OPTIMIZING MIXED QUANTUM CHANNELS VIA PROJECTED GRADIENT 2 DYNAMICS

Matthew M. Lin
Bing-Ze Lu

NCTS/Math
Technical Report
2025-002

OPTIMIZING MIXED QUANTUM CHANNELS VIA PROJECTED GRADIENT DYNAMICS

MATTHEW M. LIN* AND BING-ZE LU†

Abstract. Designing a mixed quantum channel is challenging due to the complexity of the transformations and the probabilistic mixtures of more straightforward channels involved. Fully characterizing a quantum channel generally requires preparing a complete set of input states, such as a basis for the state space, and measuring the corresponding output states. In this work, we begin by investigating a single input-output pair using projected gradient dynamics. This approach applies optimization flows constrained to the Stiefel manifold and the probabilistic simplex to identify the original quantum channel. The convergence of the flow is guaranteed by its relationship to the Zariski topology. We present numerical investigations of models adapted to various scenarios, including those with multiple input-output pairs, highlighting the flexibility and efficiency of our proposed method.

Key words. probabilistic simplex, projected gradient dynamics, quantum channel, Stiefel manifold

AMS subject classifications. 15A72, 58D15, 65F55, 65H10, 65K05

1. Introduction.

Recent advances in quantum simulators and processors have significantly improved hardware capabilities and measurement techniques. However, fully characterizing quantum dynamics or channels remains a fundamental challenge. To address this, quantum process tomography (QPT), also known as channel identification, provides a systematic framework for reconstructing an unknown quantum process from experimental data. By determining how quantum systems evolve in response to various input and output states, QPT allows us to mathematically describe the process via the notation Φ, which maps an input state ρ to an output state σ:

$$\sigma := \Phi(\rho).$$

In this work, we utilize prior knowledge of input and output states to recover a mixed quantum channel, if it exists, by solving the following optimization problem:

$$\text{Minimize} \quad \frac{1}{2}\|\sigma - \sum_{k=1}^{r} p_k U_k \rho U_k^*\|_F^2, \tag{1.1a}$$

$$\text{subject to} \quad U_k \in \mathcal{S}_n, \quad k = 1, \ldots, r, \tag{1.1b}$$

$$\mathbf{p} \in \Delta^{r-1}, \tag{1.1c}$$

where $\mathcal{S}_n$ denotes the set of n-by-n unitary matrices, Δ^{r-1} represents the probability simplex, denoted as:

$$\Delta^{r-1} := \left\{ \mathbf{p} \in \mathbb{R}^r \mid \mathbf{p} = [p_k] \geq 0, \sum_{k=1}^{r} p_k = 1 \right\}, \tag{1.2}$$

and $\|\cdot\|_F$ is the Frobenius norm.

This formulation, which we refer to as the optimization of mixtures of unitary operations, is a fundamental problem in quantum computing and quantum information theory. It has broad applications, including quantum channel approximation, quantum state synthesis, and noise modeling [11, 13, 4]. For example, consider the depolarizing channel:

$$\Phi(\rho) := (1-p)\rho + \frac{p}{3}\left(X\rho X + Y\rho Y + Z\rho Z\right), \tag{1.3}$$

* Department of Mathematics, National Cheng Kung University, Tainan 701, Taiwan (mhlin@mail.ncku.edu.tw)

† Oden Institute for Computational Engineering and Sciences, The University of Texas at Austin, TX 78712. (bingzelu@utexas.edu).

where p is the probability of depolarization and X, Y, Z are Pauli matrices. Without knowing p or the underlying unitary operations a priori, our method aims to approximate Φ using a convex mixture of unitary operations.

On the other hand, the discrepancy metric utilized in (1.1) is grounded in different operational paradigms and mathematical formulations. Rather than using the Frobenius norm, a more common measure is *fidelity*, defined as:

$$F(\sigma, \rho) = \left(\mathrm{Tr}\left(\sqrt{\sqrt{\sigma}\rho\sqrt{\sigma}}\right)\right)^2,$$

to assess the similarity between two quantum states σ and ρ. However, directly optimizing fidelity can be challenging due to the complexity introduced by the square-root and trace operations. To avoid these difficulties, we adopt the Frobenius norm as a computationally efficient alternative in (1.1). Unlike fidelity, the Frobenius norm simplifies optimization by avoiding complex matrix operations such as nested square roots and matrix traces. Although minimizing the Frobenius norm does not explicitly maximize fidelity, empirical results suggest that it often yields high-fidelity approximations, mainly when the target state and the approximate state are sufficiently close. This makes it a practical option for rebuilding quantum channels while ensuring computational efficiency.

Specifically, this work employs a gradient flow-based method that actively refines the required number of unitary operations. Note that this value r in (1.1) quantifies the complexity of decomposition and is essential to characterize the minimal resources required for tasks such as the quantum channel approximation and state synthesis. To solve the optimal problem (1.1), one crucial aspect is to determine the minimum number r of unitary operations needed for an accurate decomposition. A similar but more theoretical discussion of the minimal decomposition of quantum channels is provided by Lancien and Winter [10], who examine the approximation of quantum channels through completely positive maps with low Kraus rank, providing insights into how these decompositions can be optimized to minimize operational complexity. Beginning with a higher rank r, we demonstrate how to dynamically utilize the gradient flow method to adjust this parameter during the computation process. This approach seeks to make the approximation of a quantum channel computationally feasible using the fewest possible unitary operations.

The remainder of this paper is organized as follows. In Section 2, we present the application of the projected gradient flow to solve the optimization problem. In Section 3, we provide a thorough analysis of the proposed method, proving that the objective function consistently decreases as expected, and establishing key convergence results. In Section 4, we validate our theoretical insight through numerical experiments, including a practical application centered on recovering the depolarizing quantum channel (1.3). Finally, Section 5 offers concluding remarks.

