POLITECNICO
MILANO 1863

SCUOLA DI INGEGNERIA INDUSTRIALE
E DELL'INFORMAZIONE

Executive Summary of the Thesis

Classical and Quantum approaches to Computational Optimal Transport

Laurea Magistrale in Mathematical Engineering

Author: Juan Carlos Gonzalez Espitia

Advisor: Prof. Giulia Cavagnari

Co-advisor: Prof. Francesca Ieva

External advisors: Dr. Jannis Born (IBM Research Zurich), Prof. Dr. Marianna Rapsomaniki (IBM Research Zurich)

Academic year: 2023-2024

---

1. Introduction

The Optimal Transport (OT) problem originates in the late 18th century after Gaspard Monge studied the problem of transporting construction materials efficiently from a source to a target location. In this thesis work we study a more modern framework formulated by, among many others, Leonid Kantorovich in the 1940s. The problem reads as follows. Let $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$ be two probability measures in $\mathbb{R}^d$. We want to find an optimal transport plan $\tilde{\gamma} \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d)$ such that

$$\tilde{\gamma} = \underset{\gamma \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d)}{\arg\min} \int_{\mathbb{R}^d \times \mathbb{R}^d} |x-y|^2 \, d\gamma(x,y), \quad (1)$$

subjected to the constrain that $(\pi_X)_{\#}\gamma = \mu$ and $(\pi_Y)_{\#} = \nu$. Here, the operator $\pi_X : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ is defined as $\pi_X(x, y) := x$; the operator $\pi_Y : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ is defined as $\pi_Y(x, y) := y$; and the push-forward of any Borel map $T : \mathbb{R} \to \mathbb{R}$ is defined as $(T_{\#}\ \mu)(E) = T^{-1}(\mu(E))$ for all $E \in Y$ Borel measurable. This constraint is equivalent to requiring that the marginals of $\tilde{\gamma}$ coincide with the source and target distributions, μ and ν.

In the specific context of Equation (1), Yann Brenier proved that such optimization problem has a unique solution when μ is absolutely continuous with respect to $\lambda(\mathbb{X}^d)$, the Lebesgue measure on $\mathbb{R}^d$, and μ and ν have finite second moments. Additionally, he proved that there exists a convex function $\varphi : \mathbb{R}^n \to \mathbb{R}$ such that it induces an optimal transport plan $\bar{\gamma} \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d)$ corresponding to the solution of the Kantorovich Problem (1) as $\bar{\gamma} = (Id \times \nabla\varphi)_{\#}\mu$. Unfortunately, despite being simple in formulation, OT is computationally expensive because it poses a minimization problem involving probability measures in product spaces. Brenier's theorem and a result on convex duality ease some computational limitations and make it feasible to find numerical solutions in small dimensions. Recently, OT has gained a lot of attention in the field of computational biology for its potential use to model the effect of drugs at a cellular level [4]. These approaches rely on the usage of Input Convex Neural Networks (ICNNs) [1] to explore the space of convex functions and to approximate a solution of (1) by means of Brenier's theorem [3].

