

A fournir au format WORD exclusivement

Titre : Méthodes hybrides couplant apprentissage par renforcement et méthodes de contrôle optimal des EDP pour l'informatique Quantique

Directeur(s) de Thèse : Emmanuel Franck, Chercheur INRIA
Yannick Privat, Professeur univ. Strasbourg

Unité(s) d'Accueil(s) : IRMA

Établissement de rattachement : université de Strasbourg

Collaboration(s) (s'il y a lieu) :

Rattachement à un programme (s'il y a lieu) :

Résumé (1500 caractères au maximum) :

Ce projet de thèse a trait à l'informatique quantique : on s'intéresse à la possibilité d'encoder une porte logique comme la porte de Hadamard ou la porte "not" à partir de systèmes de type Qudit. On aborde ces questions à l'aide de problèmes de contrôle optimal, typiquement des problèmes de temps minimal. Le modèle dynamique sous-jacent est donné par l'équation de Lindblad, sur la matrice de densité décrivant l'état du système quantique. On cherche à utiliser le contrôle pour effectuer le plus efficacement possible cette opération. Néanmoins, cette question est ardue, notamment en raison de l'émergence d'un phénomène physique appelé décohérence, s'opposant à l'action du contrôle. Nous envisageons à la fois une étude théorique permettant de mieux cerner les contours du problème (existence d'un contrôle optimal, estimation des paramètres garantissant l'efficacité des contrôles, dérivation d'un algorithme numérique efficace) et une étude numérique fondée sur la combinaison d'algorithmes traditionnels de type BFGS ou gradient et de méthodes d'apprentissage adaptées au problème et à sa dimension potentiellement grande en fonction des molécules étudiées.

Projet de thèse : Méthodes hybrides couplant apprentissage par renforcement et méthodes de contrôle optimal des edp pour l'informatique Quantique.

Thèse encadrée par Emmanuel Franck & Yannick Privat

26 janvier 2023

Résumé

Ce projet de thèse a trait à l'informatique quantique : on s'intéresse à la possibilité d'encoder une porte logique comme la porte de Hadamard ou la porte "not" à partir de systèmes de type *Qudit*. On aborde ces questions à l'aide de problèmes de contrôle optimal, typiquement des problèmes de temps minimal. Le modèle dynamique sous-jacent est donné par l'équation de Lindblad, sur la matrice de densité décrivant l'état du système quantique. On cherche à utiliser le contrôle pour effectuer le plus efficacement possible cette opération. Néanmoins, cette question est ardue, notamment en raison de l'émergence d'un phénomène physique appelé décohérence, s'opposant à l'action du contrôle. Nous envisageons à la fois une étude théorique permettant de mieux cerner les contours du problème (existence d'un contrôle optimal, estimation des paramètres garantissant l'efficacité des contrôles, dérivation d'un algorithme numérique efficace) et une étude numérique fondée sur la combinaison d'algorithmes traditionnels de type BFGS ou gradient et de méthodes d'apprentissage adaptées au problème et à sa dimension potentiellement grande en fonction des molécules étudiées.

1 Description du projet

1.1 Contexte

En informatique classique, on encode une information sur un bit qui peut prendre comme valeur 1 ou 0. Dans le domaine quantique, l'approche la plus usuelle consiste à encoder les valeurs 0 et 1 sur le spin d'une particule. Par la propriété de superposition des objets quantiques, un *Qubit* est susceptible de porter à la fois l'état 0 et l'état 1. Avec un système quantique à N *Qubits*, on peut donc effectuer 2^N calculs en parallèle. Les qubits constituent ainsi des systèmes quantiques à deux niveaux. Cependant, il existe des systèmes multiniveaux, appelés *qudits*, correspondant à un système à d niveaux, qui devraient présenter un certain nombre d'avantages. En voici quelques-uns : un nombre réduit de portes élémentaires pour effectuer des opérations, des taux d'erreurs inférieurs à ceux des qubits ; des opérations plus complexes (appelées portes logiques) peuvent être appliquées à un seul *Qudit*.