2. Gradient flows. Building upon the standard Euclidean algorithm [1, 8, 2], we introduce a continuous-time flow to solve (1.1). The main advantage of this algorithm is that the individual iterate also stays on the given constraints and simultaneously yields the optimal solution once it converges. Given a fixed rank r, the nearest problem given in (1.1) is to find a probability parameter $\mathbf{p} = (p_1, \ldots, p_r)$ and unitary matrices $U_k \in \mathbb{C}^{n\times n}$, $k = 1, \ldots, r$, such that the objective function

$$f(\mathbf{p}, U_1, \ldots, U_k) := \frac{1}{2}\|\sigma - \sum_{k=1}^{r} p_k U_k \rho U_k^* \|_F^2 \tag{2.1}$$

is minimized. The function f in (2.1) is not analytic unless it is a zero function. Although the function f is not holomorphic (i.e., complex differentiable), we can still compute its derivatives concerning the real and imaginary parts of each variable. To proceed, let $U^{\mathfrak{R}}$ and $U^{\mathfrak{I}}$ denote the real and imaginary parts of the complex matrix U, respectively. Using the concept of Wirtinger derivatives, the following result provides explicit expressions for the components of the derivatives of f concerning the real variables.

THEOREM 2.1. *For $k = 1, \ldots, t$, the components of the derivative of f with respect to the real and imaginary parts, $U_k^{\mathfrak{R}}$ and $U_k^{\mathfrak{I}}$, of U_k and p_k are given as follows:*

$$\begin{cases} \dfrac{\partial f}{\partial U_k^{\mathfrak{R}}} &= 2\mathfrak{Re}\left(p_k\left(\mathcal{A}_k - \sigma\right)U_k\rho\right), \\ \dfrac{\partial f}{\partial U_k^{\mathfrak{I}}} &= 2\mathfrak{Im}\left(p_k\left(\mathcal{A}_k - \sigma\right)U_k\rho\right), \\ \dfrac{\partial f}{\partial p_k} &= \mathfrak{Re}\left(\langle \mathcal{A}_k - \sigma, U_k\rho U_k^*\rangle\right) + p_k\|\rho\|_F^2, \end{cases} \tag{2.2}$$

where $\mathcal{A}_k$ is defined as:

$$\mathcal{A}_k := \sum_{j \neq k} p_j U_j \rho U_j^*.$$

Proof. The objective function g can be equivalently expressed as:

$$\begin{aligned} f(\mathbf{p}, U_1, \ldots, U_k) = \frac{1}{2}\Big(&\langle \sigma - \mathcal{A}_k, -p_k U_k \rho \overline{U_k}^\top\rangle \\ &+\langle -p_k U_k \rho \overline{U_k}^\top, \sigma - \mathcal{A}_k\rangle + \|\sigma - \mathcal{A}_k\|_F^2 + p_k^2\|\rho\|_F^2\Big), \end{aligned} \tag{2.3}$$

where the inner product is defined as:

$$\langle X, Y\rangle := \sum_{i,j=1}^{n} x_{ij}\overline{y_{ij}}$$

for $X, Y \in \mathbb{C}^{n\times n}$. Correspondingly, we let the real-valued inner product over the real field is

$$\langle X, Y\rangle_{\mathbb{R}} := \sum_{i,j=1}^{n} x_{ij} y_{ij}.$$

From direct computation, the derivatives of g are obtained as

$$\begin{aligned} \frac{\partial f}{\partial U_k}\cdot \Delta U &= \langle p_k\overline{(\mathcal{A}_k - \sigma)U_k\rho}, \Delta U\rangle_{\mathbb{R}}, \\ \frac{\partial f}{\partial \overline{U_k}}\cdot \Delta U &= \langle p_k(\mathcal{A}_k - \sigma)U_k\rho, \Delta U\rangle_{\mathbb{R}}, \\ \frac{\partial f}{\partial p_k} &= \mathfrak{Re}\left(\langle \mathcal{A}_k - \sigma, U_k\rho\overline{U_k}^\top\rangle\right) + p_k\|\rho\|_F^2. \end{aligned}$$

Using the properties of Wirtinger derivatives [3, 9], the partial derivatives of g with respect to $U_k^{\mathfrak{R}}$ and $U_k^{\mathfrak{I}}$ are derived as

$$\begin{aligned} \frac{\partial f}{\partial U_k^{\mathfrak{R}}} &= \frac{\partial f}{\partial U_k} + \frac{\partial f}{\partial \overline{U_k}}, \\ \frac{\partial f}{\partial U_k^{\mathfrak{I}}} &= \imath\left(\frac{\partial f}{\partial U_k} - \frac{\partial f}{\partial \overline{U_k}}\right), \end{aligned}$$

which yields the result in (2.2). This completes the proof. ∎

Theorem 2.1 establishes the fundamental derivative information necessary to construct a descent flow for the optimization process. However, the optimization problem posed in (1.1) is a constrained optimization problem in which the descent flow must respect the constraints imposed. Specifically, the flow must be limited to the feasible region defined by the constraints. To explore this in greater depth, let $\gamma(t) = (\mathbf{p}(t), U_1(t), \ldots, U_r(t)$ represent a smooth curve within the domain $\Delta^{r-1} \times \mathcal{S}_{n_1} \times \cdots \times \mathcal{S}_{n_r}$, where $t \in \mathbb{R}$, and assume that $\gamma(0) = (\mathbf{p}, U_1, \ldots, U_r)$ lies in the interior of this domain. This assumption ensures that the initial point of the curve follows all necessary constraints, enabling us to determine how the descent flow advances within the permissible domain.

On the other hand, we observe that the derivative of $f(\gamma(t))$ is given by:

$$\begin{aligned}(2.4)\qquad &\frac{df(\gamma(t))}{dt}\\ &= \sum_{k=1}^{r} \frac{\partial f}{\partial p_k}\frac{dp_k}{dt} + \sum_{k=1}^{r} \operatorname{tr}\left(\left[\frac{\partial f}{\partial U_k^{\mathfrak{R}}}\right]^{\top} \frac{dU_k^{\mathfrak{R}}}{dt}\right) + \sum_{k=1}^{r} \operatorname{tr}\left(\left[\frac{\partial f}{\partial U_k^{\mathfrak{I}}}\right]^{\top} \frac{dU_k^{\mathfrak{I}}}{dt}\right),\end{aligned}$$

where the superscript "$\top$" denotes the transpose of the matrix.