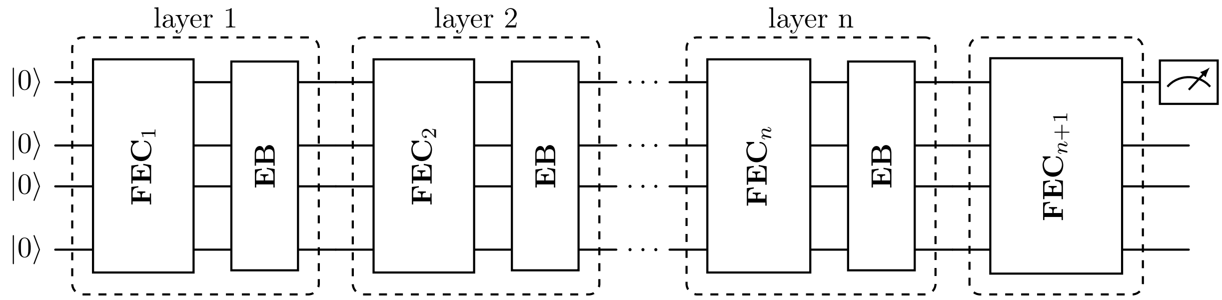

Figure 1: Variational Quantum Circuit used to perform a regression task. The circuit is composed by n layers of one parametrized Fully Entangling Circuit (FEC) followed by one Encoding Block (EB), and an additional FEC at the end. Only the first qubit is measured. On each Encoding Block we input the data as an X-rotation of x radians on each qubit, with x the value of the corresponding data point. The parameters of the circuit are present on the FECs only. Each FEC is made of two independent rotations $R(\theta_x, \theta_y, \theta_z)$ per qubit, and CNOT gates connecting qubits at distances 1 and 2. For a circuit of 4 qubits, each FEC layer has 24 parameters. The total number of parameters of the circuit is $6 * \#qubits * (\#layers + 1)$ and their values are optimized classically via Stochastic Gradient Descend.

Motivated by that application, we study some computational approaches to solve equation (1) using classical and quantum methods.
On the one hand, we propose a new simple classical algorithm to approximate continuous transport maps from discrete samples of $\mu \in \text{supp}(\mu)$ via convex interpolation.
On the other hand, we reframe the convex interpolation used to solve the OT problem as a linear system, and solve it via a Variational Quantum Linear Solver (VQLS). By considering these two approaches we show the feasibility of finding optimal transport maps between probability measures using quantum hardware. However, we stress the challenges and constrains of doing so, and discuss why that implementation is impractical in the quantum devices available today.

2. Methodologies

This work is divided into four different methodologies:

1. Implementation of a Variational Quantum Regression Circuit to perform general regression tasks and analysis of the capabilities to perform convex regression.
2. Development of a classical Convex Interpolation method based on linear solvers.
3. Application of the Variational Quantum Linear Solver to solve the Convex interpolation problem.
4. Retrieving the Optimal continuous Transport Map between empirical measures using the convex splines approach solved via VQLS.

2.1. Variational Quantum Regression Circuit

The ultimate goal of computational OT is to solve Problem (1) using computational tools. When the setting includes data, a common strategy is to rely in Machine Learning methods that implement Neural Networks (NNs) to approximate the solution of the optimization problem. In a analogous way, we try to approximate Solution (1) using Variational Quantum Circuits instead of NNs architectures.
A Variational Quantum Circuit (VQC) is a type of Quantum Circuit (QC) that implements Quantum Gates whose effects are determined by parameters θ. An example of a parameterized quantum gate is $RZ(\theta)$, that rotates the input state θ radians with respect to the z. Usually a VQC is initialized with random circuits, and then they are progressively adjusted so that they improve a certain task. This process is known as learning. In regression, VQCs are trained classically by minimizing the Mean Squared Error (MSE) of the output with respect to the expected result for a particular input x. Explicitely, if we have a training data set $\mathcal{D} = \{x_i, y_i\}_{i=1}^{I}$, the training involves finding a

set of coefficients $\hat{\theta}$ such that

$$\mathrm{MSE}(\mathrm{VQC};\mathcal{D},\theta)=\frac{1}{I}\sum_{i=1}^{I}|VQC(x_i;\theta)-y_i|^2 \tag{2}$$

is minimized.
To solve (2) we need to compute the gradient of the VQC, which is possible thanks to the shift-rule [6], and therefore, a VQC is an appropriate method for solving regression tasks.
We implemented a data re-uploading VQC as depicted and explained in Figure (1). By alternating n layers of Fully Entangling (parametrized) Circuits followed by Data Encoding Blocks, we approximate the desired function by means of Fourier sums, as discussed in [8]. For our purposes, we investigated the potential of this kind of circuits to perform convex regression, and to potentially serve as a quantum replacement for the ICNN.

