Local structure of multi-dimensional martingale optimal transport

Hadrien De March†

May 25, 2018

Abstract

This paper analyzes the support of the conditional distribution \( P_X^* \) of optimal martingale transport plans in \( \mathbb{R}^d \), for arbitrary dimension \( d \geq 1 \). In the context of a distance coupling in dimension larger than 2, previous results established by Ghoussoub, Kim & Lim [9] show that \( P_X^* \) is concentrated on its own Choquet boundary. Moreover, when the target measure is atomic, they prove that the support of \( P_X^* \) is concentrated on \( d + 1 \) points, and conjecture that this result is valid for arbitrary target measure.

We provide a structure result of the support of \( P_X^* \) for general Lipschitz couplings. Using tools from algebraic geometry, we provide sufficient conditions for finiteness of this conditional support, together with (optimal) lower bounds on the maximal cardinality for a given coupling function. More results are obtained for specific examples of coupling functions based on distance functions. In particular, we show that the above conjecture of Ghoussoub, Kim & Lim is not valid beyond the context of atomic target distributions.

Key words. Martingale optimal transport, local structure, differential structure, support.

1 Introduction

The problem of martingale optimal transport was introduced as the dual of the problem of robust (model-free) superhedging of exotic derivatives in financial mathematics, see Beiglböck, Henry-Labordère & Penkner [2] in discrete time, and Galichon, Henry-Labordère & Touzi [8] in continuous-time. Previously the robust superhedging problem was introduced by Hobson [17],

*The author gratefully acknowledges the financial support of the ERC 321111 Rofirm, and the Chairs Financial Risks (Risk Foundation, sponsored by Société Générale) and Finance and Sustainable Development (IEF sponsored by EDF and CA).

†CMAP, École Polytechnique, hadrien.de-march@polytechnique.org.
and was addressing specific examples of exotic derivatives by means of corresponding solutions of the Skorokhod embedding problem, see [5 15 16], and the survey [14].

Our interest in the present paper is on the multi-dimensional martingale optimal transport. Given two probability measures \( \mu, \nu \) on \( \mathbb{R}^d \), with finite first order moment, martingale optimal transport differs from standard optimal transport in that the set of all interpolating probability measures \( \mathcal{P}(\mu, \nu) \) on the product space is reduced to the subset \( \mathcal{M}(\mu, \nu) \) restricted by the martingale condition. We recall from Strassen [22] that \( \mathcal{M}(\mu, \nu) \neq \emptyset \) if and only if \( \mu \leq \nu \) in the convex order, i.e. \( \mu(f) \leq \nu(f) \) for all convex functions \( f \). Notice that the inequality \( \mu(f) \leq \nu(f) \) is a direct consequence of the Jensen inequality, the reverse implication follows from the Hahn-Banach theorem.

This paper focuses on showing the differential structure of the support of optimal probabilities for the martingale optimal transport problem. In the case of optimal transport, a classical result by Rüschendorf [21] states that if the map \( y \mapsto c_x(x_0, y) \) is injective, then the optimal transport is unique and supported on a graph, i.e. we may find \( T : \mathcal{X} \rightarrow \mathcal{Y} \) such that \( \mathbb{P}^*[Y = T(X)] = 1 \) for all optimal coupling \( \mathbb{P}^* \in \mathcal{P}(\mu, \nu) \). The corresponding result in the context of the one-dimensional martingale transport problem was obtained by Beiglböck-Juillet [4], and further extended by Henry-Labordère & Touzi [12]. Namely, under the so-called martingale Spence-Mirrlees condition, \( c_x \) strictly convex in \( y \), the left-curtain transport plan is optimal and concentrated on two graphs, i.e. we may find \( T_d, T_u : \mathcal{X} \rightarrow \mathcal{Y} \) such that \( \mathbb{P}^*[Y \in \{T_d(X), T_u(X)\}] = 1 \) for all optimal coupling \( \mathbb{P}^* \in \mathcal{M}(\mu, \nu) \). In this case we get similarly the uniqueness by a convexity argument.

An important issue in optimal transport is the existence and the characterization of optimal transport maps. Under the so-called twist condition (also called Spence-Mirrlees condition in the economics litterature) it was proved that the optimal transport is supported on one graph. In the context of martingale optimal transport on the line, Beiglböck & Juillet introduced the left-monotone martingale interpolating measure as a remarkable transport plan supported on two graphs, and prove its optimality for some classes of coupling functions. Ghoussoub, Kim & Lim conjectured that in higher dimensional Martingale Optimal Transport for distance coupling, the optimal plans will be supported on \( d+1 \) graphs. We prove here that there is no hope of extending this property beyond the case of atomic measure. This is obtained using the reciprocal property of the structure theorem of this paper, which serves as a counterexample generator.

A first such study in higher dimension was performed by Lim [19] under radial symmetry that allows in fact to reduce the problem to one-dimension. A more "higher-dimensional specific" approach was achieved by Ghoussoub, Kim & Lim [9]. Their main structure result is that for the Euclidean distance coupling, the supports of optimal kernels will be concentrated on their own Choquet boundary (i.e. the extreme points of the closure of their convex hull).

Our subsequent results differ from [9] from two perspectives. First, we prove that with the same techniques we can easily prove much more precise results on the local structure of the optimal Kernel, in particular, we prove that they are concentrated on \( 2d \) (possibly degenerate)
graphs, which is much more precise than a concentration on the Choquet boundary. Our main structure result states that the optimal kernels are supported on the intersection of the graph of the partial gradient $c_x(x_0, \cdot)$ with the graph of an affine function $A_{x_0} \in \text{Aff}_d$. Second, we prove a reciprocal property, i.e. that for any subset of such graph intersection \( \{ c_x(x_0, Y) = A(Y) \} \) for \( A \in \text{Aff}_d \), we may find marginals such that this set is an optimizer for these marginals. Thanks to this reciprocal property we prove that Conjecture 2 in [9] that we mentioned above is wrong. They prove this conjecture in the particular case in which the second marginal $\mu$ is atomic, however in view of our results it only works in this particular case, as we produce counterexamples in which $\mu$ and $\nu$ are dominated by the Lebesgue measure. Indeed, we prove that the support of the conditional kernel is characterized by an algebraic structure independent from the support of $\nu$, then when this support is atomic, very particular phenomena happen. Thus the intuition suggests that finding this kind of solution for an atomic approximation of a non-atomic $\nu$ is not a stable approach, as in the limit there are generally $2d$ points in the kernel.

The paper is organized as follows. Section 2 gives the main results: Subsection 2.1 states the Assumption and the main structure theorem, Subsection 2.2 applies this theorem to show the relation between finiteness of the conditional support and the algebraic geometry of its derivatives, Subsection 2.3 gives the maximal cardinality that is universally reachable for the support up to choosing carefully the marginals, and finally Subsection 2.4 shows how the structure theorem applied to classical couplings like powers of the Euclidean distance allows to give precise descriptions and properties of the conditional supports of optimal plans. Finally Section 3 contains all the proofs to the results in the previous sections, and Section 4 provides some numerical experiments.

**Notation** We fix an integer $d \geq 1$. For $x \in \mathbb{R}$, we denote $sg(x) := 1_{x > 0} - 1_{x < 0}$. If $f : \mathbb{R} \rightarrow \mathbb{R}$ we denote by $\text{fix}(f)$ the set of fixed points of $f$. A function $f : \mathbb{R}^d \rightarrow \mathbb{R}^d$ is said to be super-linear if $\lim_{|y| \to \infty} \frac{|f(y)|}{|y|} = \infty$. Let a function $f : \mathbb{R}^d \rightarrow \mathbb{R}$ and $x_0 \in \mathbb{R}^d$, we say that $f$ is super-differentiable (resp. sub-differentiable) at $x_0$ if we may find $p \in \mathbb{R}^d$ such that $f(x) - f(x_0) \leq p \cdot (x - x_0) + o(x - x_0)$ (resp. $\geq$) when $x \rightarrow x_0$, in this condition, we say that $p$ belongs to the super-gradient $\partial^+ f(x_0)$ (resp. sub-gradient $\partial^- f(x_0)$) of $f$ at $x_0$. This local notion extends the classical global notion of super-differential (resp. sub) for concave (resp. convex) functions.

For $x \in \mathbb{R}^d$, $r \geq 0$, and $V$ an affine subspace of dimension $d'$ containing $x$, we denote $S_V(x, r)$ the dim $V - 1$ dimensional sphere in the affine space $V$ for the Euclidean distance, centered in $x$ with radius $r$. We denote by $\text{Aff}_d$ the set of Affine maps from $\mathbb{R}^d$ to itself. Let $A \in \text{Aff}_d$, notice that its derivative $\nabla A$ is constant over $\mathbb{R}^d$, we abuse notation and denote $\nabla A$ for the matrix representation of this derivative. Let $M \in \mathcal{M}_d(\mathbb{R})$, a real matrix of size $d$, we denote $\det M$ the determinant of $M$, ker $M$ is the kernel of $M$, $\text{Im} M$ is the image of this matrix, and $\text{Sp}(M)$ is the set of all complex eigenvalues of $M$. We also denote $\text{Com}(M)$ the comatrix of $M$: for $1 \leq i, j \leq d$, $\text{Com}(M)_{i,j} = (-1)^{i+j} \det M^{i,j}$, where $M^{i,j}$ is the matrix of
size \( d - 1 \) obtained by removing the \( i^{th} \) line and the \( j^{th} \) row of \( M \). Recall the useful comatrix formula:

\[
Com(M)^t M = M Com(M)^t = (\det M) I_d.
\]

As a consequence, whenever \( M \) is invertible, \( M^{-1} = \frac{1}{\det M} Com(M)^t \). Throughout this paper, \( \mathbb{R}^d \) is endowed with the Euclidean structure, the Euclidean norm of \( x \in \mathbb{R}^d \) will be denoted \( |x| \), the \( p \)-norm of \( x \) will be denoted \( |x|_p := \left( \sum_{i=1}^{d} |x_i|^p \right)^{\frac{1}{p}} \). We denote \( (e_i)_{1 \leq i \leq d} \) the canonical basis of \( \mathbb{R}^d \). Let \( B \subset E \) with \( E \) a vector space, we denote \( B^* := B \setminus \{0\} \), and \( |B| \) the possibly infinite cardinal of \( B \). If \( V \) is a topological affine space and \( B \subset V \) is a subset of \( V \), \( \text{int} B \) is the interior of \( B \), \( \text{cl} B \) is the closure of \( B \), \( \text{aff} B \) is the smallest affine subspace of \( V \) containing \( B \), \( \text{conv} B \) is the convex hull of \( B \), \( \text{dim}(B) := \dim(\text{aff} B) \), and \( \text{ri} B \) is the relative interior of \( B \), which is the interior of \( B \) in the topology of \( \text{aff} B \) induced by the topology of \( V \). We also denote by \( \partial B := \text{cl} B \setminus \text{ri} B \) the relative boundary of \( B \), and if \( V \) is endowed with a euclidean structure, we denote by \( \text{proj}_B(x) \) the orthogonal projection of \( x \in V \) on \( \text{aff} B \). A set \( B \) is said to be discrete if it consists of isolated points.

We denote \( \Omega := \mathbb{R}^d \times \mathbb{R}^d \) and define the two canonical maps

\[
X : (x, y) \in \Omega \mapsto x \in \mathbb{R}^d \quad \text{and} \quad Y : (x, y) \in \Omega \mapsto y \in \mathbb{R}^d.
\]

For \( \varphi, \psi : \mathbb{R}^d \to \mathbb{R} \), and \( h : \mathbb{R}^d \to \mathbb{R}^d \), we denote

\[
\varphi \oplus \psi := \varphi(X) + \psi(Y), \quad \text{and} \quad h \otimes := h(X) \cdot (Y - X),
\]

with the convention \( \infty - \infty = \infty \).

For a Polish space \( X \), we denote by \( \mathcal{P}(X) \) the set of all probability measures on \( (X, \mathcal{B}(X)) \). For \( P \in \mathcal{P}(X) \), we denote by \( \text{supp} P \) the smallest closed support of \( P \). Let \( Y \) be another Polish space, and \( P \in \mathcal{P}(X \times Y) \). The corresponding conditional kernel \( P_x \) is defined by:

\[
P(dx, dy) = \mu(dx) P_x(dy), \quad \text{where} \quad \mu := P \circ X^{-1}.
\]

Let \( n \geq 0 \) and a field \( \mathbb{K} \) (\( \mathbb{R} \) or \( \mathbb{C} \) in this paper), we denote \( \mathbb{K}_n[X] \) the collection of all polynomials on \( \mathbb{K} \) of degree at most \( n \). The set \( \mathbb{C}^{\text{hom}}[X] \) is the collection of homogeneous polynomials of \( \mathbb{C}[X] \). Similarly for \( k \geq 1 \), we define \( \mathbb{K}_n[X_1, ..., X_d] \) the collection of multivariate polynomials on \( \mathbb{K} \) of degree at most \( n \). We denote the monomial \( X^\alpha := X_1^{\alpha_1}...X_d^{\alpha_d} \), and \( |\alpha| = \alpha_1 + \ldots + \alpha_d \) for all integer vector \( \alpha \in \mathbb{N}^d \). For two polynomial \( P \) and \( Q \), we denote \( \text{gcd}(P, Q) \) their greatest common divider. Finally, we denote \( \mathbb{P}^d := (\mathbb{C}^{d+1})^* / \mathbb{C}^* \) the projective plan of degree \( d \).

**The martingale optimal transport problem** Throughout this paper, we consider two probability measures \( \mu \) and \( \nu \) on \( \mathbb{R}^d \) with finite first order moment, and \( \mu \leq \nu \) in the convex order, i.e. \( \nu(f) \geq \mu(f) \) for all integrable convex \( f \). We denote by \( \mathcal{M}(\mu, \nu) \) the collection of all probability measures on \( \mathbb{R}^d \times \mathbb{R}^d \) with marginals \( \mathbb{P} \circ X^{-1} = \mu \) and \( \mathbb{P} \circ Y^{-1} = \nu \). Notice that \( \mathcal{M}(\mu, \nu) \neq \emptyset \) by Strassen [22].
An $\mathcal{M}(\mu, \nu)$—polar set is an element of $\cap_{\mathbb{P} \in \mathcal{M}(\mu, \nu)} \mathcal{N}_{\mathbb{P}}$. A property is said to hold $\mathcal{M}(\mu, \nu)$—quasi surely (abbreviated as q.s.) if it holds on the complement of an $\mathcal{M}(\mu, \nu)$—polar set.