To derive the descent flow from (2.4), we update the tuple $(\mathbf{p}(t), U^{(1)}(t), \ldots, U^{(r)}(t))$ along the trajectory defined by the Euclidean derivative of (2.2). However, since the optimization problem in (1.1) is constrained by the unitary matrix structure and the probability simplex, updates must remain within these domains. To address this, we first consider the fundamental problem:

$$\begin{aligned}(2.5\text{a})\qquad &\text{Minimize} \quad g(U),\\ (2.5\text{b})\qquad &\text{subject to} \quad U \in \mathcal{S}_n.\end{aligned}$$

The gradient of $f(U)$ is expressed as:

$$\nabla g(U) = \frac{\partial g(U)}{\partial U_k^{\mathfrak{R}}} + i\frac{\partial g(U)}{\partial U_k^{\mathfrak{I}}},$$

and the steepest descent direction at $U \in \mathcal{S}_n$ is determined using the real-valued inner product and norm:

$$(2.6)\qquad \langle A, B\rangle_r = \operatorname{Re}(\operatorname{tr}(A^* B)),$$

$$(2.7)\qquad \|A\|_r = \sqrt{\langle A, A\rangle_r},$$

where A and B are $n \times n$ complex matrices and $\operatorname{Re}(\cdot)$ denotes the real part of a complex number. Like (2.4), we see that the steepest descent direction starting from a point $U \in \mathcal{S}_n$ is determined by

$$(2.8)\qquad \xi_U = \underset{\xi \in T_U\mathcal{S}_n, \|\xi\|_U = 1}{\operatorname{argmin}} \langle \nabla g(U), \xi\rangle_U = \frac{-\operatorname{Proj}_{T_U\mathcal{S}_n}(\nabla g(U))}{\|\operatorname{Proj}_{T_U\mathcal{S}_n}(\nabla g(U))\|_U},$$

where $\operatorname{Proj}_{T_U\mathcal{S}_n}(\nabla f(U))$ is the projection of $\nabla f(U)$ onto the tangent space $T_U\mathcal{S}_n$.

Recall that the tangent space $T_U\mathcal{S}_n$ attached to the point U is characterized as:

$$\begin{aligned}T_U\mathcal{S}_n &= \{Z \in \mathbb{C}^{n\times n} : U^* Z + Z^* U = 0\}\\ &= U\mathcal{H}_n^{\perp},\end{aligned}$$

where $\mathcal{H}_n^{\perp}$ is the space of skew-Hermitian matrices (see [7] for further details). Using this, the steepest descent flow on $\mathcal{S}_n$ is given by:

$$(2.9)\qquad \frac{dU(t)}{dt} = -U \operatorname{skew}(U^* \nabla g(U)),$$

where $\operatorname{skew}(A) = \frac{1}{2}(A - A^*)$ represents the skew-Hermitian part of the matrix A.

Similarly, let $\mathbf{p}(t) = [p_i(t)]$ with $\mathbf{p}(0) > 0$ evolve on the probability simplex Δ^{r-1}. To preserve the trace-one property and ensure reducing the objective value in (**??**), we enforce:

$$(2.10)\qquad \sum_{i=1}^{r} \frac{dp_i(t)}{dt} = 0 \quad \text{for all } t \geq 0.$$

Combining these results, we obtain the modified continuous-time descent system for minimizing f given in (1.1):

$$\text{(2.11)} \quad \begin{cases} \dfrac{dU_k}{dt}(t) = -U_k \operatorname{skew}\left(U_k^* \dfrac{\partial f}{\partial U_k}\right), \\ \dfrac{dp_k}{dt}(t) = -\dfrac{\partial f}{\partial p_k} + \dfrac{1}{r}\displaystyle\sum_{\ell=1}^{r} \dfrac{\partial f}{\partial p_\ell}, \end{cases}$$

for $k = 1, \ldots, r$.

By construction, we observe that

$$\frac{d\|U_k\|_F^2}{dt} = 2\operatorname{Re}\left(\left\langle U_k(t), \frac{dU_k(t)}{dt}\right\rangle\right) = 0,$$

where the last equality follows from the fact that $\operatorname{Re}(\langle A, B\rangle) = 0$ for any Hermitian matrix A and skew-Hermitian matrix B. This implies that the trajectory $U_k(t)$ remains bounded in its Frobenius norm for all t and $k = 1, \ldots, r$. On the other hand, note that the optimization problem described in (1.1) imposes two constraints on the coefficients p_k: they must be nonnegative and they must sum to one. In contrast, the dynamics described in (2.11) inherently preserves only the property of summing to one. However, ensuring non-negativity is a manageable challenge. One can address this issue by using standard techniques for solving ordinary differential equations. A well-known and effective approach is to utilize the event detection feature available in MATLAB's built-in "ode" solver. This feature allows us to define a custom event function that monitors the system during integration. Specifically, we construct the event function to detect the exact moment, denoted $\hat{t}$, when any coefficient $p_{\hat{k}}(\hat{t})$, for some index $\hat{k}$, becomes zero. Identifying this critical moment is important because continuing the integration beyond this moment would violate the non-negativity constraint. Furthermore, once $p_{\hat{k}}(\hat{t})$ reaches zero, its contribution to the term $p_{\hat{k}}(\hat{t})U_{\hat{k}}(\hat{t})\rho U_{\hat{k}}(\hat{t})^*$ becomes redundant in evaluating the optimal value. At this point, we pause the iteration and restart the process from the current state, excluding the coefficient $p_{\hat{k}}$ and its associated unitary matrix $U_{\hat{k}}$ from further iterations. This approach ensures the solution's boundedness and enhances computational efficiency by adaptively reducing the problem's dimensionality. We outline the complete procedure in Algorithm 2.1.

Algorithm 2.1 Modified Continuous-Time Descent Flow

1: **Input:** Initial values $\{p_k(0), U_k(0)\}$
2: **Output:** Optimized values $\{p_k, U_k\}$
3: **while** Optimization has not converged **do**
4: Use an ODE solver to integrate (2.11) with initial values $\{p_k(0), U_k(0)\}$.
5: **if** there exists an index $\hat{k}$ and time $\hat{t}$ such that $p_{\hat{k}}(\hat{t}) = 0$ **then**
6: Remove the component $(p_{\hat{k}}, U_{\hat{k}})$ from further optimization.
7: Restart integration with updated initial values $\{p_k(0), U_k(0)\} \leftarrow \{p_k(\hat{t}), U_k(\hat{t})\}$ for $k \neq \hat{k}$.
8: **end if**
9: **end while**

3. Convergent Analysis. When solving (1.1) using the continuous-time differential system (2.11), it is essential to analyze the convergence of its dynamic behavior. To this end, we first note that using the Frobenius norm, the objective function satisfies $f(\mathbf{p}, U_1, \ldots, U_k) \geq 0$ for all $(\mathbf{p}, U_1, \ldots, U_k)$. From (2.11), let $\gamma(t) = (\mathbf{p}(t), U_1(t), \ldots, U_k(t))$. Below, we demonstrate the diminishing behavior of the objective function f by using Algorithm 2.1.

