ARTICLE

The algebraic degree of the Wasserstein distance

Chiara Meroni[1] · Bernhard Reinke[2,4] · Kexin Wang[3]

Abstract
Given two rational univariate polynomials, the Wasserstein distance of their associated measures is an algebraic number. We determine the algebraic degree of the squared Wasserstein distance, serving as a measure of algebraic complexity of the corresponding optimization problem. The computation relies on Galois theory and on the combinatorial structure of a specific subpolytope of the Birkhoff polytope, invariant under a transformation induced by complex conjugation. Proofs and computations also draw on notions from graph theory and elimination ideals.

Keywords Wasserstein distance · algebraic degree · conjugation invariant Birkhoff polytope · minimal polynomial · bipartite graphs · Galois theory

Mathematics Subject Classification 52B12 · 49Q22 · 12F10 · 68W30 · 14Q20

1 Introduction

Given two univariate polynomials $p, q \in \mathbb{Q}[z]$ of degree d with simple roots, consider the optimization problem

$$\text{minimize } \frac{1}{d} \sum_{1 \le i \le d} \|\alpha_i - \beta_{\sigma(i)}\|^2 \quad \text{over } \sigma \in \text{Sym}(d), \tag{1}$$

where $\alpha_1, \ldots, \alpha_d$ and $\beta_1, \ldots, \beta_d$ are the roots of p and q respectively. Here $\|\cdot\|$ denotes the Euclidean norm and $\text{Sym}(d)$ is the set of permutations of d elements. The optimal value of (1), denoted by W_2^2, is the square of the Wasserstein distance of the measures associated to p and

✉ Kexin Wang
  kexin_wang@g.harvard.edu

  Chiara Meroni
  chiara.meroni@eth-its.ethz.ch

  Bernhard Reinke
  bernhard.reinke@mis.mpg.de; bernhard.reinke@univ-amu.fr

1 ETH Institute for Theoretical Studies, ETH Zurich, Scheuchzerstrasse 70, Building SEW, 8006 Zürich, Switzerland

2 Max Planck Institute for Mathematics in the Sciences, Inselstrasse 22, 04103 Leipzig, Germany

3 Harvard University, 29 Oxford Street, Cambridge, MA 02138, USA

4 Aix Marseille Univ, CNRS, Marseille I2M, France

q. It follows from the definition that this minimum is an algebraic number, namely it is a root of a univariate polynomial with rational coefficients. There exists a unique such polynomial that is monic and of minimal degree; we denote the minimal degree by WDdegree(p, q). We will refer to WDdegree(p, q) as the *Wasserstein distance degree* of p and q. The goal of this paper is to provide a formula for WDdegree(p, q), which is an index of the complexity of the optimization problem (1), relevant in optimal transport. In the context of computer algebra in fact we cannot solve exactly such a minimization, but we can encode it in a polynomial that on the other hand can be computed exactly. Moreover, the exact computation might contain more information about the numeric solution to the problem or its geometry. We propose this algebraic approach to an intrinsically metric problem in the spirit of *metric algebraic geometry* [4], confident that the interplay between the different points of view will bring new insights for both communities.

The Wasserstein distance is a notion of distance between probability distributions on a metric space. It can be thought of as the cost of turning one distribution into another, and it is therefore intimately connected to optimal transport problems [23, Chapter 7]. The optimization problem (1) is a special case of the assignment problem [20, Chapter 17], a well-studied problem in combinatorial optimization.

Adopting a broader perspective, we can interpret (1) as a polynomial optimization problem where we minimize a polynomial with rational coefficients in $2d$ variables under $2d$ constraints:

$$\text{minimize } f(x_1, \ldots, x_d, y_1, \ldots, y_d) \text{ subject to } f_k(x_1, \ldots, x_d) = 0, \ g_k(y_1, \ldots, y_d) = 0,$$

where $f(x_1, \ldots, x_d, y_1, \ldots, y_d) = \frac{1}{d} \sum_i \|x_i - y_i\|^2$ and f_k, g_k are Vieta's formulae [22] for $k = 1, \ldots, d$, namely the equations of degree k that relate the roots of p, q to their coefficients. It is of general interest to compute the algebraic degree of the optimal solution of a polynomial optimization problem. Indeed, this degree encodes the *algebraic complexity* required to write the optimum *exactly* in terms of the input. This concept has been investigated in the literature in several contexts, e.g., semidefinite programming [7, 8, 16, 18]. Alternatively, when the objective function is the squared Euclidean distance [6, 17], the algebraic degree of the minimum (ED degree) measures the complexity of optimizing the Euclidean distance over an algebraic variety, problem arising in many applications. Of similar fashion is the maximum likelihood degree, central in algebraic statistics [9, 10, 21].

Algebraic techniques allow us to compute the algebraic degree of the optimization problem when the constraints are generic, see [15]. However, in many relevant scenarios (e.g., those mentioned above), the imposed conditions are not generic enough. Then, [15, Theorem 2.2] provides an upper bound for the algebraic degree. Our case fits in this family of nongeneric problems, and we can conclude that WDdegree(p, q) $\leq (d!)^2$ for any p, q as in (1). As anticipated, this bound is not tight, and we will refine it in what follows.

We begin with the setup in Sect. 2 where we define the Wasserstein distance for measures, and then interpret it in terms of associated polynomials, along the lines of [1]. In Sect. 3, we introduce a transformation of the Birkhoff polytope induced by complex conjugation and study its invariant subpolytope (Definition 3.1). We give a graph interpretation of its vertices and characterize them in Theorem 3.7 and Proposition 3.11. Exploiting the combinatorial information together with standard Galois theory, we obtain the formulae for the minimal polynomial, or in degenerate cases for a multiple of it, and the algebraic degree of the squared Wasserstein distance in Theorems 4.5 and 4.6. In Sect. 4.2, we interpret the computation of the minimal polynomial via an elimination of ideal and describe the algorithm to compute it. The code is available at https://mathrepo.mis.mpg.de/WassersteinDegree/.

We conclude the introduction with an example.

Example 1.1 Consider the univariate cubic polynomials $p(z) = 2z^3 + z^2 - 3z - 7$ and $q(z) = 3z^3 - z^2 - 5z + 4$. They both have one real root, α_1 and β_1 respectively, and a pair of complex conjugate roots, α_2, α_3 and β_2, β_3 respectively. Then,

$$W_2^2(p, q) = \min_{\sigma \in \mathrm{Sym}(3)} \frac{1}{3} \sum_{1 \le i \le 3} \|\alpha_i - \beta_{\sigma(i)}\|^2$$

$$= \|\alpha_1 - \beta_2\|^2 + \|\alpha_2 - \beta_1\|^2 + \|\alpha_3 - \beta_3\|^2 = 2.020013\dots$$

is the squared Wasserstein distance of p, q. This real number is algebraic: indeed, it is a root of the following univariate monic irreducible polynomial with rational coefficients:

$$
\begin{aligned}
z^{18} &- \frac{25}{18}z^{17} - \frac{6611}{432}z^{16} + \frac{9658735}{314928}z^{15} + \frac{860802973}{5038848}z^{14} - \frac{656593649875}{816293376}z^{13} + \frac{129841846201105}{1586874322944}z^{12} \\
&+ \frac{28152774676858415}{3570467226624}z^{11} - \frac{46036000403242535}{3570467226624}z^{10} - \frac{64389115881803301815}{1156831381426176}z^9 \\
&+ \frac{98713596333179561 4481}{4627325525704704}z^8 - \frac{18852321680214198665 5495}{749626735164162048}z^7 + \frac{570696146952180632862839}{26986562465909833728}z^6 \\
&+ \frac{2978098003861411543837955}{40479843698864750592}z^5 + \frac{67477581240353261955871 9717}{364318593289782755328}z^4 \\
&- \frac{12067578021592857861973473 5}{39346408075296537575424}z^3 - \frac{118527995532692096431865660 6075}{35411767267768838178816}z^2 \\
&+ \frac{1848261661275845551117166614375}{3187059054099019543609344}z + \frac{18255232646137865373249774012390625}{18357460151610352571189 82144}.
\end{aligned}
\tag{2}
$$

Hence, the Wasserstein distance degree of p, q is 18. We will discuss in the rest of the paper how to obtain this minimal polynomial.

2 Finitely supported probability measures

On a metric space, the distance between two probability measures can be quantified using the notion of Wasserstein distance. In [1], it has been used to introduce a new metric on the space of hypersurfaces of degree d in $\mathbb{CP}^n$. We will recall the definition of Wasserstein distance for finitely supported measures and relate the Wasserstein distance of univariate polynomials with an optimization problem on the Birkhoff polytope. For the general theory, we refer to [23]. In this section, we consider univariate polynomials with coefficients in $\mathbb{C}$. We associate probability measures to finitely supported probability measures induced by polynomials in the following way.

Definition 2.1 Let $p \in \mathbb{C}[z]$ be a polynomial of degree d. Its associated measure $\mu(p)$ is

$$\mu(p) = \frac{1}{d} \sum_{p(x)=0} m(p, x)\delta_x,$$

where $m(p, x)$ is the multiplicity of p at $x \in \mathbb{C}$.