For a derivative contract defined by a non-negative coupling function $c : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}_+$, the martingale optimal transport problem is defined by:

$$S_{\mu, \nu}(c) := \sup_{\mathbb{P} \in \mathcal{M}(\mu, \nu)} \mathbb{P}[c].$$

The corresponding robust superhedging problem is

$$I_{\mu, \nu}(c) := \inf_{(\varphi, \psi, h) \in D_{\mu, \nu}(c)} \mu(\varphi) + \nu(\psi),$$

where

$$D_{\mu, \nu}(c) := \{ (\varphi, \psi, h) \in L^1(\mu) \times L^1(\nu) \times L^1(\mu, \mathbb{R}^d) : \varphi \oplus \psi + h \geq c \}. \quad (1.4)$$

The following inequality is immediate:

$$S_{\mu, \nu}(c) \leq I_{\mu, \nu}(c). \quad (1.5)$$

This inequality is the so-called weak duality. For upper semi-continuous coupling, Beiglböck, Henry-Labordère, and Penckner [2], and Zaev [23] proved that strong duality holds, i.e. $S_{\mu, \nu}(c) = I_{\mu, \nu}(c)$. For any Borel coupling function, De March [6] extended the quasi sure duality result to the multi-dimensional context, and proved the existence of a dual minimizer.

## 2 Main results

### 2.1 Main structure theorem

We denote $\mathcal{K}$ the collection of closed convex subsets of $\mathbb{R}^d$, which is a Polish space when endowed with the Wijsman topology (see Beer [1]). De March & Touzi [7] proved that we may find a Borel mapping $I : \mathbb{R}^d \rightarrow \mathcal{K}$ such that $\{ I(x) : x \in \mathbb{R}^d \}$ is a partition of $\mathbb{R}^d$, $Y \in \text{cl } I(X)$, $\mathcal{M}(\mu, \nu)$—a.s. and $\text{cl } I(X) = \text{cl conv supp } \hat{\mathbb{P}}_X$, $\mu$—a.s. for some $\hat{\mathbb{P}} \in \mathcal{M}(\mu, \nu)$. As the map $I$ is Borel, $I(X)$ is a random variable, let $\eta := \mu \circ I^{-1}$ be the push forward of $\mu$ by $I$. It was proved in [6] that the optimal transport disintegrates on all the 'components' $I(X)$. The following conditions are needed throughout this paper.

**Assumption 2.1.** (i) The coupling $c$ is locally Lipschitz and sub-differentiable in the first variable $x \in I$, uniformly in the second variable $y \in \text{cl } I$, $\eta$—a.s.

(ii) The conditional probability $\mu_I := \mu \circ (X|I(X))^{-1}$ is dominated by the Lebesgue measure on $I$, $\eta$—a.s.

An important question in optimal transport theory is the structure of the support of the conditional distribution of optimal transport plans.
Theorem 2.2. Let \( c : \Omega \to \bar{\mathbb{R}}_+ \) be upper semianalytic with \( S_{\mu,\nu}(c) < \infty \), then under Assumption \( 2.1 \) we may find \( (A_x)_{x \in \mathbb{R}^d} \subset \text{Aff}_d \) such that for all \( \mathbb{P}^* \in \mathcal{M}(\mu, \nu) \) optimal for \( (1.2) \),
\[
x \in \text{ri conv sup} \mathbb{P}^*, \quad \text{and sup} \mathbb{P}^* \subset \{ c_x(x, Y) = A_x(Y) \} \quad \text{for } \mu - \text{a.e. } x \in \mathbb{R}^d.
\]
Conversely, let a compact \( S \subset \{ c_x(x_0, Y) = A(Y) \} \) for some \( x_0 \in \mathbb{R}^d \) and \( A \in \text{Aff}_d \), be such that \( x_0 \in \text{int conv } S \), \( c \) is \( C^{2,0} \cap C^{1,1} \) in the neighborhood of \( S_0 \), and \( c_{xy}(S_0) - \nabla A \subset GL_d(\mathbb{R}) \), then \( S_0 \) has a finite cardinal \( k \geq d + 1 \) and we may find \( \mu_0, \nu_0 \in \mathcal{P}(\mathbb{R}^d) \) with \( C^1 \) densities such that
\[
\mathbb{P}^*(dx, dy) := \mu_0(dx) \sum_{i=1}^k \lambda_i(x) \delta_{T_i(x)}(dy)
\]
is the unique solution to \( (1.2) \), with \( (T_i)_{1 \leq i \leq k} \subset C^1(\text{supp } \mu_0, \mathbb{R}^d) \) such that \( S_0 = \{ T_i(x_0) \}_{1 \leq i \leq k} \), and \( (\lambda_i)_{1 \leq i \leq k} \subset C^1(\text{supp } \mu_0) \).

The proof of this theorem is reported in Subsection 3.1.

Remark 2.3. Notice that even though the set \( S_0 := \{ c_x(x_0, Y) = A(Y) \} \) for \( x_0 \in \mathbb{R}^d \) and \( A \in \text{Aff}_d \) may contain more than \( d + 1 \) points, it is completely determined by \( d + 1 \) affine independent points \( y_1, \ldots, y_{d+1} \in S_0 \), as the equations \( c_x(x_0, y_i) = A(y_i) \) determine completely the affine map \( A \).

The first statement of Theorem 2.2 is well known, it is already used in \[12\] (to establish Theorem 5.1), \[3\] (see Theorem 7.1), and \[9\] (for Theorem 5.5). However, the converse implication is new and we will show in the next subsections how it gives crucial information about the structure of martingale optimal transport for classical coupling functions. This converse implication will serve as a counterexample generator, similar to counterexample 7.3.2 in \[3\], which could have been found by an immediate application of the converse implication in Theorem 2.2.

Beiglböck & Juillet \[3\] and Henry-Labordère & Touzi \[12\] solved the problem in dimension 1 for the distance coupling or for couplings satisfying the "Spence-Mirless condition" (i.e. \( \frac{\partial^2}{\partial x \partial y} c > 0 \)), in these particular cases, the support of the optimal probabilities is contained in two points in \( y \) for \( x \) fixed. See also Beiglböck, Henry-Labordère & Touzi \[3\]. Some more precise results have been provided by Ghoussoub, Kim, and Lim \[9\]: they show that for the distance coupling, the image can be contained in its own Choquet boundary, and in the case of minimization, they show that in some particular cases the image consists of \( d + 1 \) points, which provides uniqueness. They conjecture that this remains true in general. The subsequent theorem will allow us to prove that this conjecture is wrong, and that the properties of the image can be found much more precisely.

### 2.2 Algebraic geometric finiteness criterion

#### 2.2.1 Transversality of multivariate polynomial families

Algebraic geometry is the study of algebraic varieties, which are the sets of zeros of multivariate polynomials. When the coupling \( c \) is smooth, the set \( \{ c_x(x_0, Y) = A(Y) \} \) for \( x_0 \in \mathbb{R}^d \) and
We say that the family $\Pi$ then $d_i$.

We start with the one dimensional case. We emphasize that the sufficient condition

Theorem 2.7.

The varieties defined by the polynomials intersect transversally. The ordering of the polynomials

Definition 2.4.

in Definition 2.4 does not matter.

$\partial X^\infty$ projective space has dimension $d$.

$\partial X^\infty$

Here $d_i$.

Example 2.5.

Let $d_i$.

Definition 2.4.

Let $d_i$.

$\partial X^\infty$

If $d_i$.

Irr $\partial X^\infty$.

Indeed, let $d_i$.

Example 2.5.

We denote $\partial X^\infty$.

The result is proved.

The notion is also invariant by linear change of variables. For example, $(X^3 + XY + 3, Y^3 - X^2 + X)$ is transversal because the homogeneous polynomial family $(X^3, Y^3)$ is transversal by Example 2.5 above.

Example 2.6. Let $d_i$.

If $d_i$.

$\partial X^\infty$

non transversal, $P_2 = \partial R \Pi$ with $R \in \partial X^\infty$ and $\Pi|P_1$. Then $\Pi|\gcd(P_1, P_2) \neq 1$.

Let $k, d \in \mathbb{N}$ and $(P_1, \ldots, P_k)$ be $k$ homogeneous polynomials in $\mathbb{R}[X_0, X_1, \ldots, X_d]$, we define the set of common zeros of $(P_1, \ldots, P_k)$: $Z(P_1, \ldots, P_k) = \{x \in \mathbb{R}^d : P_i(x) = 0, \text{ for all } 1 \leq i \leq k\}$. An element $x \in \mathbb{R}^d$ is a single common root of $P_1, \ldots, P_k$ if $x \in Z(P_1, \ldots, P_k)$, and the vectors $\nabla P_i(x)$ are linearly independent in $\mathbb{R}^d$.

2.2.2 Criteria for finite support of conditional optimal martingale transport

We start with the one dimensional case. We emphasize that the sufficient condition (i) below corresponds to a local version of [12].

Theorem 2.7. Let $d = 1$ and let $S_0 = \{c_x(x_0, Y) = A(Y)\}$, for some $A \in \text{aff}(\mathbb{R}, \mathbb{R})$, such that $x_0 \in \text{ri conv}S_0$, and $c : \Omega \longrightarrow \mathbb{R}$. 

(i) If \( y \mapsto c_x(x_0, y) \) is strictly convex or strictly concave for some \( x_0 \in \mathbb{R} \), then \( |S_0| \leq 2 \).

(ii) If for all \( y_0 \in \mathbb{R} \), we can find \( k(y_0) \geq 2 \) such that \( y \mapsto c_x(x_0, y_0) \) is \( k(y_0) \) times differentiable in \( y_0 \) and \( c_{xy^k}(x_0, y_0) \neq 0 \), then \( S_0 \) is discrete. If furthermore \( c_x(x_0, \cdot) \) is super-linear in \( y \), then \( S_0 \) is finite.

**Proof.** (i) The intersection of a strictly convex or concave curve with a line is two points or one if they intersect.

(ii) We suppose that \( S_0 \) is not discrete. Then we have \( (y_n) \in S_0^N \) a sequence of distinct elements converging to \( y_0 \in \mathbb{R} \). In \( y_0 \), \( f : y \mapsto c_x(x_0, y) \) is \( k \) times differentiable for some \( k \geq 2 \) and \( f^{(k)}(y_0) = c_{xy^k}(x_0, y_0) \neq 0 \). We have \( f(y_n) = A(y_n) \). Passing to the limit \( y_n \to y_0 \) we get \( f(y_0) = A(y_0) \). Now we subtract and get \( f(y_n) - f(y_0) = \nabla A(y_n - y_0) \). We finally apply Taylor-Young around \( y_0 \) to get

\[
(f'(y_0) - \nabla A)(y_n - y_0) + \sum_{i=2}^k \frac{f^{(i)}(y_0)}{i!}(y_n - y_0)^i + o(|y_n - y_0|^k) = 0
\]

This is impossible for \( y_n \) close enough to \( y_0 \), as one of the terms of the expansion at least is nonzero. If furthermore \( c_x(x_0, \cdot) \) is super-linear in \( y \), \( S_0 \) is bounded, and therefore finite. \( \square \)

Our next result is a weaker version of Theorem 2.7 (i) in higher dimension.

**Proposition 2.8.** Let \( x_0 \in \mathbb{R}^d \) such that for \( y \in \mathbb{R}^d \), \( c_x(x_0, y) = \sum_{i=1}^d P_i(y)u_i \), with for \( 1 \leq i \leq d \), \( P_i \in \mathbb{R}[Y_1, ..., Y_d] \) and \( (u_i)_{1 \leq i \leq d} \) a basis of \( \mathbb{R}^d \). We suppose that the \( P_i \) have degrees \( \deg(P_i) \geq 2 \) and \( (P_i)_{1 \leq i \leq d} \) is a transversal family of \( \mathbb{R}[Y_1, ..., Y_d] \). Then if \( S_0 = \{c_x(x_0, Y) = A(Y)\} \) for some \( x_0 \in \mathbb{R}^d \), and \( A \in \text{Aff}_d \), we have

\[
|S_0| \leq \deg(P_1) \ldots \deg(P_d).
\]

The proof of this proposition is reported in Subsection 3.2.

**Remark 2.9.** This bound is optimal as we see with the example: \( P_i = (Y_i - 1)(Y_i - 2) \ldots (Y_i - k_i) \), for \( 1 \leq i \leq d \). Then \( \{1, 2, \ldots, k_i\} \times \ldots \times \{1, \ldots, k_d\} = \{c_x(x_0, Y) = A(Y)\} \). (For \( A = 0 \)) And this set has cardinal \( k_1 \ldots k_d = \deg(P_1) \ldots \deg(P_d) \). But this bound is not always reached when we fix the polynomials as we can see in the example \( d = 1 \) and \( P = X^4 \), we can add any affine function to it, it will never have more than 2 real zeros even if its degree is 4.

The following example illustrates this theorem in dimension 2.

**Example 2.10.** Let \( d = 2 \) and \( c : (x, y) \in \mathbb{R}^2 \times \mathbb{R}^2 \mapsto x_1(y_1^2 + 2y_2^2) + x_2(2y_1^2 + y_2^2) \). Then \( c_x(x, y) = (y_1^2 + 2y_2^2)e_1 + (2y_1^2 + y_2^2)e_2 \) for all \( (x, y) \), where \( (e_1, e_2) \) is the canonical basis of \( \mathbb{R}^2 \). Let \( A \in \text{Aff}_2 \), \( A = A_1e_1 + A_2e_2 \). The equation \( c_x(x_0, y) = A(y) \) can be written

\[
\begin{align*}
\begin{cases}
y_1^2 + 2y_2^2 = A_1(e_1)y_1 + A_1(e_2)y_2 + A_1(0) \\
n_1^2 + y_2^2 = A_2(e_1)y_1 + A_2(e_2)y_2 + A_2(0).
\end{cases}
\end{align*}
\]

These equations are equations of ellipses \( C_1 \) of axes ratio \( \sqrt{2} \) oriented along \( e_1 \), and \( C_2 \) of axes ratio \( \sqrt{2} \) oriented along \( e_2 \). Then we see visually on Figure 1 that in the nondegenerate case, \( C_1 \) and \( C_2 \) are determined by three affine independent points \( y_1, y_2, y_3 \in \{c_x(x_0, Y) = A(Y)\} \), and that a fourth point \( y' \) naturally appears in the intersection of the ellipses.
Figure 1: Solution of \( c_x(x_0, Y) = A(Y) \) for \( c(x, y) = x_1(y_1^2 + 2y_2^2) + x_2(2y_1^2 + y_2^2) \).