THEOREM 3.1. *Let $\gamma(t) = (\mathbf{p}(t), U_1(t), \ldots, U_k(t))$ represent the flow defined by (2.11). Then, the objective value $f(\gamma(t))$ in (1.1) does not increase over time along this trajectory $\gamma(t)$.*

Proof. Let us explain the result by first analyzing the inner product:

$$
\begin{aligned}
&\operatorname{tr}\left(\left[\frac{\partial f}{\partial U_k^{\mathfrak{R}}}\right]^\top \frac{dU_k^{\mathfrak{R}}}{dt}\right) + \operatorname{tr}\left(\left[\frac{\partial f}{\partial U_k^{\mathfrak{I}}}\right]^\top \frac{dU_k^{\mathfrak{I}}}{dt}\right) \\
&= \operatorname{Re}\left(\left\langle \frac{\partial f}{\partial U_k}, -U_k \frac{\left(U_k^* \frac{\partial f}{\partial U_k} - \frac{\partial f}{\partial U_k}^* U_k\right)}{2} \right\rangle\right) \\
&= \frac{1}{2}\left[-\left\langle \frac{\partial f}{\partial U_k}, \frac{\partial f}{\partial U_k}\right\rangle + \operatorname{Re}\left(\left\langle U_k^* \frac{\partial f}{\partial U_k}, \frac{\partial f}{\partial U_k}^* U_k \right\rangle\right)\right] \\
&\leq \frac{1}{2}\left[-\left\langle \frac{\partial f}{\partial U_k}, \frac{\partial f}{\partial U_k}\right\rangle + \left\| U_k^* \frac{\partial f}{\partial U_k}\right\|_F \left\| \frac{\partial f}{\partial U_k}^* U_k \right\|_F\right] = 0.
\end{aligned} \tag{3.1}
$$

Here, the inequality follows from the Cauchy-Schwarz inequality, ensuring that the result is non-positive.

Second, we observe that for the coefficients p_k, the corresponding inner product satisfies

$$
\begin{aligned}
\sum_{k=1}^r \left\langle \frac{\partial f}{\partial p_k}, \frac{dp_k}{dt}\right\rangle &= \sum_{k=1}^r \left\langle \frac{\partial f}{\partial p_k}, \frac{\partial f}{\partial p_k} + \frac{1}{r}\sum_{\ell=1}^r \frac{\partial f}{\partial p_\ell}\right\rangle \\
&= -\left(\sum_{\ell=1}^r \left\|\frac{\partial f}{\partial p_\ell}\right\|_F^2 - \frac{1}{r}\sum_{\ell=1}^r \frac{\partial f}{\partial p_\ell} \sum_{m=1}^r \frac{\partial f}{\partial p_m}\right) \\
&\leq -\left(1 - \frac{1}{r}\right)\sum_{\ell=1}^r \left\|\frac{\partial f}{\partial p_\ell}\right\|_F^2 \leq 0,
\end{aligned} \tag{3.2}
$$

where the final inequality follows from the Cauchy–Schwarz inequality. Finally, by computing the derivative of f along the trajectory of the solution $\gamma(t)$, we find that:

$$
\frac{d}{dt} f(\gamma(t)) \leq 0,
$$

by applying the results established in (2.4), (3.1), and (3.2) and completes the proof. ∎

COROLLARY 3.2. *Let $\gamma(t) = (\mathbf{p}(t), U_1(t), \ldots, U_k(t))$ represent the flow defined by (2.11), and let $f(\gamma(t))$ denote the objective value in (1.1). Then $\frac{df(\gamma(t))}{dt} = 0$ if and only if $\frac{d\gamma(t)}{dt} = 0$.*

Proof. From (2.4), we observe that $\frac{d\gamma(t)}{dt} = 0$ implies $\frac{df(\gamma(t))}{dt} = 0$.

Furthermore, from (3.1) and (3.2), we deduce that $\frac{df(\gamma(t))}{dt} = 0$ implies

$$
\begin{cases}
\operatorname{Re}\left(\left\langle \frac{\partial f}{\partial U_k}, \frac{dU_k}{dt}\right\rangle\right) = 0, \quad \text{for all } k = 1, \ldots, r, \\
\sum_{k=1}^r \left\langle \frac{\partial f}{\partial p_k}, \frac{dp_k}{dt}\right\rangle = 0,
\end{cases} \tag{3.3}
$$

Additionally, we observe that

$$\begin{aligned}
\left\langle \frac{dU_k}{dt}, \frac{dU_k}{dt} \right\rangle &= \left\langle -U_k \frac{\left(U_k^* \frac{\partial f}{\partial U_k} - \frac{\partial f}{\partial U_k}^* U_k \right)}{2}, -U_k \frac{\left(U_k^* \frac{\partial f}{\partial U_k} - \frac{\partial f}{\partial U_k}^* U_k \right)}{2} \right\rangle \\
&= \frac{1}{4} \left\langle U_k^* \frac{\partial f}{\partial U_k} - \frac{\partial f}{\partial U_k}^* U_k, U_k^* \frac{\partial f}{\partial U_k} - \frac{\partial f}{\partial U_k}^* U_k \right\rangle \\
&= \frac{1}{2} \left(\left\| \frac{\partial f}{\partial U_k} \right\|_F^2 - \mathrm{Re} \left(\left\langle \frac{\partial f}{\partial U_k}, U_k \frac{\partial f}{\partial U_k}^* U_k \right\rangle \right) \right) \\
&= - \mathrm{Re} \left(\left\langle \frac{\partial f}{\partial U_k}, \frac{dU_k}{dt} \right\rangle \right) = 0, \qquad (3.4)
\end{aligned}$$

if $\frac{df(\gamma(t))}{dt} = 0$. Finally, from (3.2), we also have

$$\sum_{k=1}^{r} \left\langle \frac{\partial f}{\partial p_k}, \frac{dp_k}{dt} \right\rangle = 0,$$

indicating that $\left\| \frac{\partial f}{\partial p_\ell} \right\|_F = 0$, i.e., $\frac{\partial f}{\partial p_\ell} = 0$ for $k = 1, \dots, r$, which completes the proof. ∎

