{\rho}_{ML}$ n'est pas inversible. Nous pouvons de plus énoncer le théorème suivant.

Théorème 3. Soit $\mathcal{D}_{\mathcal{H}}$ l'ensemble des matrices densité de dimension $N_{\mathcal{H}}$ et la fonction de log-vraisemblance (3.21). Supposons que l'espace vectoriel engendré par les $E^{(n)}$ parcourt l'ensemble des matrices Hermitiennes de dimension $N_{\mathcal{H}}$. Notons ρ^* l'unique maximum de f sur $\mathcal{D}_{\mathcal{H}}$, et supposons que, en plus des conditions d'optimalité du théorème 1, on a $\ker(\lambda^* I_{\mathcal{H}} - \nabla f(\rho^*)) = \ker(I - P^*)$. Alors, pour tout opérateur Hermitien A , sa moyenne Bayésienne, définie par $\text{Tr}(A\hat{\rho}_{BM})$ (où $\hat{\rho}_{BM}$ est défini par l'équation (3.40)), admet le développement asymptotique :

$$\text{Tr}(A\hat{\rho}_{BM}) = \text{Tr}(A\rho^*) + \mathcal{O}\left(\frac{1}{\Omega}\right), \quad (3.44)$$

et sa variance Bayésienne (définie par l'équation (3.41)) :

$$\sigma_{BM}^2(A) = \text{Tr}\left(A_{\parallel} \mathbf{R}^{-1}(A_{\parallel})\right) + \mathcal{O}\left(\frac{1}{\Omega}\right), \quad (3.45)$$

où :

- pour tout opérateur B , $B_{\parallel}$ représente la projection orthogonale de B sur l'espace tangent en ρ^* à la sous-variété des matrices Hermitiennes de rang égal à celui de ρ^* et de trace un. On l'obtient par la formule :

$$B_{\parallel} = B - \frac{\text{Tr}(BP^*)}{\text{Tr}(P^*)} P^* - (I_{\mathcal{H}} - P^*) B (I_{\mathcal{H}} - P^*);$$

lorsque ρ^* est de rang plein, $B_{\parallel} = B - \text{Tr}(B) I_{\mathcal{H}}/N_{\mathcal{H}}$, car $P^* = I_{\mathcal{H}}$;

- le super-opérateur linéaire $\mathbf{R}$ correspond au Hessian de la restriction de f au seul ensemble des matrices Hermitiennes de même rang que ρ^* et de trace nulle. Ainsi, pour tout opérateur Hermitien X :

$$\begin{aligned} \mathbf{R}(X) &= \sum_{n=1}^{\Omega} \frac{\text{Tr}(X E_{\parallel}^{(n)}) E_{\parallel}^{(n)}}{\text{Tr}^2(\rho^* E^{(n)})} \\ &\quad + \frac{1}{2} [(\lambda^* I_{\mathcal{H}} - \nabla f(\rho^*)) X (\rho^*)^+ + (\rho^*)^+ X (\lambda^* I_{\mathcal{H}} - \nabla f(\rho^*))], \end{aligned} \quad (3.46)$$

avec $(\rho^*)^+$ le pseudo-inverse de Penrose de ρ^* ; la restriction de $X \mapsto \text{Tr}(X\mathbf{R}(X))$ aux opérateurs X appartenant à l'espace tangent à $\mathcal{D}_{\mathcal{H}}$ en ρ^* est définie positive. Ainsi, la restriction de $\mathbf{R}$ à cet espace tangent est inversible, et peut-être vue comme l'analogue de l'information de Fisher : son inverse en $A_{\parallel}$ est noté ici $\mathbf{R}^{-1}(A_{\parallel})$.

En utilisant les résultats de ce théorème et les équations (3.42) et (3.43), on peut facilement retrouver la valeur de l'incertitude de l'estimation de la valeur moyenne de chaque observable A par Maximum de Vraisemblance.

3.5 Validations expérimentales

Dans cette dernière section, nous utilisons le processus de tomographie d'état que nous venons de présenter sur deux systèmes physiques. Ces deux systèmes, présentés en annexe A, sont, d'abord, une expérience de boîte à photons où le système est considéré dans une base de Fock et lu de manière discrète, puis un qubit supraconducteur, dont le signal de fluorescence, diffusif, est acquis par l'expérimentateur.

3.5.1 Cas d'un champ de photons : mesures discrètes

Configuration

Nous utilisons ici les résultats provenant de l'expérience "boîte à photons" pour montrer la pertinence de la méthode de tomographie d'état quantique présentée dans ce chapitre. Les données physiques sont les mêmes qu'exposées en annexe, hormis le nombre de photons thermiques fixé ici à $n_{\text{th}} = 0.06$.

Nous disposons de $\Omega = 1390$ trajectoires de mesures, acquises entre les instants $t = -172$ ms et $t = 344$ ms. L'intervalle de temps entre deux mesures étant de $T_c = 86$ μ s, cela correspond à 6000 mesures par trajectoire. Entre $t = -172$ ms et $t = 0^-$, le champ de la cavité est initialement dans un état thermique, avec un nombre moyen de photons égal à n_{th} . A l'instant $t = 0^-$, un unique photon est injecté dans le champ. A cause des imperfections expérimentales, il y a en réalité 1.26 photons injectés en moyenne. Ensuite, le champ est encore mesuré pendant 344 ms, une période de temps assez longue par rapport au temps de demi-vie $T_c = 65$ ms des photons dans la cavité. Le nombre moyen de photons dans la cavité à l'instant t est donc, théoriquement, $\langle \mathbf{N} \rangle_t$ tel que :

$$\langle \mathbf{N} \rangle_t = \begin{cases} n_{\text{th}} & \text{for } t < 0 \\ n_{\text{th}} + (\langle \mathbf{N} \rangle_0 - n_{\text{th}}) e^{-t/T_c}, & \text{for } t \geq 0 \end{cases} \quad (3.47)$$

où $\langle \mathbf{N} \rangle_0 = 1.26$.

Processus d'estimation

Ici, les états quantiques ρ_k et les états quantiques adjoints E_k étant diagonaux dans la base de Fock, ils sont représentés par des vecteurs, comme détaillé en annexe. Ces états sont tronqués, avec un maximum de 7 photons : ρ_k et E_k sont donc des vecteurs en 8 dimensions.

Pour chaque t_k entre -100 ms et 150 ms, nous estimons l'état $\rho_k = \rho(t_k)$ grâce au Maximum de Vraisemblance, utilisant la méthode exposée dans ce chapitre. Cette estimation est notée $\hat{\rho}_{ML}(t_k)$. Cette estimation est faite grâce aux mesures effectuées entre les instants t_k et 344 ms. Dans ce processus d'estimation, on ne prend pas en compte, dans le cas où t_k serait antérieur à l'injection du photon, cette injection dans les équations permettant de générer les états adjoints $E_k^{(n)}$. A tout instant, le rang de $\hat{\rho}_{ML}(t)$ est strictement inférieur à 8 (entre 2 et 5), et nous avons vérifié que ces états satisfaisaient la caractérisation de l'équation (3.26). Nous donnons également l'intervalle de confiance à 95 % de cette estimation. Nous comparons ces résultats avec l'estimation théorique du nombre de photons de l'équation (3.47).

Résultats

Les résultats de cette simulation sont exposés en figure 3.1. Commentons rapidement les résultats obtenus par Maximum de Vraisemblance. A l'instant $t = 0$, le nombre de photons calculé ainsi atteint son maximum, avec 1.237 ± 0.045 , ce qui correspond au moment où le photon est injecté dans la cavité. Nous insistons sur le fait que le Maximum de Vraisemblance retrouve cet état de fait, alors même que l'information de l'injection n'est jamais prise en compte dans ce modèle d'estimation.

Lorsque $t > 0$, on s'attend à ce que le nombre de photons calculé par Maximum de Vraisemblance décroisse, ce qui est effectivement le cas. Pour $t < 0$, on s'aperçoit que l'estimation recolle assez vite à la réalité, alors même que l'estimation par Maximum de Vraisemblance ne prend pas en compte l'information de l'injection. En effet, pour $t < -50$ ms, le nombre de photons estimé se stabilise à 0.042 ± 0.015 , ce qui est assez proche du nombre de photons thermiques $n_{\text{th}} = 0.06$ prédit par le modèle. Cette période de 50 ms nécessaire au processus pour s'adapter au changement de modèle peut être réduite en prenant en compte un nombre de trajectoires Ω plus grand.

Nous avons donc montré que notre processus de tomographie est robuste aux changements ponctuels de modèle. De plus, il est assez logique d'obtenir un pic d'incertitude de l'estimation juste avant l'injection de photons, puisqu'à ces instants, le processus dispose de très peu de mesures enregistrées lorsque le photon n'est pas encore injecté.

3.5.2 Cas d'un qubit : mesures diffusives

Rappel des équations

Nous cherchons ici à exploiter des données de mesures provenant d'un qubit supraconducteur, tel que décrit dans l'annexe A.2. Le signal de fluorescence du qubit est mesuré, et sa décohérence n'est pas lue, de sorte que le système évolue bien selon les équations (A.4) et (A.5).

Nous réutilisons également, pour intégrer ces équations différentielles, le schéma de discrétisation simple, présenté en introduction de cette thèse, dont nous rappelons ici la forme :

$$\rho_{t+\Delta t} = \frac{\mathbf{K}_{t,\Delta \mathbf{y}_t,\Delta t}(\rho_t)}{\text{Tr}(\mathbf{K}_{t,\Delta \mathbf{y}_t,\Delta t}(\rho_t))}$$

avec $\Delta \mathbf{y}_t = (\Delta I_t, \Delta Q_t)$, $\Delta I_t = \int_t^{t+\Delta t} dI_{t'}$ et $\Delta Q_t = \int_t^{t+\Delta t} dQ_{t'}$.

Les applications $\mathbf{K}_{t,\Delta \mathbf{y}_t,\Delta t}$ sont la version discrétisée de l'ensemble complet d'applications de Kraus $\mathbf{K}_{t,d\mathbf{y}_t,dt}$ vérifiant :

$$\mathbf{K}_{t,d\mathbf{y}_t,dt}(\rho) = M_{t,d\mathbf{y}_t,dt} \rho M_{t,d\mathbf{y}_t,dt}^\dagger + [(1-\eta) \gamma_1 \sigma_- \rho \sigma_+ + \gamma_\phi \sigma_z \rho \sigma_z] dt,$$

avec :

$$M_{t,d\mathbf{y}_t,dt} = I_{\mathcal{H}} + \left(-\frac{i}{\hbar} \hat{H}_t - \frac{1}{2} [\gamma_1 \sigma_+ \sigma_- + \gamma_\phi I_{\mathcal{H}}] \right) dt + \sqrt{\eta \gamma_1 / 2} \sigma_- (dI_t + idQ_t).$$

L'opérateur σ_+ est l'opérateur de création : $\sigma_+ = (\sigma_-)^\dagger$. Nous utilisons le schéma de discrétisation simple, qui donne pour loi de probabilité approchée :

$$\mathbb{P}(\Delta \mathbf{y}_t \in [\mathbf{s}\sqrt{\Delta t}; (\mathbf{s} + d\mathbf{s})\sqrt{\Delta t}] | \rho_t) = \text{Tr}(\mathbf{K}_{t,\Delta \mathbf{y}_t,\Delta t}(\rho_t)) \exp\left(\frac{\|\mathbf{s}\|^2}{2}\right) \frac{d\mathbf{s}_I d\mathbf{s}_Q}{2\pi},$$

nous pouvons utiliser le travail théorique réalisé dans ce chapitre pour la tomographie d'état par Maximum de Vraisemblance, à partir des Ω trajectoires de mesures discrétisées $(\Delta I_0^{(n)}, \Delta I_{\Delta t}^{(n)}, \dots, I_{(T_n-1)\Delta t}^{(n)})$ et $(\Delta Q_0^{(n)}, \Delta Q_{\Delta t}^{(n)}, \dots, Q_{(T_n-1)\Delta t}^{(n)})$, pour $n = 1, \dots, \Omega$.

De même, cette discrétisation peut être utilisée pour calculer la suite d'états adjoints E_t du système. Nous avons la formule suivante, similaire à celle de l'équation (3.18), permettant de calculer par récurrence inverse l'état adjoint de la trajectoire de mesure n :

$$E_{T_n \Delta t}^{(n)} = \frac{I_{\mathcal{H}}}{\text{Tr}(I_{\mathcal{H}})}, \quad E_{k \Delta t}^{(n)} = \frac{\mathbf{K}_{t, \Delta y_{k \Delta t}, \Delta t}^* (E_{(k+1) \Delta t}^{(n)})}{\text{Tr}(\mathbf{K}_{t, \Delta y_{k \Delta t}, \Delta t}^* (E_{(k+1) \Delta t}^{(n)}))}$$

où l'application de Kraus adjointe est définie par :

$$\mathbf{K}_{t, \Delta y_{k \Delta t}, \Delta t}^* (\rho) = M_{t, \Delta y_{k \Delta t}, \Delta t}^\dagger \rho M_{t, \Delta y_{k \Delta t}, \Delta t} + [(1 - \eta) \gamma_1 \sigma_+ \rho \sigma_- + \gamma_\phi \sigma_z \rho \sigma_z] dt,$$

Nous pouvons appliquer la méthode de tomographie d'état que nous avons établie dans les précédentes sections de ce chapitre.

Tomographie d'état quantique pour un qubit

Pour un système à deux niveaux, l'estimation par Maximum de Vraisemblance développée dans les sections précédentes peut être décrite de manière plus simple, en utilisant pour décrire les états et états adjoints leurs coordonnées dans la sphère de Bloch. Ainsi :

$$\rho = \frac{I_{\mathcal{H}} + x\sigma_x + y\sigma_y + z\sigma_z}{2} \quad (3.48)$$

$$E = \frac{I_{\mathcal{H}} + e_x\sigma_x + e_y\sigma_y + e_z\sigma_z}{2}. \quad (3.49)$$

Les coordonnées (x, y, z) et (e_x, e_y, e_z) représentent alors une nouvelle façon d'exprimer la valeur de ρ et de E . De plus, l'ensemble $\mathcal{D}_{\mathcal{H}}$ des matrices densité admissibles correspond à la sphère unitaire de $\mathbb{R}^3$ vérifiant $x^2 + y^2 + z^2 \leq 1$, et le produit de Frobenius entre deux opérateurs de $\mathcal{H}_{\text{herm}}$ de trace une correspond au produit scalaire euclidien classique sur $\mathbb{R}^3$. Le gradient en $\hat{\rho}_{ML}$ de la fonction de log-vraisemblance (3.23) devient alors :

$$\nabla f(\rho_{ML}) = \nabla f_{ML} = \sum_{n=1}^{\Omega} \frac{1}{1 + x_{ML} e_x^{(n)} + y_{ML} e_y^{(n)} + z_{ML} e_z^{(n)}} \begin{pmatrix} e_x^{(n)} \\ e_y^{(n)} \\ e_z^{(n)} \end{pmatrix}.$$

La caractérisation de $\hat{\rho}_{ML}$ donnée en (3.26) devient alors la suivante : si cette matrice densité est à l'intérieur de $\mathcal{D}_{\mathcal{H}}$, c'est-à-dire si $x^2 + y^2 + z^2 < 1$, alors $\nabla f_{ML} = 0$; si

$x^2 + y^2 + z^2 = 1$, alors $\nabla f_{ML} = \lambda_{ML} \begin{pmatrix} x_{ML} \\ y_{ML} \\ z_{ML} \end{pmatrix}$, avec $\lambda_{ML} \leq 0$. Le super-opérateur $\mathbf{R}$ défini par

l'équation (3.46) devient alors une matrice positive semi-définie de dimension 3. Sa valeur est la suivante, selon la position de ρ_{ML} dans $\mathcal{D}_{\mathcal{H}}$:

— si $x^2 + y^2 + z^2 < 1$, nous obtenons :

$$\mathbf{R} = \sum_{n=1}^{\Omega} \frac{\begin{pmatrix} e_x^{(n)} e_x^{(n)} & e_x^{(n)} e_y^{(n)} & e_x^{(n)} e_z^{(n)} \\ e_y^{(n)} e_x^{(n)} & e_y^{(n)} e_y^{(n)} & e_y^{(n)} e_z^{(n)} \\ e_z^{(n)} e_x^{(n)} & e_z^{(n)} e_y^{(n)} & e_z^{(n)} e_z^{(n)} \end{pmatrix}}{(1 + x_{ML} e_x^{(n)} + y_{ML} e_y^{(n)} + z_{ML} e_z^{(n)})^2}$$

La plupart du temps (dès que le nombre de trajectoires prises en compte est suffisant), cet opérateur est de rang plein, et il peut être inversé. Ainsi, pour chaque opérateur A de la forme $A = \frac{a\sigma_x + b\sigma_y + c\sigma_z}{2}$, associé au vecteur (a, b, c) , la formule (3.45) devient :

$$\sigma_{ML}^2(A) = \mathbf{a}^\top \mathbf{R}^{-1} \mathbf{a}, \quad \text{où } \mathbf{a} = \begin{pmatrix} a \\ b \\ c \end{pmatrix}$$

— si $x^2 + y^2 + z^2 = 1$, et lorsque $\lambda_{ML} > 0$ est assez grand, nous obtenons :

$$\mathbf{R} = \sum_{n=1}^{\Omega} \frac{\begin{pmatrix} e_{\parallel x}^{(n)} e_{\parallel x}^{(n)} & e_{\parallel x}^{(n)} e_{\parallel y}^{(n)} & e_{\parallel x}^{(n)} e_{\parallel z}^{(n)} \\ e_{\parallel y}^{(n)} e_{\parallel x}^{(n)} & e_{\parallel y}^{(n)} e_{\parallel y}^{(n)} & e_{\parallel y}^{(n)} e_{\parallel z}^{(n)} \\ e_{\parallel z}^{(n)} e_{\parallel x}^{(n)} & e_{\parallel z}^{(n)} e_{\parallel y}^{(n)} & e_{\parallel z}^{(n)} e_{\parallel z}^{(n)} \end{pmatrix}}{\left(1 + x_{ML} e_x^{(n)} + y_{ML} e_y^{(n)} + z_{ML} e_z^{(n)}\right)^2} + \lambda_{ML} \begin{pmatrix} 1 - x_{ML} x_{ML} & -x_{ML} y_{ML} & -x_{ML} z_{ML} \\ -y_{ML} x_{ML} & 1 - y_{ML} y_{ML} & -y_{ML} z_{ML} \\ -z_{ML} x_{ML} & -z_{ML} y_{ML} & 1 - z_{ML} z_{ML} \end{pmatrix}$$

où $(e_{\parallel x}^{(n)}, e_{\parallel y}^{(n)}, e_{\parallel z}^{(n)})$ sont les coordonnées de la projection de $E^{(n)}$ sur le sous-espace tangent à $\mathcal{D}_{\mathcal{H}}$ en $\hat{\rho}_{ML}$ sur $\mathcal{H}_{\text{herm}}$. Ainsi, on définit $s^{(n)} = e_x^{(n)} x_{ML} + e_y^{(n)} y_{ML} + e_z^{(n)} z_{ML}$ tel que $e_{\parallel \xi}^{(n)} = e_{\xi}^{(n)} - s^{(n)} \xi_{ML}$, pour tout $\xi = x, y, z$. Le rang de l'opérateur $\mathbf{R}$ est inférieur ou égal à 2. Son inverse, apparaissant dans la formule (3.45), correspond alors au pseudo-inverse de Moore-Penrose. Il est évalué sur le vecteur associé à la projection $A_{\parallel}$ de l'opérateur A :

$$A_{\parallel} = \frac{(a - s x_{ML}) \sigma_x + (b - s y_{ML}) \sigma_y + (c - s z_{ML}) \sigma_z}{2},$$

avec $s = a x_{ML} + b y_{ML} + c z_{ML}$.

Estimation de l'état initial d'un qubit avec mesures hétérodynes de fluorescence

Enfin, nous testons la tomographie d'état quantique par Maximum de Vraisemblance sur le système ainsi décrit. Ici, nous supposons que le Hamiltonien $\hat{H}_t = 0$, que l'efficacité de détection du signal de fluorescence vaut $\eta = 0.24$, et que les constantes de temps de la mesure et du déphasage valent $\gamma_1 = (4.15 \mu\text{s})^{-1}$ et $\gamma_{\phi} = (70 \mu\text{s})^{-1}$.

Nous utilisons ici un nombre relativement faible de trajectoires, $\Omega = 4.10^4$, qui ont toutes la même durée $T_n = T = 9.4 \mu\text{s}$. Ici, la période d'échantillonnage $\Delta t = 200 \text{ ns} \approx \gamma_1^{-1}/20$ est bien plus petite que les temps caractéristiques des matrices de mesure. L'utilisation du schéma de discrétisation simple est donc pertinente. Chaque trajectoire est supposée commencer au temps $t = 0$ à partir du même état initial $\bar{\rho}$ proche de l'état pur $(|g\rangle + |e\rangle)/\sqrt{2}$, correspondant à $x = 1$.

Dans un premier temps, l'objectif est de vérifier la valeur de $\bar{\rho}$, en utilisant la tomographie par Maximum de Vraisemblance. Nous utilisons, pour chaque trajectoire de mesure n , les 2×47 mesures discrétisées $\Delta I_0, \dots, \Delta I_{46\Delta t}$ et $\Delta Q_0, \dots, \Delta Q_{46\Delta t}$. A partir de ces mesures, nous calculons les états adjoints $E_0^{(n)}$ de l'état correspondant à chaque

trajectoires, puis en déduisant, après maximisation de la log-vraisemblance, une estimation par Maximum de Vraisemblance de chaque coordonnée de $\bar{\rho}$, accompagnée d'un intervalle de vraisemblance à 95% pour chacune, grâce à la matrice $\mathbf{R}$ obtenue :

$$\begin{aligned}\text{Tr}(\bar{\rho}_0 \sigma_x) &= 0.99 \pm 0.06 \\ \text{Tr}(\bar{\rho}_0 \sigma_y) &= -0.03 \pm 0.07 \\ \text{Tr}(\bar{\rho}_0 \sigma_z) &= -0.10 \pm 0.19\end{aligned}$$

Cette estimation de l'état initial est cohérente avec nos attentes. En effet, l'état est préparé en $x = 1$, à partir d'un état thermique avec moins de 1% d'excitation.

Dans un second temps, nous cherchons à montrer que la tomographie par Maximum de Vraisemblance permet de reconstruire les états moyens successifs du système. Pour cela, nous réalisons, comme précédemment, la tomographie de l'état $\bar{\rho}_t$, moyenne des états $\rho_t^{(n)}$ du système à l'instant t , en utilisant, parmi les Ω trajectoires de mesures, les mesures acquises entre les instants t et T . Nous ne réalisons ce travail que pour chaque instant t entre 0 et 5 μs , faute d'information suffisante pour prendre en compte des moments ultérieurs. Nous notons les coordonnées de ces estimations, respectivement $x_{ML}(t)$, $y_{ML}(t)$ et $z_{ML}(t)$. Sur les trois figures 3.2, 3.3 et 3.4, nous les faisons apparaître en ligne rouge, accompagné de leur intervalle de confiance à 95%, figurant de chaque côté de la courbe rouge en tirets bleus.

Enfin, pour prouver la pertinence de ces estimations, nous les comparons, dans le cas de $x_{ML}(t)$ et $y_{ML}(t)$, avec ce que permet d'estimer la moyennisation des signaux de mesures. En effet, puisque :

$$\begin{aligned}\langle dI_t \rangle &= \sqrt{\eta\gamma_1/2} x(t) dt \\ \langle dQ_t \rangle &= \sqrt{\eta\gamma_1/2} y(t) dt\end{aligned}$$

où $x(t)$ et $y(t)$ sont les coordonnées moyennes des états à l'instant t . On peut alors réaliser de cette moyennisation une autre estimation fiable de deux des coordonnées de l'état $\bar{\rho}_t$. Nous utilisons pour cette estimation par moyennisation un nombre beaucoup plus grand de trajectoires, en l'occurrence 3.10^6 . Ces estimations par moyenne sont représentées sur les figures par des cercles noirs. Ainsi, on peut conclure que notre méthode par Maximum de Vraisemblance donne des résultats tout à fait valables, dans le sens où une estimation plus classique par moyennisation du signal de fluorescence rentre systématiquement dans l'intervalle de confiance prédit par ce Maximum de Vraisemblance.

La figure 4 est similaire à la figure 2, à la seule différence que le nombre de trajectoires utilisées pour obtenir l'estimation par moyennisation est de seulement 4.10^4 . La quantité d'information disponible est donc ici la même que pour l'estimation par Maximum de Vraisemblance. On s'aperçoit alors que les cercles noirs ne donnent plus une estimation suffisamment précise des coordonnées en x et y de $\bar{\rho}_t$. Au contraire, le bruit inhérent à l'estimation par Maximum de Vraisemblance semble être beaucoup plus faible. Cela montre, comme cela était attendu, que l'exploitation de toutes les mesures entre les instants t et T donne, grâce au Maximum de Vraisemblance, un résultat bien meilleur que l'estimation par moyennisation, qui n'utilise que les mesures à l'instant t .

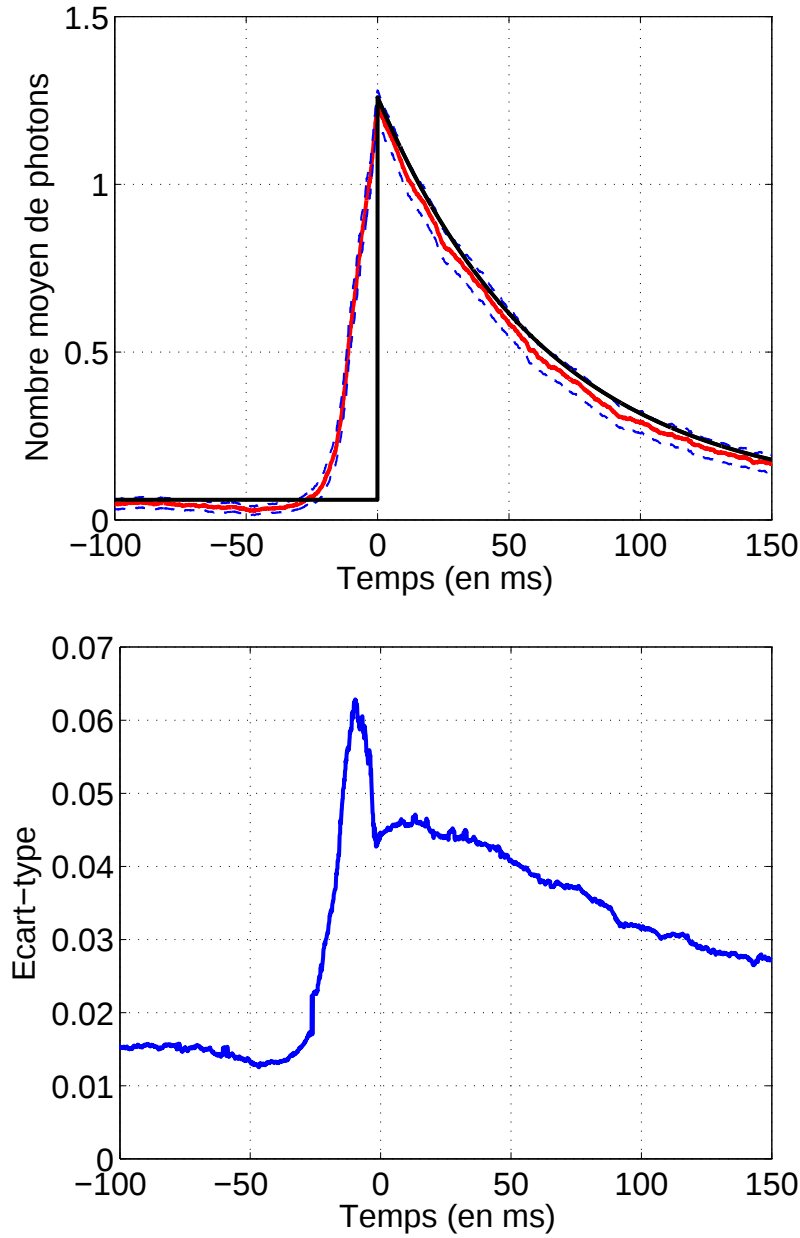


FIGURE 3.1 – En haut : la ligne rouge correspond au nombre de photons $\text{Tr}(\mathbf{N}\hat{\rho}_{ML}(t))$ calculé par Maximum de Vraisemblance, grâce aux résultats de mesures entre les instants t et 344 ms. Les deux tirets bleus correspondent à l'intervalle de confiance à 95 %, et ont pour coordonnées $\text{Tr}(\mathbf{N}\hat{\rho}_{ML}(t)) \pm 2\sqrt{\sigma_{ML}^2(\mathbf{N}, t)}$. La ligne noire correspond au nombre théorique de photons, donné par l'équation (3.47).

En bas : la valeur de l'intervalle de confiance $2\sqrt{\sigma_{ML}^2(\mathbf{N}, t)}$, obtenu grâce à l'équation (3.45).

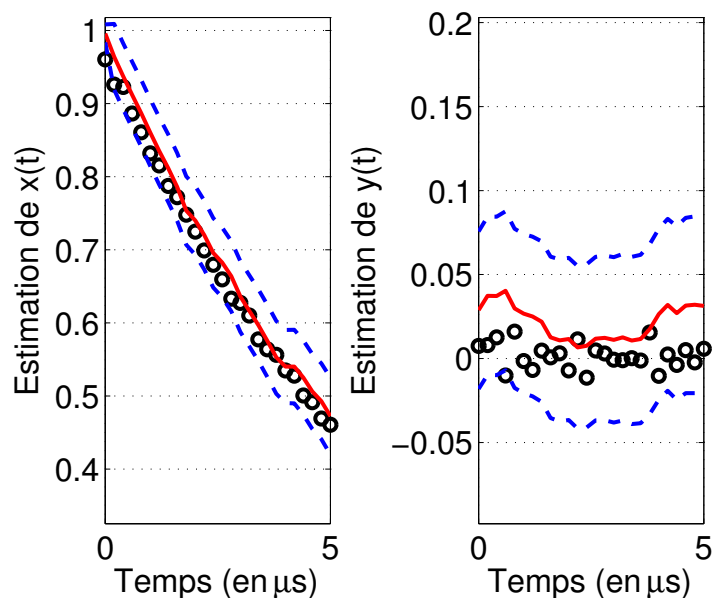


FIGURE 3.2 – Comparaison de l'estimation par Maximum de Vraisemblance de x (à gauche, rouge et tirets bleus) et de y (à droite, idem) utilisant les données de mesures provenant de $\Omega = 4.10^4$ trajectoires produites par expérience; avec une estimation par moyennisation sur l'ensemble des signaux de fluorescence (cercles noirs), utilisant un nombre bien plus important de données : $\Omega^* = 3.10^6$, expérimentales elles aussi. Les lignes rouges représentent $x_{ML}(t)$ et $y_{ML}(t)$; les tirets bleus correspondent aux intervalles de confiance de l'estimation par Maximum de Vraisemblance ($x_{ML}(t) \pm 2\sqrt{\sigma_{ML}^2(\sigma_x, t)}$ et $y_{ML}(t) \pm 2\sqrt{\sigma_{ML}^2(\sigma_y, t)}$), obtenus par (3.45). Les cercles noirs représentent l'estimation par moyennisation du signal de ces mêmes coordonnées (voir le texte pour plus de détails).

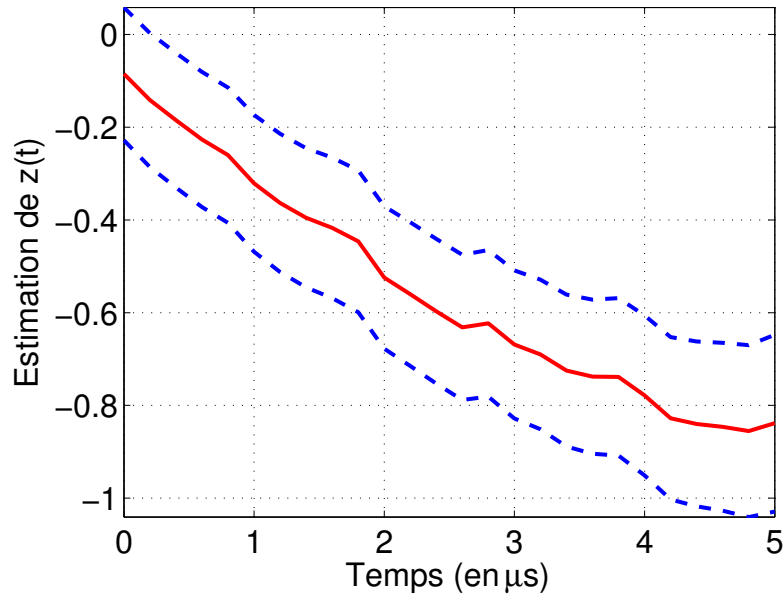


FIGURE 3.3 – Estimation par Maximum de Vraisemblance de la coordonnée z de l'état $\bar{p}_t$, utilisant les données de mesures provenant de $\Omega = 4.10^4$ trajectoires produites par expérience. La ligne rouge correspond à $z_{ML}(t)$, et les tirets bleus à l'intervalle de confiance de l'estimation par Maximum de Vraisemblance ($z_{ML}(t) \pm 2\sqrt{\sigma_{ML}^2(\sigma_z, t)}$), obtenu par (3.45).

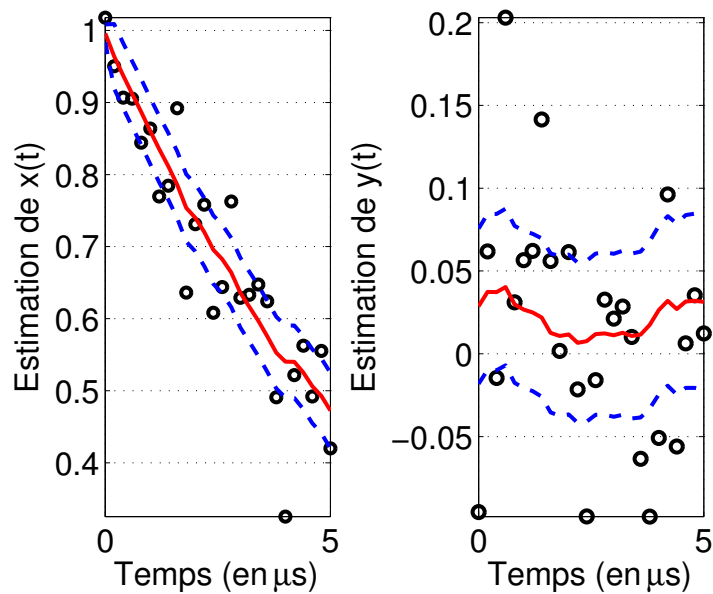


FIGURE 3.4 – Même chose que pour la figure 3.2, mais l'estimation par moyennisation est ici réalisée avec le même ensemble de trajectoires de mesures que pour le Maximum de Vraisemblance. Ce dernier apparaît significativement plus précis, prouvant la pertinence de cette méthode.

Chapitre 4

Estimation de paramètres et filtres particuliers

Nota Bene. Ce chapitre est inspiré du papier [SCIB⁺15] soumis dans le cadre de la *Conference on Decision and Control 2015*, écrit par Pierre Six, Philippe Campagne-Ibarcq*, Landry Bretheau*, Benjamin Huard* et Pierre Rouchon.

4.1 Introduction

L'estimation de paramètres au sein des chaînes de Markov à modèle caché est un sujet déjà bien établi (voir, par exemple, [CMR05]). Il y a vingt ans, Mabuchi [Mab96] a présenté des méthodes basées sur le principe du Maximum de Vraisemblance pour l'estimation de paramètres Hamiltoniens. Gambetta et Wiseman [GW01] ont ensuite détaillé une première adaptation de la technique des filtres particuliers, avec pour objectif l'estimation de paramètres dans les systèmes quantiques ouverts. Les formulations qui en découlent ont été analysées dans [CG09], en utilisant le formalisme standard des filtres quantiques. Récemment, Negretti et Mølmer [NM13] ont repris cette idée pour obtenir les équations générales définissant un filtre particulier quantique pour des systèmes évoluant selon une équation maîtresse stochastique dont le bruit équivaut à des processus de Wiener (on retrouve les systèmes diffusifs étudiés dans cette thèse).

Dans tous ces travaux, des simulations réalistes permettent de comprendre l'intérêt de tels filtres pour l'estimation de paramètres dans un espace continu. Dans [KY13], des filtres similaires sont utilisés pour choisir entre un nombre fini de paramètres : en l'occurrence, différentes structures possibles pour un réseau quantique. L'estimation Bayésienne utilisée dans l'expérience de contrôle en boucle fermée décrite dans [BWT⁺12] est en fait un cas particulier du cadre général des filtres particuliers quantiques, puisqu'ici aucune cohérence n'apparaissant entre les différents niveaux d'énergie, les probabilités manipulées restent classiques.

Dans ce chapitre, nous commençons par donner une preuve de convergence en espérance des filtres quantiques particuliers utilisés pour l'estimation de paramètres physiques, et ce

*. Laboratoire Pierre Aigrain, Ecole Normale Supérieure-PSL Research University, CNRS, Université Pierre et Marie Curie-Sorbonne Universités, Université Paris Diderot-Sorbonne Paris Cité, 24 rue Lhomond, 75231 Paris Cedex 05, France

même lorsque l'état initial des filtres construits par l'expérimentateur ne correspond pas exactement à la valeur de l'état initial du système expérimental. Nous montrons également que l'estimation par filtres particuliers quantiques revient à utiliser la technique du Maximum de Vraisemblance, sur un nombre restreint de valeurs de paramètre testées. Ensuite, nous détaillons les équations différentielles stochastiques définissant les filtres particuliers dans le cas de systèmes à mesures diffusives, puis mentionnons l'existence de schémas de discrétisation de l'équation d'évolution de ces systèmes, permettant l'implémentation des filtres particuliers pour ce type de système. Enfin, nous montrons que ces filtres sont capables de retrouver la valeur de plusieurs constantes physiques intervenant dans l'évolution d'un qubit supraconducteur, à partir de données expérimentales réelles.

4.2 Robustesse des filtres particuliers

L'estimation par filtre particulière est très liée au processus d'estimation par Maximum de Vraisemblance, bien que, dans ce cas, on n'ait pas affaire à une véritable optimisation. Pourtant, le formalisme particulier des filtres particuliers nous est très utile pour construire des résultats de robustesse de l'estimation par Maximum de Vraisemblance.

Ainsi, nous présentons dans cette section la discrétisation du processus d'estimation de paramètres par filtres particuliers, établi dans [GW01], [CG09] et [NM13] pour les systèmes à mesures continues ; puis, nous donnons plusieurs résultats mathématiques concernant la validité et la robustesse de cette estimation.

4.2.1 Filtres particuliers

Dans cette section, nous décidons de simplifier à l'extrême la prise en compte de l'incertitude sur les paramètres, puisque nous considérons que la valeur réelle $\bar{p}$ de ceux-ci ne peut prendre que deux valeurs, notées a et b . Considérons également une seule trajectoire de mesures discrètes, simplement notée $(y_1, \dots, y_T)$. Rappelons que cette trajectoire est générée par le filtre de Belavkin discrétisé suivant, dont l'état est noté ρ_k :

$$\mathbb{P}(y_k = y \mid \rho_0, y_1, \dots, y_{k-1}) = \mathbb{P}(y_k = y \mid \rho_{k-1}) = \text{Tr} \left(\mathbf{K}_{k,y}^{\bar{p}} (\rho_{k-1}) \right) \quad (4.1)$$

$$\rho_k = \frac{\mathbf{K}_{k,y_k}^{\bar{p}} (\rho_{k-1})}{\text{Tr} \left(\mathbf{K}_{k,y_k}^{\bar{p}} (\rho_{k-1}) \right)} \quad (4.2)$$

Ici, nous avons ajouté $\bar{p}$ à la notation $\mathbf{K}_{k,y}^{\bar{p}}$ des applications de Kraus décrivant l'évolution du système, puisque la valeur du paramètre pris en compte aura une importance dans les prochaines équations.

Puisqu'il n'y a ici que deux valeurs possibles de paramètres, le processus d'estimation équivaut en définitive à une discrimination entre a et b . Cette discrimination est bien entendu établie à partir des résultats de mesure $(y_1, \dots, y_T)$. Nous pouvons donc définir les probabilités π_k^a et π_k^b , qui représentent respectivement la probabilité que le paramètre réel $\bar{p}$ vaille a ou b à l'instant k , c'est-à-dire en tenant compte des mesures $(y_1, \dots, y_k)$. Bien entendu, ces probabilités vérifient $\pi_k^a + \pi_k^b = 1$ à chaque instant. Formellement, ces probabilités s'expriment comme :

$$\text{pour } p \in \{a; b\}, \quad \pi_k^p = \mathbb{P}(\bar{p} = p \mid \pi_0^a, \pi_0^b, \hat{\rho}_0, y_1, \dots, y_k) \quad (4.3)$$

Ici, $\hat{\rho}_0$ est l'estimation de l'état initial faite par l'expérimentateur, à distinguer de l'état initial ρ_0 du filtre réel défini par les équations (4.1) et (4.2). Cette estimation peut avoir

à être réalisée si la vraie valeur ρ_0 de cet état initial est inconnue de l'expérimentateur. π_0^a et π_0^b représentent l'état initial des probabilités π_k^a et π_k^b . Ces probabilités sont définies arbitrairement par l'expérimentateur, prenant en compte l'information dont il dispose a priori sur la valeur des paramètres, c'est-à-dire avant que des mesures sur le système physique ne soient enregistrées.

