

A reduced-order model for parametrized Optimal Transport problems

Elise Bonnet-Weill*, Virginie Ehrlacher[†], Luca Nenna^{‡§}

April 10, 2026

Abstract

In this work, we aim at efficiently solving a parametrized family of optimal transport problems by using model order reduction methods. We propose a reduced-order model by adding to the primal (respectively dual) version of the high-fidelity model the additional constraint to live in a non negative sub cone (resp. in subspaces) of small dimension. The reduced-order model then reads as a linear program with a small number of degrees of freedom and constraints. We identify explicit conditions under which this reduced-order model has at least one solution. We propose two a posteriori error estimations that bounds the error between the optimal values of the high-fidelity problem and the reduced-order model. As one of these estimations requires the computation of non linear terms (with respect to the reduction of dimension), we use an Empirical Interpolation Method (EIM) (see e.g. [33] or [2]) to numerically efficiently compute this estimation. We apply the whole methodology on a simple 1D example and on a problem of color transfer between images, and compare its performances to Sinkhorn algorithm.

Notations

We present here notations that are used throughout the article without recalling their meaning.

- Given two real matrices \mathbf{B} and \mathbf{E} of dimension $\mathcal{N}_x \times \mathcal{N}_y$, where $\mathcal{N}_x \in \mathbb{N}^*$ and $\mathcal{N}_y \in \mathbb{N}^*$ we denote by $\langle \mathbf{B}, \mathbf{E} \rangle$ the Frobenius scalar product between \mathbf{E} and \mathbf{B} .
- For $\mathcal{N} \in \mathbb{N}^*$, we define the ℓ_1 norm on $\mathbb{R}^{\mathcal{N}}$ saying that for every vector $\mathbf{w} \in \mathbb{R}^{\mathcal{N}}$,

$$\|\mathbf{w}\|_1 := \sum_{i=1}^{\mathcal{N}} |\mathbf{w}_i|.$$

- For $\mathcal{N} \in \mathbb{N}^*$, we define the ℓ_∞ norm on $\mathbb{R}^{\mathcal{N}}$ saying that for every vector $\mathbf{w} \in \mathbb{R}^{\mathcal{N}}$,

$$\|\mathbf{w}\|_\infty := \max_{1 \leq i \leq \mathcal{N}} |\mathbf{w}_i|.$$

- For a real number $x \in \mathbb{R}$, $\lfloor x \rfloor$ designates the integer part of x .

1 Introduction

The numerical simulation of complex physical, economic, or biological systems often requires the repeated solution of parametric problems under varying conditions — a task that rapidly becomes computationally prohibitive when state-of-the-art discretization methods are employed. In this context,

*CERMICS, CNRS, ENPC, Institut Polytechnique de Paris, ParMA, Inria Saclay, Marne-la-Vallée, France. email : elise.bonnet-weill@enpc.fr

[†]CERMICS, CNRS, ENPC, Institut Polytechnique de Paris, Marne-la-Vallée, France

[‡]Université Paris-Saclay, CNRS, Laboratoire de mathématiques d'Orsay, ParMA, Inria Saclay, 91405, Orsay, France. email: luca.nenna@universite-paris-saclay.fr

[§]Institut Universitaire de France (IUF).

Model Order Reduction (MOR) has emerged as a fundamental strategy to construct low-dimensional surrogate models that accurately reproduce the behavior of high-fidelity numerical simulations at a fraction of their computational cost. Among the various MOR strategies, Reduced Basis (RB) methods have established themselves as a mathematically rigorous and computationally efficient framework, particularly well-suited to parametric partial differential equations (PDEs). The aim of the present work is to propose a new reduced basis strategy to build reduced-order models to accelerate the resolution of parametrized Optimal Transport problems. We first give some introduction to reduced basis methods and numerical approaches for the resolution of optimal transport problems and then outline the main contributions of our paper.

Reduced Basis Methods. The reduced basis methodology was firstly introduced in the pioneering works of [1] and [38] in the context of structural mechanics, and subsequently made mathematically rigorous through the contributions of [42]. The central idea is to approximate the solution manifold — the set of all solutions to a parametric problem as the parameter varies over its admissible domain — by a low-dimensional affine subspace, the so-called reduced basis space, constructed from a carefully selected set of high-fidelity snapshot solutions. The RB approximation is then sought in this surrogate space via a Galerkin or Petrov–Galerkin projection, yielding a system of dramatically reduced size that can be solved orders of magnitude faster than the original truth discretization. A theoretical foundation for the convergence of the greedy construction algorithm was later laid by [8], and further refined by [6]. A cornerstone of the RB methodology is the availability of rigorous and efficiently computable a posteriori error estimators, which certify the accuracy of the reduced solution with respect to the high-fidelity reference and drive the greedy algorithm used for basis construction. Such estimators, typically based on dual norms of residuals, were developed for parabolic problems by [21] and for finite-volume evolution schemes by [23]. These estimators enable an offline-online computational decomposition: the expensive pre-computations (basis generation, affine decomposition of operators) are performed once during an offline phase, while the online phase — consisting in solving the reduced system and evaluating the error bound for a new parameter instance — is performed at marginal cost, independent of the dimension of the high-fidelity space. The method has since been systematically exposed in textbook form by [24] and by [43], and its general theory for affinely parametrized elliptic coercive PDEs is thoroughly surveyed in [45]. Extensions to non-affinely parametrized problems, such as those arising in geometry optimization, have been addressed through the Empirical Interpolation Method (EIM) introduced by [2], which approximates non-affine coefficient functions by a low-rank collateral expansion and permits the offline-online decomposition to be maintained. Non-affine and nonlinear parametric dependence was treated within the full RB framework by [20]. Beyond the classical greedy construction, the Proper Orthogonal Decomposition (POD) provides an alternative, singular-value-decomposition-based approach to extracting dominant modes from a set of snapshots [50, 22].

Numerical Methods for Optimal Transport. Optimal Transport (OT) is a classical mathematical theory, dating back to Monge [37] and Kantorovich [27], that formalizes the problem of finding the most efficient way to redistribute one probability measure into another with respect to a given cost function. The monographs by Villani [48, 49] and Santambrogio [46], and the computational review by Peyré and Cuturi [40], provide comprehensive accounts of the theory and its modern applications, which span machine learning, image processing, computational fluid mechanics, economics, and biology.

At the heart of the theory lies the Kantorovich dual problem, whose solutions — the Kantorovich potentials — are functions whose gradients (in the quadratic cost case) recover the optimal transport map through Brenier’s theorem. The regularity theory of these potentials is well-understood under smoothness assumptions on the source and target densities, tracing back to Caffarelli’s work on the Monge–Ampère equation and generalized to a large class of cost functions by [32]. Quantitative stability of optimal transport maps and their associated potentials with respect to perturbations of the input measures has been studied by [19] and, in a quantitative and linearized form amenable to data science applications, by Mérigot, Delalande and Chazal [35]. This stability theory is fundamental: it provides the mathematical basis for expecting the map sending a parametric family of measures to the corresponding Kantorovich potentials — to exhibit regularity in the parameter, a prerequisite for any successful model reduction strategy.

From a numerical point of view, the discrete formulation of OT leads to a linear programming (LP)

problem — the classical transportation problem — whose computational complexity fast increases with the dimension of the problem. This has motivated the development of more scalable approaches. The entropic regularization of OT, introduced in its modern form in [11, 18, 5], replaces the LP by a strictly convex optimization problem solvable via the Sinkhorn–Knopp matrix-scaling algorithm at a speed faster than LP-based solvers. The resulting Sinkhorn distances have become ubiquitous in machine learning, despite the bias introduced by regularization. De-biased versions, including the Sinkhorn divergences of [17], restore (some) properties of a genuine distance while retaining computational efficiency.

Alternative numerical approaches include semi-discrete OT methods, which handle the case where one measure is continuous and the other discrete through Laguerre cell decompositions and Newton-type or quasi-Newton solvers, as developed by [34] and [31]; methods based on the dynamic Benamou–Brenier fluid mechanics formulation [4], which reformulates OT as a convex optimization over flows and has led to efficient proximal splitting algorithms; and approaches grounded in the theory of semi-smooth Newton methods for the Monge–Ampère equation. A unified treatment of discretization methods and their algorithms, covering the assignment problem, entropic regularization, and semi-discrete OT, is provided in the handbook chapter by [36]. Sliced Wasserstein distances offer yet another computationally attractive alternative based on one-dimensional projections.

Parametric Optimal Transport and Model Reduction. Despite this rich landscape of numerical methods, the problem of computing Kantorovich potentials repeatedly for varying parameter values — arising for instance in uncertainty quantification, shape optimization, multi-marginal OT, or Wasserstein barycenter computation — remains computationally intensive. Each parameter query requires solving an independent optimization problem of large scale, rendering parameter sweeps, sensitivity analyses, or real-time applications out of reach with standard solvers. The object of primary interest in this paper is therefore the parametric map which sends a parametric family of measures to the corresponding family of Kantorovich potentials. This is a fundamentally different perspective from the recent literature on MOR for transport-dominated PDEs [9, 7], which uses OT as a tool to improve reduced basis constructions for PDE solutions that are merely dominated by advection; here, the optimal transport problem itself is parametric, and the Kantorovich potential is the quantity of interest.

This shift in viewpoint introduces specific mathematical challenges. First, Kantorovich potentials are only defined up to an additive constant, so the solution map is multi-valued and must be normalized appropriately for a well-posed reduction problem. Second, the derivation of rigorous a posteriori error estimators for the Kantorovich potential — whose accuracy is typically measured in terms of the duality gap or the Wasserstein distance — requires adapting the residual-based methodology of classical RB methods to the OT setting. While the stability theory of Optimal Transport provides qualitative guidance, quantitative and computable error bounds for reduced approximations of Kantorovich potentials have not yet been established.

Recent efforts have begun to address the broader challenge of combining model reduction with optimal transport, forming a rich and rapidly growing literature. A pioneering contribution in this direction is [25], who proposed to use solutions of Monge–Kantorovich mass transfer problems to extract advection modes — low-order representations of transport phenomena that are intrinsically elusive for standard POD-based methods. This seminal work demonstrated that OT provides a geometrically meaningful way to encode advection-dominated dynamics into reduced descriptions, and motivated a subsequent line of research on registration-based model reduction. This registration viewpoint is formalized in [47] into a general equation-independent framework for parameterized MOR, constructing parameter-dependent bijective mappings that align coherent solution features — shocks, shear layers — into a fixed reference configuration before applying linear compression. We refer the reader to extensions of this registration-based approach using OT in [26, 10]. Complementing these registration-based approaches, the theoretical groundwork for nonlinear model reduction in Wasserstein spaces was laid in [15], using Wasserstein barycenters and tangent-space approximations for conservative PDEs. Several extensions were considered, we refer the reader to [3, 7, 14, 12, 39]. We highlight in particular the work [29] where the authors proposed a ROM framework integrating OT theory and neural-network methods via kernel POD with the Wasserstein distance.

However, all of these contributions — whether they use OT to construct registration mappings for PDE solutions, or address model reduction in Wasserstein spaces for PDE solutions — treat op-

timal transport as a computational tool for handling transport-dominated PDE fields. None of them addresses the problem of directly and efficiently computing parametric Kantorovich potentials themselves, which is the focus of the present work. The only work we are aware of in this direction is the contribution [30] where parametrized optimal transport problems were solved by means of neural networks and machine learning techniques.

Contributions of this work. In this paper, we develop a novel reduced basis method for parametric optimal transport problems. More precisely, given a parametric family of optimal transport problems where the measures vary with a parameter, our framework constructs a low-dimensional approximation space for the associated family of Kantorovich potentials through an offline procedure (typically POD). To the best of our knowledge, the reduced basis approximation of parametric Kantorovich potentials, with certified a posteriori error estimation, has not yet been addressed in the literature. Our reduced-order model actually reads as a primal-dual approach with a reduced number of degrees of freedom. As a by-product, we also obtain reduced approximations of the parametric optimal transport plans. The key contributions are:

- (i) a well-posed formulation of the parametric Kantorovich dual suited to RB approximation;
- (ii) the derivation of computable and rigorous a posteriori error estimators enabling reliable online evaluation;
- (iii) an efficient offline-online decomposition that makes the online cost independent of the high-fidelity discretization dimension.

We demonstrate the efficiency and accuracy of the proposed approach on a range of parametric OT benchmarks, achieving speedups of several orders of magnitude, in particular for color transfer applications.

Outline. The remainder of the paper is organized as follows. In Section 2, we present the high-fidelity parametric optimal transport problem to be reduced. In Section 3, we present the reduced order model. Section 4 is devoted to the derivation of a posteriori error estimators and their mathematical analysis. Lastly, in Section 5, we illustrate the efficiency of our approach on various numerical test cases, in particular for color transfer applications.

2 High-fidelity Optimal Transport Problem

2.1 Presentation of the prototypical problem

In this section, we introduce the parametrized family of discrete optimal transport problems under consideration. We then highlight the continuous dependence of the optimal values with respect to the parameter.

Let $X = \{x_1, \dots, x_{N_x}\}$ and $Y = \{y_1, \dots, y_{N_y}\}$ be two finite sets of size $N_x \in \mathbb{N}^*$ and $N_y \in \mathbb{N}^*$, respectively, $c : X \times Y \rightarrow \mathbb{R}_+$ be a continuous cost function and $\mathbf{C} = \{\mathbf{C}_{ij}\}_{\substack{1 \leq i \leq N_x \\ 1 \leq j \leq N_y}} \in \mathbb{R}^{N_x \times N_y}$ the

associated cost matrix: $\mathbf{C}_{i,j} := c(x_i, y_j)$ for $1 \leq i \leq N_x$ and for $1 \leq j \leq N_y$.

Let $K_x, K_y \in \mathbb{N}^*$, and let $\{\boldsymbol{\mu}_k\}_{1 \leq k \leq K_x}$ and $\{\boldsymbol{\nu}_l\}_{1 \leq l \leq K_y}$ be two sets of probability measures on X and Y defined such that for $1 \leq k \leq K_x$, $\boldsymbol{\mu}_k = \sum_{i=1}^{N_x} \mu_k^i \delta_{x_i}$ and for $1 \leq l \leq K_y$, $\boldsymbol{\nu}_l = \sum_{j=1}^{N_y} \nu_l^j \delta_{y_j}$. We define the set of parameters \mathcal{A} by:

$$\mathcal{A} := \left\{ \boldsymbol{\alpha} = (\boldsymbol{\alpha}^x, \boldsymbol{\alpha}^y) \in \mathbb{R}_+^{K_x + K_y}, \sum_{k=1}^{K_x} \boldsymbol{\alpha}_k^x = 1, \sum_{l=1}^{K_y} \boldsymbol{\alpha}_l^y = 1 \right\}. \quad (1)$$

For a given parameter $\boldsymbol{\alpha} = (\boldsymbol{\alpha}^x, \boldsymbol{\alpha}^y) \in \mathcal{A}$, we define the parametrised measures $\boldsymbol{\mu} : \mathcal{A} \rightarrow \mathcal{P}(X)$ (respectively $\boldsymbol{\nu} : \mathcal{A} \rightarrow \mathcal{P}(Y)$) as the convex combinations of $\{\boldsymbol{\mu}_k\}_{1 \leq k \leq K_x}$ (resp. $\{\boldsymbol{\nu}_l\}_{1 \leq l \leq K_y}$) with

respect to the weights α^x (resp. α^y), that is

$$\begin{aligned}\mu(\alpha) &= \sum_{k=1}^{K_x} \alpha_k^x \mu_k, \\ \nu(\alpha) &= \sum_{l=1}^{K_y} \alpha_l^y \nu_l.\end{aligned}\tag{2}$$

The parametrized high-fidelity (HF) optimal transport problem writes as follows:

$$\begin{aligned}I(\alpha) := & \min_{\mathbf{\Pi} \in \mathbb{R}_+^{\mathcal{N}_x \times \mathcal{N}_y}} \langle \mathbf{\Pi}, \mathbf{C} \rangle. \\ & \sum_{j=1}^{\mathcal{N}_y} \mathbf{\Pi}_{i,j} = \mu_i(\alpha) \text{ for } 1 \leq i \leq \mathcal{N}_x \\ & \sum_{i=1}^{\mathcal{N}_x} \mathbf{\Pi}_{i,j} = \nu_j(\alpha) \text{ for } 1 \leq j \leq \mathcal{N}_y\end{aligned}\tag{3}$$

Here, it will be more convenient to rewrite this problem in the standard form of a linear program, as in [40]: we cast $\mathbf{\Pi} \in \mathbb{R}^{\mathcal{N}_x \times \mathcal{N}_y}$ (respectively $\mathbf{C} \in \mathbb{R}^{\mathcal{N}_x \times \mathcal{N}_y}$) as a vector $\boldsymbol{\pi} \in \mathbb{R}^{\mathcal{N}_x \mathcal{N}_y}$ (resp. $\mathbf{c} \in \mathbb{R}^{\mathcal{N}_x \mathcal{N}_y}$) by stacking its columns, and define the matrix \mathbf{A} , encoding the constraints, by

$$\mathbf{A} = \begin{pmatrix} \mathbb{1}_{\mathcal{N}_x}^T \otimes \mathbf{Id}_{\mathcal{N}_y} \\ \mathbf{Id}_{\mathcal{N}_x} \otimes \mathbb{1}_{\mathcal{N}_y}^T \end{pmatrix},\tag{4}$$

where \mathbf{Id}_\bullet denotes the identity matrix of \mathbb{R}^\bullet , $\mathbb{1}_\bullet$ the one vector of \mathbb{R}^\bullet , and \otimes the Kronecker product. The problem (3) re-writes, then, in its standard form:

$$\boxed{I(\alpha) := \min_{\boldsymbol{\pi} \in \mathbb{R}_+^{\mathcal{N}_x \mathcal{N}_y}} \langle \boldsymbol{\pi}, \mathbf{c} \rangle. \\ \mathbf{A}\boldsymbol{\pi} = \begin{pmatrix} \boldsymbol{\mu}(\alpha) \\ \boldsymbol{\nu}(\alpha) \end{pmatrix}}\tag{P_\alpha}$$

The dual associated problem is :

$$\boxed{J(\alpha) = \max_{(\boldsymbol{\varphi}, \boldsymbol{\psi}) \in \mathbb{R}^{\mathcal{N}_x + \mathcal{N}_y}} \langle \boldsymbol{\varphi}, \boldsymbol{\mu}(\alpha) \rangle + \langle \boldsymbol{\psi}, \boldsymbol{\nu}(\alpha) \rangle. \\ \mathbf{A}^T \begin{pmatrix} \boldsymbol{\varphi} \\ \boldsymbol{\psi} \end{pmatrix} \leq \mathbf{c}}\tag{D_\alpha}$$

The problem P_α always has solutions, as it is a minimization problem in finite dimensions of a continuous function on a non-empty, closed and bounded set. By strong duality, we deduce that the problem D_α also has solutions and that for any parameter $\alpha \in \mathcal{A}$, the optimal values $I(\alpha)$ and $J(\alpha)$ are finite, and $I(\alpha) = J(\alpha)$.