More concretely, if $p(z)$ factors as $\prod_{1 \le i \le d}(z - \alpha_i)$, then $\mu(p) = \frac{1}{d}\sum_{1 \le i \le d} \delta_{\alpha_i}$. In particular, if p has no multiple roots, then $\mu(p)$ is the uniform counting measure on the roots of p. We will be mostly interested in the case where p has no multiple roots.

Definition 2.2 Let μ_1, μ_2 be finitely supported probability measures on $\mathbb{C}$. A *transport plan* from μ_1 to μ_2 is a probability measure λ on $\mathbb{C} \times \mathbb{C}$ such that $\pi_{1*}\lambda = \mu_1$ and $\pi_{2*}\lambda = \mu_2$, where the pushforwards $\pi_{i*}\lambda$ are measures on $\mathbb{C}$ defined via

$$\pi_{1*}(\lambda)[A] = \lambda[A \times \mathbb{C}], \qquad \pi_{2*}(\lambda)[B] = \lambda[\mathbb{C} \times B]$$

for any $A, B \subseteq \mathbb{C}$. We write $\Pi(\mu_1, \mu_2)$ for the set of transport plans between μ_1 and μ_2. The *cost* of a transport plan is defined to be

$$\text{cost}(\lambda) = \int \|x - y\|^2 d\lambda(x, y).$$

The squared *Wasserstein distance* between μ_1, μ_2 is the minimal cost of a transport plan:

$$W_2^2(\mu_1, \mu_2) = \min_{\lambda \in \Pi(\mu_1, \mu_2)} \text{cost}(\lambda).$$

Remark 2.3 If μ_1, μ_2 are finitely supported probability measures, then every transport plan from μ_1 to μ_2 will have support contained in $\text{supp}\, \mu_1 \times \text{supp}\, \mu_2$. In particular, they will also have finite support. In this setting we do not have to deal with most measure theoretic issues. See for example [23, Chapter 7] for the general definition.

We continue Example 1.1 in order to familiarize with the definition.

Example 2.4 As in Example 1.1, consider the univariate polynomials $p(z) = 2z^3 + z^2 - 3z - 7$ and $q(z) = 3z^3 - z^2 - 5z + 4$ and denote by $\alpha_1, \alpha_2, \alpha_3$ and $\beta_1, \beta_2, \beta_3$ their roots, respectively. According to Definition 2.2, the squared Wasserstein distance of p and q is the minimum of $\int \|x - y\|^2 d\lambda(x, y)$, where λ is any probability measure on $\mathbb{C} \times \mathbb{C}$ such that

$$\lambda[A \times \mathbb{C}] = \left(\frac{1}{3}(\delta_{\alpha_1} + \delta_{\alpha_2} + \delta_{\alpha_3}) \right)[A] = \begin{cases} 0 & \text{if } \alpha_i, \alpha_j, \alpha_k \notin A, \{i, j, k\} = \{1, 2, 3\} \\ \frac{1}{3} & \text{if } \alpha_i \in A, \alpha_j, \alpha_k \notin A, \{i, j, k\} = \{1, 2, 3\} \\ \frac{2}{3} & \text{if } \alpha_i, \alpha_j \in A, \alpha_k \notin A, \{i, j, k\} = \{1, 2, 3\} \\ 1 & \text{if } \alpha_i, \alpha_j, \alpha_k \in A, \{i, j, k\} = \{1, 2, 3\} \end{cases}$$

$$\lambda[\mathbb{C} \times B] = \left(\frac{1}{3}(\delta_{\beta_1} + \delta_{\beta_2} + \delta_{\beta_3}) \right)[B] = \begin{cases} 0 & \text{if } \beta_i, \beta_j, \beta_k \notin B, \{i, j, k\} = \{1, 2, 3\} \\ \frac{1}{3} & \text{if } \beta_i \in B, \beta_j, \beta_k \notin B, \{i, j, k\} = \{1, 2, 3\} \\ \frac{2}{3} & \text{if } \beta_i, \beta_j \in B, \beta_k \notin B, \{i, j, k\} = \{1, 2, 3\} \\ 1 & \text{if } \beta_i, \beta_j, \beta_k \in B, \{i, j, k\} = \{1, 2, 3\} \end{cases}$$

$$(3)$$

for any $A, B \subset \mathbb{C}$.

For measures associated to polynomials, it will be convenient to work with doubly stochastic matrices, namely those square matrices with nonnegative real entries, each of whose rows and columns add up to 1. For $d \in \mathbb{N}$, we let B_d denote the d-th Birkhoff polytope, i.e., the set of all doubly stochastic matrices of size $d \times d$. It is a classical result due to Birkhoff [3] and von Neumann [24] that this set is a convex polytope, with vertices the permutation matrices of size $d \times d$. For more details on the Birkhoff polytope and its properties, we refer the reader to [25].

The following result highlights the fact that in the context of measures associated to univariate polynomials, each transport plan corresponds to a doubly stochastic matrix. Therefore, instead of working with probability measures on $\mathbb{C} \times \mathbb{C}$, we can focus on Birkhoff polytopes.

Lemma 2.5 *Let $p, q \in \mathbb{C}[z]$ be polynomials of degree d with roots $\alpha_1, \ldots, \alpha_d$ and $\beta_1, \ldots, \beta_d$ respectively. Suppose that p and q have no multiple roots. Then the following is a bijection:*

$$\lambda : B_d \to \Pi(\mu(p), \mu(q))$$

$$M \mapsto \lambda(M) := \frac{1}{d} \sum_{1 \leq i, j \leq d} M_{ij} \delta_{\alpha_i, \beta_j}.$$

Proof Since we assume p and q have no multiple roots, it follows that the map

$$M \mapsto \frac{1}{d} \sum_{1 \leq i, j \leq d} M_{ij} \delta_{\alpha_i, \beta_j}$$

defines a bijection between $\mathbb{R}_{\geq 0}^{d \times d}$ and the set of *finite* measures supported on the points $\{\alpha_1, \ldots, \alpha_d\} \times \{\beta_1, \ldots, \beta_d\}$. For $M \in \mathbb{R}_{\geq 0}^{d \times d}$ and $1 \leq i \leq d$, we have

$$\pi_{1*}(\lambda(M))[\{\alpha_i\}] = (\lambda(M))[\{\alpha_i\} \times \mathbb{C}] = \sum_{1 \leq j \leq d} (\lambda(M))[\{(\alpha_i, \beta_j)\}] = \frac{1}{d} \sum_{1 \leq j \leq d} M_{ij}.$$

From this we see that $\pi_{1*}(\lambda(M)) = \mu(p)$ if and only if all row sums of M are equal to 1. Similarly $\pi_{2*}(\lambda(M)) = \mu(q)$ if and only if all column sums are equal to 1. So $\lambda(M)$ is a transport plan if and only if $M \in B_d$. □

Remark 2.6 If p and q are allowed to have multiple roots, then $\Pi(\mu(p), \mu(q))$ is given by a transportation polytope (see [11] for an overview). However, the map defined in Lemma 2.5 is still a surjection if we list the roots with multiplicity.

We use Lemma 2.5 in our consideration of the Wasserstein distance for polynomials. From now on, we consider univariate polynomials $p, q \in \mathbb{C}[z]$ with factorizations $p(z) = \prod_{1 \leq i \leq d}(z - \alpha_i)$, $q(z) = \prod_{1 \leq i \leq d}(z - \beta_i)$, without multiple roots. For $M \in B_d$, we write

$$\text{cost}(M, p, q) = \text{cost}(\lambda(M)) = \frac{1}{d} \sum_{1 \leq i, j \leq d} M_{ij} \|\alpha_i - \beta_j\|^2. \tag{4}$$

By these considerations and using the bijection in Lemma 2.5 we interpret the Wasserstein distance between two polynomials as an optimization problem:

$$\text{W}_2^2(p, q) := \text{W}_2^2(\mu(p), \mu(q)) = \min_{M \in B_d} \text{cost}(M, p, q).$$

We highlight the role of the lemma in our running example.

Example 2.7 Assume the setup from Example 1.1 and Example 2.4. Instead of expressing the cost function as an integral, we can turn it into a finite sum via Lemma 2.5. Indeed, a measure λ on $\mathbb{C} \times \mathbb{C}$ that satisfies (3) must be supported on the product $\{\alpha_1, \alpha_2, \alpha_3\} \times \{\beta_1, \beta_2, \beta_3\}$. Therefore, the integral in Example 2.4 becomes

$$\frac{1}{3} \sum_{1 \leq i, j \leq 3} M_{i,j} \|\alpha_i - \beta_j\|^2,$$

where the coefficients $M_{i,j}$ encode the mass of the measure λ at the point (α_i, β_j). If the $M_{i,j}$ are interpreted as entries of a matrix M, the latter must have row sums and column sums equal to 1 in order to satisfy the conditions in (3). Equivalently, M must be a doubly stochastic matrix, hence a point of the third Birkhoff polytope B_3, which is the convex hull of the $3! = 6$ permutation matrices of size 3×3. This gives rise to a polytope in $\mathbb{R}^9$ of dimension 4. The affine hull of B_3 is in fact cut out by the linear conditions that row sums and column sums are equal to 1.