Une simple estimation par Maximum de Vraisemblance consisterait, à l'instant k , à choisir le paramètre p admissible qui maximise la probabilité de toutes les mesures accessibles à cet instant, c'est-à-dire :

$$\mathbb{P}(y_1, \dots, y_k \mid \hat{\rho}_0, \bar{p} = p).$$

Or, ce que nous présentons ici est quasiment équivalent, puisque, définies par l'équation (4.3), les probabilités π_k^p sont données par :

$$\pi_k^p = \mathbb{P}(y_1, \dots, y_k \mid \pi_0^a, \pi_0^b, \hat{\rho}_0, \bar{p} = p) \frac{\mathbb{P}(\bar{p} = p \mid \pi_0^a, \pi_0^b, \hat{\rho}_0)}{\mathbb{P}(y_1, \dots, y_k \mid \pi_0^a, \pi_0^b, \hat{\rho}_0)} \quad (4.4)$$

Lorsque les probabilités initiales sont posées, on peut en déduire que :

$$\pi_k^p \propto \pi_0^p \mathbb{P}(\bar{p} = p \mid \pi_0^a, \pi_0^b, \hat{\rho}_0)$$

Ainsi, lorsqu'on choisit, comme ici, une distribution de probabilité $(\pi_0^p)_p$ uniforme, sélectionner comme estimation la valeur à laquelle le filtre particulaire donne la probabilité π_k^p la plus grande à l'instant k revient à une estimation par Maximum de Vraisemblance sur $(y_1, \dots, y_k)$ sur un nombre fini de possibilités - ici a ou b .

Nous considérerons ici qu'aucune connaissance préalable n'existe sur la valeur de $\bar{p}$, de sorte que $\pi_0^a = \pi_0^b = 1/2$, mais notre analyse reste valable dans un cas plus général, tant qu'aucune de ces probabilités initiales n'est nulle. Désormais, pour alléger les formules de cette sous-section, nous omettrons de mentionner π_0^a , π_0^b et l'état initial estimé $\hat{\rho}_0$ dans l'expression des futures probabilités conditionnelles.

Pour définir l'évolution de ces probabilités, il nous faut tout d'abord, grâce à la loi de Bayes, que ces probabilités sont itérées par :

$$\begin{aligned} \text{pour } p \in \{a; b\}, \quad \pi_k^p &= \mathbb{P}(\bar{p} = p \mid y_1, \dots, y_k) \\ &= \frac{\mathbb{P}(y_k \mid \bar{p} = p, y_1, \dots, y_{k-1}) \mathbb{P}(\bar{p} = p \mid y_1, \dots, y_{k-1})}{\mathbb{P}(y_k \mid y_1, \dots, y_{k-1})} \\ &= \pi_{k-1}^p \frac{\mathbb{P}(y_k \mid \bar{p} = p, y_1, \dots, y_{k-1})}{\mathbb{P}(y_k \mid y_1, \dots, y_{k-1})} \end{aligned} \quad (4.5)$$

La valeur de la probabilité conditionnelle $\mathbb{P}(y_k \mid \bar{p} = p, y_1, \dots, y_{k-1})$ peut alors être calculée en construisant le filtre de Belavkin construit avec le paramètre p . Nous devons donc, pour itérer les probabilités π_k^a et π_k^b , générer deux filtres approchés, bâtis avec des applications de Kraus partielles faisant intervenir les valeurs a et b respectivement. L'état de ces deux filtres à l'instant k est écrit $\hat{\rho}_k^a$ et $\hat{\rho}_k^b$. Il est calculé par :

$$\text{pour } p \in \{a; b\}, \quad \hat{\rho}_0^p = \hat{\rho}_0, \quad \hat{\rho}_k^p = \frac{\mathbf{K}_{k,y_k}^p(\hat{\rho}_{k-1}^p)}{\text{Tr}(\mathbf{K}_{k,y_k}^p(\hat{\rho}_{k-1}^p))} \quad (4.6)$$

L'utilisation de ces filtres approchés, incrémentés dans la pratique de façon parallèle, permet finalement de donner la loi d'itération des probabilités conditionnelles :

$$\text{pour } p \in \{a; b\}, \quad \pi_k^p = \pi_{k-1}^p \frac{\text{Tr} \left(\mathbf{K}_{k, y_k}^p \left(\hat{\rho}_{k-1}^p \right) \right)}{\sum_{q \in \{a; b\}} \pi_{k-1}^q \text{Tr} \left(\mathbf{K}_{k, y_k}^q \left(\hat{\rho}_{k-1}^q \right) \right)} \quad (4.7)$$

Nous avons donc défini un filtre particulier, d'état $(\hat{\rho}_k^a, \hat{\rho}_k^b, \pi_k^a, \pi_k^b)$ à l'instant k , itéré grâce aux mesures $(y_1, \dots, y_k)$, qui détermine toute notre information sur le système à cet instant. Ici, les mesures proviennent du filtre de Belavkin standard correspondant à la valeur de paramètre correcte $\bar{p}$. Voyons comment faire du filtre étendu d'état $(\hat{\rho}_k^a, \hat{\rho}_k^b, \pi_k^a, \pi_k^b)$ un filtre quantique capable d'avoir produit lui-même les mesures y_k .

4.2.2 Description du système par filtre étendu

L'information dont l'expérimentateur dispose est de deux sortes ici : les probabilités π_k^a et π_k^b décrivent sa connaissance de la valeur de $\bar{p}$, et les états conditionnels $\hat{\rho}_k^a$ et $\hat{\rho}_k^b$ contiennent l'information sur l'état du système, dans chacune des cas de figure : que $\bar{p}$ vaille a ou b .

Nous montrons dans cette sous-section comment l'extension de l'espace de Hilbert associé au système, qui donne naissance à des états et des filtres qualifiés d'étendus, permet de prendre en compte simultanément ces deux types d'information, et ainsi d'énoncer des résultats de convergence s'appliquant à l'estimation de paramètres par filtres particuliers.

Notion d'état étendu

Puisque deux valeurs de paramètre sont possibles, on peut représenter l'information dont on dispose sur la valeur de $\bar{p}$ par un opérateur densité sur l'espace de Hilbert complexe de dimension 2. Cet espace est noté $\mathcal{P}$. L'incertitude sur la valeur du paramètre $\bar{p}$ étant de nature purement statistique, cet opérateur densité est simplement diagonal. A l'instant k , cette diagonale est simplement (π_k^a, π_k^b) .

Pour prendre en compte en même temps l'incertitude sur la valeur du paramètre, et l'état du système dans chacune de ces éventualités, on peut définir un nouvel état approché, dit étendu, et noté $\hat{\Xi}_k$. Cet état étendu est un opérateur de $\mathcal{H} \otimes \mathcal{P}$. Sa trace partielle par rapport à $\mathcal{H}$ est simplement la matrice diagonale sur $\mathcal{P}$ dont nous parlions dans le paragraphe précédent, générée par (π_k^a, π_k^b) . Cet état étendu permettant d'obtenir les états quantiques construits par les filtres correspondant aux valeurs possibles de paramètres a et b , $\hat{\Xi}_k$ est donc construit par la formule :

$$\hat{\Xi}_k = \left(\begin{array}{c|c} \pi_k^a \hat{\rho}_k^a & 0 \\ \hline 0 & \pi_k^b \hat{\rho}_k^b \end{array} \right) \quad (4.8)$$

Nous rappelons que les états $\hat{\rho}_k^a$ et $\hat{\rho}_k^b$ sont les états des filtres approchés de l'équation (4.6).

Avant de déterminer l'équation d'évolution de cet état étendu, notons que l'état $\hat{\Xi}_k$ construit par la formule (4.8) est un opérateur densité de $\mathcal{H} \otimes \mathcal{P}$. En effet, il est facile de montrer que sa trace vaut 1, et qu'elle est une matrice positive, puisque diagonale par blocs d'opérateurs positifs.

Évolution du filtre étendu approché

Donnons tout d'abord ce qui peut être qualifié d'état initial d'un futur filtre "étendu", qui aurait $\hat{\Xi}_k$ comme état. Cet état initial a pour formule :

$$\hat{\Xi}_0 = \left(\begin{array}{c|c} \pi_0^a \hat{\rho}_0 & 0 \\ \hline 0 & \pi_0^b \hat{\rho}_0 \end{array} \right) \quad (4.9)$$

Nous cherchons maintenant à donner une expression de $\hat{\Xi}_k$ en fonction de l'état précédent $\hat{\Xi}_{k-1}$. Nous présentons d'abord une expression construite par les éléments de base de cet état, c'est-à-dire $(\hat{\rho}_{k-1}^a, \hat{\rho}_{k-1}^b, \pi_{k-1}^a, \pi_{k-1}^b)$. En servant des équations (4.6) et (4.7), nous pouvons montrer, en utilisant de plus la linéarité des applications de Kraus, que :

$$\hat{\Xi}_k = \left(\begin{array}{c|c} \pi_{k-1}^a \mathbf{K}_{k,y_k}^a (\hat{\rho}_{k-1}^a) & 0 \\ \hline 0 & \pi_{k-1}^b \mathbf{K}_{k,y_k}^b (\hat{\rho}_{k-1}^b) \end{array} \right) / C_k, \quad (4.10)$$

$$\text{où } C_k = \pi_{k-1}^a \text{Tr} \left(\mathbf{K}_{k,y_k}^a (\hat{\rho}_{k-1}^a) \right) + \pi_{k-1}^b \text{Tr} \left(\mathbf{K}_{k,y_k}^b (\hat{\rho}_{k-1}^b) \right) \quad (4.11)$$

Notons que l'élément de normalisation C_k n'est que la trace de la matrice qu'elle divise. En cela, l'équation que nous venons de présenter prend la même forme que l'itération d'un filtre quantique standard. Mais ici, en apparence, il n'y a pas de super-opérateur de Kraus qui génère directement $\hat{\Xi}_k$ depuis $\hat{\Xi}_{k-1}$. Ici, l'itération est faite avec $\mathbf{K}_{k,y_k}^a$ et $\mathbf{K}_{k,y_k}^b$. Construisons le super-opérateur qui permette une telle itération. Rappelons tout d'abord qu'à tout instant k , pour tout $y = 1 \dots Y$, et pour tout p égal à a ou b , il existe un ensemble fini $(M_{k,l}^p)_{l \in I_{k,y}^p}$ d'opérateurs sur $\mathcal{H}$ tels que, pour tout opérateur densité ρ :

$$\mathbf{K}_{k,y}^p (\rho) = \sum_{l \in I_{k,y}^p} M_{k,l}^p \rho (M_{k,l}^p)^\dagger$$

Supposons que, pour tout k , et tout y , les ensembles $I_{k,y}^a$ et $I_{k,y}^b$ ont le même cardinal (on peut toujours augmenter artificiellement le cardinal de $I_{k,y}^p$ en créant des matrices $M_{k,l}^p$ nulles). Pour alléger les prochaines écritures, nous choisirons désormais $I_{k,y}^a = I_{k,y}^b = I_{k,y}$. Construisons ensuite un super-opérateur partiel sur l'espace étendu $\mathcal{H} \otimes \mathcal{P}$, noté $\tilde{\mathbf{K}}_{k,y}$, tel que, pour tout opérateur ξ sur cet espace, on ait :

$$\tilde{\mathbf{K}}_{k,y} (\xi) = \sum_{l \in I_{k,y}} \left(\begin{array}{c|c} M_{k,l}^a & 0 \\ \hline 0 & M_{k,l}^b \end{array} \right) \xi \left(\begin{array}{c|c} M_{k,l}^a & 0 \\ \hline 0 & M_{k,l}^b \end{array} \right)^\dagger \quad (4.12)$$

Ainsi, si ξ est diagonal par blocs, avec pour blocs σ^a et σ^b opérateurs sur $\mathcal{H}$, $\tilde{\mathbf{K}}_{k,y} (\xi)$ restera diagonale par blocs, avec pour blocs $\mathbf{K}_{k,y}^a (\sigma^a)$ et $\mathbf{K}_{k,y}^b (\sigma^b)$. Par conséquent, la création de ce super-opérateur "étendu" permet de donner une nouvelle forme à l'équation d'évolution de l'état étendu $\hat{\Xi}_k$:

$$\hat{\Xi}_k = \frac{\tilde{\mathbf{K}}_{k,y_k} (\hat{\Xi}_{k-1})}{\text{Tr} \left(\tilde{\mathbf{K}}_{k,y_k} (\hat{\Xi}_{k-1}) \right)} \quad (4.13)$$

Il est clair que l'expression des super-opérateurs étendus de l'équation (4.12) montre que ce sont des applications de Kraus partielles. Montrons enfin que l'ensemble des applications de Kraus $(\tilde{\mathbf{K}}_{k,y})_{y=1 \dots Y}$ ainsi formulées pour l'instant k forme un ensemble

complet d'applications de Kraus sur l'espace $\mathcal{H} \otimes \mathcal{P}$. Les matrices qui composent ces applications de Kraus vérifient :

$$\sum_{y=1}^Y \sum_{l \in I_{k,y}} \left(\begin{array}{c|c} M_{k,l}^a & 0 \\ \hline 0 & M_{k,l}^b \end{array} \right)^\dagger \left(\begin{array}{c|c} M_{k,l}^a & 0 \\ \hline 0 & M_{k,l}^b \end{array} \right) = \left(\begin{array}{c|c} \sum_{y=1}^Y \sum_{l \in I_{k,y}} (M_{k,l}^a)^\dagger M_{k,l}^a & 0 \\ \hline 0 & \sum_{y=1}^Y \sum_{l \in I_{k,y}} (M_{k,l}^b)^\dagger M_{k,l}^b \end{array} \right) = I_{\mathcal{H} \otimes \mathcal{P}} \quad (4.14)$$

Ainsi, à chaque instant k , l'ensemble $(\tilde{\mathbf{K}}_{k,y})_{y=1\dots Y}$ est un ensemble complet de super-opérateurs de Kraus sur l'espace de Hilbert étendu $\mathcal{H} \otimes \mathcal{P}$. Le filtre dont $\hat{\Xi}_k$ est l'état, itéré par l'équation (4.13), a donc la structure d'un filtre quantique approché.

Filtre étendu réel

Voyons enfin si nous pouvons définir un état étendu réel, correspondant à l'évolution physique exacte du système, qui puisse évoluer avec les mêmes applications de Kraus que l'état étendu approché. Nous appelons l'état à l'instant k d'un tel système, appelé état étendu réel Ξ_k .

Tout d'abord, supposons que la valeur réelle du paramètre $\bar{p}$ est a . Si nous reprenons pour Ξ_k la même construction que celle de $\hat{\Xi}_k$, l'état initial Ξ_0 sera :

$$\Xi_0 = \left(\begin{array}{c|c} \rho_0 & 0 \\ \hline 0 & 0 \end{array} \right) \quad (4.15)$$

et l'état courant réel Ξ_k :

$$\Xi_k = \left(\begin{array}{c|c} \rho_k & 0 \\ \hline 0 & 0 \end{array} \right) \quad (4.16)$$

où ρ_0 est l'état initial du système physique réel, et ρ_k est l'état du système physique réel, itéré avec le filtre quantique des équations (4.1) et (4.2). Remarquons d'abord que, lorsque la valeur y_k de la mesure à l'instant k est connue, nous itérons l'état étendu réel avec la même application de Kraus étendue que pour l'état étendu approché, à savoir :

$$\Xi_k = \frac{\tilde{\mathbf{K}}_{k,y_k}(\Xi_{k-1})}{\text{Tr}(\tilde{\mathbf{K}}_{k,y_k}(\Xi_{k-1}))} \quad (4.17)$$

Pour que ce filtre étendu soit bien un filtre quantique discret, il faut que les super-opérateurs de Kraus qui le structurent permettent de définir, à partir de l'état Ξ_{k-1} , la même loi de probabilité sur la valeur de la mesure y_k que la loi de probabilité de l'équation (4.1), c'est-à-dire que l'on doit vérifier à chaque instant k :

$$\mathbb{P}(y_k = y \mid \rho_0, y_1, \dots, y_{k-1}) = \mathbb{P}(y_k = y \mid \Xi_{k-1}) = \text{Tr}(\tilde{\mathbf{K}}_{k,y}(\Xi_{k-1})) \quad (4.18)$$

Or, nous venons de montrer que cette trace était égale à la trace que l'équation de probabilité initiale (4.1) exige, à savoir $\text{Tr}(\mathbf{K}_{k,y}^a(\rho_{k-1}^a))$.

Ainsi, nous avons donc transformé un problème de discrimination entre deux processus de mesure (déterminés par les valeurs de paramètres a et b), en un problème de discrimination entre deux états initiaux. En effet, le système physique produit les résultats

$(y_1, \dots, y_T)$ grâce au filtre discret quantique suivant :

$$\Xi_0 = \left(\begin{array}{c|c} \rho_0 & 0 \\ \hline 0 & 0 \end{array} \right), \quad \Xi_k = \frac{\tilde{K}_{k,y_k}(\Xi_{k-1})}{\text{Tr}(\tilde{K}_{k,y_k}(\Xi_{k-1}))} \quad (4.19)$$

$$\mathbb{P}(y_k = y \mid \Xi_{k-1}) = \text{Tr}(\tilde{K}_{k,y}(\Xi_{k-1})) \quad (4.20)$$

et le filtre approché que l'expérimentateur peut construire pour simuler le comportement du système à chaque instant est incrémenté par :

$$\hat{\Xi}_0 = \left(\begin{array}{c|c} \pi_0^a \hat{\rho}_0 & 0 \\ \hline 0 & \pi_0^b \hat{\rho}_0 \end{array} \right), \quad \hat{\Xi}_k = \frac{\tilde{K}_{k,y_k}(\hat{\Xi}_{k-1})}{\text{Tr}(\tilde{K}_{k,y_k}(\hat{\Xi}_{k-1}))} \quad (4.21)$$

4.2.3 Robustesse de l'estimation par Maximum de Vraisemblance

Une pseudo-métrique : la fidélité

Pour donner des résultats de stabilité sur les filtres Markoviens représentant l'évolution d'un système quantique, il nous faut définir une pseudo-métrique sur l'ensemble $\mathcal{D}_{\mathcal{H}}$ des opérateurs densité sur l'espace de Hilbert $\mathcal{H}$. Cette pseudo-métrique définira le niveau de proximité des états quantiques que représentent deux matrices densité différentes. La pseudo-métrique que nous utiliserons ici se nomme la *fidélité*, fréquemment utilisée en théorie de l'information quantique, et ses propriétés sont plus détaillées dans [NC00].

Prenons deux matrices densité ρ et σ sur $\mathcal{H}$. Pour calculer la fidélité entre ces deux matrices, remarquons d'abord que, de toute matrice positive A de $\mathcal{L}(\mathcal{H})$, on peut définir une racine carrée, notée $\sqrt{A}$ ou $A^{\frac{1}{2}}$, elle aussi matrice positive de $\mathcal{L}(\mathcal{H})$. En effet, toute matrice positive A admet une décomposition de la forme :

$$A = QDQ^\dagger$$

où Q est une matrice unitaire sur $\mathcal{H}$, et D est une matrice diagonale, aux éléments diagonaux $(D_1, \dots, D_n)$, n étant la dimension de $\mathcal{H}$ en tant que $\mathbb{C}$ -espace vectoriel. La matrice carrée de A est alors donnée par l'équation :

$$\sqrt{A} = Q\sqrt{D}Q^\dagger$$

où $\sqrt{D}$ est la matrice diagonale générée par $(\sqrt{D_1}, \dots, \sqrt{D_n})$.

Une fois cette définition donnée, nous pouvons alors écrire l'expression de la fidélité $F(\rho, \sigma)$ entre les opérateurs densité ρ et σ :

$$\forall (\rho, \sigma) \text{ de } \mathcal{D}_{\mathcal{H}}, \quad F(\rho, \sigma) = \text{Tr} \left(\sqrt{\sqrt{\rho}\sigma\sqrt{\rho}} \right) \quad (4.22)$$

La fonction de fidélité vérifie plusieurs propriétés. Tout d'abord, nous démontrons en annexe B.5 que cette application est symétrique : pour tous opérateurs densité ρ et σ sur $\mathcal{H}$, on a $F(\rho, \sigma) = F(\sigma, \rho)$. Nous pouvons également démontrer que la valeur de cette fidélité est toujours comprise entre 0 et 1. La fidélité est nulle lorsque les supports des deux matrices densité sont orthogonaux entre eux, ce qui est une condition équivalente à $\rho\sigma = 0_{\mathcal{H}}$ l'opérateur nul de $\mathcal{H}$. Au contraire, la fidélité vaut 1 lorsque les états physiques représentés par ρ et σ sont les mêmes, donc lorsque ces matrices sont égales.

Robustesse d'un filtre quantique par rapport à son état initial

Nous reprenons dans cette sous-section un résultat précédemment publié par Pierre Rouchon, dans [SDSR12] et [Rou11].

Nous considérons ici un filtre quantique en temps discret quelconque. L'action de ce filtre est décrite, à l'instant k , par l'ensemble complet d'applications de Kraus $(\mathbf{K}_{k,y})_{y=1\dots Y}$. L'état du système physique associé à ce filtre est alors, à l'instant k , noté ρ_k . Ces applications de Kraus sont supposées connues à l'expérimentateur, mais l'état initial ρ_0 , lui, ne l'est pas, et on doit alors en donner une estimation $\hat{\rho}_0$, qui permet d'itérer, en utilisant les mêmes applications de Kraus, un état approché $\hat{\rho}_k$ du système.

Nous donnons ici un résultat à propos de la fidélité, à chaque instant, entre l'état réel ρ_k du système, et l'estimation $\hat{\rho}_k$ qui en est faite grâce au filtre approché. Ce résultat, dont la démonstration est donnée dans [Rou11], utilise la structure particulière des applications de Kraus.

Théorème 4. Soit un filtre quantique réel, d'état initial ρ_0 , générant grâce aux applications de Kraus $(\mathbf{K}_{k,y})_{y=1\dots Y}$ les résultats de mesure y_k et les états courants ρ_k ; soit également un filtre approché de ce premier filtre, généré à partir de $\hat{\rho}_0$, itéré avec les mêmes applications de Kraus partielles, produisant les états courants approchés $\hat{\rho}_k$.

Alors, à chaque instant k , la fidélité entre état réel et état approché augmente en espérance par rapport à la valeur de y_k , c'est-à-dire :

$$\mathbb{E}(F(\rho_k, \hat{\rho}_k) \mid \rho_{k-1}, \hat{\rho}_{k-1}) \geq F(\rho_{k-1}, \hat{\rho}_{k-1})$$

Ce résultat signifie qu'en pratique, un filtre quantique a tendance à oublier de petites erreurs sur sa condition initiale au cours du temps. Ce résultat est très important, puisque, dans chaque expérience, l'état initial ne peut jamais être connu parfaitement. Nous nous assurons donc ici qu'il n'est pas probable que l'état d'un filtre approché diverge de l'état du filtre dont il reprend les mesures.

Puisque nous avons transformé plus haut notre problème d'estimation de paramètres en problème d'estimation d'état initial, nous allons nous servir de ce théorème pour donner un résultat de robustesse des filtres étendus, que l'on reliera à notre capacité à discriminer deux valeurs de paramètres à partir de résultats de mesure.

Robustesse du filtre étendu

Le second résultat que nous présentons est simplement une application du théorème 4 au cas particulier des filtres étendus, dans le cadre d'un choix de paramètres entre les valeurs a et b . En effet, nous avons construit plus haut un filtre quantique étendu, d'état initial Ξ_0 , générant avec les applications de Kraus $(\tilde{\mathbf{K}}_{k,y})_{y=1\dots Y}$ les mesures y_k et l'état courant Ξ_k .

De manière analogue à ce qui a été présenté dans le paragraphe précédent, nous développons à partir de ce filtre un filtre approché d'état initial $\hat{\Xi}_0$ et d'état courant $\hat{\Xi}_k$. La fidélité entre les états étendus Ξ_k et $\hat{\Xi}_k$ croît donc en espérance à chaque instant. Rappelons que a est la valeur réelle des paramètres, que Ξ_k et $\hat{\Xi}_k$ sont diagonales par blocs, d'expression $\Xi_k = \text{diag}(\rho_k, 0)$ et $\hat{\Xi}_k = \text{diag}(\pi_k^a \hat{\rho}_k^a, \pi_k^b \hat{\rho}_k^b)$. On peut facilement montrer que la fidélité de ces deux états étendus est alors :

$$F(\Xi_k, \hat{\Xi}_k) = \pi_k^a F(\rho_k, \hat{\rho}_k^a)$$

L'exploitation du théorème 4 sur les états étendus Ξ_k et $\hat{\Xi}_k$ permet alors de formuler le théorème suivant.

Théorème 5. Soit un filtre quantique réel, d'état initial ρ_0 , générant grâce aux applications de Kraus $(K_{k,y}^{\bar{p}})_{y=1\dots Y}$ les résultats de mesure y_k et les états courants ρ_k ; $\bar{p}$ est la valeur exacte de paramètres physiques inconnus, dont nous savons que $\bar{p} \in \{a; b\}$. Soit également les filtres approchés de ce premier filtre, généré à partir d'un état initial $\hat{\rho}_0$ approché, et itérés avec les applications de Kraus partielles correspondant respectivement aux valeurs de paramètres a et b , produisant les états courants approchés $\hat{\rho}_k^a$ et $\hat{\rho}_k^b$. Soit enfin π_0^a et π_0^b les probabilités (non nulles) que $\bar{p}$ vaille a ou b avant l'expérience, et π_k^a et π_k^b les probabilités de ces valeurs après prise en compte Bayésienne des résultats de mesure $(y_1, \dots, y_k)$.

Alors, à chaque instant k , le produit $\pi_k^{\bar{p}} F(\rho_k, \hat{\rho}_k^{\bar{p}})$, qui lie l'état du système physique et l'état approché correspondant à la valeur de paramètres correcte, augmente en espérance sur la valeur de y_k :

$$\mathbb{E}(\pi_k^{\bar{p}} F(\rho_k, \hat{\rho}_k^{\bar{p}}) \mid \rho_{k-1}, \hat{\rho}_{k-1}^{\bar{p}}, \pi_{k-1}^{\bar{p}}) \geq \pi_{k-1}^{\bar{p}} F(\rho_{k-1}, \hat{\rho}_{k-1}^{\bar{p}}) \quad (4.23)$$

Lorsque l'expérimentateur a une connaissance parfaite de l'état initial ρ , l'état du filtre correspondant à $\bar{p}$ est égal à l'état du système physique, et nous avons simplement :

$$\mathbb{E}(\pi_k^{\bar{p}} \mid \rho_{k-1}, \hat{\rho}_{k-1}^{\bar{p}}) \geq \pi_{k-1}^{\bar{p}} \quad (4.24)$$

Ces deux résultats montrent qu'en moyenne, la technique d'estimation de paramètres par filtres particulières approchés tend à sélectionner la bonne valeur du paramètre. Une piste de réflexion future sur le sujet pourrait être de trouver des conditions suffisantes sur le processus de mesure et sur les états initiaux réel ρ_0 et estimé $\hat{\rho}_0$ pour que la valeur $\pi_{k-1}^{\bar{p}} F(\rho_{k-1}, \hat{\rho}_{k-1}^{\bar{p}})$ converge presque sûrement vers 1 lorsque le nombre de mesures prises en compte tend vers l'infini.

4.3 Généralisation de la notion de robustesse

4.3.1 Extension à un espace de valeurs de paramètre possibles

Les résultats que nous avons présentés dans la section précédente supposaient seulement deux valeurs possibles de paramètre. Il est facilement vérifiable que tous ces résultats peuvent être étendus au cas où la valeur $\bar{p}$ appartient à un nombre fini de valeurs possibles, notées $p_1, \dots, p_r$.¹

4.3.2 Extension aux trajectoires de mesure successives

Nous étendons maintenant la méthode des filtres particulières au cas où l'on dispose désormais de Ω trajectoires de mesure, indexées par $n \in \{1; \dots; \Omega\}$. La trajectoire n est écrite $(y_k^{(n)})_{n \leq \Omega}$. Notons que les longueurs des trajectoires T_n peuvent être différentes, tout comme l'état initial de chaque trajectoire $\rho_0^{(n)}$, et même les applications de Kraus partielles $(K_{k^{(n)}, y_k^{(n)}}^{(n)})$ ayant produit chacune de ces trajectoires.

Nous commençons par concaténer les résultats de mesure en une seule trajectoire de mesure $(\bar{y}_k)_{k=1\dots T}$, en notant $T = \sum_{n=1}^{\Omega} T_n$, telle que :

$$(\bar{y}_k)_{k=1\dots T} = (y_1^{(1)}, \dots, y_{T_1}^{(1)}, y_1^{(2)}, \dots, y_{T_2}^{(2)}, \dots, y_1^{(\Omega)}, \dots, y_{T_\Omega}^{(\Omega)}) \quad (4.25)$$

1. Comment parler d'un ensemble non fini de possibilités ici ?

Nous pouvons montrer que cette trajectoire concaténée peut être considérée comme une unique trajectoire quantique de longueur T , produit d'un filtre quantique discret d'état initial $\rho_0^{(1)}$. Pour faire la jonction entre les trajectoires $n-1$ et n , il suffit d'appliquer, avant l'application de Kraus partielle $\mathbf{K}_{1^{(n)}, y_1^{(n)}}$, une autre application de Kraus notée $\mathbf{K}_0^{(n)}$, qui transforme l'état $\rho_{T_{n-1}}^{(n-1)}$ en $\rho_0^{(n)}$ ². L'adjonction de cette application de Kraus, complète, est complètement déterministe.³

Ainsi, nous avons construit une trajectoire de mesure $(\bar{y}_k)_{k=1\dots T}$ unique, qui est la sortie d'un filtre quantique en temps discret, et à laquelle les théorèmes de robustesse de la section précédente s'appliquent. Les filtres particuliers construits de cette façon, servant à discriminer parmi $(p_1, \dots, p_r)$, donnent naissance à des probabilités $\pi_{k,n}^{p_r}$, probabilités de chaque valeur de paramètre p_r conditionnellement aux trajectoires de mesure antérieures strictement à la trajectoire n , et aux mesures $(y_1^{(n)}, \dots, y_{k-1}^{(n)})$. Ces probabilités restent donc les mêmes lors du passage d'une trajectoire à une autre, dans le cadre de la trajectoire concaténée, alors que les états conditionnels $\hat{\rho}_{T_n}^{(n)}$ prennent la valeur de l'état initial de la trajectoire suivante.

4.4 Filtre particulière à temps continu et discrétisation

Dans la section 4.2, nous avons vu comment l'utilisation de filtres particuliers permettait de sélectionner, à partir de deux valeurs possibles, la valeur exacte d'un paramètre physique intervenant dans le processus de mesure d'un état quantique, lorsque ce processus est réalisé par des mesures discrètes. Voyons ici comment ce schéma est adapté aux systèmes physiques à mesures diffusives.

4.4.1 Filtre particulière continu

Rappelons tout d'abord que l'évolution temporelle des systèmes diffusifs, dont une description mathématique plus précise est donnée dans [BG09], est gouvernée par une équation maîtresse stochastique, qui s'écrit :

$$d\rho_t = \left(-\frac{i}{\hbar} [\hat{H}(t), \rho_t] + \sum_{\nu=1}^M \mathcal{L}_\nu(\rho_t) + \sum_{\lambda=1}^M \mathcal{L}_\lambda(\rho_t) \right) dt + \sum_{\nu=1}^M \sqrt{\eta_\nu} \left(\mathcal{M}_\nu(\rho_t) - \overline{\mathcal{M}}_\nu(\rho_t) \rho_t \right) dW_t^\nu \quad (4.26)$$

où ρ_t (opérateur densité de $\mathcal{H}$) est l'état du système au temps t , où $\hat{H}(t)$ est l'Hamiltonien du système, et où, pour chaque canal de diffusion ν de 1 à M :

2. En effet, il est possible de définir, pour toute matrice densité σ , une application de Kraus $\mathbf{K}^\sigma$ telle que, pour toute matrice densité ρ , $\mathbf{K}^\sigma(\rho) = \sigma$. Il suffit d'écrire la décomposition spectrale $\sigma = \sum_{k=1}^m \lambda_k |\psi_k\rangle\langle\psi_k|$ de σ , puis de construire l'application de Kraus $\mathbf{K}^\sigma$ avec l'ensemble des matrices de mesure M_{ik} , où $i = 1, \dots, N_{\mathcal{H}}$ et $k = 1, \dots, m$, construites avec $M_{ik} = \sqrt{\lambda_k} |\psi_k\rangle\langle i|$, avec $(|i\rangle)_{i=1, \dots, N_{\mathcal{H}}}$ une base orthonormée quelconque de $\mathcal{H}$. On remarque que $\mathbf{K}^\sigma$ est alors bien une application de Kraus complète, car $\sum_{i=1}^{N_{\mathcal{H}}} \sum_{k=1}^m M_{ik}^\dagger M_{ik} = I_{\mathcal{H}}$.

3. Que faire si l'état initial $\rho_0^{(n)}$ n'est pas connu ?

- $\mathcal{L}_\nu(\cdot)$ est un super-opérateur appelé Lindbladien du canal de mesure ν , tel que :

$$\mathcal{L}_\nu(\rho) = L_\nu \rho L_\nu^\dagger - \frac{1}{2} (L_\nu^\dagger L_\nu \rho + \rho L_\nu^\dagger L_\nu)$$

pour tout opérateur ρ de $\mathcal{H}$;

- L_ν est un opérateur de $\mathcal{H}$, qui n'est pas nécessairement Hermitien, et qui décrit le processus de mesure imputable au canal ν ; L_λ est un opérateur de mesure non lue ;
- $\mathcal{M}_\nu(\cdot)$ est l'application de mesure du canal ν , vérifiant pour tout ρ :

$$\mathcal{M}_\nu(\rho) = L_\nu \rho + \rho L_\nu^\dagger,$$

et l'application $\overline{\mathcal{M}}_\nu(\cdot)$ est une application renvoyant sa trace : $\overline{\mathcal{M}}_\nu(\rho) = \text{Tr}(\mathcal{M}_\nu(\rho))$;

- $\eta_\nu \in]0; 1]$ est l'efficacité de détection du canal de mesure ν : $\eta_\nu = 1$ correspond à une mesure diffusive parfaite.
- W_t^ν est un processus de Wiener, processus de Lévy vérifiant $\mathbb{E}(dW_t^\nu dW_{t'}^\nu) = \delta_{t,t'} dt$, décrivant le bruit quantique intervenant dans la sortie produite par le canal de mesure ν . Les canaux de mesure sont indépendants entre eux, et chaque canal est relié au signal de sortie y_t^ν par :

$$dy_t^\nu = \sqrt{\eta_\nu} \overline{\mathcal{M}}_\nu(\rho_t) dt + dW_t^\nu \quad (4.27)$$

Supposons maintenant que nous cherchons à déterminer la valeur réelle $\bar{p}$ d'un paramètre, et que l'expérimentateur sait que $\bar{p}$ peut prendre un nombre fini de valeurs, notées $(p_1, \dots, p_r)$. Ce paramètre peut intervenir dans l'expression du Hamiltonien, des opérateurs de mesure L_ν , ou des efficacités de détection η_ν . Par conséquent, à partir de la trajectoire de mesure $t \mapsto y_t$ obtenue par le système physique, on peut définir r filtres approchés, ayant, avec $p \in \{p_1; \dots; p_r\}$, pour état ρ_t^p à l'instant t :

$$\begin{aligned} d\rho_t^p = & \left(-\frac{i}{\hbar} [\hat{H}^p(t), \rho_t^p] + \sum_{\nu=1}^M \mathcal{L}_\nu^p(\rho_t^p) + \sum_{\lambda=1}^{\Lambda} \mathcal{L}_\lambda^p(\rho_t^p) \right) dt \\ & + \sum_{\nu=1}^M \sqrt{\eta_\nu^p} \left(\mathcal{M}_\nu^p(\rho_t^p) - \overline{\mathcal{M}}_\nu^p(\rho_t^p) \rho_t^p \right) dW_t^\nu \end{aligned} \quad (4.28)$$

où le Lindbladien $\mathcal{L}_\nu^p(\cdot)$ et les opérateurs de mesure $\mathcal{M}_\nu^p(\cdot)$ et $\overline{\mathcal{M}}_\nu^p(\cdot)$ ont la même définition que plus haut, avec comme opérateur de mesure L_ν^p .

Notons désormais π_t^p la probabilité conditionnelle du paramètre p . Nous créons, de façon similaire à ce qui a été présenté en section 4.2, un état étendu noté ξ_t sur l'espace de Hilbert $\mathcal{H} \otimes \mathcal{P}$, avec $\mathcal{P}$ isomorphe à $\mathbb{C}^r$. Cet état est simplement diagonal par blocs sur $\mathcal{P}$, avec pour éléments diagonaux $\pi_t^{p_j} \rho_t^{p_j}$. En s'inspirant de ce qui a été fait dans [CG09], la recherche d'une équation diffusive stochastique de la forme (4.26) pour l'état étendu ξ_t , qui conserve le fait que cette matrice est diagonale par blocs, avec des blocs proportionnels égaux à $\pi_t^{p_j} \rho_t^{p_j}$, implique que les probabilités $\pi_t^{p_j}$ évoluent selon :

$$d\pi_t^{p_j} = \pi_t^{p_j} \left(\sum_{\nu=1}^M \left(C_\nu^{p_j}(\rho_t^{p_j}) - \overline{C}_t^\nu \right) (dy_t^\nu - \overline{C}_t^\nu) \right) \quad (4.29)$$

où la quantité $C_\nu^{p_j}(\rho_t^{p_j})$ est égale à :

$$C_\nu^{p_j}(\rho_t^{p_j}) = \sqrt{\eta_\nu^{p_j}} \overline{\mathcal{M}}_\nu^{p_j}(\rho_t^{p_j})$$

et où $\bar{C}_t^\nu$ en est la quantité "moyenne" :

$$\bar{C}_t^\nu = \sum_{j=1}^r \pi_t^{p_j} C_\nu^{p_j} (\rho_t^{p_j})$$

Nous avons donc donné un premier formalisme de filtre particulière en temps continu. Voyons ensuite comment utiliser le schéma de discrétisation des systèmes diffusifs présenté en paragraphe 2.2.4 pour implémenter numériquement ces filtres particuliers dans le cas des systèmes quantiques diffusifs.

4.4.2 Filtre particulière discrétisé

Pour définir un filtre particulière qui puisse être implémenté numériquement, dans le cas des systèmes diffusifs, nous pouvons également nous servir des schémas de discrétisation présenté dans le premier chapitre de cette thèse.

Premièrement, nous pouvons montrer que l'équation différentielle stochastique qui gouverne l'évolution des systèmes diffusifs peut être réécrite sous la forme habituelle des filtres de Belavkin en temps discret. En effet, un système diffusif, d'état initial, peut être décrit par l'équation :

$$\rho_{t+dt} = \frac{\mathbf{K}_{t, d\mathbf{y}_t, dt}(\rho_t)}{\text{Tr}(\mathbf{K}_{t, d\mathbf{y}_t, dt}(\rho_t))}$$

où $d\mathbf{y}_t$ est un vecteur réunissant tous les canaux de mesure : $d\mathbf{y}_t = (dy_t^1, \dots, dy_t^M)$, et où le super-opérateur $\mathbf{K}_{t, d\mathbf{y}_t, dt}$ a une structure d'application de Kraus partielle.

La discrétisation de cette équation se fait ainsi : notons Δt la période de discrétisation du schéma, Δy_k^ν le signal de mesure du canal ν intégré entre les instants $(k-1)\Delta t$ et $k\Delta t$, et $\Delta \mathbf{y}_k = (\Delta y_k^1, \dots, \Delta y_k^M)$. Rappelons que le schéma de discrétisation d'ordre 1 prend alors la forme :

$$\mathbf{K}_{t, \Delta \mathbf{y}_t, \Delta t}(\rho_t) = M_{t, \Delta \mathbf{y}_t, \Delta t} \rho_t M_{t, \Delta \mathbf{y}_t, \Delta t}^\dagger + \sum_{\nu=1}^M (1 - \eta_\nu) L_\nu \rho_t L_\nu^\dagger \Delta t + \sum_{\lambda=1}^\Lambda L_\lambda \rho_t L_\lambda^\dagger \Delta t \quad (4.30)$$

où la matrice $M_{t, \Delta \mathbf{y}_t, \Delta t}$ a pour expression :

$$M_{t, \Delta \mathbf{y}_t, \Delta t} = I_{\mathcal{H}} - \left(i\hat{H}(t) + \sum_{\nu=1}^M \frac{L_\nu^\dagger L_\nu}{2} + \sum_{\lambda=1}^\Lambda \frac{L_\lambda^\dagger L_\lambda}{2} \right) \Delta t + \sum_{\nu=1}^M \sqrt{\eta_\nu} L_\nu \Delta y_t^\nu$$

Le schéma de discrétisation devient alors :

$$\rho_t = \frac{\mathbf{K}_{t, \Delta \mathbf{y}_t, \Delta t}(\rho_{t-\Delta t})}{\text{Tr}(\mathbf{K}_{t, \Delta \mathbf{y}_t, \Delta t}(\rho_{t-\Delta t}))}$$

Lorsque l'état en question est celui du système physique, la densité de probabilité des mesures discrétisées $\Delta y_t^1, \dots, \Delta y_t^M$, connaissant l'état $\rho_{t-\Delta t}$, est donnée par :

$$\frac{\text{Tr}(\mathbf{K}_{t, \Delta \mathbf{y}_t, \Delta t}(\rho_{t-\Delta t}))}{\sqrt{2\pi}(\Delta t)^M} \exp\left(-\frac{\|\Delta \mathbf{y}_t\|^2}{2\Delta t}\right)$$

Pour une mesure $\Delta \mathbf{y}_t$ donnée, la densité de probabilité de cette mesure est seulement proportionnelle à $\text{Tr}(\mathbf{K}_{t, \Delta \mathbf{y}_t, \Delta t}(\rho_{t-\Delta t}))$. Nous pouvons donc utiliser la méthode des filtres particuliers discrets de la même façon, pour les systèmes diffusifs.