LEMMA 3.3. *Let $\phi : \mathbb{R}^n \to \mathbb{R}$ be a function satisfying $\phi(\mathbf{x}) \geq 0$. Furthermore, suppose that its derivative along any trajectory $\mathbf{x}(t)$ governed by*

$$\frac{d\mathbf{x}(t)}{dt} = h(\mathbf{x}(t)) \tag{3.5}$$

satisfies $\frac{d}{dt}\phi(\mathbf{x}(t)) \leq 0$. Furthermore, suppose that at some particular time $\hat{t}$, the condition

$$\left. \frac{d\phi(\mathbf{x}(t))}{dt} \right|_{\hat{t}} = 0$$

holds if and only if

$$\left. \frac{d\mathbf{x}(t)}{dt} \right|_{\hat{t}} = 0,$$

or equivalently, $h(\mathbf{x}(\hat{t})) = 0$, and assume that for all $t \geq 0$, the trajectory $\mathbf{x}(t) \subset \mathbb{R}^n$ is compact and the equilibrium points of the dynamical system (3.5) are isolated. Then, for any initial condition $\mathbf{x}_0$, the solution $\mathbf{x}(t)$ converges to one of these equilibrium points, denoted by hat$\mathbf{x}$, i.e.,

$$\lim_{t \to \infty} \mathbf{x}(t) = \hat{\mathbf{x}}. \tag{3.6}$$

Equivalently, the ω-limit set of $\mathbf{x}(t)$ consists solely of the point $\mathbf{x}^*$.

Proof. First, we show that the ω-limit set is contained in $\left\{ \mathbf{x} : \frac{d}{dt}\phi(\mathbf{x}(t)) = 0 \right\}$. By the definition of the ω-limit point, there exists a strictly increasing sequence $t_k \to \infty$ and a particular point $\hat{\mathbf{x}} \in \mathbb{R}^n$ for which

$$\mathbf{x}(t_k) \to \hat{\mathbf{x}} \quad \text{as } k \to \infty.$$

By assumption, $\phi(\mathbf{x}(t))$ is a continuous and non-increasing function along the trajectory $\mathbf{x}(t)$, which is defined on a compact set. Consequently, for any strictly increasing sequence $s_k \to \infty$, the sequence

$\{\phi(\mathbf{x}(s_k))\}$ is non-increasing. Since $\{\phi(\mathbf{x}(s_k))\}$ is non-increasing and bounded (due to the compactness of the set and continuity of $\{\phi(\mathbf{x}(t))\}$), it converges to a limit, denoted by

$$\lim_{k\to\infty} \phi(\mathbf{x}(s_k)) = \phi^*.$$

Because $\mathbf{x}(t_k) \to x_\infty$ and ϕ is continuous, we have

$$\lim_{k\to\infty} \phi\big(\mathbf{x}(t_k)\big) = \phi(\hat{\mathbf{x}}).$$

Without loss of generality, we assume that for all $k \geq 1$, $t_k \leq s_k$. Since $\{\phi(\mathbf{x}(t))\}$ is a non-increasing sequence for all $t \geq 0$, it follows that $\phi(t_k) \geq \phi(s_k)$ for all $k \geq 1$. Consequently, we have $\phi(\hat{\mathbf{x}}) \geq \phi^*$. Similarly, we can select specific subsequences $\{t_{k_j}\}$ and $\{s_{k_j}\}$ such that $t_{k_j} \geq s_{k_j}$ and show that $\phi^* \geq \phi(\hat{\mathbf{x}})$. Together, these results imply $\phi(\hat{\mathbf{x}}) = \phi^*$. Thus, for any strictly increasing sequence $\{r_k\}$ with $r_k \to \infty$ as $k \to \infty$, we have $\phi(\mathbf{x}(r_k)) \to \phi^*$ as $k \to \infty$. Therefore, $\phi(\mathbf{x}(t)) \to \phi^*$ as $t \to \infty$. Hence, ϕ^* must be a local minimum of $\mathbf{x}(t)$.

On another note, we have

$$\frac{d}{dt}\phi(\mathbf{x}(t)) = \langle \nabla\phi(\mathbf{x}(t)), h(\mathbf{x}(t))\rangle \leq 0 \quad \text{for all } t \geq 0.$$

Furthermore,

$$\lim_{k\to\infty} \langle \nabla\phi(\mathbf{x}(t_k)), h(\mathbf{x}(t_k))\rangle = \langle \nabla\phi(\hat{\mathbf{x}}), h(\hat{\mathbf{x}})\rangle = 0,$$

since $\phi(\hat{\mathbf{x}}) = \phi^*$ is a local minimum. This implies that $h(\hat{\mathbf{x}}) = 0$, i.e., $\hat{\mathbf{x}}$ is an equilibrium of the dynamical system given by (3.5).

We will then demonstrate that $\mathbf{x}(t)$ converges to $\hat{\mathbf{x}}$ as $t \to \infty$. Since the flow $\mathbf{x}(t)$ is continuous and bounded for all $t \geq 0$, suppose that $\mathbf{x}(t)$ has two distinct ω-limit points, $\hat{\mathbf{x}}_1$ and $\hat{\mathbf{x}}_2$, with $\hat{\mathbf{x}}_1 < \hat{\mathbf{x}}_2$. Then, there exist two sequences $\{p_n\}$ and $\{q_n\}$ such that $\mathbf{x}(p_n) \to \hat{\mathbf{x}}_1$ and $\mathbf{x}(q_n) \to \hat{\mathbf{x}}_2$ as $n \to \infty$. Let $\mathbf{y}$ be a point in $(\hat{\mathbf{x}}_1, \hat{\mathbf{x}}_2)$. Then $\mathbf{y}$ must be an ω-limit point. Otherwise, there exist two positive numbers ϵ and t_ℓ such that $|\mathbf{x}(t) - \mathbf{y}| > \epsilon$ for all $t \geq t_\ell$. This contradicts the fact that the flow between $(\hat{\mathbf{x}}_1, \hat{\mathbf{x}}_2)$ must be continuous. Therefore, this implies that every point in $[\hat{\mathbf{x}}_1, \hat{\mathbf{x}}_2]$ is an ω-limit point and an equilibrium point, which contradicts our assumption that the equilibrium points are isolated. This completes the proof. ∎

Furthermore, our convergence analysis counts on the following result for geometrically isolated solutions of a generic polynomial system [12, Theorem 7.1.1].