3 Invariant Birkhoff polytope

We now restrict to polynomials with real coefficients and study the effects of complex conjugation, denoted as usual by $\bar{z}$. If $p, q \in \mathbb{R}[z]$, then complex conjugation will induce an involution on the transport plans form $\mu(p)$ to $\mu(q)$. We compute this involution for the Birkhoff polytope under the identification of Lemma 2.5.

Assume $p, q \in \mathbb{R}[z]$ with roots α_i and β_i as above. We denote by ϕ and ψ the permutations that swap complex conjugate pairs of roots of p and q respectively, namely,

$$\begin{aligned} \alpha_{\phi(i)} &= \overline{\alpha_i}, \\ \beta_{\psi(i)} &= \overline{\beta_i}. \end{aligned} \tag{5}$$

We will call P_ϕ, P_ψ the permutation matrices associated to ϕ and ψ respectively. Then the pair (ϕ, ψ) defines a map $\iota : B_d \to B_d$ via

$$(\iota M)_{ij} = M_{\phi(i)\psi(j)} = (P_\phi M P_\psi)_{ij}.$$

We can see that ι is a linear isometry of the Birkhoff polytope. In fact, every combinatorial symmetry of B_d is of the form $M \mapsto P_g M P_h$ or $M \mapsto P_g M^\top P_h$ for some permutations $g, h \in \mathrm{Sym}(d)$, see [2]. In particular, ι defines an involution on the Birkhoff polytope. This suggests the following definition.

Definition 3.1 Let $p, q \in \mathbb{R}[z]$ be polynomials of degree d with distinct roots and take $\phi, \psi \in \mathrm{Sym}(d)$ to be permutations that satisfy (5). Consider the operation $\iota = (\phi, \psi)$. The associated *ι-invariant Birkhoff polytope*, denoted B_d^ι, is the set of all doubly stochastic matrices that are invariant under ι.

Observe that the cost function does not change under ι, namely

$$\mathrm{cost}(M, p, q) = \mathrm{cost}(\iota M, p, q) = \mathrm{cost}\left(\frac{M + \iota M}{2}, p, q\right).$$

Therefore, instead of solving an optimization problem over the Birkhoff polytope, we can find the squared Wasserstein distance using the ι-invariant Birkhoff polytope:

$$\mathrm{W}_2^2(p, q) = \frac{1}{d} \min_{M \in B_d^\iota} \mathrm{cost}(M, p, q).$$

The cost function remains linear for $M \in B_d^\iota$, hence the minimum is attained at a vertex of B_d^ι. The ι-invariant Birkhoff polytope can reduce the complexity of the optimization problem, since it is potentially lower dimensional. We give the formula for the dimension of B_d^ι in terms of the number of real and complex roots of the polynomials p, q.

For completeness, we compute the dimension of the ι-invariant Birkhoff polytopes. Let us introduce some notation. We denote the number of real roots of a polynomial $p \in \mathbb{R}[z]$ by $p_\mathbb{R}$, and the number of *pairs* of its complex conjugate roots by $p_\mathbb{C}$.

Lemma 3.2 *Suppose $\iota = (\phi, \psi)$ is the involution associated with p, q. Then, the dimension of the ι-invariant Birkhoff polytope B_d^ι is*

$$\dim B_d^\iota = (p_\mathbb{R} + p_\mathbb{C})(q_\mathbb{R} + q_\mathbb{C}) - (p_\mathbb{R} + p_\mathbb{C} + q_\mathbb{R} + q_\mathbb{C} - 1) + p_\mathbb{C} q_\mathbb{C}.$$

Proof By definition, the affine span of B_d^ι is

$$V = \left\{ M \in \mathbb{R}^{d \times d} \mid M_{ij} = M_{\phi(i)\psi(j)}, \forall i, j \in [d], \sum_{1 \le i \le d} M_{ij} = 1, \right.$$

$$\forall j \in [d], \sum_{1 \le j \le d}^{d} M_{ij} = 1, \forall i \in [d] \Bigg\},$$

where $[d] = \{1, \ldots, d\}$. The first family of conditions implies that there are at most

$$(p_\mathbb{R} + 2p_\mathbb{C})(q_\mathbb{R} + 2q_\mathbb{C}) - p_\mathbb{C}q_\mathbb{R} - p_\mathbb{R}q_\mathbb{C} - p_\mathbb{C}q_\mathbb{C} = (p_\mathbb{R} + p_\mathbb{C})(q_\mathbb{R} + q_\mathbb{C}) + p_\mathbb{C}q_\mathbb{C}$$

free parameters for the points in V, since ϕ and ψ switch complex conjugate pairs. The second and third type of conditions in the definition of V impose that each column and each row of M sums to 1. Since $\sum_{1 \le j \le d} M_{ij} = \sum_{1 \le j \le d} M_{\phi(i)j}$ for $i \in [d]$, and $\sum_{1 \le i \le d} M_{ij} = \sum_{1 \le i \le d} M_{i\psi(j)}$ for $j \in [d]$, there are exactly $m = p_\mathbb{R} + p_\mathbb{C} + q_\mathbb{R} + q_\mathbb{C}$ distinct linear equations among row-sums and column-sums. However, only $p_\mathbb{R} + p_\mathbb{C} + q_\mathbb{R} + q_\mathbb{C} - 1$ are linearly independent. Indeed, consider the distinct relations $\ell_{i_1}, \ldots, \ell_{i_{p_\mathbb{R}+p_\mathbb{C}}}, \ell_{j_1}, \ldots, \ell_{j_{q_\mathbb{R}+q_\mathbb{C}}}$, where the i indices correspond to row-sums and the j indices to column-sums. The k-th summand in $\ell_{j_{q_\mathbb{R}+q_\mathbb{C}}}$ can be expressed in terms of ℓ_{i_k}, for all k. On the other hand, the equation

$$\sum_{1 \le k \le p_\mathbb{R}+p_\mathbb{C}} \eta_k \ell_{i_k} + \sum_{1 \le k \le q_\mathbb{R}+q_\mathbb{C}-1} \mu_k \ell_{j_k} = 0$$

must be trivial. To see this, notice that $\eta = 0$ for every k, since each of them is the only coefficient of $M_{i_k, j_{q_\mathbb{R}+q_\mathbb{C}}}$ in the sum. Then we are left only with column-sums, and these are trivially independent since each of them involves some term that does not appear in the others. So $\mu_k = 0$ too, and we conclude that

$$\dim B_d^\iota = \dim V = (p_\mathbb{R} + p_\mathbb{C})(q_\mathbb{R} + q_\mathbb{C}) + p_\mathbb{C}q_\mathbb{C} - (p_\mathbb{R} + p_\mathbb{C} + q_\mathbb{R} + q_\mathbb{C} - 1).$$

$\square$

In the following subsections, we will focus on the vertices of this polytope. In order to get a better understanding, we will introduce a graph associated with the vertices of B_d^ι. To conclude this part, we construct some examples of ι-invariant Birkhoff polytopes, which we will review later using the graph interpretation. The code used to compute these examples is available at https://mathrepo.mis.mpg.de/WassersteinDegree/Birkhoff.

Example 3.3 Let $d = 2$, then p, q are quadratic polynomials with real coefficients. The Birkhoff polytope B_2 is a segment in $\mathbb{R}^4$. There are three possible cases for the structure of the roots of p, q: both polynomials have both roots real; one has both roots real and one has both roots complex; both polynomials have both roots complex. In the first case, we have identity matrices $P_\phi = P_\psi = \mathbb{1}_2$, hence $B_2^\iota = B_2$. In the second case, we assume that p has complex roots and q real roots, then $P_\phi = \left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right)$, $P_\psi = \mathbb{1}_2$, and the polytope becomes a point: $B_2^\iota = \left\{ \frac{1}{2} \left(\begin{smallmatrix} 1 & 1 \\ 1 & 1 \end{smallmatrix}\right) \right\}$. Finally, in the case where the roots of both polynomials are complex, $P_\phi = P_\psi = \left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right)$ hence ι fixes the two vertices of the Birkhoff polytope, so $B_2^\iota = B_2$.