2.2. Classical Convex Splines

Inspired by the method proposed in [5] we formulate a linear system to solve the convex interpolation problem and use it as an alternative to ICNNs. Given a couple of points, $(x_0,s_0),(x_1,s_1)\in\mathbb{R}^2$, and $s_0',s_1'\in\mathbb{R}$, we want to find $s:\mathbb{R}\to\mathbb{R}$, $s\in C^2(\mathbb{R})$, such that $s(x_0)=x_0, s(x_1)=x_1, s'(x_0)=s_0'$, and $s'(x_1)=s_1'$. Defining $\xi=\frac{x-x_0}{x_1-x_0}$, we write s by means of the following parametrization:

$$s(\xi)=\lambda^n\left(\frac{a}{n(n-1)}\left(\xi-\frac{1-\gamma}{2}\right)^n+\frac{b}{n(n-1)}\left(\xi-\frac{1+\gamma}{2}\right)^n\right)+c\xi+d,$$

where $a,b,c,d\in\mathbb{R}, 0\le\gamma\le 1$, $n\ge 4$ are parameters that control the shape of the curve.
Taking the first derivative of s, imposing the boundary conditions and performing some algebraic manipulations, we can write the problem as $A\zeta=\eta$, where $\zeta=[a,b,c,d]$ is the vector of unknowns, $\eta=[s_0,s_1,s_0',s_1']$, and A is the following matrix:

$$A=\begin{bmatrix}\alpha_0 & \beta_0 & 0 & 1\\ \beta_0 & \alpha_0 & 1 & 1\\ \alpha_1 & -\beta_1 & 1 & 0\\ \beta_1 & -\alpha_1 & 1 & 0\end{bmatrix},$$

with,

$$\alpha_0=\lambda^n\bar{\alpha}_0,\qquad \bar{\alpha}_0=\frac{1}{n(n-1)}\left(\frac{1-\gamma}{2}\right)^n$$
$$\beta_0=\lambda^n\bar{\beta}_0,\qquad \bar{\beta}_0=\frac{1}{n(n-1)}\left(\frac{1+\gamma}{2}\right)^n$$
$$\alpha_1=\lambda^{n-1}\bar{\alpha}_1,\qquad \bar{\alpha}_1=\frac{-1}{(n-1)}\left(\frac{1-\gamma}{2}\right)^{n-1}$$
$$\beta_1=\lambda^{n-1}\bar{\beta}_1,\qquad \bar{\beta}_1=\frac{1}{(n-1)}\left(\frac{1+\gamma}{2}\right)^{n-1}.$$

Therefore, s_0' and s_1' are free parameters that determine the solution of the problem. Indeed, a solution is admissible if and only if the slopes satisfy the following conditions:

$$\frac{\phi_1\bar{s}-\phi_3 s_1'}{\phi_2}<s_0'<\frac{\phi_1\bar{s}-\phi_2 s_1'}{\phi_3} \tag{3}$$
$$\frac{\phi_1\bar{s}-\phi_2 s_0'}{\phi_3}<s_1'<\frac{\phi_1\bar{s}-\phi_3 s_0'}{\phi_2}, \tag{4}$$

with $\phi_1=\bar{\beta}_1-\bar{\alpha}_1$, $\phi_2=\bar{\beta}_0-\bar{\alpha}_1-\bar{\alpha}_0$, $\phi_3=\bar{\alpha}_0+\bar{\beta}_1-\bar{\beta}_0$, and $\bar{s}=\frac{x_1-x_0}{\lambda}$.
Starting from the previous parametrization, we propose a modification to the Slope Averaging Method (SAM) from [5]. Given a dataset with the support points and masses of the empirical measures $\hat{\mu}$ and $\hat{\nu}$, we choose the segment's slopes as a weighted average of adjacent empirical slopes weighted by their relative distance. We call this method the Slopes Weighted Averaging Method (SWAM). Concretely, let us define $\Delta^+(k)=x_{k+1}-x_k$, $\Delta^-(k)=x_k-x_{k-1}$, and

$$\hat{s}^+(k)=\begin{cases}\frac{s_{k+1}-s_k}{\Delta^+(k)} & \text{if } \Delta^+>0\\ 0 & \text{if } \Delta^+=0\end{cases}$$
$$\hat{s}^-(k)=\begin{cases}\frac{s_k-s_{k-1}}{\Delta^-(k)} & \text{if } \Delta^->0\\ 0 & \text{if } \Delta^-=0\end{cases},$$

then, we select the slopes as:

$$\begin{cases}s_1'=(1-\sigma\ sign(\hat{s}_1^+))\ \hat{s}_1^+\\ s_k'=\frac{\Delta^+(k)\ \hat{s}^+(k)+\Delta^-(k)\ \hat{s}^-(k)}{\Delta^-(k)+\Delta^+(k)}, & 2\le k\le K-1\\ s_K'=(1+\sigma\ sign(\hat{s}_K^-))\ \hat{s}_K^-\end{cases}$$