Lemme 3. La discrétisation de l'équation d'évolution dans le filtre particulaire des équations (4.28) et (4.29) permet d'écrire, pour tout $l \in \{p_1; \dots; p_r\}$:

$$\rho_t^{p_l} = \frac{\mathbf{K}_{t, \mathbf{dy}_t, dt}^{p_l} (\rho_{t-dt}^{p_l})}{\text{Tr} \left(\mathbf{K}_{t, \mathbf{dy}_t, dt}^{p_l} (\rho_{t-dt}^{p_l}) \right)} \quad (4.31)$$

$$\pi_t^{p_l} = \frac{\pi_{t-dt}^{p_l} \text{Tr} \left(\mathbf{K}_{t, \mathbf{dy}_t, dt}^{p_l} (\rho_{t-dt}^{p_l}) \right)}{\sum_{l'=1}^r \pi_{t-dt}^{p_{l'}} \text{Tr} \left(\mathbf{K}_{t, \mathbf{dy}_t, dt}^{p_{l'}} (\rho_{t-dt}^{p_{l'}}) \right)} \quad (4.32)$$

où $\mathbf{dy}_t = (dy_t^1, \dots, dy_t^M)$ et où la version discrétisée de l'application $\mathbf{K}_{t, \mathbf{dy}_t, dt}^{p_l}$ est donnée par :

$$\rho \mapsto \mathbf{K}_{t, \Delta \mathbf{y}_t, \Delta t}^{p_l}(\rho) = M_{t, \Delta \mathbf{y}_t, \Delta t}^{p_l} \rho (M_{t, \Delta \mathbf{y}_t, \Delta t}^{p_l})^\dagger + \sum_{\nu=1}^M (1 - \eta_\nu^{p_l}) L_\nu^{p_l} \rho (L_\nu^{p_l})^\dagger \Delta t + \sum_{\lambda=1}^\Lambda L_\lambda^{p_l} \rho (L_\lambda^{p_l})^\dagger \Delta t$$

et où l'opérateur $M_{t, \Delta \mathbf{y}_t, \Delta t}^{p_l}$ sur $\mathcal{H}$ est donné par la formule :

$$M_{t, \Delta \mathbf{y}_t, \Delta t}^{p_l} = I_{\mathcal{H}} - \left(i \hat{H}^{p_l}(t) + \sum_{\nu=1}^M \frac{(L_\nu^{p_l})^\dagger L_\nu^{p_l}}{2} + \sum_{\lambda=1}^\Lambda \frac{(L_\lambda^{p_l})^\dagger L_\lambda^{p_l}}{2} \right) \Delta t + \sum_{\nu=1}^M \sqrt{\eta_\nu^{p_l}} L_\nu^{p_l} \Delta y_t^\nu$$

De la même façon, l'utilisation conjointe de ce théorème et du formalisme mathématique employé dans [APR14] ouvre la voie à un résultat analogue au théorème 5 pour les systèmes diffusifs.

4.5 Validations expérimentales

Dans cette dernière section, nous utilisons les filtres particuliers pour l'estimation de paramètres, sur des données réelles. Le système dont les trajectoires de mesure proviennent (voir l'annexe A.2) est un qubit supraconducteur, dont le canal de fluorescence est observé. La mesure est donc ici diffusive et l'espace de Hilbert $\mathcal{H}$ en question isomorphe à $\mathbb{C}^2$. Par conséquent, le système évolue selon l'équation maîtresse stochastique (4.26), où $M = 2$ et $\Lambda = 1$. L'efficacité de détection $\eta_1 = \eta_2 = \eta$ est l'efficacité de détection totale du canal de fluorescence ($\eta > 0$), et le canal L_λ correspond à un canal de déphasage non lu. Les opérateurs de mesure associés sont :

$$L_1 = \frac{1}{\sqrt{2T_1}} \frac{\sigma_x - i\sigma_y}{2}, \quad L_2 = iL_1, \quad L_3 = \frac{1}{\sqrt{2T_\phi}} \sigma_z, \quad (4.33)$$

où σ_x , σ_y et σ_z sont les matrices de Pauli habituelles [NC00]. Les valeurs de $T_1 = 4.14 \mu s$ et de $T_1 = 35 \mu s$ ont été déterminées par des expériences de calibration a priori, utilisant des protocoles de Rabi ou de Ramsey. En revanche, η n'est pas connu avec beaucoup de précision. Une première calibration de cette efficacité de détection, reposant sur la moyenne des signaux de décohérence, avait permis d'obtenir $\bar{\eta} = 0.36 \pm 0.02$.

Nous commençons donc par obtenir une meilleure estimation de η grâce aux filtres particuliers, puis nous montrons que la valeur de T_1 peut également être retrouvée, de la même façon, avec une précision correcte. Nous disposons ici de $\Omega = 3 \times 10^6$ trajectoires quantiques de $10 \mu s$, commençant toutes par le même état initial, estimé à $\hat{\rho}_0 = (I_{\mathcal{H}} + \sigma_x)/2$ avec une bonne précision. Les trajectoires de mesure que nous utilisons

ont une période d'échantillonnage $\Delta t = 200$ ns. Au sein de chaque trajectoire, nous disposons donc des mesures discrétisées $(\Delta y_1^1, \dots, \Delta y_{50}^1)$ et $(\Delta y_1^2, \dots, \Delta y_{50}^2)$, obtenues par $\Delta y_k^\nu = \int_{t'=(k-1)\Delta t}^{t'=k\Delta t} dy_t^\nu$, dy_t^ν provenant de l'équation différentielle (4.27). Le filtre particulaire que nous utilisons est simplement celui du lemme 3, où les éléments infinitésimaux dt et dy_t^ν sont remplacés par les discrétisations Δt et Δy_t^ν . Enfin, pour chaque filtre, les trajectoires de mesure utilisées sont concaténées, comme expliqué en sous-section 4.3.2.

4.5.1 Estimation de l'efficacité de détection

Nous utilisons donc le filtre particulaire pour tester différentes valeurs $\eta^1, \dots, \eta^r$. Ces valeurs sont d'abord choisies autour de la valeur attendue $\hat{\eta} = 0.26$, et très différentes, pour valider la pertinence expérimentale du processus d'estimation. Ensuite, nous augmentons le nombre de trajectoires utilisées et resserrons les valeurs testées autour des valeurs que le filtre particulaire précédent a sélectionnées. Nous développons ainsi trois filtres particuliers de précision croissante, dont les résultats sont à lire sur les figures 4.1, 4.2 et 4.3. Sur chaque figure, l'axe des abscisses représentent le nombre de trajectoires déjà prises en compte dans le jeu de de probabilités $\pi_k^{p_i}$. Ces probabilités sont affichées en axe des ordonnées.

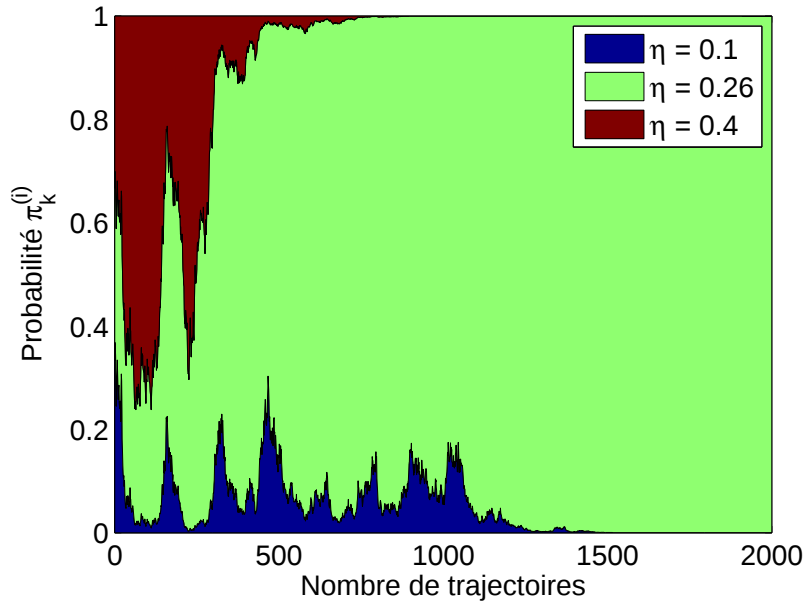


FIGURE 4.1 – Première estimation de $\bar{\eta}$, avec comme valeurs testées $\eta^1 = 0.10$, $\eta^2 = 0.26$ (on sait que la véritable valeur $\bar{\eta}$ en est proche), et $\eta^3 = 0.40$. Seulement 2000 trajectoires sont nécessaires pour sélectionner 0.26 et éliminer les deux autres valeurs.

4.5.2 Estimation du temps de relaxation

Nous cherchons maintenant à utiliser les filtres particuliers pour retrouver la valeur du temps de relaxation $\bar{T}_1$. Nous supposons dans ce cas la valeur $\hat{\eta} = 0.2425$ estimée plus haut. Nous devrions retrouver une estimation de l'ordre de $T_1 = 4.15 \mu s$. Commençons par implémenter un filtre particulaire testant les valeurs $2.0 \mu s$, $3.3 \mu s$ et $4.1 \mu s$, construit grâce au schéma de discrétisation simple des équations (2.42) et (2.44), dont nous nous

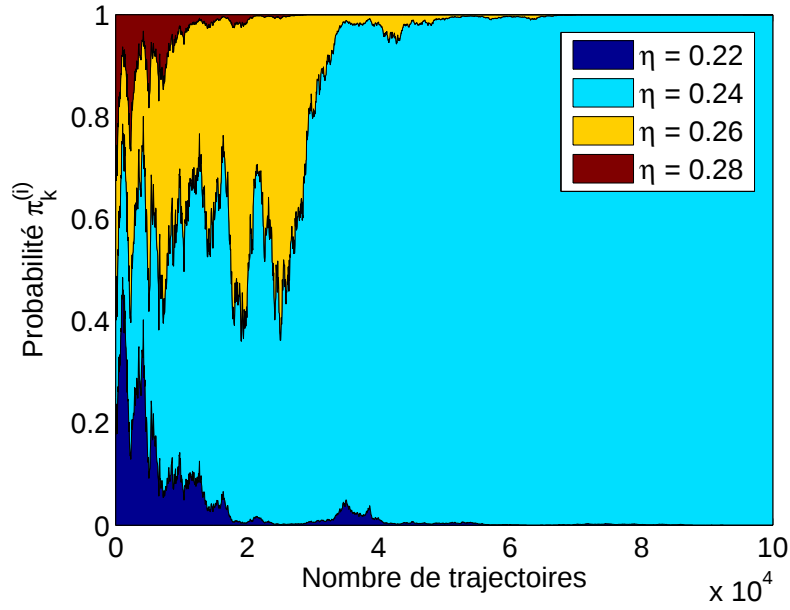


FIGURE 4.2 – Seconde estimation de $\bar{\eta}$, beaucoup plus resserrée autour de 0.26. Nous laissons un intervalle de 0.02 entre chaque valeur. On s’aperçoit que $\bar{\eta}$ est en réalité plus proche de 0.24 que de 0.26. Le nombre de trajectoires nécessaires pour la discrimination, 1×10^5 , est beaucoup plus grand.

étions déjà servir pour les filtres particuliers portant sur η . Les résultats sont donnés en figure 4.4.

Comme on le voit sur la figure, le filtre particulier ne converge pas vers la valeur attendue. En effet, comme nous l’avons expliqué dans l’introduction de cette thèse, le schéma de discrétisation simple favorise, lors d’une procédure d’estimation par Maximum de Vraisemblance, les valeurs de paramètres donnant de l’amplitude aux opérateurs de mesure. Il est donc normal qu’ici, les basses valeurs de T_1 soient avantagés, comme il est normal que l’estimation qui portait sur η n’ait pas été affectée. Nous utiliserons désormais le schéma de discrétisation modifié, construit avec la matrice de mesure $M_{\Delta \mathbf{y}_t}$ définie par l’équation (2.47). Les résultats sont présentés dans les figures 4.5, 4.6 et 4.7, et aboutissent à une estimation satisfaisante $\hat{T}_1 = 3.95 \pm 0.05 \mu s$.

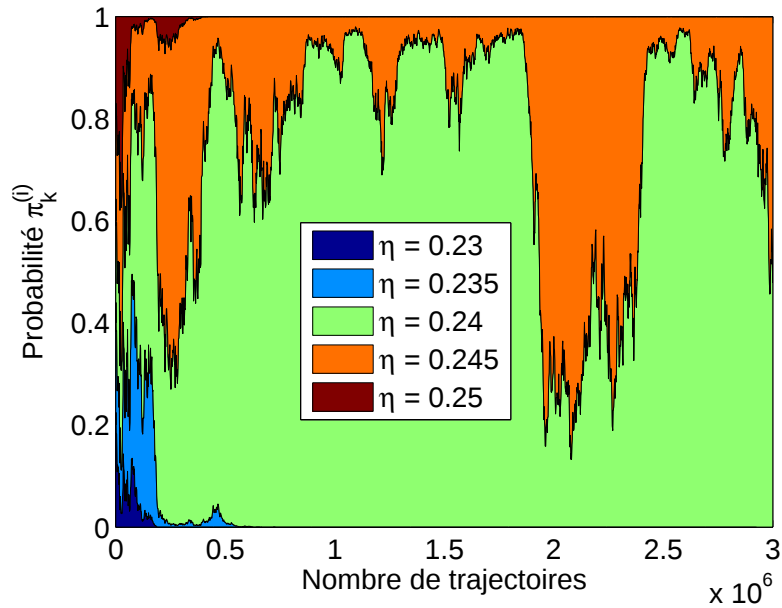


FIGURE 4.3 – La dernière estimation est encore plus resserrée, avec des intervalles de 0.005. Toutes les trajectoires disponibles sont prises en compte, mais le filtre n’arrive pas à choisir entre 0.240 et 0.245.

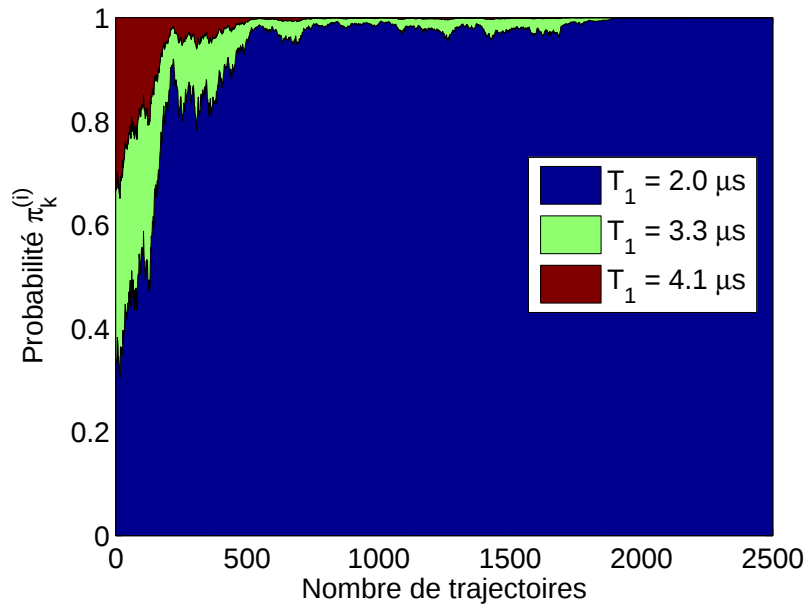


FIGURE 4.4 – Estimation de $\bar{T}_1$ réalisée, avec un schéma de discrétisation d’ordre 1. On voit que le filtre converge très rapidement (seules 2500 trajectoires sont représentées ici) vers la plus basse des valeurs testées, $2.0 \mu\text{s}$, qui est complètement irréaliste, compte tenu de la calibration faite avant l’expérience.

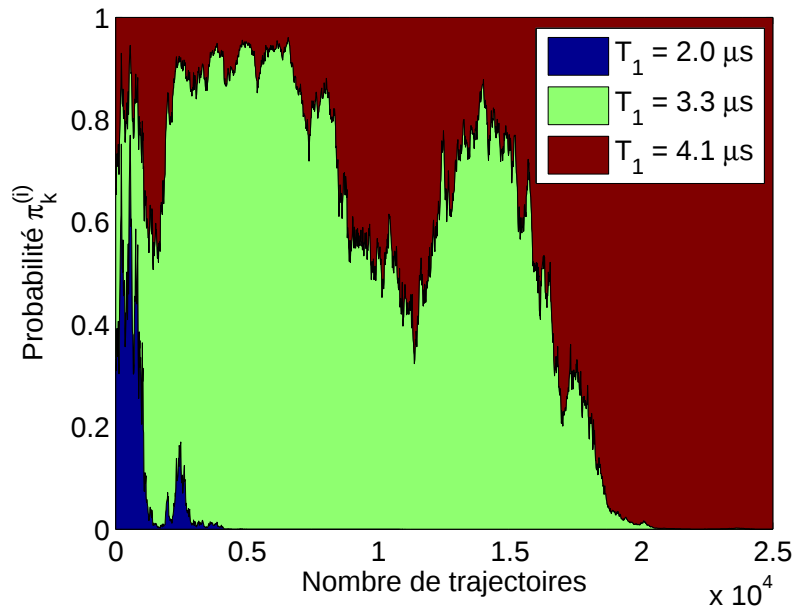


FIGURE 4.5 – Première estimation de $\bar{T}_1$ utilisant le schéma de discrétisation d'ordre 2. Nous testons ici les mêmes valeurs que précédemment, pour comparer la pertinence des deux schémas de mesure. Environ 25×10^3 trajectoires sont nécessaires pour sélectionner la valeur que l'on attendait.

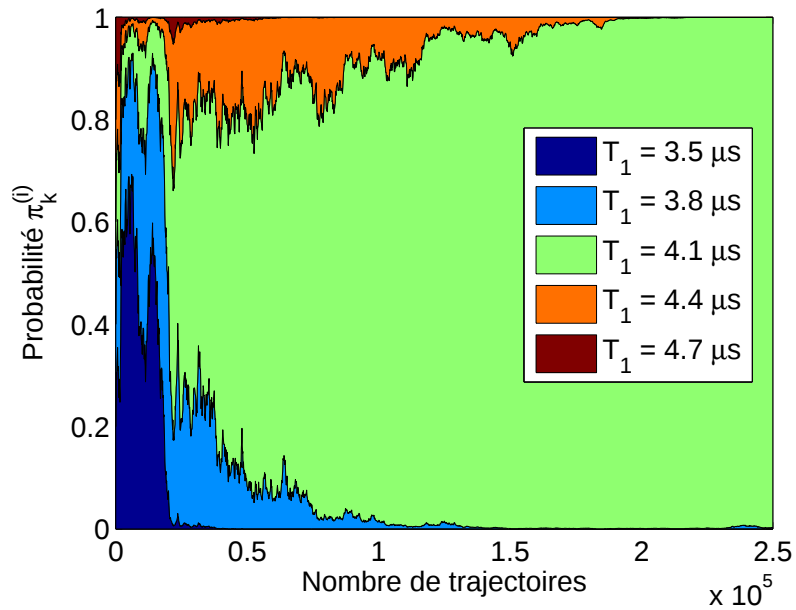


FIGURE 4.6 – Seconde estimation de $\bar{T}_1$, resserrée autour de $4.1 \mu s$. Nous laissons un intervalle de $0.3 \mu s$ entre chaque valeur.

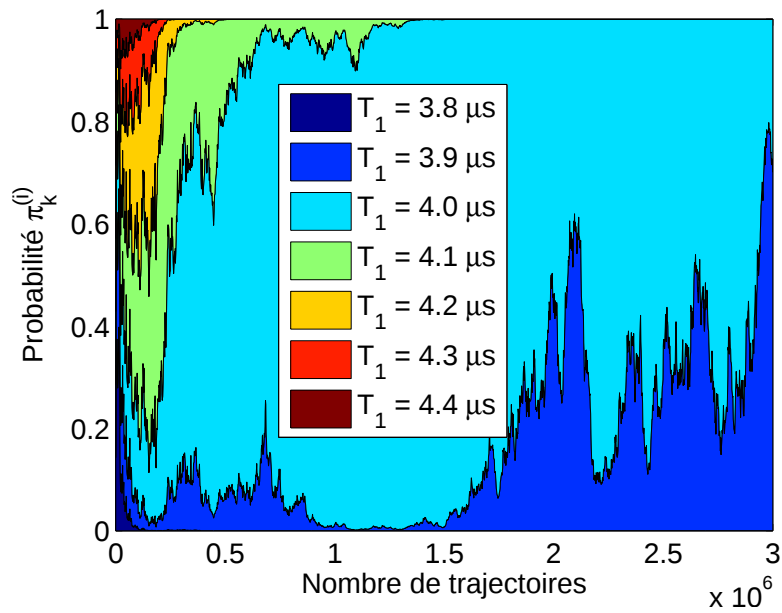


FIGURE 4.7 – Dernière estimation de $\bar{T}_1$, est encore plus resserrée, avec des intervalles de $0.1 \mu\text{s}$ entre chaque valeur testée. On s'aperçoit que le filtre n'arrive pas à choisir entre $3.9 \mu\text{s}$ et $4.0 \mu\text{s}$. Cette estimation est assez proche de ce qui était attendu, à savoir $4.15 \mu\text{s}$.

Chapitre 5

Tomographie de processus quantique

5.1 Introduction

5.1.1 Motivations

Dans le précédent chapitre, nous avons vu que l'utilisation de la technique du Maximum de Vraisemblance dans le cadre d'une tomographie de processus quantique pouvait être réalisée, lorsque nous testons un nombre fini de valeurs de paramètres, grâce à un filtre particulière quantique. Des résultats de convergence de cette estimation peuvent alors être formulés. Utilisée telle quelle, lorsque le paramètre à estimer est un scalaire, cette technique permet de discriminer entre plusieurs valeurs possibles, ce qui définit un intervalle de confiance pour l'estimation. La prochaine estimation est alors un choix entre plusieurs valeurs de cette intervalle. Ces filtres permettent également de relever l'apparition d'événements ponctuels, comme des pannes, lors desquelles la valeur d'un paramètre change de manière rapide. En répétant l'opération quelques fois, l'estimation de l'expérimentateur arrive facilement à des niveaux de précision satisfaisants.

Cependant, quand le paramètre p à estimer recouvre un grand nombre de grandeurs physiques simultanément, c'est-à-dire lorsque celui-ci évolue dans un espace de dimension trop grande, le nombre de valeurs possibles à tester devient rapidement trop grand. On retrouve là le fléau de la dimension (*curse of dimensionality*), courant en apprentissage statistique. On doit donc ici considérer le principe du Maximum de Vraisemblance comme une optimisation sur l'ensemble de toutes les valeurs de paramètres p physiquement admissibles : c'est ce que nous présentons ici.

5.1.2 Cadre de la maximisation

Rappelons le cadre dans lequel s'effectue l'estimation. Ici, nous disposons toujours de Ω trajectoires de mesures discrètes, de longueurs respectives T_n , car ces trajectoires sont indexées par n allant de 1 à Ω . La n -ième trajectoire de mesure est notée $y^{(n)} = (y_1^{(n)}, \dots, y_{T_n}^{(n)})$. Surtout, nous considérons dans ce chapitre que l'état initial réel du système est suffisamment bien connu. Par conséquent, l'état physique que l'expérimentateur reconstruit à partir des résultats de mesure est itéré à partir de l'état initial réel ρ_0 , et non pas à partir d'une quelconque approximation de celui-ci. Un paramètre p , constant tout au long de l'expérience, intervient dans le processus de mesure. Sa valeur exacte $\bar{p}$ est inconnue, et nous cherchons à l'estimer à partir des trajectoires de mesure $y^{(n)}$.

Ainsi, puisque l'instant k de la trajectoire n est aussi écrit $k^{(n)}$, nous notons $\hat{\rho}_{k^{(n)}}^p$ l'état

reconstruit par l'expérimentateur, à cet instant, en supposant une valeur de paramètres p . La probabilité conditionnelle de l'ensemble des résultats de mesure (à savoir les Ω trajectoires), en supposant une valeur de paramètres p arbitraire, est donnée par :

$$\prod_{n=1}^{\Omega} \prod_{k=1}^{T_n} \text{Tr} \left(\mathbf{K}_{\tilde{y}_k^{(n)}}^p \left(\hat{\rho}_{k-1}^{p(n)} \right) \right)$$

Dans cette formule, l'expression $\tilde{y}_k^{(n)}$ utilisée pour définir l'application de Kraus partielle recouvre, pour alléger les notations, deux informations : l'instant k et le numéro de trajectoire n (car l'expression d'une application de Kraus partielle peut dépendre du temps), et le résultat de la mesure à ce moment, $y_k^{(n)}$. Bien entendu, les applications de Kraus partielles dépendent du paramètre p choisi, et cette dépendance se retrouve donc dans l'état $\hat{\rho}_{k(n)}^p$ (sauf lorsque celui-ci est l'état initial). Cette dépendance implicite nous empêche donc de réaliser une maximisation analytique de la probabilité dont nous venons de mentionner la formule ; or, l'estimation par Maximum de Vraisemblance consiste à choisir comme estimation $\hat{p}_{ML}$ de $\bar{p}$ la valeur qui maximise cette probabilité.

Pour que l'expression de cette probabilité puisse être prise en compte numériquement, il nous faut plutôt considérer son logarithme. Ainsi, nous cherchons dans ce chapitre à déterminer la valeur $\hat{p}_{ML}$ qui maximise :

$$p \mapsto f(p) = \sum_{n=1}^{\Omega} \sum_{k=1}^{T_n} \log \left(\text{Tr} \left(\mathbf{K}_{\tilde{y}_k^{(n)}}^p \left(\hat{\rho}_{k-1}^{p(n)} \right) \right) \right) \quad (5.1)$$

Le calcul du gradient et du Hessian de la fonction f étant impossible à réaliser de manière directe, nous allons transformer ce problème d'optimisation par l'utilisation de la technique du gradient par l'adjoint.

5.2 Optimisation grâce à l'adjoint

5.2.1 Écriture du Lagrangien et optimalité

Comme nous l'avons dit plus haut, le problème que nous rencontrons est le fait que la dépendance en p des états reconstruits $\hat{\rho}_{k(n)}^p$ est très difficile à prendre en compte. En effet, ceux-ci sont définis par récurrence, grâce à l'équation :

$$\hat{\rho}_{k(n)}^p = \frac{\mathbf{K}_{\tilde{y}_k^{(n)}}^p \left(\hat{\rho}_{k-1}^{p(n)} \right)}{\text{Tr} \left(\mathbf{K}_{\tilde{y}_k^{(n)}}^p \left(\hat{\rho}_{k-1}^{p(n)} \right) \right)} \quad (5.2)$$

Lorsque k atteint la dizaine, il devient donc très difficile de calculer analytiquement la moindre dérivée en p de $\hat{\rho}_{k(n)}^p$. Nous allons donc poser différemment le problème d'optimisation. Jusqu'ici, les états $\hat{\rho}_{k(n)}^p$ étaient des fonctions implicites de p , variant avec lui au cours d'une maximisation réalisée *uniquement* sur p . Désormais, nous considérerons ces états comme des variables libres. Pour bien marquer cette différence, nous noterons X_k^n la nouvelle variable libre jouant le rôle de $\hat{\rho}_{k(n)}^p$. De même, on écrit $\mathbf{X}$, pour simplifier les formules, l'ensemble de toutes les variables libres, $(X_k^n)_{\substack{k=0 \dots T_n-1 \\ n=1 \dots \Omega}}$. Enfin, notons $\tilde{f}$ la fonction de maximisation, qui dépend désormais de p et de $\mathbf{X}$:

$$(p, \mathbf{X}) \mapsto \tilde{f}(p, \mathbf{X}) = \sum_{n=1}^{\Omega} \sum_{k=1}^{T_n} \log \left(\text{Tr} \left(\mathbf{K}_{\tilde{y}_k^{(n)}}^p \left(X_{k-1}^n \right) \right) \right) \quad (5.3)$$

Notre problème devient un problème d'optimisation sous contraintes, qui s'écrit :

$$(\hat{p}_{ML}, \hat{\mathbf{X}}_{ML}) = \arg \max_{(p, \mathbf{X})} \tilde{f}(p, \mathbf{X}), \text{ sous les contraintes suivantes,} \quad (5.4)$$

$$\text{pour } n = 1 \dots \Omega, \left\{ \begin{array}{l} h_0^n(p, \mathbf{X}) = X_0^n - \rho_0 \\ \forall k = 1 \dots T_n - 1, \quad h_k^n(p, \mathbf{X}) = X_k^n - \frac{\mathbf{K}_{\tilde{y}_k^{(n)}}^p(X_{k-1}^n)}{\text{Tr}(\mathbf{K}_{\tilde{y}_k^{(n)}}^p(X_{k-1}^n))} \end{array} \right. \quad (5.5)$$

Ce problème est complètement équivalent au problème d'optimisation initial, puisque les contraintes $(h_k^n(\cdot))_{k,n}$ définissent complètement, pour un p donné, la valeur de toutes les variables X_k^n .

Pour résoudre ce problème d'optimisation sous contraintes, il nous faut alors introduire des multiplicateurs de Lagrange pour chacune des contraintes de type $h_k^n(p, \mathbf{X}) = 0$. Notons λ_k^n le multiplicateur correspondant à cette contrainte. Il s'agit d'un opérateur Hermitien sur $\mathcal{H}$, car la fonction h_k^n renvoie, lorsque les X_k^n sont Hermitiens, des opérateurs Hermitiens sur $\mathcal{H}$. Notons enfin $\mathbf{\Lambda}$ l'ensemble de tous les multiplicateurs : $\mathbf{\Lambda} = (\lambda_k^n)_{\substack{k=0, \dots, T_n-1 \\ n=1, \dots, \Omega}}$.

Le Lagrangien du problème d'optimisation énoncé par les équations (5.4) et (5.5) prend la formule suivante :

$$\mathcal{L}(p, \mathbf{X}, \mathbf{\Lambda}) = \tilde{f}(p, \mathbf{X}) - \sum_{n=1}^{\Omega} \sum_{k=0}^{T_n-1} \text{Tr}(\lambda_k^n h_k^n(p, \mathbf{X})) \quad (5.6)$$

5.2.2 Lagrangien et conditions d'optimalité

Comment les conditions d'optimalité du problème sous contraintes peuvent-elles trouver leur expression à l'aide du Lagrangien $\mathcal{L}$ venant d'être introduit ? Voyons tout d'abord ce qu'implique l'existence d'un maximum local du Lagrangien. Ce maximum local est noté $(\hat{p}, \hat{\mathbf{X}}, \hat{\mathbf{\Lambda}})$. En ce point, nous avons donc :

$$\left. \frac{\partial \mathcal{L}}{\partial p} \right|_{(\hat{p}, \hat{\mathbf{X}}, \hat{\mathbf{\Lambda}})} = 0, \quad \left. \frac{\partial \mathcal{L}}{\partial \mathbf{X}} \right|_{(\hat{p}, \hat{\mathbf{X}}, \hat{\mathbf{\Lambda}})} = 0, \quad \left. \frac{\partial \mathcal{L}}{\partial \mathbf{\Lambda}} \right|_{(\hat{p}, \hat{\mathbf{X}}, \hat{\mathbf{\Lambda}})} = 0$$

Ces conditions nous assurent donc que, au maximum local, les contraintes du problème d'optimisation originel sont respectées, c'est-à-dire que $h_k^n(\hat{p}, \hat{\mathbf{X}}) = 0$, pour tout k et n admissible. Montrons maintenant que $\hat{p}$ est alors nécessairement un maximum local de la fonction $p \mapsto f(p)$. Puisque nous avons dit que, pour tout paramètre p , la valeur de $\mathbf{X}$ satisfaisant toutes les contraintes $h_k^n(p, \mathbf{X}) = 0$ était entièrement déterminée, il existe une fonction implicite $p \mapsto \mu(p)$ générant les états $\mathbf{X}$, c'est-à-dire :

$$\forall p \text{ admissible}, \forall n = 1, \dots, \Omega, \forall k = 0, \dots, T_n - 1, \quad h_k^n(p, \mu(p)) = 0 \quad (5.7)$$

Ainsi, $\nabla_{\mathbf{\Lambda}} \mathcal{L}|_{(\hat{p}, \hat{\mathbf{X}}, \hat{\mathbf{\Lambda}})} = 0$ implique que :

$$\hat{\mathbf{X}} = \mu(\hat{p})$$

Supposons, pour alléger les notations, que le paramètre p soit un scalaire. Lorsque p a une valeur vectorielle, il sera possible de reprendre les calculs qui vont être faits ci-dessous

pour chacune de ses composantes. Tout d'abord, la dérivée partielle en p du Lagrangien est calculé en tout point $(p, \mathbf{X}, \mathbf{\Lambda})$ par :

$$\left. \frac{\partial \mathcal{L}}{\partial p} \right|_{(p, \mathbf{X}, \mathbf{\Lambda})} = \left. \frac{\partial \tilde{f}}{\partial p} \right|_{(p, \mathbf{X})} - \sum_{n=1}^{\Omega} \sum_{k=0}^{T_n-1} \text{Tr} \left(\hat{\lambda}_k^n \left. \frac{\partial h_k^n}{\partial p} \right|_{(p, \mathbf{X})} \right) \quad (5.8)$$

Or, nous souhaitons calculer la valeur, considérant le maximum local $(\hat{p}, \hat{\mathbf{X}}, \hat{\mathbf{\Lambda}})$ du Lagrangien, de $f'(\hat{p})$. Utilisons ensuite le fait que, pour tout paramètre p :

$$f(p) = \tilde{f}(p, \mu(p))$$

pour écrire que :

$$f'(\hat{p}) = \left. \frac{\partial \tilde{f}}{\partial p} \right|_{(\hat{p}, \hat{\mathbf{X}})} + \mu'(\hat{p}) \cdot \left. \frac{\partial \tilde{f}}{\partial \mathbf{X}} \right|_{(\hat{p}, \hat{\mathbf{X}})} \quad (5.9)$$

Ensuite, nous utilisons la définition implicite, donnée en équation (5.7), de la fonction μ , pour écrire que :

$$\forall p \text{ admissible, } \left. \frac{\partial h_k^n}{\partial p} \right|_{(p, \mu(p))} + \mu'(\hat{p}) \cdot \left. \frac{\partial h_k^n}{\partial \mathbf{X}} \right|_{(p, \mu(p))} = 0 \quad (5.10)$$

C'est en particulier vrai pour le point $(\hat{p}, \hat{\mathbf{X}})$. Couplées ensemble, les deux dernières équations permettent d'écrire que :

$$\left. \frac{\partial \mathcal{L}}{\partial p} \right|_{(\hat{p}, \hat{\mathbf{X}}, \hat{\mathbf{\Lambda}})} = f'(\hat{p}) - \mu'(\hat{p}) \cdot \left. \frac{\partial \tilde{f}}{\partial \mathbf{X}} \right|_{(\hat{p}, \hat{\mathbf{X}})} + \mu'(\hat{p}) \cdot \sum_{n=1}^{\Omega} \sum_{k=0}^{T_n-1} \text{Tr} \left(\hat{\lambda}_k^n \left. \frac{\partial h_k^n}{\partial \mathbf{X}} \right|_{(\hat{p}, \hat{\mathbf{X}})} \right) \quad (5.11)$$

où $\hat{\lambda}_k^n$ est la composante pour la trajectoire n et l'instant k du multiplicateur global $\hat{\mathbf{\Lambda}}$.

Enfin, il est aisé de constater que les deux derniers termes de cette équation sont simplement proportionnels à $\left. \frac{\partial \mathcal{L}}{\partial \mathbf{X}} \right|_{(\hat{p}, \hat{\mathbf{X}}, \hat{\mathbf{\Lambda}})}$, de sorte qu'en conclusion :

$$\left. \frac{\partial \mathcal{L}}{\partial p} \right|_{(\hat{p}, \hat{\mathbf{X}}, \hat{\mathbf{\Lambda}})} = f'(\hat{p}) = 0 \quad (5.12)$$

La recherche d'un maximum local du Lagrangien défini par l'équation (5.6), noté $(\hat{p}, \hat{\mathbf{X}}, \hat{\mathbf{\Lambda}})$, permet donc d'obtenir un maximum local de la fonction f . En supposant que, sur l'ensemble des valeurs de paramètres admissibles, ce maximum local soit un maximum global, cette valeur $\hat{p}$ est tout simplement la valeur de l'estimation faite par Maximum de Vraisemblance :

$$\hat{p}_{ML} = \hat{p} \quad (5.13)$$

5.2.3 Calcul du gradient et états adjoints

Nous présentons ici la façon dont on réalise l'optimisation nous permettant de trouver un maximum local du Lagrangien. Le principe est le suivant : nous partons d'un $\hat{p}_0$ admissible. En utilisant ce $\hat{p}_0$, nous cherchons les valeurs $\hat{\mathbf{X}}_0$ et $\hat{\mathbf{\Lambda}}_0$ qui permettent de vérifier :

$$\left. \frac{\partial \mathcal{L}}{\partial \mathbf{X}} \right|_{(\hat{p}_0, \hat{\mathbf{X}}_0, \hat{\mathbf{\Lambda}}_0)} = 0, \quad \left. \frac{\partial \mathcal{L}}{\partial \mathbf{\Lambda}} \right|_{(\hat{p}_0, \hat{\mathbf{X}}_0, \hat{\mathbf{\Lambda}}_0)} = 0 \quad (5.14)$$

Or, nous avons montré dans la sous-section précédente que ces deux conditions engendraient nécessairement :

$$\left. \frac{\partial \mathcal{L}}{\partial p} \right|_{(\hat{p}_0, \hat{\mathbf{X}}_0, \hat{\Lambda}_0)} = f'(\hat{p}_0)$$

Ainsi, nous réalisons directement une optimisation sur la fonction f , puisque nous avons accès à la valeur de son gradient en un point donné. Nous notons $\hat{p}_j$ les valeurs de paramètres estimées que l'optimisation produit successivement. Nous présenterons dans la sous-section suivante la façon dont nous itérons $\hat{p}_j$ à partir de $\hat{p}_{j-1}$, et nous présentons ici la façon dont, pour une valeur de paramètre $\tilde{p}$ quelconque, les valeurs $\tilde{\mathbf{X}}$ et $\tilde{\Lambda}$ permettant au triplet $(\tilde{p}, \tilde{\mathbf{X}}, \tilde{\Lambda})$ de vérifier les conditions de l'équation (5.14) sont obtenues, ainsi que la formule de la dérivée de f en $\tilde{p}$, en exploitant la structure générale des applications de Kraus.