Now we give a general result. If \( k \geq 1 \), we denote

\[
c_{x_i,y^k}(x_0,y_0)[Y^k] := \sum_{1 \leq j_1, \ldots, j_k \leq d} c_{x_i,y^k}(x_0,y_0)Y_{j_1} \ldots Y_{j_k},
\]

the homogeneous multivariate polynomial of degree \( k \) associated to the Taylor term of the expansion of the map \( c_x(x_0, \cdot) \) around \( y_0 \) for \( 1 \leq i \leq d \).

We now provide the extension of Theorem 2.7 (ii) to higher dimension.

**Theorem 2.11.** Let \( x_0 \in \mathbb{R}^d \) and \( S_0 = \{c_x(x_0, Y) = A(Y)\} \) for some \( A \in \text{Aff}_d \). Assume that for all \( y_0 \in \mathbb{R}^d \) and any \( 1 \leq i \leq d \), \( c_{x_i}(x_0, \cdot) \) is \( k_i \geq 2 \) times differentiable at the point \( y_0 \) and that \( \left(c_{x_i,y^k}(x_0,y_0)[Y^{k_i}]\right)_{1 \leq i \leq d} \) is a transversal family of \( \mathbb{R}[Y_1, \ldots, Y_d] \), then \( S_0 \) consists of isolated points. If furthermore \( c_{x_i}(x_0, \cdot) \) is super-linear in \( y \), then \( S_0 \) is finite.

The proof of this theorem is reported in Subsection 3.2.

### 2.3 Largest support of conditional optimal martingale transport plan

The previous section provides a bound on the cardinal of the set \( S_0 \) in the polynomial case, which could be converted to a local result for a sufficiently smooth function, as it behaves locally like a multivariate polynomial. However, with the converse statement of the structure Theorem 2.2, we may also bound this cardinality from below. Let \( c \) be a \( C^{1,2} \) coupling function, and \( x_0 \in \mathbb{R}^d \), we denote

\[
N_c(x_0) := \sup_{P \in \mathbb{R}[Y_1, \ldots, Y_d]^d} |Z^1_\mathbb{R}(H_c(x_0) + P)|, \quad \text{where} \quad H_c(x_0) := \left(c_{x_i,y^2}(x_0,x_0)[Y^2]\right)_{1 \leq i \leq d}.
\]

where we denote by \( Z^1_\mathbb{R}(Q_1, \ldots, Q_d) \) the set of real (finite) single common zeros of the multivariate polynomials \( Q_1, \ldots, Q_d \in \mathbb{R}[Y_1, \ldots, Y_d] \).

**Definition 2.12.** We say that \( c \) is second order transversal at \( x_0 \in \mathbb{R}^d \) if \( c \) is differentiable at \( x = x_0 \) and twice differentiable at \( y = x_0 \), and \( H_c(x_0) \) is a transversal family of \( \mathbb{R}[Y_1, \ldots, Y_d] \).
Theorem 2.13. Let \( c : \Omega \rightarrow \mathbb{R} \) be second order transversal and \( C^{2,0} \cap C^{1,3} \) in the neighborhood of \((x_0, x_0)\) for some \( x_0 \in \mathbb{R}^d \). Then, we may find \( \mu_0, \nu_0 \in \mathcal{P}(\mathbb{R}^d) \) with \( C^1 \) densities, and a unique \( \mathbb{P}^* \in \mathcal{M}(\mu_0, \nu_0) \) such that

\[
S_{\mu_0, \nu_0}(c) = \mathbb{P}^*[c] \quad \text{and} \quad \text{supp} \mathbb{P}^* = N_c(x_0), \mu - \text{a.s.}
\]

The proof of this result is reported in subsection 3.3.

Theorem 2.13 shows the importance of the determination of the numbers \( N_c(x_0) \). We know by Remark 2.9 that for some coupling \( c : \Omega \rightarrow \mathbb{R} \), the upper bound is reached: \( N_c(x_0) = 2^d \). We conjecture that this bound is reached for all coupling which is second order transversal at \( x_0 \). An important question is whether there exists a criterion on coupling functions to have the differential intersection limited to \( d + 1 \) points, similarly to the Spence-Mirless condition in one dimension. It has been conjectured in [9] in the case of minimization for the distance coupling. Theorem 2.17 together with Theorem 2.2 proves that this conjecture is wrong. Now we prove that even for much more general second order transversal coupling functions, there is no hope to find such a criterion for \( d \) even.

Theorem 2.14. Let \( x_0 \in \mathbb{R}^d \), and \( c \) second order transversal and \( C^{1,2} \) at \((x_0, x_0)\), then

\[
d + 1 + 1_{\{d \text{ even}\}} \leq N_c(x_0) \leq 2^d.
\]

2.4 Support of optimal plans for classical couplings

2.4.1 Euclidean distance based coupling functions

Theorem 2.2 shows the importance of sets \( S_0 = \{c_x(x_0, Y) = A(Y)\} \) for \( x_0 \in \text{ri conv } S_0 \), and \( A \in \text{Aff}_d \). We can characterize them precisely when \( c : (x, y) \in \mathbb{R}^d \times \mathbb{R}^d \rightarrow f(|x - y|) \) for some \( f \in C^1(\mathbb{R}^+, \mathbb{R}) \). In view of Remark 2.3, the following result gives the structure of \( S_0 \) as a function of \( d + 1 \) known points in this set. Let \( g : t > 0 \rightarrow -f'(t)/t \), notice that

\[
c_x(x, y) = g(|y - x|)(y - x), \quad \text{on } \{X \neq Y\}.
\]

Furthermore, \( c(x, y) \) is differentiable in \( x = y \) if and only if \( f'(0) = 0 \), in this case \( c_x(x, x) = 0 \). We fix \( S_0 := \{c_x(x_0, Y) = A(Y)\} \), for some \( x_0 \in \text{int conv } S_0 \), and \( A \in \text{Aff}_d \). The next theorem gives \( S_0 \) as a function of \( A \) and \( x_0 \). For \( a \notin Sp(\nabla A) \), let \( y(a) := x_0 + (aI_d - \nabla A)^{-1}A(x_0) \). For \( a \in Sp(\nabla A) \), if the limit exists, we write \( |y(a)| < \infty \) and denote \( y(a) := \lim_{t \to a} y(t) \).

Theorem 2.15. Let \( S_0 := \{c_x(x_0, Y) = A(Y)\} \) for \( x_0 \in \text{ri conv } S_0 \), and \( A \in \text{Aff}_d \). Then

\[
S_0 = \cup_{(t, \rho) \in A} S^0_t \cup \{y(t) : t \in \text{fix}(g \circ |y - x_0|)\},
\]

where \( S^0_t := S_{V_t} \left( p_t, \sqrt{\rho^2 - |p_t - x_0|^2} \right) \), with \( V_t := y(t) + \ker(tI_d - \nabla A) \), \( p_t := \text{proj}_{V_t}(x_0) \), and \( A := \{(t, \rho) : t \in Sp(\nabla A), |y(t)| < \infty, g(\rho) = t, \text{ and } \rho \geq |p_t - x_0|\} \).
Therefore, by Theorem 2.2, we may find
\[ S(i) \]
(i) The elements in the spheres \( y \) and the optimization reduces to a minimization with the marginals chosen by the problem: Theorem 4.2 in [19] states that the mass optimizer \( P \) and if \( c \) and \( d \) are still holds for \( 2 \) are trivial, for other values, we have the following theorems.

**Corollary 2.16.** \( S_0 \) contains at least \( 2d \) possibly degenerate points counted with multiplicity.

The proofs of Theorem 2.15 and Corollary 2.16 are reported in Subsection 3.5.

### 2.4.2 Powers of Euclidean distance cost

In this section we provide calculations in the case where \( f \) is a power function. The particular cases \( p = 0, 2 \) are trivial, for other values, we have the following theorems.

**Theorem 2.17.** Let \( c := |X - Y|^p \). Let \( S_0 := \{c_x(x_0, Y) = A(Y)\} \), for some \( x_0 \in \text{int conv} \ S_0 \), and \( A \in \text{Aff}_d \). Then if \( p \leq 1 \), \( S_0 \) contains \( 2d \) possibly degenerate points counted with multiplicity, and if \( 1 < p < 2 - \frac{2}{n} \) or \( p > 2 + \frac{2}{n} \), \( S_0 \) contains \( 2d + 1 \) possibly degenerate points counted with multiplicity.

The proof of this theorem is reported in Subsection 3.5.

**Remark 2.18.** In both cases, for almost all choice of \( y_0, \ldots, y_d \in \mathbb{R}^d \) as the first elements of \( S_0 \), determining the Affine mapping \( A \), we have \( d_i = 0 \) for all \( i \), and \( c_{xy}(x_0, S_0) - \nabla A \subset GL_d(\mathbb{R}^d) \). Then for \( -\infty < p \leq 1 \), and \( p \neq 0 \), \( |S_0| = 2d \), and for \( 1 < p < 2 - \frac{2}{n} \) or \( p > 2 + \frac{2}{n} \), \( |S_0| = 2d + 1 \). Therefore, by Theorem 2.12, we may find \( \mu, \nu \in \mathcal{P}(\mathbb{R}^d) \) with \( C^1 \) densities such that the associated optimizer \( \mathbb{P} \in \mathcal{M}(\mu, \nu) \) of the MOT problem [1,2] satisfies \( |\text{supp} \mathbb{P}_X| = 2d, \mu-\text{a.s.} \) if \( p \leq 1 \), and \( |\text{supp} \mathbb{P}_X| = 2d + 1, \mu-\text{a.s.} \) if \( p > 1 \).

**Remark 2.19.** Based on numerical experiments, we conjecture that the result of Theorem 2.17 still holds for \( 2 - \frac{2}{n} \leq p \leq 2 + \frac{2}{n} \), and \( p \neq 2 \). See Section 4.

**Remark 2.20.** Assumption 2.12 implies that \( c \) is subdifferentiable. Then we can deal with coupling functions \( c := -|X - Y|^p \) with \( 0 < p \leq 1 \) only by evacuating the problem on \{\( X = Y \}\}. If \( 0 < p \leq 1 \), it was proved by Lim [19] that in this case the value \{\( X = Y \)\} is preferentially chosen by the problem: Theorem 4.2 in [19] states that the mass \( \mu \land \nu \) stays put (i.e. this common mass of \( \mu \) and \( \nu \) is concentrated on the diagonal \{\( X = Y \)\} by the optimal coupling) and the optimization reduces to a minimization with the marginals \( \mu - \mu \land \nu \) and \( \nu - \mu \land \nu \). Therefore, \( c \) is differentiable on all the points concerned by this other optimization, and the supports are given by \( \text{supp} \mathbb{P}_x \subset \{c_x(x, Y) = A_x(Y)\} \cup \{x\} \), for \( \mu-\text{a.e.} \) \( x \in \mathbb{R}^d \). Then the supports are exactly given by the ones from the maximisation case with eventually adding the diagonal.
Notice that Remark 2.20 together with Theorem 2.2 and Theorem 2.17 prove that Conjecture 2 in [9] is wrong, and explains the counterexample found by Lim [20], Example 2.9.

2.4.3 One and infinity norm coupling

For \( \varepsilon \in \mathcal{E}^1 := \{-1, 1\}^d \), we denote \( Q^1_\varepsilon := \prod_{1 \leq i \leq d} \varepsilon_i (0, \infty) \) the quadrant corresponding to the sign vector \( \varepsilon \). Similarly, for \( \varepsilon \in \mathcal{E}^\infty := \{\pm \varepsilon_i\}_{1 \leq i \leq d} \), we denote \( Q^\infty_\varepsilon := \{y \in \mathbb{R}^d : \varepsilon \cdot y > \|y - (\varepsilon \cdot y)_{\|\cdot\|\infty}\|\} \) the quadrant corresponding to the signed basis vector \( \varepsilon \).

**Proposition 2.21.** Let \( c := |X - Y|_p \) with \( p \in \{1, \infty\} \), and \( S_0 := \{c_x(x_0, Y) = A(Y)\} \) for some \( x_0 \in \text{ri conv} S_0 \), and \( A \in \text{Aff}_d \), with \( r := \text{rank} \nabla A \). Then, we may find \( 2 \leq k \leq 1_p=2^d + 1_{p=\infty}2r \), \( \varepsilon_1, \ldots, \varepsilon_k \in \mathcal{E}^p \), and \( y_1, \ldots, y_k \in \mathbb{R}^d \) such that

\[
S_0 = \bigcup_{i=1}^k (x_0 + Q^p_{\varepsilon_i}) \cap (y_i + \ker A).
\]

In particular, \( S_0 \) is concentrated on the boundary of its convex hull.

This Proposition will be proved in Subsection 3.4. The case \( r = d \) is of particular interest.

**Remark 2.22.** Notice that the gradient of \( c \) is locally constant where it exists (i.e. if \( c \) is differentiable at \( (x_0, y_0) \), then \( c \) is differentiable at \( (x, y) \) and \( \nabla c(x, y) = \nabla c(x_0, y_0) \) for \( (x, y) \) in the neighborhood of \( (x_0, y_0) \)). Then if \( r = d \), \( c(x_0, S_0) - \nabla A = -\nabla A \in GL_d(\mathbb{R}) \), \( S_0 \) is finite and \( |S_0| \leq 1_{p=2^d} + 1_{p=\infty}2d \). The bound is sharp (consider for example \( A := x_0 + I_d \)). Therefore, by Theorem 2.22, we may find \( \mu, \nu \in \mathcal{P}(\mathbb{R}^d) \) with \( C^1 \) densities such that the associated optimizer \( \mathbb{P} \in \mathcal{M}(\mu, \nu) \) of the MOT problem (1.2) satisfies \( |\text{supp} \mathbb{P}_X| = 1_{p=2^d} + 1_{p=\infty}2d \), \( \mu \)-a.s.

2.4.4 Concentration on the Choquet boundary

Recall that a set \( S_0 \) is included in its own Choquet boundary if \( S_0 \subset \text{Ext}(\text{cl conv}(S_0)) \), i.e. any point of \( S_0 \) is extreme in \( \text{cl conv}(S_0) \). A result showed in [9] is that the image of the optimal transport is concentrated in its own Choquet boundary for distance coupling. We prove that this is a consequence of the structure Theorem 2.2 and we generalize this observation to some other cases.