With this method, we are able to interpolate points guaranteeing convexity. In case a set of points don't admit a convex interpolating curve, in our computational implementation we still allow for non-convex interpolating smooth curves, however, the method ensures convexity when possible.

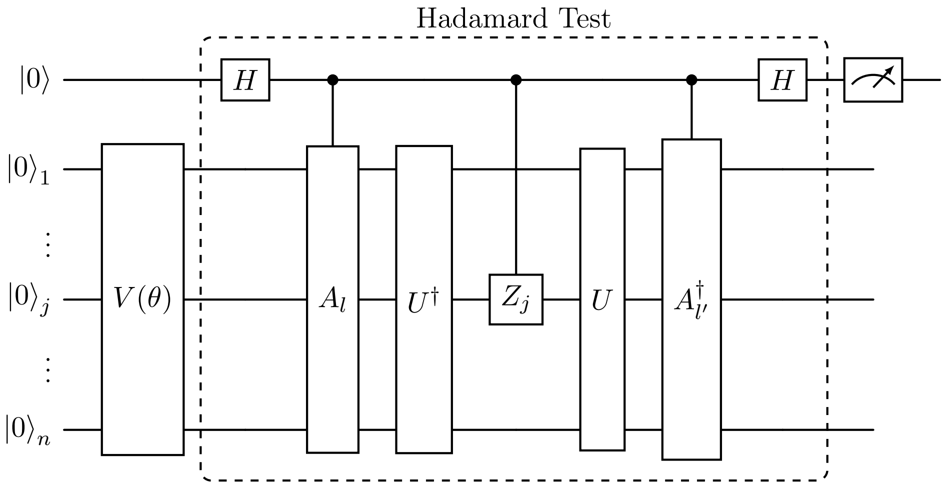

Figure 2: Circuit used for the Variational Quantum Linear Solver as described in [2].

2.3. Variational Quantum Linear Solver

The advantage of the formulation from Section (2.2) is that it re-frames our original problem in the context of solving a linear system. In this section we see how we can solve linear systems using an approach known as a Variational Quantum Linear Solver (VQLS) and introduced by [2].

In this framework, given a matrix A, a known vector b, and an unknown vector x, instead of solving $Ax = b$, the algorithm tries to prepare a quantum state $|x(\theta)\rangle = V(\theta)|0\rangle$ such that it corresponds to the classical solution of the problem. In particular, here we solve:

$$|\varphi(\theta)\rangle = \frac{1}{K} A|x\rangle \approx |b\rangle,$$

where K is a normalization constant introduced to guarantee the that $|x(\theta)\rangle$ the resulting quantum state is properly normalized.

Therefore, we are interested in constructing a Variational Quantum Circuit (VQC) $V(\theta)$ such that the similarity between $\varphi(\theta)$ and $|b\rangle$ is maximized, or alternatively, that $C = 1-|\langle b|\varphi(\theta)|\rangle^2$, is minimized. And so, the optimal parameters for the VQC are:

$$\hat{\theta} = \frac{1}{2}\left[1 - \frac{1}{n}\frac{\langle 0| V^\dagger(\theta) U |0\rangle \langle 0| U^\dagger A V |0\rangle}{\langle 0| V A^\dagger A V |0\rangle}\right],$$

with U a unitary operator such that $U|0\rangle = |b\rangle$. We compute the involved expectation values using the circuit depicted in Figure 2 for the real part, and introduce a $\pi/2$ rotation after the first Hadamard Gate to estimate the imaginary part of the expectation values.

We use classical optimization methods to search for the set of parameters θ such that the cost C is minimized.