2.2 Regularity of the cost with respect to the parameter α

We now turn our attention on the dependence of the optimal values with respect to the parameter α , showing, in particular, that, $I(\alpha)$ (or $J(\alpha)$) is continuous in α . Notice that we will exploit this property in Section 4.2 to derive an a posteriori error estimation. To get the desired result, we first derive a bound on $|I(\alpha) - I(\alpha')|$ which depends on the solutions of the dual problem D_α at $\alpha \in \mathcal{A}$ and $\alpha' \in \mathcal{A}$.

Proposition 2.1. *Let $\alpha \in \mathcal{A}$ and $\alpha' \in \mathcal{A}$ be two parameters. Let $(\boldsymbol{\varphi}_*(\alpha), \boldsymbol{\psi}_*(\alpha))$ and $(\boldsymbol{\varphi}_*(\alpha'), \boldsymbol{\psi}_*(\alpha'))$ be solutions of (D_α) and $(D_{\alpha'})$ at α and α' respectively. Then:*

$$|J(\alpha) - J(\alpha')| \leq C_\varphi \|\boldsymbol{\mu}(\alpha) - \boldsymbol{\mu}(\alpha')\|_1 + C_\psi \|\boldsymbol{\nu}(\alpha) - \boldsymbol{\nu}(\alpha')\|_1,$$

where $C_\varphi = \max(\|\boldsymbol{\varphi}_*(\alpha)\|_\infty, \|\boldsymbol{\varphi}_*(\alpha')\|_\infty)$ and $C_\psi = \max(\|\boldsymbol{\psi}_*(\alpha)\|_\infty, \|\boldsymbol{\psi}_*(\alpha')\|_\infty)$.

Proof. The pair $(\varphi_*(\alpha), \psi_*(\alpha))$ is optimal for (D_α) , and also admissible for $(D_{\alpha'})$. Thus

$$J(\alpha) - J(\alpha') \leq \langle \varphi_*(\alpha), \mu(\alpha) - \mu(\alpha') \rangle + \langle \psi_*(\alpha), \nu(\alpha) - \nu(\alpha') \rangle. \quad (5)$$

By analogy, since $(\varphi_*(\alpha'), \psi_*(\alpha'))$ is optimal for $(D_{\alpha'})$, and admissible for (D_α) , we obtain :

$$J(\alpha) - J(\alpha') \geq \langle \varphi_*(\alpha'), \mu(\alpha) - \mu(\alpha') \rangle + \langle \psi_*(\alpha'), \nu(\alpha) - \nu(\alpha') \rangle. \quad (6)$$

Now, using norm inequalities, we bound from above the absolute value of the right hand side terms of (5) and (6).

$$|\langle \varphi_*(\alpha), \mu(\alpha) - \mu(\alpha') \rangle + \langle \psi_*(\alpha), \nu(\alpha) - \nu(\alpha') \rangle| \leq \quad (7)$$

$$\|\varphi_*(\alpha)\|_\infty \|\mu(\alpha) - \mu(\alpha')\|_1 + \|\psi_*(\alpha)\|_\infty \|\nu(\alpha) - \nu(\alpha')\|_1, \quad (8)$$

and

$$|\langle \varphi_*(\alpha'), \mu(\alpha) - \mu(\alpha') \rangle + \langle \psi_*(\alpha'), \nu(\alpha) - \nu(\alpha') \rangle| \leq \quad (9)$$

$$\|\varphi_*(\alpha')\|_\infty \|\mu(\alpha) - \mu(\alpha')\|_1 + \|\psi_*(\alpha')\|_\infty \|\nu(\alpha) - \nu(\alpha')\|_1. \quad (10)$$

Since the right hand sides of (7) and (9) are bounded from above by

$$\max(\|\varphi_*(\alpha)\|_\infty, \|\varphi_*(\alpha')\|_\infty) \|\mu(\alpha) - \mu(\alpha')\|_1 + \max(\|\psi_*(\alpha)\|_\infty, \|\psi_*(\alpha')\|_\infty) \|\nu(\alpha) - \nu(\alpha')\|_1,$$

we get the desired result. \square

Notice that the bound given in (2.1) depends on one hand on the ℓ_1 norm of the difference of parametrized marginals, and we prove in the following lemma that this difference is bounded by a term depending continuously on α . On the other hand, it depends on the ℓ_∞ norms of Kantorovitch potentials, and we explain in the sequel how to bound it using c -transforms.

Lemma 2.2. *Let $\alpha \in \mathcal{A}$ and $\alpha' \in \mathcal{A}$ be parameters. Then*

$$\|\mu(\alpha) - \mu(\alpha')\|_1 \leq K_x \|\alpha - \alpha'\|_\infty, \quad (11)$$

and

$$\|\nu(\alpha) - \nu(\alpha')\|_1 \leq K_y \|\alpha - \alpha'\|_\infty \quad (12)$$

Proof. By definition,

$$\mu(\alpha) - \mu(\alpha') = \sum_{k=1}^{K_x} (\alpha_k^x - \alpha_k'^x) \mu_k.$$

Then, since the μ_k are probability measures for all $1 \leq k \leq K_x$, we get

$$\|\mu(\alpha) - \mu(\alpha')\|_1 \leq \|\alpha^x - \alpha'^x\|_\infty \sum_{k=1}^{K_x} \|\mu_k\|_1 \leq K_x \|\alpha - \alpha'\|_\infty.$$

The bound on $\|\nu(\alpha) - \nu(\alpha')\|_1$ is obtained analogously. \square

In what follows, we use c -transforms (see e.g. Chapter 1 of [46]) to build potentials $(\varphi(\alpha), \psi(\alpha))$ that are solution of (D_α) , and whose norms are bounded by a constant that does not depend on α . Let us first recall the definition of a c -transform in the discrete setting:

Definition 2.3. *The discrete c -transform of a vector $\varphi \in \mathbb{R}^{\mathcal{N}_x}$ is the vector $\varphi^c \in \mathbb{R}^{\mathcal{N}_y}$ defined for every $1 \leq j \leq \mathcal{N}_y$ by :*

$$\varphi_j^c := \min_{1 \leq i \leq \mathcal{N}_x} C_{i,j} - \varphi_i. \quad (13)$$

The discrete \bar{c} -transform of a vector $\psi \in \mathbb{R}^{\mathcal{N}_y}$ is the vector $\psi^{\bar{c}} \in \mathbb{R}^{\mathcal{N}_x}$ defined for every $1 \leq i \leq \mathcal{N}_x$ by :

$$\psi_i^{\bar{c}} := \min_{1 \leq j \leq \mathcal{N}_y} C_{i,j} - \psi_j. \quad (14)$$

One interest of the c -transform is that, given an admissible pair $(\varphi(\boldsymbol{\alpha}), \psi(\boldsymbol{\alpha}))$ of $(D_{\boldsymbol{\alpha}})$, the pair $(\varphi(\boldsymbol{\alpha}), \varphi^c(\boldsymbol{\alpha}))$ is still admissible and increases the cost (the same holds for \bar{c} -transforms). Thus, given a solution $(\varphi_*(\boldsymbol{\alpha}), \psi_*(\boldsymbol{\alpha}))$, one can assume that it is virtually improved by c -transform and \bar{c} -transform. Moreover, c -transforms share the same modulus of continuity of the cost function, as we show in the following.

Lemma 2.4. *Let $\varphi \in \mathbb{R}^{\mathcal{N}_x}$. Then, denoting by $\varphi^c = (\varphi_j^c)_{1 \leq j \leq \mathcal{N}_y}$ the c -transform of φ , it holds that, for $1 \leq j, j' \leq \mathcal{N}_y$,*

$$|\varphi_j^c - \varphi_{j'}^c| \leq 2\|\mathbf{C}\|_{\infty}. \quad (15)$$

Proof. Let $1 \leq j, j' \leq \mathcal{N}_y$, then for every $1 \leq i \leq \mathcal{N}_x$,

$$\mathbf{C}_{ij} - \varphi_i^c \leq \mathbf{C}_{ij'} - \varphi_i + 2\|\mathbf{C}\|_{\infty}. \quad (16)$$

passing to the minimum in i in the left hand side of the inequality, and then in the right hand side, we obtain :

$$\varphi_j^c \leq \varphi_{j'}^c + 2\|\mathbf{C}\|_{\infty}, \quad (17)$$

and by interverting the roles of j and j' , we obtain the desired result. \square

Remark 2.5. *We also have that, for $\psi \in \mathbb{R}^{\mathcal{N}_y}$, for $1 \leq i, i' \leq \mathcal{N}_x$,*

$$|\psi_i^c - \psi_{i'}^c| \leq 2\|\mathbf{C}\|_{\infty}. \quad (18)$$

Corollary 2.6. *For any parameter $\boldsymbol{\alpha} \in \mathcal{A}$, there exists a solution $(\varphi_*(\boldsymbol{\alpha}), \psi_*(\boldsymbol{\alpha}))$ satisfying*

$$0 \leq \varphi_*(\boldsymbol{\alpha}) \leq 2\|\mathbf{C}\|_{\infty}, \quad (19)$$

and

$$-3\|\mathbf{C}\|_{\infty} \leq \psi_*(\boldsymbol{\alpha}) \leq \|\mathbf{C}\|_{\infty}. \quad (20)$$

In particular, $\|\varphi_*(\boldsymbol{\alpha})\|_{\infty} \leq 2\|\mathbf{C}\|_{\infty}$ and $\|\psi_*(\boldsymbol{\alpha})\|_{\infty} \leq 3\|\mathbf{C}\|_{\infty}$.

Proof. Let $(\varphi_*(\boldsymbol{\alpha}), \psi_*(\boldsymbol{\alpha}))$ be a solution of $(D_{\boldsymbol{\alpha}})$. Without loss of generality, we can improve this pair via c -transforms, so that they satisfy the continuity bound given in lemma 2.4 and remark 2.5. Moreover, since for every constant $\lambda \in \mathbb{R}$, $(\varphi_*(\boldsymbol{\alpha}) - \lambda, \psi_*(\boldsymbol{\alpha}) + \lambda)$ is still an optimal pair, we can suppose that $\min_{1 \leq i \leq \mathcal{N}_x} \varphi_*(\boldsymbol{\alpha})_i = 0$. We thus obtain that for $1 \leq i \leq \mathcal{N}_x$,

$$|\varphi_*(\boldsymbol{\alpha})_i - 0| \leq 2\|\mathbf{C}\|_{\infty}. \quad (21)$$

And since $\varphi_*(\boldsymbol{\alpha}) \geq 0$, we obtain the bound on $\varphi_*(\boldsymbol{\alpha})$:

$$0 \leq \varphi_*(\boldsymbol{\alpha}) \leq 2\|\mathbf{C}\|_{\infty}.$$

In order to bound $\psi_*(\boldsymbol{\alpha})$, we use the fact that it is defined as the \bar{c} -transform of $\varphi_*(\boldsymbol{\alpha})$, that is, for every $1 \leq j \leq \mathcal{N}_y$,

$$\psi_*(\boldsymbol{\alpha})_j = \min_{1 \leq i \leq \mathcal{N}_x} \mathbf{C}_{ij} - \varphi_*(\boldsymbol{\alpha})_i.$$

and we deduce from the bound on $\varphi_*(\boldsymbol{\alpha})$:

$$\min_{1 \leq i \leq \mathcal{N}_x} \mathbf{C}_{ij} - 2\|\mathbf{C}\|_{\infty} \leq \psi_*(\boldsymbol{\alpha})_j \leq \min_{1 \leq i \leq \mathcal{N}_x} \mathbf{C}_{ij}.$$

We deduce

$$-3\|\mathbf{C}\|_{\infty} \leq \psi_*(\boldsymbol{\alpha})_j \leq \|\mathbf{C}\|_{\infty}. \quad \square$$

Now, we aggregate the previous results to show that $J(\boldsymbol{\alpha})$ depends continuously on the parameter $\boldsymbol{\alpha}$.

Remark 2.7. In the proof of Corollary 2.6, the roles of $\varphi_*(\alpha)$ and $\psi_*(\alpha)$ can be inverted. Thus we can also conclude that for any parameter $\alpha \in \mathcal{A}$, there exists a solution $(\varphi_*(\alpha), \psi_*(\alpha))$ satisfying

$$0 \leq \psi_*(\alpha) \leq 2\|\mathbf{C}\|_\infty, \quad (22)$$

and

$$-3\|\mathbf{C}\|_\infty \leq \varphi_*(\alpha) \leq \|\mathbf{C}\|_\infty. \quad (23)$$

Proposition 2.8. Let $\alpha \in \mathcal{A}$ and $\alpha' \in \mathcal{A}$ be two parameters. Then

$$|J(\alpha) - J(\alpha')| \leq \|\mathbf{C}\|_\infty (2 \max(K_x, K_y) + 3 \min(K_x + K_y)) \|\alpha - \alpha'\|_\infty.$$

Proof. Let $(\varphi_*(\alpha), \psi_*(\alpha))$ and $(\varphi_*(\alpha'), \psi_*(\alpha'))$ be a pair of potential solutions to D_α at the parameter α . They satisfy the bound of property 2.1:

$$|J(\alpha) - J(\alpha')| \leq \max(\|\varphi_*(\alpha)\|_\infty, \|\varphi_*(\alpha')\|_\infty) \|\mu(\alpha) - \mu(\alpha')\|_1 + \max(\|\psi_*(\alpha)\|_\infty, \|\psi_*(\alpha')\|_\infty) \|\nu(\alpha) - \nu(\alpha')\|_1$$

We can suppose without loss of generality that both pairs of potentials satisfy the conditions of corollary 2.6, and thus

$$|J(\alpha) - J(\alpha')| \leq 2\|\mathbf{C}\|_\infty \|\mu(\alpha) - \mu(\alpha')\|_1 + 3\|\mathbf{C}\|_\infty \|\nu(\alpha) - \nu(\alpha')\|_1.$$

Using the continuous dependency of the differences marginals built in lemma 2.2, we obtain

$$|J(\alpha) - J(\alpha')| \leq \|\mathbf{C}\|_\infty (2K_x + 3K_y) \|\alpha - \alpha'\|_\infty.$$

Now, we can also find a pair $(\varphi_*(\alpha), \psi_*(\alpha))$ satisfying Remark 2.7. Again, using Lemma 2.2, we obtain

$$|J(\alpha) - J(\alpha')| \leq \|\mathbf{C}\|_\infty (2K_y + 3K_x) \|\alpha - \alpha'\|_\infty.$$

Taking the minimum of the right hand □

Now, our aim is to solve the problem (P_α) for many parameters $\alpha \in \mathcal{A}$. However, in high dimensions, solving (P_α) for only one parameter may already be computationally onerous. In the following section, we build a problem of smaller dimensions that approximates (P_α) , namely the reduced-order model, which enables us to approximate solutions of (P_α) at a computational cost that does not depend on the dimensions of the problem (P_α) .

3 Reduced-order model for the discrete optimal transport problem and its dual

In this section, we introduce the reduced-order linear programs that approximate the problems (P_α) and (D_α) . It consists in building a non-negative cone (respectively two vector spaces) of smaller dimension by using snapshots methods on the primal problem (P_α) (respectively the dual (D_α)), and to rewrite the problem where the minimization is performed on these cone or spaces.

We begin by explaining how these cone and spaces are defined. We then build the reduced-order model, which reads as a linear program with a small number of variables and constraints. We then provide sufficient conditions under which the reduced-order model has a least one solution. Finally, we propose an alternative method to build subspaces \mathcal{U} and \mathcal{V} of small dimension.

3.1 Construction of the reduced bases

For the approximation of the primal solutions is selected in a reduced non-negative cone $W_+ \subset \mathbb{R}_+^{\mathcal{N}_x \mathcal{N}_y}$ spanned by a family of vectors $\{\mathbf{w}_r\}_{1 \leq r \leq R} \subset \mathbb{R}_+^{\mathcal{N}_x \mathcal{N}_y}$ for $R \ll \mathcal{N}_x \mathcal{N}_y$. Concerning the dual solutions we approximate them by looking at the elements of reduced subspaces $\mathcal{U} \subset \mathbb{R}^{\mathcal{N}_x}$ and $\mathcal{V} \subset \mathbb{R}^{\mathcal{N}_y}$ spanned by two families of vectors $\{\mathbf{u}_n\}_{1 \leq n \leq N} \subset \mathbb{R}^{\mathcal{N}_x}$ and $\{\mathbf{v}_m\}_{1 \leq m \leq M} \subset \mathbb{R}^{\mathcal{N}_y}$ for $N \ll \mathcal{N}_x$ and $M \ll \mathcal{N}_y$.

The reduced variables will then be written as :

$$\begin{aligned}
\widehat{\boldsymbol{\pi}}(\boldsymbol{\alpha}) &= \sum_{r=1}^R \mathbf{p}_r(\boldsymbol{\alpha}) \mathbf{w}_r \text{ with } \mathbf{p}(\boldsymbol{\alpha}) \in \mathbb{R}_+^R \\
\widehat{\boldsymbol{\varphi}}(\boldsymbol{\alpha}) &= \sum_{n=1}^N \mathbf{a}_n(\boldsymbol{\alpha}) \mathbf{u}_n \text{ with } \mathbf{a}(\boldsymbol{\alpha}) \in \mathbb{R}^N \\
\widehat{\boldsymbol{\psi}}(\boldsymbol{\alpha}) &= \sum_{m=1}^M \mathbf{b}_m(\boldsymbol{\alpha}) \mathbf{v}_m \text{ with } \mathbf{b}(\boldsymbol{\alpha}) \in \mathbb{R}^M
\end{aligned} \tag{24}$$

In the following, we will denote by $\widehat{\mathbf{W}}$ the matrix whose columns are the vectors $\{\mathbf{w}_r\}_{1 \leq r \leq R}$, and by $\widehat{\mathbf{U}}$ (respectively $\widehat{\mathbf{V}}$) the matrix composed by $\{\mathbf{u}_n\}_{1 \leq n \leq N}$ (resp. by $\{\mathbf{v}_m\}_{1 \leq m \leq M}$).

Using these notations, the reduced solution may be rewritten as matrix-vector multiplication:

$$\begin{aligned}
\widehat{\boldsymbol{\pi}}(\boldsymbol{\alpha}) &= \widehat{\mathbf{W}} \mathbf{p}(\boldsymbol{\alpha}) \\
\widehat{\boldsymbol{\varphi}}(\boldsymbol{\alpha}) &= \widehat{\mathbf{U}} \mathbf{a}(\boldsymbol{\alpha}) \\
\widehat{\boldsymbol{\psi}}(\boldsymbol{\alpha}) &= \widehat{\mathbf{V}} \mathbf{b}(\boldsymbol{\alpha})
\end{aligned} \tag{25}$$

The Primal reduced space \mathcal{W}_+ We admit here that the set of parameters \mathcal{A} is a convex set generated by a finite number of extreme points. Let $\mathcal{A}_{\text{ext}} \subset \mathcal{A}$ be the set of extreme points of \mathcal{A} . Let $\mathcal{A}_{\text{train}}^P$ be a finite training set containing the set of extreme points, namely $\mathcal{A}_{\text{ext}} \subset \mathcal{A}_{\text{train}}^P \subset \mathcal{A}$. Then \mathcal{W}_+ is defined as the cone generated by the solutions to the problem $(P_{\boldsymbol{\alpha}})$ for all the parameters $\boldsymbol{\alpha} \in \mathcal{A}_{\text{train}}^P$, that is

$$\mathcal{W}_+ := \text{cone}_+ \{ \boldsymbol{\pi}_*(\boldsymbol{\alpha}) \}_{\boldsymbol{\alpha} \in \mathcal{A}_{\text{train}}^P}, \tag{26}$$

and $\widehat{\mathbf{W}}$ is the matrix formed by stacking the transport plans $\{ \boldsymbol{\pi}_*(\boldsymbol{\alpha}) \}_{\boldsymbol{\alpha} \in \mathcal{A}_{\text{train}}^P}$ columnwise. We explain in Section 3.3 that the inclusion $\mathcal{A}_{\text{ext}} \subset \mathcal{A}_{\text{train}}^P$ guarantees the non-emptiness of the feasible set.