LEMMA 3.4. *Let $P(\mathbf{z};\mathbf{q})$ be a system of polynomials with variables $\mathbf{z} \in \mathbb{C}^n$ and parameters $\mathbf{q} \in \mathbb{C}^m$. Define $\mathcal{N}(\mathbf{q})$ as the number of geometrically isolated solutions satisfying the condition:*

$$\mathcal{N}(\mathbf{q}) := \#\left\{\mathbf{z} \in \mathbb{C}^n \,\middle|\, P(\mathbf{z};\mathbf{q}) = 0, \det\left(\frac{\partial P}{\partial \mathbf{z}}(\mathbf{z};\mathbf{q})\right) \neq 0\right\}.$$

The following properties hold:

1. *$\mathcal{N}(\mathbf{q})$ is finite and remains constant, denoted as $\mathcal{N}$, for almost all $\mathbf{q} \in \mathbb{C}^m$;*
2. *For all $\mathbf{q} \in \mathbb{C}^m$, it follows that $\mathcal{N}(\mathbf{q}) \leq \mathcal{N}$;*
3. *The subset of $\mathbb{C}^m$ where $\mathcal{N}(\mathbf{q}) = \mathcal{N}$ is a Zariski open set. In other words, the exceptional subset of $\mathbf{q} \in \mathbb{C}^m$ where $\mathcal{N}(\mathbf{q}) < \mathcal{N}$ is an affine algebraic set contained within an algebraic set of dimension $n - 1$.*

Note that the set $\mathbb{R}^n$ is Zariski dense in $\mathbb{C}^n$ [6]. Thus, the properties described above hold for almost all parameters $q \in \mathbb{R}^m$, although the number of isolated solutions of real value of the function varies and is no longer constant. Despite this imperfection, this result is sufficient for our purposes, as it establishes the necessary conditions for the subsequent discussion.

By utilizing Lemma 3.4 and Lemma 3.3, along with the established condition of the boundedness for this dynamical flow (2.11), we can demonstrate the following convergence property.

THEOREM 3.5. *Let $\gamma(t) = (\mathbf{p}(t), U_1(t), \ldots, U_k(t))$ represent the flow defined in equation (2.11). Let γ^* be an ω-limit point of the flow $\gamma(t)$. Then, we have*

$$\lim_{t\to\infty} \gamma(t) = \gamma^* \tag{3.7}$$

almost surely for any initial value $\gamma(0)$.

4. Numerical Experiments. In this section, we present three experiments demonstrating a decreasing trend of the objective function along the defined trajectory while addressing stability concerns. We have implemented our proposed method in MATLAB (version 2024b). For numerical integration, we utilized the *ode15s* function with an absolute tolerance (Abstol) and a relative tolerance (Retol) set to 10^{-12}, which allows the integrator to select the time step size adaptively. The program terminates once the objective function reaches 10^{-17}, and we report the corresponding silhouettes. Despite the non-uniqueness of the approximated quantum channel, we demonstrate how our method can effectively approximate the original channel.

It is known that a mixed unitary quantum channel can have multiple ways of being decomposed. If a quantum channel admits the decomposition

$$\Phi(X) = \sum_{k=1}^{r} p_k U_k X U_k^*,$$

its corresponding Choi representation is given by

$$C(\Phi) = \sum_{k=1}^{r} \mathrm{vec}(\sqrt{p_k} U_k)\mathrm{vec}(\sqrt{p_k} U_k)^*.$$

These two representations are equivalent [5, 13]. Thus, in our subsequent discussion, we use the Choi representation to determine whether different decompositions correspond to the same quantum channel.

Example 1 This example illustrates the effectiveness of our proposed method through two experiments. The first experiment tackles the following optimization problem (1.1) The second experiment solves a similar problem but over multiple pairs of input and output quantum states:

$$\text{Minimize} \quad \frac{1}{2}\sum_{j=1}^{m} \left\| \sigma_j - \sum_{k=1}^{r} p_k\, U_k\, \rho_j\, U_k^* \right\|_F^2, \tag{4.1a}$$

$$\text{subject to} \quad U_k \in \mathcal{S}_n, \quad k = 1, \ldots, r, \tag{4.1b}$$

$$\mathbf{p} \in \Delta^{r-1}, \tag{4.1c}$$

Assume that we have a mixed unitary quantum channel Φ over $\mathbb{C}^{5\times 5}$ that is unknown to our program. We define Φ by randomly generating five unitary matrices U_k and a set of probabilities p_k (which sum up to 1), so that

$$\Phi(X) = \sum_{k=1}^{5} p_k\, U_k\, X\, U_k^*.$$

Next, we create the one-shot data. Let $\rho \in \mathbb{C}^{5\times 5}$ be a randomly generated positive-definite Hermitian matrix, which we regard as the input quantum data. We then set $\sigma = \Phi(\rho)$ to be the corresponding output quantum data. Without knowing the initial number $r = 5$, our numerical procedure sets the initial data with $R = 10$, which is twice the exact low rank, and randomly generates the initial data

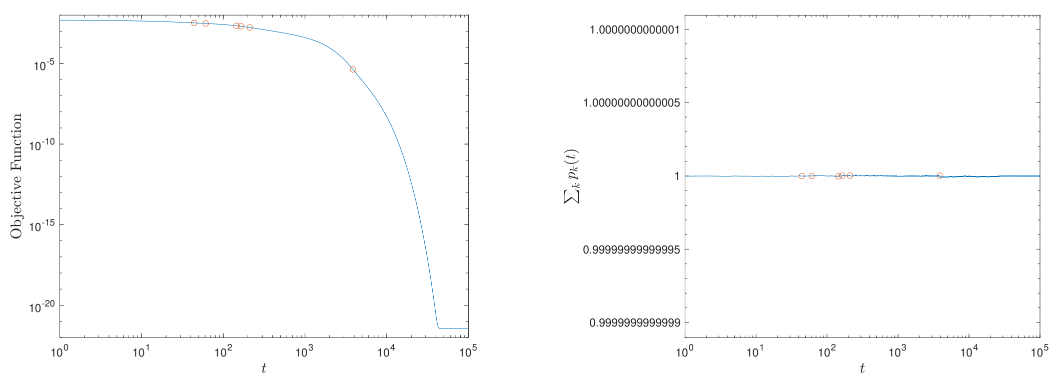
$$\{p_k(0), U_k(0)\}_{k=0}^{R}$$

corresponding to the prescribed structure in (1.1).