Example 3.4 Let $d = 3$, then p, q are cubic polynomials with real coefficients. The polytope B_3 has dimension 4 in $\mathbb{R}^9$. There are again three cases of roots configurations for p, q:

1. p, q have all roots real, then $P_\phi = P_\psi = \mathbb{1}_3$, so $B_3^\iota = B_3$.
2. q has all roots real and p has exactly one real root. We will assume that the real root of p is α_1. Then $P_\phi = \left(\begin{smallmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{smallmatrix}\right)$, $P_\psi = \mathbb{1}_3$. The 6 vertices of the Birkhoff polytope collapse to form a triangle

$$B_3^\iota = \mathrm{conv}\left\{ \begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & \frac{1}{2} & \frac{1}{2} \end{pmatrix}, \begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{2} \\ 1 & 0 & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} \end{pmatrix}, \begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & \frac{1}{2} & \frac{1}{2} \\ 1 & 0 & 0 \end{pmatrix} \right\}.$$

Table 1 Dimensions of the invariant Birkhoff polytopes for $d = 4$

$(p_{\mathbb{R}}, p_{\mathbb{C}})$	$(q_{\mathbb{R}}, q_{\mathbb{C}})$	dim B_d^ι	$| \text{vertices}(B_d^\iota) |$
(4,0)	(4,0)	9	24
(4,0)	(2,1)	6	12
(4,0)	(0,2)	3	6
(2,1)	(2,1)	5	13
(2,1)	(0,2)	4	12
(0,2)	(0,2)	5	8

3. p, q both have exactly one real root. We will assume that these are α_1, β_1. Then $P_\phi = P_\psi = \left(\begin{smallmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{smallmatrix} \right)$. The associated invariant Birkhoff polytope is now the square

$$B_3^\iota = \text{conv} \left\{ \mathbb{1}_3, \left(\begin{smallmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{smallmatrix} \right), \left(\begin{smallmatrix} 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} \end{smallmatrix} \right), \left(\begin{smallmatrix} 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & 0 \end{smallmatrix} \right) \right\}.$$

Example 3.5 Let $d = 4$, then p, q are quartic polynomials with real coefficients. The Birkhoff polytope B_4 has dimension 9 in $\mathbb{R}^{16}$. Assume that p has only real roots, and q has only complex roots. Then, as predicted in Lemma 3.2, the invariant Birkhoff polytope is a 3-dimensional octahedron. This and all the other cases are listed in Table 1.

3.1 Graph interpretation

The aim of this subsection is to introduce a graph interpretation of the vertices of B_d^ι. This interpretation will be used to realize all the vertices of B_d^ι as minimizer of our optimization problem (1) and we shall see that the graph automorphisms are closely related to the Wasserstein distance degree.

Recall that the ι-invariant Birkhoff polytope B_d^ι is the image of the standard Birkhoff polytope via the map $v \mapsto \frac{v + \iota(v)}{2}$. Any vertex of the ι-invariant Birkhoff polytope B_d^ι is necessarily the image of a vertex of the Birkhoff polytope. Otherwise, by Krein–Milman theorem [12], it would be possible to express it as the convex combination of other points in B_d^ι. Since the vertices of the classical Birkhoff polytope are permutation matrices, a vertex $M \in \mathbb{R}^{d \times d}$ of B_d^ι satisfies $M_{ij} \in \{0, \frac{1}{2}, 1\}$. Hence, the matrix $2M$ has integer entries.

We associate a bipartite graph Γ with d nodes on the left side (L) and d nodes on the right side (R), such that $2M$ is its adjacency matrix. Namely, the vertices of the graph, denoted $V(\Gamma) = [d] \times \{L, R\}$, satisfy the property that (i, L) and (j, R) are connected by $2M_{ij}$ edges. Such an edge will be denoted by (i, j), where the first entry indicates a vertex (i, L) and the second entry indicates a vertex (j, R). Since M is a doubly stochastic matrix, every vertex in Γ has degree 2. Γ is then a disjoint union of cycles, and because it is bipartite, each cycle has even length.

Example 3.6 We continue here our running example, and refer to Example 1.1 and Fig. 1 for the setting. From the figure, we construct the permutation matrix P associated to p, q, with $P_{i,j} = 1$ exactly when α_i and β_j are connected. Then,
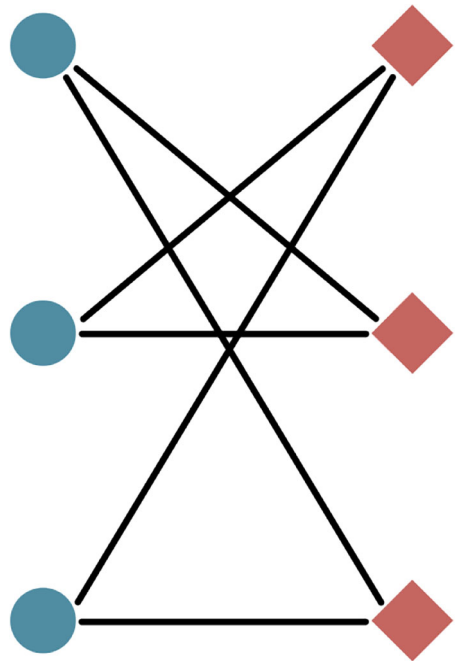
$$M = \frac{P + \iota(P)}{2} = \frac{1}{2} \left(\begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} + \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix} \right).$$

Fig. 1 The roots of p (blue dots) and the roots of q (red diamonds), from Example 1.1. The segments joining them represent a pairing that realizes the minimum for the Wasserstein distance of p, q
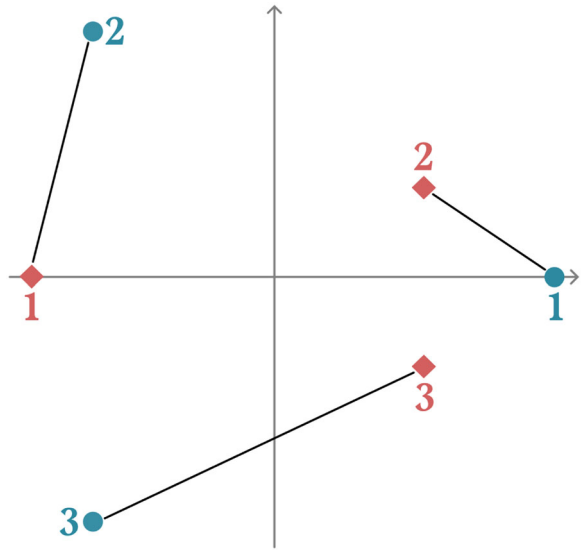
Fig. 2 Graph Γ associated to the roots configuration in Example 3.6. The setting is also displayed in Fig. 1

We associate to p, q a graph as follows. Consider a bipartite graph Γ with vertices that represent the roots of p on one side and the roots of q on the other side, and adjeciency matrix $2M$. In our specific case, we obtain the graph Γ from Fig. 2. The bipartite graph has 6 edges which form a unique cycle.

The invariance of M under ι gives a graph automorphism that we will by abuse of notation also call ι. Explicitly, $\iota(i, L) = (\phi(i), L)$ and $\iota(j, R) = (\psi(j), R)$. Note that ι is also an involution and it acts independently on both sides of the bipartite graph Γ.

Let us introduce some types of cycles in Γ, governed by the involution:

1. 2-cycles between fixed points of ι;
2. pairs of 2-cycles that are interchanged by ι;
3. cycles of length $2(2k + 1)$, $k > 0$, including exactly one fixed point of ι on each side;
4. cycles of length $4k$, $k > 0$, including exactly two fixed points of ι on one side and none on the other.

The following result states that these are the only possible disjoint cycles in our graphs.

Theorem 3.7 *Let M be a vertex of B_d^ι and Γ be the associated graph. Then, the disjoint cycles in Γ are of type 1–4. Conversely, let Γ be a bipartite graph with d vertices on each side, such that all its vertices have degree 2. If all the disjoint cycles of Γ are of type 1–4, then the graph corresponds to a vertex $M \in B_d^\iota$.*

Proof Suppose that M is a vertex of B_d^ι corresponding to the graph Γ. Since M is the image of a permutation matrix via $\frac{\cdot + \iota(\cdot)}{2}$, the entry of M corresponding to a pair of fixed points of ι is either 0 or 1. Hence, if there is an edge in Γ between two fixed points of ι, then there must be two such edges. The latter case describes cycles C in Γ of type 1.

If a cycle C contains no fixed point of ι, then C must have length 2. Assume for contradiction that C has length $2k > 2$. We label its edges in cyclic order by $e_1, \ldots, e_{2k}$. Recall that a perfect matching of a graph is a subset $\mathcal{M}$ of edges such that every vertex is adjacent to exactly one edge in $\mathcal{M}$. The edge subsets $\{e_1, e_3, \ldots, e_{2k-1}\}$, $\{e_2, e_4, \ldots, e_{2k}\}$ (and those for $\iota(C)$ if C is not set-wise fixed) into two distinct ι-invariant perfect matchings $\mathcal{M}_1, \mathcal{M}_2$. We define a matrix M_1 with $M_1(i, j) = M(i, j)$ if $(i, j) \notin C$, $M_1(i, j) = 2M(i, j)$ if $(i, j) \in \mathcal{M}_1$, and $M_1(i, j) = 0$ if $(i, j) \in \mathcal{M}_2$, similarly M_2 for the perfect matching $\mathcal{M}_2$. The matrices M_1, M_2 are in B_d^ι since the perfect matchings are ι-invariant. Hence, $2M = M_1 + M_2$, contradicting the vertex assumption. Such a cycle of length 2 must therefore be (one cycle of the pair) of type 2.

Next, suppose that C is a cycle of length at least 4. Then, it must contain a fixed point of ι and hence C is set-wise fixed by ι. Since ι is an involution on C, its action on C can be identified with an involution in the appropriate dihedral group, namely the graph automorphisms of C. Using the classification of such involutions, we get that ι corresponds to either a reflection, and therefore C contains exactly 2 fixed points, or to the identity, hence all the points of C are fixed. We already observed that between two fixed points there cannot be simple edges, so the second option leads to a contradiction. Hence, C must have exactly 2 fixed points and they can either lie on the same side (cycle of type 4) or one on each side (cycle of type 3).