Note that, as usual in basis reduction, it is possible to reduce the dimension of \mathcal{W}_+ with a non-negative matrix factorization. In this case, the solutions at the extreme points \mathcal{A}_{ext} must be added to $\widehat{\mathbf{W}}$.

The Dual reduced spaces \mathcal{U} and \mathcal{V} Let us assume that we are given a data set of dual solutions $\{(\boldsymbol{\varphi}_*(\boldsymbol{\alpha}), \boldsymbol{\psi}_*(\boldsymbol{\alpha}))\}_{\boldsymbol{\alpha} \in \mathcal{A}_{\text{train}}^D}$ for a finite training subset $\mathcal{A}_{\text{train}}^D \subset \mathcal{A}$. The dual reduced space \mathcal{U} (resp. \mathcal{V}) is the space generated by $\{(\boldsymbol{\varphi}_*(\boldsymbol{\alpha}))\}_{\boldsymbol{\alpha} \in \mathcal{A}_{\text{train}}^D}$ and $\mathbf{1}_{N_x}$ (resp. by $\{(\boldsymbol{\psi}_*(\boldsymbol{\alpha}))\}_{\boldsymbol{\alpha} \in \mathcal{A}_{\text{train}}^D}$ and $\mathbf{1}_{N_y}$). We explain in Section 3.3 why the $\mathbf{1}_{\bullet}$ vectors are added. We propose later in Section 3.4 an alternative way of building \mathcal{U} and \mathcal{V} , as the corresponding method requires further understanding of the reduced-order model.

3.2 The reduced-order model

We now explain how to use the reduced cones and bases in order to obtain a reduced-order model. In particular we proceed in two steps: (1) we build a semi-reduced-order model, where the ambient space of the high-fidelity primal $(P_{\boldsymbol{\alpha}})$ is reduced (2) we further reduce the dimension of the space variable of the dual of this semi-reduced-order model.

Notice that by inverting the two steps (we first reduce the dimension of the variables of the dual and then reduce the dimension of the variables of the associated primal) we obtain in the end the same fully reduced-order model, so only the semi-reduced-order models will differ. We explain how both of them are obtained in what follows.

Reduction on Primal-then-Dual We propose here a semi-reduced-order model where the variables of the primal are searched in a non-negative cone of smaller dimension. It is obtained as rewriting $(P_{\boldsymbol{\alpha}})$ with the added constraint that the solutions should belong to the non-negative cone \mathcal{W}_+ :

$$\begin{aligned}
I_{R_P}(\boldsymbol{\alpha}) &:= \inf_{\widehat{\boldsymbol{\pi}} \in \mathcal{W}_+} \langle \widehat{\boldsymbol{\pi}}, \mathbf{c} \rangle. \\
\mathbf{A} \widehat{\boldsymbol{\pi}} &= \begin{pmatrix} \boldsymbol{\mu}(\boldsymbol{\alpha}) \\ \boldsymbol{\nu}(\boldsymbol{\alpha}) \end{pmatrix}
\end{aligned} \tag{27}$$

If we use the notations introduced previously (25), this problem can be reformulated in the following form :

$$I_{RP}(\alpha) := \inf_{\mathbf{p} \in \mathbb{R}_+^R} \langle \mathbf{p}, \widehat{\mathbf{W}}^T \mathbf{c} \rangle. \quad (RPP_\alpha)$$

$$\mathbf{A} \widehat{\mathbf{W}} \mathbf{p} = \begin{pmatrix} \boldsymbol{\mu}(\alpha) \\ \boldsymbol{\nu}(\alpha) \end{pmatrix}$$

The dual of this semi-reduced-order problem writes

$$J_{RP}(\alpha) = \sup_{(\boldsymbol{\varphi}, \boldsymbol{\psi}) \in \mathbb{R}^{\mathcal{N}_x + \mathcal{N}_y}} \langle \boldsymbol{\varphi}, \boldsymbol{\mu}(\alpha) \rangle + \langle \boldsymbol{\psi}, \boldsymbol{\nu}(\alpha) \rangle. \quad (RPD_\alpha)$$

$$\widehat{\mathbf{W}}^T \mathbf{A}^T \begin{pmatrix} \boldsymbol{\varphi} \\ \boldsymbol{\psi} \end{pmatrix} \leq \widehat{\mathbf{W}}^T \mathbf{c}$$

We now add to this semi-reduced-order dual problem the constraint that the Kantorovitch potentials should belong to the subspaces \mathcal{U} and \mathcal{V} . We obtain the following reduced dual:

$$J_R(\alpha) := \sup_{(\widehat{\boldsymbol{\varphi}}, \widehat{\boldsymbol{\psi}}) \in \mathcal{U} \times \mathcal{V}} \langle \widehat{\boldsymbol{\varphi}}, \boldsymbol{\mu}(\alpha) \rangle + \langle \widehat{\boldsymbol{\psi}}, \boldsymbol{\nu}(\alpha) \rangle$$

$$\widehat{\mathbf{W}}^T \mathbf{A}^T \begin{pmatrix} \widehat{\boldsymbol{\varphi}} \\ \widehat{\boldsymbol{\psi}} \end{pmatrix} \leq \widehat{\mathbf{W}}^T \mathbf{c}$$

$$= \sup_{(\mathbf{a}, \mathbf{b}) \in \mathbb{R}^{N+M}} \langle \mathbf{a}, \widehat{\mathbf{U}}^T \boldsymbol{\mu}(\alpha) \rangle + \langle \mathbf{b}, \widehat{\mathbf{V}}^T \boldsymbol{\nu}(\alpha) \rangle,$$

$$\widehat{\mathbf{W}}^T \mathbf{A}^T \widehat{\mathbf{S}} \begin{pmatrix} \mathbf{a} \\ \mathbf{b} \end{pmatrix} \leq \widehat{\mathbf{W}}^T \mathbf{c}$$
(28)

where

$$\widehat{\mathbf{S}} := \begin{pmatrix} \widehat{\mathbf{U}} & \mathbf{0} \\ \mathbf{0} & \widehat{\mathbf{V}} \end{pmatrix}. \quad (29)$$

The primal reduced problem then writes

$$I_R(\alpha) := \inf_{\widehat{\boldsymbol{\pi}} \in \mathcal{W}_+} \langle \widehat{\boldsymbol{\pi}}, \mathbf{c} \rangle,$$

$$\widehat{\mathbf{S}}^T \mathbf{A} \widehat{\boldsymbol{\pi}} = \widehat{\mathbf{S}}^T \begin{pmatrix} \boldsymbol{\mu}(\alpha) \\ \boldsymbol{\nu}(\alpha) \end{pmatrix}$$

$$= \inf_{\mathbf{p} \in \mathbb{R}_+^R} \langle \mathbf{p}, \widehat{\mathbf{W}}^T \mathbf{c} \rangle.$$

$$\widehat{\mathbf{S}}^T \mathbf{A} \widehat{\mathbf{W}} \mathbf{p} = \widehat{\mathbf{S}}^T \begin{pmatrix} \boldsymbol{\mu}(\alpha) \\ \boldsymbol{\nu}(\alpha) \end{pmatrix}$$
(30)

We can re-write this problem in the standard form of linear programming : let $\widehat{\mathbf{c}} := \widehat{\mathbf{W}}^T \mathbf{c}$, $\widehat{\boldsymbol{\mu}}(\alpha) := \widehat{\mathbf{U}}^T \boldsymbol{\mu}(\alpha)$ and $\widehat{\boldsymbol{\nu}}(\alpha) := \widehat{\mathbf{V}}^T \boldsymbol{\nu}(\alpha)$, and let $\widehat{\mathbf{A}} := \widehat{\mathbf{S}}^T \mathbf{A} \widehat{\mathbf{W}}$.

The reduced problem then writes

$$I_R(\alpha) := \inf_{\mathbf{p} \in \mathbb{R}_+^R} \langle \mathbf{p}, \widehat{\mathbf{c}} \rangle, \quad (RP_\alpha)$$

$$\widehat{\mathbf{A}} \mathbf{p} = \begin{pmatrix} \widehat{\boldsymbol{\mu}}(\alpha) \\ \widehat{\boldsymbol{\nu}}(\alpha) \end{pmatrix}$$

and its dual

$$J_R(\alpha) := \inf_{\mathbf{a} \in \mathbb{R}^N, \mathbf{b} \in \mathbb{R}^M} \langle \mathbf{a}, \widehat{\boldsymbol{\mu}}(\alpha) \rangle + \langle \mathbf{b}, \widehat{\boldsymbol{\nu}}(\alpha) \rangle, \quad (RD_\alpha)$$

$$\widehat{\mathbf{A}}^T \begin{pmatrix} \mathbf{a} \\ \mathbf{b} \end{pmatrix} \leq \widehat{\mathbf{c}}$$

For convenience in the proofs of the properties that follow, we will sometimes refer to the form (30) of the reduced-order model.

The dimensions of the high-fidelity model (P_α) and the reduced-order model are compared in table 1.

Problem	Degrees of Freedom	Constraints
High-Fidelity	$\mathcal{N}_x \mathcal{N}_y$	$\mathcal{N}_x + \mathcal{N}_y$
Redced-order Model	R	$N + M$

Table 1: Dimensions comparison.

Reduction on dual-then-primal Here we reduce the problem by first reducing the dimensions of the variables of the dual high-fidelity problem (D_α), and then obtaining the reduced-order model by reducing again on the dimensions of the space variables of the primal problem. So we begin by adding to the high-fidelity dual problem the constraint that the Kantorovitch potential should belong to the subspaces \mathcal{U} and \mathcal{V} :

$$J_{RD}(\alpha) = \sup_{\substack{(\hat{\varphi}, \hat{\psi}) \in \mathcal{U} \times \mathcal{V} \\ \mathbf{A}^T \begin{pmatrix} \hat{\varphi} \\ \hat{\psi} \end{pmatrix} \leq \mathbf{c}}} \langle \hat{\varphi}, \boldsymbol{\mu}(\alpha) \rangle + \langle \hat{\psi}, \boldsymbol{\nu}(\alpha) \rangle. \quad (31)$$

Using the notations introduced in 25 and in 29, we rewrite this problem with reduced dimensions

$$J_{RD}(\alpha) = \sup_{\substack{(\mathbf{a}, \mathbf{b}) \in \mathbb{R}^{N+M} \\ \mathbf{A}^T \hat{\mathbf{S}} \begin{pmatrix} \mathbf{a} \\ \mathbf{b} \end{pmatrix} \leq \mathbf{c}}} \langle \mathbf{a}, \mathbf{U}^T \boldsymbol{\mu}(\alpha) \rangle + \langle \mathbf{b}, \mathbf{V}^T \boldsymbol{\nu}(\alpha) \rangle. \quad (RD D_\alpha)$$

The primal version of this linear program writes

$$I_{RD}(\alpha) := \inf_{\substack{\boldsymbol{\pi} \in \mathbb{R}_+^{\mathcal{N}_x \mathcal{N}_y} \\ \hat{\mathbf{S}}^T \mathbf{A} \boldsymbol{\pi} = \hat{\mathbf{S}}^T \begin{pmatrix} \boldsymbol{\mu}(\alpha) \\ \boldsymbol{\nu}(\alpha) \end{pmatrix}}} \langle \boldsymbol{\pi}, \mathbf{c} \rangle, \quad (RD P_\alpha)$$

We now reduce the dimension of the cone of variables :

$$I_R(\alpha) := \inf_{\substack{\hat{\boldsymbol{\pi}} \in \mathcal{W}_+ \\ \hat{\mathbf{S}}^T \mathbf{A} \hat{\boldsymbol{\pi}} = \hat{\mathbf{S}}^T \begin{pmatrix} \boldsymbol{\mu}(\alpha) \\ \boldsymbol{\nu}(\alpha) \end{pmatrix}}} \langle \hat{\boldsymbol{\pi}}, \mathbf{c} \rangle, \quad (32)$$

and we recognize the reduced-order model (30).

3.3 Existence of solution for the reduced-order model

The high-fidelity problem (P_α) reads as a minimization problem of a continuous function in an non empty compact set, as the vector $\boldsymbol{\mu}(\alpha) \otimes \boldsymbol{\nu}(\alpha)$ is admissible and the feasible set is a closed subset of the probability vectors of dimension $\mathcal{N}_x \mathcal{N}_y$. Therefore, it has solutions. In the reduced-order model, the function to minimize is continuous, but the conditions on the feasible set are a priori not satisfied anymore. We propose in what follows a condition on the cone \mathcal{W}_+ to ensure that the feasible set is not empty, and a condition on the matrices $\hat{\mathbf{U}}$ and $\hat{\mathbf{V}}$ to guarantee compactness.

Non emptiness of the feasible set We first exhibit that \mathcal{A} is a convex polytope, thus can be generated by finitely many extreme points, and so does the family of marginals $\left\{ \begin{pmatrix} \boldsymbol{\mu}(\boldsymbol{\alpha}) \\ \boldsymbol{\nu}(\boldsymbol{\alpha}) \end{pmatrix} \right\}_{\boldsymbol{\alpha} \in \mathcal{A}}$. We then deduce that the reduced-order model ($RP_{\boldsymbol{\alpha}}$) is feasible for any $\boldsymbol{\alpha} \in \mathcal{A}$ if the reduced cone \mathcal{W}_+ contains a transport plan solution of the discrete high-fidelity problem ($P_{\boldsymbol{\alpha}}$) at each extreme point of \mathcal{A} .

Proposition 3.1. *The set \mathcal{A} is a finite intersection of half-spaces, and it is bounded. In other words, it is a convex polytope.*

Corollary 3.2. *\mathcal{A} is generated by the set of its extreme points, denoted \mathcal{A}_{ext} , and this set is finite.*

This corollary is a classical property of the analysis of convex polytopes. For a proof we refer for instance the reader to Chapter 1 of [28]. We now show that the set $\left\{ \begin{pmatrix} \boldsymbol{\mu}(\boldsymbol{\alpha}) \\ \boldsymbol{\nu}(\boldsymbol{\alpha}) \end{pmatrix} \right\}_{\boldsymbol{\alpha} \in \mathcal{A}}$ inherits the same structure as \mathcal{A} .

Lemma 3.3. *Let $\boldsymbol{\alpha} \in \mathcal{A}$ and its convex decomposition with extreme points be $\boldsymbol{\alpha} = \sum_{\boldsymbol{\alpha}_{\text{ext}} \in \mathcal{A}_{\text{ext}}} \lambda_{\text{ext}} \boldsymbol{\alpha}_{\text{ext}}$.*

Then

$$\begin{pmatrix} \boldsymbol{\mu}(\boldsymbol{\alpha}) \\ \boldsymbol{\nu}(\boldsymbol{\alpha}) \end{pmatrix} = \sum_{\boldsymbol{\alpha}_{\text{ext}} \in \mathcal{A}_{\text{ext}}} \lambda_{\text{ext}} \begin{pmatrix} \boldsymbol{\mu}(\boldsymbol{\alpha}_{\text{ext}}) \\ \boldsymbol{\nu}(\boldsymbol{\alpha}_{\text{ext}}) \end{pmatrix} \quad (33)$$

Proof. Let us remark that the set of parametrized marginals $\left\{ \begin{pmatrix} \boldsymbol{\mu}(\boldsymbol{\alpha}) \\ \boldsymbol{\nu}(\boldsymbol{\alpha}) \end{pmatrix} \right\}_{\boldsymbol{\alpha} \in \mathcal{A}}$ can be seen as the image of \mathcal{A} by the linear application $\mathbf{M} := \begin{pmatrix} \boldsymbol{\mu}_1 & \dots & \boldsymbol{\mu}_{K_x} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \dots & \mathbf{0} & \boldsymbol{\nu}_1 & \dots & \boldsymbol{\nu}_{K_y} \end{pmatrix}$. Now, let $\boldsymbol{\alpha} \in \mathcal{A}$. With the previous notations,

$$\begin{pmatrix} \boldsymbol{\mu}(\boldsymbol{\alpha}) \\ \boldsymbol{\nu}(\boldsymbol{\alpha}) \end{pmatrix} = \mathbf{M}\boldsymbol{\alpha}. \quad (34)$$

But since $\boldsymbol{\alpha} = \sum_{\boldsymbol{\alpha}_{\text{ext}} \in \mathcal{A}_{\text{ext}}} \lambda_{\text{ext}} \boldsymbol{\alpha}_{\text{ext}}$, we also have

$$\begin{pmatrix} \boldsymbol{\mu}(\boldsymbol{\alpha}) \\ \boldsymbol{\nu}(\boldsymbol{\alpha}) \end{pmatrix} = \sum_{\boldsymbol{\alpha}_{\text{ext}} \in \mathcal{A}_{\text{ext}}} \lambda_{\text{ext}} \mathbf{M}\boldsymbol{\alpha}_{\text{ext}} = \sum_{\boldsymbol{\alpha}_{\text{ext}} \in \mathcal{A}_{\text{ext}}} \lambda_{\text{ext}} \begin{pmatrix} \boldsymbol{\mu}(\boldsymbol{\alpha}_{\text{ext}}) \\ \boldsymbol{\nu}(\boldsymbol{\alpha}_{\text{ext}}) \end{pmatrix}. \quad (35)$$

□

We are now able to give a feasibility condition.