By employing Algorithm 2.1, we generate the flow

$$p_k(t), U_k(t)_{k=0}^{R}.$$

Figure 1a verifies that the constructed flow monotonically decreases the objective function, as rigorously established in Theorem 3.1. The red circles in Figures 1a and 1b mark critical moments when any $p_k(t)$ approaches zero, prompting a program restart to ensure the solution remains feasible. Importantly, Figure 1b demonstrates that the sum of $p_k(t)$ consistently equals 1 throughout all iterations, despite the observed minor fluctuations. These fluctuations arise due to numerical integration and rounding errors; however, their magnitudes remain relatively small and close to zero, ensuring the overall reliability of the approach.



(a) The evolution of the objective values of (1.1)

(b) The evolution of $\sum_{k=1} p_k(t)$

Fig. 1: Numerical results of solving (1.1)

The presence of five red circles in Figures 1a and 1b indicates five restarts, which leads to a progressive reduction of the low-rank approximation to 5. Since σ is made up of precisely five quantum channels, the reduction of R from 10 to 5 highlights the effectiveness of our method in identifying and eliminating redundant channels. Additionally, the objective function shows a substantial decrease, ultimately reaching 10^{-20} as t nears 10^5, highlighting the robustness and accuracy of our algorithm in optimizing the problem at hand.

Next, we investigate how multiple datasets can aid in recovering the initial channel Φ. It is important to note that the optimization problem in (1.1) may admit multiple solutions due to the degrees of freedom in U_k exceeding the amount of information the data provides. Based on the contributions of Choi and Jamiołkowski, different quantum channels can correspond to distinct Choi matrix representations. To assess the robustness of our method, we run the optimization 20 times with different initial guesses, then compute the difference between each resulting optimal channel $\hat{\Phi}$ and the original channel Φ using their corresponding Choi matrices; specifically, we evaluate the deviation through the Frobenius norm $\|C(\Phi) - C(\hat{\Phi})\|_F$. The results are shown in 2, where the x-axis represents the difference between the computed optimal channel and the original quantum channel. Although our experiments still reveal that each run successfully reduces the objective function to 10^{-20}, the resulting quantum channel still exhibits deviations from the true channel.

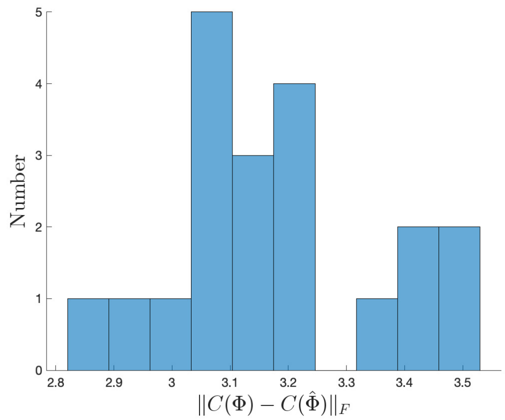
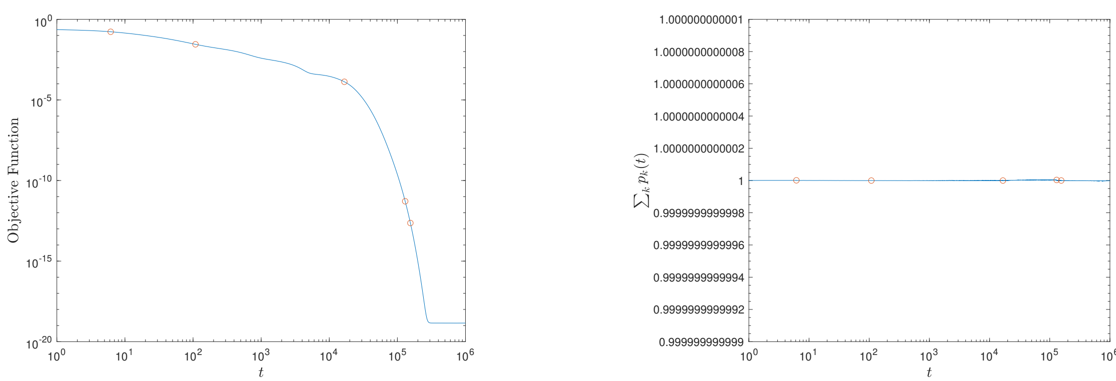


Fig. 2: Values of $\|C(\Phi) - C(\hat{\Phi})\|_F$ collecting from 20 runs

To address this issue, we aim to evaluate the impact of providing additional data pairs to enhance the likelihood of accurately approximating the original quantum channel. This is done by solving the optimization problem in (4.1). Establishing the dynamical system for (4.1) is similar to that given in (2.11), except that we sum over all data pairs; therefore, we omit the entire process for brevity. Specifically, we use the same setup but with $m = 20$ pairs of input and output quantum states. Our primary objective is to verify that the proposed method continues to decrease the objective function while maintaining the sum-to-one property among the p_k. These features are demonstrated in Figure 3a for the descent behavior and in Figure 3b for the sum-to-one property, both of which confirm the applicability of our method to the multi-shot problem in (4.1).



(a) The evolution of the objective values of (4.1)

(b) The evolution of $\sum_{k=1} p_k$

Fig. 3: Numerical results of solving (4.1), where the five circles label the restart occurrence.

Next, we evaluate whether additional data enhances the program's capacity to reconstruct the original quantum channel. We use 100 data points, i.e., $\{\sigma_j, \rho_j\}_{j=1}^{100}$, in each experiment, as expressed in (4.1). This procedure is repeated 20 times. The distribution of the differences $\|C(\Phi) - C(\hat{\Phi})\|_F$ across the 20 runs is shown in Figure 4. In contrast to Figure 2, where the x-axis covers a wider

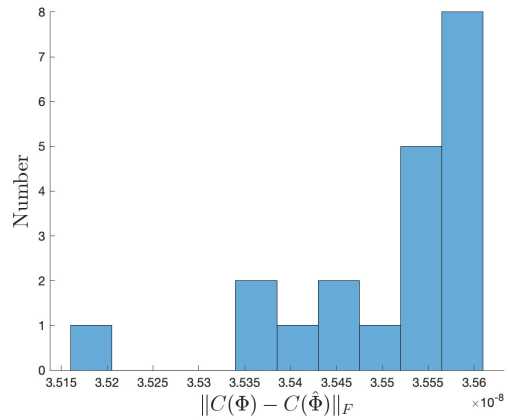


Fig. 4: Distribution of $\|C(\Phi) - C(\hat{\Phi})\|_F$ collected from 20 runs.

range, the x-axis in Figure 4 is centered around 3.5×10^{-8}.The findings indicate that the reconstructed channels, derived from multiple data pairs, closely resemble the original channel in terms of their Choi matrix representations. Therefore, we conclude that our method can effectively reconstruct an unknown mixed unitary quantum channel when enough data pairs are available.