Conversely, let M be the matrix corresponding to the bipartite graph Γ with cycles of types 1–4. Then, M is a doubly stochastic matrix and it is invariant under ι so $M \in B_d^\iota$. We prove that M is a vertex. Assume that $M = \frac{N+L}{2}$ for some $N, L \in B_d^\iota$. Note that $N_{ij} = L_{ij} = 0$ whenever $M_{ij} = 0$, and $N_{ij} = L_{ij} = 1$ whenever $M_{ij} = 1$. This is the case for (i, j) with the edge between (i, L), (j, R) in a cycle of type 1 or 2. Let C in Γ be a cycle of type 3 or 4. Let $(i_1, j_1), (i_2, j_2), \ldots, (i_\ell, j_\ell)$ be the indices of entries of M corresponding to edges in C. Without loss of generality, assume $i_1 = i_2 = \phi(i_1)$ and $j_1 = \psi(j_2)$. The only two positions in the i_1 row filled with nonzero entries in N or L are (i_1, j_1) and (i_1, j_2). Let $L_{i_1, j_1} = \epsilon$, then $L_{i_1, j_2} = 1 - \epsilon$. Since L is invariant under ι, we must have $\epsilon = 1 - \epsilon = \frac{1}{2}$. Similarly,

$N_{i_1,j_1} = N_{i_1,j_2} = \frac{1}{2}$. Since the row and column sums of M, N, L are 1 and $N_{ij} = L_{ij} = 0$ whenever $M_{ij} = 0$, we have $N_{i_k,j_k} = L_{i_k,j_k} = M_{i_k,j_k} = \frac{1}{2}$ for all $k = 1, \ldots, \ell$. Hence, $L = N = M$ and M is a vertex. $\qquad\square$

We are interested in the cardinality of the group automorphisms of the graph Γ that commute with ι and preserve both sides. Such number relates to the algebraic degree of the Wasserstein distance, as it will be explained in Sect. 4. The formula for the cardinality relies on the number of cycles of each type. We denote the number of cycles of type 1 and of pairs of cycles of type 2 by c_1 and c_2 respectively; the number of cycles in the third class with length $2(2k+1)$ is $c_{3,k}$ and the number of cycles in the fourth class with length $4k$ is $c_{4,k,L}$ or $c_{4,k,R}$, depending on the side of the fixed points.

Corollary 3.8 *Let Γ be the bipartite graph corresponding to a vertex M in B_d^ι and let S be the set of disjoint cycles (or type 2 cycle pairs) in Γ. The group of automorphisms $\mathrm{Aut}_\iota(\Gamma)$ of Γ that commute with ι and preserve both sides has cardinality*

$$
|\mathrm{Aut}_\iota(\Gamma)| = \prod_{C \in S} |\mathrm{Aut}_\iota(C)| c_1! \, c_2! \prod_k c_{3,k}! \, c_{4,k,L}! \, c_{4,k,R}!
$$
$$
= 2^{c_2 + \sum_k (c_{3,k} + 2c_{4,k,L} + 2c_{4,k,R})} c_1! \, c_2! \prod_k (c_{3,k}! \, c_{4,k,L}! \, c_{4,k,R}!).
$$

Proof The automorphisms in $\mathrm{Aut}_\iota(\Gamma)$ respect the classification of disjoint cycles in Γ. We can freely permute inside these classifications, hence obtain the factorial terms in the formula. We are left to compute the automorphisms that commute with ι inside each disjoint cycle of the corresponding type:

1. trivial symmetry
2. two automorphisms: ι and the trivial symmetry.
3. two automorphisms: ι and the trivial symmetry.
4. Kleinean group of symmetry, namely we are allowed to swap the two fixed points and also act by ι.

$\qquad\square$

Example 3.9 Let $d = 3$ and assume that p, q both have exactly 1 real root. Then, from case 3 in Example 3.4 we know that B_3^ι is a square. Its representation in terms of the associated graphs is shown in Figure 3. The graph Γ corresponding to either of the bottom vertices consists of one cycle of type 3 with length 6. The graph automorphisms $\mathrm{Aut}_\iota(\Gamma)$ are just the identity and ι itself. On the other hand, the graph Γ corresponding to either of the top vertices in Figure 3 consists of one cycle of type 1 and one pair of cycles of type 2. The graph automorphisms $\mathrm{Aut}_\iota(\Gamma)$ also consist of the identity and ι itself.

3.2 Realization of the vertices

Any vertex of the ι-invariant Birkhoff polytope can be a minimizer for the cost funciton $\mathrm{cost}(M, p, q) = \frac{1}{d} \sum_{1 \leq i, j \leq d} M_{ij} \|\alpha_i - \beta_j\|^2$ for (p, q) in some non-empty open set of monic polynomials without multiple roots. Note that we always consider the roots of the polynomials p, q as ordered. The choice of a specific order does not alter the results, but an order must be fixed. This is implicitly assumed also for the open sets of monic polynomials, since the ordering of the roots of p (resp. q) induces a coherent ordering of the roots of

Fig. 3 The invariant square B_3^ι from Example 3.9. Its vertices correspond to bipartite graphs Γ. The purple squares represent real roots, the green triangles represent complex roots. The involution ι exchanges pairs of green triangles
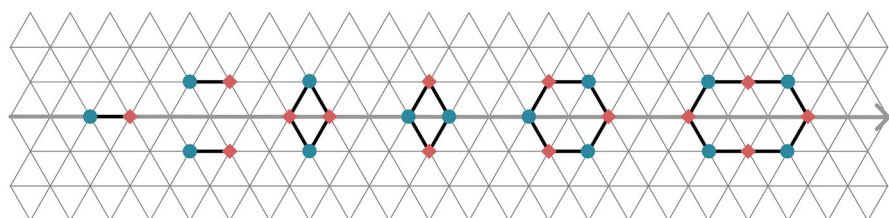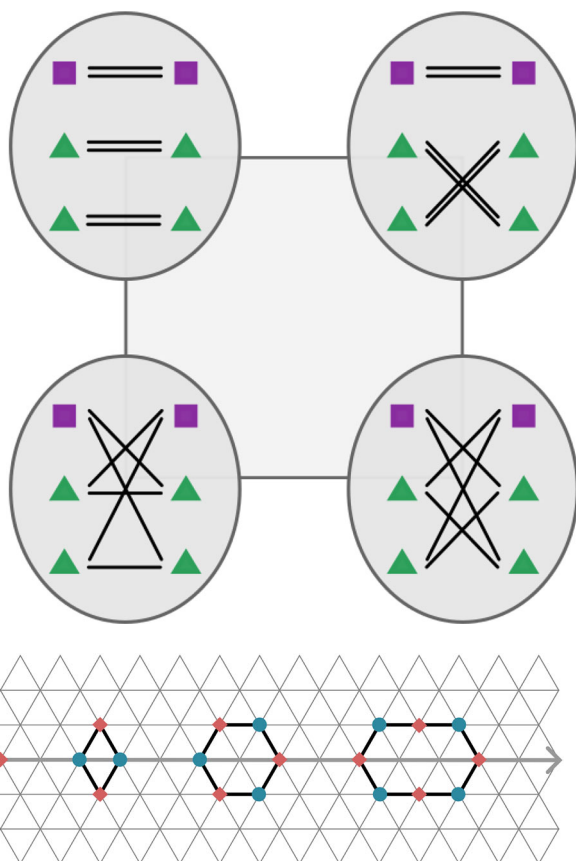
Fig. 4 Disjoint cycles in Γ as subgraphs of the Eisenstein lattice. From left to right, we have disjoint cycles of types 1, 2 (pair), 4, 4, 3, 4

polynomials in a sufficiently small neighborhood of p (resp. q). In this open neighborhood the involutions ϕ, ψ induced by complex conjugation remain the same.

From now on, we work under the assumptions that $p, q \in \mathbb{R}[z]$ factorize as $p(z) = \prod_{1 \leq i \leq d}(z - \alpha_i)$, $q(z) = \prod_{1 \leq j \leq d}(z - \beta_j)$, and that the involutions ϕ, ψ of $\mathrm{Sym}(d)$ satisfy $\overline{\alpha_i} = \alpha_{\phi(i)}$, $\overline{\beta_i} = \beta_{\psi(i)}$. We define the following set.

Definition 3.10 We define $U_{(\phi,\psi)}^M$ to be the set of univariate polynomials $p, q \in \mathbb{R}[z]$ for which M is the unique minimizer for $\mathrm{cost}(M, p, q)$ in B_d^ι.

The uniqueness requirement in the definition of $U_{(\phi,\psi)}^M$ implies that M is a vertex of B_d^ι. The following result holds.

Proposition 3.11 *For every vertex $M \in B_d^\iota$, the set $U_{(\phi,\psi)}^M$ is open and non-empty.*

Proof The roots of a polynomial vary continuously as its coefficients vary. So, if M is the unique minimizer for the cost function of two polynomials p, q with roots ordered corresponding to the involution ϕ, ψ, then there is an open set of polynomial pairs around (p, q) such that any polynomial in the pair has distinct roots (ordered to satisfy the involutions ϕ, ψ), and M is the unique minimizer of the cost function for all polynomial pairs in

the open set. The unique minimizer requirement is an open condition, since it means that $\text{cost}(M, p, q) < \text{cost}(M', p, q)$ for any other vertex $M' \in B_d^\iota$.