Proposition 3.4. *Suppose that for any extreme point of \mathcal{A} $\boldsymbol{\alpha}_{\text{ext}}$, the solution $\boldsymbol{\pi}(\boldsymbol{\alpha}_{\text{ext}})$ of the high-fidelity discrete problem $P_{\boldsymbol{\alpha}}$ belongs to \mathcal{W}_+ . Then, for any $\boldsymbol{\alpha} \in \mathcal{A}$, the problem $RP_{\boldsymbol{\alpha}}$ is feasible.*

Proof. Let $\boldsymbol{\alpha} \in \mathcal{A}$ be a given parameter, and let $\boldsymbol{\alpha} = \sum_{\boldsymbol{\alpha}_{\text{ext}} \in \mathcal{A}_{\text{ext}}} \lambda_{\text{ext}} \boldsymbol{\alpha}_{\text{ext}}$ be its convex decomposition over the extreme points of \mathcal{A} . Let $\boldsymbol{\pi} = \sum_{\boldsymbol{\alpha}_{\text{ext}} \in \mathcal{A}_{\text{ext}}} \lambda_{\text{ext}} \boldsymbol{\pi}(\boldsymbol{\alpha}_{\text{ext}})$. Our goal is to prove that $\boldsymbol{\pi}$ is feasible, that

is $\boldsymbol{\pi} \in \mathcal{W}_+$, and. $\widehat{\mathbf{S}}^T \mathbf{A}\boldsymbol{\pi} = \widehat{\mathbf{S}}^T \begin{pmatrix} \boldsymbol{\mu}(\boldsymbol{\alpha}) \\ \boldsymbol{\nu}(\boldsymbol{\alpha}) \end{pmatrix}$ (we refer to the feasibility conditions based on the formulation (30) of the reduced order model). Let us first notice that by hypothesis, $\boldsymbol{\pi}$ is a convex (thus positive) combination of points of \mathcal{W}_+ , and thus $\boldsymbol{\pi} \in \mathcal{W}_+$. It remains to prove that $\boldsymbol{\pi}$ satisfies the constraints. One one hand, we can rewrite the left hand side of the constraint equality :

$$\mathbf{A}\boldsymbol{\pi} = \sum_{\boldsymbol{\alpha}_{\text{ext}} \in \mathcal{A}_{\text{ext}}} \lambda_{\text{ext}} \mathbf{A}\boldsymbol{\pi}(\boldsymbol{\alpha}_{\text{ext}}) = \sum_{\boldsymbol{\alpha}_{\text{ext}} \in \mathcal{A}_{\text{ext}}} \lambda_{\text{ext}} \begin{pmatrix} \boldsymbol{\mu}(\boldsymbol{\alpha}_{\text{ext}}) \\ \boldsymbol{\nu}(\boldsymbol{\alpha}_{\text{ext}}) \end{pmatrix}, \quad (36)$$

and on the other hand, by lemma 3.7, we rewrite the right-hand side of the constraint equality

$$\begin{pmatrix} \boldsymbol{\mu}(\boldsymbol{\alpha}) \\ \boldsymbol{\nu}(\boldsymbol{\alpha}) \end{pmatrix} = \sum_{\boldsymbol{\alpha}_{\text{ext}} \in \mathcal{A}_{\text{ext}}} \lambda_{\text{ext}} \begin{pmatrix} \boldsymbol{\mu}(\boldsymbol{\alpha}_{\text{ext}}) \\ \boldsymbol{\nu}(\boldsymbol{\alpha}_{\text{ext}}) \end{pmatrix}, \quad (37)$$

thus $\boldsymbol{\pi}$ is admissible and the feasible set of (30) is not empty. □

Remark 3.5. Notice that if the conditions of Proposition 3.4 holds, then the semi-reduced problem $(R_P P_\alpha)$ is feasible too.

Remark 3.6. Since the matrix $\widehat{\mathbf{W}}$ is obtained by stacking columnwise the transport plans $\{\pi_*(\alpha)\}_{\alpha \in \mathcal{A}_{train}^P}$, $\mathbf{A}\widehat{\mathbf{W}}$ is obtained by stacking columnwise the corresponding constraints $\left\{ \begin{pmatrix} \mu(\alpha) \\ \nu(\alpha) \end{pmatrix} \right\}_{\alpha \in \mathcal{A}_{train}^P}$. Saying that the feasible set of the semi-reduced problem is not empty at $\alpha \in \mathcal{A}$ is equivalent to the fact that there exists a positive combination of columns of $\mathbf{A}\widehat{\mathbf{W}}$ that is equal to $\begin{pmatrix} \mu(\alpha) \\ \nu(\alpha) \end{pmatrix}$.

We finish by giving an explicit expression of the extreme points of \mathcal{A} , since it can be convenient for numerical applications.

Lemma 3.7. The set $\mathcal{A} := \left\{ \alpha := (\alpha^x, \alpha^y) \in \mathbb{R}_+^{K_x+K_y}, \sum_{k=1}^{K_x} \alpha_k^x = 1, \sum_{l=1}^{K_y} \alpha_l^y = 1 \right\}$ has $K_x \times K_y$ extreme points, defined by $\mathcal{A}_{ext} = \{(\mathbf{e}_k, \mathbf{f}_l)\}_{\substack{1 \leq k \leq K_x \\ 1 \leq l \leq K_y}}$, where $\{\mathbf{e}_k\}_{1 \leq k \leq K_x}$ and $\{\mathbf{f}_l\}_{1 \leq l \leq K_y}$ are the elementary basis vectors of \mathbb{R}^{K_x} and \mathbb{R}^{K_y} .

Proof. We begin by showing the generative property of \mathcal{A}_{ext} . We then prove that the elements of \mathcal{A}_{ext} are extreme points by noticing each point α of \mathcal{A}_{ext} , can be expressed as one unique convex combination of points of \mathcal{A}_{ext} .

Let $(\alpha^x, \alpha^y) = \left(\{\alpha_k^x\}_{1 \leq k \leq K_x}, \{\alpha_l^y\}_{1 \leq l \leq K_y} \right)$ be an element of \mathcal{A} . Let us show that α is a convex combination of $\{(\mathbf{e}_k, \mathbf{f}_l)\}_{\substack{1 \leq k \leq K_x \\ 1 \leq l \leq K_y}}$ associated with the coefficients $\beta = \{\alpha_k^x \alpha_l^y\}_{\substack{1 \leq k \leq K_x \\ 1 \leq l \leq K_y}}$:

$$\sum_{k'=1}^{K_x} \sum_{l'=1}^{K_y} \alpha_{k'}^x \alpha_{l'}^y (\mathbf{e}_{k'}, \mathbf{f}_{l'}) = (\alpha^x, \alpha^y). \quad (38)$$

We first check that the coefficients of β form a convex combination. This arises from the fact that the coefficients of α^x and α^y are themselves convex combinations, so for any $1 \leq k' \leq K_x$ and $1 \leq l' \leq K_y$, $\beta_{k',l'} \geq 0$, and

$$\sum_{k'=1}^{K_x} \sum_{l'=1}^{K_y} \beta_{k',l'} = \sum_{k'=1}^{K_x} \sum_{l'=1}^{K_y} \alpha_{k'}^x \alpha_{l'}^y = 1. \quad (39)$$

Now, notice that we can decompose α^x and α^y as sums of elementary vectors: $\alpha^x = \sum_{k=1}^{K_x} \alpha_k^x \mathbf{e}_k$,

and $\alpha^y = \sum_{l=1}^{K_y} \alpha_l^y \mathbf{f}_l$.

Thus, for $1 \leq k \leq K_x$,

$$\begin{aligned} \left(\sum_{k'=1}^{K_x} \sum_{l'=1}^{K_y} \alpha_{k'}^x \alpha_{l'}^y \mathbf{e}_{k'} \right)_k &= \left(\sum_{k'=1}^{K_x} \alpha_{k'}^x \mathbf{e}_{k'} \right)_k \\ &= (\alpha_k^x \mathbf{e}_k)_k \\ &= \alpha_k^x, \end{aligned} \quad (40)$$

and for $1 \leq l \leq K_y$,

$$\begin{aligned} \left(\sum_{k'=1}^{K_x} \sum_{l'=1}^{K_y} \alpha_{k'}^x \alpha_{l'}^y \mathbf{f}_{l'} \right)_l &= \left(\sum_{l'=1}^{K_y} \alpha_{l'}^y \mathbf{f}_{l'} \right)_l \\ &= (\alpha_l^y \mathbf{f}_l)_l \\ &= \alpha_l^y \end{aligned} \quad (41)$$

This concludes the fact that $\{(\mathbf{e}_k, \mathbf{f}_l)\}_{\substack{1 \leq k \leq K_x \\ 1 \leq l \leq K_y}}$ is a generative set of \mathcal{A} .

Now, suppose that, for $1 \leq k \leq K_x$ and $1 \leq l \leq K_y$, the point $(\mathbf{e}_k, \mathbf{f}_l)$ is expressed as a convex combination of $\{(\mathbf{e}_k, \mathbf{f}_l)\}_{\substack{1 \leq k \leq K_x \\ 1 \leq l \leq K_y}}$ associated to the coefficients $\{\beta_{k',l'}\}$. Then, in particular :

$$\mathbf{e}_k = \sum_{k'=1}^{K_x} \left(\sum_{l'=1}^{K_y} \beta_{k',l'} \right) \mathbf{e}_{k'}, \quad (42)$$

By linear independence of the elementary vectors $\mathbf{e}_{k'}$, if $k' \neq k$,

$$\sum_{l'=1}^{K_y} \beta_{k',l'} = 0, \quad (43)$$

and by non-negativity of the coefficients of β , we obtain that $\beta_{k',l'} = 0$ if $k' \neq k$.

By a similar computation, since the $\{\mathbf{f}_{l'}\}_{1 \leq l' \leq K_y}$ are independent, by non-negativity of the coefficients of β , for every $1 \leq k' \leq K_x$, $1 \leq l' \leq K_y$, $\beta_{k',l'} = 0$ if $l' \neq l$.

Since the coefficients of β sum to one, this means that the only coefficient that can be non zero, namely $\beta_{k,l}$, is equal to 1. So we have proven that every element of $\{(\mathbf{e}_k, \mathbf{f}_l)\}_{\substack{1 \leq k \leq K_x \\ 1 \leq l \leq K_y}}$ can be written

as one unique convex combination of elements of this set. \square

Compactness of the feasible set We draw here a simple condition to guaranty the compactness of the reduced-order model (RP_α) .

Lemma 3.8. *If the vector $\mathbb{1}_{N_x}$ is one column of $\widehat{\mathbf{U}}$ or if $\mathbb{1}_{N_y}$ is one column of $\widehat{\mathbf{V}}$, the set of feasible solutions of (RP_α) is compact.*

Proof. The feasible set of the reduced problem written on the form (30) is closed as it is the intersection between a closed cone, \mathcal{W}_+ , and the reciprocal image of a point by a linear application, $\text{Im}^{-1}(\widehat{\mathbf{S}}^T \mathbf{A}) \left\{ \begin{pmatrix} \widehat{\boldsymbol{\mu}}(\boldsymbol{\alpha}) \\ \widehat{\boldsymbol{\nu}}(\boldsymbol{\alpha}) \end{pmatrix} \right\}$.

If $\mathbb{1}_{N_x}$ is a column of $\widehat{\mathbf{U}}$, for instance, the last one, then

$$\widehat{\mathbf{S}}_N^T = (\widehat{\mathbf{U}}^T | 0)_N = (\mathbb{1}_{N_x} | 0). \quad (44)$$

Recall now that, by definition of \mathbf{A} in (4), for any $\boldsymbol{\pi} \in \mathbb{R}_+^{N_x N_y}$,

$$(\mathbf{A}\boldsymbol{\pi})_i = \sum_{j=1}^{N_y} \pi_{i+jN_x}. \quad (45)$$

Thus, the condition $(\widehat{\mathbf{S}}^T \mathbf{A})_N \boldsymbol{\pi} = (\widehat{\mathbf{S}}^T \mathbf{A})_N \begin{pmatrix} \widehat{\boldsymbol{\mu}}(\boldsymbol{\alpha}) \\ \widehat{\boldsymbol{\nu}}(\boldsymbol{\alpha}) \end{pmatrix}$ reads explicitly as the following equality:

$$\sum_{i=1}^{N_x} (\mathbf{A}\boldsymbol{\pi})_i = \sum_{k=1}^{N_x \times N_y} \pi_k = \sum_{i=1}^{N_x} \boldsymbol{\mu}(\boldsymbol{\alpha})_i = 1. \quad (46)$$

Since $\boldsymbol{\pi} \in \mathbb{R}_+^{N_x N_y}$, the fact that its coefficients sum to 1 implies that the feasible set is bounded.

If $\mathbb{1}_{N_y}$ is one column of $\widehat{\mathbf{V}}$, the same proof holds with very similar computations. \square

Remark 3.9. *Adding the $\mathbf{1}_\bullet$ vector is in fact a way to say we are looking for solutions that are probability vectors.*

Existence of a solution We end this Section by giving an existence result for the reduced-order model.

Proposition 3.10. *Under the conditions of Properties 3.4 and 3.8, the reduced-order primal has a solution.*

Proof. The reduced-order primal is a minimization problem of a continuous function. The conditions of 3.4 and 3.8 guaranty that the feasible set is not empty and compact. We deduce that there exists a solution to this problem. \square

Remark 3.11. *By strong duality theory (see e.g. Chapter 3 of [28]), it is sufficient to prove that the reduced-order problem (RP_α) has solution to guaranty that its dual version (RD_α) has solutions too.*

3.4 An alternative way to choose \mathcal{U} and \mathcal{V}

We propose here a way to choose \mathcal{U} and \mathcal{V} of dimension K_x and K_y , that guarantees that the optimal costs of the reduced-order model (RP_α) and of the semi-reduced order model (RP_α) are the same. We begin by giving a sufficient condition to obtain the equality of the optimal value. We then present a way to build $\widehat{\mathbf{U}}$ and $\widehat{\mathbf{V}}$ satisfying this condition with Gram-Schmidt algorithm.

3.4.1 Rank condition to make the reduced-order model equivalent to the semi-reduced-order model

Let us begin by giving a short analysis of the structure of the constraint matrix of the half reduced problem (RP_α) that motivates the search of $\widehat{\mathbf{U}}$ and $\widehat{\mathbf{V}}$ of small dimension. As defined in 3.1, $\widehat{\mathbf{W}}$ is obtained by stacking columnwise solutions of high-fidelity problems for different parameters. In other terms, if the corresponding training set is defined as $\mathcal{A}_{train}^P = \{\alpha_r\}_{1 \leq r \leq R}$, then:

$$\widehat{\mathbf{W}} := (\pi_*(\alpha_1) | \cdots | \pi_*(\alpha_R)). \quad (47)$$

One can now easily see that the constraint matrix $\mathbf{A}\widehat{\mathbf{W}}$ can be written as

$$\mathbf{A}\widehat{\mathbf{W}} = \begin{pmatrix} \boldsymbol{\mu}(\alpha_1) | \cdots | \boldsymbol{\mu}(\alpha_R) \\ \boldsymbol{\nu}(\alpha_1) | \cdots | \boldsymbol{\nu}(\alpha_R) \end{pmatrix} =: \begin{pmatrix} (\mathbf{A}\widehat{\mathbf{W}})_x \\ (\mathbf{A}\widehat{\mathbf{W}})_y \end{pmatrix}, \quad (48)$$

where $(\mathbf{A}\widehat{\mathbf{W}})_x$ (respectively $(\mathbf{A}\widehat{\mathbf{W}})_y$) denotes the upper part (resp. lower part) of $\mathbf{A}\widehat{\mathbf{W}}$. We can now derive an upper bound of the rank of $(\mathbf{A}\widehat{\mathbf{W}})_x$ and $(\mathbf{A}\widehat{\mathbf{W}})_y$:

Lemma 3.12. $\text{rank}((\mathbf{A}\widehat{\mathbf{W}})_x) \leq K_x$ and $\text{rank}((\mathbf{A}\widehat{\mathbf{W}})_y) \leq K_y$.

Proof. In this proof, we only focus on $(\mathbf{A}\widehat{\mathbf{W}})_x$, but the same reasoning can be applied on $(\mathbf{A}\widehat{\mathbf{W}})_y$. For every parameter α , the marginal $\boldsymbol{\mu}(\alpha)$ is a convex (thus linear) combination of the K_x marginals $\{\boldsymbol{\mu}_k\}_{1 \leq k \leq K_x}$. In particular, since every column of $(\mathbf{A}\widehat{\mathbf{W}})_x$ is a linear combination of this family of marginals, whose rank is at most K_x . \square

This rank results motivates us to find some $\widehat{\mathbf{U}}$ and $\widehat{\mathbf{V}}$ that would have at most K_x and K_y columns for the reduced problem, and that would however satisfy that the optimal value fo the half reduced problem (RP_α) and the reduced problem (RP_α) would be the same. The following property gives a sufficient condition on $\widehat{\mathbf{U}}$ and $\widehat{\mathbf{V}}$ to guarantee that these optimal values are equal.

Proposition 3.13. *Let $\alpha \in \mathcal{A}$ be a parameter. suppose that the semi-reduced order model (RP_α) is feasible, and that following rank conditions are satisfied :*

$$\text{rank}((\mathbf{A}\widehat{\mathbf{W}})_x) = \text{rank}(\widehat{\mathbf{U}}^T (\mathbf{A}\widehat{\mathbf{W}})_x), \quad (49)$$

and

$$\text{rank}((\mathbf{A}\widehat{\mathbf{W}})_y) = \text{rank}(\widehat{\mathbf{V}}^T (\mathbf{A}\widehat{\mathbf{W}})_y). \quad (50)$$

Then the half reduced- problem (RP_α) and the reduced problem (RP_α) have the same optimal value.

Before proving this Proposition, we need the following simple linear algebra Lemma.

Lemma 3.14. *Let $\mathbf{B} \in \mathbb{R}^{D_1 \times D_2}$ and $\mathbf{E} \in \mathbb{R}^{D_2 \times D_3}$ be matrices such that $\text{Rank}(\mathbf{B}) = \text{Rank}(\mathbf{BE})$. Then, there exists a matrix $\mathbf{F} \in \mathbb{R}^{D_3 \times D_2}$ such that :*

$$\mathbf{B} = \mathbf{BEF}. \quad (51)$$

Proof. In general, the inclusion $\text{Im}(\mathbf{BE}) \subset \text{Im}(\mathbf{B})$ holds. But since by hypothesis, the rank of the two matrices of consideration are equal. we have equality of the image spaces:

$$\text{Im}(\mathbf{BE}) = \text{Im}(\mathbf{B}). \quad (52)$$

Thus, for every $1 \leq d \leq D_2$, by hypothesis, $\mathbf{Be}_d \in \text{Im}(\mathbf{B}) = \text{Im}(\mathbf{BE})$ (here $\{\mathbf{e}_d\}_{1 \leq d \leq D_2}$ are the canonical basis' vector). Therefore, there exists a vector $\mathbf{f}_d \in \mathbb{R}^{D_3}$ such that $\mathbf{Be}_d = \mathbf{BEf}_d$. Stacking every $\{\mathbf{e}_d\}_{1 \leq d \leq D_2}$ columnwise, we obtain

$$\mathbf{B} = \mathbf{BId}_{D_2} = \mathbf{BEF}, \quad (53)$$

where \mathbf{F} is obtained by stacking columnwise the $\{\mathbf{f}_i\}_{1 \leq i \leq m}$. \square

We now ready to prove Proposition 3.13.