Dérivée en Λ du Lagrangien

Reprenons d'abord pour une valeur de paramètres $\tilde{p}$ la condition de nullité du gradient du Lagrangien par rapport à Λ . Soit Λ et $\mathbf{X}$ des valeurs quelconques des multiplicateurs de Lagrange (notés λ_k^n), et des états reconstruits (notés X_k^n). Pour une évolution infinitésimale $\delta\Lambda$ des multiplicateurs de Lagrange, le Lagrangien devient :

$$\mathcal{L}(\hat{p}_j, \mathbf{X}, \Lambda + \delta\Lambda) = \mathcal{L}(\hat{p}_j, \mathbf{X}, \Lambda) - \sum_{n=1}^{\Omega} \sum_{k=0}^{T_n-1} \text{Tr}(\delta\lambda_k^n h_k^n(\hat{p}_j, \mathbf{X})) \quad (5.15)$$

La nullité du gradient de $\mathcal{L}$ par rapport aux multiplicateurs de Lagrange, au point $(\tilde{p}, \tilde{\mathbf{X}}, \tilde{\Lambda})$, implique simplement que, pour toute trajectoire n et tout instant k entre 0 et $T_n - 1$, on ait :

$$h_k^n(\tilde{p}, \tilde{\mathbf{X}}) = 0,$$

ce qui implique que les états $\tilde{X}_k^n$ qui constituent $\tilde{\mathbf{X}}$ soient simplement les états à l'instant k de la trajectoire n , reconstruits avec la valeur de paramètres $\tilde{p}$. Pour conclure, on a, pour tout $n = 1, \dots, \Omega$:

$$\tilde{X}_0^n = \rho_0, \quad \text{et } \forall k = 1, \dots, T_n - 1, \quad \tilde{X}_k^n = \frac{\mathbf{K}_{\tilde{y}_k^{(n)}}^{\tilde{p}}(\tilde{X}_{k-1}^n)}{\text{Tr}\left(\mathbf{K}_{\tilde{y}_k^{(n)}}^{\tilde{p}}(\tilde{X}_{k-1}^n)\right)} \quad (5.16)$$

Dérivée en $\mathbf{X}$ du Lagrangien

Voyons désormais quelles équations nous pouvons tirer de la contrainte de nullité du gradient du Lagrangien par rapport à $\mathbf{X}$. En particulier, ces équations doivent nous permettre de calculer les multiplicateurs de Lagrange $\tilde{\lambda}_k^n$, formant le multiplicateur global $\tilde{\Lambda}$, à partir de $\tilde{p}$ et de $\tilde{\mathbf{X}}$. Lorsque nous passons de $\tilde{\mathbf{X}}$ à $\tilde{\mathbf{X}} + \delta\mathbf{X}$, l'évolution infinitésimale correspondante du Lagrangien a pour forme :

$$\begin{aligned} \mathcal{L}(\tilde{p}, \tilde{\mathbf{X}} + \delta\mathbf{X}, \Lambda) = & \sum_{n=1}^{\Omega} \left[\sum_{k=0}^{T_n-1} \log \left(\text{Tr} \left(\mathbf{K}_{\tilde{y}_{k+1}}^{\tilde{p}} \left(\tilde{X}_k^n + \delta X_k^n \right) \right) \right) - \right. \\ & \sum_{k=0}^{T_n-1} \text{Tr} \left(\lambda_k^n \left(\tilde{X}_k^n + \delta X_k^n \right) \right) + \text{Tr}(\lambda_0^n \rho_0) + \\ & \left. \sum_{k=0}^{T_n-2} \text{Tr} \left(\lambda_{k+1}^n \frac{\mathbf{K}_{\tilde{y}_{k+1}}^{\tilde{p}} \left(\tilde{X}_k^n + \delta X_k^n \right)}{\text{Tr} \left(\mathbf{K}_{\tilde{y}_{k+1}}^{\tilde{p}} \left(\tilde{X}_k^n + \delta X_k^n \right) \right)} \right) \right] \quad (5.17) \end{aligned}$$

Il est difficile, à partir de cette équation, d'extraire le gradient du Lagrangien par rapport à un état particulier δX_k^n . En effet, on peut réécrire l'équation (5.17), au premier ordre en $\delta\mathbf{X}$, comme :

$$\begin{aligned} \mathcal{L}(\tilde{p}, \tilde{\mathbf{X}} + \delta\mathbf{X}, \Lambda) = & \mathcal{L}(\tilde{p}, \tilde{\mathbf{X}}, \Lambda) + \\ & \sum_{n=1}^{\Omega} \left[\sum_{k=0}^{T_n-1} \frac{\text{Tr} \left(\mathbf{K}_{\tilde{y}_{k+1}}^{\tilde{p}} (\delta X_k^n) \right)}{\text{Tr} \left(\mathbf{K}_{\tilde{y}_{k+1}}^{\tilde{p}} \left(\tilde{X}_k^n \right) \right)} - \text{Tr}(\lambda_k^n \delta X_k^n) + \right. \\ & \left. \sum_{k=0}^{T_n-2} \text{Tr} \left(\lambda_{k+1}^n \frac{\mathbf{K}_{\tilde{y}_{k+1}}^{\tilde{p}} (\delta X_k^n) \text{Tr} \left(\mathbf{K}_{\tilde{y}_{k+1}}^{\tilde{p}} \left(\tilde{X}_k^n \right) \right) - \mathbf{K}_{\tilde{y}_{k+1}}^{\tilde{p}} \left(\tilde{X}_k^n \right) \text{Tr} \left(\mathbf{K}_{\tilde{y}_{k+1}}^{\tilde{p}} (\delta X_k^n) \right)}{\text{Tr}^2 \left(\mathbf{K}_{\tilde{y}_{k+1}}^{\tilde{p}} \left(\tilde{X}_k^n \right) \right)} \right) \right] \quad (5.18) \end{aligned}$$

Il nous faut "factoriser" chacun de ces termes en éléments δX_k^n , c'est-à-dire réécrire l'équation (5.18) comme :

$$\mathcal{L}(\tilde{p}, \tilde{\mathbf{X}} + \delta\mathbf{X}, \Lambda) = \mathcal{L}(\tilde{p}, \tilde{\mathbf{X}}, \Lambda) + \sum_{n=1}^{\Omega} \sum_{k=0}^{T_n-1} \text{Tr} \left(r_k^n(\tilde{p}, \tilde{\mathbf{X}}, \Lambda) \delta X_k^n \right) \quad (5.19)$$

en cherchant ensuite à déterminer les multiplicateurs de Lagrange $\tilde{\lambda}_k^n$ tels que $r_k^n(\tilde{p}, \tilde{\mathbf{X}}, \Lambda) = 0$ pour tout n et k admissibles. Pour trouver une telle forme, il nous faut utiliser le formalisme des applications de Kraus adjointes, déjà décrit en sous-section 3.2.3. Notons $\mathbf{K}_{\tilde{y}_{k+1}}^{\tilde{p},*}(\cdot)$ l'adjoint de l'application de Kraus partielle $\rho \mapsto \mathbf{K}_{\tilde{y}_{k+1}}^{\tilde{p}}(\rho)$. En utilisant la relation $\text{Tr}(A\mathbf{K}(B)) = \text{Tr}(\mathbf{K}^*(A)B)$ pour tous opérateurs A et B , et toute application de Kraus $\mathbf{K}$, on peut transformer l'équation (5.18) comme :

$$\begin{aligned} \mathcal{L}(\tilde{p}, \tilde{\mathbf{X}} + \delta\mathbf{X}, \Lambda) = & \mathcal{L}(\tilde{p}, \tilde{\mathbf{X}}, \Lambda) + \\ & \sum_{n=1}^{\Omega} \left[\sum_{k=0}^{T_n-1} \text{Tr} \left(\left[\frac{\mathbf{K}_{\tilde{y}_{k+1}}^{\tilde{p},*}(I_{\mathcal{H}})}{\text{Tr} \left(\mathbf{K}_{\tilde{y}_{k+1}}^{\tilde{p}} \left(\tilde{X}_k^n \right) \right)} - \lambda_k^n \right] \delta X_k^n \right) + \right. \\ & \left. \sum_{k=0}^{T_n-2} \text{Tr} \left(\left[\frac{\mathbf{K}_{\tilde{y}_{k+1}}^{\tilde{p},*}(\lambda_{k+1}^n)}{\text{Tr} \left(\mathbf{K}_{\tilde{y}_{k+1}}^{\tilde{p}} \left(\tilde{X}_k^n \right) \right)} - \frac{\text{Tr} \left(\lambda_{k+1}^n \mathbf{K}_{\tilde{y}_{k+1}}^{\tilde{p}} \left(\tilde{X}_k^n \right) \right) \mathbf{K}_{\tilde{y}_{k+1}}^{\tilde{p},*}(I_{\mathcal{H}})}{\text{Tr}^2 \left(\mathbf{K}_{\tilde{y}_{k+1}}^{\tilde{p}} \left(\tilde{X}_k^n \right) \right)} \right] \delta X_k^n \right) \right] \quad (5.20) \end{aligned}$$

où, on le rappelle, $I_{\mathcal{H}}$ est l'opérateur identité sur $\mathcal{H}$. La nullité du gradient par rapport à $\mathbf{\Lambda}$ du Lagrangien en $(\tilde{p}, \tilde{\mathbf{X}}, \tilde{\mathbf{\Lambda}})$ engendre finalement les relations suivantes, pour toute trajectoire n :

$$\tilde{\lambda}_{T_n-1}^n = \frac{\mathbf{K}_{\tilde{y}_{T_n}}^{\tilde{p},*}(I_{\mathcal{H}})}{\text{Tr}\left(\mathbf{K}_{\tilde{y}_{T_n}}^{\tilde{p}}(\tilde{X}_{T_n-1}^n)\right)}, \quad \text{et } \forall k = 0, \dots, T_n - 2, \quad (5.21)$$

$$\tilde{\lambda}_k^n = \frac{\mathbf{K}_{\tilde{y}_{k+1}}^{\tilde{p},*}(\tilde{\lambda}_{k+1}^n + I_{\mathcal{H}}) - \text{Tr}(\tilde{\lambda}_{k+1}^n \tilde{X}_{k+1}^n) \mathbf{K}_{\tilde{y}_{k+1}}^{\tilde{p},*}(I_{\mathcal{H}})}{\text{Tr}\left(\mathbf{K}_{\tilde{y}_{k+1}}^{\tilde{p}}(\tilde{X}_k^n)\right)} \quad (5.22)$$

On obtient donc les multiplicateurs de Lagrange correspondant à une trajectoire donnée par une récurrence inverse. Ces multiplicateurs sont des matrices Hermitiennes, comme les états reconstruits $\tilde{X}_k^n$, mais ne sont pas nécessairement des opérateurs positifs.¹

Formule du gradient

Pour calculer la valeur du gradient de la fonction de log-vraisemblance f en une valeur de paramètres $\tilde{p}$ quelconque, nous calculons tout d'abord la valeur de $\tilde{\mathbf{X}}$ correspondant à $\tilde{p}$ grâce à l'équation (5.16), puis celle des états adjoints $\tilde{\mathbf{\Lambda}}$ à partir des équations (5.21) et (5.22). Utilisant ensuite l'équation (5.12), on obtient finalement :

$$\begin{aligned} f'(\tilde{p}) = \frac{\partial \mathcal{L}}{\partial p} \Big|_{(\tilde{p}, \tilde{\mathbf{X}}, \tilde{\mathbf{\Lambda}})} &= \sum_{n=1}^{\Omega} \left[\sum_{k=1}^{T_n} \frac{\text{Tr}\left(\mathbf{D}\mathbf{K}_{\tilde{y}_k}^{\tilde{p}}(\tilde{X}_{k-1}^n)\right)}{\text{Tr}\left(\mathbf{K}_{\tilde{y}_k}^{\tilde{p}}(\tilde{X}_{k-1}^n)\right)} \right. \\ &\quad \left. + \sum_{k=1}^{T_n-1} \text{Tr} \left(\tilde{\lambda}_k^n \left(\frac{\mathbf{D}\mathbf{K}_{\tilde{y}_k}^{\tilde{p}}(\tilde{X}_{k-1}^n)}{\text{Tr}\left(\mathbf{K}_{\tilde{y}_k}^{\tilde{p}}(\tilde{X}_{k-1}^n)\right)} - \frac{\mathbf{K}_{\tilde{y}_k}^{\tilde{p}}(\tilde{X}_{k-1}^n) \text{Tr}\left(\mathbf{D}\mathbf{K}_{\tilde{y}_k}^{\tilde{p}}(\tilde{X}_{k-1}^n)\right)}{\text{Tr}^2\left(\mathbf{K}_{\tilde{y}_k}^{\tilde{p}}(\tilde{X}_{k-1}^n)\right)} \right) \right) \right] \quad (5.23) \end{aligned}$$

où $\mathbf{D}\mathbf{K}$ est l'application qui génère le gradient en p de l'application de Kraus K ; pour toute matrice densité ρ , on a :

$$\mathbf{D}\mathbf{K}_{\tilde{y}_k}^{\tilde{p}}(\rho) = \frac{d}{dp} \left(\mathbf{K}_{\tilde{y}_k}^{\tilde{p}}(\rho) \right) \Big|_{p=\tilde{p}}$$

Il existe également des techniques, de nature similaire au gradient par l'adjoint, qui permettent de calculer le Hessien d'une fonction, en un temps proportionnel au nombre de variables pris en compte dans le Hessien en question. Nous n'avons pas utilisé ces techniques, puisque l'algorithme d'optimisation que nous avons choisi génère une approximation du Hessien de f , mais on peut consulter par exemple [PG08] pour plus de détails.

1. L'éventualité d'une divergence de ces multiplicateurs lorsque k décroît vers 0 n'a pas été étudiée, mais, dans la validation expérimentale que nous présentons plus loin, les multiplicateurs n'ont pas divergé ni convergé vers 0 de manière significative.

5.3 Validations expérimentales : champ de photons et mesures discrètes

5.3.1 Configuration

Le but de cette section est de présenter une estimation de paramètres multiples réalisées sur des trajectoires de mesure acquises sur un système physique réel. En l'occurrence, ce système provient de l'expérience de boîte à photons présenté en annexe A.1. Ici, le nombre de photons thermiques est $n_{\text{th}} = 0.08$ photons, et les états ρ_k sont représentés comme des vecteurs. Nous disposons d'un nombre Ω de trajectoires de mesure, provenant des instants $t = 0$ à $t > 600$ ms, c'est-à-dire un temps presque dix fois supérieur au temps de demi-vie d'un photon dans la cavité. Par conséquent, nous n'utiliserons que les mesures établies dans des instants où l'état du système couvre le nombre de populations le plus large, pour maximiser l'information acquise sur la valeur des paramètres physiques du système. Ainsi, nous n'utilisons que les mesures acquises entre les instants $t = 0$ et $t = 2T_c = 130$ ms, ce qui correspond à 1512 résultats de mesure par trajectoire.

L'état initial du système un état cohérent de nombre moyen de photons $\langle \mathbf{N} \rangle_0 = 12.01$ photons. Puisque nous tronquons l'espace de Hilbert de sorte que le nombre maximal de photons pris en compte est $N_{\text{max}} = 15$, les populations $\rho_n(0)$ à l'instant initial valent :

$$\rho_n(0) = C_0 \frac{\langle \mathbf{N} \rangle_0^n}{n!},$$

C_0 étant une constante de normalisation, de sorte que $\sum_{n=0}^{N_{\text{max}}} \rho_n(0) = 1$.

5.3.2 Cadre de l'estimation

Ici, nous souhaitons donner une estimation par Maximum de Vraisemblance des matrices de mesure. Puisqu'il y a $N_R = 4$ valeurs de phase de Ramsey différentes, et $N_{\text{max}} + 1 = 16$ populations possibles, cela fait en théorie 64 paramètres à estimer de manière indépendante.

Rappelons que ces matrices de mesure sont toutes construites à partir de $\mathbb{P}(e | \phi_R, n)$ (puisque $\mathbb{P}(g | \phi_R, n) = 1 - \mathbb{P}(e | \phi_R, n)$), où ϕ_R est la phase de Ramsey, et que la physique de l'expérience nous apprend que :

$$\mathbb{P}(g | \phi_R, n) = \frac{1 + A + B(n) \sin(\phi(n) - \phi_R)}{2}$$

En fait, l'expression physique de cette probabilité prend en compte l'angle qu'acquiert l'atome durant sa trajectoire entre son émission et sa réception, qui est une variable stochastique notée ϵ , à travers la formule :

$$\mathbb{P}(g | \phi_R, n) = \frac{1 + A + B(n) \sin(\phi_\epsilon(n) - \phi_R)}{2} \quad (5.24)$$

On peut alors linéariser le sinus, et profiter de la linéarité apparaissant en physique quantique, nous permettant d'affirmer quand l'absence d'information sur la valeur de ϵ , $\mathbb{P}(g | \phi_R, n) = \int_\epsilon \mathbb{P}(g | \phi_R, n, \epsilon) dP(\epsilon)$. On obtient alors :

$$\begin{aligned} \mathbb{P}(g | \phi_R, n) &= \frac{1 + A + B[\langle \sin(\phi_\epsilon(n)) \rangle_\epsilon \cos(\phi_R) - \langle \cos(\phi_\epsilon(n)) \rangle_\epsilon \sin(\phi_R)]}{2} \\ &= \frac{1 + A + [B_s(n) \cos(\phi_R) - B_c(n) \sin(\phi_R)]}{2} \end{aligned} \quad (5.25)$$

Avec cette formulation, nous voyons qu'il ne reste que $1 + 2(N_{\max} + 1)$ paramètres à estimer. En supposant que la valeur de A est déjà connue avec une précision acceptable, nous nous donnons comme objectif de réaliser l'estimation de $p = (B_s(n), B_c(n))_{n=0\dots 15}$.

5.3.3 Résultats

L'optimisation est réalisée à partir de $\hat{p}_0$, construit simplement en prenant $B_s(n) = B_c(n) = 0$ pour tout n . Nous disposons d'une première estimation réalisée par l'équipe d'Igor Dotsenko ayant réalisé l'expérience, notée $\hat{p}_{\text{ini}}$.

Ainsi, nous avons commencé par utiliser un algorithme simple, en prenant comme direction de montée le gradient de la fonction de log-vraisemblance, en utilisant un pas sélectionné par recherche linéaire. Puis, lorsque nous avons atteint une estimation dont la log-vraisemblance était meilleure que celle de $\hat{p}_{\text{ini}}$, nous avons commencé à utiliser la méthode de Broyden-Fletcher-Goldfarb-Shanno [Avr03], qui se base sur une estimation itérée B_j du Hessien de la fonction à maximiser, pour donner comme direction de maximisation :

$$\hat{p}_{j+1} = \hat{p}_j - \alpha_j B_j^{-1} \nabla f(\hat{p}_j), \quad (5.26)$$

où α_j est obtenu par recherche linéaire. Cette technique n'est utilisée qu'à partir d'un certain moment dans l'optimisation, car le Hessien au point courant $\hat{p}_j$ doit être assez proche du Hessien au point maximal pour que la méthode du BFGS fonctionne.

Les conditions d'arrêt de la maximisation sont simplement la nullité du gradient de f , puisqu'a priori, toutes les valeurs de $B_s(n)$ et $B_c(n)$ sont autorisées entre -1 et 1 , et que ces valeurs n'ont jamais été atteintes. Nous donnons alors une estimation de l'incertitude sur chacune de ces valeurs, grâce à l'inverse du Hessien calculé pendant la maximisation. Ces résultats sont donnés en figure 5.1.

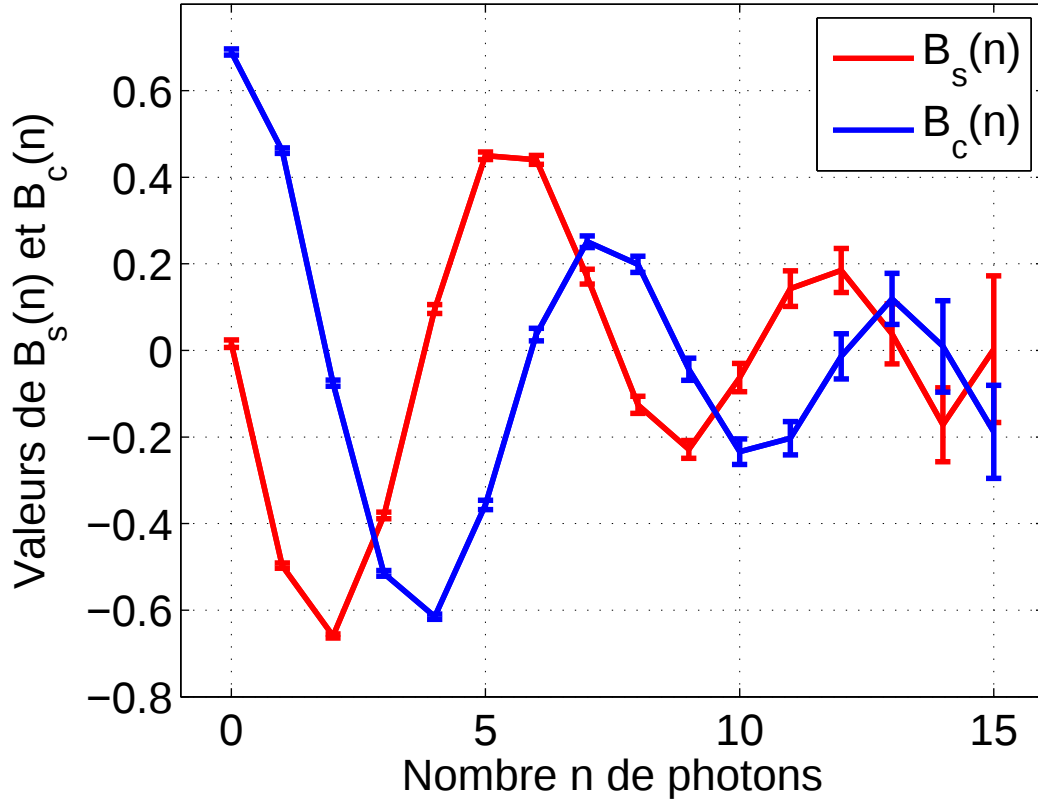


FIGURE 5.1 – Résultat de l'estimation par Maximum de Vraisemblance des paramètres $B_s(n)$ (en rouge) et $B_c(n)$ (en bleu). On retrouve bien le caractère d'oscillation atténuée que nous attendions, puisque ces coefficients sont l'intégrale d'une sinusoïde sur un petit intervalle de valeurs, et que les mesures réalisées par un atome sur le système sont moins discriminantes pour les grands nombres de photons. Les lignes correspondent aux valeurs pour différentes valeurs n , et les intervalles indiquent l'intervalle de confiance à 95 %, c'est-à-dire $B_s(n) \pm 2\sqrt{\sigma_{ML}^2(E_s(n))}$ (respectivement $B_c(n) \pm 2\sqrt{\sigma_{ML}^2(E_c(n))}$, $E_s(n)$ et $E_c(n)$ étant les opérateurs permettant de mesurer $B_s(n)$ ou $B_c(n)$ dans une valeur de paramètres p .

Chapitre 6

Conclusion

Dans cette thèse, après avoir rappelé le cadre général de l'évolution déterministe d'un système quantique, nous avons présenté les différentes caractéristiques des mesures discrètes et diffusives : les lois de probabilité générant ces mesures, et l'évolution induite de l'état du système. Les systèmes diffusifs étant régis par une équation maîtresse stochastique sous forme différentielle, son utilisation requiert l'existence d'un schéma numérique capable d'être implémenté informatiquement. Le premier apport de ce thèse a donc été de présenter un schéma de discrétisation, capable de fournir, pour une période d'échantillonnage $\Delta t > 0$ donnée et pour tout instant t , une loi de probabilité sur la valeur des signaux de mesures intégrés entre t et $t + \Delta t$, ainsi que l'évolution de l'état du système, causée par cette mesure, conditionnellement à la valeur mesurée. Nous avons montré de plus que le schéma de discrétisation proposé converge en loi vers le schéma d'évolution décrit par les équations différentielles diffusives standard.

Dans le deuxième chapitre, nous avons étudié le problème de l'estimation de la valeur $\bar{\rho}$ d'un état quantique initial, par Maximum de Vraisemblance sur l'ensemble des trajectoires de mesure $(Y^{(n)})_{n=1,\dots,\Omega}$ acquises sur un système après la préparation de $\bar{\rho}$, ce qui est appelé tomographie d'état quantique. Nous avons commencé par nous intéresser aux systèmes à mesures discrètes, puis avons introduit le formalisme des états quantiques adjoints pour rendre la fonction de log-vraisemblance très facile à calculer. Nous avons ensuite donné les conditions d'arrêt de la maximisation de la fonction de log-vraisemblance. Nous avons montré que le schéma de discrétisation pouvait être utilisé pour réaliser cette tomographie d'état sur des systèmes à mesures diffusives. Après avoir rappelé que l'estimation par Maximum de Vraisemblance tend à être non biaisée (consistance) et à variance minimale lorsque le nombre de mesures prises en compte tend vers l'infini¹, nous avons introduit la notion d'information de Fisher pour déterminer la valeur de cette variance (qui est aussi une estimation de l'erreur $\hat{\rho}_{ML} - \bar{\rho}$), lorsque l'état estimé $\hat{\rho}_{ML}$ est un point intérieur de l'espace des points acceptables. En l'occurrence, cette condition correspond à avoir une matrice densité estimée inversible, donc de rang plein. Lorsque la matrice densité estimée par le Maximum de Vraisemblance n'est pas inversible, nous avons étudié la valeur de l'estimation Bayésienne $\hat{\rho}_{BM}$ de $\bar{\rho}$ et de la variance Bayésienne, ayant recours au développement asymptotique autour de $\hat{\rho}_{ML}$ des intégrales de Laplace apparaissant dans les formules de celles-ci, pour donner la valeur de la variance de l'estimation par Maximum de Vraisemblance, que $\hat{\rho}_{ML}$ soit de rang plein ou non. Enfin, nous avons appliqué

1. Nous parlons ici de la moyenne et de la variance de la variable $Y = (Y^{(n)})_{n=1,\dots,\Omega} \mapsto \hat{\rho}_{ML}(Y)$, où Y , étant générée à partir de $\bar{\rho}$, dépend de lui. Le nombre de mesures tend vers l'infini quand Ω tend vers l'infini.

la tomographie d'état quantique à deux types de systèmes : un qubit supraconducteur dont la décohérence est mesurée de manière diffuse, et un champ électromagnétique mesuré projectivement par chaque atome traversant la cavité dans laquelle ce champ est enfermé.

Dans le troisième chapitre, nous avons introduit la notion d'estimation par Maximum de Vraisemblance pour l'estimation de la valeur réelle $\bar{p}$ de certains paramètres physiques intervenant dans le processus de mesure des systèmes quantiques. Notre étude théorique a, là aussi, commencé par le cadre des mesures discrètes. Nous avons introduit le formalisme des filtres particuliers pour montrer que, lorsqu'on teste plusieurs valeurs de paramètres dont $\bar{p}$, la probabilité que le filtre particulière quantique attribue à $\bar{p}$ est une sous-martingale². Ce résultat suppose de connaître la valeur de l'état initial du système physique, mais nous donnons également un résultat de robustesse du filtre particulière lorsque cet état initial n'est pas connu. Enfin, ces filtres particuliers peuvent être implémentés pour des systèmes diffusifs, grâce au schéma de discrétisation. C'est ce qui est fait en annexe, pour l'estimation de l'efficacité de détection de la mesure hétérodyne de décohérence d'un qubit supraconducteur, puis pour son temps de relaxation. A cette occasion, nous voyons que le choix du schéma de discrétisation peut avoir une importance sur l'estimation de certains paramètres.

Enfin, dans le dernier chapitre, nous nous sommes attachés à l'estimation de paramètres évoluant dans un espace de dimension multiple. Dans ce cas, l'utilisation incrémentale des filtres particuliers est trop longue à mettre en place, puisque l'ensemble des paramètres à tester augmente exponentiellement avec la dimension de l'espace des valeurs de paramètres admissibles. La fonction de log-vraisemblance faisant intervenir la valeur des états successifs du système, il nous a fallu introduire la technique du gradient par l'adjoint pour réaliser la maximisation de cette fonction. Nous présentons l'implémentation de cette technique courante dans le cas particulier de la tomographie de processus quantique. Enfin, nous avons utilisé ces calculs pour réaliser l'estimation par Maximum de Vraisemblance des coefficients intervenant dans les matrices de mesure du champ électromagnétique par les atomes de Rydberg, au sein d'une expérience de type "boîte à photons", à partir de vraies trajectoires de mesure.

Au moment de clore cette thèse, nous voyons deux principales pistes de développement au travail qui vient d'être réalisé. Tout d'abord, nous pensons que le fait d'utiliser les développements asymptotiques d'intégrales de Laplace multi-dimensionnelles pourraient avoir un intérêt pour le *compress sensing* quantique [GLF⁺10] et aussi pour la prise en compte de parcimonie souvent observée pour les états quantiques [GKD12], lorsque la dimension de l'espace de Hilbert sous-jacent est a priori bien plus grande que le rang de la matrice densité estimée ; ce qui a été le cas ici, lorsque nous avons estimé l'état initial du système "boîte à photons" par Maximum de Vraisemblance. Le second développement que nous voyons, et sur lequel nous avons commencé à nous pencher avant le début de la rédaction de cette thèse, est l'adaptation des techniques de tomographie de processus quantique exposées ici au cas où la valeur des paramètres recherchés varie lentement le long d'une seule et longue trajectoire de mesures quantiques.

2. A tout instant k , cette probabilité augmente en espérance de k à $k + 1$, l'espérance étant calculée sur la valeur de la mesure entre k et $k + 1$.

Bibliographie

- [AGZV85] V.I. Arnold, S.M. Gusein-Zade, and A.N. Varchenko. *Singularities of Differentiable Maps, Vol. II*. Birkhäuser, Boston, 1985.
- [APR14] H. Amini, C. Pellegrini, and P. Rouchon. Stability of continuous-time quantum filters with measurement imperfections. *Russian Journal of Mathematical Physics*, 21(3) :297–315–, 2014.
- [Avr03] M. Avriel. *Nonlinear Programming : Analysis and Methods*. Dover Publishing, 2003.
- [BD11] Escher B.M., R. L. Matos Filho , and L. Davidovich. General framework for estimating the ultimate precision limit in noisy quantum-enhanced metrology. *Nature Physics* 7,, 7(406), 2011.
- [Bel92] V.P. Belavkin. Quantum stochastic calculus and quantum nonlinear filtering. *Journal of Multivariate Analysis*, 42(2) :171–201, 1992.
- [BG09] A. Barchielli and M. Gregoratti. *Quantum Trajectories and Measurements in Continuous Time : the Diffusive Case*. Springer Verlag, 2009.
- [BH86] N. Bleistein and R.A. Handelsman. *Asymptotic Expansions of Integrals*. Dover, 1986.
- [BHIW86] J. Bergquist, R. Hulet, W. Itano, and D. Wineland. Observation of quantum jumps in a single atom. *Physical Review Letters*, 57 :1699–1702, 1986.
- [BK10] Robin Blume-Kohout. Optimal, reliable estimation of quantum states. *New Journal of Physics*, 12(4) :043034–, 2010.
- [BKB95] T. Basché, S. Kummer, and C. Bräuchle. Direct spectroscopic observation of quantum jumps of a single molecule. *Nature*, 373 :132–134, 1995.
- [Buz04] V. Buzek. *Lecture Notes in Physics*, volume 649, chapter Quantum tomography from incomplete data via MaxEnt principle, pages 180–234. Springer-Verlag Berlin, 2004.
- [BV09] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 2009.
- [BvHJ07] L. Bouten, R. van Handel, and M. R. James. An introduction to quantum filtering. *SIAM J. Control Optim.*, 46 :2199–2241, 2007.
- [BWT⁺12] S. Brakhane, Alt W., T.Kampschulte, M. Martinez-Dorantes, R. Reimann, S. Yoon, A. Widera, and D. Meschede. Bayesian feedback control of a two-atom spin-state in an atom-cavity system. *Phys. Rev. Lett.*, 109(17) :173601–, October 2012.
- [Car93] H. Carmichael. *An Open Systems Approach to Quantum Optics*. Springer-Verlag, 1993.
- [CG09] Bradley A. Chase and J. M. Geremia. Single-shot parameter estimation via continuous quantum measurement. *Phys. Rev. A*, 79(2) :022314–, February 2009.
- [CI15] Philippe Campagne-Ibarcq. *Measurement back-action and feedback in superconducting circuits*. PhD thesis, Ecole Normale Supérieure, 2015.
- [CIBF⁺14] P. Campagne-Ibarcq, L. Bretheau, E. Flurin, A. Auffèves, F. Mallet, and B. Huard. Observing interferences between past and future quantum states in resonance fluorescence. *Phys. Rev. Lett.*, 112 :180402, May 2014.

- [CIJC⁺16] P. Campagne-Ibarcq, S. Jezouin, N. Cottet, P. Six, L. Bretheau, F. Mallet, A. Sarlette, P. Rouchon, and B. Huard. Using spontaneous emission of a qubit as a resource for feedback control. *Phys. Rev. Lett.*, 117(6) :060502, August 2016.
- [CISB⁺16] P. Campagne-Ibarcq, P. Six, L. Bretheau, A. Sarlette, M. Mirrahimi, P. Rouchon, and B. Huard. Observing quantum state diffusion by heterodyne detection of fluorescence. *Phys. Rev. X*, 6 :011002, 2016.
- [CMR05] O. Cappé, E. Moulines, and T. Ryden. *Inference in Hidden Markov Models*. Springer series in statistics, 2005.
- [CTDL77] C. Cohen-Tannoudji, B. Diu, and F. Laloë. *Mécanique Quantique*, volume I& II. Hermann, Paris, 1977.
- [DDKG12] R. Demkowicz-Dobrzanski, J. Kolodynski, and M. Guta. The elusive heisenberg limit in quantum-enhanced metrology. *Nature Communications*, 3 :1063–, 2012.
- [dLRT⁺14] G. de Lange, D. Ristè, M. J. Tiggelman, C. Eichler, L. Tornberg, G. Johansson, A. Wallraff, R. N. Schouten, and L. DiCarlo. Reversing quantum trajectories with analog feedback. *Phys. Rev. Lett.*, 112(8) :080501–, February 2014.
- [DMB⁺09] I. Dotsenko, M. Mirrahimi, M. Brune, S. Haroche, J.-M. Raimond, and P. Rouchon. Quantum feedback by discrete quantum non-demolition measurements : towards on-demand generation of photon-number states. *Physical Review A*, 80 : 013805-013813, 2009.
- [DMOH11] R. Douc, E. Moulines, J. Olsson, and R. Van Handel. Consistency of the maximum likelihood estimator for general hidden markov models. *The Annals of Statistics*, 39(1) :474–513, 2011.
- [GJM13] S. Gammelmark, B. Julsgaard, and K. Mølmer. Past quantum states of a monitored system. *Phys. Rev. Lett.*, 111(16) :160401–, October 2013.
- [GJNC12] John E. Gough, Matthew R. James, Hendra I. Nurdin, and Joshua Combes. Quantum filtering for systems driven by fields in single-photon states or superposition of coherent states. *Phys. Rev. A*, 86(4) :043819–, October 2012.
- [GKD12] M. Guta, T. Kypraios, and I. Dryden. Rank based model selection for multiple ions quantum tomography. *New Journal of Physics*, 14 :105002–, 2012.
- [GKG⁺07] S. Gleyzes, S. Kuhr, C. Guerlin, J. Bernu, S. Deléglise, U. Busk Hoff, M. Brune, J.-M. Raimond, and S. Haroche. Quantum jumps of light recording the birth and death of a photon in a cavity. *Nature*, 446 :297–300, 2007.
- [GLF⁺10] David Gross, Yi-Kai Liu, Steven T. Flammia, Stephen Becker, and Jens Eisert. Quantum state tomography via compressed sensing. *Phys. Rev. Lett.*, 105(15) :150401–, October 2010.
- [GP92] N. Gisin and I. C. Percival. The quantum-state diffusion model applied to open quantum systems. *Journal of Physics A : Mathematical and General*, 25 :5677, 1992.
- [GW01] J. Gambetta and H. M. Wiseman. State and dynamical parameter estimation for open quantum systems. *Phys. Rev. A*, 64(4) :042105–, September 2001.
- [GW15] I. Guevara and H. Wiseman. Quantum state smoothing. *Phys. Rev. Lett.*, 115 :180407, Oct 2015.

- [Hares] Serge Haroche. Cours au college de france. 2009-2010 : Synthese et Reconstructions d'etats quantiques.
- [HR06] S. Haroche and J.M. Raimond. *Exploring the Quantum : Atoms, Cavities and Photons*. Oxford University Press, 2006.
- [HSM⁺13] M. Hatridge, S. Shankar, M. Mirrahimi, F. Schackert, K. Geerlings, T. Brecht, K. M. Sliwa, B. Abdo, L. Frunzio, S. M. Girvin, R. J. Schoelkopf, and M. H. Devoret. Quantum back-action of an individual variable-strength measurement. *Science*, 339(6116) :178–181, 2013.
- [JCRH15] A. N. Jordan, A. Chantasri, P. Rouchon, and B. Huard. Anatomy of fluorescence : Quantum trajectory statistics from continuously measuring spontaneous emission. *arXiv :1511.06677*, 2015.
- [KGG⁺07] S. Kuhr, S. Gleyzes, C. Guerlin, J. Bernu, U. B. Hoff, S. Deléglise, S. Osnaghi, M. Brune, J.-M. Raimond, S. Haroche, E. Jacques, P. Bosland, and B. Visentin. Ultrahigh finesse fabry-pérot superconducting resonator. *Appl. Phys. Lett.*, 90 :164101, 2007.
- [KY13] Y. Kato and N. Yamamoto. Estimation and initialization of quantum network via continuous measurement on single node. In *Decision and Control (CDC), 2013 IEEE 52nd Annual Conference on*, pages 1904–1909, 2013.
- [KYG⁺07] J. Koch, T. M. Yu, J. Gambetta, A. A. Houck, D. I. Schuster, J. Majer, A. Blais, M. H. Devoret, S. M. Girvin, and R. J. Schoelkopf. Introducing the Transmon : a new superconducting qubit from optimizing the Cooper Pair Box. *Physical Review A*, 76 :1–19, February 2007.
- [LC98] E. L. Lehmann and G. Castella. *Theory of Point Estimation*. Springer Texts in Statistics, 1998.
- [LR09] A. I. Lvovsky and M. G. Raymer. Continuous-variable optical quantum-state tomography. *Rev. Mod. Phys.*, 81(1) :299–332, March 2009.
- [Mab96] H Mabuchi. Dynamical identification of open quantum systems. *Quantum and Semiclassical Optics : Journal of the European Optical Society Part B*, 8(6) :1103–, 1996.
- [Meh04] M. L. Mehta. *Random matrices (3rd ed)*. Elsevier, Academic Press, 2004.
- [Mil63] J. Milnor. *Morse Theory*. Princeton University Press, 1963.
- [MWMS13] K. W. Murch, S. J. Weber, C. Macklin, and I. Siddiqi. Observing single quantum trajectories of a superconducting quantum bit. *Nature*, 502 :211–214, October 2013.
- [NBS⁺10] P. Neumann, J. Beck, M. Steiner, F. Rempp, H. Fedder, P.R. Hemmer, J. Wrachtrup, and F. Jelezko. Single-shot readout of a single nuclear spin. *Science*, 5991 :542, 2010.
- [NC00] M.A. Nielsen and I.L. Chuang. *Quantum Computation and Quantum Information*. Cambridge University Press, 2000.
- [NM13] A Negretti and K Mølmer. Estimation of classical parameters via continuous probing of complementary quantum observables. *New Journal of Physics*, 15(12) :125002–, 2013.
- [NSD86] W. Nagourney, J. Sandberg, and H. Dehmelt. Shelved optical electron amplifier : Observation of quantum jumps. *Physical Review Letters*, 56 :2797–2799, 1986.

- [OJTBG07] Alexei Ourjoumtsev, Hyunseok Jeong, Rosa Tualle-Brouri, and Philippe Grangier. Generation of optical /‘schrodinger cats/’ from photon number states. *Nature*, 448(7155) :784–786, August 2007.
- [Pel08] C. Pellegrini. Markov chains approximation of jump-diffusion stochastic master equations. *arXiv :0803.2593*, 2008.
- [PG99] S. Peil and G. Gabrielse. Observing the quantum limit of an electron cyclotron : Qnd measurements of quantum jumps between fock states. *Physical Review Letters*, 83 :1287–1290, 1999.
- [PG08] D. I. Papadimitriou and Giannakoglou. Direct, adjoint and mixed approaches for the computation of hessian in airfoil design problems. *Int.. J. Numer. Meth. Fluids*, 56 :1929–1943, 2008.
- [PR04] M.G.A Paris and J. Rehacek. *Quantum State Estimation*. Springer, 2004.
- [PSB⁺11] H. Paik, D. I. Schuster, L. S. Bishop, G. Kirchmair, G. Catelani, A. P. Sears, B. R. Johnson, M. J. Reagor, L. Frunzio, L. I. Glazman, S. M. Girvin, M. H. Devoret, and R. J. Schoelkopf. Observation of High Coherence in Josephson Junction Qubits Measured in a Three-Dimensional Circuit QED Architecture. *Physical Review Letters*, 107(24) :240501, December 2011.
- [Pur46] E.M. Purcell. Proceedings of the american physical society. *Physical Review*, 69 :674, 1946.
- [RGP⁺14] Th Rybarczyk, S. Gerlich, B. Peaudecerf, M. Penasa, B. Julsgaard, K. Moelmer, S. Gleyzes, M. Brune, JM Raimond, S. Haroche, and I. Dotsenko. Past quantum state analysis of the photon number evolution in a cavity. *arXiv :1409.0958 [quant-ph]*, 2014.
- [Rou11] P. Rouchon. Fidelity is a sub-martingale for discrete-time quantum filters. *IEEE Transactions on Automatic Control*, 56(11) :2743–2747, 2011.
- [RPP⁺15] T. Rybarczyk, B. Peaudecerf, M. Penasa, S. Gerlich, B. Julsgaard, K. Mølmer, S. Gleyzes, M. Brune, J. M. Raimond, S. Haroche, and I. Dotsenko. Forward-backward analysis of the photon-number evolution in a cavity. *Phys. Rev. A*, 91 :062116, Jun 2015.
- [RR15] P. Rouchon and J. F. Ralph. Efficient quantum filtering for quantum feedback control. *Phys. Rev. A*, 91(1) :012118–, January 2015.
- [SCIB⁺15] P. Six, Ph. Campagne-Ibarcq, L. Bretheau, B. Huard, and P. Rouchon. Parameter estimation from measurements along quantum trajectories. *CDC 2015*, *arXiv :1503.06149 [math.OC]*, 2015.
- [SCID⁺16] P. Six, Ph. Campagne-Ibarcq, I. Dotsenko, A. Sarlette, B. Huard, and P. Rouchon. Quantum state tomography with non-instantaneous measurements, imperfections and decoherence. *Phys. Rev. A*, 93 :012109, January 2016.
- [SDSR12] A. Somaraju, I. Dotsenko, C. Sayrin, and P. Rouchon. Design and stability of discrete-time quantum filters with measurement imperfections. In *American Control Conference*, pages 5084–, 2012.
- [Sha11] Lin Shaowei. *Algebraic Methods for Evaluating Integrals in Bayesian Statistics*. PhD thesis, University of California, Berkeley, 2011.
- [SNBT86] T. Sauter, W. Neuhauser, R. Blatt, and P. Toschek. Observation of quantum jumps. *Physical Review Letters*, 57 :1696–1698, 1986.