**Proposition 2.23.** Let \( c : \Omega \longrightarrow \mathbb{R} \) be a coupling function, \( A \in \text{Aff}_d \), \( S_0 \subset \{c_x(x_0, Y) = A(Y)\} \), and \( x_0 \in \text{ri conv} S_0 \). \( S_0 \) is concentrated in its own Choquet boundary in the following cases:

(i) the map \( y \mapsto c_x(x_0, y) \cdot u \) is strictly convex for some \( u \in \mathbb{R}^d \); 
(ii) \( c : (x, y) \mapsto |x - y|^p \), with \( 1 < p < \infty \); 
(iii) \( c : (x, y) \mapsto |x - y|^p \), with \( -\infty < p \leq 1 \); 
(iv) \( c : (x, y) \mapsto |x - y|^p \), with \( 1 < p < 2 - \frac{2}{\alpha} \) or \( p > 2 + \frac{2}{\alpha} \), and \( p \min_{y \in S_0} |y - x_0|^p \) is a double root of the polynomial \( \det(\nabla A - XI_d)^2 - |p|^{\frac{1}{p-1}}X^\frac{1}{p-1}|\text{Com}(\nabla A - XI_d)^t A(0)|^2 \).

Furthermore, if \( c : (x, y) \mapsto |x - y|^p \), with \( 1 < p < 2 - \frac{2}{\alpha} \) or \( p > 2 + \frac{2}{\alpha} \), and \( S_0 \) is not concentrated on its own Choquet boundary, then we may find a unique \( y_0 \in S_0 \) such that \( |y_0 - x_0| = \min_{y \in S_0} |y - x_0| \), and \( S_0 \setminus \{y_0\} \) is concentrated on its own Choquet boundary.
The proof of this proposition is reported in Subsection 3.6.

**Remark 2.24.** If $p = 1$ or $p = \infty$, there are counterexamples to Proposition 2.23 (ii), as $S_0$ may contain a non-trivial face of itself, see Proposition 2.21.

## 3 Proofs of the main results

### 3.1 Structure theorem

**Proof of Theorem 2.2** By Theorem 3.5 (i) in [6], (and using the notation therein), the quasiprobabilistic super-hedging problem may be decomposed in pointwise robustness, separate problems attached to each component, and we may find functions $(\varphi, h) \in L^0(\mathbb{R}^d) \times L^0(\mathbb{R}^d, \mathbb{R}^d)$, and $(\psi_K)_{K \in \{0, 1\}} \subset L^0_+(\mathbb{R}^d)$ with $\psi_1(Y) \in L^0_+\{\Omega\}$, and dom $\psi_1 = J_\theta$, $\eta$–a.s. for some $\theta \in \hat{T}(\mu, \nu)$, such that $c \leq \varphi(Y) + \psi_1(Y) + h^\theta$, and $S_{\mu, \nu}(c) = S_{\mu, \nu}(\varphi(Y) + \psi_1(Y) + h^\theta)$. Then applying the theorem to $c' := \varphi(Y) + \psi_1(Y) + h^\theta$, $S_{\mu, \nu}(c) = S_{\mu, \nu}(\varphi(Y) + \psi_1(Y) + h^\theta) = \sup_{\mathbb{P} \in M(\mu, \nu)} S_{\mu, \nu}(\varphi(Y) + \psi_1(Y) + h^\theta)$. Then if $\mathbb{P} \in M(\mu, \nu)$ is optimal for $S_{\mu, \nu}(c)$, then $\mathbb{P} \{c = \varphi(Y) + h^\theta\} = 1$, $\eta$–a.s. By Lemma 3.17 in [6] the regularity of $c$ in Assumption 2.21 (i) guarantees that we may choose $\varphi$ to be locally Lipschitz on $I$, and $h$ locally bounded on $I$. In view of Assumption 2.21 (ii), $\varphi$ is differentiable $\mu_1$–a.e. by the Rademacher Theorem. Then after possibly restricting to an irreducible component, we may suppose that we have the following duality: for any $x, y \in \mathbb{R}^d$,

$$\varphi(x) + \psi(y) + h(x) \cdot (y - x) - c(x, y) \geq 0,$$

with equality if and only if $(x, y) \in \Gamma := \{c \leq \varphi(Y) + h^\theta = c < \infty\}$, concentrating all optimal coupling for $S_{\mu, \nu}(c)$.

Let $x_0 \in \text{ri conv dom } \psi$ such that $\varphi$ is differentiable in $x_0$. Let $y_1, \ldots, y_k \in \Gamma_{x_0}$ such that $\sum_{i=1}^k \lambda_i y_i = x_0$, convex combination. We complete $(y_1, \ldots, y_k)$ in a barycentric basis $(y_1, \ldots, y_k, y_{k+1}, \ldots, y_l)$ of ri conv dom $\psi$. Let $x \in \text{ri conv dom } \psi$ in the neighborhood of $x_0$, and let $(\lambda'_i)$ such that $x = \sum_{i=1}^l \lambda'_i y_i$, convex combination. We apply (3.7), both in the equality and in the inequality case:

$$\varphi(x) + \sum_{i=1}^l \lambda'_i \psi(y_i) \geq \sum_{i=1}^l \lambda'_i c(x, y_i), \quad \varphi(x_0) + \sum_{i=1}^l \lambda'_i \psi(y_i) + h(x_0) \cdot (x - x_0) = \sum_{i=1}^l \lambda'_i c(x_0, y_i).$$

By subtracting these equations, we get

$$\varphi(x) - \varphi(x_0) - h(x_0) \cdot (x - x_0) \geq \sum_{i=1}^l \lambda'_i (c(x, y_i) - c(x_0, y_i)).$$

As $c$ is Lipschitz in $x$, and $\lambda'_i \rightarrow \lambda_i$ when $x \rightarrow x_0$, we get:

$$(\nabla \varphi(x_0) - h(x_0)) \cdot (x - x_0) + o(x - x_0) \geq \sum_{i=1}^k \lambda_i (c(x, y_i) - c(x_0, y_i)).$$
Then, $x \mapsto \sum_{i=1}^{k} \lambda_i c(x, y_i)$ is super-differentiable at $x_0$, and $\nabla \varphi(x_0) - h(x_0)$ belongs to its super-gradient. As $x \mapsto c(x, y)$ is sub-differentiable by Assumption 2.11 (i), it implies that $x \mapsto c(x, y_i)$ is differentiable at $x_0$ for all $i$ such that $\lambda_i > 0$, and therefore

$$\nabla \varphi(x_0) - h(x_0) = \sum_{i=1}^{k} \lambda_i c(x_0, y_i).$$

(3.8)

Now we want to prove that we may find $A_x \in \text{Aff}_d$ such that $A_x(y) = c_x(x, y)$ for all $y \in \Gamma_x$.

Let $y_1^0, \ldots, y_m^0 \in \Gamma_{x_0}$ generating $\text{aff} \Gamma_{x_0}$ and such that $x \in \text{ri conv}(y_1^0, \ldots, y_m^0)$, let $y \in \Gamma_{x_0}$. $A_x$ is defined in a unique way if $\nabla A = 0$ on $(\text{aff} \Gamma_{x_0} - x_0)^\perp$ by its values on $(y_1^0, \ldots, y_m^0)$. Now we prove that $A_x(y) = c_x(x_0, y)$. As $y \in \text{aff}(y_1^0, \ldots, y_m^0)$, we may find $(\mu_i)$ so that $\sum_{i=1}^{m} \mu_i y_i^0 = y$, and $\sum_{i=1}^{m} \mu_i = 1$. For $\varepsilon > 0$ small enough, $x_0 - \varepsilon(y - x_0) \in \text{ri conv}(y_1^0, \ldots, y_m^0)$. Then $x_0 - \varepsilon(y - x_0) = \sum_{i=1}^{m} \lambda_i y_i$ with $\lambda_i^0 y_i$ with $\lambda_i^0 > 0$. We take the convex combination: $x_0 = \frac{1}{1+\varepsilon} (x_0 - \varepsilon(y - x_0)) + \frac{\varepsilon}{1+\varepsilon} y$, and $x_0 = \sum_{i=1}^{m} (\frac{1}{1+\varepsilon} \lambda_i^0 + \frac{\varepsilon}{1+\varepsilon} \mu_i) y_i^0$. We suppose that $\varepsilon$ is small enough so that $\lambda_i^0 := \frac{1}{1+\varepsilon} \lambda_i^0 + \frac{\varepsilon}{1+\varepsilon} \mu_i > 0$. Then applying (3.8) for $(y_i) = (y_i^0)$ and $(\lambda_i) = (\lambda_i^0)$,

$$\nabla \varphi(x_0) - h(x_0) = \sum_{i=1}^{m} \lambda_i \varphi_{x_0}(x_0, y_i) = \sum_{i=1}^{m} \frac{1}{1+\varepsilon} \lambda_i c(x_0, y_i) + \frac{\varepsilon}{1+\varepsilon} c(x_0, y).$$

By subtracting, we get $c_x(x_0, y) = A_{x_0} \left( \frac{1+\varepsilon}{\varepsilon} \sum_{i=1}^{m} (\lambda_i^0 - \frac{\varepsilon}{1+\varepsilon} \lambda_i) y_i \right) = A_{x_0}(y)$. Now doing this for all $x \in \mathbb{R}^d$ so that $\varphi$ is differentiable in $x$, by domination of $\mu_{I^1}$ by Lebesgue, this holds for $\mu_{I^1-} a.e. x \in \mathbb{R}^d$, $\eta$-a.s. and therefore $\mu-a.s.$.

Now we prove the converse statement. Let $S_0 \subset \{ A(Y) = c_x(x_0, Y) \}$ be a closed bounded subset of $\Omega$ for some $x_0 \in \mathbb{R}^d$, and $A \in \text{Aff}_d$ such that $x_0 \in \text{int conv} S_0$, $c$ is $C^2_0 \cap C^{1,1}$ in the neighborhood of $S_0$, and $c_{xy}(S_0) - \nabla A \subset GL_d(\mathbb{R})$. First, we show that $S_0$ is finite. Indeed, we suppose to the contrary that $|S_0| = \infty$, we can find a sequence $(y_n)_{n \geq 1} \subset S_0$ with distinct elements. As $S_0$ is closed bounded, and therefore compact, we may extract a subsequence $(y_{n(i)})$ converging to $y_i \in S_0$. We have $c_{xy}(x, y_{n(i)}) = A(y_{n(i)})$, and $c_{xy}(x, y_i) = A(y_i)$. We subtract and get $c_{xy}(x, y_{n(i)}) - c_{xy}(x, y_i) = \nabla A(y_{n(i)} - y_i) = 0$, and using Taylor-Young around $y_i$, $c_{xy}(x, y_i)(y_{n(i)} - y_i) + o(|y_{n(i)} - y_i|) - \nabla A(y_{n(i)} - y_i) = 0$. As $y_{n(i)} \neq y_i$ for $n$ large enough , we may write $u_n := \frac{y_{n(i)} - y_i}{|y_{n(i)} - y_i|}$. As $u_n$ stands in the unit sphere which is compact, we can extract a subsequence $(u_{n(i)})$, converging to a unit vector $u$. As we have $c_{xy}(x, y_i) u_{n(i)} = o(1)$ - $\nabla A u_{n(i)} = 0$, we may pass to the limit $n \to \infty$, and get:

$$(c_{xy}(x_0, y) - \nabla A) u = 0.$$ 

As $u \neq 0$, we get the contradiction: $c_{xy}(x_0, y) - \nabla A \notin GL_d(\mathbb{R})$.

Now, we denote $S_0 = \{ y_i \}_{1 \leq i \leq k}$ where $k := |S_0|$. For $r > 0$ small enough, the balls $B((x_0, y_i), r)$ are disjoint, $c_{xy}(\cdot) - \nabla A \subset GL_d(\mathbb{R})$ on these balls by continuity of the determinant, and $c$ is $C^2_0 \cap C^{1,1}$ on these balls. Now we define appropriate dual functions. Let $M > 0$ large enough so that on the balls, $(M - 1)I_d - (\nabla A + \nabla A^T) - c_{xx}$ is positive semidefinite.

We set $h(X) := (X - x_0)^T \nabla A - A(x_0)$, and $\varphi(X) := \frac{1}{2} M|X - x_0|^2$. Now for $1 \leq i \leq k$, $c_x(x, y_i) - \nabla A : (y_i - x_0) = \nabla \varphi(x_0) - h(x_0), (x, y) \mapsto c_x(x, y) - \nabla A : (y - x) is C^1$, and its partial
derivative with respect to $y$, $c_{xy} - \nabla A$ is invertible on the balls. Then by the implicit functions Theorem, we may find a mapping $T_i \in C^1(\mathbb{R}^d, \mathbb{R}^d)$ such that for $x \in \mathbb{R}^d$ in the neighborhood of $x_0$,

$$c_x(x, T_i(x)) - \nabla A \cdot (T_i(x) - x) = \nabla \varphi(x) - h(x). \quad (3.9)$$

Its gradient at $x_0$ is given by $\nabla T_i(x_0) = (c_{xy}(x_0, y_i) - \nabla A)^{-1}(MI_d - (\nabla A + \nabla A^t) - c_{xx}(x_0, y_i))$. This matrix is invertible, and therefore by the local inversion theorem, $T_i$ is locally a $C^1$-diffeomorphism in the neighborhood of $x_0$. We shrink the radius $r$ of the balls so that each $T_i$ is a diffeomorphism on $B := X(\bar{B}((x_0, y_i), r))$ (independent of $i$). Let $B_i := T_i(B)$, for $y \in B_i$, let $\psi(y) := c(T_i^{-1}(y), y) - \varphi(T_i^{-1}(y)) - h(T_i^{-1}(y)) \cdot (y - T_i^{-1}(y))$. These definitions are not interfering, as we supposed that the balls $B_i$ are not overlapping.

Let $\Gamma := \{(x, T_i(x)) : x \in B, 1 \leq i \leq k\}$. By definition of $\psi$, $c = \varphi \oplus \psi + h^\otimes$ on $\Gamma$. Now let $(x, y) \in B \times B_i$, for some $i$. $(x_0, y) \in \Gamma$, for some $x_0 \in B$. Let $F := \varphi \oplus \psi + h^\otimes - c$, we prove now that $F(x, y) \geq 0$, with equality if and only if $x = x_0$ (i.e. $(x, y) \in \Gamma$). $F(x_0, y) = 0$, and $F_x(x_0, y) = 0$ by (3.9). However, $F_{xx}(X, Y) = MI_d - (\nabla A + \nabla A^t) - c_{xx}(X, Y)$ which is positive definite on $B \times B_i$, and therefore we get

$$F(x, y) = F(x, y) - F(x_0, y) = \int_{x_0}^x F_x(z, y) \cdot dz = \int_{x_0}^x (F_x(z, y) - F_x(x_0, y)) \cdot dz$$

$$= \int_{x_0}^x \int_{x_0}^z dw \cdot F_{xx}(w, y) \cdot dz \geq 0.$$

Where the last inequality follows from the fact that $F_{xx}$ is positive definite and $dw$ and $dz$ are two vectors collinear with $(x - x_0)$. It also proves that $F(x, y) = 0$ if and only if $(x, y) \in \Gamma$.