Proof. We are going to show that the feasible sets of problems $(R_P P_\alpha)$ and (RP_α) are the same. First, a feasible point of $(R_P P_\alpha)$ is feasible for (RP_α) . So, we have to a feasible point (RP_α) is also feasible for $(R_P P_\alpha)$. For sake of clarity, we only focus on the upper part of the constraint equality, that is we prove that if $\mathbf{p}(\alpha)$ satisfies:

$$\mathbf{U}^T(\mathbf{A}\widehat{\mathbf{W}})_x \mathbf{p}(\alpha) = \mathbf{U}^T \boldsymbol{\mu}(\alpha), \quad (54)$$

then

$$(\mathbf{A}\widehat{\mathbf{W}})_x \mathbf{p}(\alpha) = \boldsymbol{\mu}(\alpha). \quad (55)$$

By hypothesis, and since the rank is not affected by transposition, the rank equality $\text{rank}((\mathbf{A}\widehat{\mathbf{W}})_x^T) = \text{rank}((\mathbf{A}\widehat{\mathbf{W}})_x^T \mathbf{U})$ holds. From Lemma 3.14, and by transposing again, we deduce that there exists a matrix \mathbf{F} such that $(\mathbf{A}\widehat{\mathbf{W}})_x = \mathbf{F}^T \mathbf{U}^T (\mathbf{A}\widehat{\mathbf{W}})_x$. So, let $\alpha \in \mathcal{A}$ be a parameter, and $\mathbf{p}(\alpha) \in \mathbb{R}^R$ satisfying:

$$\mathbf{U}^T(\mathbf{A}\widehat{\mathbf{W}})_x \mathbf{p}(\alpha) = \mathbf{U}^T \boldsymbol{\mu}(\alpha). \quad (56)$$

Since we supposed that the feasible set of $(R_P P_\alpha)$ is not empty, there exists a vector $\mathbf{s}(\alpha)$ such that

$$\boldsymbol{\mu}(\alpha) = (\mathbf{A}\widehat{\mathbf{W}})_x \mathbf{s}(\alpha) = \mathbf{F}^T \mathbf{U}^T (\mathbf{A}\widehat{\mathbf{W}})_x \mathbf{s}(\alpha). \quad (57)$$

And therefore, mutliplying (56) by \mathbf{F}^T , we obtain

$$(\mathbf{A}\widehat{\mathbf{W}})_x \mathbf{p}(\alpha) = \mathbf{F}^T \mathbf{U}^T (\mathbf{A}\widehat{\mathbf{W}})_x \mathbf{p}(\alpha) = \mathbf{F}^T \mathbf{U}^T \boldsymbol{\mu}(\alpha) = \mathbf{F}^T \mathbf{U}^T (\mathbf{A}\widehat{\mathbf{W}})_x \mathbf{s}(\alpha) = \boldsymbol{\mu}(\alpha). \quad (58)$$

An equivalent reasoning leads to the fact that $\mathbf{p}(\alpha)$ satisfies the lower part of the constraint equality. \square

We shortly notice that if we find $\widehat{\mathbf{U}}$ and $\widehat{\mathbf{V}}$ satisfying the rank assumptions, there is ne need to add constant vectors to them to ensure boundedness of the problem as we did in Section 3.3.

Corollary 3.15. *Under the hypotheses of Property 3.13, the reduced-problem (RP_α) is bounded.*

Proof. The feasible set of (RP_α) is equal to the feasible set of $(R_P P_\alpha)$, which is bounded. \square

Algorithm 1 Gram-Schmidt process

- 1: **Inputs:** $\{\boldsymbol{\mu}_k\}_{1 \leq k \leq K_x}$
 - 2: **Outputs:** $\widehat{\mathbf{U}} := \{\mathbf{u}_k\}_{1 \leq k \leq K_x}$
 - 3: **Initialisation:**
 - 4: $\mathbf{u}_1 = \boldsymbol{\mu}_1$
 - 5: **for** $k = 2, \dots, K_x$ **do**
 - 6: $\mathbf{u}_k = \boldsymbol{\mu}_k - \sum_{z=1}^{k-1} \frac{\mathbf{u}_z^T \boldsymbol{\mu}_k}{\|\mathbf{u}_z\|^2} \mathbf{u}_z$
 - 7: **end for**
-

3.4.2 Gram-Schmidt algorithm to build the matrices $\widehat{\mathbf{U}}$ and $\widehat{\mathbf{V}}$

In this part, we propose a way, namely a Gram-Schmidt algorithm, to build $\widehat{\mathbf{U}}$ and $\widehat{\mathbf{V}}$ that satisfy the rank condition of Proposition 3.13. Moreover, $\widehat{\mathbf{U}}$ and $\widehat{\mathbf{V}}$ will be of dimension $\mathbb{R}^{K_x \times \mathcal{N}_x}$ (resp. $\mathbb{R}^{K_y \times \mathcal{N}_y}$). As before, only focus on the construction of $\widehat{\mathbf{U}}$, because the construction of $\widehat{\mathbf{V}}$ is analogous. To make the following method work, we place ourselves in that case where the generating family $\{\boldsymbol{\mu}_k\}_{1 \leq k \leq K_x}$ is a family of independent vectors (therefore of maximal rank K_x). We first detail the construction in Algorithm 1, and then we explain why $\widehat{\mathbf{U}}$ built this way satisfies the condition of Property 3.13.

Let us verify that the matrix $\widehat{\mathbf{U}}$ built this way satisfies the expected rank condition. This derives from direct consequences of the Gram-Schmidt orthogonalisation process.

Lemma 3.16. *Let $\{\mathbf{u}_k\}_{1 \leq k \leq K_x}$ built via Gram-Schmidt algorithm 1 applied to the family of independent vectors $\{\boldsymbol{\mu}_k\}_{1 \leq k \leq K_x}$. Then*

1. For every $1 \leq k \neq k' \leq K_x$, $\mathbf{u}_k \perp \mathbf{u}_{k'}$
2. For $1 \leq k \leq K_x$, $\text{Im}(\boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_k) = \text{Im}(\mathbf{u}_1, \dots, \mathbf{u}_k)$
3. For $2 \leq k \leq K_x$, $\mathbf{u}_k \in \text{Im}(\boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_{k-1})^\perp$
4. For $1 \leq k \leq K_x$, $\mathbf{u}_k^T \boldsymbol{\mu}_k \neq 0$.

Proof. We prove this lemma recursively: for $k=1$, $\mathbf{u}_1 = \boldsymbol{\mu}_1$, thus $\text{Im}(\mathbf{u}_1) = \text{Im}(\boldsymbol{\mu}_1)$ and $\mathbf{u}_1^T \boldsymbol{\mu}_1 = \|\boldsymbol{\mu}_1\|^2 > 0$.

Let us now suppose the lemma true up to rank $k-1 \geq 1$. Let $1 \leq k' \leq k-1$. Using item 1 up to rank $k-1$ we have :

$$\mathbf{u}_{k'}^T \mathbf{u}_k = \mathbf{u}_{k'}^T \boldsymbol{\mu}_k - \sum_{z=1}^{k-1} \frac{\mathbf{u}_z^T \boldsymbol{\mu}_k}{\|\mathbf{u}_z\|^2} \mathbf{u}_{k'}^T \mathbf{u}_z = \mathbf{u}_{k'}^T \boldsymbol{\mu}_k - \frac{\mathbf{u}_{k'}^T \boldsymbol{\mu}_k}{\|\mathbf{u}_{k'}\|^2} \mathbf{u}_{k'}^T \mathbf{u}_{k'} = 0, \quad (59)$$

which proves 1 up to rank k .

Now, by item 2 at rank $k-1$, \mathbf{u}_k can be seen as a vector combination of $\boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_{k-1}$ and $\boldsymbol{\mu}_k$. Thus $\mathbf{u}_k \in \text{Im}(\boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_k)$, and again by item 2 on rank $k-1$,

$$\text{Im}(\mathbf{u}_1, \dots, \mathbf{u}_{k-1}) = \text{Im}(\boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_{k-1}), \quad (60)$$

thus,

$$\text{Im}(\mathbf{u}_1, \dots, \mathbf{u}_k) \subset \text{Im}(\boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_k). \quad (61)$$

For the converse inclusion, we know that

$$\boldsymbol{\mu}_k = \mathbf{u}_k + \sum_{z=1}^{k-1} \frac{\mathbf{u}_z^T \boldsymbol{\mu}_k}{\|\mathbf{u}_z\|^2} \mathbf{u}_z, \quad (62)$$

thus $\boldsymbol{\mu}_k \in \text{Im}(\mathbf{u}_1, \dots, \mathbf{u}_{k-1}, \mathbf{u}_k)$. Again by item 2 applied on $k-1$ we obtain

$$\text{Im}(\boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_k) \subset \text{Im}(\mathbf{u}_1, \dots, \mathbf{u}_k). \quad (63)$$

We deduce equality from the two inclusions, and item 2 is true at rank k .

Item 3 is a consequence of items 1 and 2 : since $\mathbf{u}_k \perp \mathbf{u}_{k'}$ for $1 \leq k' \leq k-1$, $\mathbf{u}_k \in \text{Im}(\mathbf{u}_1, \dots, \mathbf{u}_{k-1})^\perp$. By item 2, $\text{Im}(\mathbf{u}_1, \dots, \mathbf{u}_{k-1})^\perp = \text{Im}(\boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_{k-1})^\perp$, thus $\mathbf{u}_k \in \text{Im}(\boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_{k-1})^\perp$.

Finally, item 4 is a consequence of items 1 and 2 conjugated to the fact that, by hypothesis, $\boldsymbol{\mu}_k$ is independent of $\boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_{k-1}$. By independence of the family $(\boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_k)$, and by item 2, $\mathbf{u}_k \neq 0$. Indeed, otherwise we would have :

$$0 = \boldsymbol{\mu}_k - \sum_{z=1}^{k-1} \frac{\mathbf{u}_z^T \boldsymbol{\mu}_k}{\|\mathbf{u}_z\|^2} \mathbf{u}_z, \quad (64)$$

thus $\boldsymbol{\mu}_k \in \text{Im}(\mathbf{u}_1, \dots, \mathbf{u}_{k-1}) = \text{Im}(\boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_{k-1})$, and that would contradict the independency hypothesis. Let us now write the scalar product and use item 1 :

$$\mathbf{u}_k^T \boldsymbol{\mu}_k = \mathbf{u}_k^T \left(\mathbf{u}_k + \sum_{z=1}^{k-1} \frac{\mathbf{u}_z^T \boldsymbol{\mu}_k}{\|\mathbf{u}_z\|^2} \mathbf{u}_z \right) = \|\mathbf{u}_k\|^2 > 0, \quad (65)$$

which gives item 4 at rank k . □

Corollary 3.17. *Let $\widehat{\mathbf{U}}$ be the matrix build with Gram-Schmidt algorithm 1 on the family $\{\boldsymbol{\mu}(\boldsymbol{\alpha})\}_{\boldsymbol{\alpha} \in \mathcal{A}}$. Then $\widehat{\mathbf{U}}^T (\boldsymbol{\mu}_0 | \dots | \boldsymbol{\mu}_{K_x})$ is of rank K_x .*

Proof. Using the orthogonality conclusion 3 of the lemma 3.16, we have :

$$\widehat{\mathbf{U}}^T (\boldsymbol{\mu}_0 | \dots | \boldsymbol{\mu}_{K_x}) = \begin{pmatrix} \mathbf{u}_1^T \boldsymbol{\mu}_1 & \dots & \mathbf{u}_1^T \boldsymbol{\mu}_{K_x} \\ \vdots & & \vdots \\ \mathbf{u}_{K_x}^T \boldsymbol{\mu}_1 & \dots & \mathbf{u}_{K_x}^T \boldsymbol{\mu}_{K_x} \end{pmatrix} = \begin{pmatrix} \mathbf{u}_1^T \boldsymbol{\mu}_1 & \dots & \mathbf{u}_1^T \boldsymbol{\mu}_{K_x} \\ 0 & \mathbf{u}_2^T \boldsymbol{\mu}_2 & \dots & \mathbf{u}_2^T \boldsymbol{\mu}_{K_x} & \vdots \\ \vdots & \ddots & \ddots & & \\ 0 & \dots & 0 & \mathbf{u}_{K_x}^T \boldsymbol{\mu}_{K_x} \end{pmatrix}. \quad (66)$$

By conclusion 4 in the lemma 3.16, this upper triangular matrix has non zero diagonal and is thus of maximal rank K_x . □

Corollary 3.18. *Let $\widehat{\mathbf{U}}$ be the matrix build with Gram-Schmidt algorithm 1 on the family $\{\boldsymbol{\mu}(\boldsymbol{\alpha})\}_{\boldsymbol{\alpha} \in \mathcal{A}}$. Then $\text{rank}(\widehat{\mathbf{U}}^T (\mathbf{A}\widehat{\mathbf{W}})_x) = \text{rank}((\mathbf{A}\widehat{\mathbf{W}})_x)$.*

To prove Corollary 3.18 we use the following lemma of linear algebra.

Lemma 3.19. *Let D_1, D_2, D_3 be positive integers and $\mathbf{E} \in \mathbb{R}^{D_1 \times D_2}$ and $\mathbf{F} \in \mathbb{R}^{D_2 \times D_3}$ be matrices. Suppose that $\text{rank}(\mathbf{E}\mathbf{F}) = \text{rank}(\mathbf{F})$. Then, for any positive integer Q , for any matrix $\mathbf{M} \in \mathbb{R}^{D_3 \times Q}$, $\text{rank}(\mathbf{E}\mathbf{F}\mathbf{M}) = \text{rank}(\mathbf{F}\mathbf{M})$.*

Proof. In general, $\ker(\mathbf{F}) \subset \ker(\mathbf{E}\mathbf{F})$. But, since $D_2 = \text{rank}(\mathbf{F}) + \dim(\ker(\mathbf{F})) = \text{rank}(\mathbf{E}\mathbf{F}) + \dim(\ker(\mathbf{E}\mathbf{F}))$, the rank equality implies $\ker(\mathbf{F}) = \ker(\mathbf{E}\mathbf{F})$. Thus

$$\ker(\mathbf{F}\mathbf{M}) = \{\mathbf{f} \in \mathbb{R}^Q, \mathbf{M}\mathbf{f} \in \ker(\mathbf{F})\} = \{\mathbf{f} \in \mathbb{R}^Q, \mathbf{M}\mathbf{f} \in \ker(\mathbf{E}\mathbf{F})\} = \ker(\mathbf{E}\mathbf{F}\mathbf{M}). \quad (67)$$

Therefore, $\text{rank}(\mathbf{E}\mathbf{F}\mathbf{M}) = \text{rank}(\mathbf{F}\mathbf{M})$. □

We now prove Corollary 3.18.

Proof. Since every column of $\text{rank}((\mathbf{A}\widehat{\mathbf{W}})_x)$ is a convex combination of the family of vectors $\{\boldsymbol{\mu}_k\}_{1 \leq k \leq K_x}$, there exists a matrix \mathbf{M} such that :

$$(\boldsymbol{\mu}_0 | \dots | \boldsymbol{\mu}_{K_x}) \mathbf{M} = (\boldsymbol{\mu}(\boldsymbol{\alpha}_1) | \dots | \boldsymbol{\mu}_{K_x}) = (\mathbf{A}\widehat{\mathbf{W}})_x. \quad (68)$$

From corollary 3.17, the hypothesis of lemma 3.19 hold with 3.19 with $\mathbf{E} = \widehat{\mathbf{U}}^T$ and $\mathbf{F} = (\boldsymbol{\mu}_0 | \dots | \boldsymbol{\mu}_{K_x})$. Thus, taking \mathbf{M} as defined above, we obtain :

$$\text{rank}(\widehat{\mathbf{U}}^T (\mathbf{A}\widehat{\mathbf{W}})_x) = \text{rank}(\widehat{\mathbf{U}}^T (\boldsymbol{\mu}_0 | \dots | \boldsymbol{\mu}_{K_x}) \mathbf{M}) = \text{rank}((\boldsymbol{\mu}_0 | \dots | \boldsymbol{\mu}_{K_x}) \mathbf{M}) = \text{rank}(\mathbf{A}\widehat{\mathbf{W}})_x. \quad (69)$$

□

4 A posteriori error estimations

In the previous section, we focused on building a reduced-order model (RP_α) from the high-fidelity model (P_α), and we found conditions ensuring that the model built had solutions. Now, let $\alpha \in \mathcal{A}$ be a parameter, suppose that we have computed numerically the optimal value $I_R(\alpha)$ of the reduced-order model at α . We would like to estimate how far this value is from the optimal cost $I(\alpha)$ of the high-fidelity model. To do so, we search for quantities that bound the value $|I(\alpha) - I_R(\alpha)|$, and that can be computed without having to solve the high-fidelity model at α . We call such a quantity an a posteriori error estimation. We present two error estimations. The first one is built using C-transforms of a pair of potentials solution of the reduced-order model (RD_α). The second one is based on the continuity of the model with respect to the parameter α .

4.1 Error estimation with c -transforms

For this estimation, we suppose that the reduced-order model (RP_α) and the half reduced-order model on primal ($RP_\alpha P_\alpha$) share the same optimal value (this is satisfied when the conditions of Property 3.13 hold). Since the half reduced-order model ($RP_\alpha P_\alpha$) corresponds to the minimization problem described by the high-fidelity model (P_α) to which we add the constraint to belong to the subcone \mathcal{W}_+ , its optimal value is higher than the optimal value of the high-fidelity problem. We thus deduce directly the following inequality for every parameter $\alpha \in \mathcal{A}$:

$$I_R(\alpha) - I(\alpha) \geq 0. \quad (70)$$

We thus focus on bounding the quantity $I_R(\alpha) - I(\alpha)$ from above. Let us first recall, that, by strong duality, the optimal values of the primal and dual of the high-fidelity model (resp. of the reduced-order model) are equal : $J(\alpha) = I(\alpha)$ and $J_R(\alpha) = I_R(\alpha)$. Now, let $(\mathbf{a}^*(\alpha), \mathbf{b}^*(\alpha))$ be a solution of (RD_α) and $\hat{\varphi}_*(\alpha)$ and $\hat{\psi}_*(\alpha)$ the associated Kantorovitch potentials:

$$\begin{aligned} \hat{\varphi}_*(\alpha) &:= \sum_{n=1}^N \mathbf{a}_n^*(\alpha) \mathbf{u}_n \\ \hat{\psi}_*(\alpha) &:= \sum_{m=1}^M \mathbf{b}_m^*(\alpha) \mathbf{v}_m. \end{aligned} \quad (71)$$

Since this pair is optimal, the following equality holds :

$$I_R(\alpha) = \langle \hat{\varphi}_*(\alpha), \boldsymbol{\mu}(\alpha) \rangle + \langle \hat{\psi}_*(\alpha), \boldsymbol{\nu}(\alpha) \rangle. \quad (72)$$

We now use the c -transform and \bar{c} -transform defined in 2.3 to create admissible points of the dual high-fidelity model (D_α). Indeed, the pairs $(\hat{\varphi}_*(\alpha), \hat{\varphi}_*^c(\alpha))$ and $(\hat{\psi}_*^{\bar{c}}(\alpha), \hat{\psi}_*(\alpha))$ are admissible for (D_α) (see e.g. [46] for a proof). Let us focus on the pair $(\hat{\varphi}_*(\alpha), \hat{\varphi}_*^c(\alpha))$. Since it is admissible, we have the following inequality :

$$I(\alpha) \geq \langle \hat{\varphi}_*(\alpha), \boldsymbol{\mu}(\alpha) \rangle + \langle \hat{\varphi}_*^c(\alpha), \boldsymbol{\nu}(\alpha) \rangle. \quad (73)$$

We now subtract (72) to (73), and we obtain the following expression :

$$I_R(\alpha) - I(\alpha) \leq \langle \hat{\psi}_*(\alpha) - \hat{\varphi}_*^c(\alpha), \boldsymbol{\nu}(\alpha) \rangle. \quad (74)$$

By performing the same computation with $(\hat{\psi}_*^{\bar{c}}(\alpha), \hat{\psi}_*(\alpha))$, we obtain :

$$I_R(\alpha) - I(\alpha) \leq \langle \hat{\varphi}_*(\alpha) - \hat{\psi}_*^{\bar{c}}(\alpha), \boldsymbol{\mu}(\alpha) \rangle. \quad (75)$$

We thus obtain the following a posteriori error estimation that bounds from above the difference $I_R(\alpha) - I(\alpha)$:

$$I_R(\alpha) - I(\alpha) \leq \min \left(\langle \hat{\varphi}_*^c(\alpha) - \hat{\psi}_*(\alpha), \boldsymbol{\nu}(\alpha) \rangle, \langle \hat{\psi}_*^{\bar{c}}(\alpha) - \hat{\varphi}_*(\alpha), \boldsymbol{\mu}(\alpha) \rangle \right). \quad (76)$$

Since this estimation uses c -transforms and \bar{c} -transforms, in general, it depends non linearly on the parameter α . We show in what follow that the exact computation of this estimation is of order $N_x N_y$. We then propose a way to approach this quantity by a variant of the Empirical Interpolation Method (EIM).