2.4. Approximating continuous OT maps via convex splines

Atomic measures lie outside the range of validity of Brenier's theorem because they are not absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}$. However, for the case of empirical measures, it may still be possible to retrieve the continuous optimal transport map T solving the extended continuous problem, if their associated continuous measures satisfy the regularity conditions of the theorem. One possible way to retrieve the continuous transport map is to use a scheme based in numerical integration and differentiation.

Assume that $\mu, \nu \in \mathcal{P}(\mathbb{R})$ are non-atomic measures, and that μ is absolutely continuous with respect to Lebesgue's measure. We denote as $\hat{\mu}$ and $\hat{\nu}$ their empirical distributions. First, we solve the discrete optimal transport problem using a combinatorial approach [7]. The solution to that problem is a discrete collection of points $\hat{T}$ that can be interpreted as the (discrete) optimal transport map linking the empirical measures. Then, we integrate naively and get another collection of points $\hat{\varphi}$ to which we apply our convex splines approach. The resulting (continuous) curve should induce a transportation map $\bar{T} = \hat{\varphi}'$. Therefore, we derive this curve obtain $\bar{T}$ as an approximator of the real continuous

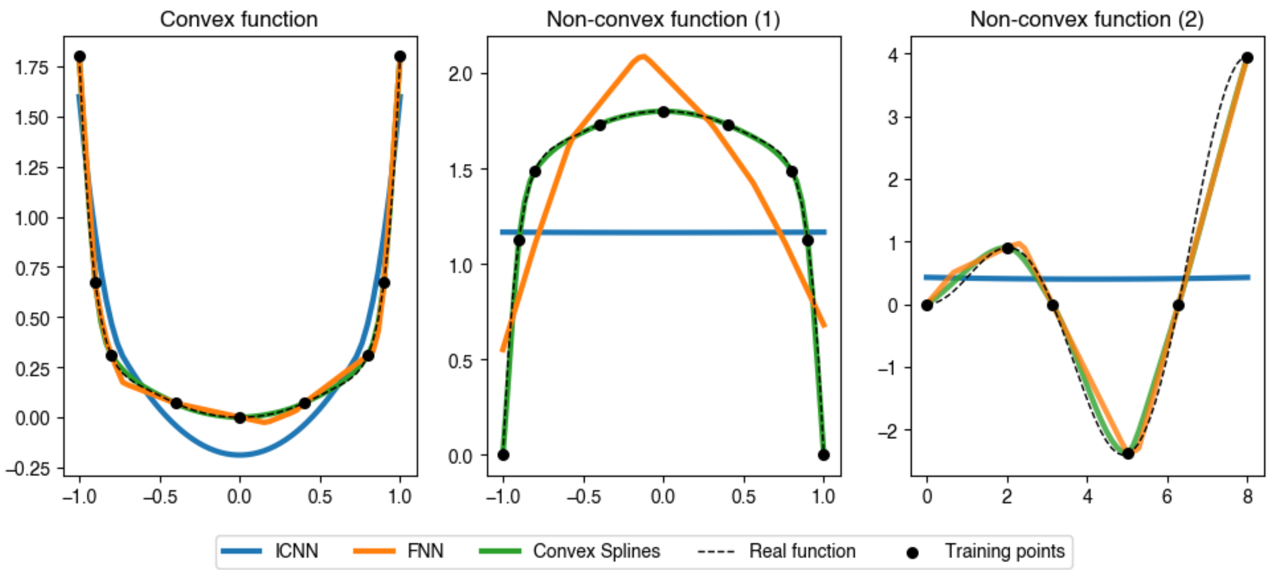

Figure 3: Comparison of ICNN, FNN, and convex splines for the regression of three different functions. On the left we plot a (strictly) convex function in $[-1,1]$ defined by $f(x) = 0.4x^2 + 0.5x^4 - 1.6x^6 + 2.5x^{10}$. In the center we see a nowhere convex function in $[-1,1]$ defined by $f(x) = -(0.4x^2 + 0.5x^4 - 1.6x^6 + 2.5x^{10}) + 1.8$. Finally, in the right we see the function $f(x) = \left(\frac{x}{2}\right) \sin(x)$ for $x \in [0,8]$ that is only convex in -approximately- $[0, 1.076] \cup [3.643, 6.578]$.

transport map T.
The rationale of the method is that, in the continuous case, for μ and ν regular, there exists a unique transport plan $T : \mathbb{R} \to \mathbb{R}$ being the solution of the OT problem, and that additionally, $T = \nabla\varphi$ for some $\varphi : \mathbb{R} \to \mathbb{R}$ convex. Because such map exists, we can approximate it using discrete maps, in the same way as we approximate the continuous distribution using their empirical distributions.