Ici on s'intéresse à la possibilité d'encoder une porte logique comme la porte de Hadamard ou la porte "not" à partir de système de type *Qudit*. Pour cela, on manipule le système à l'aide d'impulsions électriques micro-ondes. Les qubits et les qudits sont capables de stocker des informations quantiques pendant une certaine période de temps appelée *temps de cohérence*. Lorsque le système est en interaction avec l'environnement, le système quantique a tendance à perdre son caractère quantique, à travers un processus appelé *décohérence*. Il est donc important de déterminer des impulsions électriques qui permettent de générer la porte avant que la décohérence n'intervienne. D'un point de vue mathématique, l'évolution dynamique de ces systèmes est décrite par l'équation de *Lindblad* [1] qui permet d'obtenir les probabilités en fonction du temps de chaque état possible du système.

Cette équation traduit l'évolution de la matrice de densité, décrivant l'ensemble des états du système quantique. Elle s'écrit sous forme diagonale :

$$\frac{d\rho}{dt} = -i[H, \rho] + \sum_{i,j=1}^{d^2-1} g_{ij}(F_i \rho F_i^\top - \frac{1}{2}\{F_i^\top F_j, \rho\}), \quad (1)$$

où

- d est la dimension de l'espace euclidien considéré.
- ρ est une matrice de densité (i.e. une matrice hermitienne définie positive et de trace 1) décrivant l'état du système quantique.
- H est le Hamiltonien du système (un opérateur hermitien dans l'espace euclidien considéré).
- $[a, b]$ désigne le commutateur $ab - ba$.
- $\{a, b\}$ désigne l'anticommutateur $ab + ba$.
- les g_{ij} sont des coefficients scalaires, les $(F_i)_i$ forment une base orthonormée de l'espace des opérateurs sur l'espace euclidien considéré.

En l'absence d'interaction avec l'environnement, c'est-à-dire lorsqu'on ne considère pas le deuxième terme de la somme du second membre de (1), ce système est hamiltonien et possède des propriétés de conservation (de positivité et de la trace de la matrice de densité). La prise en compte de ces interactions conduit à l'ajout du dernier terme de la somme dans le second membre de (1), rendant le système dissipatif.

Pour un système de dimension d , on a $d - 1$ contrôles (les impulsions électriques, des fonctions de $L^\infty(0, T)$, avec $T > 0$ désignant la durée de l'expérience) qui agissent sur le système comme une composante affine du Hamiltonien. Il s'agit donc d'un problème de contrôle optimal quadratique dont la dimension de l'espace des états et des contrôles varie selon la molécule considérée et peut rapidement devenir importante. Par exemple, pour une molécule $TbPc_2$, on a $d = 4$ [4] tandis que pour une molécule Tb_2Pc_3 , on a $d = 8$.

Des premiers résultats numériques obtenus par des chercheurs de l'Institut de Physique et de Chimie des Matériaux de Strasbourg (IPCMS) semblent montrer que les méthodes de gradient classiques [5] sont très efficaces pour d petit (i.e. au-dessous de $d = 6, 7$), mais pas en grande dimension (i.e. supérieure à $d = 8$). Il y a des difficultés algorithmiques, mais aussi physiques. En effet, les résultats numériques suggèrent que plus la dimension de l'espace euclidien augmente, plus le temps après lequel l'état quantique du système est définitivement perdu appelé *temps de décohérence* semble devenir petit, ce qui nécessite de déterminer des stratégies de contrôle destinées à agir rapidement sur le système. Cela apparaît comme une des difficultés majeures en grande dimension. Depuis peu, des approches de type apprentissage par renforcement ont été étudiées pour pallier ce problème : il s'agit d'appliquer des méthodes de type gradient de politique [9] pour contrôler l'équation de Lindblad simulée. Ces méthodes ont l'avantage d'introduire des mécanismes d'exploration de l'espace des contrôles et de calculer un contrôle en boucle fermée (i.e. dépendant de l'état du système) qui apparaît plus robuste lorsqu'il est implémenté expérimentalement.