Evaluation of the computational We focus here on the computational cost of the term $\langle \widehat{\varphi}_*^c(\boldsymbol{\alpha}) - \widehat{\psi}_*(\boldsymbol{\alpha}), \boldsymbol{\nu}(\boldsymbol{\alpha}) \rangle$. We can deduce by analogy the number of operations needed to evaluate $\langle \widehat{\psi}_*^c(\boldsymbol{\alpha}) - \widehat{\varphi}_*(\boldsymbol{\alpha}), \boldsymbol{\mu}(\boldsymbol{\alpha}) \rangle$.

We begin by replacing $\widehat{\psi}_*(\boldsymbol{\alpha})$ with its expression in (71), and we express $\boldsymbol{\nu}(\boldsymbol{\alpha})$ as a convex combination of $\{\boldsymbol{\nu}^l\}_{l=1}^{K_y}$. We obtain a weighted sum of MK_y terms:

$$\langle \widehat{\psi}_*(\boldsymbol{\alpha}), \boldsymbol{\nu}(\boldsymbol{\alpha}) \rangle = \sum_{m=1}^M \sum_{l=1}^{K_y} \alpha_l^y b_m \langle \mathbf{v}_m^*, \boldsymbol{\nu}^l \rangle. \quad (77)$$

The expression $\langle \widehat{\varphi}_*^c(\boldsymbol{\alpha}), \boldsymbol{\nu}(\boldsymbol{\alpha}) \rangle$ can be decomposed as follows:

$$\langle \widehat{\varphi}_*^c(\boldsymbol{\alpha}), \boldsymbol{\nu}(\boldsymbol{\alpha}) \rangle = \sum_{l=1}^{K_y} \alpha_l^y \langle \widehat{\varphi}_*^c(\boldsymbol{\alpha}), \boldsymbol{\nu}^l \rangle. \quad (78)$$

In general, the c -transform operation does not depend linearly on the parameter $\boldsymbol{\alpha}$, and its exact evaluation requires $N_y N_x$ operations.

EIM for fast evaluation of the error estimation For a given $\boldsymbol{\alpha} \in \mathcal{A}$, the exact evaluations of $\widehat{\varphi}_*^c(\boldsymbol{\alpha})$ and of $\langle \widehat{\varphi}_*^c(\boldsymbol{\alpha}), \boldsymbol{\nu}(\boldsymbol{\alpha}) \rangle$ both require around $N_x N_y$ operations. In this section we implement a variant of the Empirical Interpolation Method (EIM) that enables to estimate $\widehat{\varphi}_*^c(\boldsymbol{\alpha})$ at a reduced computational cost. (see e.g. [33] or [2] for a general description of the EIM). More specifically, here our aim is to approximate $\widehat{\varphi}_*^c(\boldsymbol{\alpha})$ as a linear combination of M^{eim} independent vectors $\{\mathbf{q}^\ell\}_{1 \leq \ell \leq M^{\text{eim}}}$:

$$\widehat{\varphi}_*^c(\boldsymbol{\alpha}) \approx \sum_{\ell=1}^{M^{\text{eim}}} \beta_\ell(\boldsymbol{\alpha}) \mathbf{q}^\ell. \quad (79)$$

The coefficients $\{\beta_\ell(\boldsymbol{\alpha})\}_{1 \leq \ell \leq M^{\text{eim}}}$ are found by solving a linear system of size M^{eim^2} , where the right hand side vector is formed of M^{eim} fast evaluations of the c -transform of $\widehat{\varphi}_*(\boldsymbol{\alpha})$:

$$\sum_{\ell=1}^{M^{\text{eim}}} \mathbf{q}_j^\ell \beta_\ell(\boldsymbol{\alpha}) = \widehat{\varphi}_{*j}^{\text{Fast-C}}(\boldsymbol{\alpha}), \text{ for } j \in J^{\text{eim}}, \quad (80)$$

where

$$\widehat{\varphi}_{*j}^{\text{Fast-C}} := \arg \min_{i \in I^{\text{eim}}} \mathbf{C}_{i,j} - \widehat{\varphi}_{*i}(\boldsymbol{\alpha}). \quad (81)$$

How to find $\{\mathbf{q}^\ell\}_{1 \leq \ell \leq M^{\text{eim}}}$, the evaluation points and how to build a fast evaluation are determined by the offline phase of the EIM described in Algorithm 2.

In the end, we have the following approximation:

$$\langle \widehat{\varphi}_*^c(\boldsymbol{\alpha}), \boldsymbol{\nu}(\boldsymbol{\alpha}) \rangle \approx \sum_{l=1}^{K_y} \sum_{\ell=1}^{M^{\text{eim}}} \beta_\ell \alpha_l^y \langle \mathbf{q}^\ell, \boldsymbol{\nu}^l \rangle. \quad (82)$$

In term of computational costs, the determination of $\{\beta_\ell\}_{1 \leq \ell \leq M^{\text{eim}}}$ requires of order $M^{\text{eim}} M^{\text{eim}}$ for finding the right-hand side, then of order M^{eim^2} to solve the system (it has a triangular structure). The final scalar product requires around $K_y M^{\text{eim}}$ operations.

4.2 Error estimation from continuous dependency in $\boldsymbol{\alpha}$

In this part, we use the continuity Property 2.8 to build an upper bound of the quantity $|I(\boldsymbol{\alpha}) - I'_R(\boldsymbol{\alpha})|$. Let $\mathcal{A}_{\text{train}} \subset \mathcal{A}$ be a training set, and suppose that we know the optimal cost of the high-fidelity problem $I(\boldsymbol{\alpha}')$ for every parameter $\boldsymbol{\alpha}' \in \mathcal{A}_{\text{train}}$ of this training set. Let $\boldsymbol{\alpha} \in \mathcal{A}$ be a parameter, then, for very $\boldsymbol{\alpha}' \in \mathcal{A}_{\text{train}}$, we have :

$$|I(\boldsymbol{\alpha}) - I_R(\boldsymbol{\alpha})| \leq |I(\boldsymbol{\alpha}) - I(\boldsymbol{\alpha}')| + |I(\boldsymbol{\alpha}') - I_R(\boldsymbol{\alpha})|. \quad (83)$$

Algorithm 2 EIM (offline phase)

- 1: **Inputs:** $\mathcal{G} := \{\hat{\varphi}(\alpha)\}_{\alpha \in \mathcal{A}_{train}}$, $\mathcal{G}^c := \{\hat{\varphi}^C(\alpha)\}_{\alpha \in \mathcal{A}_{train}}$ $M^{\text{eim}}, M'^{\text{eim}}$
 - 2: **Outputs:** $I^{\text{eim}} \subset \{1, \dots, N_x\}$, $J^{\text{eim}} \subset \{1, \dots, N_y\}$, $\{\mathbf{q}^1, \dots, \mathbf{q}^{M^{\text{eim}}}\} \in \mathbb{R}^{N_y \times M^{\text{eim}}}$
 - 3: **Initialisation:**
 - 4: $\mathbf{g}^1 = \arg \max_{\hat{\varphi}^C \in \mathcal{G}^c} |\hat{\varphi}^C|$
 - 5: $j_1 = \arg \max_{1 \leq j \leq N_y} |\mathbf{g}_j^1|$
 - 6: $\mathbf{q}^1 = \frac{\mathbf{g}^1}{\mathbf{g}_{j_1}^1}$
 - 7: $J_1[\hat{\varphi}] = \hat{\varphi}_{j_1} \mathbf{q}^1$
 - 8: $J^{\text{eim}} \leftarrow \{j_1\}$
 - 9: **for** $m^{\text{eim}} = 2, \dots, M^{\text{eim}}$ **do**
 - 10: $\mathbf{g}^{m^{\text{eim}}} = \arg \max_{\hat{\varphi}^C \in \mathcal{G}^c} \|\hat{\varphi}^C - J_{m^{\text{eim}}-1}[\hat{\varphi}^C]\|$
 - 11: $j_{m^{\text{eim}}} = \arg \max_{1 \leq j \leq N_y} |\mathbf{g}_j^{m^{\text{eim}}}|$
 - 12: $\mathbf{q}^{m^{\text{eim}}} = \frac{\mathbf{g}^{m^{\text{eim}}} - J_{m^{\text{eim}}-1}[\mathbf{g}^{m^{\text{eim}}}]}{(\mathbf{g}^{m^{\text{eim}}} - J_{m^{\text{eim}}-1}[\mathbf{g}^{m^{\text{eim}}}]_{j_{m^{\text{eim}}}})}$
 - 13: $J_{m^{\text{eim}}}[\hat{\varphi}^C] = \sum_{\ell=1}^{m^{\text{eim}}} \beta_\ell \mathbf{q}^\ell$,
 - 14: where β solves $\sum_{\ell=1}^{m^{\text{eim}}} \beta_\ell \mathbf{q}_{j_\ell}^\ell = \hat{\varphi}_{j_{m^{\text{eim}}}}^C$ for $1 \leq \ell \leq m^{\text{eim}}$
 - 15: $J^{\text{eim}} \leftarrow J \cup \{j_{m^{\text{eim}}}\}$
 - 16: **end for**
 - 17: **for** $1 \leq m_1^{\text{eim}} \leq M^{\text{eim}}$ **do**
 - 18: **for** $1 \leq m_2^{\text{eim}} \leq M'^{\text{eim}}$ **do**
 - 19: $i_{m_1^{\text{eim}}, m_2^{\text{eim}}} = \arg \min_{1 \leq i \leq N_x} \mathbf{C}_{i, j_{m_2^{\text{eim}}}} - \mathbf{g}_{j_{m_2^{\text{eim}}}}^{m_1^{\text{eim}}}$
 - 20: $I^{\text{eim}} \leftarrow \{i_{m_1^{\text{eim}}, m_2^{\text{eim}}}\}$
 - 21: **end for**
 - 22: **end for**
-

Using Property 2.8, we obtain

$$|I(\boldsymbol{\alpha}) - I_R(\boldsymbol{\alpha})| \leq |I(\boldsymbol{\alpha}') - I_R(\boldsymbol{\alpha}')| + \|\mathbf{C}\|_\infty (2 \max(K_x, K_y) + 3 \min(K_x, K_y)) \|\boldsymbol{\alpha} - \boldsymbol{\alpha}'\|_\infty. \quad (84)$$

We can then control this difference by choosing a suitable $\boldsymbol{\alpha}' \in \mathcal{A}_{\text{train}}$, for example $\boldsymbol{\alpha}' \in \arg \min_{\boldsymbol{\alpha}' \in \mathcal{A}_{\text{train}}} (|I(\boldsymbol{\alpha}') - I_R(\boldsymbol{\alpha}')| + \|\mathbf{C}\|_\infty (2 \max(K_x, K_y) + 3 \min(K_x, K_y)) \|\boldsymbol{\alpha} - \boldsymbol{\alpha}'\|_\infty)$ or $\boldsymbol{\alpha}' \in \arg \min_{\boldsymbol{\alpha}' \in \mathcal{A}_{\text{train}}} \|\boldsymbol{\alpha} - \boldsymbol{\alpha}'\|_\infty$. The complexity of this estimation depends on the cardinal $\#\mathcal{A}_{\text{train}}$ of $\mathcal{A}_{\text{train}}$ and of the way to choose $\boldsymbol{\alpha}'$. For the proposed $\boldsymbol{\alpha}'$ it will be of order $\#\mathcal{A}_{\text{train}}$.

5 Numerical results

In this section, we first use a toy example to give some numerical results about our method. More precisely, we analyse on this example simultaneously the time gain and precision obtained with our method in comparison to solving the high-fidelity problem with the linear programming solver `linprog` of the Python library `scipy.optimize`. We compare these results to the time gain and precision of the classical Sinkhorn algorithm. We then test the behaviors of the two a posteriori error estimations we built. Finally, we apply our reduction method to a problem of 3D optimal transport applied to a transfer of color palette. Every numerical experiment can be found on the two notebooks https://colab.research.google.com/drive/10G9mqaIrASvS_ucC-V9sngRppLmto0nr?usp=sharing and https://colab.research.google.com/drive/1m7X29BdsbFLV5iXbk7oe5_OKY_lsga76?usp=sharing.

Remark 5.1. *In each presented reduced-order model, the dual reduced basis is built using the method presented in Section 3.4. Thus, when the size of a reduced-basis is mentioned, it refers to the dimension of the primal reduced-cone.*

5.1 Analysis of the numerical results of the reduced-order model in dimension 1

In the sequel, we use a simple example of parametrized optimal transport problem to analyze numerically the performances of the reduced-order model (RP_α) in term of time gain and precision with respect to the high-fidelity model (P_α). We begin by describing the high-fidelity model built. We then briefly display at a fixed parameter solutions of the reduced-order model for different sizes of reduced bases, and compare it to a reference solution of the high-fidelity model.

The setting The family of optimal transport problems we aim at studying is a 1-dimensional optimal transport problem on the domains $X = \{x_i = -1 + \frac{2i}{N_x}\}_{1 \leq i \leq N_x}$ and $Y = \{y_j = -1 + \frac{2j}{N_y}\}_{1 \leq j \leq N_y}$.

We chose the quadratic cost $\mathbf{C}_{ij} = \|y_j - x_i\|^2$, for $1 \leq i \leq N_x$, $1 \leq j \leq N_y$. We chose the family of marginals to be convex combinations of 2 normalized gaussians given by:

$$\boldsymbol{\mu}_0^i = \frac{e^{-\frac{(x_i - m_0^\mu)^2}{2\sigma_0^{\mu 2}}}}{\sum_{i^n=0}^{N_x} e^{-\frac{(x_{i^n} - m_0^\mu)^2}{2\sigma_0^{\mu 2}}}}, \boldsymbol{\mu}_1^i = \frac{e^{-\frac{(x_i - m_1^\mu)^2}{2\sigma_1^{\mu 2}}}}{\sum_{i^n=0}^{N_x} e^{-\frac{(x_{i^n} - m_1^\mu)^2}{2\sigma_1^{\mu 2}}}} \text{ for } 1 \leq i \leq N_x, \quad (85)$$

and

$$\boldsymbol{\nu}_0^j = \frac{e^{-\frac{(y_j - m_0^\nu)^2}{2\sigma_0^{\nu 2}}}}{\sum_{j^n=0}^{N_y} e^{-\frac{(y_{j^n} - m_0^\nu)^2}{2\sigma_0^{\nu 2}}}}, \boldsymbol{\nu}_1^j = \frac{e^{-\frac{(y_j - m_1^\nu)^2}{2\sigma_1^{\nu 2}}}}{\sum_{j^n=0}^{N_y} e^{-\frac{(y_{j^n} - m_1^\nu)^2}{2\sigma_1^{\nu 2}}}} \text{ for } 1 \leq j \leq N_y, \quad (86)$$

where $m_0^\mu = -\frac{1}{2}$, $m_1^\mu = \frac{1}{2}$, $m_0^\nu = -\frac{1}{2}$, $m_1^\nu = \frac{1}{2}$, and $\sigma_0^\mu = \sigma_1^\mu = \sigma_0^\nu = \sigma_1^\nu = \frac{1}{2}$. Figure 1 gives an example of a parametrized measure obtained in this configuration. We set the discretizations parameters at $N_x = N_y = 100$.

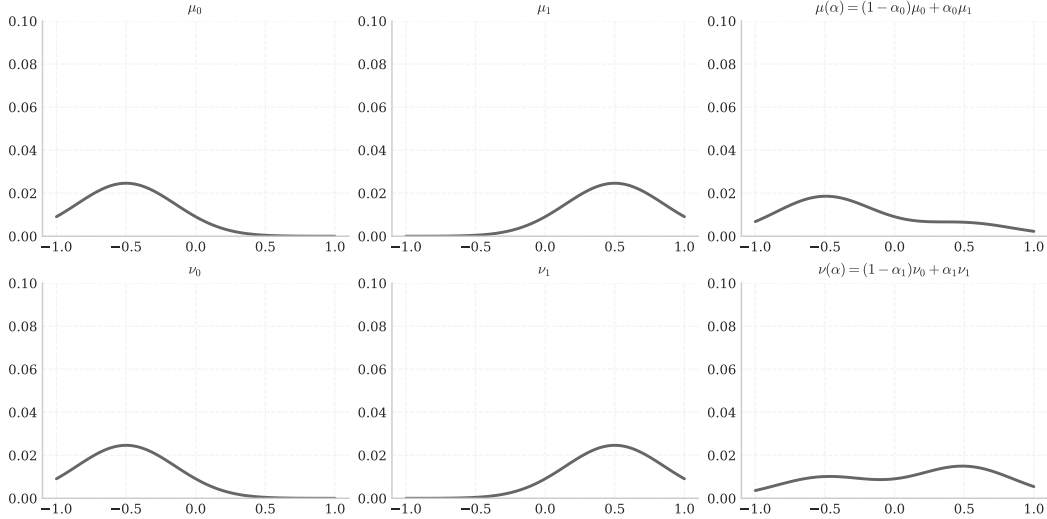


Figure 1: Visualization of the parametrized marginals

Snapshot grid $\mathcal{A}_{\text{train}}^P$ We built the training set $\mathcal{A}_{\text{train}}^P$ as a regular grid of $[0, 1]^2$. We then made the precision vary from $4 = 2^2$ to $400 = 20^2$ snapshots.

Shape of solutions for the ROM On Figure 2 we show the general shape of the solutions of the reduced-order model (ROM) obtained with different sizes of reduced bases (RB), and we compare it to the solution of the high-fidelity model. By definition, solutions of the ROM are positive (even convex) combinations of solutions of the high-fidelity model, and that is why in practice they have a sort of filament shape (and not the shape of a map anymore). The bigger the reduced bases are, the more the reduced solutions get a shape that looks similar to a map.