- [SPL⁺14] L. Sun, A. Petrenko, Z. Leghtas, B. Vlastakis, G. Kirchmair, K. M. Sliwa, A. Narla, M. Hatridge, S. Shankar, J. Blumoff, L. Frunzio, M. Mirrahimi, M. H. Devoret, and R. J. Schoelkopf. Tracking photon jumps with repeated quantum non-demolition parity measurements. *Nature*, 511 :444–448, July 2014.
- [SR16] P. Six and P. Rouchon. Asymptotic expansions of laplace integrals for quantum state tomography. *arXiv :1607.00948 [quant-ph]*, 2016.
- [Tsa09] Mankei Tsang. Time-symmetric quantum theory of smoothing. *Phys. Rev. Lett.*, 102(25) :250403–, June 2009.
- [vdV98] A. W. van der Vaart. *Asymptotic Statistics*. Cambridge Series in Statistical and Probabilistic Mathematics, 1998.
- [VLM⁺10] A.N. Vamivakas, C.Y. Lu, C. Matthiesen, Y. Zhao, S. Falt, A. Badolato, and M. Atature. Observation of spin-dependent quantum jumps via quantum dot resonance fluorescence. *Nature*, 7313 :297, 2010.
- [VPS⁺14] U. Vool, I.M. Pop, K. Sliwa, B. Abdo, C. Wang, T. Brecht, Y.Y. Gao, S. Shankar, M. Hatridge, G. Catelani, M. Mirrahimi, L. Frunzio, R.J. Schoelkopf, L.I. Glazman, and M.H. Devoret. Non-poissonian quantum jumps of a fluxonium qubit due to quasiparticle excitations. *Physical Review Letters*, 113 :247001, 2014.
- [VSS11] R. Vijay, D. H. Slichter, and I. Siddiqi. Observation of Quantum Jumps in a Superconducting Artificial Atom. *Physical Review Letters*, 106(11) :110502, March 2011.
- [Wat09] S. Watanabe. *Algebraic Geometry and Statistical Learning Theory*. Cambridge University Press, 2009.
- [WCD⁺14] S. J. Weber, A. Chantasri, J. Dressel, A. N. Jordan, K. W. Murch, and I. Siddiqi. Mapping the optimal route between two quantum states. *Nature*, 511(7511) :570–573, July 2014.
- [WM93] H.M. Wiseman and G.J. Milburn. Interpretation of quantum jump and diffusion processes illustrated on the bloch sphere. *Physical Review A*, 47 :1652–66, 1993.
- [WM09] H.M. Wiseman and G.J. Milburn. *Quantum Measurement and Control*. Cambridge University Press, 2009.
- [WMS⁺15] S. J. Weber, K. W. Murch, M. E. Schwartz, N. Roch, and I. Siddiqi. Quantum trajectories of superconducting qubits. *ArXiv e-prints*, June 2015.
- [YSLG⁺08] Y. Yang, Z. Shi-Liang, S. Guozhu, W. Xueda, D. Ning, C. Jian, W. Peiheng, and H. Siyuan. Quantum jumps between macroscopic quantum states of a superconducting qubit coupled to a microscopic two-level system. *Physical Review Letters*, 101 :157001, 2008.

Annexe A

Systèmes expérimentaux utilisés

A.1 Boîte à photons

Le premier dispositif expérimental dont nous avons exploité les données a été développé au laboratoire Kastler Brossel, d’abord à l’École Normale Supérieure, puis au Collège de France. Le système quantique dont il est question dans cette expérience est un champ électromagnétique piégé dans un mode d’une cavité micro-ondes supraconductrice. Un des papiers [SCID⁺16] que nous avons écrits durant cette thèse voit ses résultats illustrés par des trajectoires de mesure provenant de cette expérience. Cette expérience est présentée plus en détail dans l’ouvrage de Serge Haroche et Jean-Michel Raimond [HR06], et les explications ci-dessous sont largement reprises des papiers [RGP⁺14, RPP⁺15].

A.1.1 Dispositif expérimental

Un schéma du dispositif expérimental, provenant de [RGP⁺14], est donné en figure A.1. Le système à proprement parler est un champ micro-ondes piégé dans une cavité supra-conductrice à haute qualité, dont la fréquence de résonance est de $\omega_c/2\pi = 51$ GHz [KGG⁺07] (lettre **C** sur le schéma). La cavité est refroidie à 0.8 K, et le temps de demi-vie des photons à l’intérieur de la cavité est $T_c = 65$ ms. La mesure du champ est réalisée par un flux continu d’atomes de Rydberg, préparés a priori (lettre **B** sur le schéma) à partir d’un bain d’atomes de Rubidium évoluant à une vitesse de 250 m/s. La fréquence de transition atomique entre les états circulaires de nombres quantiques principaux 50 (état $|g\rangle$) et 51 (état $|e\rangle$) est proche de ω_c . Des atomes sortis du bain traversent la cavité à chaque intervalle de temps $T_a = 86 \mu\text{s}$, même si, au cours d’un même intervalle, des cas de non-détection, d’atomes absents ou multiples sont possibles.

Après interaction avec le système-champ, l’atome voit son état mesuré projectivement par le détecteur, en **D** sur le schéma. En moyenne, environ 0.30 atomes sont mesurés par intervalle de temps.

A.1.2 Cadre mathématique

Ce système étant mesuré de manière discrète, aux instants t_k , le temps sera parfois simplement indexé par k , l’intervalle entre deux instant successifs étant $t_{k+1} - t_k = T_c = 65$ ms. Au cours de cet intervalle, deux phénomènes ont lieu. Tout d’abord, le système interagit avec son environnement, ce qui provoque une génération et une destruction spontanées de photons, représentées par l’opérateur L .

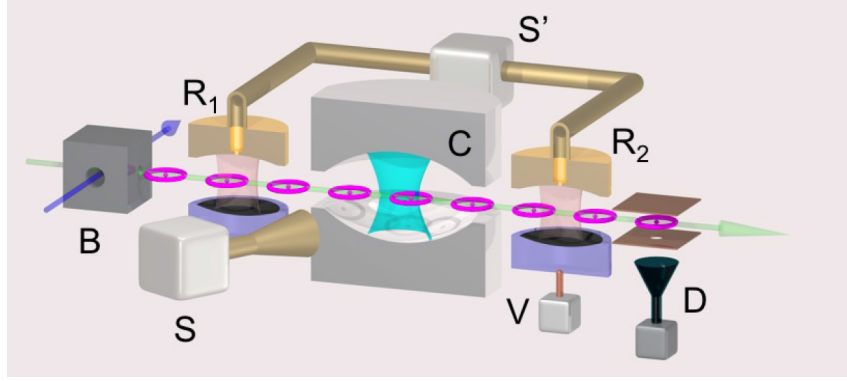


FIGURE A.1 – Schéma du dispositif expérimental "boîte à photons". La cavité optique **C** est alimentée par une source micro-ondes **S**. Un flux d'atomes de Rydberg circulaires, excités en **B**, traverse le mode de la cavité et est mesuré par **D**. L'interféromètre de Ramsey, constitué des cavités large-bande **R₁** et **R₂**, permet à la mesure de discriminer alternativement les populations de photons de la cavité.

Ensuite, l'interaction dispersive d'un atome avec n photons dans la cavité introduit un déphasage $\phi(n) \approx \phi_0(n + 1/2)$ entre $|e\rangle$ et $|g\rangle$ dans l'état de l'atome. ϕ_0 est le déphasage par photon [HR06]. L'information sur $\phi(n)$, et donc sur le nombre de photons dans la cavité, est lue grâce à un interféromètre de Ramsey, généré par deux cavités à large bande (**R₁** et **R₂** sur le schéma) de part et d'autre du système. Ces cavités appliquent une pulsation Rabi d'angle $\pi/2$ entre les états $|g\rangle$ et $|e\rangle$. Finalement, étant donné un nombre n de photons dans le champ, la probabilité conditionnelle de détecter l'atome dans l'état g ou e est donnée par :

$$\mathbb{P}(g | \phi, n) = \frac{1 + A + B(n) \sin(\phi(n) - \phi_0/2 - \phi)}{2}$$

$$\mathbb{P}(e | \phi, n) = \frac{1 - A - B(n) \sin(\phi(n) - \phi_0/2 - \phi)}{2}$$

où ϕ est la phase de l'interféromètre de Ramsey. La valeur non nulle de $A = 0.03$ est due aux imperfections de mesure, tout comme le fait que B dépende de n . La suite des $B(n)$ est décroissante en n , avec $B(0) = 0.71$. La valeur de ϕ_0 est déterminée avant l'enregistrement des mesures à $\phi_0 \approx \pi/4$, en réglant le *detuning* entre l'atome et la cavité. Enfin, pour faire en sorte que tous les nombres de photons puissent avoir une influence sur la génération de la trajectoire de mesures, la valeur de ϕ est réglée pour prendre successivement quatre valeurs, la valeur changeant à chaque intervalle de temps T_a . Ces valeurs prennent les valeurs de $\phi_1 = 0$, $\phi_2 = 0.26\pi$, $\phi_3 = 0.519\pi$, et $\phi_4 = 0.773\pi$.

Ici, on considère que le nombre de photons est mesuré assez fortement pour ne pas prendre en compte les cohérences dans la représentation de l'état du système. Ainsi, cet état, à l'instant t_k , est simplement écrit ρ_k , où $\rho_k = (\rho_k(0), \dots, \rho_k(N_{\max}))$ est un vecteur représentant les différentes probabilités de population à l'instant t_k . On l'appellera vecteur densité dans le sens où toutes ses composantes sont positives ou nulles, et où la somme de ces composantes vaut 1.

Pour itérer l'état entre les instants t_k et t_{k+1} , notons d'abord $P_a(m)$ la probabilité de mesurer m atomes pendant un intervalle de temps : on considère que $P_a(m) = 0$ si $m \geq 3$. Après mesure à l'instant t_k de y_k^e atomes dans l'état $|e\rangle$ et y_k^g dans l'état $|g\rangle$, l'état devient

ρ_k^+ à l'instant t_k^+ , avec :

$$\begin{aligned} \rho_k^+(n) &= \frac{P_a(y_k^e + y_k^g) (\mathbb{P}(e | \phi_k, n))^{y_k^e} (\mathbb{P}(g | \phi_k, n))^{y_k^g} \rho_k(n)}{\sum_{n'=0}^{N_{\max}} P_a(y_k^e + y_k^g) (\mathbb{P}(e | \phi_k, n'))^{y_k^e} (\mathbb{P}(g | \phi_k, n'))^{y_k^g} \rho_k(n')} \\ &= \frac{(\mathbb{P}(e | \phi_k, n))^{y_k^e} (\mathbb{P}(g | \phi_k, n))^{y_k^g} \rho_k(n)}{\sum_{n'=0}^{N_{\max}} (\mathbb{P}(e | \phi_k, n'))^{y_k^e} (\mathbb{P}(g | \phi_k, n'))^{y_k^g} \rho_k(n')} \end{aligned} \quad (\text{A.1})$$

Ensuite, entre les instants t_k et t_{k+1} , l'interaction non maîtrisée avec l'environnement du champ entraîne une évolution Hamiltonienne de l'état, représentée par la matrice L , telle que :

$$\rho_{k+1}(n) = \sum_{n'=0}^{N_{\max}} L_{n,n'} \rho_k^+(n')$$

La valeur des coefficients de la matrice L est, pour tout $n \geq 0$:

$$\begin{aligned} L_{n,n} &= 1 - \frac{T_a}{T_c} [(1 + n_{\text{th}})n + n_{\text{th}}(n + 1)] \\ L_{n,n+1} &= \frac{T_a}{T_c} (1 + n_{\text{th}})(n + 1) \\ L_{n+1,n} &= \frac{T_a}{T_c} n_{\text{th}}(n + 1) \end{aligned}$$

où $n_{\text{th}} = 0.08$ est le nombre de photons thermiques présents en moyenne dans la cavité laissée au repos.

Lors de la troncature de l'espace de Hilbert, des accommodements doivent être réalisés pour conserver $\sum_{n'=0}^{N_{\max}} L_{n',n} = 1$, lorsque $n = N_{\max}$. Ainsi, en notant $\mathbf{y}_k = (y_k^e, y_k^g)$, l'application $\mathbf{M}_{k,\mathbf{y}_k,\phi_k}$ qui vérifie, pour tout vecteur densité ρ :

$$\mathbf{M}_{k,\mathbf{y}_k,\phi_k}(\rho) = \sigma, \quad \text{avec } \sigma(n) = \sum_{p=0}^{N_{\max}} L_{n,p} (\mathbb{P}(e | \phi_k, p))^{y_k^e} (\mathbb{P}(g | \phi_k, p))^{y_k^g} \rho_k(p)$$

tient un rôle équivalent aux applications de Kraus, puisque, à l'instant t_k , pour tout résultat admissible $\mathbf{y}$ (en notant par abus notation $\text{Tr}(\rho)$ la somme des éléments de ρ) :

$$\mathbb{P}(\mathbf{y} | \rho_k, \phi_k) \propto \text{Tr}(\mathbf{M}_{k,\mathbf{y},\phi_k}(\rho_k)), \quad (\text{A.2})$$

et puisque, après la mesure $\mathbf{y}_k$ effectuée :

$$\rho_{k+1} = \frac{\mathbf{M}_{k,\mathbf{y}_k,\phi_k}(\rho_k)}{\text{Tr}(\mathbf{M}_{k,\mathbf{y}_k,\phi_k}(\rho_k))} \quad (\text{A.3})$$

A.2 Qubit supraconducteur

La second dispositif expérimental sur lequel nous avons travaillé provient du laboratoire Pierre Aigrain de l'École Normale Supérieure. Cette expérience, dont le système physique étudié est un qubit supraconducteur, a été l'objet de la thèse de Philippe Campagne-Ibarcq [CI15]. Au cours de cette thèse, nous avons écrit deux papiers relatifs à cette expérience, sur l'observation de ces systèmes [CISB⁺16] et sur leur contrôle [CIJC⁺16], dans lesquels les procédés de reconstruction d'état présentés en introduction ont été utilisés.

La principale source de décohérence pour un qubit est l'émission (ou l'absorption) spontanée d'un photon. Ces événements ne pouvant être parfaitement connus de l'expérimentateur, nous utiliserons ici une matrice densité pour décrire l'état du qubit au cours de l'expérience. On peut néanmoins récupérer une partie de cette information, en observant le signal de fluorescence émis par le qubit. Dans ce cas, la succession d'états occupés par le qubit au cours du temps est fonction de la réalisation d'un processus stochastique attaché au signal de mesure ; cette succession d'états prend alors à chaque réalisation une valeur différente de la trajectoire moyenne d'états [Car93, WM09, BG09].

En 1992, Wiseman et Milburn ont montré que la mesure hétérodyne du signal de fluorescence d'un qubit engendrait une diffusion continue de son état quantique [GP92, WM93]. Les sauts quantiques ont déjà été observés expérimentalement, de manière indirecte, en utilisant l'information contenue par un système tiers, comme des niveaux d'énergie supplémentaires [BHIW86, NSD86, SNBT86, BKB95], des qubits couplés [GKG⁺07, YSLG⁺08, VLM⁺10, NBS⁺10, VPS⁺14, SPL⁺14] ou des oscillateurs harmoniques [PG99, VSS11]. De plus, on avait déjà observé (voir [MWMS13, HSM⁺13, WCD⁺14, dLRT⁺14, WMS⁺15]) des trajectoires de mesures quantiques diffusives en utilisant l'information envoyée dans un système tiers. Ici, la mesure diffusive hétérodyne du signal électromagnétique émis par la relaxation du qubit est réalisée sans aucun système tiers, procédé qui avait été décrit dans [WM93].

A.2.1 Dispositif expérimental

Le système physique donnant lieu au qubit est un qubit *transmon*, dont la fréquence de transition f_q est environ de 6.3 GHz, placé dans une cavité¹ de cuivre, réfrigéré en continu à 30 mK (voir [KYG⁺07, PSB⁺11]). L'intérêt de cette cavité est double. Premièrement, une ligne de transmission en sortie, couplée fortement avec la cavité, reçoit une grande partie de l'émission de photon provenant de la relaxation du qubit [Pur46]. On note le taux de relaxation du qubit γ_1 , avec $\gamma_1 = (4.15 \mu s)^{-1}$. Deuxièmement, la cavité est utilisée pour réaliser la mesure projective du qubit, qui nous permettra de valider a posteriori la pertinence des reconstructions que nous allons effectuer.

Le signal de fluorescence est amplifié en sortie par un JPC (*Josephson Parametric Converter*), puis reconverti en un signal de fréquence $f_h = 100$ Mhz, et enfin démodulé numériquement et intégré, pour ses deux quadratures dI et dQ sur des pas de temps de $dt = 200$ ns.

A.2.2 Cadre mathématique

L'espace de Hilbert $\mathcal{H}$ dans lequel évolue l'état du qubit est donc isomorphe à $\mathbb{C}^2$. Nous notons $|g\rangle$ la fonction d'onde correspondant au niveau d'énergie le plus bas, et $|e\rangle$ celle associée à un qubit excité. Nous utiliserons, pour décrire l'état du système, le formalisme des matrices de Pauli σ_x , σ_y et σ_z . L'état ρ_t du système à l'instant t est donc décrit par les coordonnées (x_t, y_t, z_t) , où :

$$\rho_t = \frac{I + x_t \sigma_x + y_t \sigma_y + z_t \sigma_z}{2}$$

Dans toutes les expériences présentées ici, l'état initial ρ_0 sera supposé préparé dans l'une de ces trois configurations :

1. Traduction de *off-resonant* ?

- dans l'état d'excitation maximal : $\rho_0 \approx |e\rangle\langle e|$;
- dans un état cohérent : $\rho_0 \approx (I + \sigma_x)/2$;
- dans l'état de repos : $\rho_0 \approx |g\rangle\langle g|$.

Nous ne donnons que des relations d'égalité approximative ici, car la préparation de ces états physiques n'est jamais parfaite : l'estimation de l'état initial *réel* $\bar{\rho}_0$ étant un des sujets de cette thèse.

Les mesures acquises sur le système étant de nature diffusive, l'évolution du système est gouvernée par l'équation différentielle stochastique maîtresse, déjà présentée dans l'introduction. Ici, la mesure hétérodyne de fluorescence donne naissance à deux canaux de mesure dI_t et dQ_t . L'évolution de déphasage de l'état est considérée comme un canal de mesure non lue, si bien que l'état évolue selon :

$$\begin{aligned} d\rho_t = & \left(-\frac{i}{\hbar} [\hat{H}_t, \rho_t] + \gamma_1 \mathcal{D}[\sigma_-, \rho_t] + \gamma_\phi \mathcal{D}[\sigma_z, \rho_t] \right) dt \\ & + \sqrt{\eta\gamma_1/2} \mathcal{M}[\sigma_-, \rho_t] dW_{I,t} \\ & + \sqrt{\eta\gamma_1/2} \mathcal{M}[i\sigma_-, \rho_t] dW_{Q,t} \end{aligned} \quad (\text{A.4})$$

où $\sigma_- = (\sigma_x - i\sigma_y)/2$ est l'opérateur d'annihilation ; $\mathcal{D}[\sigma, \rho] = \sigma\rho\sigma^\dagger - (\sigma^\dagger\sigma\rho + \rho\sigma^\dagger\sigma)/2$ est l'application Lindbladienne et $\mathcal{M}[\sigma, \rho] = \sigma\rho + \rho\sigma^\dagger$ l'opérateur de mesure, pour tout opérateur σ et tout opérateur densité ρ . Les bruits $dW_{I,t}$ et $dW_{Q,t}$ sont des processus de Wiener indépendants entre eux, de variance dt , intervenant dans chaque canal de mesure via :

$$\begin{aligned} dI_t &= \sqrt{\eta\gamma_1/2} \text{Tr}(\mathcal{M}[\sigma_-, \rho_t]) dt + dW_{I,t} \\ dQ_t &= \sqrt{\eta\gamma_1/2} \text{Tr}(\mathcal{M}[i\sigma_-, \rho_t]) dt + dW_{Q,t} \end{aligned} \quad (\text{A.5})$$

On retrouve l'équation maîtresse stochastique décrivant l'évolution des systèmes diffusifs, des équations (2.39) et (2.40), avec $L_1 = \sqrt{\gamma_1/2}\sigma_-$, $L_2 = iL_1$, $L_3 = \sqrt{\gamma_\phi}\sigma_z$, $\eta_1 = \eta_2 = \eta$, $dy_t^1 = dI_t$, $dy_t^2 = dQ_t$ et où dy_t^3 est un bruit ne contenant aucune information sur le système.

Annexe B

Développements mathématiques

B.1 Hamiltonien de trace nulle

Nous montrons ici brièvement que, pour décrire l'évolution Hamiltonienne dans un système quantique, on peut toujours choisir, à chaque instant, un opérateur Hamiltonien de trace nulle.

Prenons le cas le plus général, celui où l'état est mixte, et décrit par un opérateur densité $\rho(t)$ sur $\mathcal{H}$, espace de Hilbert de dimension finie. Considérons pour le moment que le système est uniquement soumis à un Hamiltonien $t \mapsto \hat{H}(t)$ quelconque, de sorte que, en adaptant l'équation (2.1) :

$$d\rho(t) = -\frac{i}{\hbar} [\hat{H}(t), \rho(t)] dt \quad (\text{B.1})$$

Écrivons alors $\hat{J}(t)$ la projection orthogonale de $\hat{H}(t)$ sur l'ensemble des opérateurs de trace nulle :

$$\hat{J}(t) = \hat{H}(t) - \frac{\text{Tr}(\hat{H}(t)) I_{\mathcal{H}}}{N_{\mathcal{H}}}$$

où $N_{\mathcal{H}}$ est la dimension de $\mathcal{H}$. En injectant cette équation dans (B.1), on obtient :

$$d\rho(t) = -\frac{i}{\hbar} [\hat{J}(t), \rho(t)] dt + \frac{i \text{Tr}(\hat{H}(t))}{\hbar N_{\mathcal{H}}} [I_{\mathcal{H}}, \rho(t)] dt \quad (\text{B.2})$$

$$= -\frac{i}{\hbar} [\hat{J}(t), \rho(t)] dt \quad (\text{B.3})$$

Ainsi, $\hat{H}(t)$ et $\hat{J}(t)$ étant permutables à l'envi, on pourra toujours choisir dans cette thèse un Hamiltonien de trace nulle.¹

B.2 Mesure discrète par un système tiers

Nous souhaitons ici donner des arguments mathématiques permettant de comprendre pourquoi le cas d'une mesure discrète par couplage avec l'environnement, puis mesure de celui-ci, se traite, de la même façon que les mesures discrètes imparfaites appliquées

1. Il est possible de vérifier que ce résultat reste valable en présence de mesures diffusives (dans un système régi par l'équation (2.40)), puisque l'évolution Hamiltonienne et l'évolution due à la mesure s'ajoutent simplement.

directement au système, grâce au formalisme des applications de Kraus. L'évolution de l'état due à la mesure et la loi de probabilité qui gouverne celle-ci sont détaillées dans le paragraphe 2.2.4.

Ici, le système expérimental évolue dans l'espace de Hilbert $\mathcal{H}$, et l'environnement dans l'espace de Hilbert $\mathcal{E}$, de dimension finie lui aussi. Initialement, les deux systèmes ne sont pas intriqués, de sorte que l'état conjoint s'écrit (voir (2.33)) :

$$\rho_{\mathcal{H} \otimes \mathcal{E}} = \rho_{\mathcal{H}} \otimes \rho_{\mathcal{E}}$$

Ensuite, une évolution Hamiltonienne, antérieure à la mesure projective, représentée par l'opérateur d'évolution, permet l'intrication des deux systèmes. La mesure projective s'exerce donc sur l'état intriqué :

$$\rho_{\mathcal{H} \otimes \mathcal{E}}^- = U (\rho_{\mathcal{H}} \otimes \rho_{\mathcal{E}}) U^\dagger$$

Supposons désormais qu'on réalise à cet instant une mesure projective sur l'environnement $\mathcal{E}$, grâce aux projecteurs orthogonaux $(P_j)_{j=1 \dots J}$. Quelle est la probabilité de chaque résultat de mesure μ , et comment cette mesure affecte l'état du système $\rho_{\mathcal{H}}$? Commençons par noter tout d'abord $\tilde{P}_j = I_{\mathcal{H}} \otimes P_j$. La probabilité d'obtenir le résultat μ est alors :

$$\begin{aligned} \mathbb{P}(\text{résultat } \mu \mid \rho_{\mathcal{H} \otimes \mathcal{E}}^-) &= \text{Tr}(\tilde{P}_\mu \rho_{\mathcal{H} \otimes \mathcal{E}}^- \tilde{P}_\mu^\dagger) \\ &= \text{Tr}((I_{\mathcal{H}} \otimes P_\mu) U (\rho_{\mathcal{H}} \otimes \rho_{\mathcal{E}}) U^\dagger (I_{\mathcal{H}} \otimes P_\mu)^\dagger) \end{aligned} \quad (\text{B.4})$$

Donnons-nous une base orthonormale de $\mathcal{E}$, notée $(|j_{\mathcal{E}}\rangle)_{j=1 \dots \dim(\mathcal{E})}$. Cette base peut être choisie de sorte qu'il existe une partition $(K_1, \dots, K_J)$ de $\{1; \dots, \dim(\mathcal{E})\}$, telle que pour tout l inférieur à J , et pour tout $j \in K_k$, on ait $P_l |j_{\mathcal{E}}\rangle = |j_{\mathcal{E}}\rangle$. Réécrivons U en somme d'opérateurs non intriqués :

$$U = \sum_{j, j'=1}^{\dim(\mathcal{E})} U_{jj'} \otimes |j\rangle \langle j'|,$$

de sorte que la probabilité de mesurer le résultat μ s'écrit :

$$\begin{aligned} \mathbb{P}(\text{résultat } \mu \mid \rho_{\mathcal{H} \otimes \mathcal{E}}^-) &= \sum_{j, j'=1}^{\dim(\mathcal{E})} \sum_{k, k'=1}^{\dim(\mathcal{E})} \text{Tr}(U_{jj'} \rho_{\mathcal{H}} U_{kk'}^\dagger \otimes P_\mu |j\rangle \langle j'| \rho_{\mathcal{E}} |k'\rangle \langle k| P_\mu) \\ &= \sum_{j \in K_\mu, k \in K_\mu} \sum_{j', k'=1}^{\dim(\mathcal{E})} \text{Tr}(U_{jj'} \rho_{\mathcal{H}} U_{kk'}^\dagger \otimes |j\rangle \langle j'| \rho_{\mathcal{E}} |k'\rangle \langle k|) \\ &= \sum_{j \in K_\mu} \sum_{j', k'=1}^{\dim(\mathcal{E})} \text{Tr}(U_{jj'} \rho_{\mathcal{H}} U_{jk'}^\dagger) \langle j' | \rho_{\mathcal{E}} |k'\rangle \end{aligned} \quad (\text{B.5})$$

On peut alors écrire $\rho_{\mathcal{E}}$ comme une somme de projecteurs orthogonaux pondérés : $\rho_{\mathcal{E}} = \sum_{i=1}^{\text{rg}(\rho_{\mathcal{E}})} \lambda_i |\lambda_i\rangle \langle \lambda_i|$. On finit alors par obtenir :

$$\begin{aligned} \mathbb{P}(\text{résultat } \mu \mid \rho_{\mathcal{H} \otimes \mathcal{E}}^-) &= \sum_{j \in K_\mu} \sum_{j', k'=1}^{\dim(\mathcal{E})} \sum_{i=1}^{\text{rg}(\rho_{\mathcal{E}})} \text{Tr}(U_{jj'} \rho_{\mathcal{H}} U_{jk'}^\dagger) \left(\sqrt{\lambda_i} \langle j' | \lambda_i \rangle \right) \left(\sqrt{\lambda_i} \langle \lambda_i | k' \rangle \right) \\ &= \sum_{j \in K_\mu} \sum_{i=1}^{\dim(\mathcal{E}) \text{rg}(\rho_{\mathcal{E}})} \text{Tr}(V_{ji} \rho_{\mathcal{H}} V_{ji}^\dagger) \end{aligned} \quad (\text{B.6})$$

avec $V_{ji} = \sum_{j'=1}^{\dim(\mathcal{E})} U_{jj'} \left(\sqrt{\lambda_i} \langle j' | \lambda_i \rangle \right)$. De même, après la mesure, l'état du système devient $\rho_{\mathcal{H}}^+$, obtenu par trace partielle sur $\tilde{P}_\mu \rho_{\mathcal{H} \otimes \mathcal{E}} \tilde{P}_\mu^\dagger$:

$$\begin{aligned} \rho_{\mathcal{H}}^+ &= \text{Tr}_{\mathcal{E}} \left(\tilde{P}_\mu \rho_{\mathcal{H} \otimes \mathcal{E}} \tilde{P}_\mu^\dagger \right) \\ &= \frac{\sum_{j \in K_\mu}^{\dim(\mathcal{E})} \sum_{i=1}^{\text{rg}(\rho_{\mathcal{E}})} V_{ji} \rho_{\mathcal{H}} V_{ji}^\dagger}{\sum_{j \in K_\mu}^{\dim(\mathcal{E})} \sum_{i=1}^{\text{rg}(\rho_{\mathcal{E}})} \text{Tr} \left(V_{ji} \rho_{\mathcal{H}} V_{ji}^\dagger \right)} \end{aligned} \quad (\text{B.7})$$

On retrouve alors le schéma général des mesures discrètes effectuées par mesure projective sur le système lui-même. En effet, les matrices V_{ji} définissent un ensemble *complet* d'applications de Kraus partielles, puisque :

$$\begin{aligned} \sum_{\mu=1}^J \sum_{j \in K_\mu}^{\dim(\mathcal{E})} \sum_{i=1}^{\text{rg}(\rho_{\mathcal{E}})} V_{ji}^\dagger V_{ji} &= \sum_{j,j',k'=1}^{\dim(\mathcal{E})} \sum_{i=1}^{\text{rg}(\rho_{\mathcal{E}})} U_{j'j}^\dagger U_{jk'} \left(\sqrt{\lambda_i} \langle \lambda_i | j' \rangle \right) \left(\sqrt{\lambda_i} \langle k' | \lambda_i \rangle \right) \\ &= \sum_{j',k'=1}^{\dim(\mathcal{E})} \delta_{j',k'} I_{\mathcal{H}} \langle k' | \rho_{\mathcal{H}} | j' \rangle \\ &= \sum_{j'=1}^{\dim(\mathcal{E})} I_{\mathcal{H}} \langle j' | \rho_{\mathcal{H}} | j' \rangle = I_{\mathcal{H}} \end{aligned} \quad (\text{B.8})$$

B.3 Processus de Wiener

Nous présentons ici les principales caractéristiques des processus de Wiener, processus stochastiques à temps continu, dont nous avons besoin pour décrire l'évolution des systèmes quantiques à mesures diffusives.

Le processus de Wiener est un processus de Lévy. A ce titre, en considérant que l'initialisation du processus se fait à l'instant $t = 0$, et en notant W_t la valeur de ce processus au temps $t \geq 0$, on a les propriétés suivantes :

- $\mathbb{P}(W_0 = 0) = 1$;
- les accroissements sont indépendants : pour tout $N \geq 2$, et pour tous instants $0 \leq t_1 < t_2 < \dots < t_N$, les accroissements $W_{t_{k+1}} - W_{t_k}$ sont indépendants entre eux ;
- les accroissements sont stationnaires : pour toute durée $s > 0$, il existe une distribution $N(s)$ sur $\mathbb{R}$ telle que, pour tout instant $0 \leq t_1$, $W_{t_1+s} - W_{t_1}$ est égale en loi à $N(s)$;
- la fonction $t \mapsto W_t$ est Càdlàg, c'est-à-dire qu'elle est continue presque sûrement à droite ($\forall \epsilon > 0$ et $\forall t \geq 0$, on a $\lim_{h \rightarrow 0^+} \mathbb{P}(|W_{t+h} - W_t| > \epsilon) = 0$), et qu'elle admet toujours une limite à gauche.

De plus, un processus de Lévy est appelé processus de Wiener s'il vérifie les deux conditions suivantes :

- pour tout instant $0 \leq t_1$, et toute durée $s > 0$, $W_{t_1+s} - W_{t_1}$ est égale en loi à $N(s)$, une loi Gaussienne de moyenne nulle et de variance s ;
- la trajectoire $t \mapsto W_t$ est presque sûrement continue.

Nous notons $dW_t = W_{t+dt} - W_t$. Dans cette thèse, nous utiliserons parfois en même temps plusieurs réalisations du processus de Wiener, notées $t \mapsto W_t^\nu$. Ces réalisations sont indépendantes. Ainsi, pour tous réels t_1 et t_2 positifs, et pour tous entiers ν_1 et ν_2 , l'espérance de $dW_{t_1}^{\nu_1} dW_{t_2}^{\nu_2}$ vaut $\delta_{\nu_1, \nu_2} \delta_{t_1, t_2} dt$.

Les processus de Wiener sont très utilisés pour modéliser des mouvements browniens.

B.4 Schéma de discrétisation des équations de mesures diffusives

Cette section a pour but de justifier les schémas de discrétisation des diverses formes de l'équation maîtresse stochastique régissant les systèmes à mesures diffusives, présentés dans le paragraphe 2.2.4. Nous ne montrons formellement la convergence en loi vers l'équation différentielle stochastique maîtresse que dans le cas du dernier schéma de discrétisation, impliquant $\sigma(t)$, puisqu'il est le seul à constituer une véritable densité de probabilité lorsque $\Delta t > 0$. Nous nous inspirons ici des démonstrations réalisées dans [JCRH15] et [APR14].

Ici, pour simplifier les calculs, nous traitons du cas où le système n'est observé que par un seul canal de mesure, dont l'opérateur associé est noté L et l'efficacité de mesure $\eta > 0$. Le système est soumis à un Hamiltonien $\hat{H}(t)$, et nous nous plaçons à l'instant t .

Lorsque l'état est représenté, à l'instant t , par l'opérateur densité ρ_t , et que le système est soumis à une évolution Hamiltonienne $\hat{H}(t)$ et à un unique canal de mesure représenté par L et $\eta > 0$, celui-ci évolue selon :

$$d\rho_t = \left(-\frac{i}{\hbar} [\hat{H}(t), \rho_t] + \mathcal{L}(\rho_t) \right) dt + \sqrt{\eta} (\mathcal{M}(\rho_t) - \rho_t \overline{\mathcal{M}}(\rho_t)) dW_t \quad (\text{B.9})$$

où l'élément différentiel $d\rho_t$ est donné par $\rho_{t+dt} = \rho_t + d\rho_t$, où $\mathcal{L}(\rho) = L\rho L^\dagger - (L^\dagger L\rho + \rho L^\dagger L)/2$ est le Lindbladien de la mesure L , $\mathcal{M}(\rho) = L\rho + \rho L^\dagger$ et $\overline{\mathcal{M}}(\rho) = \text{Tr}(L\rho + \rho L^\dagger)$ et où dW_t est un processus de Wiener de variance dt , lié au signal de mesure dy_t par :

$$dy_t = \sqrt{\eta \overline{\mathcal{M}}}(\rho_t) dt + dW_t \quad (\text{B.10})$$

De cette équation différentielle, nous avons présenté trois schémas de discrétisation, bâtis sur la structure suivante :

$$\rho_{t+\Delta t} = \frac{\mathbf{K}_{t,\Delta y_t}(\rho_t)}{\text{Tr}(\mathbf{K}_{t,\Delta y_t}(\rho_t))}, \quad (\text{B.11})$$

$$\mathbb{P}(\Delta y_t \in [s; s+ds] \sqrt{\Delta t} \mid \rho_t) = \frac{\text{Tr}(\mathbf{K}_{t,\Delta y_t}(\rho_t))}{\sqrt{2\pi}} \exp\left(-\frac{s^2}{2}\right) ds \quad (\text{B.12})$$

avec différentes façons de construire l'application de Kraus $\mathbf{K}_{t,\Delta y_t}$. Nous avons proposé en (2.44) une application de Kraus de base, notée $\mathbf{K}_{t,\Delta y_t}^{\text{sim}}$, grâce à laquelle se construit le schéma de discrétisation simple, et à partir de laquelle nous créons deux schémas de discrétisation alternatifs. A priori, seul le schéma de l'équation (2.48), construit grâce à la matrice $\sigma(t)$, correspond à une vraie loi de probabilité pour $\Delta t > 0$, puisque l'intégrale de ses probabilités fait 1. Nous montrons donc que ce dernier schéma converge en loi vers le processus défini par les équations différentielles (B.9) et (B.10), lorsque Δt tend vers 0^+ .

Deux conditions sont nécessaires pour la convergence en loi du schéma de discrétisation des équations (B.11) et (B.12) vers l'équation différentielle stochastique maîtresse lorsque Δt tend vers 0 :

- pour toute fonction scalaire f de classe C^2 à support compact sur l'ensemble des matrices densité de $\mathcal{H}$, on doit avoir :

$$\lim_{\Delta t \rightarrow 0^+} \mathbb{E} \left(\frac{f(\rho_{t+\Delta t}) - f(\rho_t)}{\Delta t} \middle| \rho_t \right) = \mathbb{E} \left(\frac{f(\rho_{t+dt}) - f(\rho_t)}{dt} \middle| \rho_t \right),$$

où le terme de gauche est calculé via le schéma de discrétisation, par développement limité sur Δt , et le terme de droite via l'équation différentielle stochastique, grâce au générateur de Markov de f ;

- le schéma de discrétisation vérifie une hypothèse de *tightness* (voir [Pel08]) : pour un intervalle $\Delta t > 0$, il existe un réel $Z > 0$ tel que, pour tous instants $k_1\Delta t$ et $k_2\Delta t$, avec $k_1 < k_2$, on a, pour $k_1 < k < k_2$,

$$\mathbb{E} \left(\|\rho_{k_2\Delta t} - \rho_{k\Delta t}\|^2 \|\rho_{k\Delta t} - \rho_{k_1\Delta t}\|^2 \right) \leq Z (k_2 - k_1)^2 (\Delta t)^2 \quad (\text{B.13})$$

Nous montrons d'abord que la première de ces deux conditions est vérifiée par tous les schémas de discrétisation. En effet, cette espérance prend la valeur $Af(\rho_t)$ lorsque calculée avec l'équation maîtresse stochastique, en utilisant les règles de calcul d'Itô², et le fait que l'espérance de dW_t est nulle :

$$\begin{aligned} Af(\rho_t) = \nabla f|_{\rho_t} \cdot & \left[-\frac{i}{\hbar} [\hat{H}(t), \rho_t] + \mathcal{L}(\rho_t) \right] \\ & + \frac{\eta}{2} \nabla^2 f|_{\rho_t} \cdot \left[\left(\mathcal{M}(\rho_t) - \rho_t \overline{\mathcal{M}}(\rho_t) \right), \left(\mathcal{M}(\rho_t) - \rho_t \overline{\mathcal{M}}(\rho_t) \right) \right] \end{aligned} \quad (\text{B.14})$$

Nous calculons maintenant la valeur de cette espérance, grâce au schéma de discrétisation simple. Ainsi, nous choisissons ici $\mathbf{K}_{t,\Delta y_t} = \mathbf{K}_{t,\Delta y_t}^{\text{sim}}$. Ce calcul est réalisé par développement limité. On note ici $\Delta y_t = s\Delta t$. Tout d'abord, on cherche la valeur de $\mathbf{K}_{t,s\Delta t}(\rho_t)$:

$$\begin{aligned} \mathbf{K}_{t,s\Delta t}(\rho_t) = \rho_t + s\sqrt{\eta}\mathcal{M}(\rho_t)\sqrt{\Delta t} \\ + \left(-\frac{i}{\hbar} [\hat{H}(t), \rho_t] + \mathcal{L}(\rho_t) + (s^2 - 1)\eta L\rho_t L^\dagger \right) \Delta t + \mathcal{O}((\Delta t)^{3/2}) \end{aligned}$$

Sa trace prend donc la valeur :

$$\text{Tr}(\mathbf{K}_{t,s\Delta t}(\rho_t)) = 1 + s\sqrt{\eta}\overline{\mathcal{M}}(\rho_t)\sqrt{\Delta t} + (s^2 - 1)\eta \text{Tr}(L\rho_t L^\dagger) \Delta t + \mathcal{O}((\Delta t)^{3/2}),$$

et l'inverse de cette trace :

$$\begin{aligned} \frac{1}{\text{Tr}(\mathbf{K}_{t,s\Delta t}(\rho_t))} = 1 - s\sqrt{\eta}\overline{\mathcal{M}}(\rho_t)\sqrt{\Delta t} \\ + \eta \left(s^2 \overline{\mathcal{M}}^2(\rho_t) - (s^2 - 1)\text{Tr}(L\rho_t L^\dagger) \right) \Delta t + \mathcal{O}((\Delta t)^{3/2}) \end{aligned}$$

Ainsi, $\rho_{t+\Delta t}$ est donné par :

$$\begin{aligned} \rho_{t+\Delta t} = \rho_t + s\sqrt{\eta} \left(\mathcal{M}(\rho_t) - \rho_t \overline{\mathcal{M}}(\rho_t) \right) \sqrt{\Delta t} \\ + \left(-\frac{i}{\hbar} [\hat{H}(t), \rho_t] + \mathcal{L}(\rho_t) + (s^2 - 1)\eta (L\rho_t L^\dagger - \rho_t \text{Tr}(L\rho_t L^\dagger)) \right. \\ \left. - s^2 \eta \overline{\mathcal{M}}(\rho_t) \left(\mathcal{M}(\rho_t) - \rho_t \overline{\mathcal{M}}(\rho_t) \right) \right) \Delta t + \mathcal{O}((\Delta t)^{3/2}) \end{aligned}$$

2. référence ito

Ainsi, pour toute fonction f de classe C^2 , la variation $f(\rho_{t+\Delta t}) - f(\rho_t)$ est donnée par :

$$\begin{aligned} f(\rho_{t+\Delta t}) - f(\rho_t) = & \nabla f|_{\rho_t} \cdot \left[s\sqrt{\eta} \left(\mathcal{M}(\rho_t) - \rho_t \overline{\mathcal{M}}(\rho_t) \right) \sqrt{\Delta t} \right. \\ & + \left(-\frac{i}{\hbar} [\hat{H}(t), \rho_t] + \mathcal{L}(\rho_t) + (s^2 - 1)\eta \left(L\rho_t L^\dagger - \rho_t \text{Tr}(L\rho_t L^\dagger) \right) \right. \\ & \quad \left. \left. - s^2 \eta \overline{\mathcal{M}}(\rho_t) \left(\mathcal{M}(\rho_t) - \rho_t \overline{\mathcal{M}}(\rho_t) \right) \right) \Delta t \right] \\ & + \frac{\eta}{2} \nabla^2 f|_{\rho_t} \cdot \left[\left(\mathcal{M}(\rho_t) - \rho_t \overline{\mathcal{M}}(\rho_t) \right), \left(\mathcal{M}(\rho_t) - \rho_t \overline{\mathcal{M}}(\rho_t) \right) \right] \Delta t + \mathcal{O}((\Delta t)^{3/2}) \end{aligned}$$

Or, compte tenu de l'équation (B.12), le calcul de l'espérance $Bf(\rho_t)$ revient au calcul de l'espérance de $\text{Tr}(\mathbf{K}_{t,s\sqrt{\Delta t}}(\rho_t)) (f(\rho_{t+\Delta t}) - f(\rho_t)) / \Delta t$, où s est la réalisation d'une loi Gaussienne de moyenne nulle et de variance 1. On trouve finalement :

$$\begin{aligned} Bf(\rho_t) = & \nabla f|_{\rho_t} \cdot \left[-\frac{i}{\hbar} [\hat{H}(t), \rho_t] + \mathcal{L}(\rho_t) \right] \\ & + \frac{\eta}{2} \nabla^2 f|_{\rho_t} \cdot \left[\left(\mathcal{M}(\rho_t) - \rho_t \overline{\mathcal{M}}(\rho_t) \right), \left(\mathcal{M}(\rho_t) - \rho_t \overline{\mathcal{M}}(\rho_t) \right) \right] + \mathcal{O}(\sqrt{\Delta t}) \end{aligned}$$

Nous venons finalement de prouver que :

$$Af(\rho_t) = Bf(\rho_t) + \mathcal{O}(\sqrt{\Delta t}) \quad (\text{B.15})$$

Ce résultat est valable dans le cas des deux autres schémas de discrétisation. En effet, ceux-ci font intervenir dans l'expression du développement limité en Δt de $\mathbf{K}_{t,s\Delta t}(\rho_t)$ des termes en $(\Delta t)^2$ au minimum, ce qui n'invalide pas la formule ci-dessus.