Dans le cadre de cette thèse, en collaboration avec l'IPCMS, nous souhaitons compléter et étendre les résultats ci-dessus à l'aide d'analyse théorique de problèmes de contrôle optimal : nous chercherons à préciser la structure des contrôles optimaux et à comprendre dans quelle mesure la dimension du problème de contrôle (le paramètre d dans l'équation de Lindblad) influence sa structure ainsi que le temps minimal de contrôlabilité. Nous pensons que la compréhension de ces questions débouchera sur l'amélioration des méthodes de calculs actuellement utilisées pour résoudre ces problèmes.

1.2 Modélisation et analyse

Nous souhaitons considérer de front le problème en petite et grande dimensions, d'un point de vue théorique et numérique, ces deux approches étant complémentaires. Nous comptons débiter par l'étude des systèmes de petite dimension qui présentent l'avantage d'être numériquement peu coûteux à simuler.

Afin de bien cerner les difficultés et d'initier la réflexion sur la détermination d'algorithmes efficaces, nous souhaitons d'abord étudier théoriquement le problème de contrôle optimal et réfléchir à sa modélisation : en effet, on souhaite non seulement atteindre un état cible mais aussi maintenir la cohérence du système, ce qui pose des questions de formalisation du problème d'optimisation. Nous

proposons une (nouvelle) version du problème ci-dessous, mais souhaitons également mettre ce projet à profit pour réfléchir à des alternatives pertinentes.

Il est notable que, jusqu'à présent, ce type de problème a été principalement étudié d'un point de vue numérique, en appliquant des méthodes issues de l'apprentissage non spécifiques à ce problème. Dans cette thèse, nous souhaitons améliorer l'efficacité des approches utilisées en tenant compte de la structure physique du problème. En particulier, à l'aide d'une analyse mathématique destinée à rendre compte de l'influence des paramètres du problème sur la solution et de la sensibilité des quantités d'intérêt (le temps de décohérence, l'état au temps final, etc.), nous envisageons de tester des algorithmes hybrides, couplant des approches de renforcement à des algorithmes de type point-fixe (par exemple, des algorithmes de gradient), rendant compte des conditions d'optimalité.

Il est alors pertinent de chercher à contrôler un tel système, en visant à se rapprocher au mieux d'un état donné, tout en maintenant la cohérence du système quantique. Une première version de ce problème s'écrit :

$$\inf_{T>0, \|u\|_{L^\infty} \leq M} \text{Tr}(C^\top \rho_u(T)) + \alpha T$$

où ρ_u désigne l'équation de Lindblad contrôlée, obtenue à partir de (1) en ajoutant dans le second membre un terme de contrôle sous la forme d'une dynamique hamiltonienne affine en le contrôle, $\alpha > 0$ est une pondération destinée à tenir compte du temps de contrôle et à le prendre en compte dans la minimisation.

Plusieurs approches seront considérées et comparées : on souhaite comparer les approches utilisées jusqu'alors avec une nouvelle approche visant à minimiser un critère tenant compte de l'état final du système bien sûr, mais également du temps de contrôle vu comme une inconnue du problème, de sorte à ne pas dégrader la cohérence du système.

Les questions prospectives auxquelles ce travail vise à répondre concernent :

- la compréhension de la dépendance du contrôle optimal par rapport à la dimension d du système, ainsi que des paramètres du problème.
- en profitant des informations obtenues grâce à la question précédente, l'amélioration des algorithmes existants/la détermination d'un algorithme de résolution performant, en tentant de pallier les difficultés numériques constatées jusqu'alors.