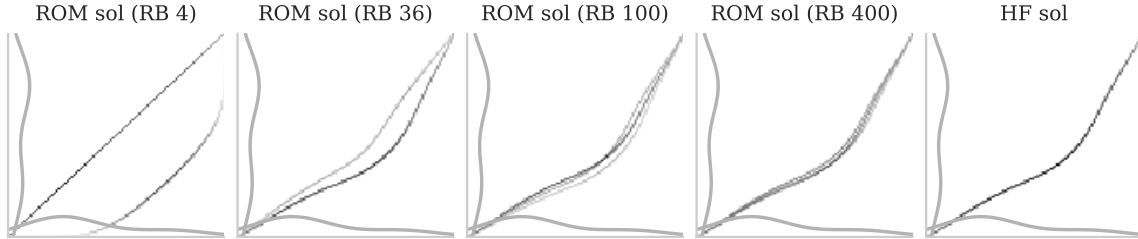


Figure 2: Left : solutions of the reduced order-model (ROM) for different sizes of reduced bases (RB). Right : solution of the high-fidelity model (HF) of size (10000×200) .

Evolution of the error with respect with the number of snapshots taken Figure 3 describes the evolution of mean error between the optimal value of the High-fidelity problem (P_α) and the reduced-order model (RP_α) with respect to the number of snapshots taken to build the reduced basis. As expected, the error decreases, as the snapshot grid gets finer.

Comparison of performances of the reduced-order model with those of Sinkhorn We show on Figure 4 the time gain and accuracy evolution of the reduced-order model for different sizes of reduced bases. We observe that, the bigger the reduced-basis gets, the less the error between the reduced-order model and the high-fidelity model is, but also the less the time gain is. We also plot the results obtained with a Sinkhorn method, for different values of entropy ε . It appears that for the best precision points of both methods, the time gain of the reduced-order model is greater.

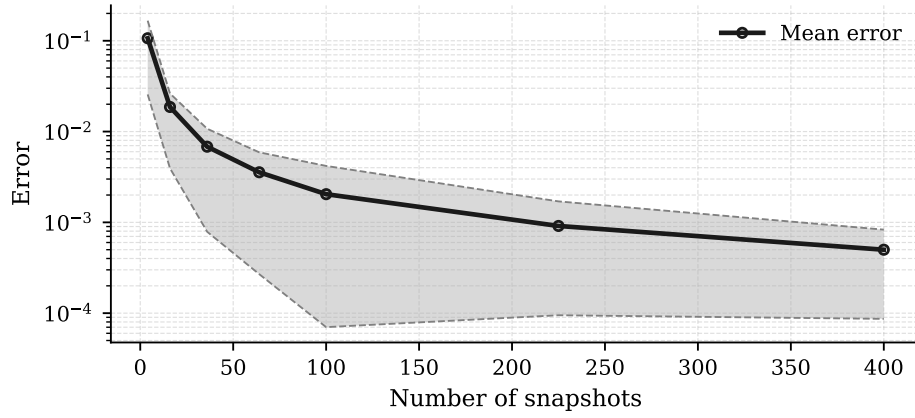


Figure 3: Evolution of the mean error over a test set of 50 random parameters with respect to the number of snapshots taken. The grey area is delimited by the minimal and maximal error obtained on this test set.

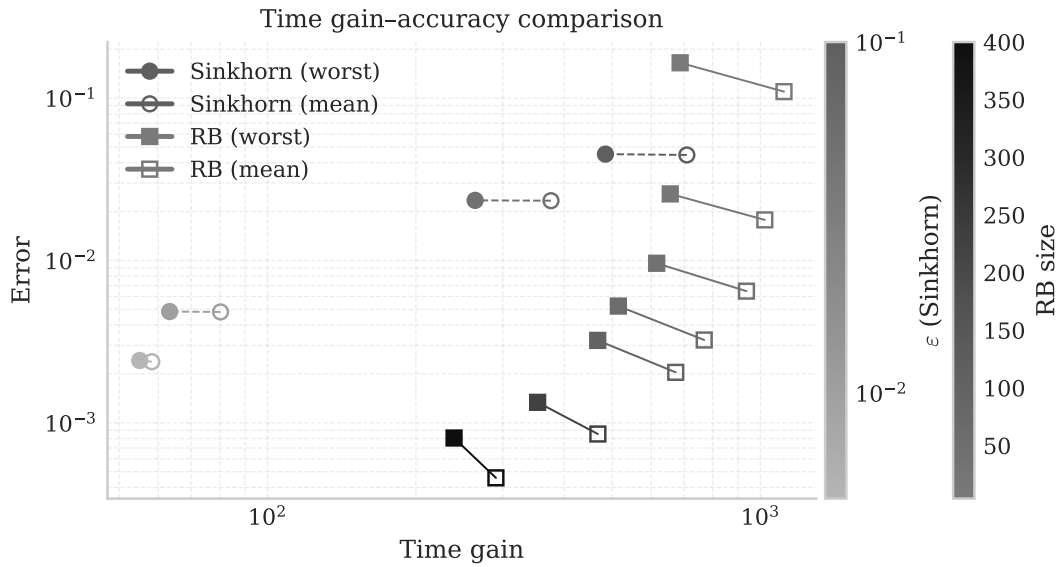


Figure 4: Comparison of precision and time gain between the reduced-order model and the Sinkhorn method. We tested the method on a test set of 50 random parameters, and plot in each case worst and mean performance.

A posteriori error estimations with c -transforms In the sequel, we test the first a posteriori error estimation build in Section 4. We begin by comparing the exact error estimation to its fast evaluation. We then compare the real error of the reduced-order model with the estimated one. It appears that the a posteriori error estimation may be far from the real error. We thus suggest a way to estimate the gap between the real error and the estimation in order to obtain a more accurate error estimation.

Comparison between the exact a posteriori error estimation and its fast evaluation We test here the accuracy of the fast estimation presented in Section 4.1 that uses an EIM interpolation. Figure 5 shows the evolution of the gap $|\text{fast error estimation} - \text{exact error estimation}|$ with respect to the size M^{eim} of the EIM basis. More precisely, it shows the mean, minimum and maximum gaps computed on a test set of 50 random parameters. It appears that this difference goes rapidly beyond the threshold 10^{-1} but then tends to diminish slowly.

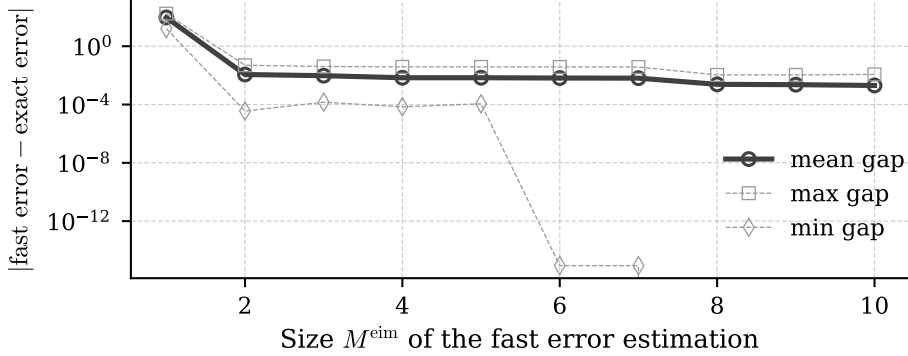


Figure 5: Gap between the fast evaluation and the exact evaluation of the a posteriori error estimation in function of the size of the EIM estimation taken. The curves are built as a mean/min/max of this gap over a test set of 50 random parameters.

Precision of the error estimation Figure 6 compares on a test set of 50 random parameters the mean exact error and the mean error obtained with the a posteriori error estimation using c -transforms. In the following, we compare the exact a posteriori error estimation and its fast approximation to the real error between the cost of the High-fidelity model and the reduced-order model. On figure 6, we observe that both fast - and exact error estimations behave poorly when the matrices \mathbf{U} and \mathbf{V} are built via the method proposed in 3.4.

Corrected error estimation In order to get a better estimation, we estimate using a training set \mathcal{A}_{est} three constants C_{max} , C_{min} and C_{mean} that correspond to the maximum (respectively the minimum and the mean) of the ratio between the real error and the error approximation evaluated on every parameter of \mathcal{A}_{est} . We then multiply the a posteriori error estimation by these constants. Using this method, we obtain a more precise estimation, as shown on figure 7.

A posteriori error estimations using continuity Figure 8 shows the a posteriori error estimation obtained using the bound computed in Section. 4.2. The bigger the training set, the finer the estimation, but we are still losing several order of magnitude in comparison to the real error. We may improve numerically the estimation by computing constant corrections as we did in Section 5.1. We show the resulting estimation on Figure 9.

5.2 ROM in high dimensions

The reduction method that we proposed in Section 3 formally requires to build a matrix $\widehat{\mathbf{W}}$ of size $\mathcal{N}_x \mathcal{N}_y \times R$, which can be numerically challenging when \mathcal{N}_x and \mathcal{N}_y are large. We show in this section

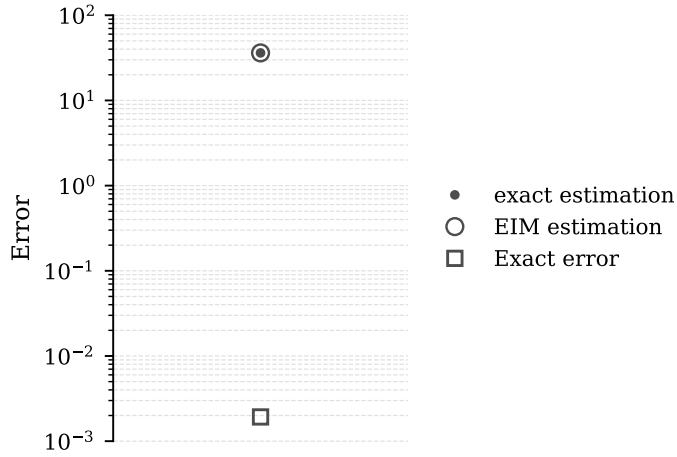


Figure 6: Comparison of the exact error and the a posteriori error estimation. Here we represent a mean of errors computed on a test set of 50 random parameters.

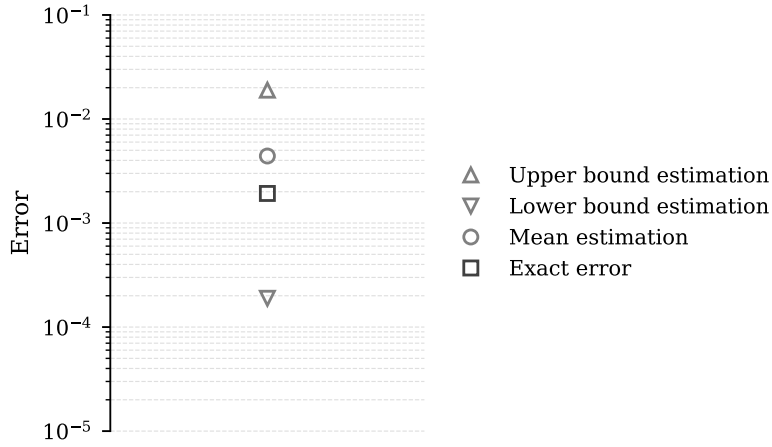


Figure 7: Error correction with a multiplicative constant. The values correspond to a mean over 50 random parameters. The training set used to approximate the constant is of size 100.

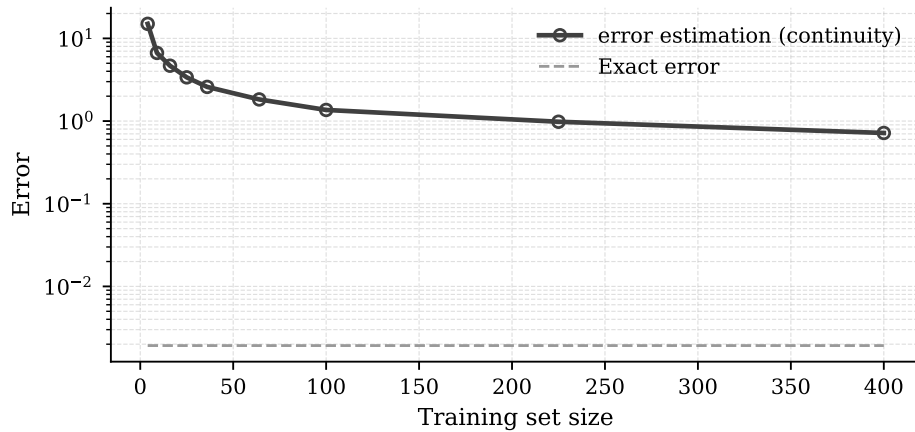


Figure 8: A posteriori error estimation using continuity in function of the size of the training set.

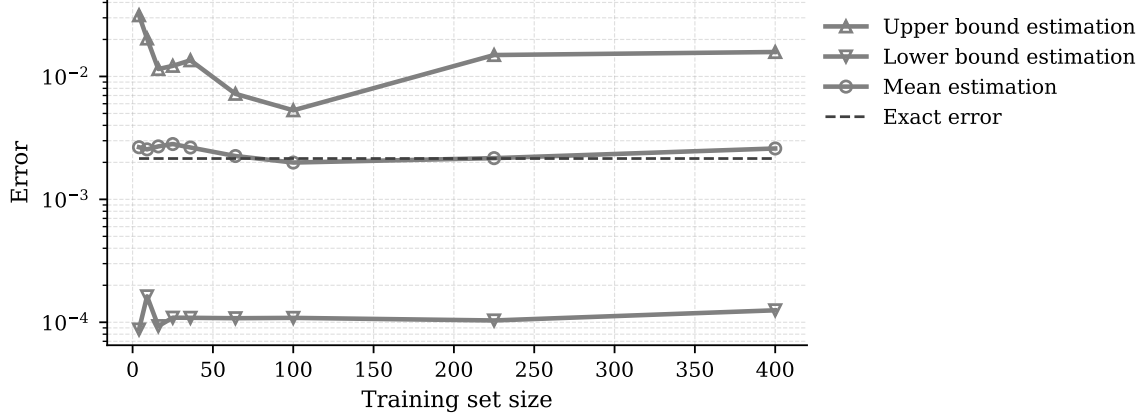


Figure 9: Error correction with a multiplicative constant. The training set used to approximate the constant is of size 100.

that we do not need to compute the matrix $\widehat{\mathbf{W}}$ to build the reduced-order model (RP_α). We then propose an efficient way to reconstruct a map from a reduced-order solution.

Building the reduced-order model without computing \mathbf{W} We show here that there is no need to compute explicitly $\widehat{\mathbf{W}}$ to build the reduced order model (RP_α). The matrix $\widehat{\mathbf{W}}$ has the form

$$\widehat{\mathbf{W}} = (\pi_*(\alpha_1)| \cdots | \pi_*(\alpha_R)), \quad (87)$$

where $\{\pi_*(\alpha_r)\}_{1 \leq r \leq R}$ are solutions of the high-fidelity problems $\{P_{\alpha_r}\}_{1 \leq r \leq R}$. In the reduced-order model (RP_α), $\widehat{\mathbf{W}}$ appears in the cost in the form $\widehat{\mathbf{W}}^T \mathbf{c}$ and in the constraints in the form $\mathbf{A} \widehat{\mathbf{W}}$. But both of these matrix can be reformulated. On one hand, we have

$$\widehat{\mathbf{W}}^T \mathbf{c} = (\pi_*(\alpha_1)^T \mathbf{c} | \cdots | \pi_*(\alpha_R)^T \mathbf{c})^T = (I(\alpha_1) | \cdots | I(\alpha_R))^T, \quad (88)$$

Thus $\widehat{\mathbf{W}}^T \mathbf{c}$ is obtained by stacking the optimal costs of the high-fidelity problems at $\{\alpha_r\}_{1 \leq r \leq R}$. On the other hand,

$$\mathbf{A} \widehat{\mathbf{W}} = \begin{pmatrix} \mu(\alpha_1) | \cdots | \mu(\alpha_R) \\ \nu(\alpha_1) | \cdots | \nu(\alpha_R) \end{pmatrix}, \quad (89)$$

thus $\mathbf{A} \widehat{\mathbf{W}}$ can even be obtained without solving any problem.

Reconstructing solutions in high-dimensions from solutions of (RP_α) In some applications (see for instance Section 5.3), one may need to build from a solution of the reduced-order model a transport map approximating the solution of the high-fidelity model (P_α). As show in Figure (2), the solutions of the the reduced-order model does not have the shape of a map. A similar issue is also encountered when solving Optimal Transport with Sinkhorn method, regularized transport [16] or with sliced optimal transport [41] [44]. In [16], the transport map is built from a transport plan $\{\Pi_{ij}\}_{1 \leq i \leq \mathcal{N}_x, 1 \leq j \leq \mathcal{N}_y}$ as follows

$$\mathbf{T}_i := \left[\frac{\sum_{1 \leq j \leq \mathcal{N}_y} \Pi_{ij} j}{\sum_{1 \leq j \leq \mathcal{N}_y} \Pi_{ij}} \right] \quad \text{for } 1 \leq i \leq \mathcal{N}_x. \quad (90)$$

We adapt this computation to the reduced order model. Indeed, given a solution $\widehat{\mathbf{p}}_*(\boldsymbol{\alpha})$ of the reduced-order model (RP_α), one can build \mathbf{T}_* in the following way

$$\begin{aligned} \mathbf{T}_{*i} &:= \left[\frac{\sum_{1 \leq j \leq N_y} \sum_{1 \leq r \leq R} \mathbf{p}_*^r(\boldsymbol{\alpha}) \mathbf{\Pi}_{*ij}(\boldsymbol{\alpha}_r) j}{\sum_{1 \leq j \leq N_y} \sum_{1 \leq r \leq R} \mathbf{p}_*^r(\boldsymbol{\alpha}) \mathbf{\Pi}_{*ij}(\boldsymbol{\alpha}_r) j} \right] \\ &= \left[\frac{\sum_{1 \leq r \leq R} \mathbf{p}_*^r(\boldsymbol{\alpha}) \left(\sum_{1 \leq j \leq N_y} \mathbf{\Pi}_{*ij}(\boldsymbol{\alpha}_r) j \right)}{\sum_{1 \leq r \leq R} \mathbf{p}_*^r(\boldsymbol{\alpha}) \left(\sum_{1 \leq j \leq N_y} \mathbf{\Pi}_{*ij}(\boldsymbol{\alpha}_r) \right)} \right], \end{aligned} \quad (91)$$

where $\{\mathbf{\Pi}_*(\boldsymbol{\alpha}_r)\}_{1 \leq r \leq R}$ are the matrix versions of the the transport plans $\{\pi_*(\boldsymbol{\alpha}_r)\}_{1 \leq r \leq R}$. This way, after precomputations, we reduce the initial building cost $N_x N_y$ to $R N_x$.