We are left to construct the roots of a pair of polynomials of degree d such that M is the unique minimizer for the pair, to prove that $U_{(\phi, \psi)}^M$ is non-empty. Let Γ be the graph associated to M. Recall that Γ consists of disjoint cycles of four different types with cardinalities $c_1, c_2, c_{3,k}, c_{4,k,L}, c_{4,k,R}$. Denote the cycles by $C_1, \ldots, C_\ell$. Let $\omega = e^{\frac{\pi i}{3}}$; we will do our construction on the Eisenstein lattice $\mathbb{Z}[1, \omega]$. The idea is to repeat our construction for each disjoint cycle (or pairs of disjoint cycles, if they are of type 2) and make sure that distinct disjoint cycles are far away from each other.

- C_k is of type 1: assign the value $2dk$ and $2dk+1$ to the corresponding real roots of p and q respectively.
- C_k is of type 2: assign the values $2dk+\omega$, $2dk+\overline{\omega}$ to the pair of roots of p, and $2dk+\omega+1$, $2dk+\overline{\omega}+1$ to the pair of roots of q. In this way, the vertex in Γ corresponding to $2dk+\omega$ connects to the one of $2dk+\omega+1$, and analogously the vertex in Γ corresponding to $2dk+\overline{\omega}$ connects to the one of $2dk+\overline{\omega}+1$.
- C_k is of type 3: assume its length is $2(2j+1)$. Assign the values $2dk$, $2dk+\omega$, $2dk+\omega+1$, $\ldots$, $2dk+\omega+2j-1$, $2dk+2j$, $2dk+\overline{\omega}+2j-1$, $\ldots$, $2dk+\overline{\omega}$ to the roots of p, q alternated.
- C_k is of type 4: assume its length is $4j$ and that the fixed points are on the left side. Assign the values $2dk$, $2dk+\omega$, $2dk+\omega+1$, $\ldots$, $2dk+\omega+2j-2$, $2dk+2j-1$, $2dk+\overline{\omega}+2j-2$, $\ldots$, $2dk+\overline{\omega}$ to the roots of p, q alternated.

See Fig. 4 for an illustration.

By the above construction, for any pair of (i, j) such that $M_{i,j} \neq 0$, we have $\|\alpha_i - \beta_j\| = 1$. Any two distinct lattice points in the Eisenstein lattice have distance at least 1, so M is a minimizer for the cost function of p, q. Suppose there exists another vertex $N \in B_d^\iota$ with $\text{cost}(M) = \text{cost}(N)$ and denote the graph corresponding to N by Γ'. Then, $\|\alpha_i - \beta_j\| = 1$ whenever $N_{ij} \neq 0$. Since any two lattice points in disjoint cycles (or pairs, for type 2) have at least distance d, the disjoint cycles of Γ' have the same vertices of the disjoint cycles of Γ. Inside each disjoint cycle (or pairs, for type 2), the only way to achieve $\|\alpha_i - \beta_j\| = 1$ whenever $N_{ij} \neq 0$ is to assign the roots to p, q exactly as for M. Hence, $M = N$ and M is the unique minimizer for the cost function of p, q. $\square$

Remark 3.12 Note that the set $U_{(\phi, \psi)}^M$ is also semialgebraic, since it is a projection of an appropriate semialgebraic sets arising from the conditions that the pairs of factorized polynomials have a coherent enumeration of their roots, and from M being the unique minimizer. More explicitly, these conditions are polynomial discriminants and inequalities of type $\text{cost}(M, p, q) \leq \text{cost}(M', p, q)$.

4 Algebraic degree formulae

Based on the combinatorial data of the previous section, here we construct a formal candidate of the minimal polynomial for the squared Wasserstein distance. The actual minimal polynomial will always divide the formal candidate. We start by reviewing useful tools from Galois theory.

Consider the ring $A = \mathbb{Q}[x_1, \ldots, x_d, y_1, \ldots y_d]$. We denote by $\cdot$ the following group action of $G = \text{Sym}(d) \times \text{Sym}(d)$ on A:

$$(g_1, g_2) \cdot x_i = x_{g_1(i)},$$

$$(g_1, g_2) \cdot y_j = y_{g_2(j)}.$$

Let $\sigma_1, \ldots, \sigma_d$ be the elementary symmetric polynomials in the variables $x_1, \ldots x_d$, and $\tau_1, \ldots, \tau_d$ be the elementary symmetric polynomials in $y_1, \ldots y_d$. We denote by A^G the ring of invariants under the group action G. Then the following holds.

Lemma 4.1 *The ring of invariants A^G is the polynomial ring $\mathbb{Q}[\sigma_1, \ldots, \sigma_d, \tau_1, \ldots \tau_d]$. In particular, the generators $\sigma_1, \ldots, \sigma_d, \tau_1, \ldots \tau_d$ are algebraically independent.*

Proof Let us consider the subgroup $H = \mathrm{Sym}(d) \times 1$ of G, so H permutes the coordinates x_i. By the fundamental theorem of symmetric polynomials (over the ring $\mathbb{Q}[y_1, \ldots, y_d]$), we have $A^H = \mathbb{Q}[\sigma_1, \ldots \sigma_d, y_1, \ldots y_d]$ with algebraically independent generators σ_i, y_j. Such H is normal in G, and G/H acts on A^H by permuting the coordinates y_i. Using the fundamental theorem of symmetric polynomials over the ring $\mathbb{Q}[\sigma_1, \ldots, \sigma_d]$, we obtain

$$A^G = (A^H)^{(G/H)} = \mathbb{Q}[\sigma_1, \ldots \sigma_d, y_1, \ldots y_d]^{\mathrm{Sym}(d)} = \mathbb{Q}[\sigma_1, \ldots \sigma_d, \tau_1, \ldots \tau_d]$$

with algebraically independent generators. □

From this, we get that the inclusion $\mathbb{Q}(\sigma_1, \ldots \sigma_d, \tau_1, \ldots \tau_d) \subset \mathbb{Q}(x_1, \ldots, x_d, y_1, \ldots y_d)$ is a Galois field extension with Galois group G.

Lemma 4.2 *Let $f \in A$ and let $\{f_1 = f, f_2, \ldots, f_k\}$ be the orbit of f under G. Consider the following product in $A[t]$:*

$$h := \prod_{i=1}^{k} (t - f_i). \tag{6}$$

Then, h actually lies in $A^G[t]$ and it is an irreducible polynomial in $\mathbb{Q}(\sigma_1, \ldots \sigma_d, \tau_1, \ldots \tau_d)[t]$ of degree k in t.

Proof This is a classical exercise in Galois theory about Galois conjugates, see [5, p. 573]. We give a proof sketch for convenience. The action of G permutes the factors of h, so h lies in $(A[t])^G = A^G[t]$. Suppose $h = h_1 h_2$ is a factorization of h in $\mathbb{Q}(x_1, \ldots, x_d, y_1, \ldots y_d)$, and assume without loss of generality that $(t - f_1)$ divides h_1 in $\mathbb{Q}(x_1, \ldots, x_d, y_1, \ldots, y_d)[t]$. Applying the Galois action, we get that every linear factor $(t - f_i)$ divides h_1. Since the f_i are distinct, we get that h_1 must have degree k in t, so the factorization is trivial. □

4.1 Algebraic degrees

Given the polynomials p, q, the involutions $\phi, \psi \in \mathrm{Sym}(d)$ and a vertex M of B_d^ι, consider the polynomial

$$f_M := \frac{1}{d} \sum_{1 \le i, j \le d} M_{ij} (x_i - y_j)(x_{\phi(i)} - y_{\psi(j)}) \subset \mathbb{Q}[x_1, \ldots, x_d, y_1, \ldots y_d], \tag{7}$$

and the map

$$\begin{aligned} \Phi_{p,q} : \mathbb{Q}[x_1, \ldots, x_d, y_1, \ldots y_d] &\to \mathbb{C} \\ x_i &\mapsto \alpha_i, \\ y_i &\mapsto \beta_i. \end{aligned} \tag{8}$$

We observe the following.

Lemma 4.3 *Given a vertex $M \in B_d^\iota$, we have $\Phi_{p,q}(f_M) = \text{cost}(M, p, q)$.*

Proof This follows directly from $\|\alpha_i - \beta_j\|^2 = (\alpha_i - \beta_j)\overline{(\alpha_i - \beta_j)} = (\alpha_i - \beta_j)(\alpha_{\phi(i)} - \beta_{\psi(j)})$ and comparing Eqs. (4) and (7). □

We also have an action of G on $[d] \times \{L, R\}$ via:

$$(g_1, g_2) \cdot (i, L) = (g_1(i), L),$$
$$(g_1, g_2) \cdot (j, R) = (g_2(j), R).$$

By abuse of notation, we identify the above action of G on $[d] \times \{L, R\}$ with the resulting permutation group. For $M \in B_d^\iota$ and Γ associated to M as in Sect. 3.1, we then view $\text{Aut}_\iota(\Gamma)$ as a subgroup of G. In fact we have the following.