La difficulté de tels problèmes réside dans la compétition entre l'objectif visé et la dissipation naturelle, qui suggère qu'il est nécessaire de contrôler rapidement le système quantique avant que la dissipation ne devienne prédominante (phénomène de *décohérence*, voir par exemple [2]). S'ajoute à ces considérations le fait que des contraintes physiques doivent être considérées sur les contrôles (par exemple, on peut supposer qu'ils sont ponctuellement majorés par une constante uniforme).

Nous comptons alors aborder les questions suivantes, dont les réponses sont également d'intérêt pratique et devraient permettre de guider les études numériques à venir :

- ↪ **Contrôlabilité.** Est-il toujours possible d'atteindre un état stationnaire donné de l'équation en faisant agir le(s) contrôle(s), compte tenu des contraintes imposées? Peut-on prouver que les bornes imposées aux contrôles doivent dépasser un certain seuil pour que la propriété de contrôlabilité exacte soit satisfaite? Pour cette question, nous avons récemment développé dans un autre contexte (le ferromagnétisme) des techniques dont nous pensons qu'elles peuvent être adaptées au contexte qui nous intéresse.
- ↪ **Existence d'un contrôle optimal.** Les problèmes de contrôle optimal ont-ils des solutions? Même question si l'on impose au temps de contrôle T de ne pas dépasser un certain seuil T_1 .
- ↪ **Analyse qualitative.**
 - Quelle est l'influence des paramètres du problème sur la solution? (la taille maximale M des contrôles admissibles, le paramètre α dans la définition de la fonctionnelle, etc.)
 - Comment déduire des conditions d'optimalité données par le principe du maximum de Pontryagin la structure des contrôles optimaux et obtenir un algorithme de minimisation efficace utilisant ces informations?
 - En utilisant le point précédent, peut-on caractériser/estimer le temps de décohérence en fonction de la dimension d du système quantique?

On souhaite également certifier les algorithmes de gradient associés et déterminer si possible des lois de comportement du temps de décohérence en fonction de la dimension. Cela pourrait donner ainsi des pistes supplémentaires sur la façon d'aborder le problème en grande dimension.

1.3 1^{ère} approche : renforcement, principe du maximum de Pontryagin et méthode PINNs

À notre connaissance, les approches de type renforcement ont été étudiées pour des problèmes de contrôle de systèmes quantique en basse dimension [8]-[10]-[11]. Des résultats très préliminaires (non publiés pour le moment) obtenus à l’IPCMS témoignent d’une bonne robustesse de ce type d’approche par rapport à la dimension.

Cependant, il est notable que les contrôles obtenus sont moins efficaces. Cela se traduit de façon pratique par la difficulté de ces méthodes à maintenir la cohérence du système quantique. De plus, les contrôles générés semblent également moins réguliers en temps que ceux générés par les méthodes de gradient ce qui peut également poser des difficultés de convergence des algorithmes.