Now, we define $C^1$ mappings $\lambda_i : B \rightarrow (0, 1]$ such that $\sum_{i=1}^k \lambda_i(x) T_i(x) = x$. We may do this because we assumed that $x \in \text{int conv } S_0$, and therefore, by continuity, up to reducing $B$ again, $x \in \text{int conv } \{T_1(x), ..., T_k(x)\}$ for all $x \in B$. Finally let $\mu_0 \in \mathcal{P}(\mathbb{R}^d)$ such that $\text{supp } \mu_0 = B$ with $C^\infty$ density $f$ (take for example a well chosen wavelet). Now for $1 \leq i \leq k$, we define $\nu_0$ on $B_i$ by $\nu_0(dy) = \lambda_i(T_i^{-1}(y)) f(T_i^{-1}(y)) |det \nabla T_i(T_i^{-1}(y))|^{-1}$. Then $P(dx, dy) := \mu_0(dx) \otimes \sum_{i=1}^k \lambda_i(x) \delta_{T_i(x)}(dy)$ is supported on $\Gamma$, is in $\mathcal{M}(\mu_0, \nu_0)$. As $\varphi$, and $\psi$ are continuous, and therefore bounded, and as $\mu_0$ and $\nu_0$ are compactly supported, $P^*[c] = \mu_0[\varphi] + \nu_0[\psi]$, and therefore $P^*$ is an optimizer for $S_{\mu_0, \nu_0}(c)$.

Now we prove that this is the only optimizer. Let $P$ be an optimizer for $S_{\mu_0, \nu_0}(c)$. Then $P[\Gamma] = 1$, and therefore $P(dx, dy) = \mu_0(dx) \otimes \sum_{i=1}^k \gamma_i(x) \delta_{T_i(x)}(dy)$, for some mappings $\gamma_i$. Let $1 \leq i \leq k$, as for $y \in B_i$, there is only one $x := T_i^{-1}(y) \in B$ such that $(x, y) \in \Gamma$. Then we may apply the Jacobian formula: $\nu_0(dy) = \gamma_i(T_i^{-1}(y)) f(T_i^{-1}(y)) |det \nabla T_i(T_i^{-1}(y))|^{-1}$. As this density in also equal to $\nu_0(dy) = \lambda_i(T_i^{-1}(y)) f(T_i^{-1}(y)) |det \nabla T_i(T_i^{-1}(y))|^{-1}$, and as $f(T_i^{-1}(y)) |det \nabla T_i(T_i^{-1}(y))|^{-1} > 0$, we deduce that $\lambda_i(T^{-1}(Y)) = \gamma_i(T^{-1}(Y))$, $\nu_0$-a.s. and $\lambda_i = \gamma_i, \mu_0$-a.s. and therefore $P = P^*$. \hfill $\square$

**Remark 3.1.** We have $A_x(x) = \nabla \varphi(x) - h(x)$ in the previous proof. Under the stronger assumption that $\varphi$ and $h$ are $C^1$, we can get the previous result much easier. As for $(x, y) \in \mathbb{R}^d$,

$$\varphi(x) + \psi(y) + h(x) \cdot (y - x) - c(x, y) \geq 0,$$
with equality for \((x, y) \in \Gamma\). When \(y_0\) is fixed, \(x_0\) such that \((x_0, y_0) \in \Gamma\) is a critical point of \(x \mapsto \varphi(x)+\psi(y_0)+h(x)-(y_0-x)-c(x, y_0)\). Then we get \(c_x(x_0, y_0) = \nabla h(x_0)(y_0-x_0) + \nabla \varphi(x_0) - h(x_0)\)
by the first order condition.

We see that we have in this case \(A_{x_0}(y) := \nabla h(x_0)(y - x_0) + \nabla \varphi(x_0) - h(x_0)\), and \(\Gamma_{x_0} \subset \{c_x(x_0, Y) = A_{x_0}(Y)\}\), for \(\mu\)--a.e. \(x_0 \in \mathbb{R}^d\).

### 3.2 Proof of the support cardinality bounds

We first introduce some notions of Algebraic geometry. Recall \(\mathbb{P}^d := (\mathbb{C}^{d+1})^*/\mathbb{C}^*\), the \(d\)--dimensional projective space which complements the space with points at infinity. Recall that there is an isomorphism \(\mathbb{P}^d \cong \mathbb{C}^d \cup \mathbb{P}^{d-1}\), where \(\mathbb{P}^{d-1}\) are the 'points at infinity'. Then we may consider the points for which \(x_0 = 0\) as 'at infinity' because the surjection of \(\mathbb{P}^d\) in \(\mathbb{C}^d\) is given by \((x_0, x_1, ..., x_d) \mapsto (x_1/x_0, ..., x_d/x_0)\) so that when \(x_0 = 0\), we formally divide by zero and then consider that the point is sent to infinity. The isomorphism \(\mathbb{P}^d \cong \mathbb{C}^d \cup \mathbb{P}^{d-1}\) follows from the easy decomposition:

\[
\mathbb{P}^d = \{(x_0, ..., x_d) \in \mathbb{C}^{d+1}, x_0 \neq 0\}/\mathbb{C}^* \cup \{(0, x_1, ..., x_d), (x_1, ..., x_d) \in \mathbb{C}^d \setminus \{0\}\}/\mathbb{C}^*
\]

\[
= \{(1, x_1/x_0, ..., x_d/x_0), (x_0, ..., x_d) \in \mathbb{C}^{d+1}, x_0 \neq 0\}
\]

\[
\cup \{(0, x_1, ..., x_d), (x_1, ..., x_d) \in (\mathbb{C}^d)^*/\mathbb{C}^*\}
\]

\[
\cong \mathbb{C}^d \cup (\mathbb{C}^d)^*/\mathbb{C}^* \cong \mathbb{C}^d \cup \mathbb{P}^{d-1}.
\]

The points in the projective space \(\mathbb{P}^d\) in the equivalence class of \(\{x_0 = 0\}\) are called points at infinity.

**Definition 3.2.** The map \(P = \sum_{n \in \mathbb{N}_0, |n| \leq \deg(P)} a_n X^n \mapsto P^{\mathrm{proj}} := \sum_{n \in \mathbb{N}_0, |n| \leq \deg(P)} a_n X^n X_0^{\deg(P) - |n|}
defines an isomorphism between \(\mathbb{C}[X_1, ..., X_d]\) and \(\mathbb{C}^{\mathrm{hom}}[X_0, X_1, ..., X_d]\). Let \((P_1, ..., P_k)\) be \(k \geq 1\) polynomials in \(\mathbb{R}[X_1, ..., X_d]\), we define the set of common projective zeros of \((P_1, ..., P_k)\) by \(Z^{\mathrm{proj}}(P_1, ..., P_k) := Z(P_1^{\mathrm{proj}}, ..., P_k^{\mathrm{proj}})\).

This allows us to define the zeros of a nonhomogeneous polynomial in the projective space.

We finally report the following well-known result which will be needed for the proofs of Proposition 2.3 and Theorem 2.11.

**Theorem 3.3** (Bezout). Let \(d \in \mathbb{N}\) and \((P_1, ..., P_d)\) be \(d\) transversal polynomials in \(\mathbb{R}[X_1, ..., X_d]\). Then \(|Z^{\mathrm{proj}}(P_1, ..., P_d)| = \deg(P_1) \cdots \deg(P_d)\), where the roots are counted with multiplicity.

See Hartshorne [11] or Harris [10].

Notice that we have the identity \(P^{\mathrm{hom}} = P^{\mathrm{proj}}(X_0 = 0)\). Then \(P^{\mathrm{hom}}\) may be interpreted as the restriction to infinity of \(P^{\mathrm{proj}}\) and we deduce the following characterization of transversality. We believe that this is a standard algebraic geometry result, but we could not find precise references. For this reason, we report the proof for completeness.

**Proposition 3.4.** Let \(P_1, ..., P_d \in \mathbb{R}[X_1, ..., X_d]\). Then the following assertions are equivalent:
(i) \((P_1, \ldots, P_d)\) is transversal;
(ii) \(Z^{\text{proj}}(P_1, \ldots, P_d)\) contains no points at infinity;
(iii) \(Z(P_1^{\text{hom}}, \ldots, P_d^{\text{hom}}) = \{0\}\).

**Proof.** The equivalence (ii) \(\iff\) (iii) immediately follows from the fact that in the projective space, a point at infinity \(x \in \mathbb{P}^d\) is characterized by \(x_0 = 0\), thus annihilating the coefficients of \(P_1, \ldots, P_d\) that does not have maximum degree.

Now we prove (i) \(\implies\) (iii). By definition of transversality, we have that \((P_1^{\text{hom}}, \ldots, P_d^{\text{hom}})\) is transversal by the fact that \((P_1, \ldots, P_d)\) is transversal. By Theorem 3.3, \((P_1^{\text{hom}}, \ldots, P_d^{\text{hom}})\) has exactly \(\deg P_1^{\text{hom}} \cdots \deg P_d^{\text{hom}}\) common roots counted with multiplicity. However, by their homogeneity property, 0 is a root of order \(\deg P_1^{\text{hom}} \cdots \deg P_d^{\text{hom}}\), therefore it is the only common root of these multivariate polynomials.

Finally we prove that (iii) \(\implies\) (i). In order to prove this implication, we assume to the contrary that (i) does not hold. Then one of the polynomials among \((P_1^{\text{hom}}, \ldots, P_d^{\text{hom}})\) belongs to the ideal generated by the others. We assume without loss of generality that this polynomial is \(P_1^{\text{hom}}\). Then \(Z^{\text{proj}}(P_1^{\text{hom}}, \ldots, P_d^{\text{hom}}) = Z^{\text{proj}}(P_1^{\text{hom}}, \ldots, P_{d-1}^{\text{hom}})\), and the dimension of this projective variety is higher than \(d - (d - 1) = 1\). Then we may find some \(x \in Z^{\text{proj}}(P_1^{\text{hom}}, \ldots, P_{d-1}^{\text{hom}})\) which is different from \(z = (1, 0, \ldots, 0)\), as if \(z\) was the only zero, the dimension of \(Z^{\text{proj}}(P_1^{\text{hom}}, \ldots, P_{d-1}^{\text{hom}})\) would be 0. Now we consider \(x' := (1, x_1, \ldots, x_d) \in \mathbb{P}^d\). As the polynomial \(P_1^{\text{hom}}\) are homogeneous, the value of the first coefficient does not change their value. Therefore, \(x' \in Z^{\text{proj}}(P_1^{\text{hom}}, \ldots, P_{d-1}^{\text{hom}}) = Z^{\text{proj}}(P_1^{\text{hom}}, \ldots, P_{d-1}^{\text{hom}})\). Finally as we assumed that \(x \neq z\), we have that \(x' \neq z\). As \(x'\) is non-zero and not at infinity by the fact that \(x_0 = 1\), \(x'\) is a finite non-zero common root in \(Z(P_1^{\text{hom}}, \ldots, P_d^{\text{hom}})\), therefore contradicting (iii).

**Proof of Proposition 2.28** We suppose that \((P_i)_{1 \leq i \leq d}\) is a transversal family of \(\mathbb{R}[Y_1, \ldots, Y_d]\). For \(1 \leq i \leq d\), let \(A_i := u_i \cdot A \in \text{aff}(\mathbb{R}^d, \mathbb{R})\). If we project on the vectors \(u_i\) we get:

\[P_i(y) = A_i(y), \quad i = 1, \ldots, d.\]

Thanks to the transversality of \((P_i)\), the \(\tilde{P}_i\) which are defined on the projective space \(\mathbb{P}^d\) for \(1 \leq i \leq d\) by

\[\tilde{P}_i(Z_0, Z_1, \ldots, Z_d) := P_i^{\text{proj}}(Z_1, \ldots, Z_d) + \nabla A_i Z_0^{k-1} + A_i(0)Z_0^k\]

also form a transversal family: we see that this fact holds by computing these polynomials at \(Z_0 = 0\). By Bezout Theorem 3.3 there are only \(k^d\) roots to this polynomial. They may be complex, infinite, or multiple, but the upper bound holds.

**Proof of Theorem 2.11** We suppose that \(S_0\) is not discrete. Then we have \((y_n) \in S_0^n\) a sequence of distinct elements converging to \(y_0 \in \mathbb{R}^d\). We denote \(P_i(Y_1, \ldots, Y_d) := c_{x_i, y_0}(x_0, y_0) Y^{k_i}\) for \(1 \leq i \leq d\). We know that \((P_i)_{1 \leq i \leq d}\) is a transversal family of \(\mathbb{R}[Y_1, \ldots, Y_d]\). We have \(f(y_n) := c_x(x_0, y_n) = A(y_n)\). Passing to the limit \(y_n \to y_0\), we get \(f(y_0) = A(y_0)\). Now subtracting the terms, we get \(f(y_n) - f(y_0) = \nabla A(y_n - y_0)\), and applying Taylor-Young around
We consider from the definition of $P_1$, which is a C\(^{1}\) of all these neighborhoods $N$ below on this set by $m$.