2.5. The OT problem and the VQLS

We connect the estimation of continuous transport plans from empirical measure with the VQLS using the methodology presented in Section 2.4. In this setting it is enough to replace the classical linear solver routine by the VQLS. In this work we proceed in that way.

3. Results and Discussion

As a first result we verify the usefulness of our convex splines method. In Figure 3 we depict three different functions: a convex function on the left, a nowhere convex function in the middle, and a non-convex function with a region of convexity on the right. Our method is shown in green, while in blue and orange we observe the ICNN and FNN architectures used as a baseline. In the three scenarios we used the black points as training data. In the convex case, we evidence that the three methods succeed in approximating the function with a convex function. However, the result of the FNN is not smooth, and the result of the ICNN is not accurate. By incorporating the convexity assumptions implicitly, our method is able to overpass the others, by profiting of the extra information included in the model. In the non-convex scenarios, we see that (1) our method is still able to interpolate the points with a smooth curve and that, in the regions where it is possible, it guarantees convexity; (2) while the FNN has a good accuracy overall, it fails to capture the convexity; (3) because the output of the ICNN has to be everywhere convex, it returns a constant value when incapable of performing the task. With this validation, we proceed with the convex splines as our selected method.
In Figure 4 we see the result of the application of our algorithm to approximate continuous transport maps from empirical measures. The procedure, detailed on the caption of the image, allows us to reconstruct the real continuous optimal transport map from the continuous distribution μ to the continuous distribution ν by only using empirical approximates of them $\hat{\mu}$ and $\hat{\nu}$.