Lemma 4.4 *For $g \in G$, $g(f_M) = f_M \Leftrightarrow g \in \text{Aut}_\iota(\Gamma)$, where Γ is defined as in Sect. 3.1.*

In other words, the stabilizer of f_M under the action by G can be identified with $\text{Aut}_\iota(\Gamma)$.

Proof We first decompose f_M into three components: f_M is a quadratic polynomial in the variables $x_1, \ldots x_d, y_1, \ldots y_d$. We can decompose $f_M = f_{M,2,0} + f_{M,1,1} + f_{M,0,2}$ by the multidegree in $x_1, \ldots x_d$ and $y_1, \ldots y_d$. The action of G preserves multidegrees, so $g \in G$ fixes f_M if it fixes all summands in the multidegree decomposition.

Let us first show that $g(f_{M,2,0}) = f_{M,2,0}$ for $g = (g_1, g_2)$ if and only if g_1 commutes with ϕ. Indeed, we have $f_{M,2,0} = \frac{1}{d} \sum_{1 \leq i,j \leq d} M_{ij} x_i x_{\phi(i)} = \frac{1}{d} \sum_{1 \leq i \leq d} x_i x_{\phi(i)}$ since M is doubly stochastic. So the coefficient of x_i^2 is $\frac{1}{d}$ if i is a fixed point of ϕ, and 0 otherwise. For $x_i x_j$ with $i \neq j$ the coefficient is $\frac{2}{d}$ if i and j are interchanged by ϕ, and 0 otherwise. So $g(f_{M,2,0}) = f_{M,2,0}$ if and only if g_1 sends fixed points of ϕ to fixed points of ϕ, and pairs interchanged by ϕ to pairs interchanged by ϕ; namely, if and only if g_1 commutes with ϕ. We can see similarly that $g(f_{M,0,2}) = f_{M,0,2}$ if and only if g_2 commutes with ψ.

Suppose now that $g \in G$ commutes with ι. Let us show that $g(f_{M,1,1}) = f_{M,1,1}$ if and only if $g \in \text{Aut}_\iota(\Gamma)$. Since $M \in B_d^\iota$, we can write

$$f_{M,1,1} = -\frac{1}{d} \sum_{1 \leq i,j \leq d} M_{ij}(x_i y_{\psi(j)} + x_{\phi(i)} y_j)$$

$$= -\frac{1}{d} \left(\sum_{1 \leq i,j \leq d} M_{ij} x_i y_{\psi(j)} + \sum_{1 \leq i,j \leq d} M_{ij} x_{\phi(i)} y_j \right).$$

Now we can substitute $j \mapsto \psi(j)$ in the first sum and $i \mapsto \phi(i)$ in the second sum to obtain

$$f_{M,1,1} = -\frac{1}{d} \sum_{1 \leq i,j \leq d} (M_{i\psi(j)} + M_{\phi(i)j}) x_i y_j = -\frac{2}{d} \sum_{1 \leq i,j \leq d} M_{i\psi(j)} x_i y_j,$$

where the last equality comes from $M \in B_d^\iota$. Hence, for $g = (g_1, g_2) \in G$, we have

$$g(f_{M,1,1}) = -\frac{2}{d} \sum_{1 \leq i,j \leq d} M_{i\psi(j)} x_{g_1(i)} y_{g_2(j)} = -\frac{2}{d} \sum_{1 \leq i,j \leq d} M_{g_1^{-1}(i)\psi(g_2^{-1}(j))} x_i y_j.$$

Therefore, $g(f_{M,1,1}) = f_{M,1,1}$ if and only if $M_{i\psi(j)} = M_{g_1^{-1}(i)\psi(g_2^{-1}(j))}$ for all i, j. Since g commutes with ι, then ψ commutes with g_2^{-1}. We can substitute $j \mapsto \psi(j)$ to see that $g(f_{M,1,1}) = f_{M,1,1}$ is equivalent to g preserving Γ. □

We now consider the orbit of f_M under G, we denote this set by $O_G(f_M)$. Let us denote

$$h_M := \prod_{\tilde{f} \in O_G(f_M)} (\tilde{f} - t) \in A[t]. \tag{9}$$

Applying Lemma 4.2, we get that h_M is irreducible in $\mathbb{Q}(\sigma_1, \dots \sigma_d, \tau_1, \dots \tau_d)[t]$ and by the Orbit-Stabilizer Theorem, it has degree $\frac{(d!)^2}{|\mathrm{Aut}_\iota(\Gamma)|}$. By abuse of notation, we also denote by $\Phi_{p,q} : \mathbb{Q}[x_1, \dots, x_d, y_1, \dots, y_d, t] \to \mathbb{C}[t]$ the extension of the map defined in Eq. 8. We obtain the following result.

Theorem 4.5 *Let $p, q \in \mathbb{Q}[z]$ be monic polynomials of degree d, each with distinct roots.*

Suppose $M \in B_n^\iota$ is a minimizing vertex for $\mathrm{cost}(M, p, q)$ *and Γ is the graph associated to M.*

Then, $\Phi_{p,q}(h_M) \in \mathbb{Q}[t]$ and $\mathrm{W}_2^2(p, q)$ is a root of $\Phi_{p,q}(h_M)$.

In particular, the algebraic degree of $\mathrm{W}_2^2(p, q)$ is bounded by $\frac{(d!)^2}{|\mathrm{Aut}_\iota(\Gamma)|}$.

Proof By Vieta's formulae, $\Phi_{p,q}$ maps σ_i and τ_j (up to sign) to coefficients of p and q. So $\Phi_{p,q}(A^G) \subset \mathbb{Q}$. Since, by Lemma 4.2 $h_M \in A^G[t]$, the claim follows from the above construction. □

Let us show that this bound is sharp on a large set. Recall the open set of polynomial pairs $U_{(\phi,\psi)}^M$ defined in Definition 3.10, where polynomial roots are ordered satisfying the involution ϕ, ψ and M is the unique minimizer for the cost function.

Theorem 4.6 *For every vertex $M \in B_d^\iota$, there exists a dense set $U \subseteq U_{(\phi,\psi)}^M$ of polynomial pairs (p, q) such that $\Phi_{p,q}(h_M)$ is irreducible. In particular, the algebraic degree of $\mathrm{W}_2^2(p, q)$ is equal to $\frac{(d!)^2}{|\mathrm{Aut}_\iota(\Gamma)|}$ for these polynomials, so the bound of Theorem 4.5 is sharp.*

Proof The extended map $\Phi_{p,q} : \mathbb{Q}[\sigma_1, \dots, \sigma_d, \tau_1, \dots, \tau_d, t] \to \mathbb{Q}[t]$, is the specialization map sending σ_i and τ_j (up to sign) to the coefficients of p and q. We know that $h_M \in A^G[t]$ is irreducible by Lemma 4.2. By Hilbert's irreducibility theorem (see e.g. [19, Theorem 46] or [13, Chapter 9, Corollary 2.5]), the specialization $\Phi_{p,q}(h_M)$ is irreducible for a dense set of monic polynomials $p, q \in \mathbb{Q}[z]$. In particular, on the open set $U_{(\phi,\psi)}^M$ that admits M as the unique minimizer, we have that $\Phi_{p,q}(h_M)$ is the minimal polynomial of $\mathrm{W}_2^2(p, q)$ for a dense open subset U of $U_{(\phi,\psi)}^M$. Hence, our bound in Theorem 4.5 is sharp on U. □

Example 4.7 Suppose $p, q \in \mathbb{Q}[z]$ are generic monic polynomials of degree d, both with only real roots. The only possible type (cf. Theorem 3.7) for the disjoint cycles of the graph associated to vertices of B_d^ι is type 1. Therefore, we get $c_1 = d$, $c_2 = c_{3,k} = c_{4,k,L} = c_{4,k,R} = 0$ and $|\mathrm{Aut}_\iota(\Gamma)| = c_1! = d!$. The algebraic degree of $W_2^2(p, q)$ is

$$\mathrm{WDdegree}(p, q) = \frac{(d!)^2}{d!} = d!,$$

for generic p, q, by Theorem 4.6.

For degree 1 polynomials, WDdegree $= 1$ trivially. When $d = 2$, if $(p_\mathbb{R}, p_\mathbb{C}) = (q_\mathbb{R}, q_\mathbb{C})$ and p, q are generic, then WDdegree $= 2$. In the remaining cases, the Wasserstein distance degree is 1. The values of the Wasserstein distance degree for $d = 3, 4$ are displayed in Table 2. Degree 4 is the first interesting case where polynomials with the same number of real and complex roots give rise to multiple open sets where we get different algebraic degrees for $\mathrm{W}_2^2(p, q)$.

Table 2 Values of WDdegree(p, q) for a dense set of pairs (p, q)

Degree	$(p_{\mathbb{R}}, p_{\mathbb{C}}), (q_{\mathbb{R}}, q_{\mathbb{C}})$	WDdegree(p, q)
3	(3,0), (3,0)	6
3	(3,0), (1,1)	9
3	(1,1), (1,1)	18
4	(4,0), (4,0)	24
4	(4,0), (2,1)	72
4	(4,0), (0,2)	18
4	(2,1), (2,1)	36, 144, 288
4	(2,1), (0,2)	18, 288
4	(0,2), (0,2)	72

Note that Theorem 4.5 gives the algebraic degree of the squared Wasserstein distance for a dense set of pairs of polynomials. However, in special cases, it provides only an upper bound, as exhibited in the next example.