Proof of Theorem 2.13

3.3 Lower bound for a smooth coupling function

Proof of Theorem 2.13 By Taylor expansion of $c_x$ in $y$ in the neighborhood of $x_0$, we get for $0 < \varepsilon < 1$ and $h \in \mathbb{R}^d$ that

$$c_x(x_0, x_0 + \varepsilon h) = P(\varepsilon h) + c_{xy}(x_0, x_0)\varepsilon h + c_x(x_0, x_0) + \varepsilon^3 R_\varepsilon(h),$$

where, recalling the notation (2.6), $P_i(Y) := \frac{1}{2}c_{xyy}(x_0, x_0)[Y^2]$ and the Taylor-Lagrange remainder

$$R_\varepsilon(h) = \int_0^1 (1 - t)^2 c_{xyyy}(x_0, x_0 + t\varepsilon h)[h^3]dt$$

which is a $C^1$ function with derivative and value uniformly bounded in $\varepsilon$ on any compact. By Proposition 2.8, we see that $N_c(x_0)$ is finite by second order transversality of $c$ at $(x_0, x_0)$. We consider from the definition of $N_c(x_0)$ an affine map $A \in \text{Aff}_d$ such that the $d$-tuple of multivariate polynomials of degree one $A(X_1, ..., X_d)$ satisfies

$$|Z(P_i + A(X_1, ..., X_d)_i : 1 \leq i \leq d) \cap \{y \in \mathbb{R}^d : \nabla(P + A(X_1, ..., X_d))(y) \in GL_d(\mathbb{R})\}| = n$$

$$:= N_c(x_0).$$

Let $A_\varepsilon := \frac{1}{\varepsilon^2}A + c_{xy}(x_0, x_0)$, and $f_\varepsilon(h) := \frac{1}{\varepsilon^2}(c_x(x_0, x_0 + \varepsilon h) - A_\varepsilon(x_0 + \varepsilon h))$, with Taylor expansion is given by

$$f_\varepsilon(h) = P(h) - A(h) + \varepsilon R_\varepsilon(h) =: Q(h) + \varepsilon R_\varepsilon(h).$$

Let us now show that for $\varepsilon$ small enough, $f_\varepsilon$ has $n$ zeros $\{h_1, ..., h_n\}$ on $\mathbb{R}^d$ and that for any of these zeros, $\nabla f_\varepsilon(h_i) \in GL_d(\mathbb{R})$. Let $M > 0$ bounding the $n$ real zeros of $Q$, we set $S := B_{M+1}$ which is compact. We will work in this compact so that $R_\varepsilon$ is bounded and Lipschitz uniformly in $\varepsilon$. For all $1 \leq i \leq 2^d$, we have $\nabla Q(y_i) \in GL_d(\mathbb{R})$. Then for $\varepsilon$ small enough, $\nabla(Q(y_i) + \varepsilon R_\varepsilon(y_i)) \in GL_d(\mathbb{R})$. Let $h$ in a neighborhood $N_i$ of $y_i$, and assume that $\varepsilon$ is small enough so that we have $\nabla(Q(h) + \varepsilon R_\varepsilon(h)) = \nabla f_\varepsilon(h) \in GL_d(\mathbb{R})$. We denote $N$ the union of all these neighborhoods $N_i$, $Q$ is nonzero on $S \setminus N$, as it is a closed set, $|Q|$ is bounded from below on this set by $m > 0$. We take $\varepsilon$ small enough to have $|\varepsilon R_\varepsilon| \leq \frac{m}{2}$. Then for $1 \leq i \leq 2^d$, $|f_\varepsilon|^2$ reaches its minimum $h_i$ on $N_i$ in its interior as $|f_\varepsilon(y_i)| = |\varepsilon R_\varepsilon(y_i)|^2 < m^2$. Therefore,
we consider the determinants of submatrices of $\nabla f_\varepsilon(h_i)f_\varepsilon(h_i)$. We have $\nabla f_\varepsilon(h_i) \in GL_d(\mathbb{R})$, so that $f_\varepsilon(h_i) = 0$. We finally assume that the neighborhood we have taken is small enough to make them disjoint. The result is proved.

Now the theorem is just an application of Theorem 2.2 as the mapping $y \mapsto c_\varepsilon(x_0, y) - H_\varepsilon - \varepsilon_c$ has $n$ distinct zeros $x_0 + \varepsilon h_i$, it is $C^2$ in their neighborhood and $c_{xy}(x_0, x_0 + \varepsilon h_i) = \varepsilon^2 \nabla f_\varepsilon(h_i) \in GL_d(\mathbb{R})$.

As a preparation for the proof of Theorem 2.14 we need to prove the following lemma.

**Lemma 3.5.** Let $(P_1, ..., P_d)$ be a transversal family in $\mathbb{R}^2[X_1, ..., X_d]$. Then the multivariate polynomial $\det(\nabla P_1, ..., \nabla P_d)$ is non-zero.

**Proof.** We suppose to the contrary that $\det(\nabla P) = 0$, where we denote $P = (P_1, ..., P_d)$. We claim that we may find $y_0 \in \mathbb{R}^d$, and a map $u : \mathbb{R}^d \to \mathcal{S}_1(0)$ which is $C^\infty$ in the neighborhood of $y_0$ and such that $u(y) \in \ker(\nabla P(y))$ for $y$ in this neighborhood. Then we solve the differential equation $y'(t) = u(y(t))$ with initial condition $y(0) = y_0$. As a consequence of the regularity of $u$ in the neighborhood of $y_0$, by the Cauchy-Lipschitz theorem, this dynamic system has a unique solution for $t$ in a neighborhood $[-\varepsilon, \varepsilon]$ of 0, where $\varepsilon > 0$. However, we notice that $P(y(t))$ is constant in $t$, indeed, $\frac{d(P(y(t)))}{dt} = \nabla P(y(t))u(y(t)) = 0$. Since $|y'(t)| = 1$, this solution is non constant, then $P - P(y_0)$ has an infinity of roots: $y([0, 0])$. However, as $P$ is non-constant, $P - P(y_0)$ is also transversal, which is in contradiction with the fact that it has an infinity of zeros by the Bezout Theorem 3.3.

It remains to prove the existence of $y_0 \in \mathbb{R}^d$, and a map $u : \mathbb{R}^d \to \mathbb{R}^d$, $C^\infty$ in the neighborhood of $y_0$, such that $u(y) \in \ker(\nabla P(y))$ for $y$ in this neighborhood. For all $i < d$, we consider the determinants of submatrices of $\nabla P$ which have size $i$. Let $r \geq 0$ the biggest such $i$ so that at least one of these determinants is not the zero polynomial. By the fact that $\det(\nabla P) = 0$, and that the polynomials are non-constant by transversality, we have $0 < r < d - 1$. We fix one of these non-zero polynomial determinants. Let $x_0 \in \mathbb{R}^d$ such that this determinant is non-zero at $y_0$. As this determinant is continuous in $y$, it is non-zero in the neighbourhood of $y_0$. Therefore, $\nabla P$ has exactly rank $r$ in the neighbourhood of $y_0$. Now we show that this consideration allows to find a continuous map $y \mapsto u(y)$, such that $u(y)$ is a unit vector in $\ker(\nabla P)$. Notice that $\ker(\nabla P) = \im(\nabla P^t)^\perp$. We consider $r$ columns of $\nabla P^t$ that are used for the non-zero determinant. We apply the Gramm-Schmidt orthogonalisation algorithm on them. We get $u_1(y), ..., u_r(y)$, an orthonormal basis of $\im(\nabla P(y)^t)^\perp$, defined and $C^\infty$ in the neighbourhood of $y_0$. Then let $u_0 \in \ker(\nabla P(y_0))$, a unit vector. The map

$$u(y) := \frac{u_0 - \sum_{i=1}^r \langle u_0, u_i(y) \rangle u_i(y)}{\|u_0 - \sum_{i=1}^r \langle u_0, u_i(y) \rangle u_i(y)\|}$$

is well defined, $C^\infty$, and in $\im(\nabla P(y)^t)^\perp = \ker(\nabla P(y))$ in the neighbourhood of $y_0$, and therefore satisfies the conditions of the claim. \qed
Proof of Theorem 2.14: Let \( P_i := (X_1, ..., X_d)_{c_{i,y}}(x_0, x_0)(X_1, ..., X_d)^t \). Let \( y_1, ..., y_{d+1} \in \mathbb{R}^d \), affine independent. We may find \( A \in \text{Aff}_d \) such that \( A(y_i) = P(y_i) \) for all \( i \), where we denote \( P := (P_i)_{1 \leq i \leq d} \).

Step 1: Now we prove that \( \nabla (P(y'_i) - A) \) is invertible at points \( y'_i \) at the neighborhood of \( y_i \). First we get an explicit expression of \( \nabla A \) as a function of the \( y_i \). Let \( Y = \text{Mat}(y_i - y_{d+1}, i = 1, ..., d) \), the matrix with columns \( y_i - y_{d+1} \), using the equality \( \nabla A y_i + A(0) = P(y_i) \), we get the identity \( \nabla A Y = M \), where we denote \( M := \text{Mat}(P(y_i) - P(y_{d+1}), i = 1, ..., d) \). Then we get the result \( \nabla A = M Y^{-1} \) (\( Y \) is invertible as the \( y_i \) are affine independent). Then having \( \nabla P(y_{d+1}) - \nabla A \) invertible is equivalent to having \( \nabla P(y_{d+1}) Y - M \) invertible. Notice that \( \nabla P(y_{d+1}) Y - M = -\text{Mat}(\tilde{P}(y_i), i = 1, ..., d) \), where \( \tilde{P} = P - P(y_{d+1}) - \nabla P(y_{d+1}) \cdot (Y - y_{d+1}) \), and that the multivariate polynomials \( \tilde{P}_i \) are transversal, as they only differ from the \( P_i \) by degree one polynomials. Consider the multivariate polynomial \( D := \det(\nabla \tilde{P}) \). Let \( 1 \leq i \leq d \), by Lemma 3.3 we may find \( y'_i \) in the neighborhood of \( y_i \) such that \( D(y'_i) \neq 0 \), and therefore \( \nabla \tilde{P}(y'_i) \) is invertible. Thanks to this invertibility, we may perturb the \( y'_i \) to make \( M' := \text{Mat}(\tilde{P}(y'_i), i = 1, ..., d) \) invertible. As \( S\text{p}(M') \) is finite, for \( \lambda > 0 \) small enough, \( M' + \lambda I_d \) is invertible. For \( 1 \leq i \leq d \), we may find \( y''_i \) in the neighborhood of \( y'_i \) so that \( \tilde{P}(y''_i) = \tilde{P}(y'_i) + \lambda e_i + o(\lambda) \), thanks to the invertibility of \( \nabla \tilde{P}(y'_i) \). Then for \( \lambda \) small enough, \( (P(y''_i), i = 1, ..., d) = M' + \lambda I_d + o(\lambda) \) is invertible.

We were able, by perturbing the \( y_i \) for \( i \neq d + 1 \) to make \( \nabla (P(y'_{d+1}) - A) \) invertible. By continuity, this invertibility property will still hold if we perturb again sufficiently slightly the \( y_i \). Then we redo the same process, replacing \( y'_{d+1} \) by another \( y'_i \). We suppose that the perturbation is sufficiently small so that all the invertibilities hold in spite of the successive perturbations of the \( y_i \). Finally, we found \( y'_1, ..., y'_{d+1} \) affine independent so that \( P(y'_i) = A(y'_i) \) and \( \nabla P(y'_i) - \nabla A \) is invertible for all \( 1 \leq i \leq d + 1 \).

Step 2: Then \( N_c(x_0) \geq d + 1 \) because \( y'_1, ..., y'_{d+1} \) are \( d + 1 \) single real roots of \( P + A = H_c(x_0) + A \), and \( A \in \text{Aff}_d \), which may be identified to \( R[Y_1, ..., Y_d]^d \). As the \( P_i - A_i \) are real multivariate polynomials, all non-real zeros have to be coupled with their complex conjugate. Recall that by Theorem 3.3 there are exactly \( 2^d \) zeros to this system. There are no zeros at infinity by Proposition 3.4 and there is an even number of non-real zeros by the invariance by conjugation observation. Then there must be an even number of real roots. As the \( y'_i \) are simple roots by invertibility of the derivative of \( P - A \) at these points, there must be an even number of real roots, counted with multiplicity. If \( d \) is even, \( d + 1 \) is odd, which proves the existence of a possibly multiple \( d + 2 - \text{th} \) zero \( y_0 \), distinct from the \( y_i \). We assume, up to renumbering, that \( y'_0, ..., y'_d \) are affine independent, and we perturb again \( y'_0, ..., y'_d \) to make \( y_0 \) a single zero. We need to check that \( y'_{d+1} \) is still a single zero of \( P - A \). Indeed, the map \( (y'_1, ..., y'_{d+1}) \mapsto A \) if locally a diffeomorphism around \( (y_1, ..., y_{d+1}) \), then by the implicit functions Theorem, we may write \( y'_{d+1} = F(y'_1, ..., y'_d, A) = F(y'_1, ..., y'_d, A(y'_0, ..., y'_d)) \), where \( F \) is a local smooth function. Then \( y'_{d+1} \) remains a single zero if the perturbation of \( y_0, ..., y_d \) is small enough. The result is proved, if \( d \) is even we may find \( d + 2 \) single zeros to \( P - A \).

The reverse inequality is a simple application of Proposition 2.8.  \( \square \)
3.4 Characterization for the p-distance

For \( p \geq 1 \) and \( x \in \mathbb{R}^d \), we have \( c(\cdot, y) \) differentiable on \( (\mathbb{R}^d)^* \) with

\[
c_x(x, y) = \frac{1}{|x - y|^p} \sum_{i=1}^{d} |x_i - y_i|^{p-1} \frac{x_i - y_i}{|x_i - y_i|} e_i.
\]

For \( p = 1 \) and \( p = \infty \), it takes a simpler form.

If \( p = 1 \), \( c(\cdot, y) \) is differentiable on \((\mathbb{R}^*)^d\) and \( c_x(x, y) = \sum_{i=1}^{d} \frac{x_i - y_i}{|x_i - y_i|} e_i \).

If \( p = \infty \), \( c(\cdot, y) \) is differentiable on \( \{x' \in \mathbb{R}^d, |x'_i - y_i| > |x'_j - y_j|, j \neq i, \text{ for some } 1 \leq i \leq d\} \), let \( i := \operatorname{argmax}_{1 \leq j \leq d}(|x_j - y_j|) \), we have \( c_x(x, y) = \frac{x_i - y_i}{|x_i - y_i|} e_i \).

**Proof of Proposition 2.21** We start with the case \( p = 1 \). We suppose without loss of generality that \( x_0 = 0 \). Recall that \( c(\cdot, y) \) is differentiable on \((\mathbb{R}^*)^d\) and \( c_x(0, y) = \sum_{i=1}^{d} \frac{y_i}{|y_i|} e_i \). Then the equation that we get is \( A(y) = \sum_{i=1}^{d} s(y)|e_i| \). Let \( E := \{\sum_{i=1}^{d} s(y)|e_i| : y \in S_0\} \subset \varepsilon \in \{-1,1\}^d \). We have \( E \subset \operatorname{Im} A \), which is an affine space of dimension \( r \). Then there are \( r \) coordinates \( i_1, \ldots, i_r \) that are chosen arbitrarily in \( \operatorname{Im} A \), and the other coordinates are affine functions of the previous one. We denote \( I := (i_1, \ldots, i_r) \) and \( T := (1, \ldots, d) / I \). Thus, \( \operatorname{card}(\operatorname{Im} A \cap \{-1,1\}^T) = \operatorname{card}((-1,1)^d) = 2^r \). As \( 0 \in \operatorname{ri} S_0 \), \( r \geq 1 \). Now, for all \( \varepsilon \in E \), let \( y_\varepsilon \in S_0 \) such that \( c_x(0, y_\varepsilon) = \varepsilon \). Then if \( y := y_\varepsilon + y_0 \in Q^d \) with \( y_0 \in \ker \nabla A \), we have \( A(y) = c_x(0, y) \), and therefore \( y \in S_0 \), proving the first part of the result.