5.3 Application : Reduced-bases applied to color transfer

In this section, we present an application of the reduction method to a color transfer problem. The principle of color transfer is to modify an image so that its color histogram matches a target histogram (named color palette). To do so, [41] and [44] propose to match the histogram of the original image to the target palette with a map that minimizes the total transport cost between these two histograms. In the present case, we propose a variation of this problem: given an image and K_y color histograms, we would like to transfer the color histogram of our image to color palettes built as convex combinations of the K_y histograms. We begin by introducing the transport problem solved to perform the color transfer. We then compare the results obtained with the reduced-order model to what is obtained solving the high-fidelity model.

Principle of color transfer Given a source image Im , we denote by $\boldsymbol{\mu} \in \mathbb{R}^{\mathcal{N}_x^3}$ its color histogram. Let $\{\text{Target}_l\}_{1 \leq l \leq K_y}$ be target palettes and $\{\boldsymbol{\nu}_l\}_{1 \leq l \leq K_y} \in (\mathbb{R}^{\mathcal{N}_y^3})^{K_y}$ be their corresponding color histograms. Given a convex combination $\boldsymbol{\nu}(\boldsymbol{\alpha}) := \sum_{1 \leq l \leq K_y} \alpha_l \boldsymbol{\nu}_l$ of these color histograms, we would like to find a

map $T(\boldsymbol{\alpha}) : \mathbb{R}^{\mathcal{N}_x^3} \rightarrow \mathbb{R}^{\mathcal{N}_y^3}$ that transports $\boldsymbol{\mu}$ onto $\boldsymbol{\nu}(\boldsymbol{\alpha})$ and that minimize the total transport cost :

$$\begin{aligned} \min_{T : \mathbb{R}^{\mathcal{N}_x^3} \rightarrow \mathbb{R}^{\mathcal{N}_y^3}} \sum_{i: T(i)=j} \boldsymbol{\mu}_i = \boldsymbol{\nu}(\boldsymbol{\alpha})_j, \text{ for } j = (j_1, j_2, j_3), 1 \leq j_1, j_2, j_3 \leq N_y \quad \sum_{i = (i_1, i_2, i_3), 1 \leq i_1, i_2, i_3 \leq N_x} \mathbf{C}_{iT(i)}, \end{aligned} \quad (92)$$

where $\mathbf{C} : \mathbb{R}^{\mathcal{N}_x^3 \times \mathcal{N}_y^3} \rightarrow \mathbb{R}_+$. In practice, for $i = (i_1, i_2, i_3)$ and $j = (j_1, j_2, j_3)$, $\{\mathbf{C}_{i,j} = |i_1 - j_1|^2 + |i_2 - j_2|^2 + |i_3 - j_3|^2\}$. As underlined in [40], this discrete Monge problem is possibly not well defined. We relax this problem with the Kantorovitch formulation

$$\begin{aligned} \min_{\mathbf{\Pi} \in \mathbb{R}_+^{\mathcal{N}_x^3 \times \mathcal{N}_y^3}} \langle \mathbf{\Pi}, \mathbf{C} \rangle. \end{aligned} \quad (93)$$

$$\begin{aligned} \sum_j \mathbf{\Pi}_{i,j} &= \boldsymbol{\mu}_i \text{ for } i = (i_1, i_2, i_3), 1 \leq i_1, i_2, i_3 \leq N_x, \\ \sum_i \mathbf{\Pi}_{i,j} &= \boldsymbol{\nu}_j(\boldsymbol{\alpha}) \text{ for } j = (j_1, j_2, j_3), 1 \leq j_1, j_2, j_3 \leq N_y \end{aligned}$$

Problem (93) reads as a 3D and two-marginal optimal transport problem. To solve it and build a reduced order model, we use the Sinkhorn method designed in [13]. Once an optimal transport plan has been built, it is possible to obtain a map following the method described in Section 5.2.

Results In the following numerical tests, $N_x = N_y = 64$ so that $N_x^3 = N_y^3 \approx 2.6 \cdot 10^5$. Table 2 compares the time taken on a commodity laptop to solve one transfer color problem (Optimal transport problem + building a map) with Sinkhorn method and with the reduced-basis approach (Here, the reduced-order model is built with only 3 snapshots). The reduced model achieves a significant speed-up over the high-fidelity Sinkhorn method.

Method	Time (s)	Speed-up
Solving Reduced model	0.0208	$\times 333$
High-fidelity (Sinkhorn)	6.93	1

Table 2: Computation time comparison. The value corresponds to the mean computation time on a test set of 5 parameters.

On Figure 10, we compare the results of the reduced basis method to the results obtained via solving the high-fidelity model. At first sight, the images obtained with both methods look similar. On Figure 11, we exhibit that one can still observe differences. Indeed, on this example, the red tones appear less vivid in the picture obtained with the reduced-order model.

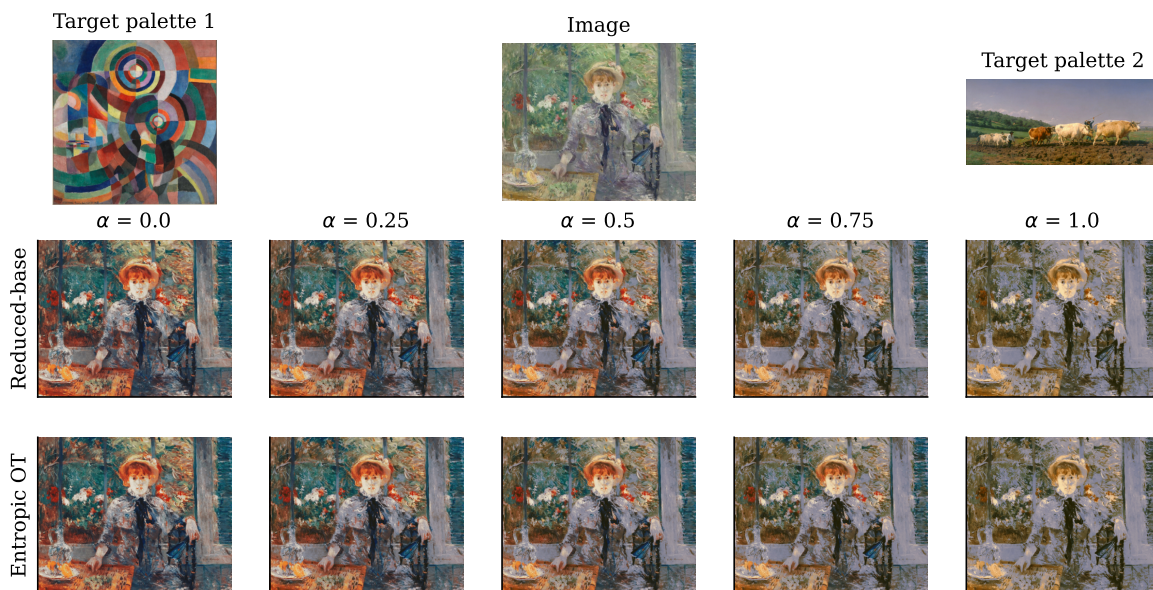


Figure 10: Color transfer obtained with a reduced-basis method (line 2) and via. These pictures were obtained with a reduced basis with only 3 snapshots. For $0 \leq \alpha \leq 1$, the target palette is computed as $(1 - \alpha)(\text{target palette 1}) + \alpha(\text{target palette 2})$.

Sources : Image (Berthe Morisot, *Après le déjeuner*, 1881), Target palette 1 (Sonia Delaunay, *Prismes électriques*, 1914), Target palette 2 (Rosa Bonheur, *Labourage nivernais*, 1849).

Acknowledgments The third author’s work benefited from the support of the FMJH Program PGMO, Université Paris-Saclay and from the ANR project GOTA (ANR-23-CE46-0001). We acknowledge the financial support of European Research Council (ERC) under the European Union’s Horizon 2020 Research and Innovation Programme – Grant Agreement n°101077204 HighLEAP.

References

- [1] Bo O Almroth, Perry Stern, and Frank A Brogan. Automatic choice of global shape functions in structural analysis. *Aiaa Journal*, 16(5):525–528, 1978.
- [2] Maxime Barrault, Yvon Maday, Ngoc Cuong Nguyen, and Anthony T Patera. An ‘empirical interpolation’ method: application to efficient reduced-basis discretization of partial differential equations. *Comptes Rendus Mathématique*, 339(9):667–672, 2004.

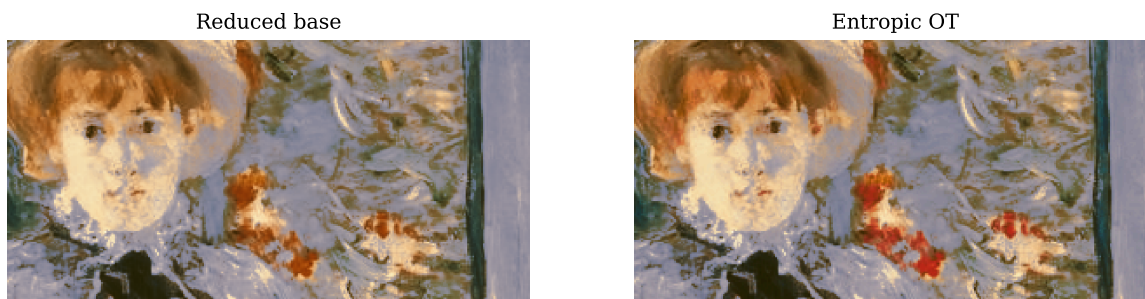


Figure 11: Detail of the resulting images for $\alpha = 0.9$.

- [3] Beatrice Battisti, Tobias Blickhan, Guillaume Enchery, Virginie Ehrlacher, Damiano Lombardi, and Olga Mula. Wasserstein model reduction approach for parametrized flow problems in porous media. *ESAIM: Proceedings and Surveys*, 73:28–47, 2023.
- [4] Jean-David Benamou and Yann Brenier. A computational fluid mechanics solution to the monge-kantorovich mass transfer problem. *Numerische Mathematik*, 84(3):375–393, 2000.
- [5] Jean-David Benamou, Guillaume Carlier, Marco Cuturi, Luca Nenna, and Gabriel Peyré. Iterative bregman projections for regularized transportation problems. *SIAM Journal on Scientific Computing*, 37(2):A1111–A1138, 2015.
- [6] Peter Binev, Albert Cohen, Wolfgang Dahmen, Ronald DeVore, Guergana Petrova, and Przemyslaw Wojtaszczyk. Convergence rates for greedy algorithms in reduced basis methods. *SIAM journal on mathematical analysis*, 43(3):1457–1472, 2011.
- [7] Tobias Blickhan. A registration method for reduced basis problems using linear optimal transport. *SIAM Journal on Scientific Computing*, 46(5):A3177–A3204, 2024.
- [8] Annalisa Buffa, Yvon Maday, Anthony T Patera, Christophe Prud’Homme, and Gabriel Turinici. A priori convergence of the greedy algorithm for the parametrized reduced basis method. *ESAIM: Mathematical modelling and numerical analysis*, 46(3):595–603, 2012.
- [9] Nicolas Cagniard, Yvon Maday, and Benjamin Stamm. Model order reduction for problems with large convection effects. In *Contributions to partial differential equations and applications*, pages 131–150. Springer, 2018.
- [10] Simona Cucchiara, Angelo Iollo, Tommaso Taddei, and Haysam Telib. Model order reduction by convex displacement interpolation. *Journal of Computational Physics*, 514:113230, 2024.
- [11] Marco Cuturi. Sinkhorn distances: Lightspeed computation of optimal transport. *Advances in neural information processing systems*, 26, 2013.
- [12] Maxime Dalery, Geneviève Dusson, Virginie Ehrlacher, and Alexei Lozinski. Nonlinear reduced basis using mixture wasserstein barycenters: application to an eigenvalue problem inspired from quantum chemistry. *Constructive Approximation*, pages 1–41, 2026.
- [13] Julie Delon, Bruno Galerne, Agnes Desolneux, Valentin De Bortoli, and Lucía Bouza. Stori, 2024. https://colab.research.google.com/github/storimaging/Notebooks/blob/main/ContrastAndColor/TP_color_transfer_with_Sinkhorn.ipynb.
- [14] Minh-Hieu Do, Jean Feydy, and Olga Mula. Sparse wasserstein barycenters and application to reduced order modeling. *Journal of Scientific Computing*, 102(3):64, 2025.
- [15] Virginie Ehrlacher, Damiano Lombardi, Olga Mula, and François-Xavier Vialard. Nonlinear model reduction on metric spaces. application to one-dimensional conservative pdes in wasserstein spaces. *ESAIM: Mathematical Modelling and Numerical Analysis*, 54(6):2159–2197, 2020.

- [16] Sira Ferradans, Nicolas Papadakis, Gabriel Peyré, and Jean-François Aujol. Regularized discrete optimal transport. *SIAM Journal on Imaging Sciences*, 7(3):1853–1882, 2014.
- [17] Jean Feydy, Thibault Séjourné, François-Xavier Vialard, Shun-ichi Amari, Alain Trounev, and Gabriel Peyré. Interpolating between optimal transport and mmd using sinkhorn divergences. In *The 22nd international conference on artificial intelligence and statistics*, pages 2681–2690. PMLR, 2019.
- [18] Alfred Galichon and Bernard Salanié. Cupid’s invisible hand: Social surplus and identification in matching models. *The Review of Economic Studies*, 89(5):2600–2629, 2022.
- [19] Nicola Gigli. On hölder continuity-in-time of the optimal transport map towards measures along a curve. *Proceedings of the Edinburgh Mathematical Society*, 54(2):401–409, 2011.
- [20] Martin A Grepl, Yvon Maday, Ngoc C Nguyen, and Anthony T Patera. Efficient reduced-basis treatment of nonaffine and nonlinear partial differential equations. *ESAIM: Modélisation mathématique et analyse numérique*, 41(3):575–605, 2007.
- [21] Martin A Grepl and Anthony T Patera. A posteriori error bounds for reduced-basis approximations of parametrized parabolic partial differential equations. *ESAIM: Modélisation mathématique et analyse numérique*, 39(1):157–181, 2005.
- [22] Bernard Haasdonk. Convergence rates of the pod–greedy method. *ESAIM: Mathematical modelling and numerical Analysis*, 47(3):859–873, 2013.
- [23] Bernard Haasdonk and Mario Ohlberger. Reduced basis method for finite volume approximations of parametrized linear evolution equations. *ESAIM: Mathematical Modelling and Numerical Analysis*, 42(2):277–302, 2008.
- [24] Jan S Hesthaven, Gianluigi Rozza, Benjamin Stamm, et al. *Certified reduced basis methods for parametrized partial differential equations*, volume 590. Springer, 2016.
- [25] Angelo Iollo and Damiano Lombardi. Advection modes by optimal mass transfer. *Physical Review E*, 89(2):022923, 2014.
- [26] Angelo Iollo and Tommaso Taddei. Mapping of coherent structures in parameterized flows by learning optimal transportation with gaussian models. *Journal of Computational Physics*, 471:111671, 2022.
- [27] Leonid V Kantorovich. On the translocation of masses. *Journal of mathematical sciences*, 133(4), 2006.
- [28] Howard Karloff. *Linear programming*. Springer Science & Business Media, 2008.
- [29] Moaad Khamlich, Federico Pichi, and Gianluigi Rozza. Optimal transport–inspired deep learning framework for slow-decaying kolmogorov-width problems: Exploiting sinkhorn loss and wasserstein kernel. *SIAM Journal on Scientific Computing*, 47(2):C235–C264, 2025.
- [30] Julien Lacombe, Julie Digne, Nicolas Courty, and Nicolas Bonneel. Learning to generate wasserstein barycenters. *Journal of Mathematical Imaging and Vision*, 65(2):354–370, 2023.
- [31] Bruno Lévy. A numerical algorithm for l_2 semi-discrete optimal transport in 3d. *ESAIM: Mathematical Modelling and Numerical Analysis*, 49(6):1693–1715, 2015.
- [32] Xi-Nan Ma, Neil S Trudinger, and Xu-Jia Wang. Regularity of potential functions of the optimal transportation problem. *Archive for rational mechanics and analysis*, 177(2):151–183, 2005.
- [33] Yvon Maday, Ngoc Cuong Nguyen, Anthony T Patera, and George SH Pau. A general, multipurpose interpolation procedure: the magic points. *Communications on pure and applied analysis*, 8(1), 2007.
- [34] Quentin Mérigot. A multiscale approach to optimal transport. 30(5):1583–1592, 2011.

- [35] Quentin Mérigot, Alex Delalande, and Frederic Chazal. Quantitative stability of optimal transport maps and linearization of the 2-wasserstein space. In *International Conference on Artificial Intelligence and Statistics*, pages 3186–3196. PMLR, 2020.
- [36] Quentin Merigot and Boris Thibert. Optimal transport: discretization and algorithms. In *Handbook of numerical analysis*, volume 22, pages 133–212. Elsevier, 2021.
- [37] Gaspard Monge. Mémoire sur la théorie des déblais et des remblais. *Mem. Math. Phys. Acad. Royale Sci.*, pages 666–704, 1781.
- [38] Ahmed K Noor and Jeanne M Peters. Reduced basis technique for nonlinear analysis of structures. *Aiaa journal*, 18(4):455–462, 1980.
- [39] Benjamin Peherstorfer. Model reduction for transport-dominated problems via online adaptive bases and adaptive sampling. *SIAM Journal on Scientific Computing*, 42(5):A2803–A2836, 2020.
- [40] Gabriel Peyré and Marco Cuturi. Computational optimal transport. *Foundations and Trends in Machine Learning*, 11(5-6):355–607, 2019.
- [41] François Pitié, Anil C Kokaram, and Rozenn Dahyot. Automated colour grading using colour distribution transfer. *Computer Vision and Image Understanding*, 107(1-2):123–137, 2007.
- [42] Christophe Prud’Homme, Dimitrios V Rovas, Karen Veroy, Luc Machiels, Yvon Maday, Anthony T Patera, and Gabriel Turinici. Reliable real-time solution of parametrized partial differential equations: Reduced-basis output bound methods. *J. Fluids Eng.*, 124(1):70–80, 2002.
- [43] Alfio Quarteroni, Andrea Manzoni, and Federico Negri. *Reduced basis methods for partial differential equations: an introduction*. Springer, 2015.
- [44] Julien Rabin, Julie Delon, and Yann Gousseau. Regularization of transportation maps for color and contrast transfer. *2010 IEEE International Conference on Image Processing*, pages 1933–1936, 2010.
- [45] Gianluigi Rozza, Dinh Bao Phuong Huynh, and Anthony T Patera. Reduced basis approximation and a posteriori error estimation for affinely parametrized elliptic coercive partial differential equations: application to transport and continuum mechanics. *Archives of computational methods in engineering*, 15(3):229–275, 2008.
- [46] Filippo Santambrogio. *Optimal transport for applied mathematicians*. Springer, 2015.
- [47] Tommaso Taddei. A registration method for model order reduction: data compression and geometry reduction. *SIAM Journal on Scientific Computing*, 42(2):A997–A1027, 2020.
- [48] Cédric Villani. *Topics in optimal transportation*, volume 58. American Mathematical Soc., 2021.
- [49] Cédric Villani et al. *Optimal transport: old and new*, volume 338. Springer, 2009.
- [50] Stefan Volkwein. Proper orthogonal decomposition: Theory and reduced-order modelling. *Lecture Notes, University of Konstanz*, 4(4):1–29, 2013.