Example 4.8 Consider $p = z^3 - 1$ with only one real root and $q = z^3 - 5z^2 + 4z + 3$ with three real roots. Let ι be the associated involution and M the optimal vertex of B_3^ι. Theorem 4.5 predicts an upper bound WDdegree$(p, q) \leq \deg h_M = 9$. Indeed, the minimal polynomial of the squared Wasserstein distance in this case is

$$27t^3 - 540t^2 + 3483t - 7231,$$

hence the Wasserstein distance degree is 3. This is due to the fact that p is special, in the sense that one of its roots is a rational (in this case even integer) number.

Remark 4.9 If $p, q \in \mathbb{Q}[z]$ (with possibly multiple roots), and $(p, q) \in \overline{U^M_{(\phi, \psi)}}$, then we still get that $W_2^2(p, q)$ is a root of $\phi_{p,q}(h_M)$, so we still obtain a bound for the Wasserstein distance degree. However, this bound is far from being sharp in general.

4.2 Wasserstein distance degree via elimination

While the Wasserstein distance degree is studied from combinatoric perspective, the algebraic degree of two specific polynomials can also be understood via elimination of ideals. We describe here how to compute the minimal polynomial of the Wasserstein distance of two given polynomials via elimination. For the theory of elimination ideals, we refer to [14].

Let $p, q \in \mathbb{Q}[z]$ be polynomials of degree d with simple roots; we denote by p_i, q_i the coefficient of the monomial z^i in the univariate polynomial p, q respectively. Consider the ideal in $\mathbb{Q}[x_1, \ldots, x_d, y_1, \ldots, y_d]$ generated by Vieta's formulae, i.e., all the relations between the elementary symmetric polynomials in the roots of p, q and their coefficients:

$$I = \left\langle \sum_{1 \leq i \leq d} x_i + \frac{p_{d-1}}{p_d}, \ldots, \prod_{1 \leq i \leq d} x_i - (-1)^d \frac{p_0}{p_d}, \right.$$

$$\left. \sum_{1 \leq i \leq d} y_i + \frac{q_{d-1}}{q_d}, \ldots, \prod_{1 \leq i \leq d} y_i - (-1)^d \frac{q_0}{q_d} \right\rangle.$$

The ideal I is zero dimensional. By construction, the variety associated to it consists of all tuples of points in $\mathbb{C}^{2d}$ such that the first d coordinates are the roots of p, and the last d

coordinates are the roots of q. Counting all permutations of the two sets of coordinates, we get that deg $I = (d!)^2$. Notice that in the case where p, q are allowed to have double roots, the degree of I decreases, because some permutations coincide, but $(d!)^2$ remains a valid upper bound.

Assume that the minimizer of the cost function $\mathrm{cost}(\cdot, p, q)$ is attained at the vertex M of B_d^ι, and let h_M be the invariant irreducible polynomial defined in (9). Consider a second ideal

$$J = I + \langle t - f_M(x, y) \rangle \subset \mathbb{Q}[x_1, \ldots, x_d, y_1, \ldots, y_d, t].$$

Using the isomorphism $\mathbb{Q}[x_1, \ldots, x_d, y_1, \ldots, y_d, t]/\langle t - f_M(x, y) \rangle = \mathbb{Q}[x_1, \ldots, x_d, y_1, \ldots, y_d]$, we see that $\mathbb{Q}[x_1, \ldots, x_d, y_1, \ldots, y_d, t]/J \cong \mathbb{Q}[x_1, \ldots, x_d, y_1, \ldots, y_d, t]/I$, in particular, J is also zero dimensional and has the same degree as I. Points in $V(J)$ are of the form $(\alpha, \beta, f_M(\alpha, \beta))$ for (α, β) in $V(I)$. The elimination ideal $E = J \cap \mathbb{Q}[t]$ has dimension dim $J = 0$ and is explicitly generated by some polynomial. During the elimination process or, more geometrically, during this projection, some of the $(d!)^2$ points in the variety of J get mapped to the same point. Points in $V(E)$ are $\tilde{f}(\alpha, \beta)$ for some fixed $(\alpha, \beta) \in V(I)$ and every $\tilde{f} \in O_G(f_M)$. So, for a dense set of pairs (p, q), E is generated by $\Phi_{p,q}(h_M)(t)$, where $\Phi_{p,q}$ is the natural extension of the specialization map in (8), and it is the minimal polynomial of $W_2^2(p, q)$. In general, the generator of E contains as an irreducible factor the minimal polynomial of $W_2^2(p, q)$.

References

1. Antonini, P., Cavalletti, F., Lerario, A.: Optimal transport between algebraic hypersurfaces. Geometric and Functional Analysis, 1–70 (2025)
2. Baumeister, B., Ladisch, F.: A property of the Birkhoff polytope. Algebr. Comb. **1**(2), 275–281 (2018)
3. Birkhoff, G.: Tres observaciones sobre el algebra lineal. Re- vista Facultad de Ciencias Exactas, Puras y Aplicadas Universidad Nacional de Tucuman, Serie A **5**(20), 147–151 (1946)
4. Breiding, P., Kohn, K., Sturmfels, B.: Metric Algebraic Geometry. Oberwolfach Seminars. Birkhäuser, Basel (2024)
5. David, S., Foote, R.M.: Abstract Algebra, 3rd edn. John Wiley & Sons Inc, Hoboken (2004)
6. Draisma, J., Horobeţ, E., Ottaviani, G., Sturmfels, B., Thomas, R.R.: The Euclidean distance degree of an algebraic variety. J. Soc. Found. Comput. Math. **16**(1), 99–149 (2016)
7. Hans-Christian Graf von Bothmer and Kristian Ranestad: A general formula for the algebraic degree in semidefinite programming. Bull. Lond. Math. Soc. **41**(2), 193–197 (2009)
8. Hiep, D.T., Giao, N.T., Ngoc, V., Mai, N.T.: A characterization of the algebraic degree in semidefinite programming. Collect. Math. **74**(2), 443–455 (2023)
9. Hoşten, S., Khetan, A., Sturmfels, B.: Solving the likelihood equations. J. Soc. Found. Comput. Math. **5**(4), 389–407 (2005)
10. Huh, J., Sturmfels, B.: Likelihood geometry. In: Combinatorial Algebraic Geometry, volume **2108** of Lecture Notes in Math., pages 63–117. Springer, Cham (2014)
11. Jesús, A., Loera, D., Kim, E. D.: Combinatorics and geometry of transportation polytopes: an update. In: Discrete Geometry and Algebraic Combinatorics, volume **625** of Contemp. Math., pages 37–76. Amer. Math. Soc., Providence, RI (2014)
12. Krein, M., Milman, D.: On extreme points of regular convex sets. Studia Math. **9**, 133–138 (1940)
13. Lang, S.: Fundamentals of Diophantine Geometry. Springer-Verlag, New York (1983)
14. Michałek, M., Sturmfels, B.: Invitation to Nonlinear Algebra, volume **211** of Graduate Studies in Mathematics. American Mathematical Society, Providence [2021] (2021)
15. Nie, J., Ranestad, K.: Algebraic degree of polynomial optimization. SIAM J. Optim. **20**(1), 485–502 (2009)

16. Nie, J., Ranestad, K., Sturmfels, B.: The algebraic degree of semidefinite programming. Math. Program. **122**(2), 379–405 (2010)
17. Ottaviani, G., Sodomaco, L.: The distance function from a real algebraic variety. Comput. Aided Geometric Design **82**(101927), 20 (2020)
18. Rostalski, P., Sturmfels, B.: Dualities in convex algebraic geometry. Rendiconti di Matematica e delle sue Applicazioni. Serie VII **30**(3–4), 285–327 (2010)
19. Schinzel, A.: Polynomials with Special Regard to Reducibility, volume **77** of Encyclopedia of Mathematics and its Applications. Cambridge University Press, Cambridge, With an appendix by Umberto Zannier (2000)
20. Schrijver, A.: Combinatorial optimization. Polyhedra and Efficiency. Vol. **A**, volume **24** of Algorithms and Combinatorics. Springer-Verlag, Berlin (2003)
21. Sullivant, S.: Algebraic Statistics. Graduate Studies in Mathematics, vol. 194. American Mathematical Society, Providence (2018)
22. Vieta, F.: Opera mathematica. Elzevier, Leiden, 1646. Edited by Frans van Schooten
23. Villani, C.: Topics in Optimal Transportation. Graduate Studies in Mathematics, vol. 58. American Mathematical Society, Providence (2003)
24. von Neumann, J.: A certain zero-sum two-person game equivalent to the optimal assignment problem. In: Kuhn, H.W., Tucker, A.W. (eds.) Contributions to the Theory of Games (AM-28), vol. II, pp. 5–12. Princeton University Press, Princeton (1953)
25. Ziegler, G.M.: Lectures on Polytopes. Graduate Texts in Mathematics, vol. 152. Springer-Verlag, New York (1995)