Now we prove that \( S_0 \subset \tilde{\operatorname{conv}} S_0 \). Let us suppose to the contrary that \( y \in \operatorname{ri} \operatorname{conv} S_0 \cap S_0 \). Let \( y_1, \ldots, y_n \in S_0 \) such that \( y = \sum_{i=1}^{n} \lambda_i y_i \), convex combination. Then \( c_x(0, y) = \sum_{i=1}^{n} \lambda_i c_x(0, y_i) \). As \( |c_x(0, y)| = \sum_{i=1}^{n} \lambda_i |c_x(0, y_i)| = \sqrt{d} \), we are in a case of equality in Cauchy-Schwartz inequality. \( \varepsilon := c_x(0, y), c_x(0, y_1), \ldots, c_x(0, y_n) \) are all non-negative multiples of the same unit vector, and therefore all equal as they have the same norm. Then \( y, y_1, \ldots, y_n \in Q^d \), and \( y, y_1, \ldots, y_n \in y_\varepsilon + \ker \nabla A \). As we may apply the same to any \( y' \in y_\varepsilon + \ker \nabla A \), these vectors cannot be written as convex combinations of elements of \( S_0 \) from another affine space. Therefore, \( (y_\varepsilon + \ker \nabla A \cap \tilde{S}_0 = (y_\varepsilon + \ker \nabla A \cap Q^d \) is a relative face of \( \operatorname{conv} S_0 \). As we supposed that \( y \in \operatorname{ri} \operatorname{conv} S_0 \), we have \( (y_\varepsilon + \ker \nabla A \cap Q^d = \operatorname{ri} \operatorname{conv} S_0 \), as \( \operatorname{ri} \operatorname{conv} S_0 \) is a relative face of \( \operatorname{conv} S_0 \) (which constitute a partition of \( \operatorname{conv} S_0 \), see Hiriart-Urruty-Lemaréchal [13]). This is impossible as \( 0 \in \operatorname{ri} \operatorname{conv} S_0 \), whence the required contradiction.

The proof of the case \( p = \infty \) is similar to the proof of Proposition 2.21, replacing by \( \operatorname{card}((-1,1)^d(e_i)_{1 \leq i \leq d}) = 2d \) instead of \( 2^d \), and by \( |c_x(0, y)| = 1 \) instead of \( \sqrt{d} \). □

3.5 Characterization for the Euclidean p-distance coupling

By the fact that \( \operatorname{int} \operatorname{conv} S_0 \) contains \( x_0 \), we may find \( y_1, \ldots, y_{d+1} \in S_0 \) that are affine independent. Then we may find unique barycenter coefficients \( (\lambda_i)_i \) such that \( x_0 = \sum_{i=1}^{d+1} \lambda_i y_i \). For some \( y_1, \ldots, y_{d+1} \in S_0 \). For all \( a \in \mathbb{R} \), we define

\[
y'(a) := G(a) \sum_{i=1}^{d+1} \frac{\lambda_i}{a - a_i} y_i, \quad \text{with} \quad G(a) = \left(\sum_{i=1}^{d+1} \frac{\lambda_i}{a - a_i}\right)^{-1}, \quad \text{and} \quad a_i := g(|y_i - x_0|) \tag{3.11}
\]
where \( \{b_1, \ldots, b_r\} := \{a_1, \ldots, a_{d+1}\} \) with \( r \leq d + 1 \) and \( b_1 < \ldots < b_r \), and \( d_i := |\{j : a_j = b_i\}| - 1 \), the multiplicity of each \( b_i \) for all \( i \).

**Proposition 3.6.** We have \( y'(a) = y(a) \) for all \( a \notin Sp(\nabla A) \). In particular the map \( y' \) is independent of the choice of \( y_1, \ldots, y_{d+1} \in S_0 \). Furthermore, \( G(a) = \frac{(a-a_1) \ldots (a-a_{d+1})}{\det(aI_d - \nabla A)} \) where \( \gamma_1 < \ldots < \gamma_{r-1} \) are eigenvalues of \( \nabla A \). Finally if \( x_0 \in \operatorname{int} \text{conv}(y_1, \ldots, y_{d+1}) \), then we have \( b_1 < \gamma_1 < b_2 < \ldots < \gamma_{r-1} < b_r \).

**Proof.** We suppose that \( x_0 = 0 \) for simplicity. Let \( a \notin Sp(\nabla A) \), \( y(a) \) is the unique vector such that

\[
(aI_d - \nabla A)y(a) = A(0) \tag{3.12}
\]

We now find the barycentric coordinates of \( y(a) \). For any \( i \), \( A(y_i) = a_i y_i \) with \( a_i := g(|y_i|) \). As \( (y_i)_i \) is a barycentric basis, we may find unique \( (\lambda_i(a))_i \in \mathbb{R} \) such that \( y(a) = \sum_i \lambda_i(a)y_i \), and \( 1 = \sum_i \lambda_i(a) \). Then we apply \( A \) and get \( A(y(a)) = \sum_i \lambda_i(a)A(y_i) \), so that \( ay(a) = \sum_i \lambda_i(a)ay_i \). Subtracting the previous equality on \( y(a) \), we get \( 0 = \sum_i \lambda_i(a)(a-a_i)y_i \). As \( (y_i)_i \) is a barycentric basis, it is a family or rank \( d \). Then, by the fact that \( \sum_{i=1}^{d+1} \lambda_i y_i = 0 \), we have \( \lambda_i(a-a_i)_i \leq 0 \) and \( (\lambda_i(a-a_i))_i \leq 0 \) are in the same 1-dimensional kernel of the matrix \( (y_1, \ldots, y_{d+1}) \). Then we may find \( G(a) \) such that \( \lambda_i(a-a_i) = G(a) \lambda_i \). Now we assume that \( a \) is not part of the \( a_i \), then we have \( \lambda_i(a) = G(a)\lambda_i \), and \( G(a) = \left( \sum_{i=1}^{d+1} \frac{\lambda_i}{a-a_i} \right)^{-1} \). Finally

\[
y(a) = y'(a) = G(a) \sum_{i=1}^{d+1} \frac{\lambda_i}{a-a_i} y_i \quad \text{with} \quad G(a) = \left( \sum_{i=1}^{d+1} \frac{\lambda_i}{a-a_i} \right)^{-1}. \tag{3.13}
\]

Now we prove that \( G(a) = \frac{(a-a_1) \ldots (a-a_{d+1})}{\det(aI_d - \nabla A)} \). We first assume that \( a_1 < \ldots < a_{d+1} \) and that \( x_0 \in \operatorname{int} \text{conv}(y_1, \ldots, y_{d+1}) \) (i.e. \( \lambda_1, \ldots, \lambda_{d+1} > 0 \)). Then \( G(a)^{-1} \) has \( d+1 \) single poles \( a_1, \ldots, a_{d+1} \), such that \( \lim_{a \to a_i} G(a)^{-1} = +\infty \), and \( \lim_{a \to a_i} G(a) = -\infty \) for all \( i \). Therefore, \( G(\gamma_i)^{-1} = 0 \) for some \( a_i < \gamma_i < a_{i+1} \) for all \( i \leq d \). Then \( \gamma_i \) is a pole of \( G \), and \( y'(a) \) goes to infinity when \( a \to \gamma_i \), as the coefficient in the affine basis \( (y_i)_i \) go to \( \pm \infty \). Therefore, \( \gamma_i \) is an eigenvalue of \( \nabla A \), as there are \( d \) such eigenvalues, we have obtained all of them. Finally, by the fact that the rational fraction \( f \) has degree 1, as the set of its roots is restricted to the \( d+1 \) numbers \( a_i \). Furthermore the \( \gamma_i \) are \( d \) poles, and \( a^{-1}G(a) \to (\sum_{i=1}^{d+1} \lambda_i)^{-1} = 1 \), when \( a \to \infty \), we deduce the rational fraction \( G(X) = \frac{(X-a_1) \ldots (X-a_{d+1})}{(X-\gamma_1) \ldots (X-\gamma_d)} = \frac{(X-a_1) \ldots (X-a_{d+1})}{\det(aI_d - \nabla A)} \).

Now if we chose other affine independent \( (y_i)_i \leq i \leq d+1 \) (this time not necessary with \( x_0 \in \operatorname{conv}(y_1, 1 \leq i \leq d+1) \)), let the associated barycenter coordinates \( \lambda_1, \ldots, \lambda_{d+1} \in \mathbb{R}^d \), we suppose that the \( (a_i)_i \) are still distinct, the poles of \( y'(a) \) are still the \( d \) distinct eigenvalues of \( \nabla A \) that are determined by the \( \gamma_i \) such that \( \lim_{a \to \gamma_i} |y'(a)| \), independent of the choice of \( (y_i)_i \) because \( y'(a) = (aI_d - \nabla A)^{-1}A(0) \) is independent of this choice. However, the numerator of the fraction can be determined in the same way than it is determined in the previous case.

Now we want to generalize this result to \( \lambda_1, \ldots, \lambda_{d+1} \in \mathbb{R}^d \), and any \( (a_i)_i \). If we stay in the open set in which \( (y_i)_i \) is an affine basis of \( \mathbb{R}^d \), the mapping \( (y_i, a_i)_i \to A \) is continuous, and
so is the mapping \((y_i)_i \mapsto (\lambda_i)_i\). Therefore, as \((y_i, a_i, \lambda_i)_i \mapsto \sum_{i=0}^d \frac{\lambda_i}{a_i}\) is continuous as well, the identity remains true for all \(a_i, y_i\) such that \((y_i)_i\) is an affine basis and \(\lambda_i \geq 0\).

Let us now focus on the multiple \(a_i\)s. We consider \(1 \leq i \leq r\) such that \(d_i > 0\). By passing to the limit \(n \to \infty\) with some distinct \(a_i^n\) converging to \(a_i\) for all \(1 \leq i \leq d\), \(d_i\) eigen values of \(\nabla A\) at least will be trapped between the \(a_i\)s, as \(a_i^n < \gamma_i + 1 < a_i^{n+1} < \cdots < \gamma_i + k < a_i + k\) becomes at the limit \(a_i = \gamma_i + a_i + 1 = \cdots = \gamma_i + k = a_i + k\). Now we prove that no other eigenvalue is equal to \(a_i\). Indeed, rewriting (3.13) that equation become

\[
y(a) = y'(a) = G(a) \sum_{i=1}^r \frac{\lambda'_i}{a - b_i} y_i \quad \text{with} \quad G(a) = \left( \sum_{i=1}^r \frac{\lambda'_i}{a - b_i} \right)^{-1}.
\]

with \(\lambda'_i := \sum_{a_j = b_i} \lambda_j\). And \(G(a) = \frac{(X - b_1)^{d_1} \cdots (X - b_r)^{d_r}}{\det(XI_d - \nabla A)}\). By a similar reasoning when the \((a_i)_i\)s are distinct, we may find \(b_1 < \gamma_1 < b_2 < \cdots < \gamma_{r-1} < b_r\), eigenvalues of \(\nabla A\). Then, as \(\deg \det(XI_d - \nabla A) = d\), and \((X - b_1)^{d_1} \cdots (X - b_r)^{d_r}\) is a divisor to \(\det(XI_d - \nabla A)\), we have \(\det(XI_d - \nabla A) = (X - \gamma_1) \cdots (X - \gamma_{r-1}) (X - b_1)^{d_1} \cdots (X - b_r)^{d_r}\).

\[\text{Remark 3.7. Notice that in Proposition 3.6, the eigenvalues of } \nabla A \text{ are given by the } \gamma_i, \text{ and by each } b_i \text{ such that } d_i > 0, \text{ which has multiplicity } d_i, \text{ in particular, these coefficients (up to their numbering) do not depend on the choice of } y_1, \ldots, y_{d+1}.\]

\[\text{Proof of Theorem 2.15 We suppose again that } x_0 = 0 \text{ for simplicity. We know that if } y \in S_0, \ c_x(0, y) = g(|y|)y = A(y). \text{ We denote } a := g(|y|) \text{ and get},
\]

\[
(aI_d - \nabla A) y = A(0) \quad \text{(3.15)}
\]

Let \(a \in \text{fix}(g \circ |y - x_0|)\), then \((aI_d - \nabla A)y(a) = A(0)\), and \(A(y(a)) = ay(a) = g(|y(a)|)y(a) = c_x(0, y(a))\), and therefore \(y(a) \in S_0\). Conversely, if \(y \in S_0\) and \(a := g(|y|)\) is not an eigenvalue of \(\nabla A\), then \(y = (aI_d - \nabla A)^{-1} A(0) = y(a)\), and finally \(g(|y(a)|) = a\), hence \(a \in \text{fix}(g \circ |y - x_0|)\).