De plus, le schéma de discrétisation impliquant la matrice $\sigma(t)$ vérifie bien la propriété de l'équation (B.13). Pour démontrer cela, montrons tout d'abord que, pour un $\Delta t > 0$ donné, il existe un $Z > 0$ tel que pour tous k_1 et $k_2 > k_1$ entiers, on ait :

$$\mathbb{E} \left(\left\| \rho_{k_2 \Delta t} - \rho_{k_1 \Delta t} \right\|^2 \middle| \rho_{k_1 \Delta t} \right) \leq Z (k_2 - k_1) \Delta t$$

Notons d'abord que :

$$\begin{aligned} \left\| \rho_{k_2 \Delta t} - \rho_{k_1 \Delta t} \right\|^2 = & \left\| \rho_{k_2 \Delta t} - \rho_{(k_2-1)\Delta t} \right\|^2 + \left\| \rho_{(k_2-1)\Delta t} - \rho_{k_1 \Delta t} \right\|^2 \\ & + 2\text{Re} \left(\langle \rho_{k_2 \Delta t} - \rho_{(k_2-1)\Delta t}, \rho_{(k_2-1)\Delta t} - \rho_{k_1 \Delta t} \rangle \right), \end{aligned}$$

et Y_t l'ensemble des mesures sur le système prises avant l'instant t : $Y_{(k_2-1)\Delta t}$ définit donc l'état $\rho_{(k_2-1)\Delta t}$. En utilisant les expressions de $\rho_{t+\Delta t}$ et de $\text{Tr}(\mathbf{K}_{t,s\Delta t}(\rho_t))$ exprimées plus haut, et la compacité de l'ensemble des matrices densité, on trouve qu'il existe $Z_1 > 0$ et $Z_2 > 0$ tel que, $\forall Y_{(k_2-1)\Delta t}$:

$$\begin{aligned} \mathbb{E} \left(\left\| \rho_{k_2 \Delta t} - \rho_{(k_2-1)\Delta t} \right\|^2 \middle| Y_{(k_2-1)\Delta t} \right) & \leq Z_1 \Delta t \\ \mathbb{E} \left(\text{Re} \left(\langle \rho_{k_2 \Delta t} - \rho_{(k_2-1)\Delta t}, \rho_{(k_2-1)\Delta t} - \rho_{k_1 \Delta t} \rangle \right) \middle| Y_{(k_2-1)\Delta t} \right) & \leq Z_2 \Delta t \end{aligned}$$

Ainsi, il existe $Z > 0$ tel que :

$$\mathbb{E} \left(\left\| \rho_{k_2 \Delta t} - \rho_{k_1 \Delta t} \right\|^2 \middle| Y_{(k_2-1)\Delta t} \right) \leq Z \Delta t + \mathbb{E} \left(\left\| \rho_{(k_2-1)\Delta t} - \rho_{k_1 \Delta t} \right\|^2 \middle| Y_{(k_2-1)\Delta t} \right)$$

Ainsi, par itération et intégration sur l'ensemble des $Y_{k\Delta t}$ possibles, on a presque sûrement :

$$\mathbb{E} \left(\|\rho_{k_2\Delta t} - \rho_{k_1\Delta t}\|^2 \right) \leq Z(k_2 - k_1)\Delta t \quad (\text{B.16})$$

Nous nous servons alors du même raisonnement que celui de [Pel08, Proposition 5] pour démontrer la validité de l'équation (B.13).

Ainsi, le schéma de discrétisation utilisant la matrice $\sigma(t)$ converge bien en loi vers le processus dirigé par l'équation différentielle stochastique maîtresse.

B.5 Autour de la fidélité

Nous rappelons que la fidélité est une des pseudo-métriques les plus fréquemment employées, avec la *trace distance*, pour comparer deux opérateurs densité sur le même espace de Hilbert $\mathcal{H}$. Pour deux opérateurs densité ρ et σ sur $\mathcal{H}$, la fidélité entre ces deux états est donnée par la formule :

$$F(\rho, \sigma) = \text{Tr} \left(\sqrt{\sqrt{\rho}\sigma\sqrt{\rho}} \right),$$

où $\sqrt{\rho}$ est la racine carrée d'une matrice positive ρ : lorsque ρ s'écrit $\rho = \sum_{j=1}^{N_{\mathcal{H}}} \lambda_j |\rho_j\rangle\langle\rho_j|$, sa matrice carrée devient :

$$\sqrt{\rho} = \sum_{j=1}^{N_{\mathcal{H}}} \sqrt{\lambda_j} |\rho_j\rangle\langle\rho_j|$$

Nous cherchons dans cette annexe à donner la démonstration de certaines propriétés de la fidélité, notamment sa symétrie, les valeurs qu'elle peut prendre, et les cas de figure lors desquels la fidélité prend ses valeurs minimales et maximales.

B.5.1 Cas d'un état pur

Nous commençons par extraire, dans le cas particulier où un des deux états dont on calcule la fidélité est pur, une formule de la fidélité un peu plus simple à interpréter. Dans le cas où ρ est un état pur, on peut le réécrire comme $\rho = |\psi\rangle\langle\psi|$, avec $|\psi\rangle$ un vecteur normé de $\mathcal{H}$. La fidélité entre ρ et un état mixte quelconque représenté par σ est dans ce cas :

$$\begin{aligned} F(\rho, \sigma) &= F(|\psi\rangle\langle\psi|, \sigma) = \text{Tr} \left(\sqrt{|\psi\rangle\langle\psi|\sigma|\psi\rangle\langle\psi|} \right) \\ &= \langle\psi|\sigma|\psi\rangle \end{aligned} \quad (\text{B.17})$$

Réciproquement, lorsque c'est l'état σ qui est pur et vaut $|\psi\rangle\langle\psi|$, la fidélité entre ρ et ce projecteur devient :

$$F(\rho, \sigma) = F(\rho, |\psi\rangle\langle\psi|) = \text{Tr} \left(\sqrt{\sqrt{\rho}|\psi\rangle\langle\psi|\sqrt{\rho}} \right) \quad (\text{B.18})$$

$$\begin{aligned} &= \text{Tr} \left(\sqrt{|\lambda\rangle\langle\lambda|} \right) \text{ avec } |\lambda\rangle = \sqrt{\rho}|\psi\rangle \\ &= \langle\lambda|\lambda\rangle = \langle\psi|\rho|\psi\rangle \end{aligned} \quad (\text{B.19})$$

Ainsi, lorsqu'un des deux états ρ et σ est pur, la fidélité est une application symétrique, et se calcule par projection sur la fonction d'onde correspondant à l'état pur de la matrice

densité représentant le second état physique. Dans ce cas particulier, la fidélité prend donc une valeur entre 0 et 1. Si la fidélité est nulle, les supports des deux matrices densité sont complètement orthogonaux, et si la fidélité vaut un, les deux matrices densité sont rigoureusement les mêmes.

Nous allons maintenant étendre ces résultats au cas plus général où les deux matrices densité peuvent représenter des états mixtes.

B.5.2 Théorème d’Uhlmann

Pour obtenir les résultats que nous cherchons sur la fidélité dans le cas général, nous devons commencer par présenter un lemme, le théorème d’Uhlmann. Ce théorème est une réécriture de la formule de la fidélité, qui utilise la notion de purification d’un état mixte.

Purification d’un état mixte

Commençons par présenter le principe de purification d’un état mixte ρ sur l’espace de Hilbert complexe $\mathcal{H}$, de dimension finie $N_{\mathcal{H}}$. Cet état mixte s’écrit $\rho = \sum_{j=1}^{N_{\mathcal{H}}} \lambda_j |\rho_j\rangle\langle\rho_j|$, où $(|\rho_1\rangle, \dots, |\rho_{N_{\mathcal{H}}}\rangle)$ forme une base orthonormée de $\mathcal{H}$. Le principe de la purification de ρ est d’intriquer $\mathcal{H}$ avec un autre espace de Hilbert $\mathcal{Q}$, tel que ρ soit la trace partielle par rapport à $\mathcal{Q}$ d’un état pur de $\mathcal{H} \otimes \mathcal{Q}$.

Choisissons un espace de Hilbert $\mathcal{Q}$ isomorphe à $\mathcal{H}$. Soit $(|j_{\mathcal{Q}}\rangle)_{j=1, \dots, N_{\mathcal{H}}}$ une base orthonormée quelconque de $\mathcal{Q}$. On définit comme une purification de ρ la fonction d’onde $|\tilde{\rho}\rangle$ sur $\mathcal{H} \otimes \mathcal{Q}$, telle que :

$$|\tilde{\rho}\rangle = \sum_{j=1}^{N_{\mathcal{H}}} \sqrt{\lambda_j} |\rho_j\rangle \otimes |j_{\mathcal{Q}}\rangle \quad (\text{B.20})$$

En effet, la trace partielle par rapport à $\mathcal{Q}$ du projecteur $|\tilde{\rho}\rangle\langle\tilde{\rho}|$ est donnée par :

$$\text{Tr}_{\mathcal{Q}}(|\tilde{\rho}\rangle\langle\tilde{\rho}|) = \sum_{j=1}^{N_{\mathcal{H}}} \langle j_{\mathcal{Q}}|\tilde{\rho}\rangle\langle\tilde{\rho}|j_{\mathcal{Q}}\rangle = \sum_{j=1}^{N_{\mathcal{H}}} \lambda_j |\rho_j\rangle\langle\rho_j| = \rho$$

Notons que la définition d’une purification d’un état mixte ρ , que nous venons de présenter, n’implique aucunement l’unicité de cette purification, puisque cette purification dépend de la base orthonormale de $\mathcal{Q}$ choisie. En effet, toute matrice unitaire U sur $\mathcal{Q}$ permet d’obtenir une autre purification $|\tilde{\rho}\rangle_U$ de ρ , donnée par :

$$|\tilde{\rho}\rangle_U = \sum_{j=1}^{N_{\mathcal{H}}} \sqrt{\lambda_j} |\rho_j\rangle \otimes U|j_{\mathcal{Q}}\rangle \quad (\text{B.21})$$

Démonstration du théorème d’Uhlmann

Nous écrivons désormais le théorème de Uhlmann, qui lie la fidélité de deux états mixtes et celle de leurs purifications.

Théorème 6. (Théorème d’Uhlmann)

Soit ρ et σ deux opérateurs densité d’un espace de Hilbert complexe $\mathcal{H}$ de dimension finie. Alors, la fidélité entre ces deux états vérifie :

$$\begin{aligned} F(\rho, \sigma) &= \max_{|\tilde{\rho}\rangle, |\tilde{\sigma}\rangle} F(|\tilde{\rho}\rangle\langle\tilde{\rho}|, |\tilde{\sigma}\rangle\langle\tilde{\sigma}|) \\ &= \max_{|\tilde{\rho}\rangle, |\tilde{\sigma}\rangle} |\langle\tilde{\rho}|\tilde{\sigma}\rangle|, \end{aligned} \quad (\text{B.22})$$

où la maximisation se fait sur l'ensemble des purifications $\tilde{\rho}$ et $\tilde{\sigma}$ de ρ et de σ .

La démonstration du théorème d'Uhlmann est reprise de [NC00]. Elle nécessite le lemme suivant.

Lemme 4. Soit A un opérateur quelconque, non nécessairement positif, sur un espace de Hilbert $\mathcal{H}$ de dimension finie, et U un opérateur unitaire sur $\mathcal{H}$. Alors, on a l'inégalité :

$$|\mathrm{Tr}(AU)| \leq \mathrm{Tr}(|A|) \quad (\text{B.23})$$

L'égalité est atteinte lorsque $U = V^\dagger$, où V est obtenue par la décomposition polaire de A : $A = |A| V$ ($|A|$ étant un opérateur positif ou nul et V un opérateur autoadjoint).

Démonstration. L'obtention de l'égalité lorsque $U = V^\dagger$ est triviale. Dans le cas général, on transforme :

$$|\mathrm{Tr}(AU)| = |\mathrm{Tr}(|A| VU)| = \left| \mathrm{Tr} \left(|A|^{\frac{1}{2}} |A|^{\frac{1}{2}} VU \right) \right|$$

L'application de l'inégalité de Cauchy-Schwarz, en considérant le produit scalaire de Frobenius sur l'ensemble des opérateurs de $\mathcal{H}$, permet de donner l'inégalité suivante :

$$|\mathrm{Tr}(AU)| \leq \sqrt{\mathrm{Tr}(|A|) \mathrm{Tr}(U^\dagger V^\dagger |A| VU)} = \mathrm{Tr}(|A|)$$

ce qui démontre le lemme. ■

On peut alors commencer à exposer la preuve du théorème d'Uhlmann.

Démonstration. Soit ρ un opérateur densité quelconque de l'espace de Hilbert complexe $\mathcal{H}$, de dimension finie $N_{\mathcal{H}}$, et soit $\mathcal{Q}$ un espace de Hilbert isomorphe à $\mathcal{H}$. Soit également $(|j_{\mathcal{H}}\rangle)_{j=1,\dots,N_{\mathcal{H}}}$ et $(|j_{\mathcal{Q}}\rangle)_{j=1,\dots,N_{\mathcal{H}}}$ deux bases orthonormales quelconques, respectivement de $\mathcal{H}$ et $\mathcal{Q}$. On construit alors le vecteur $|m\rangle$ de l'espace intriqué $\mathcal{H} \otimes \mathcal{Q}$, de la façon suivante :

$$|m\rangle = \sum_{j=1}^{N_{\mathcal{H}}} |j_{\mathcal{H}}\rangle \otimes |j_{\mathcal{Q}}\rangle$$

En utilisant la décomposition de Schmidt des vecteurs d'un espace $\mathcal{H} \otimes \mathcal{Q}$ intriqué, on peut alors trouver, pour toute purification $|\tilde{\rho}\rangle$ sur $\mathcal{H} \otimes \mathcal{Q}$ d'un opérateur densité ρ de $\mathcal{H}$, deux opérateurs unitaires $U_{\mathcal{H}}$ et $U_{\mathcal{Q}}$ de $\mathcal{H}$ et $\mathcal{Q}$, tels que :

$$|\tilde{\rho}\rangle = (\sqrt{\rho} U_{\mathcal{H}} \otimes U_{\mathcal{Q}}) |m\rangle$$

Réciproquement, pour toutes matrices unitaires $U_{\mathcal{H}}$ et $U_{\mathcal{Q}}$, le vecteur $|\tilde{\rho}\rangle$ ainsi créé est une purification de ρ . De même, $|\tilde{\sigma}\rangle$ est une purification de l'opérateur densité σ si et seulement si il existe $V_{\mathcal{H}}$ et $V_{\mathcal{Q}}$ deux opérateurs unitaires tels que :

$$|\tilde{\sigma}\rangle = (\sqrt{\sigma} V_{\mathcal{H}} \otimes V_{\mathcal{Q}}) |m\rangle$$

Ainsi exprimées, le produit scalaire de ces deux purifications a pour valeur absolue :

$$\begin{aligned} |\langle \tilde{\rho} | \tilde{\sigma} \rangle| &= \left| \langle m | \left(U_{\mathcal{H}}^\dagger \sqrt{\rho} \otimes U_{\mathcal{Q}}^\dagger \right) \left(\sqrt{\sigma} V_{\mathcal{H}} \otimes V_{\mathcal{Q}} \right) | m \rangle \right| \\ &= \left| \langle m | \left(U_{\mathcal{H}}^\dagger \sqrt{\rho} \sqrt{\sigma} V_{\mathcal{H}} \otimes U_{\mathcal{Q}}^\dagger V_{\mathcal{Q}} \right) | m \rangle \right| \end{aligned}$$

En utilisant le fait que, pour tous opérateurs A et B , de $\mathcal{H}$ et $\mathcal{Q}$ isomorphes, on ait $\langle m|A \otimes B|m\rangle = \text{Tr}(AB^T)$ (où B^T est la transposée de B), alors :

$$\begin{aligned} |\langle \tilde{\rho}|\tilde{\sigma}\rangle| &= \left| \text{Tr} \left(U_{\mathcal{H}}^\dagger \sqrt{\rho} \sqrt{\sigma} V_{\mathcal{H}} \otimes U_{\mathcal{Q}}^\dagger V_{\mathcal{Q}} \right) \right| \\ &= \left| \text{Tr} \left(U_{\mathcal{H}}^\dagger \sqrt{\rho} \sqrt{\sigma} V_{\mathcal{H}} V_{\mathcal{Q}}^T (U_{\mathcal{Q}}^\dagger)^T \right) \right| \end{aligned} \quad (\text{B.24})$$

$$= \left| \text{Tr} \left(\sqrt{\rho} \sqrt{\sigma} U \right) \right| \quad (\text{B.25})$$

où $U = V_{\mathcal{H}} V_{\mathcal{Q}}^T (U_{\mathcal{Q}}^\dagger)^T U_{\mathcal{H}}^\dagger$ est une matrice unitaire. Or, la trace $\text{Tr} \left(\left| \sqrt{\rho} \sqrt{\sigma} \right| \right)$ étant simplement égale à la fidélité $F(\rho, \sigma)$, l'utilisation du lemme 4 permet d'obtenir que :

- $|\langle \tilde{\rho}|\tilde{\sigma}\rangle| \leq F(\rho, \sigma)$ pour toutes purifications $|\tilde{\rho}\rangle$ et $|\tilde{\sigma}\rangle$;
- en choisissant $V_{\mathcal{Q}} = U_{\mathcal{Q}} = U_{\mathcal{H}} = I_{\mathcal{H}}$, et $V_{\mathcal{H}}$ tel que $\sqrt{\rho} \sqrt{\sigma} = \left| \sqrt{\rho} \sqrt{\sigma} \right| V_{\mathcal{H}}$, on obtient $|\langle \tilde{\rho}|\tilde{\sigma}\rangle| = F(\rho, \sigma)$.

Ainsi, le théorème d'Uhlmann est démontré. ■

B.5.3 Propriétés de la fidélité

Une fois le théorème d'Uhlmann démontré, il est facile de démontrer que la fidélité vérifie les conditions usuelles requises pour une pseudo-métrique :

- **symétrie** : la nouvelle formulation de la fidélité rend cette propriété évidente. Pour tous opérateurs densité ρ et σ , on a donc $F(\rho, \sigma) = F(\sigma, \rho)$;
- **bornes de la fidélité** : de même, cette nouvelle formulation montre clairement que la fidélité ne prend que des valeurs comprises entre 0 et 1, puisque les purifications sont des fonctions d'onde, par définition bornées ;
- **cas où la fidélité est nulle** : la fidélité $F(\rho, \sigma)$ est nulle si et seulement si toutes purifications de ρ et de σ sont orthogonales. Cela nous amène rapidement à conclure que la fidélité entre deux opérateurs densité est nul si et seulement si leurs supports sont orthogonaux ;
- **cas où la fidélité est maximale** : la fidélité $F(\rho, \sigma)$ est maximale, et vaut 1, lorsqu'il est possible de trouver des purifications $|\tilde{\rho}\rangle$ et $|\tilde{\sigma}\rangle$, de ρ et σ respectivement, colinéaires. Cette condition impose que les deux opérateurs densité aient le même spectre, et que chaque valeur propre soit associée à deux vecteurs colinéaires. Par conséquent, le fait que $F(\rho, \sigma) = 1$ équivaut à $\rho = \sigma$.

B.6 Maximisation de la fonction de log-vraisemblance pour la tomographie d'état

B.6.1 Projection sur l'ensemble des matrices densité admissibles

Dans cette sous-section, nous montrons comment réaliser la projection d'un opérateur linéaire ρ de $\mathcal{L}(\mathcal{H})$ sur l'ensemble des matrices densité $\mathcal{D}_{\mathcal{H}}$ associé au même espace de Hilbert. Cette projection est notée $\mathcal{P}_{\mathcal{D}_{\mathcal{H}}}$.

Puisque ρ est Hermitienne, ρ est également diagonalisable, et on peut réécrire $\rho =$

$UDU^\dagger$, avec $U^\dagger U = I_{\mathcal{H}}$ et D une matrice diagonale réelle. La projection de ρ est alors :

$$\begin{aligned}\mathcal{P}_{\mathcal{D}_{\mathcal{H}}}(\rho) &= \arg \min_{\sigma \in \mathcal{D}_{\mathcal{H}}} \text{Tr} \left((UDU^\dagger - \sigma)^2 \right) \\ &= U \arg \min_{\sigma \in \mathcal{D}_{\mathcal{H}}} \text{Tr} \left((D - \sigma)^2 \right) U^\dagger \\ &= U \left(\arg \min_{\sigma \in \mathcal{D}_{\mathcal{H}}} \left(\sum_{k=1}^{N_{\mathcal{H}}} (D_k - \sigma_{kk})^2 + \sum_{k=1}^{N_{\mathcal{H}}} \sum_{k'=1, k' \neq k}^{N_{\mathcal{H}}} |\sigma_{kk'}|^2 \right) \right) U^\dagger\end{aligned}\quad (\text{B.26})$$

Maintenant, rappelons-nous que la contrainte $\text{Tr}(\sigma) = 1$ s'applique seulement sur les coefficients σ_{kk} . Puisque pour chaque matrice positive σ de trace une, la réduction de σ à sa seule diagonale est aussi positive et de trace une, on peut supposer que $\sigma_{kk'} = 0$ pour tout k et k' différents. Nous noterons désormais simplement $\sigma_{kk} = \sigma_k$, et effectuerons la minimisation sur l'ensemble des matrices diagonalisables dans la même base que ρ .

Notre problème d'optimisation est donc devenu :

$$\arg \min_{\sigma_k \geq 0, \sum_{k=1}^{N_{\mathcal{H}}} \sigma_k = 1} \sum_{k=1}^{N_{\mathcal{H}}} (D_k - \sigma_k)^2 \quad (\text{B.27})$$

Ce problème peut être résolu avec des arguments géométriques, en remarquant que, dans une base euclidienne en $N_{\mathcal{H}}$ dimensions, dans laquelle les coefficients σ_k et D_k sont les coordonnées des points σ et D , et où $\mathcal{S} = \{\sigma / \forall k, \sigma_k \geq 0, \sum_{k=1}^{N_{\mathcal{H}}} \sigma_k = 1\}$, ce problème revient à trouver le point de $\mathcal{S}$ le plus proche de D . Ce point sera noté σ^* . On commence par trouver le point de l'hyperplan $\mathcal{T}$ défini par $\mathcal{T} = \{\sigma / \sum_{k=1}^{N_{\mathcal{H}}} \sigma_k = 1\}$ le plus proche de D . Comme le vecteur normal à cet hyperplan est $\vec{u} = (1, \dots, 1)$, ce point, noté D^* , est donné par :

$$D^* = \mathcal{P}_{\mathcal{T}}(D) = D - \lambda \vec{u}, \quad \text{tel que } \langle D - \lambda \vec{u}, \vec{u} \rangle = 0$$

La valeur du coefficient λ est donc donnée par $\lambda = \sum_{k=1}^{N_{\mathcal{H}}} D_k / N_{\mathcal{H}}$. Nous avons maintenant à minimiser la distance entre σ et D^* . Nous allons successivement modifier D^* , pour le faire parvenir en définitive à σ^* . Pour une dimension k donnée, si $D_k^* \leq 0$, alors nécessairement $\sigma_k^* = 0$. La minimisation peut alors tout aussi bien se réaliser en prenant, au lieu de D^* , sa projection sur l'espace $\{\sigma \in \mathcal{T} / \sigma_k \geq 0\}$, qui se fait grâce au vecteur normal $(1, \dots, 1, -(N_{\mathcal{H}} - 1), 1, \dots, 1)$. Lorsque plus aucune dimension du D^* ainsi itéré n'est strictement négative, ce D^* produit est simplement égal à σ^* .

Le temps de calcul de cet algorithme est au plus en $\mathcal{O}(N_{\mathcal{H}}^2)$, où $N_{\mathcal{H}}$ est la dimension de l'espace de Hilbert considéré. Dans le cadre de la maximisation présentée dans le chapitre 3, l'itération est alors simplement :

$$\hat{\rho}_{j+1} = U_j \text{diag}(\sigma_1^*, \dots, \sigma_{N_{\mathcal{H}}}^*) U_j^\dagger,$$

où $\hat{\rho}_j + \alpha_j \nabla f(\hat{\rho}_j) = U_j D_j U_j^\dagger$, et où D_j sert à obtenir l'ensemble $(\sigma_1^*, \dots, \sigma_{N_{\mathcal{H}}}^*)$.

B.6.2 Descente par algorithme EM

Dans cette sous-section, nous donnons une preuve du théorème 2. En utilisant les notations du théorème, cela revient à montrer que :

$$\forall \rho \in \mathcal{D}_{\mathcal{H}} \text{ diagonale}, \sum_{n=1}^{\Omega} \log \left(\text{Tr} \left(\sum_{m=1}^{\Omega} \frac{\rho E^{(m)}}{\text{Tr}(\rho E^{(m)})} E^{(n)} \right) \right) \geq \sum_{n=1}^{\Omega} \log \left(\text{Tr}(\rho E^{(n)}) \right)$$

Maintenant, nous écrivons les traces de manière explicite, en indexant les dimensions des matrices par q entre 1 et $N_{\mathcal{H}}$. Nous supposons également que nous disposons d'une première estimation de $\hat{\rho}_{ML}$, notée ρ^- , un opérateur diagonal. Il est possible d'introduire la valeur de ρ^- dans l'expression de $f(\rho)$, en procédant de la façon suivante :

$$\begin{aligned}
 f(\rho) &= \sum_{n=1}^{\Omega} \log \left(\text{Tr} \left(\rho E^{(n)} \right) \right) \\
 &= \sum_{n=1}^{\Omega} \left(\frac{1}{\text{Tr}(\rho^- E^{(n)})} \sum_{q=1}^{N_{\mathcal{H}}} \rho_q^- E_q^{(n)} \right) \log \left(\text{Tr} \left(\rho E^{(n)} \right) \right) \\
 &= \sum_{n=1}^{\Omega} \sum_{q=1}^{N_{\mathcal{H}}} \frac{\rho_q^- E_q^{(n)}}{\text{Tr}(\rho^- E^{(n)})} \log \left(\rho_q^- E_q^{(n)} \right) - \sum_{n=1}^{\Omega} \sum_{q=1}^{N_{\mathcal{H}}} \frac{\rho_q^- E_q^{(n)}}{\text{Tr}(\rho^- E^{(n)})} \log \left(\frac{\rho_q^- E_q^{(n)}}{\text{Tr}(\rho E^{(n)})} \right) \\
 &= Q(\rho \mid \rho^-) + H(\rho \mid \rho^-)
 \end{aligned} \tag{B.28}$$

Désormais, nous allons étudier la valeur de la différence $f(\rho) - f(\rho^-)$ et chercher un processus itératif qui nous garantit que $f(\rho) - f(\rho^-) \geq 0$.

Nous remarquons que, pour chaque n , et pour chaque matrice densité diagonale ρ , les deux ensembles $\left(\frac{\rho_q^- E_q^{(n)}}{\text{Tr}(\rho^- E^{(n)})} \right)_{q=1 \dots N_{\mathcal{H}}}$ et $\left(\frac{\rho_q E_q^{(n)}}{\text{Tr}(\rho E^{(n)})} \right)_{q=1 \dots N_{\mathcal{H}}}$ forment deux différents jeux de probabilités discrètes. Nous rappelons que l'inégalité de Gibbs postule que, pour deux distributions finies de probabilités $(p_1, \dots, p_l)$ et $(q_1, \dots, q_l)$, on a :

$$-\sum_{k=1}^l p_k \log(p_k) \leq -\sum_{k=1}^l p_k \log(q_k)$$

Il est alors rapide de trouver que, pour toute matrice densité diagonale ρ , on a :

$$H(\rho \mid \rho^-) - H(\rho^- \mid \rho^-) \geq 0$$

Quant au terme $Q(\rho \mid \rho^-)$, nous cherchons à trouver le ρ qui le maximise. On commence par remarquer que la fonction $\rho \mapsto Q(\rho \mid \rho^-)$ est concave. Ainsi, une matrice densité diagonale ρ^+ en laquelle le gradient de Q serait orthogonal à la variété des matrices densité diagonales (voir l'équation 3.26 adaptée au cas des matrices diagonales) serait l'unique maximisateur de $\rho \mapsto Q(\rho \mid \rho^-)$ sur l'ensemble des matrices densité admissibles. Voyons si un tel ρ^+ existe.

Tout d'abord, le coefficient en q du gradient $\nabla Q|_{\rho}$ est :

$$\left(\nabla Q|_{\rho} \right)_q = \frac{1}{\rho_q} \sum_{n=1}^{\Omega} \frac{\rho_q^- E_q^{(n)}}{\text{Tr}(\rho^- E^{(n)})}$$

Ainsi, puisque $\sum_{n=1}^{\Omega} \frac{\rho_q^- E_q^{(n)}}{\text{Tr}(\rho^- E^{(n)})}$ est positif pour tout q , il suffit de choisir un ρ^+ colinéaire à $\rho^- \sum_{n=1}^{\Omega} \frac{E^{(n)}}{\text{Tr}(\rho^- E^{(n)})}$ pour satisfaire les conditions d'optimalité (sous réserve que la famille $(E^{(n)})$ génère complètement l'espace des opérateurs linéaires, ce qui est le cas en pratique). Le coefficient de normalisation est rapidement trouvé, et en choisissant :

$$\rho^+ = \frac{\rho^-}{\Omega} \left(\sum_{n=1}^{\Omega} \frac{E^{(n)}}{\text{Tr}(\rho^- E^{(n)})} \right), \tag{B.29}$$

on s'assure que $Q(\rho^+ \mid \rho^-) \geq Q(\rho^- \mid \rho^-)$. Ainsi, on a bien prouvé que, pour tout ρ^- diagonale, la matrice diagonale ρ^+ définie par la formule ci-dessus vérifie bien $f(\rho^+) \geq f(\rho^-)$.

Annexe C

Développements asymptotiques de l'estimation Bayésienne

Résumé. L'estimation Bayésienne d'un état quantique mixte a pour valeur approchée l'estimation par Maximum de Vraisemblance (MaxLike), lorsque la fonction de vraisemblance est assez piquée autour de son maximum. De telles approximations s'appuient sur les développements asymptotiques d'intégrales de Laplace multi-dimensionnelles. Lorsque le maximum est sur la frontière du domaine d'intégration, comme dans le cas où l'état estimé par Maximum de Vraisemblance n'est pas inversible, de tels développements ne rentrent plus dans le cadre standard. Nous apportons ici les formules de ces développements, même lorsque le maximum ne repose pas sur la partie C^∞ de la frontière du domaine d'intégration, comme dans le cas où la dimension du noyau est supérieure ou égale à deux. Ces développements apportent aussi un intervalle de confiance pour la valeur moyenne de chaque observable, lorsque celle-ci est calculée par Maximum de Vraisemblance.

C.1 Introduction

Quand les lois de probabilité ayant produit une série de mesures observées Y et dépendant d'un paramètre continu p ont une structure analytique, la technique de reconstruction par Maximum de Vraisemblance est fréquemment utilisée (voir, par exemple, [CMR05]). Cette technique choisit comme estimation $\hat{p}_{ML}$, la valeur de p qui maximise la probabilité conditionnelle $\mathbb{P}(Y | p)$ de la mesure Y . En effet, lorsque Y correspond à un grand nombre de mesures indépendantes, la fonction $p \mapsto \mathbb{P}(Y | p)$ devient très piquée autour de sa valeur maximale, et ainsi, l'estimation par Maximum de Vraisemblance $\hat{p}_{ML}$ devient une bonne approximation de l'estimation par moyenne Bayésienne $\hat{p}_{BM}$:

$$\hat{p}_{BM} = \int_{\mathcal{D}} p \mathbb{P}(p | Y) \, dp = \frac{\int_{\mathcal{D}} p \mathbb{P}(Y | p) \mathbb{P}_0(p) \, dp}{\int_{\mathcal{D}} \mathbb{P}(Y | p) \mathbb{P}_0(p) \, dp},$$

où $\mathcal{D} \subset \mathbb{R}^{\dim p}$ est l'ensemble des valeurs physiquement admissibles de p ; $\mathbb{P}(p | Y)$ est la densité de probabilité de p sachant Y ; et $\mathbb{P}_0(p)$ est une densité de probabilité initiale quelconque pour p .

L'estimation par Maximum de Vraisemblance a l'avantage de reposer sur des algorithmes faciles à mettre en œuvre. Le gradient de $p \mapsto \mathbb{P}(Y | p)$ peut être calculé par différences

finies, et l'optimisation peut se réaliser par des techniques de quasi-Newton, qui donnent en plus une très bonne approximation du Hessien de cette même fonction. La borne de Cramér-Rao, qui peut être déduite du Hessien de la fonction de log-vraisemblance, définit une borne inférieure à l'erreur d'estimation moyenne lorsque cette matrice Hessienne est de rang plein. Néanmoins, certains problèmes peuvent apparaître, en particulier pour la tomographie d'état quantique [PR04], où la valeur à estimer p est un état quantique ρ sur un espace de Hilbert $\mathcal{H}$, un élément de l'espace convexe $\mathcal{D}_{\mathcal{H}}$ qui contient l'ensemble des matrices Hermitiennes positives ou nulles de trace un. En pratique, la valeur de $\hat{\rho}_{ML}$ est souvent de rang faible, donc non inversible. Dans ce cas, $\hat{\rho}_{ML}$ appartient à la frontière de $\mathcal{D}_{\mathcal{H}}$, comme noté dans [BK10] et observé dans [SCID⁺16].

Toutes ces raisons nous ont amené à étudier les estimations par moyenne Bayésienne (BME) dans le cas général où le paramètre p est de dimension finie et évolue dans un domaine compact $\mathcal{D}$, dont la frontière est de classe C^∞ par morceaux. Puisque la valeur de $\mathbb{P}(Y | p)$ croît (ou décroît) exponentiellement par rapport au nombre de mesures indépendantes Ω formant la mesure totale Y , nous considérons plutôt la fonction de log-vraisemblance moyenne $f(p) = \frac{1}{\Omega} \log(\mathbb{P}(Y | p))$. Ensuite, pour toutes fonctions scalaires f et g de classe C^∞ , et en supposant certaines conditions, nous cherchons un développement asymptotique, lorsque Ω tend vers l'infini, de l'intégrale de Laplace :

$$\mathcal{I}_g(\Omega) = \int_{\mathcal{D}} g(p) \exp(\Omega f(p)) \, dp. \quad (\text{C.1})$$

Ces développements asymptotiques sont étudiés de longue date et leur obtention requiert des intégrations par parties, l'utilisation du lemme de Watson, de la méthode de Laplace et de la résolution des singularités de Hironaka : voir [BH86] pour le cas $\dim p = 1$ et dans le cas général où $\dim p \geq 1$; voir aussi [AGZV85] pour le cas singulier en toute dimension, et pour son analyse beaucoup plus élaborée. Dans le cas analytique, et lorsque le maximum de f est atteint en p_{ML} point intérieur de $\mathcal{D}$, ces développements ont un terme dominant en $e^{\Omega f(p_{ML})} \frac{(\log \Omega)^k}{\Omega^\alpha}$, où k est un entier positif ou nul inférieur à $\dim p - 1$ et où α est un rationnel strictement positif [AGZV85, page 231]. Des relations fondamentales entre la géométrie algébrique et la théorie d'apprentissage statistique dans le cas singulier découlent de ces développements asymptotiques, par exemple lorsque le Hessien de f en $\hat{p}_{ML}$ n'est pas négatif défini. C'est l'objet de la théorie d'apprentissage singulier développée dans [Wat09] et [Sha11].

Il est intéressant de noter que, à notre connaissance, très peu de résultats existent lorsque $\hat{p}_{ML}$ repose sur la frontière de l'ensemble $\mathcal{D}$ d'intégration, sauf lorsque cette frontière est de classe C^∞ en p_{ML} . Dans [BH86, section 8.3], le calcul du terme principal du développement est détaillé lorsque la frontière est de classe C^∞ en $\hat{p}_{ML}$ et lorsque le Hessien de la restriction de f à cette frontière est négatif défini ; la sous-section 8.3.4 de [AGZV85] donne des indications précises montrant que, lorsque le Hessien de la restriction de f est dégénéré, un développement asymptotique existe, et est similaire à celui obtenu pour un p_{ML} à l'intérieur de $\mathcal{D}$.

Dans le cadre de l'estimation d'état quantique, cela nous assure qu'un développement asymptotique existe dans le cas où $\hat{\rho}_{ML}$ est de rang plein (il est alors à l'intérieur de $\mathcal{D} = \mathcal{D}_{\mathcal{H}}$), ou lorsque son noyau est de rang 1 (il est cette fois sur une frontière C^∞ de $\mathcal{D}$). Lorsque la dimension du noyau est strictement supérieure à 1, $\hat{\rho}_{ML}$ n'est plus sur une partie C^∞ de la frontière, et, à notre connaissance, le calcul des développements asymptotiques dans un tel cas singulier n'a pas encore été effectué. Ce papier est un premier essai vers le calcul des développements asymptotiques de la moyenne et de la

variance Bayésienne, lorsque la fonction de log-vraisemblance atteint son maximum sur la frontière de $\mathcal{D}$, c'est-à-dire lorsque $\hat{\rho}_{ML}$ est de rang faible.

L'objectif de ce papier est double. Premièrement, nous calculons le terme dominant de certains développements asymptotiques lorsque p parcourt un demi-espace. C'est l'objet de la section C.2, où nous supposons que la restriction de f à la frontière du demi-espace admet un maximum non-dégénéré en $\hat{\rho}_{ML}$ (voir le théorème 8). Ensuite, nous nous intéressons à l'estimation d'état quantique et reformulons les résultats de la première partie en termes de produits matriciels et de trace. C'est l'objet de la section C.3, où nous rappelons les structures de f et g dans ce cas et où nous exploitons la convexité et l'invariance unitaire. Nous donnons dans cette section une démonstration mathématique précise des conditions nécessaires et suffisantes d'optimalité données sans détails dans [SCID⁺16, eq. (8)] (voir le lemme 6 ci-dessous) et de l'expression de l'approximation de la variance Bayésienne correspondant à l'équation (10) dans [SCID⁺16] (voir le théorème 9).

C.2 Développement asymptotique des intégrales de Laplace

Ici, nous supposons que p est de dimension q et que $\mathcal{D} = (-1, 1)^q$. Posons $p = z$ avec $z \in \mathbb{R}^q$. Alors, on peut réécrire la formule (C.1) :

$$\mathcal{I}_g(\Omega) = \int_{z \in (-1, 1)^q} g(z) \exp(\Omega f(z)) \, dz. \quad (\text{C.2})$$

Théorème 7. Considérons l'équation (C.2), où f et g sont des fonctions analytiques de z sur un voisinage compact de $\overline{\mathcal{D}}$, l'adhérence de $\mathcal{D}$. Supposons que la fonction f admet un unique maximum sur $\overline{\mathcal{D}}$ en $z = 0$, avec $\frac{\partial^2 f}{\partial z^2} \Big|_0$ définie négative.

Lorsque $g(0) \neq 0$, on trouve le terme dominant suivant dans le développement asymptotique de $\mathcal{I}_g(\Omega)$ lorsque Ω tend vers l'infini :

$$\mathcal{I}_g(\Omega) = \left(\frac{g(0) (2\pi)^{n/2} e^{\Omega f(0)} \Omega^{-n/2}}{\sqrt{|\det \left(\frac{\partial^2 f}{\partial z^2} \Big|_0 \right)|}} \right) + O\left(e^{\Omega f(0)} \Omega^{-n/2-1}\right). \quad (\text{C.3})$$

Lorsque $g(0) = 0$, avec $\frac{\partial g}{\partial x} \Big|_0 = 0$ et $\frac{\partial g}{\partial z} \Big|_0 = 0$, on a :

$$\mathcal{I}_g(\Omega) = \left(\frac{\text{Tr} \left(- \frac{\partial^2 g}{\partial z^2} \Big|_0 \left(\frac{\partial^2 f}{\partial z^2} \Big|_0 \right)^{-1} \right) (2\pi)^{n/2}}{2 \sqrt{|\det \left(\frac{\partial^2 f}{\partial z^2} \Big|_0 \right)|}} \right) e^{\Omega f(0)} \Omega^{-n/2-1} + O\left(e^{\Omega f(0)} \Omega^{-n/2-2}\right). \quad (\text{C.4})$$

Démonstration. La fonction f étant analytique, on peut écrire $f(z) = f(0) - h(z)$, où h est une fonction analytique de z seulement, avec $h(0) = 0$, $\frac{\partial h}{\partial z} \Big|_0 = 0$ et $\frac{\partial^2 h}{\partial z^2} \Big|_0 = - \frac{\partial^2 f}{\partial z^2} \Big|_0$ définie négative.