Les algorithmes de renforcement type TD3 ou DDPG [7], qui sont des candidats naturels pour ce type de problème, sont des méthodes d’apprentissage par renforcement dites *sans modèle* qui partent du principe que l’on ne connaît pas l’environnement sur lequel on souhaite agir. Ces méthodes ne peuvent par conséquent pas calculer ni utiliser la différentielle de l’environnement par rapport au contrôle. En général, les algorithmes de RL commencent par explorer aléatoirement l’espace des contrôles admissibles pour sélectionner leur valeur en chaque temps. En grande dimension, cette exploration peut devenir très fastidieuse et rendre difficile la phase d’apprentissage du contrôle. Dans notre contexte, le modèle met en jeu un système différentiel : il est connu et différentiable par rapport à la variable de contrôle. Nous proposons de modifier l’exploration des méthodes de renforcement en utilisant une exploration *informée par la physique* pour générer les contrôles potentiels. Pour cela, nous pensons utiliser des méthodes de gradient/point fixe utilisant l’état adjoint obtenu à l’aide du principe du maximum de Pontryagin ou des principes de programmation différentiable ainsi que des réseaux de type Physics-informed neural networks (PINNs) pour prédire de courtes séquences de contrôles physiquement pertinents. Cela permettrait également d’imposer de la régularité en temps sur la variable de contrôle et ainsi de guider l’exploration de l’espace des actions. S’il s’avère (comme cela est probable) que cela n’est pas suffisant en grande dimension, on propose de coupler cette approche avec une autre de type *curriculum learning* qui consiste à mettre en œuvre une méthode d’homotopie : on résout des problèmes de Reinforcement Learning (RL) avec difficulté croissante et on utilise comme initialisation le contrôle calculé lors de l’étape précédente. Une possibilité serait de contrôler des problèmes dont l’interaction avec l’environnement est de plus en plus forte (jusqu’à l’interaction cible). Une autre solution serait de paramétrer les contrôles afin de limiter l’espace de fonctions associé et de relâcher progressivement cette paramétrisation afin d’obtenir des espaces de contrôles croissants en taille.

Les méthodes de renforcement dans l’étude préliminaire de l’IPCMS ont été utilisées pour contrôler un seul système quantique (dimension fixée, ainsi que les paramètres). On souhaiterait aussi investiguer la capacité de ces approches à apprendre des contrôles plus généraux (dépendant de la dimension, des paramètres) afin d’effectuer un seul apprentissage pour toute famille de problèmes.

1.4 2^{ème} approche : renforcement et modèle réduit

On propose également de considérer une autre approche déjà étudiée dans un autre contexte. Dans le travail récent [6], dédié au contrôle optimal de chaînes de Markov de très grandes dimensions, appliqué à l’épidémiologie, nous avons proposé une méthode qui se rapproche de l’apprentissage par renforcement pour construire des modèles réduits. Plus précisément, il s’agit de contrôler le système en construisant un modèle nonlinéaire en dimension réduite avec un réseau de neurones (ou des approches certifiées de type réduction d’ordre [3]) et de contrôler le modèle réduit à l’aide de méthodes de descente, ce qui permet d’obtenir un bon candidat pour le modèle global. Si le contrôle obtenu n’est pas assez bon pour le système d’origine, on reconstruit un modèle réduit autour des trajectoires contrôlées récemment obtenues et on itère à nouveau ce processus. Ces itérations sont nécessaires car le modèle réduit a une zone de validité restreinte et est souvent trop éloigné au début de l’algorithme de la zone d’optimalité du contrôle recherché. On alterne donc des étapes de contrôle et des étapes d’amélioration du modèle réduit.

Dans la continuité de travaux récents en analyse numérique et réduction de modèles, on souhaite construire des modèles réduits préservant les propriétés de l’équation et admettant une meilleure stabilité en temps long.

1.5 Application

L’IPCMS travaille sur un système quantique de type *Qudits* particulier. Etant donné qu’une partie conséquente de ce travail sera dédiée à la prospective, au développement d’algorithmes que l’on espère performants pour cette application, nous envisagerons de traiter d’abord des cas académiques (l’étude de systèmes quantiques de petite dimension) issus de la littérature, bien adaptés à cette étape. La connaissance théorique qui sera développée sur ces modèles sera fort utile pour valider ensuite les approches introduites, avant de les tester sur les cas plus complexes étudiés à l’IPCMS.

On commencera ainsi par l’étude d’un ensemble de **qubits** de taille croissante. L’objectif principal de ce travail sera néanmoins le traitement de systèmes contrôlés *Qudits* cibles en commençant par le cas d’un qudit (TbPc2), qui sera envisagé dans un second temps. Si nous obtenons des résultats positifs, nous projetons de collaborer avec l’IPCMS (et plus spécifiquement l’équipe de P.-A. Hervieux) pour mener, en parallèle, des vérifications expérimentales.