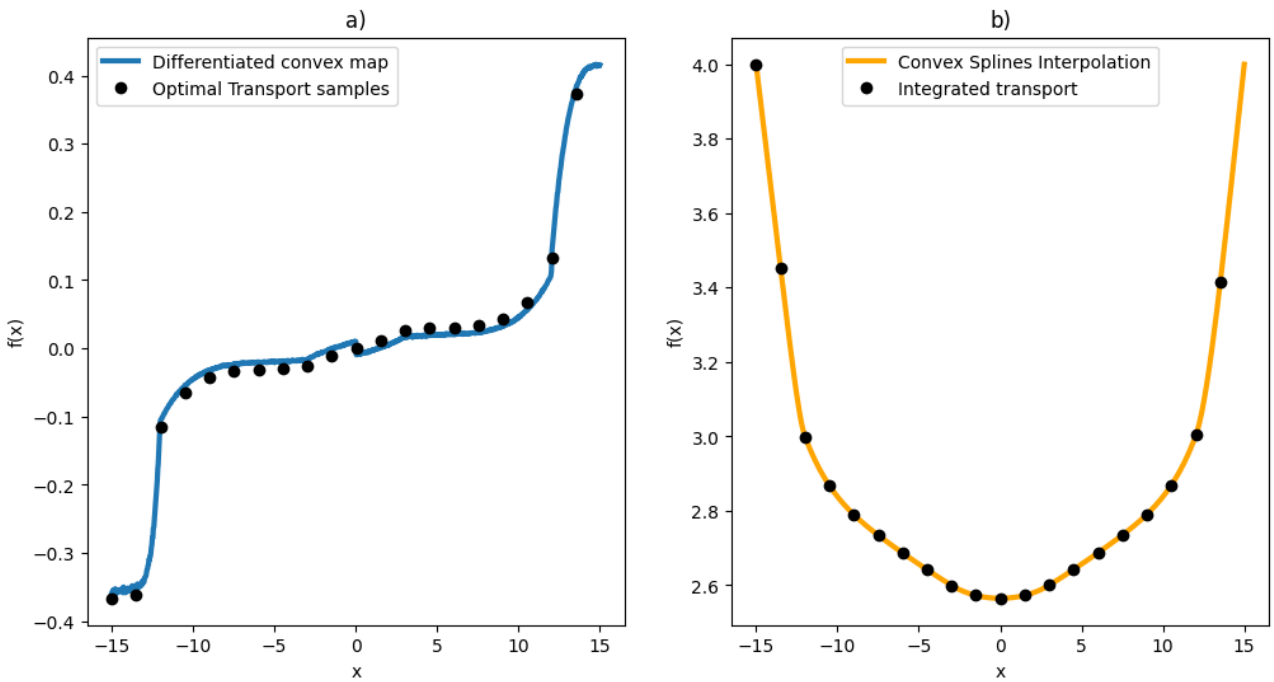

Figure 4: Result of retrieving the continuous transport plan T from discrete samples $\hat{T}$. We start with the black points on the left, that are the result of solving the discrete optimization problem between μ and ν by combinatorial methods. Then, after naive integration, we get the black points in b) to which we apply our convex splines procedure whose result is shown in orange. Finally, by differentiating the orange curve, we obtain the continuous transport map depicted in blue on the left.

Since we are able to achieve the goal, we perform this procedure as our OT solver.
Lastly, we introduce the VQLS as the linear solver for the framework previously described. This allows us to compute the OT map by relying in the implementation and optimization of a Variational Quantum Circuit (VQC). This procedure can be potentially beneficial for scenarios where the dimensionalit of the datasets is so large that it becomes prohibitivy to compute them in classical hardware.

4. Conclusion and further developments

In this works we show that, despite the difficulties, it is possible to approximate continuous Optimal Transport plans from discrete samples using convex splines interpolation, and how Variational Quantum Circuits can be incorporated to that framework via the Variational Quantum Linear Solver (VQLS). We claim that for large datasets whose associated s vector could be easily encoded into a quantum state $|b\rangle$ the VQLS approach might be beneficial thanks to its logarithmic scaling in the number of segments that is superior to the quadratic scaling of common linear solver algorithms.
As a continuation of this work, we propose the following extensions and further developments:

1. To explore different data encoding mechanisms.
2. To use a combinatorial optimization quantum algorithm to solve the discrete OT problem.
3. To train a Variational Quantum Circuit with loss penalizing non-convex predictions.

A rigorous development of the last points will greatly contribute to answering the question about the feasibility of the usage of quantum circuits to solve OT problems.

References

[1] Brandon Amos, Lei Xu, and J Zico Kolter. Input convex neural networks. In *International Conference on Machine Learning*, pages 146–155. PMLR, 2017.

[2] Carlos Bravo-Prieto, Ryan LaRose, Marco Cerezo, Yigit Subasi, Lukasz Cincio, and

Patrick J Coles. Variational quantum linear solver. *Quantum*, 7:1188, 2023.

[3] Yann Brenier. Polar factorization and monotone rearrangement of vector-valued functions. *Communications on pure and applied mathematics*, 44(4):375–417, 1991.

[4] Charlotte Bunne, Stefan G Stark, Gabriele Gut, Jacobo Sarabia Del Castillo, Mitch Levesque, Kjong-Van Lehmann, Lucas Pelkmans, Andreas Krause, and Gunnar Rätsch. Learning single-cell perturbation responses using neural optimal transport. *Nature Methods*, 20(11):1759–1768, 2023.

[5] Achille Messac and Arun Sivanandan. A new family of convex splines for data interpolation. *Computer aided geometric design*, 15(1):39–59, 1997.

[6] Kosuke Mitarai, Makoto Negoro, Masahiro Kitagawa, and Keisuke Fujii. Quantum circuit learning. *Physical Review A*, 98(3):032309, 2018.

[7] Gabriel Peyré, Marco Cuturi, et al. Computational optimal transport: With applications to data science. *Foundations and Trends® in Machine Learning*, 11(5-6):355–607, 2019.

[8] Maria Schuld, Ryan Sweke, and Johannes Jakob Meyer. Effect of data encoding on the expressive power of variational quantum-machine-learning models. *Physical Review A*, 103(3):032430, 2021.