Grâce au lemme de Morse (voir, par exemple, [Mil63]), on sait qu'il existe un difféomorphisme local sur z dans un voisinage de 0, noté $\tilde{z} = \psi(z)$, tel que $\psi(0) = 0$ et $h(z) = \frac{1}{2} \sum_{k=1}^n (\psi_k(z))^2$. De plus, on peut choisir ψ tel que $\frac{\partial \psi}{\partial z} \Big|_0 = \sqrt{- \frac{\partial^2 f}{\partial z^2} \Big|_0}$ est une matrice positive définie (donc symétrique).

Pour $\eta \in (0, 1)$ petit, il existe un $c < f(0)$ tel que, $\forall z \in (-1, 1)^n / (-\eta, \eta)^n$, $f(z) \leq c$. Par conséquent, on peut réécrire :

$$\begin{aligned} \mathcal{I}_g(\Omega) &= \int_{z \in (-\eta, \eta)^n} g(z) e^{\Omega f(z)} dz + \int_{z \in (-1, 1)^n / (-\eta, \eta)^n} g(z) e^{\Omega f(z)} dz \\ &= e^{\Omega f(0)} \left(\int_{z \in (-\eta, \eta)^n} g(z) e^{\Omega(f(z) - f(0))} dz + O\left(e^{-\Omega(f(0) - c)}\right) \right). \end{aligned}$$

Nous considérons donc seulement :

$$I_\eta(\Omega) = \int_{z \in (-\eta, \eta)^n} g(z) e^{\Omega(f(z) - f(0))} dz.$$

Puisque η est petit, nous pouvons introduire le changement de variables $\tilde{z} = \psi(z)$ qui amène :

$$I_\eta(\Omega) = \int_{\tilde{z} \in \psi((-\eta, \eta)^n)} \tilde{g}(\tilde{z}) e^{-\frac{\Omega}{2} \sum_{k=1}^n \tilde{z}_k^2} d\tilde{z},$$

où :

$$\tilde{g}(\tilde{z}) = \frac{g(\psi^{-1}(\tilde{z}))}{\sqrt{\left| \det \left(\frac{\partial^2 f}{\partial z^2} \Big|_0 \right) \right|}} (1 + \tilde{d}(\tilde{z})) \quad (\text{C.5})$$

et $\tilde{d}$ une fonction analytique vérifiant $\tilde{d}(0) = 0$. Il existe un $\tilde{\eta} > 0$ tel que $(-\tilde{\eta}, \tilde{\eta})^n \subset \psi((-\eta, \eta)^n)$. Par conséquent, de la même façon que lors du passage de $\mathcal{I}_g(\Omega)$ à $I_\eta(\Omega)$, on peut, en omettant des termes exponentiellement petits en Ω , ne considérer que le développement asymptotique de :

$$\tilde{I}_{\tilde{\eta}} = \int_{\tilde{z} \in (-\tilde{\eta}, \tilde{\eta})^n} \tilde{g}(\tilde{z}) e^{-\frac{\Omega}{2} \sum_{k=1}^n \tilde{z}_k^2} d\tilde{z}. \quad (\text{C.6})$$

Lorsque $g(0) \neq 0$, on a aussi $\tilde{g}(0) \neq 0$. Posant $\tilde{g}(\tilde{z}) = \tilde{g}(0) + \sum_{k=1}^n \tilde{z}_k \tilde{h}_k(\tilde{z})$ avec les $\tilde{h}_k$ étant des fonctions analytiques bornées sur $(-\tilde{\eta}, \tilde{\eta})^n$, on obtient :

$$\tilde{I}_{\tilde{\eta}} = \tilde{g}(0) \int_{\tilde{z} \in (-\tilde{\eta}, \tilde{\eta})^n} e^{-\frac{\Omega}{2} \sum_{k=1}^n \tilde{z}_k^2} d\tilde{z} + \int_{\tilde{z} \in (-\tilde{\eta}, \tilde{\eta})^n} \left(\sum_{k=1}^n \tilde{z}_k \tilde{h}_k(\tilde{z}) \right) e^{-\frac{\Omega}{2} \sum_{k=1}^n \tilde{z}_k^2} d\tilde{z}.$$

En omettant des termes exponentiellement petits en Ω , la première intégrale du terme de droite peut être remplacée par :

$$\int_{\tilde{z} \in (-\infty, +\infty)^n} e^{-\frac{\Omega}{2} \sum_{k=1}^n \tilde{z}_k^2} d\tilde{z} = \left(\frac{2\pi}{\Omega} \right)^{n/2}.$$

Une simple intégration par parties sur z_k permet d'obtenir :

$$\begin{aligned} \int_{\tilde{z} \in (-\tilde{\eta}, \tilde{\eta})^n} \tilde{z}_k \tilde{h}_k(\tilde{z}) e^{-\frac{\Omega}{2} \sum_{k=1}^n \tilde{z}_k^2} d\tilde{z} \\ = \frac{1}{\Omega} \int_{\tilde{z} \in (-\tilde{\eta}, \tilde{\eta})^n} \frac{\partial \tilde{h}_k}{\partial \tilde{z}_k}(\tilde{z}) e^{-\frac{\Omega}{2} \sum_{k=1}^n \tilde{z}_k^2} d\tilde{z} + O(e^{-\tilde{\eta}^2 \Omega / 2} / \Omega). \end{aligned}$$

Ceci nous amène à formuler (C.3), en utilisant le fait que $\tilde{I}_{\tilde{\eta}} = \tilde{g}(0) \left(\frac{2\pi}{\Omega} \right)^{n/2} (1 + O(1/\Omega))$ et $\tilde{g}(0) = \frac{g(0)}{\sqrt{\left| \det \left(\frac{\partial^2 f}{\partial z^2} \Big|_0 \right) \right|}}.$

Supposons maintenant que $g(0) = 0$ et que $\frac{\partial g}{\partial z}\big|_0 = 0$, et considérons alors la fonction $\tilde{g}$ présente dans (C.5). Nous avons alors $\tilde{g}(0) = 0$ et $\frac{\partial \tilde{g}}{\partial \tilde{z}}\big|_0 = 0$. De plus, en écrivant :

$$\kappa_0 = \sqrt{\left| \det \left(\frac{\partial^2 f}{\partial z^2} \right) \right|_0},$$

on obtient :

$$\kappa_0 \tilde{g}(\psi(z)) = g(z) (1 + e(z))$$

où e est une fonction analytique telle que $e(0) = 0$. Ainsi, pour tous $i, j \in \{1, \dots, n\}$,

$$\frac{\partial^2 g}{\partial z_i \partial z_j} \bigg|_0 = \kappa_0 \sum_{k, k'=1}^n \frac{\partial^2 \tilde{g}}{\partial \tilde{z}_k \partial \tilde{z}_{k'}} \bigg|_0 \frac{\partial \psi_k}{\partial z_i} \bigg|_0 \frac{\partial \psi_{k'}}{\partial z_j} \bigg|_0.$$

Puisque $\frac{\partial \psi}{\partial z} \big|_0 = \sqrt{-\frac{\partial^2 f}{\partial z^2}} \big|_0$, on peut écrire que :

$$\kappa_0 \frac{\partial^2 \tilde{g}}{\partial \tilde{z}^2} \bigg|_0 = \left(\sqrt{-\frac{\partial^2 f}{\partial z^2}} \bigg|_0 \right)^{-1} \frac{\partial^2 g}{\partial z^2} \bigg|_0 \left(\sqrt{-\frac{\partial^2 f}{\partial z^2}} \bigg|_0 \right)^{-1},$$

et ainsi :

$$\mathrm{Tr} \left(\frac{\partial^2 \tilde{g}}{\partial \tilde{z}^2} \bigg|_0 \right) = \frac{\mathrm{Tr} \left(-\frac{\partial^2 g}{\partial z^2} \bigg|_0 \left(\frac{\partial^2 f}{\partial z^2} \bigg|_0 \right)^{-1} \right)}{\sqrt{\left| \det \left(\frac{\partial^2 f}{\partial z^2} \right) \right|_0}} \quad (\text{C.7})$$

$\tilde{g}$ et ses dérivées premières par rapport à $\tilde{z}_k$ étant nulles, nous avons :

$$\tilde{g}(\tilde{z}) = \sum_{k, k'=1}^n \tilde{z}_k \tilde{z}_{k'} \tilde{b}_{k, k'}(\tilde{z}),$$

où la fonction $\tilde{b}_{k, k'}$ est analytique. Pour donner une formule à l'intégrale de l'équation (C.6), nous devons chercher les termes dominants des intégrales suivantes :

$$B_{k, k'} = \int_{\tilde{z} \in (-\tilde{\eta}, \tilde{\eta})^n} \tilde{z}_k \tilde{z}_{k'} \tilde{b}_{k, k'}(\tilde{z}) e^{-\frac{\Omega}{2} \sum_{l=1}^n \tilde{z}_l^2} d\tilde{z}.$$

Dès que $k \neq k'$, une intégration par parties par rapport à $\tilde{z}_k$, suivie d'une autre par rapport à $\tilde{z}_{k'}$, nous montre que $B_{k, k'} = O(\Omega^{-n/2-2})$. Si $k = k'$, on peut ne réaliser qu'une seule intégration par parties, par rapport à $\tilde{z}_k$:

$$\begin{aligned} & \int_{\tilde{z} \in (-\tilde{\eta}, \tilde{\eta})^n} \tilde{z}_k^2 \tilde{b}_{k, k}(\tilde{z}) e^{-\frac{\Omega}{2} \sum_{l=1}^n \tilde{z}_l^2} d\tilde{z} \\ &= \frac{1}{\Omega} \int_{\tilde{z} \in (-\tilde{\eta}, \tilde{\eta})^n} \left(\tilde{b}_{k, k}(\tilde{z}) + \tilde{z}_k \frac{\partial \tilde{b}_{k, k}}{\partial \tilde{z}_k}(\tilde{z}) \right) e^{-\frac{\Omega}{2} \sum_{l=1}^n \tilde{z}_l^2} d\tilde{z} + O(e^{-\Omega \tilde{\eta}^2/2}) \\ &= \frac{\tilde{b}_{k, k}(0)}{\Omega} \left(\frac{2\pi}{\Omega} \right)^{n/2} + O(\Omega^{-n/2-2}). \end{aligned}$$

La somme $\sum_{k, k'} B_{k, k'}$ correspond à l'intégrale $\tilde{I}_{\tilde{\eta}}$ et devient :

$$\tilde{I}_{\tilde{\eta}}(\Omega) = \frac{\sum_{k=1}^n \tilde{b}_{k, k}(0)}{\Omega} \left(\frac{2\pi}{\Omega} \right)^{n/2} + O(\Omega^{-n/2-2}).$$

Puisque $\tilde{I}_{\tilde{\eta}}$ et $e^{-\Omega f(0)} \mathcal{I}_g(\Omega)$ sont égaux, à des termes exponentiellement petits sur Ω , nous obtenons finalement l'équation (C.4) en utilisant le résultat (C.7), et $\sum_{k=1}^n \tilde{b}_{k, k}(0) = \frac{1}{2} \mathrm{Tr} \left(\frac{\partial^2 \tilde{g}}{\partial \tilde{z}^2} \bigg|_0 \right)$. ■

Nous supposons désormais que $p \in \mathbb{R}^{n+1}$, $n+1$ étant la dimension de p (n un entier positif ou nul), et que l'espace d'intégration $\mathcal{D} = (0, 1) \times (-1, 1)^n$. Posons $p = (x, z)$, avec $x \in \mathbb{R}$ et $z \in \mathbb{R}^n$. Ainsi, lorsque $g(x, z)$ est remplacé par $x^m g(x, z)$, où m est un entier positif ou nul, (C.1) devient :

$$\mathcal{I}_g(\Omega) = \int_{x \in (0,1)} \int_{z \in (-1,1)^n} x^m g(x, z) \exp(\Omega f(x, z)) \, dx \, dz. \quad (\text{C.8})$$

Théorème 8. Considérons (C.8), où f et g sont des fonctions analytiques (x, z) sur un voisinage compact de $\overline{\mathcal{D}}$, l'adhérence de $\mathcal{D}$. On suppose que f admet un unique maximum sur $\overline{\mathcal{D}}$ en $(x, z) = (0, 0)$, avec $\frac{\partial^2 f}{\partial z^2} \Big|_{(0,0)}$ défini négatif et $\frac{\partial f}{\partial x} \Big|_{(0,0)} < 0$.

Si $g(0, 0) \neq 0$, le terme dominant du développement asymptotique de $\mathcal{I}_g(\Omega)$, lorsque Ω tend vers l'infini, est :

$$\mathcal{I}_g(\Omega) = \left(\frac{g(0, 0) \, m! \, (2\pi)^{n/2} \, e^{\Omega f(0,0)} \Omega^{-m-n/2-1}}{\sqrt{\left| \det \left(\frac{\partial^2 f}{\partial z^2} \Big|_{(0,0)} \right) \right|} \left(-\frac{\partial f}{\partial x} \Big|_{(0,0)} \right)^{m+1}} \right) + O\left(e^{\Omega f(0,0)} \Omega^{-m-n/2-2} \right). \quad (\text{C.9})$$

Si $g(0, 0) = 0$, avec $\frac{\partial g}{\partial x} \Big|_{(0,0)} = 0$ et $\frac{\partial g}{\partial z} \Big|_{(0,0)} = 0$, on a :

$$\mathcal{I}_g(\Omega) = \left(\frac{\text{Tr} \left(-\frac{\partial^2 g}{\partial z^2} \Big|_{(0,0)} \left(\frac{\partial^2 f}{\partial z^2} \Big|_{(0,0)} \right)^{-1} \right) \, m! \, (2\pi)^{n/2}}{2 \sqrt{\left| \det \left(\frac{\partial^2 f}{\partial z^2} \Big|_{(0,0)} \right) \right|} \left(-\frac{\partial f}{\partial x} \Big|_{(0,0)} \right)^{m+1}} \right) e^{\Omega f(0,0)} \Omega^{-m-n/2-2} + O\left(e^{\Omega f(0,0)} \Omega^{-m-n/2-3} \right). \quad (\text{C.10})$$

Pour simplifier les formules, nous considérons ici seulement le cas analytique, même si les équivalences développées dans le théorème sont également valides dans le cas où les fonctions impliquées sont de classe C^{m+3} .

Démonstration. Dans cette preuve, nous adaptons la méthode exposée en section 8.3.4 de [AGZV85] pour les intégrales oscillantes dans un demi-espace. La fonction f étant analytique, nous avons

$$f(x, z) = f(0, 0) - x f_1(x, z) - h(z),$$

où f_1 est analytique avec $f_1(0, 0) = -\frac{\partial f}{\partial x} \Big|_{(0,0)} > 0$ et où h est une fonction analytique de z seulement, avec $h(0) = 0$, $\frac{\partial h}{\partial z} \Big|_0 = 0$ et $\frac{\partial^2 h}{\partial z^2} \Big|_0 = -\frac{\partial^2 f}{\partial z^2} \Big|_{(0,0)}$ défini positif.

Posons $\phi(x, z) = x f_1(x, z)$. Considérons l'application $(x, z) \mapsto (\tilde{x} = \phi(x, z), z)$. Il s'agit d'un difféomorphisme local sur un voisinage de $(0, 0)$ qui préserve le signe de x , c'est-à-dire tel que $x \phi(x, z) \geq 0$. De plus, le lemme de Morse (see, e.g., [Mil63]) nous montre qu'il existe un difféomorphisme local sur z sur un voisinage de 0, noté $\tilde{z} = \psi(z)$, tel que $\psi(0) = 0$ et $h(z) = \frac{1}{2} \sum_{k=1}^n (\psi_k(z))^2$ (voir, par exemple, [Mil63]). De plus, on peut choisir ψ tel que $\frac{\partial \psi}{\partial z} \Big|_0 = \sqrt{-\frac{\partial^2 f}{\partial z^2} \Big|_0}$ est une matrice positive définie (donc symétrique).

En résumé, il existe localement un difféomorphisme analytique $\Xi : V \ni (x, z) \mapsto (\tilde{x}, \tilde{z}) \in \tilde{V}$, depuis un voisinage ouvert et connexe par arcs V de 0 jusqu'à un autre voisinage ouvert de 0 tel que :

- pour tout $(x, z) \in V$, on a $\phi(x, z) > 0$ (resp. < 0 , $= 0$) quand $x > 0$ (resp. < 0 , $= 0$).
- $\forall (x, z) \in V$, $f(x, z) = -\phi(x, z) - \frac{1}{2} \sum_{k=1}^n (\psi_k(z))^2$.
- $\det \begin{pmatrix} \frac{\partial \phi}{\partial x} & \frac{\partial \phi}{\partial z} \\ \frac{\partial \psi}{\partial x} & \frac{\partial \psi}{\partial z} \end{pmatrix} \Big|_{(x,z)} = \left| \frac{\partial f}{\partial x} \Big|_{(0,0)} \right| \sqrt{\left| \det \begin{pmatrix} \frac{\partial^2 f}{\partial z^2} \Big|_{(0,0)} \end{pmatrix} \right|} (1 + d(x, z))$ où d est une fonction analytique sur V vérifiant $d(0, 0) = 0$.

Puisque V est un voisinage de 0, il existe un $\eta \in (0, 1)$ tel que $\mathcal{C}_\eta = (0, \eta) \times (-\eta, \eta)^n \subset V$. De plus, il existe $c < f(0, 0)$ tel que, $\forall (x, z) \in \mathcal{D}/\mathcal{C}_\eta$, $f(x, z) \leq c$. Par conséquent, puisque :

$$\begin{aligned} \mathcal{I}_g(\Omega) &= \int_{(x,z) \in \mathcal{C}_\eta} x^m g(x, z) e^{\Omega f(x, z)} dx dz + \int_{(x,z) \in \mathcal{D}/\mathcal{C}_\eta} x^m g(x, z) e^{\Omega f(x, z)} dx dz \\ &= e^{\Omega f(0,0)} \left(\int_{(x,z) \in \mathcal{C}_\eta} x^m g(x, z) e^{\Omega(f(x, z) - f(0,0))} dx dz + e^{-\Omega(f(0,0) - c)} \int_{(x,z) \in \mathcal{D}/\mathcal{C}_\eta} x^m g(x, z) e^{\Omega(f(x, z) - c)} dx dz \right) \\ &= e^{\Omega f(0,0)} \left(\int_{(x,z) \in \mathcal{C}_\eta} x^m g(x, z) e^{\Omega(f(x, z) - f(0,0))} dx dz + O\left(e^{-\Omega(f(0,0) - c)}\right) \right), \end{aligned}$$

nous avons simplement à étudier le développement asymptotique de :

$$I_\eta(\Omega) = \int_{(x,z) \in \mathcal{C}_\eta} x^m g(x, z) e^{\Omega(f(x, z) - f(0,0))} dx dz.$$

Puisque $\mathcal{C}_\eta \subset V$, on peut utiliser le changement de variables $(\tilde{x}, \tilde{z}) = \Xi(x, z)$ qui entraîne :

$$I_\eta(\Omega) = \int_{(\tilde{x}, \tilde{z}) \in \Xi(\mathcal{C}_\eta)} \tilde{x}^m \tilde{g}(\tilde{x}, \tilde{z}) e^{-\Omega(\tilde{x} + \frac{1}{2} \sum_{k=1}^n \tilde{z}_k^2)} d\tilde{x} d\tilde{z},$$

où :

$$\tilde{g}(\tilde{x}, \tilde{z}) = \frac{g(\Xi^{-1}(\tilde{x}, \tilde{z}))}{\left(f_1(\Xi^{-1}(\tilde{x}, \tilde{z}))\right)^m \left| \frac{\partial f}{\partial x} \Big|_{(0,0)} \right| \sqrt{\left| \det \begin{pmatrix} \frac{\partial^2 f}{\partial z^2} \Big|_{(0,0)} \end{pmatrix} \right|}} (1 + \tilde{d}(\tilde{x}, \tilde{z}))$$

et où $\tilde{d}$ est une fonction analytique avec $\tilde{d}(0, 0) = 0$. Puisque, pour tout $(\tilde{x}, \tilde{z}) \in \Xi(\mathcal{C}_\eta)$, nous avons $\tilde{x} \geq 0$, il existe un $\tilde{\eta} > 0$ tel que $\tilde{\mathcal{C}}_{\tilde{\eta}} = (0, \tilde{\eta}) \times (-\tilde{\eta}, \tilde{\eta})^n \subset \Xi(\mathcal{C}_\eta)$. Ainsi, de la même façon que pour passer de $\mathcal{I}_g(\Omega)$ à $I_\eta(\Omega)$, à des termes exponentiellement faibles en Ω près, nous pouvons étudier seulement le développement asymptotique de :

$$\tilde{I}_{\tilde{\eta}} = \int_{(\tilde{x}, \tilde{z}) \in \tilde{\mathcal{C}}_{\tilde{\eta}}} \tilde{x}^m \tilde{g}(\tilde{x}, \tilde{z}) e^{-\Omega(\tilde{x} + \frac{1}{2} \sum_{k=1}^n \tilde{z}_k^2)} d\tilde{x} d\tilde{z}. \quad (\text{C.11})$$

Quand $g(0, 0) \neq 0$, on a $\tilde{g}(0, 0) \neq 0$. Posons $\tilde{g}(\tilde{x}, \tilde{z}) = \tilde{g}(0, 0) + \tilde{x} \tilde{g}_1(\tilde{x}, \tilde{z}) + \sum_{k=1}^n \tilde{z}_k \tilde{h}_k(\tilde{x}, \tilde{z})$ avec $\tilde{g}_1$ et $\tilde{h}_k$ des fonctions analytiques bornées sur $\tilde{\mathcal{C}}_{\tilde{\eta}}$. Nous obtenons :

$$\begin{aligned} \tilde{I}_{\tilde{\eta}} &= \tilde{g}(0, 0) \int_{(\tilde{x}, \tilde{z}) \in \tilde{\mathcal{C}}_{\tilde{\eta}}} \tilde{x}^m e^{-\Omega(\tilde{x} + \frac{1}{2} \sum_{k=1}^n \tilde{z}_k^2)} d\tilde{x} d\tilde{z} \\ &\quad + \int_{(\tilde{x}, \tilde{z}) \in \tilde{\mathcal{C}}_{\tilde{\eta}}} \tilde{x}^{m+1} \tilde{g}_1(\tilde{x}, \tilde{z}) e^{-\Omega(\tilde{x} + \frac{1}{2} \sum_{k=1}^n \tilde{z}_k^2)} d\tilde{x} d\tilde{z} \\ &\quad + \int_{(\tilde{x}, \tilde{z}) \in \tilde{\mathcal{C}}_{\tilde{\eta}}} \tilde{x}^m \left(\sum_{k=1}^n \tilde{z}_k \tilde{h}_k(\tilde{x}, \tilde{z}) \right) e^{-\Omega(\tilde{x} + \frac{1}{2} \sum_{k=1}^n \tilde{z}_k^2)} d\tilde{x} d\tilde{z}. \end{aligned}$$

A des termes exponentiellement petits en Ω près, la première intégrale du terme de droite de l'équation ci-dessus peut être remplacé par :

$$\int_{(\tilde{x}, \tilde{z}) \in (0, +\infty) \times (-\infty, +\infty)^n} \tilde{x}^m e^{-\Omega(\tilde{x} + \frac{1}{2} \sum_{k=1}^n \tilde{z}_k^2)} d\tilde{x} d\tilde{z} = \frac{m!}{\Omega^{m+1}} \left(\frac{2\pi}{\Omega} \right)^{n/2}.$$

En ce qui concerne la deuxième intégrale, $m + 1$ intégrations par parties par rapport à $\tilde{x}$ sont nécessaires :

$$\begin{aligned} \int_{(\tilde{x}, \tilde{z}) \in \tilde{\mathcal{C}}_{\tilde{\eta}}} \tilde{x}^{m+1} \tilde{g}_1(\tilde{x}, \tilde{z}) e^{-\Omega(\tilde{x} + \frac{1}{2} \sum_{k=1}^n \tilde{z}_k^2)} d\tilde{x} d\tilde{z} \\ = \int_{\tilde{z} \in (-\tilde{\eta}, \tilde{\eta})^n} \left(\int_0^{\tilde{\eta}} \tilde{x}^{m+1} \tilde{g}_1(\tilde{x}, \tilde{z}) e^{-\Omega\tilde{x}} d\tilde{x} \right) e^{-\frac{\Omega}{2} \sum_{k=1}^n \tilde{z}_k^2} d\tilde{z}, \end{aligned}$$

et ces $m + 1$ intégrations par parties donnent :

$$\int_0^{\tilde{\eta}} \tilde{x}^{m+1} \tilde{g}_1(\tilde{x}, \tilde{z}) e^{-\Omega\tilde{x}} d\tilde{x} = \frac{1}{\Omega^{m+1}} \int_0^{\tilde{\eta}} \tilde{g}_{m+2}(\tilde{x}, \tilde{z}) e^{-\Omega\tilde{x}} d\tilde{x} + O(e^{-\tilde{\eta}\Omega}/\Omega)$$

avec $\tilde{g}_{m+2} = \frac{\partial^{m+1}}{\partial \tilde{x}^{m+1}} (\tilde{x}^{m+1} \tilde{g}_1(\tilde{x}, \tilde{z}))$. Nous obtenons :

$$\begin{aligned} \int_{(\tilde{x}, \tilde{z}) \in \tilde{\mathcal{C}}_{\tilde{\eta}}} \tilde{x}^{m+1} \tilde{g}_1(\tilde{x}, \tilde{z}) e^{-\Omega(\tilde{x} + \frac{1}{2} \sum_{k=1}^n \tilde{z}_k^2)} d\tilde{x} d\tilde{z} \\ = \frac{1}{\Omega^{m+1}} \int_{(\tilde{x}, \tilde{z}) \in \tilde{\mathcal{C}}_{\tilde{\eta}}} \tilde{g}_{m+2}(\tilde{x}, \tilde{z}) e^{-\Omega(\tilde{x} + \frac{1}{2} \sum_{k=1}^n \tilde{z}_k^2)} d\tilde{x} d\tilde{z} + O(e^{-\tilde{\eta}\Omega}/\Omega) \\ = O\left(\frac{1}{\Omega^{m+n/2+2}}\right), \end{aligned}$$

puisque $\int_0^{\tilde{\eta}} \tilde{g}_{m+2}(\tilde{x}, \tilde{z}) e^{-\Omega\tilde{x}} d\tilde{x}$ est de l'ordre de $1/\Omega$.

De la même façon, m intégrations par parties par rapport à $\tilde{x}$ donnent :

$$\begin{aligned} \int_{(\tilde{x}, \tilde{z}) \in \tilde{\mathcal{C}}_{\tilde{\eta}}} \tilde{x}^m \tilde{z}_k \tilde{h}_k(\tilde{x}, \tilde{z}) e^{-\Omega(\tilde{x} + \frac{1}{2} \sum_{l=1}^n \tilde{z}_l^2)} d\tilde{x} d\tilde{z} \\ = \frac{1}{\Omega^m} \int_{(\tilde{x}, \tilde{z}) \in \tilde{\mathcal{C}}_{\tilde{\eta}}} \tilde{z}_k \tilde{q}_{k,m}(\tilde{x}, \tilde{z}) e^{-\Omega(\tilde{x} + \frac{1}{2} \sum_{l=1}^n \tilde{z}_l^2)} d\tilde{x} d\tilde{z} + O(e^{-\tilde{\eta}\Omega}/\Omega), \end{aligned}$$

où $\tilde{q}_{k,m}(\tilde{x}, \tilde{z}) = \frac{\partial^m}{\partial \tilde{x}^m} (\tilde{x}^m \tilde{h}_k(\tilde{x}, \tilde{z}))$. Une seule intégration par parties par rapport à $\tilde{z}_k$ entraîne :

$$\begin{aligned} \int_{(\tilde{x}, \tilde{z}) \in \tilde{\mathcal{C}}_{\tilde{\eta}}} \tilde{z}_k \tilde{q}_{k,m}(\tilde{x}, \tilde{z}) e^{-\Omega(\tilde{x} + \frac{1}{2} \sum_{l=1}^n \tilde{z}_l^2)} d\tilde{x} d\tilde{z} \\ = \frac{1}{\Omega} \int_{(\tilde{x}, \tilde{z}) \in \tilde{\mathcal{C}}_{\tilde{\eta}}} \frac{\partial \tilde{q}_{k,m}}{\partial \tilde{z}_k}(\tilde{x}, \tilde{z}) e^{-\Omega(\tilde{x} + \frac{1}{2} \sum_{l=1}^n \tilde{z}_l^2)} d\tilde{x} d\tilde{z} + O(e^{-\tilde{\eta}^2\Omega/2}/\Omega). \end{aligned}$$

Cela implique que :

$$\begin{aligned} \int_{(\tilde{x}, \tilde{z}) \in \tilde{\mathcal{C}}_{\tilde{\eta}}} \tilde{x}^m \tilde{z}_k \tilde{h}_k(\tilde{x}, \tilde{z}) e^{-\Omega(\tilde{x} + \frac{1}{2} \sum_{kl=1}^n \tilde{z}_l^2)} d\tilde{x} d\tilde{z} \\ = \frac{1}{\Omega^{m+1}} \int_{(\tilde{x}, \tilde{z}) \in \tilde{\mathcal{C}}_{\tilde{\eta}}} \frac{\partial \tilde{r}_{k,m}}{\partial \tilde{z}_k}(\tilde{x}, \tilde{z}) e^{-\Omega(\tilde{x} + \frac{1}{2} \sum_{l=1}^n \tilde{z}_l^2)} d\tilde{x} d\tilde{z} + O(e^{-\tilde{\eta}^2\Omega/2}/\Omega) \\ = O\left(\frac{1}{\Omega^{m+n/2+2}}\right). \end{aligned}$$

Ainsi, nous obtenons finalement (C.9), puisque $\tilde{I}_{\tilde{\eta}} = \frac{\tilde{g}(0,0)m!}{\Omega^{m+1}} \left(\frac{2\pi}{\Omega}\right)^{n/2} (1 + O(1/\Omega))$ and

$$\tilde{g}(0,0) = \frac{g(0,0)}{\left(\left|\frac{\partial f}{\partial x}\right|_{(0,0)}\right)^{m+1} \sqrt{\left|\det\left(\frac{\partial^2 f}{\partial z^2}\right|_{(0,0)}\right)}}.$$

En supposant maintenant que $g(0,0) = 0$, $\frac{\partial g}{\partial x}\big|_{(0,0)} = 0$ et que $\frac{\partial g}{\partial z}\big|_{(0,0)} = 0$, et en considérant la fonction $\tilde{g}$ de l'équation (C.11), on a $\tilde{g}(0,0) = 0$, $\frac{\partial \tilde{g}}{\partial \tilde{x}}\big|_{(0,0)} = 0$ and $\frac{\partial \tilde{g}}{\partial \tilde{z}}\big|_{(0,0)} = 0$. De plus, en notant :

$$\lambda_0 = \sqrt{\left| \det \left(\frac{\partial^2 f}{\partial z^2} \bigg|_{(0,0)} \right) \right|} \left(- \frac{\partial f}{\partial x} \bigg|_{(0,0)} \right)^{m+1},$$

on a :

$$\lambda_0 \tilde{g}(\phi(x, z), \psi(z)) = g(x, z) (1 + e(x, z)),$$

où e est une fonction analytique telle que $e(0,0) = 0$. Comme dans (C.7), nous obtenons :

$$\mathrm{Tr} \left(\frac{\partial^2 \tilde{g}}{\partial \tilde{z}^2} \bigg|_{(0,0)} \right) = \frac{\mathrm{Tr} \left(- \frac{\partial^2 g}{\partial z^2} \bigg|_{(0,0)} \left(\frac{\partial^2 f}{\partial z^2} \bigg|_{(0,0)} \right)^{-1} \right)}{\sqrt{\left| \det \left(\frac{\partial^2 f}{\partial z^2} \bigg|_{(0,0)} \right) \right|} \left(- \frac{\partial f}{\partial x} \bigg|_{(0,0)} \right)^{m+1}}. \quad (\text{C.12})$$

Puisque $\tilde{g}$ et ses dérivées premières par rapport à $\tilde{x}$ et $\tilde{z}_k$ sont nulles, on a :

$$\tilde{g}(\tilde{x}, \tilde{z}) = \tilde{x}^2 \tilde{a}(\tilde{x}, \tilde{z}) + \sum_{k,k'=1}^n \tilde{z}_k \tilde{z}_{k'} \tilde{b}_{k,k'}(\tilde{x}, \tilde{z}) + \sum_{k=1}^n \tilde{x} \tilde{z}_k \tilde{c}_k(\tilde{x}, \tilde{z}),$$

où les fonctions $\tilde{a}$, $\tilde{b}_{k,k'}$ et $\tilde{c}_k$ sont analytiques. Pour donner une formule à l'intégrale dans (C.11), nous devons étudier les termes dominants des trois types d'intégrales suivants :

$$\begin{aligned} A &= \int_{(\tilde{x}, \tilde{z}) \in \tilde{\mathcal{C}}_{\tilde{\eta}}} \tilde{x}^{m+2} \tilde{a}(\tilde{x}, \tilde{z}) e^{-\Omega(\tilde{x} + \frac{1}{2} \sum_{l=1}^n \tilde{z}_l^2)} d\tilde{x} d\tilde{z}, \\ B_{k,k'} &= \int_{(\tilde{x}, \tilde{z}) \in \tilde{\mathcal{C}}_{\tilde{\eta}}} \tilde{x}^m \tilde{z}_k \tilde{z}_{k'} \tilde{b}_{k,k'}(\tilde{x}, \tilde{z}) e^{-\Omega(\tilde{x} + \frac{1}{2} \sum_{l=1}^n \tilde{z}_l^2)} d\tilde{x} d\tilde{z}, \\ C_k &= \int_{(\tilde{x}, \tilde{z}) \in \tilde{\mathcal{C}}_{\tilde{\eta}}} \tilde{x}^{m+1} \tilde{z}_k \tilde{c}_k(\tilde{x}, \tilde{z}) e^{-\Omega(\tilde{x} + \frac{1}{2} \sum_{l=1}^n \tilde{z}_l^2)} d\tilde{x} d\tilde{z}. \end{aligned}$$

Comme nous l'avons fait précédemment, $m+2$ intégrations par parties par rapport à $\tilde{x}$ montrent que $A = O(\Omega^{-m-n/2-3})$. De même, $m+1$ intégrations par parties par rapport à $\tilde{x}$ et une unique intégration par parties par rapport à $\tilde{z}_k$ donne $C_k = O(\Omega^{-m-n/2-3})$. Lorsque $k \neq k'$, m intégrations par parties par rapport à $\tilde{x}$, une intégration par parties par rapport à $\tilde{z}_k$ et une autre par rapport à $\tilde{z}_{k'}$ amènent à une expression du type $B_{k,k'} = O(\Omega^{-m-n/2-3})$. Pour $k = k'$, on commence avec m intégrations par parties par rapport à $\tilde{x}$:

$$\begin{aligned} B_{k,k} &= \int_{(\tilde{x}, \tilde{z}) \in \tilde{\mathcal{C}}_{\tilde{\eta}}} \tilde{x}^m \tilde{z}_k^2 \tilde{b}_{k,k}(\tilde{x}, \tilde{z}) e^{-\Omega(\tilde{x} + \frac{1}{2} \sum_{l=1}^n \tilde{z}_l^2)} d\tilde{x} d\tilde{z} \\ &= \frac{1}{\Omega^m} \int_{(\tilde{x}, \tilde{z}) \in \tilde{\mathcal{C}}_{\tilde{\eta}}} \tilde{z}_k^2 \tilde{q}_{k,m}(\tilde{x}, \tilde{z}) e^{-\Omega(\tilde{x} + \frac{1}{2} \sum_{k=1}^n \tilde{z}_k^2)} d\tilde{x} d\tilde{z} + O(e^{-\tilde{\eta}\Omega}/\Omega), \end{aligned}$$

où $\tilde{q}_{k,m}(\tilde{x}, \tilde{z}) = \frac{\partial^m}{\partial \tilde{x}^m} (\tilde{x}^m \tilde{b}_{k,k}(\tilde{x}, \tilde{z}))$. Remarquons que $\tilde{q}_{k,m}(0) = m! \tilde{b}_{k,k}(0)$. Une simple intégration par parties par rapport à $\tilde{z}_k$ donne :

$$\begin{aligned} &\int_{(\tilde{x}, \tilde{z}) \in \tilde{\mathcal{C}}_{\tilde{\eta}}} \tilde{z}_k^2 \tilde{q}_{k,m}(\tilde{x}, \tilde{z}) e^{-\Omega(\tilde{x} + \frac{1}{2} \sum_{l=1}^n \tilde{z}_l^2)} d\tilde{x} d\tilde{z} \\ &= \frac{1}{\Omega} \int_{(\tilde{x}, \tilde{z}) \in \tilde{\mathcal{C}}_{\tilde{\eta}}} \left(\tilde{q}_{k,m}(\tilde{x}, \tilde{z}) + \tilde{z}_k \frac{\partial \tilde{q}_{k,m}}{\partial \tilde{z}_k}(\tilde{x}, \tilde{z}) \right) e^{-\Omega(\tilde{x} + \frac{1}{2} \sum_{l=1}^n \tilde{z}_l^2)} d\tilde{x} d\tilde{z} + O(e^{-\Omega\tilde{\eta}^2/2}) \\ &= \tilde{q}_{k,m}(0) \frac{1}{\Omega^2} \left(\frac{2\pi}{\Omega} \right)^{n/2} + O(\Omega^{-n/2-3}). \end{aligned}$$

En utilisant $\tilde{q}_{k,m}(0) = m! \tilde{b}_{k,k}(0)$, la somme $A + \sum_k C_k + \sum_{k,k'} B_{k,k'}$ correspondant à l'intégrale de l'équation (C.11) devient :

$$\tilde{I}_{\tilde{\eta}}(\Omega) = \frac{\sum_{k=1}^n m! \tilde{b}_{k,k}(0)}{\Omega^{m+2}} \left(\frac{2\pi}{\Omega} \right)^{n/2} + O\left(\Omega^{-m-n/2-3}\right).$$

Puisque $\tilde{I}_{\tilde{\eta}}$ et $e^{-\Omega f(0)} \mathcal{I}_g(\Omega)$ sont égaux à des termes exponentiellement petits près, nous obtenons finalement (C.10) de (C.12), puisque $\sum_{k=1}^n \tilde{b}_{k,k}(0) = \frac{1}{2} \text{Tr} \left(\frac{\partial^2 \tilde{g}}{\partial \tilde{z}^2} \Big|_0 \right)$. ■ ■

Les développements asymptotiques des théorèmes 7 et C.2 permettent d'énoncer directement les approximations de la moyenne et de la variance Bayésiennes ci-dessous.

Corollaire 1. Considérons la fonction analytique $f(z)$ du théorème 7. Alors, on a le développement asymptotique suivant pour toute fonction analytique $g(z)$:

$$\mathcal{M}_g(\Omega) \triangleq \frac{\int_{z \in (-1,1)^n} g(z) \exp(\Omega f(z)) dz}{\int_{z \in (-1,1)^n} \exp(\Omega f(z)) dz} = g(0) + O(\Omega^{-1}). \quad (\text{C.13})$$

On a aussi :

$$\begin{aligned} \mathcal{V}_g(\Omega) &\triangleq \frac{\int_{z \in (-1,1)^n} \left(g(z) - \mathcal{M}_g(\Omega) \right)^2 \exp(\Omega f(z)) dz}{\int_{z \in (-1,1)^n} \exp(\Omega f(z)) dz} \\ &= \frac{\text{Tr} \left(- \frac{\partial^2 g}{\partial z^2} \Big|_0 \left(\frac{\partial^2 f}{\partial z^2} \Big|_0 \right)^{-1} \right)}{2\Omega} + O(\Omega^{-2}). \end{aligned} \quad (\text{C.14})$$

Considérons la fonction analytique $f(x, z)$ du théorème 8. Alors, on a le développement asymptotique suivant pour toute fonction analytique $g(x, z)$:

$$\mathcal{M}_g(\Omega) \triangleq \frac{\int_{x \in (0,1)} \int_{z \in (-1,1)^n} x^m g(x, z) \exp(\Omega f(x, z)) dx dz}{\int_{x \in (0,1)} \int_{z \in (-1,1)^n} x^m \exp(\Omega f(x, z)) dx dz} = g(0, 0) + O(\Omega^{-1}). \quad (\text{C.15})$$

On a aussi :

$$\begin{aligned} \mathcal{V}_g(\Omega) &\triangleq \frac{\int_{x \in (0,1)} \int_{z \in (-1,1)^n} x^m \left(g(x, z) - \mathcal{M}_g(\Omega) \right)^2 \exp(\Omega f(x, z)) dx dz}{\int_{x \in (0,1)} \int_{z \in (-1,1)^n} x^m \exp(\Omega f(x, z)) dx dz} \\ &= \frac{\text{Tr} \left(- \frac{\partial^2 g}{\partial z^2} \Big|_{(0,0)} \left(\frac{\partial^2 f}{\partial z^2} \Big|_{(0,0)} \right)^{-1} \right)}{2\Omega} + O(\Omega^{-2}). \end{aligned} \quad (\text{C.16})$$

Au cours de la preuve du théorème 7, nous avons montré le lemme suivant, pendant le changement de coordonnées de z à $\tilde{z}$.