Example 2 The second example examines the depolarizing channel defined in (1.3), where the parameter p controls the strength of the noise. This channel is crucial for simulating errors in quantum information processing, while in quantum error correction, it aids in the design of codes that aim to mitigate the effects of quantum noise.The second example examines the depolarizing channel defined in (1.3), where the parameter p controls the strength of the noise. This channel is essential for simulating errors in quantum information processing and plays a significant role in quantum error correction, helping to design codes that can mitigate the impacts of quantum noise. We set $p = 0.9$ to characterize the channel and evaluate our proposed method by comparing the Choi matrix representation of the learned channel with that of the actual channel. We provide 20 input data points and measure the corresponding outputs using the depolarizing channel to achieve this goal. The experiment is repeated 20 times, as done in previous studies. We then plot the difference distribution between the Choi matrix representations of the actual and approximated channels.

As shown in Figure 5, the difference in the Choi matrix representations between the approximated and actual depolarizing channels is concentrated around 3.2×10^{-9} across the 20 runs. These results demonstrate that our method effectively identifies a quantum channel that closely approximates the unknown channel, resulting in minimal error in the Choi matrix representation.

5. Conclusion. This article proposes a descent flow approach to approximate an unknown unitary quantum channel. We formulate the problem as an optimization task on complex Stiefel manifolds and construct the flow using Wirtinger derivatives. Theoretically, we prove that the flow reduces the objective function while preserving the positivity and sum-to-one properties of the probability distribution p_k, which characterizes the approximated channel. Moreover, we show that the ω-limit points obtained through our method are isolated and correspond to critical points of the objective function. These findings ensure the applicability of our approach in achieving an optimal solution. We validate the effectiveness and stability of our method numerically through experiments. Our approach is extended from single-shot to multi-shot data, and we evaluate the computed accuracy based on the Choi matrix representation of the quantum channel. The numerical results indicate that the proposed method can effectively approximate the unknown quantum channel by using multiple datasets.

References.

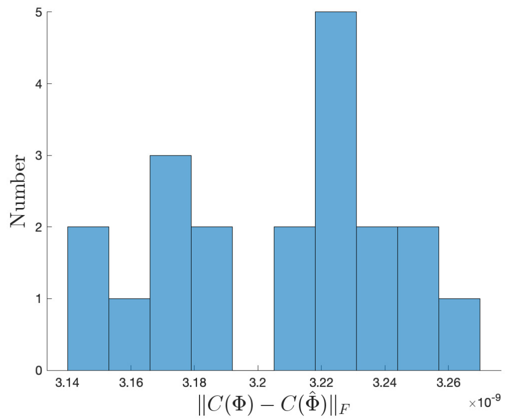


Fig. 5: Distribution of all $\|C(\Phi) - C(\hat{\Phi})\|_F$ collected from 20 runs.

[1] P.-A. Absil, R. Mahony, and R. Sepulchre, *Optimization algorithms on matrix manifolds*, Princeton University Press, Princeton, NJ, 2008, https://doi.org/10.1515/9781400830244, https://doi.org/10.1515/9781400830244. With a foreword by Paul Van Dooren.

[2] N. Boumal, *An introduction to optimization on smooth manifolds*, Cambridge University Press, Cambridge, 2023.

[3] D. H. Brandwood, *A complex gradient operator and its application in adaptive array theory*, IEE Proceedings F: Communications Radar and Signal Processing, 130 (1983), pp. 11–16.

[4] T. A. Brun, *Quantum error correction*, 02 2020, https://doi.org/10.1093/acrefore/9780190871994.013.35, https://oxfordre.com/physics/view/10.1093/acrefore/9780190871994.001.0001/acrefore-9780190871994-e-35.

[5] M.-D. Choi, *Completely positive linear maps on complex matrices*, Linear algebra and its applications, 10 (1975), pp. 285–290.

[6] B. Dong, N. Jiang, and M. T. Chu, *Nonlinear power-like iteration by polar decomposition and its application to tensor approximation*, Numer. Math., 144 (2020), pp. 729–749, https://doi.org/10.1007/s00211-020-01100-8, https://doi.org/10.1007/s00211-020-01100-8.

[7] A. Edelman, T. A. Arias, and S. T. Smith, *The geometry of algorithms with orthogonality constraints*, SIAM J. Matrix Anal. Appl., 20 (1998), pp. 303–353, https://doi.org/10.1137/S0895479895290954, https://doi.org/10.1137/S0895479895290954.

[8] B. Gao, G. Hu, Y. Kuang, and X. Liu, *An orthogonalization-free parallelizable framework for all-electron calculations in density functional theory*, SIAM J. Sci. Comput., 44 (2022), pp. B723–B745, https://doi.org/10.1137/20M1355884, https://doi.org/10.1137/20M1355884.

[9] S. G. Krantz, *Function theory of several complex variables*, AMS Chelsea Publishing, Providence, RI, 2001, https://doi.org/10.1090/chel/340, https://doi.org/10.1090/chel/340. Reprint of the 1992 edition.

[10] C. Lancien and A. Winter, *Approximating quantum channels by completely positive maps with small kraus rank*, Quantum, 8 (2024), p. 1320.

[11] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information: 10th Anniversary Edition*, Cambridge University Press, 2010.

[12] A. J. Sommese and C. W. Wampler, II, *The numerical solution of systems of polynomials arising in engineering and science*, World Scientific Publishing Co. Pte. Ltd., Hackensack, NJ, 2005, https://doi.org/10.1142/9789812567727, http://dx.doi.org.prox.lib.ncsu.edu/10.1142/9789812567727.

[13] J. Watrous, *The Theory of Quantum Information*, Cambridge University Press, 2018.