2 Collaborateurs

- Clémentine Courtès : analyse/simulation numérique, contrôle, ferromagnétisme, Univ. Strasbourg, équipe INRIA TONUS.
- Paul-Antoine Hervieux : physique quantique, ferromagnétisme, physique des plasmas, Institut de physique et chimie des matériaux de Strasbourg (IPCMS).
- Laurent Navoret : analyse numérique, apprentissage, Univ. Strasbourg, équipe INRIA TONUS.

3 Objectifs

À notre connaissance, les méthodes de renforcement pour améliorer les méthodes de résolution d’EDO/EDP ou de problèmes de contrôle optimal impliquant de tels systèmes sont peu étudiées. La plupart du temps, l’espace de contrôle étant en grande dimension, on est confronté à des difficultés d’exploration de l’espace des contrôles (d’autant plus qu’il est difficile de caractériser quel contrôle a un sens physique ou non). Les approches que l’on propose mélangent les approches de renforcement classiques, les méthodes d’analyse de problèmes d’optimisation impliquant des EDO/EDP et les approches de type informées physiquement qui nous semblent nouvelles et pourraient être appliquées à un grand nombre de problèmes mettant en jeu des EDO/EDP. Les approches basées sur le modèle réduit sont plus classiques, mais leur couplage avec des approches récentes d’apprentissage pourrait mener à de nouveaux algorithmes.

D’un point de vue plus applicatif, un des objectifs serait d’obtenir un résultat concret pour une molécule candidate particulière TbPc2 [4]. Pour le moment il ne semble pas exister de résultats pour cette molécule. Il existe donc un aspect exploratoire aussi en termes de physique.

4 Références

- 1 *A short introduction to the Lindblad master equation*, D. Manzano, AIP Advances, 2020.
- 2 *Decoherence : And the Quantum-To-Classical Transition*, M. A. Schlosshauer, The Frontiers Collection, Springer Science & Business Media, 2007 (416 pages).
- 3 *Low-rank numerical approximations for high-dimensional Lindblad equations*, C. Le Bris and P. Rouchon. Physical Review A covering atomic, molecular, and optical physics and quantum information. 2013.
- 4 *Measuring molecular magnets for quantum technologies*. E. Moreno-Pineda and W. Wernsdorfer. Nature Reviews Physics, 3(9), 645-659 (2021).
- 5 *Optimal control of coupled spin dynamics : design of NMR pulse sequences by gradient ascent algorithms*, T. Schulte-Herbrueggen et co-auteurs. Journal of magnetic resonance, 2005.
- 6 *Reduced modelling and optimal control of epidemiological individual-based models with contact heterogeneity*, C. Courtès, E. Franck, K. Lutz, L. Navoret, Y. Privat. Soumis. 2022.
- 7 *Continuous control with deep reinforcement learning*, T. P. Lillicrap, J. Hunt, A. Pritzel, N. Heess, T. Erez, Y. Tassa, D. Silver, D. Wierstra, 2015

- 8 *Measurement-Based Feedback Quantum Control with Deep Reinforcement Learning for a Double-Well Nonlinear Potential*, S. Borah, B. Sarma, M. Kewming, G. J. Milburn and J. Twamley, Physical Review letters, 2021.
- 9 *Reinforcement Learning : An Introduction*, R. S. Sutton et A. G. Barto. A Bradford Book, coll. « Adaptive Computation and Machine Learning series », 13 novembre 2018.
- 10 *Quantum optimal control of multi-level dissipative quantum systems with Reinforcement Learning*, Z. An, H. J. Song, Q. K. He, D. L. Zhou, Physical Review A, 2021
- 11 *Universal quantum control through deep reinforcement learning*, M. Y. Niu, S. Boixo, V. N. Smelyanskiy, and H. Neven, NPJ quantum information, 2021.