Lemme 5. Prenons deux fonctions réelles f et g de $z \in \mathbb{R}^n$ de classe C^2 . Supposons que 0 est un point critique régulier de f et prenons deux fonctions réelles f and g de $z \in \mathbb{R}^n$ de classe C^2 . Supposons que 0 est un point critique régulier de f et l'unique point critique de g . Prenons alors n'importe quel difféomorphisme ϕ de classe C^2 , défini localement dans un voisinage de 0 : $\tilde{z} = \phi(z)$. Alors :

$$\text{Tr} \left(- \frac{\partial^2 g}{\partial z^2} \Big|_0 \left(\frac{\partial^2 f}{\partial z^2} \Big|_0 \right)^{-1} \right) = \text{Tr} \left(- \frac{\partial^2 \tilde{g}}{\partial \tilde{z}^2} \Big|_{\phi(0)} \left(\frac{\partial^2 \tilde{f}}{\partial \tilde{z}^2} \Big|_{\phi(0)} \right)^{-1} \right)$$

où $\tilde{f}(\phi(z)) = f(z)$ et $\tilde{g}(\phi(z)) = g(z)$.

Ce lemme montre simplement que la trace présente dans la formule ci-dessus est invariante par changement de coordonnées, c'est-à-dire qu'elle est définie indépendamment des coordonnées locales choisies pour calculer les Hessiens de f et g en leur point critique commun.

C.3 Application à la tomographie d'état quantique

Comme on l'explique au chapitre 3, le paramètre p à estimer dans le cas de la tomographie quantique correspond à un opérateur densité ρ , qui est une matrice carrée à coefficients complexes appartenant à l'ensemble convexe et compact $\mathcal{D}_{\mathcal{H}}$ des matrices Hermitiennes de dimension $N_{\mathcal{H}} \times N_{\mathcal{H}}$ et de trace un. Ainsi, la fonction de log-vraisemblance a la structure suivante :

$$f(\rho) = \sum_{n=1}^{\Omega} \log \left(\text{Tr} \left(\rho E^{(n)} \right) \right), \quad (\text{C.17})$$

où les états adjoints $E^{(n)}$ appartiennent aussi à $\mathcal{D}_{\mathcal{H}}$. Pour toute matrice Hermitienne A de dimension $N_{\mathcal{H}} \times N_{\mathcal{H}}$ représentant un observable quantique, nous cherchons à établir une approximation de l'estimation Bayésienne $I_A(\Omega)$ de $\text{Tr}(\rho A)$:

$$I_A(\Omega) = \frac{\int_{\mathcal{D}_{\mathcal{H}}} \text{Tr}(\rho A) e^{\Omega f(\rho)} \mathbb{P}_0(\rho) \, d\rho}{\int_{\mathcal{D}_{\mathcal{H}}} e^{\Omega f(\rho)} \mathbb{P}_0(\rho) \, d\rho}, \quad (\text{C.18})$$

et de la variance Bayésienne $V_A(\Omega)$:

$$V_A(\Omega) = \frac{\int_{\mathcal{D}_{\mathcal{H}}} \left(\text{Tr}(\rho A) - I_A(\Omega) \right)^2 e^{\Omega f(\rho)} \mathbb{P}_0(\rho) \, d\rho}{\int_{\mathcal{D}_{\mathcal{H}}} e^{\Omega f(\rho)} \mathbb{P}_0(\rho) \, d\rho}. \quad (\text{C.19})$$

Ici, $d\rho$ représente un volume infinitésimal Euclidien de $\mathcal{D}_{\mathcal{H}}$, construit avec le produit scalaire de Frobenius sur l'ensemble des matrices Hermitiennes de dimension $N_{\mathcal{H}} \times N_{\mathcal{H}}$, et $\mathbb{P}_0 > 0$ est la densité de probabilité initiale sur ρ , sans considération de l'information apportée par les matrices $E^{(n)}$ sur la valeur de ρ . Puisque le nombre de paramètres réels décrivant ρ est grand en général, il est difficile de calculer ces intégrales, même numériquement en utilisant la méthode de Monte-Carlo.

Le lemme suivant apporte une caractérisation par invariance unitaire de toute matrice $\bar{\rho}$ en laquelle f atteint son maximum en $\mathcal{D}_{\mathcal{H}}$.

Lemme 6. Soit $\bar{\rho}$ une matrice Hermitienne de dimension $N_{\mathcal{H}} \times N_{\mathcal{H}}$ en laquelle la fonction $f : \mathcal{D}_{\mathcal{H}} \ni \rho \mapsto f(\rho) \in [-\infty, 0]$ définie par l'équation (C.17) sur $\mathcal{D}_{\mathcal{H}}$ (l'ensemble des matrices densité). Alors, $\bar{\rho}$ vérifie nécessairement les conditions suivantes :

- $\text{Tr}(\bar{\rho} E^{(n)}) > 0$ pour tout $n = 1, \dots, \Omega$;
- $[\bar{\rho}, \nabla f|_{\bar{\rho}}] = \bar{\rho} \cdot \nabla f|_{\bar{\rho}} - \nabla f|_{\bar{\rho}} \cdot \bar{\rho} = 0$, où $\nabla f|_{\bar{\rho}} = \sum_{n=1}^{\Omega} \frac{E^{(n)}}{\text{Tr}(\bar{\rho} E^{(n)})}$ est le gradient de f en $\bar{\rho}$, pour le produit de Frobenius ;
- il existe un scalaire $\bar{\lambda} > 0$ tel que $\bar{\lambda} \bar{P} = \bar{P} \nabla f|_{\bar{\rho}}$ et $\nabla f|_{\bar{\rho}} \leq \bar{\lambda} I$, où $\bar{P}$ est le projecteur orthogonal sur l'espace image de $\bar{\rho}$ et où $I_{\mathcal{H}}$ est l'opérateur identité.

Ces conditions sont également suffisantes et caractérisent un unique maximum lorsque, de plus, l'espace vectoriel généré par les matrices $E^{(n)}$ remplit l'espace des matrices Hermitiennes.

Démonstration. Lorsque f est une fonction concave en ρ , nous pouvons utiliser le critère d'optimalité standard pour un problème d'optimisation convexe (voir, par exemple, [BV09, section 4.2.3]) : $\bar{\rho}$ maximise f sur l'espace convexe compact $\mathcal{D}_{\mathcal{H}}$ si et seulement si pour tout $\rho \in \mathcal{D}_{\mathcal{H}}$, $\text{Tr}((\rho - \bar{\rho}) \nabla f|_{\bar{\rho}}) \leq 0$.

Supposons que $f(\bar{\rho})$ est maximum. Puisque $f(I) > -\infty$, on doit avoir $\text{Tr}(\bar{\rho} E^{(n)}) > 0$ pour tout n . En prenant $\rho = e^{-iH} \bar{\rho} e^{iH}$, où H est un opérateur Hermitien quelconque, nous avons :

$$\text{Tr}(e^{-iH} \bar{\rho} e^{iH} \nabla f|_{\bar{\rho}}) \leq \text{Tr}(\bar{\rho} \nabla f|_{\bar{\rho}}).$$

Lorsque H est proche de zéro, on obtient, grâce à la formule de Baker-Campbell-Hausdorff, $e^{-iH} \bar{\rho} e^{iH} = \bar{\rho} - i[H, \bar{\rho}] + O(\text{Tr}(H^2))$. L'inégalité ci-dessus implique que, pour tout H suffisamment petit, $\text{Tr}([H, \bar{\rho}] \nabla f|_{\bar{\rho}}) = \text{Tr}(H [\bar{\rho}, \nabla f|_{\bar{\rho}}]) = 0$, et, par conséquent, $\bar{\rho}$ et $\nabla f|_{\bar{\rho}}$ commutent.

Maintenant, considérons la décomposition spectrale $\bar{\rho} = U \bar{\Delta} U^\dagger$, où U est une matrice unitaire et $\bar{\Delta}$ est une matrice diagonale, dont les éléments diagonaux sont $0 \leq \bar{\Delta}_1 \leq \bar{\Delta}_2 \leq \dots \leq \bar{\Delta}_d \leq 1$. Puisque $\bar{\rho}$ et $\nabla f|_{\bar{\rho}}$ commutent, nous pouvons également noter $\nabla f|_{\bar{\rho}} = U \bar{\Lambda} U^\dagger$, où $\bar{\Lambda}$ est une matrice diagonale, d'éléments diagonaux $(\bar{\Lambda}_k)_k$. Comme ∇f est une matrice non-négative, ces coefficients $\bar{\Lambda}_k$ sont positifs ou nuls. En notant $\rho = U \Delta U^\dagger$, où Δ est une matrice diagonale non-négative de trace un quelconque, on a :

$$\text{Tr}((\rho - \bar{\rho}) \nabla f|_{\bar{\rho}}) = \text{Tr}((\Delta - \bar{\Delta}) \bar{\Lambda}) \leq 0.$$

Cela signifie que, pour tout $(\Delta_1, \dots, \Delta_d) \in [0, 1]^d$ tel que $\sum_{k=1}^d \Delta_k = 1$, nous avons :

$$\sum_{k=1}^d (\Delta_k - \bar{\Delta}_k) \bar{\Lambda}_k \leq 0.$$

Prenons alors $\epsilon > 0$, $(k_1, k_2) \in \{1, \dots, d\}^2$ tels que $\bar{\Delta}_{k_1} > 0$ et $k_2 \neq k_1$. Pour $k \in \{1, \dots, d-1\} \setminus \{k_1, k_2\}$, on choisit $\Delta_k = \bar{\Delta}_k$, et on fixe $\Delta_{k_1} = \bar{\Delta}_{k_1} - \epsilon$ avec $\Delta_{k_2} = \bar{\Delta}_{k_2} + \epsilon$. Ceci implique que $\text{Tr}(\Delta) = 1$ et que, pour $\epsilon > 0$ suffisamment petit, $\Delta_k \geq 0$ pour tout $k \in \{1, \dots, d\}$. L'inégalité précédente implique que :

$$\forall (k_1, k_2) \in \{1, \dots, d\}^2 \text{ tels que } \bar{\Delta}_{k_1} > 0 \text{ et } k_1 \neq k_2, \quad \bar{\Lambda}_{k_2} \leq \bar{\Lambda}_{k_1}.$$

Ainsi, pour tous k_1, k_2 tels que $\bar{\Delta}_{k_1} > 0$ et $\bar{\Delta}_{k_2} > 0$, $\bar{\Lambda}_{k_1} = \bar{\Lambda}_{k_2} = \bar{\Lambda} \geq 0$. Pour k_1 et k_2 tels que $\bar{\Delta}_{k_1} > 0$ et $\bar{\Delta}_{k_2} = 0$, on a également $\bar{\Lambda}_{k_2} \leq \bar{\Lambda}_{k_1} = \bar{\Lambda}$. Par conséquent, on a $\bar{\Lambda} \leq \bar{\Lambda} I$. En notant $\bar{\Theta}$ la matrice diagonale aux coefficients $\bar{\Theta}_k = 0$ (resp. $= 1$) lorsque $\bar{\Delta}_k = 0$ (resp. > 0), on a $\bar{P} = U \bar{\Theta} U^\dagger$, et nous obtenons $\bar{\Lambda} \bar{P} = \bar{P} \nabla f|_{\bar{\rho}}$. Puisque $\nabla f|_{\bar{\rho}}$ est non-négative et ne peut pas être nulle, nous trouvons $\bar{\Lambda} > 0$.

Prenons un $\bar{\rho}$ vérifiant les conditions du lemme 6. Puisqu'elles sont diagonales dans la même base, on peut assimiler $\bar{\rho}$ et $\nabla f|_{\bar{\rho}}$ aux opérateurs diagonaux $\bar{\Delta}$ et $\bar{\Lambda}$. Puisque f est concave et $\mathcal{D}_{\mathcal{H}}$ convexe, il est suffisant de prouver que $\bar{\rho}$ est un maximum local. Toute variation locale de ρ autour de $\bar{\rho}$ restant à l'intérieur de $\mathcal{D}_{\mathcal{H}}$ peut être paramétrée par l'application suivante :

$$(H, D) \mapsto e^{-iH} (\bar{\Delta} + D) e^{iH} = \rho_{H,D},$$

définie pour toute matrice Hermitienne H et toute matrice diagonale de trace nulle D telle que $\bar{\Delta} + D \geq 0$. On a le développement suivant pour H et D autour de zéro :

$$\rho_{H,D} = \bar{\Delta} + D - i[H, \bar{\Delta}] - i[H, D] - \frac{1}{2}[H, [H, \bar{\Delta}]] + O(\text{Tr}(H^3 + D^3)).$$

Ceci permet de donner le développement suivant au second ordre de $(H, D) \mapsto f(\rho_{H,D})$ autour de zéro :

$$f(\rho_{H,D}) = f(\bar{\rho}) + \text{Tr} \left(\bar{\Lambda} \left(D - i[H, \bar{\Delta}] - i[H, D] - \frac{1}{2}[H, [H, \bar{\Delta}]] \right) \right) - \sum_{\mu \in \mathcal{M}} \frac{\text{Tr}^2 \left((\rho_{H,D} - \bar{\rho}) E^{(n)} \right)}{2 \text{Tr}^2 (\bar{\rho} E^{(n)})} + O(\|\rho_{H,D} - \bar{\rho}\|^3).$$

Par hypothèse, $\bar{\Lambda}$, $\bar{\Delta}$ et D sont diagonales, donc $\text{Tr} \left(\bar{\Lambda} \left(-i[H, \bar{\Delta}] - i[H, D] \right) \right) = 0$. On peut trouver rapidement, en utilisant le fait que $\bar{\Lambda} \bar{\Theta} \leq \bar{\Delta} \leq \bar{\Lambda} I$, que $\text{Tr} (\bar{\Lambda} D) \leq 0$, puisqu'on a défini l'application pour des matrices D telles que $\bar{\Delta} + D$ est non-négative et de trace un. On a également :

$$-\text{Tr} \left(\bar{\Lambda} \left([H, [H, \bar{\Delta}]] \right) \right) = \text{Tr} \left([H, \bar{\Lambda}] [H, \bar{\Delta}] \right) = -2 \sum_{k_1 \in P, k_2 \in Q} \bar{\Delta}_{k_1} (\bar{\Lambda} - \bar{\Lambda}_{k_2}) |H_{k_1 k_2}|^2 \leq 0$$

où on note $P = \{k \mid \bar{\Delta}_k > 0\}$ et $Q = \{k \mid \bar{\Delta}_k = 0\}$.

Par conséquent :

$$f(\rho_{H,D}) \leq f(\bar{\rho}) - \sum_{n=1}^{\Omega} \frac{\text{Tr}^2 \left((\rho_{H,D} - \bar{\rho}) E^{(n)} \right)}{2 \text{Tr}^2 (\bar{\rho} E^{(n)})} + O(\|\rho_{H,D} - \bar{\rho}\|^3).$$

Puisque l'espace vectoriel généré par les matrices $E^{(n)}$ est l'ensemble $\mathcal{H}_{\text{herm}}$ des matrices Hermitiennes, la forme quadratique $X \in \mathcal{H}_{\text{herm}} \mapsto \sum_{n=1}^{\Omega} \frac{\text{Tr}^2(X E^{(n)})}{2 \text{Tr}^2(\bar{\rho} E^{(n)})}$ est non-dégénérée et f est strictement concave. Ainsi, nous avons $f(\rho) < f(\bar{\rho})$ pour tout $\rho \neq \bar{\rho}$ proche de $\bar{\rho}$. Par conséquent, $\bar{\rho}$ est un maximum local strict, et ce maximum est unique et global, car f est concave. ■

Théorème 9. Considérons la fonction de log-vraisemblance f définie par l'équation (C.17). On suppose que l'espace vectoriel engendré par la famille des $E^{(n)}$ coïncide avec l'espace $\mathcal{H}_{\text{herm}}$ des matrices Hermitiennes. Notons $\bar{\rho}$ l'unique maximum de f sur $\mathcal{D}$ et définissons le projecteur orthogonal $\bar{P}$ tel que, en plus des conditions nécessaires et suffisantes du lemme 6, nous avons $\ker(\bar{\Lambda} I - \nabla f|_{\bar{\rho}}) = \ker(I - \bar{P})$. Ainsi, pour tout opérateur Hermitien A , sa moyenne Bayésienne telle que définie par la formule (C.18) admet le développement asymptotique suivant :

$$I_A(\Omega) = \text{Tr}(A \bar{\rho}) + O(1/\Omega),$$

et sa variance Bayésienne telle que définie par la formule (C.19) est donnée par :

$$V_A(\Omega) = \text{Tr} \left(A_{\parallel} \left(\bar{\mathbf{F}} \right)^{-1} (A_{\parallel}) \right) / \Omega + O(1/\Omega^2),$$

où :

- pour tout opérateur Hermitien B , $B_{\parallel}$ représente la projection orthogonale sur l'espace tangent en $\bar{\rho}$ à la sous-variété des matrices Hermitiennes de même rang que $\bar{\rho}$ et de trace un, ce qui s'écrit :

$$B_{\parallel} = B - \frac{\text{Tr}(B \bar{P})}{\text{Tr}(\bar{P})} \bar{P} - (I - \bar{P}) B (I - \bar{P}); \quad (\text{C.20})$$

quand $\bar{\rho}$ est de rang plein, $B_{\parallel} = B - \text{Tr}(B) I/d$ car $\bar{P} = I$;

- le super-opérateur linéaire $\bar{\mathbf{F}}$ correspond au Hessien en $\bar{\rho}$ de la restriction de f à la sous-variété des matrices Hermitiennes de même rang que $\bar{\rho}$ et de trace un. Pour tout opérateur Hermitien X , ce super-opérateur a pour formule :

$$\bar{\mathbf{F}}(X) = \sum_{\mu} \frac{\text{Tr}(XY_{\mu\parallel})}{\text{Tr}^2(\bar{\rho}E^{(n)})} Y_{\mu\parallel} + (\bar{\lambda}I - \nabla f|_{\bar{\rho}}) X \bar{\rho}^+ + \bar{\rho}^+ X (\bar{\lambda}I - \nabla f|_{\bar{\rho}}), \quad (\text{C.21})$$

où $\bar{\rho}^+$ est le pseudo-inverse de Moore-Penrose de $\bar{\rho}$; la restriction de $X \mapsto \text{Tr}(X\bar{\mathbf{F}}(X))$ à l'espace tangent en $\bar{\rho}$ est positive définie; ainsi, la restriction de $\bar{\mathbf{F}}$ à cet espace tangent est inversible et peut être vue comme l'analogie de l'information de Fisher; son inverse en $A_{\parallel}$ est noté ci-dessus $(\bar{\mathbf{F}})^{-1}(A_{\parallel})$.

Démonstration. Le Hessien de f en $\rho \in \mathcal{D}$ tel que $f(\rho) > -\infty$ est noté :

$$\nabla^2 f|_{\rho}(X, Z) = - \sum_{\mu} \frac{\text{Tr}(XE^{(n)}) \text{Tr}(ZE^{(n)})}{\text{Tr}^2(\rho E^{(n)})},$$

où X et Z sont des matrices Hermitiennes quelconques. Puisque ce Hessien est négatif défini, f est strictement concave. Par conséquent, f admet un unique maximum sur $\mathcal{D}$, noté $\bar{\rho}$, et vérifie la condition du lemme 6. Prenons un voisinage infinitésimal $\mathcal{V}$ de $\bar{\rho}$ dans $\mathcal{D}$. Alors, il existe un $\epsilon > 0$ tel que, pour tout $\rho \in \mathcal{D}/\mathcal{V}$, $f(\rho) \leq f(\bar{\rho}) - \epsilon$. Pour donner une formule de $\int_{\mathcal{V}} e^{\Omega(f(\rho)-f(\bar{\rho}))} \mathbb{P}_0(\rho) d\rho$, nous utilisons les coordonnées locales suivantes, basées sur la décomposition spectrale de $\bar{\rho} = U\bar{\Delta}U^\dagger$, avec U unitaire et $\bar{\Delta}$ diagonale, de coefficients $0 = \bar{\delta}_1 \leq \bar{\delta}_2 \leq \dots \leq \bar{\delta}_d \leq 1$ vérifiant $\sum_{k=1}^d \bar{\delta}_k = 1$. Notons r le rang de $\bar{\rho}$ et supposons que $r < d$ (le cas où $r = d$ est beaucoup plus simple, puisqu'il se résout grâce au théorème 7, et est laissé aux soins du lecteur). Nous avons $\bar{\delta}_k = 0$ pour k entre 1 et $d-r$, et $\bar{\delta}_k > 0$ pour tout k entre $d-r+1$ et d . Puisque l'élément de volume infinitésimal $d\rho$ utilisé dans les équations (C.18) et (C.19) est unitairement invariant [Meh04, page 42], on peut supposer sans perte de généralité que $\bar{\rho}$ est diagonale (il faut simplement changer $\mathbb{P}_0(\bullet)$ en $\mathbb{P}_0(U \bullet U^\dagger)$ et remplacer chaque $E^{(n)}$ par $U^\dagger E^{(n)} U$ dans la définition de f de l'équation (C.17)). Considérons l'application suivante :

$$(\xi, \zeta, \omega) \mapsto \Upsilon = \exp \left(\begin{bmatrix} 0 & \omega \\ -\omega^\dagger & 0 \end{bmatrix} \right) \begin{bmatrix} \xi & 0 \\ 0 & \bar{\Delta}_r + \zeta - \frac{\text{Tr}(\xi)}{r} I_r \end{bmatrix} \exp \left(\begin{bmatrix} 0 & -\omega \\ \omega^\dagger & 0 \end{bmatrix} \right)$$

où ξ est une matrice Hermitienne de dimensions $(d-r) \times (d-r)$, ω est une matrice $(d-r) \times r$ à coefficients complexes, ζ est une matrice $r \times r$ Hermitienne de trace nulle, I_r est la matrice identité de dimension r et $\bar{\Delta} = \begin{bmatrix} 0 & 0 \\ 0 & \bar{\Delta}_r \end{bmatrix}$. Cette application est un difféomorphisme local d'un voisinage de $(0, 0, 0)$ à un voisinage de $\bar{\rho}$ sur l'ensemble des matrices Hermitiennes de trace un, puisque l'application permettant de donner l'espace tangent en zéro, donnée par :

$$(\delta\xi, \delta\zeta, \delta\omega) \mapsto \begin{bmatrix} \delta\xi & \delta\omega \bar{\Delta}_r \\ \bar{\Delta}_r \delta\omega^\dagger & \delta\zeta - \frac{\text{Tr}(\delta\xi)}{r} I_r \end{bmatrix} = \delta\rho \quad (\text{C.22})$$

est bijective (théorème d'inversion locale). Ainsi, nous pouvons écrire :

$$\begin{aligned} \int_{\mathcal{V}} e^{\Omega(f(\rho)-f(\bar{\rho}))} \mathbb{P}_0(\rho) d\rho \\ = \int_{\Upsilon^{-1}(\mathcal{V})} e^{\Omega(f(\xi, \zeta, \omega)-f(0,0,0))} \mathbb{P}_0(\xi, \zeta, \omega) J(\xi, \zeta, \omega) d\xi d\zeta d\omega \end{aligned}$$

où $f(\xi, \zeta, \omega)$ et $\mathbb{P}_0(\xi, \zeta, \omega)$ représentent $f(\Upsilon(\xi, \zeta, \omega))$ et $\mathbb{P}_0(\Upsilon(\xi, \zeta, \omega))$, et où $J(\xi, \zeta, \omega)$ est le Jacobien du changement de coordonnées.

La contrainte $\Upsilon(\xi, \zeta, \omega) \geq 0$ pouvant être écrite $\xi \geq 0$, nous introduisons un nouveau changement de coordonnées pour paramétrer $\xi \geq 0$ autour de 0 : $\Xi : (x, \sigma, \zeta, \omega) \mapsto (x\sigma = \xi, \zeta, \omega)$, où $x \geq 0$ et σ est une matrice densité de dimension $(d-r) \times (d-r)$. Dans ce cas :

$$\begin{aligned} \int_{\mathcal{V}} e^{\Omega(f(\rho) - f(\bar{\rho}))} \mathbb{P}_0(\rho) \, d\rho \\ = \int_{\Xi^{-1}(\Upsilon^{-1}(\mathcal{V}))} e^{\Omega(f(x\sigma, \zeta, \omega) - f(0, 0, 0))} \mathbb{P}_0(x\sigma, \zeta, \omega) J(x\sigma, \zeta, \omega) x^m \, dx \, d\sigma \, d\zeta \, d\omega \end{aligned}$$

où $m = (d-r+1)(d-r-1)$. Ce changement de variables est singulier, puisque pour $x = 0$, il n'est pas inversible ; cependant, puisque l'ensemble des coordonnées vérifiant $x = 0$ est de mesure nulle, cela n'a pas d'impact sur l'intégrale. Prenons $\eta > 0$ assez petit et ajustons le voisinage $\mathcal{V}$ de $\bar{\rho}$ tel que $\Xi^{-1}(\Upsilon^{-1}(\mathcal{V}))$ coïncide avec l'espace défini par $x \in (0, \eta)$, $\sigma \in \mathcal{D}_{d-r}$ et par le fait que les coefficients des parties réelles et imaginaires de ζ et de ω appartiennent à $(-\eta, \eta)$. En respectant les notations du théorème 7, nous posons $z = (\zeta, \omega)$. On a $z \in (-\eta, \eta)^n$ avec $n = 2r(d-r) + (r+1)(r-1)$ et :

$$\begin{aligned} \int_{\mathcal{V}} e^{\Omega(f(\rho) - f(\bar{\rho}))} \mathbb{P}_0(\rho) \, d\rho \\ = \int_{\sigma \in \mathcal{D}_{d-r}} \left(\int_{(x,z) \in (0,\eta) \times (-\eta,\eta)^n} e^{\Omega f(x\sigma, z)} x^m J(x\sigma, z) \mathbb{P}_0(x\sigma, z) \, dx \, dz \right) d\sigma. \end{aligned}$$

Pour chaque $\sigma \in \mathcal{D}_{d-r}$, nous utilisons le résultat (C.9), avec $J(x\sigma, z)\mathbb{P}_0(x\sigma, z)$ à la place de $g(x, z)$. On sait que $g(0, 0) = J(0, 0)\mathbb{P}_0(0, 0) > 0$. Par construction, on a :

$$f(x\sigma, z) = x f_1(x, \sigma, z) + f(0, z),$$

où $f_1(x, \sigma, z)$ est analytique par rapport à (x, z) et $f_1(0, \sigma, 0) = (\text{Tr}(\Lambda_{d-r}\sigma) - \bar{\lambda})$. Cela se montre en utilisant (C.22) et en exploitant la structure diagonale de $\nabla f|_{\bar{\rho}} = \begin{bmatrix} \Lambda_{d-r} & 0 \\ 0 & \bar{\lambda} I_r \end{bmatrix}$.

Par hypothèse, $\Lambda_{d-r} < \bar{\lambda} I_{d-r}$. Ainsi, il existe un $\epsilon' > 0$ tel que, pour tout σ , $f_1(0, \sigma, 0) < -\epsilon'$ et $\frac{\partial f}{\partial x} < -\epsilon'$ en $(x, z) = 0$, pour tout $\sigma \in \mathcal{D}_{d-r}$. Considérons maintenant le développement de $z \mapsto f(0, z)$ jusqu'à l'ordre 2 en z . En rappelant que $\delta z = (\delta\zeta, \delta\omega)$ et en utilisant (C.22), complété des termes de second ordre obtenus grâce à la formule de Baker-Campbell-Hausdorff, on trouve :

$$\delta\rho = \begin{bmatrix} \delta\omega \, \bar{\Delta}_r \, \delta\omega^\dagger & \delta\omega \, (\bar{\Delta}_r + \delta\zeta) \\ (\delta\zeta + \bar{\Delta}_r) \, \delta\omega^\dagger & \delta\zeta - \frac{\delta\omega^\dagger \delta\omega \bar{\Delta}_r + \bar{\Delta}_r \delta\omega^\dagger \delta\omega}{2} \end{bmatrix} + 0(\|\delta z\|^3).$$

Par conséquent,

$$\begin{aligned} f(0, \delta z) &= f(\bar{\rho}) + \text{Tr} \left(\nabla f|_{\bar{\rho}} \, \delta\rho \right) + \frac{1}{2} \nabla^2 f|_{\bar{\rho}} (\delta\rho, \delta\rho) + O(\|\delta\rho\|^3) \\ &= f(\bar{\rho}) - \text{Tr} \left((\bar{\lambda} I_{d-r} - \Lambda_{d-r}) \delta\omega \, \bar{\Delta}_r \, \delta\omega^\dagger \right) - \frac{1}{2} \sum_{\mu} \frac{\text{Tr}^2(\delta\rho E^{(n)})}{\text{Tr}^2(\bar{\rho} E^{(n)})}. \end{aligned} \quad (\text{C.23})$$

Cela montre que $\frac{\partial f}{\partial z}$ est nul en $(0, z)$ et que $\frac{\partial^2 f}{\partial z^2}$ est négative définie en $(0, z)$ ($\bar{\lambda} I_{d-r} > \Lambda_{d-r}$) et sa valeur indépendante de σ . Toutes les hypothèses nécessaires à l'application du

théorème (C.9) sont remplies, et on peut écrire :

$$\int_{\mathcal{D}} e^{\Omega f(\rho)} \mathbb{P}_0(\rho) d\rho = \kappa_0 e^{f(\bar{\rho})\Omega} \Omega^{-m-n/2-1} \int_{\sigma \in \mathcal{D}_{d-r}} \frac{d\sigma}{(\bar{\lambda} - \text{Tr}(\Lambda_{d-r}\sigma))^{m+1}} + O\left(e^{f(\bar{\rho})\Omega} \Omega^{-m-n/2-2}\right)$$

où $\kappa_0 = \frac{\mathbb{P}_0(\bar{\rho}) J(0,0) m! (2\pi)^{n/2}}{\sqrt{\left| \det \left(\frac{\partial^2 f}{\partial z^2} \right) \Big|_{(0,0)} \right|}}.$

De la même façon, nous avons :

$$\int_{\mathcal{D}} \text{Tr}(\rho A) e^{\Omega f(\rho)} \mathbb{P}_0(\rho) d\rho = \kappa_0 \text{Tr}(A\bar{\rho}) e^{f(\bar{\rho})\Omega} \Omega^{-m-n/2-1} \int_{\sigma \in \mathcal{D}_{d-r}} \frac{d\sigma}{(\bar{\lambda} - \text{Tr}(\Lambda_{d-r}\sigma))^{m+1}} + O\left(e^{f(\bar{\rho})\Omega} \Omega^{-m-n/2-2}\right).$$

Par conséquent, nous avons prouvé que : $I_A(\Omega) = \text{Tr}(\bar{\rho}A) + O(1/\Omega).$

Des calculs relativement faciles montrent que le développement asymptotique de $V_A(\Omega)$ se réduit à celui de l'intégrale $\int_{\mathcal{D}} \text{Tr}^2((\rho - \bar{\rho})A) e^{\Omega f(\rho)} \mathbb{P}_0(\rho) d\rho$, réalisé en utilisant (C.10) avec $g(x, \sigma, z) = J(x\sigma, z) \mathbb{P}_0(x\sigma, z) h(x\sigma, z)$, $h(x\sigma, z) = \text{Tr}^2((\Upsilon(x\sigma, z) - \bar{\rho})A)$ et $z = (\zeta, \omega)$. Puisque $\frac{\partial^2 g}{\partial z^2} \Big|_{(0,\sigma,0)} = J(0,0) \mathbb{P}_0(\bar{\rho}) \frac{\partial^2 h}{\partial z^2} \Big|_{(0,\sigma,0)}$ est indépendant de σ , on a, à partir de (C.10) :

$$\int_{\mathcal{D}} \text{Tr}^2((\rho - \bar{\rho})A) e^{\Omega f(\rho)} \mathbb{P}_0(\rho) d\rho = \kappa_0 \frac{\text{Tr} \left(-\frac{\partial^2 h}{\partial z^2} \Big|_{(0,0)} \left(\frac{\partial^2 f}{\partial z^2} \Big|_{(0,0)} \right)^{-1} \right)}{2} e^{f(\bar{\rho})\Omega} \Omega^{-m-n/2-2} \int_{\sigma \in \mathcal{D}_{d-r}} \frac{d\sigma}{(\bar{\lambda} - \text{Tr}(\Lambda_{d-r}\sigma))^{m+1}} + O\left(e^{f(\bar{\rho})\Omega} \Omega^{-m-n/2-3}\right).$$

Par conséquent, on trouve que $V_A(\Omega) = \frac{\text{Tr} \left(-\frac{\partial^2 h}{\partial z^2} \Big|_{(0,0)} \left(\frac{\partial^2 f}{\partial z^2} \Big|_{(0,0)} \right)^{-1} \right)}{2\Omega} + O(\Omega^{-2})$. Le fait que la trace au numérateur vaut $2\text{Tr} \left(A_{\parallel} (\bar{\mathbf{F}})^{-1} (A_{\parallel}) \right)$ vient des raisonnements suivants.

- La formule (C.20) est invariante par transformation unitaire. Dans le cas où $\bar{\rho} = \begin{bmatrix} 0 & 0 \\ 0 & \bar{\Delta}_r \end{bmatrix}$ est diagonal, l'espace tangent en $\bar{\rho}$ à la variété des matrices Hermitiennes de rang r est donnée par $\delta\rho$ vérifiant (C.22) avec $\delta\xi = 0$ et $(\delta\zeta, \delta\omega)$ quelconque. On peut vérifier que (C.20) permet d'écrire la décomposition par blocs $\begin{bmatrix} 0 & A_{0,r} \\ A_{0,r}^\dagger & A_r - \frac{\text{Tr}(A_r)}{r} I_r \end{bmatrix}$ pour $A_{\parallel}$ écrit $A = \begin{bmatrix} A_0 & A_{0,r} \\ A_{0,r}^\dagger & A_r \end{bmatrix}$. On peut également vérifier que $A_{\parallel}$ appartient à cet espace tangent et que $\text{Tr}(A\delta\rho) = \text{Tr}(A_{\parallel}\delta\rho)$ pour chaque élément tangent $\delta\rho$.
- Puisque $h(0, z) = \text{Tr}^2((\Upsilon(0, z) - \bar{\rho})A)$, on a :

$$\frac{\partial^2 h}{\partial z^2} \Big|_{(0,0)} (\delta z, \delta z) = 2\text{Tr}^2(\delta\Upsilon A) = 2\text{Tr}^2(\delta\Upsilon A_{\parallel})$$

avec $\delta\Upsilon = \begin{bmatrix} 0 & \delta\omega \overline{\Delta}_r \\ \overline{\Delta}_r \delta\omega^\dagger & \delta\zeta \end{bmatrix}$ et $\delta z = (\delta\zeta, \delta\omega)$. Cela signifie que $\frac{\partial^2 h}{\partial z^2}|_{(0,0)}$ est colinéaire au projecteur orthogonal sur la direction donnée par $A_\parallel$ sur l'espace tangent à $\bar{\rho}$. Cela implique que $\text{Tr} \left(\frac{\partial^2 h}{\partial z^2}|_{(0,0)} \left(\frac{\partial^2 f}{\partial z^2}|_{(0,0)} \right)^{-1} \right)$ équivaut à deux fois la valeur en $A_\parallel$ de la forme quadratique associée à l'inverse du Hessien en $\bar{\rho}$ de la restriction de f à la variété des matrices Hermitiennes de rang r et de trace un (on utilise ici le lemme 5).

- Ce Hessien a pour expression (C.21) puisque, pour $X = \delta\Upsilon = \begin{bmatrix} 0 & \delta\omega \overline{\Delta}_r \\ \overline{\Delta}_r \delta\omega^\dagger & \delta\zeta \end{bmatrix}$, on trouve :

$$\begin{aligned} \text{Tr} \left(X \left(\bar{\lambda}I - \nabla f|_{\bar{\rho}} \right) X \bar{\rho}^+ + X \bar{\rho}^+ X \left(\bar{\lambda}I - \nabla f|_{\bar{\rho}} \right) \right) \\ = 2\text{Tr} \left((\bar{\lambda}I_{d-r} - \Lambda_{d-r}) \delta\omega \overline{\Delta}_r \delta\omega^\dagger \right) \end{aligned}$$

car $\bar{\rho}^+ = \begin{bmatrix} 0 & 0 \\ 0 & \overline{\Delta}_r^{-1} \end{bmatrix}$. Nous retrouvons, grâce à (C.23), que $f(0, z) = f(\bar{\rho}) - \frac{1}{2}\text{Tr} \left(X \overline{\mathbf{F}}(X) \right)$, c'est-à-dire que $\overline{\mathbf{F}}$ est en réalité le Hessien en $\bar{\rho}$ de la restriction de f à l'ensemble des matrices Hermitiennes de rang r et de trace un.

■

■

C.4 Conclusion

Lorsque l'estimation par Maximum de Vraisemblance produit un état quantique de rang faible, nous avons apporté dans cette annexe une estimation de la moyenne et de la variance Bayésiennes pour tout observable, grâce aux développements asymptotiques de certaines intégrales de Laplace multi-dimensionnelles. Nous supposons que des développements asymptotiques similaires pourraient avoir un certain intérêt pour le *compress sensing* quantique [GLF⁺10], lorsque la dimension de l'espace de Hilbert est beaucoup plus grande que le rang de la matrice estimée.

Résumé

La communauté scientifique a réussi ces dernières années à bâtir des systèmes quantiques simples sur lesquels des séries de mesures sont acquises successivement le long de trajectoires quantiques et sans réinitialisation de l'état (opérateur densité) par l'expérimentateur. L'objet de cette thèse est d'adapter les méthodes de tomographie quantique (estimation d'état et de paramètres) à ce cadre pour prendre en compte la rétroaction de la mesure sur l'état, la décohérence et les imperfections expérimentales.

Durant le processus de mesure, l'évolution de l'état quantique est alors gouvernée par un processus de Markov à états cachés (filtres quantiques de Belavkin). Pour des mesures en temps continu, nous commençons par montrer comment discrétiser l'équation maîtresse stochastique tout en préservant la positivité et la trace de l'état quantique, et ainsi se ramener aux filtres quantiques en temps discret. Ensuite, nous développons, à partir de trajectoires de mesures en temps discret, des techniques d'estimation par maximum de vraisemblance pour l'état initial et les paramètres. Cette estimation est accompagnée de son intervalle de confiance. Lorsqu'elle concerne des valeurs de paramètres (tomographie de processus quantique), nous donnons un résultat de robustesse grâce au formalisme des filtres particuliers et nous proposons une méthode de maximisation fondée sur le calcul du gradient par l'adjoint et bien adaptée au cas multiparamétrique. Lorsque l'estimation porte sur l'état initial (tomographie d'état quantique), nous donnons une formulation explicite de la fonction de vraisemblance grâce aux états adjoints, montrons que son logarithme est une fonction concave de l'état initial et élaborons une expression intrinsèque de la variance grâce à des développements asymptotiques de moyennes bayésiennes et reposant sur la géométrie de l'espace des opérateurs densité.

Ces méthodes d'estimation ont été appliquées et validées expérimentalement pour deux types de mesures quantiques : des mesures en temps discret non destructives de photons dans le groupe d'électrodynamique quantique en cavité du LKB au Collège de France, des mesures diffusives de la fluorescence d'un qubit supraconducteur dans le groupe d'électronique quantique du LPA à l'ENS Paris.

Mots Clés

Tomographie quantique, filtres quantiques, équation maîtresse stochastique, chaînes de Markov à état caché, qubit supraconducteur, électrodynamique en cavité, maximum de vraisemblance, estimation bayésienne, calcul du gradient par l'adjoint, développement asymptotique d'intégrales de Laplace

Abstract

In recent years, the scientific community has succeeded in experimentally building simple quantum systems on which series of measurements are successively acquired along quantum trajectories, without any reinitialization of their state (density operator) by the physicist. The subject of this thesis is to adapt the quantum tomography techniques (state and parameters estimation) to this frame, in order to take into account the feedback of the measurement on the state, the decoherence and experimental imperfections.

During the measurement process, the evolution of the quantum state is then governed by a hidden-state Markov process (Belavkin quantum filters). Concerning continuous-time measurements, we begin by showing how to discretize the stochastic master equation, while preserving the positivity and the trace of the quantum state, and so reducing to discrete-time quantum filters. Then, we develop, starting from trajectories of discrete-time measurements, some maximum-likelihood estimation techniques for initial state and parameters. This estimation is coupled with its confidence interval. When it concerns the value of parameters (quantum process tomography), we provide a result of robustness using the formalism of particular filters, and we propose a maximization technique based on the calculus of gradient by adjoint method, which is well adapted to the multi-parametric case. When the estimation concerns the initial state (quantum state tomography), we give an explicit formulation of the likelihood function thanks to the adjoint states, show that its logarithm is a concave function of the initial state and build an intrinsic expression of the variance, obtained from asymptotic developments of Bayesian means, lying on the geometry of the space of density operators.

These estimation techniques have been applied and experimentally validated for two types of quantum measurements: discrete-time non-destructive measurements of photons in the group of cavity quantum electro-dynamics of LKB at Collège de France, diffusive measurements of the fluorescence of a supra-conducting qubit in the quantum electronics group of LPA at ENS Paris.

Keywords

Quantum tomography, quantum filters, stochastic master equation, hidden-state Markov chains, superconducting qubit, cavity quantum electrodynamics, maximum-likelihood, bayesian estimation, gradient calculus by adjoint method, asymptotic expansion of Laplace's integrals