Now let \(t \in \text{Sp}(\nabla A)\) such that \(|y(t)| < \infty\). Let \(y \in S_0^t\), we have \((tI_d - \nabla A)y = (tI_d - \nabla A)(y - y(t)) + A(0) = A(0)\), by passing to the limit \(a \to t\) in the equation \((aI_d - \nabla A)y(a) = A(0)\). Finally, \(|y|^2 = \sqrt{p^2 - |p_t|^2}^2 + |p_t|^2 = p^2\) by Pythagoras theorem, \(A(y) = c_x(0, y)\), and therefore \(y \in S_0\). Conversely, if \(y \in S_0\) with \(g(|y|) = t\), then we have \(y - y(t) \in \ker(tI_d - \nabla A)\), and \(|y - p_t| = \sqrt{p^2 - |p_t|^2}\) by Pythagoras theorem: by definition \(y \in S_0^t\).

\[\text{Proof of Corollary 2.16 We use the notations from Proposition 3.6 and assume that } x_0 \in \text{int conv}(y_1, \ldots, y_{d+1}) \text{. By Theorem 2.15 } S_0 \text{ contains } 2 \sum_{i=1}^r d_i \text{ degenerate points. Furthermore, for all } 1 \leq i \leq r-1, \lim_{t \to \gamma_i} |y(t) - x_0| = \infty, \text{ therefore, as } b_{i+1} \text{ is a root of } g(|y(t) - x_0|) - t \text{ between } \gamma_i \text{ and } \gamma_{i+1}, \text{ there is another root } b_i^* \text{ possibly multiple equal to } b_i, \text{ by continuity of } g. \text{ Finally we have } 2 \sum_{i=1}^r d_i + r + (r-2) = 2d \text{ elements in } S_0 \text{ at least, with possible degeneracy}.\]

\[\text{Proof of Theorem 2.17 We assume again that } x_0 = 0 \text{ for simplicity. We suppose again that } x_0 = 0 \text{ for simplicity. By identity (1.1), if we multiply (3.12) by the comatrix, we get } \det(\lambda I_d - \nabla A)y = \text{Com}(\lambda I_d - \nabla A)^t A(0). \text{ Now taking the square norm, we get: } \det(\lambda I_d -} \]
\( \nabla A)^2 |p|^\frac{2}{2-p} \lambda^{\frac{2}{2-p}} - |\text{Com}(\lambda I_d - \nabla A)^i A(0)|^2 = 0. \) The polynomial \( \chi := \det(H - XI_d)^2 - |p|^\frac{2}{2-p} \lambda^{\frac{2}{2-p}} |\text{Com}(XI_d - \nabla A)^i A(0)|^2 \) is continuous in \((y_i)_i\), then similar to the proof of Theorem 2.15 we can pass to the limit from sequences of \( y_i^n \) converging to \( y_i \) for all \( i \) such that for all \( n \geq 1 \), the vectors \( y_i^n \) have distinct norms. It follows that \( b_i \) is a \( d_i \)-eigenvalue of \( \nabla A \), and a \((2d_i - 1)\)-root of \( \chi \). By Theorem 2.15 we have

\[
S_i = S_{V_i} \left( p_i, \sqrt{b_i^2 - |p_i|^2} \right) \subset \{ c_x(0, Y) = A(Y) \}.
\]

With the radius \( \sqrt{b_i^2 - |p_i|^2} > 0 \) as there are more than one elements in the sphere. We have a single sphere as the function \( g \) is monotonic, and therefore injective.

Now we prove that if \(-\infty < p \leq 1\), then the polynomial

\[
\chi(X) := \det(XI_d - \nabla A)^2 - |p|^\frac{2}{2-p} X^{\frac{2}{2-p}} |\text{Com}(XI_d - \nabla A)^i A(0)|^2
\]

has exactly \( 2d \) positive roots, counted with multiplicity. By Corollary 2.16 it has at least \( 2d \) roots, counted with multiplicity. Now we prove that there are at most \( 2d \) roots.

By Theorem 2.15 the roots of \( \det(XI_d - \nabla A) \) all have the same sign (same than \( p \)). Consequently, the coefficients of \( \det(XI_d - \nabla A) \) are alternated or all have the same sign. The same happens for \( \det(XI_d - \nabla A)^2 \). Now using the Descartes rule for polynomials with non integer exponents, by the fact that the coefficients of \( X^{\frac{2}{2-p}} |\text{Com}(XI_d - \nabla A)^i A(0)|^2 \) are located between the ones of \( \det(XI_d - \nabla A)^2 \), and as \( \text{deg}(\det(XI_d - \nabla A)^2) = 2d \), it follows that \( \text{deg}(|\text{Com}(XI_d - \nabla A)^i A(0)|^2) = 2d - 2 \), and \( 1 < \frac{2}{2-p} \leq 2 \). Then \( \chi(X) \) has at most \( 2d \) alternations in its coefficients, and therefore it has at most \( 2d \) positive roots according to the Descartes rule.

Now, assume that \( 1 < p < 2 - \frac{2}{3} \) or \( p > 2 + \frac{2}{3} \), then

\[
\chi(X) := \det(XI_d - \nabla A)^2 - X^{\frac{2}{2-p}} |\text{Com}(XI_d - \nabla A)^i A(0)|^2
\]

has exactly \( 2d + 1 \) positive roots counted with multiplicity.

Let us first prove that the polynomial has less than \( 2d + 1 \) roots. Similar to above, the coefficients of \( \det(XI_d - \nabla A) \) are alternated. And the same happens for \( \det(XI_d - \nabla A)^2 \). Using the Descartes rule for polynomials with non integer coefficients, by the fact that the coefficients of \( X^{\frac{2}{2-p}} |\text{Com}(XI_d - \nabla A)^i A(0)|^2 \) are located between the ones of \( \det(XI_d - \nabla A)^2 \), except strictly less than \( 3 \), and as \( \text{deg}(\det(XI_d - \nabla A)^2) = 2d \), it follows that \( \text{deg}(|\text{Com}(XI_d - \nabla A)^i A(0)|^2) = 2d - 2 \) and \(-3 < \frac{2}{2-p} < 5 \). Then \( \chi(X) \) has at most \( 2d + 2 \) alternations in its coefficients. Furthermore, the sign of the coefficients in front of the extreme monomials are opposed (because \( \chi \) is a difference of positive polynomials) then the maximum number of positive roots is odd, and therefore it has at most \( 2d + 1 \) positive roots according to Descartes rule.

\(^1\)The Descartes rule states that for a polynomial with positive coefficients, the number of positive roots is dominated by the number of alternations of signs of its coefficients ordered by their associated exponents, see [18]
By Corollary 2.16 we have 2d elements in $S_0$, more precisely, which range between $b_1$ and $b_r$. Furthermore, between 0 and $b_1$ we can find some $a \in D$:

**Case 1:** We assume that $p > 2$. Then $\chi(X) \to -\infty$ when $X \to 0$ as $-|p|^{-\frac{2}{p^2}}X^{-\frac{2}{p^2}}|\text{Com}(XI_d - \nabla A)A(0)|^2$ becomes dominant.

**Case 2:** We assume that $p < 2$. Then $\chi(X) \to -\infty$ when $X \to +\infty$ as $-|p|^{-\frac{2}{p^2}}X^{-\frac{2}{p^2}}|\text{Com}(XI_d - \nabla A)A(0)|^2$ becomes dominant.

Therefore there is one more real root, on the side where the polynomial goes to $-\infty$ as there is already one. Finally $\chi$ has $2d + 1$ roots at least and less than $2d + 1$ roots, it follows that it has exactly $2d + 1$ roots. We proved the second part of the theorem.

### 3.6 Concentration on the Choquet boundary for the p-distance

**Proof of Proposition 2.23**

(i) Let $y_0, y_1, \ldots, y_k \in S_0$ such that $y_0 = \sum_{i=1}^{k} \lambda_i y_i$, convex combination. Then as $c_x(x_0, y_0) \cdot u = \sum_{i=1}^{k} \lambda_i c_x(x_0, y_i) \cdot u = \sum_{i=1}^{k} \lambda_i c_x(x_0, y_i) \cdot u\left(y_i - x_0\right)$, we have $\sum_{i=1}^{k} \lambda_i c_x(x_0, y_i) \cdot u = u \cdot A(y_0 - x_0)$.

As $y \mapsto c_x(x_0, y) \cdot u$ is strictly convex, this imposes that $\lambda_i = 1$ and $y_i = y_0$ for some $i$. Finally, $y_0$ is extreme in $S_0$, $S_0$ is concentrated in its own Choquet boundary.

(ii) We know that for any $y \in S_0$ we have $c_x(x_0, y) = A(y)$. As the situation is invariant in $x_0$, we will assume $x_0 = 0$ for notations simplicity. We consider $1 < q < +\infty$ such that $\frac{1}{p} + \frac{1}{q} = 1$. For any $y \in (\mathbb{R}^d)^*$,

$$|c_x(0, y)|_q = \left| \frac{1}{|y|^{p-1}} \sum_{i=1}^{d} \left| y_i \right|^{p-1} \frac{y_i}{|y_i|} e_i \right|_q = \left| \frac{1}{|y|^{p-1}} \left( \sum_{i=1}^{d} \left| y_i \right| \left| y_i \right|^{p-1} \right)^{\frac{1}{q}} \right| = \frac{1}{|y|^{\frac{2}{p}}} = 1,$$

as we know that $y \neq 0$ because $c$ is superdifferentiable. Then for any $y \in S_0$, we have $|Hy + v|_q = 1$. We now assume that $y_0 = \sum_{i=1}^{k} \lambda_i y_i$ is a strict convex combination with $(y_i)_{0 \leq i \leq k} \in S_0^{k+1}$.

$$1 = |A(y_0)|_q = \sum_{i=1}^{k} \lambda_i |A(y_i)|_q \leq \sum_{i=1}^{k} \lambda_i |A(y_i)|_q = \sum_{i=1}^{k} \lambda_i = 1$$

We are in a case of equality for the triangular inequality for the norm $|.|_q$. We know then that all the $\lambda_i A(y_i)$ and $A(y_0)$ are positively multiples. As we know that all their $q$-norm is $\lambda_i \neq 0$ and 1, therefore $A(y_0) = \ldots = A(y_k)$ and $\frac{\lambda_k}{|y_k|} = \ldots = \frac{\lambda_k}{|y_k|}$. It means that they all belong to the same semi straight line originated in 0. As we supposed that $y_0$ is not extreme, 0 can be included in the convex combination as we must have $1 \leq i \leq k$ such that $|y_k| > |y_0|$. Then increasing the corresponding $\lambda_i$ while decreasing all the others, 0 can be included. As $0 \in ri \text{ conv } S_0$, we can then put any element of $S_0$ in the convex combination and $S_0 \subset \{0\} + \frac{\lambda_k}{|y_k|} \mathbb{R}_+$. As $0 \in ri \text{ conv } S_0$, then $S_0 = \{0\}$ and $y_0 = 0$, which is the required contradiction because we supposed that $y_0$ is not extreme in $S_0$.

(iii) We use the notations from Theorem 2.17. We suppose again without loss of generality that $x_0 = 0$. Let $d := \dim S_0$, for any $y_1, \ldots, y_{d+1} \in S_0$ with full dimension $d$, we may find
unique barycentric coordinates \((\lambda_i)_{1\leq i \leq d+1}\) such that \(\sum_{i=0}^{d} \lambda_i y_i = 0\). Let \(y \in S_0\) such that \(p|y|^{p-2} = g(|y|) \neq Sp(\nabla A)\). By Proposition 3.16, \(y\) can be expressed as

\[ y = G(X) \sum_{i=1}^{d+1} \frac{\lambda_i}{X-a_i} y_i \quad \text{with} \quad G(X) = \left( \sum_{i=1}^{d+1} \frac{\lambda_i}{X-a_i} \right)^{-1} \].

with \(X = p|y|^{p-2} > 0\). To have \(y \in \text{conv}(S_0)\) we then need to have all the \(\frac{\lambda_i}{X-a_i}\) of the same sign. As we supposed that the \((a_i)\), is an increasing sequence, there must be a \(0 \leq i_0 \leq d - 1\) such that \(\lambda_i < 0\) if \(i \leq i_0\) and \(\lambda_i > 0\) if \(i > i_0 + 1\) (or \(\lambda_i > 0\) if \(i \leq i_0\) and \(\lambda_i < 0\) if \(i > i_0 + 1\) but we will only treat the first case as this one can be dealt with similarly). Then the idea consists in proving that \(\chi\) defined by (3.16) has no zero in \(]a_{i_0}, a_{i_0+1}[\).

First let us prove that \(G\) has no pole on \(]a_{i_0}, a_{i_0+1}[\). \(G^{-1}\) can hit 0 at most \(d\) times (It is a polynomial of degree \(d\) divided by another polynomial). It hits 0 in any \(]a_i, a_{i+1}[\) for \(i \neq i_0\), as the limits on the bounds are \(+\infty\) and \(-\infty\). This provides \(d-1\) zeros. If there where a zero in \(]a_{i_0}, a_{i_0+1}[\), it would be double, as the infinity limits at \(a_{i_0}^+\) and \(a_{i_0+1}^-\) have the same sign. Which would be a contradiction.

Finally, as the poles of \(G\) are the eigenvalues of \(\nabla A\) and do not depend on the choice of \(y_1, \ldots, y_{d+1}\), we know that there are exactly two roots of \(\chi\) between two poles. As \(a_{i_0}\) and \(a_{i_0+1}\) are two zeros surrounded by two consecutive poles, there are not other zeros between these two poles. \(\chi\) has no zero on \(]a_{i_0}, a_{i_0+1}[\).

If \(X = a_{i_0}\) or \(X = a_{i_0+1}\), then it is a zero of \(a_{i_0} - X\), and all the elements in the convex combination have same size than \(y\). By the fact that we are in the case of equality in the Cauchy-Schwartz inequality, this proves that the combination only contains one element. Hence, \(y \in S_0\) has to be extreme in \(S_0\).

Now if \(y\) corresponds to an eigenvalue of \(\nabla A\), let \(b := g(|y|)\). We suppose that \(y = \sum_{i=1}^{d+1} \mu_i y_i,\) convex combination with \(y_1, \ldots, y_{d+1} \in S_0\), affine basis. Recall that all \(y(a)\) for \(a \notin Sp(\nabla A)\) can be written \(y(a) = G(a) \sum_{i=1}^{d+1} \frac{\lambda_i y_i}{a-b_i} = G(a) \sum_{i=1}^{d+1} \frac{\lambda_i}{a-b_i} y'_i\) where \(\lambda'_i = \sum_{i=1}^{d+1} \frac{\lambda_j}{a-b_i}\) and \(y'_i = \sum_{a_i = b_i} a_i \frac{\lambda_i}{a-b_i} y_j\). Let \(i_0\) such that \(b_{i_0} = b\), let \(\{y'_1, \ldots, y'_{d+1}\} := \{y' \in \{y_1, \ldots, y_{d+1}\} : g(|y'|) = b_{i_0}\}\), \(y \in \text{aff}(y'_1, \ldots, y'_{d+1})\), therefore \(\mu_i = 0\) if \(a_i \neq b\). As \(S_1\) is a sphere, it is concentrated on its own Choquet boundary, and therefore the convex combination \(y = \sum_{i=1}^{d+1} \mu_i y_i\) is trivial, \(y = y_i\) for some \(i\) and \(\mu_i = 1\).

(iv) In the first case, if \(p|y_0|^{p-2}\) is a double root of \(\chi\) defined by (3.17), then if \(p < 2 - \frac{2}{3}\) or \(p > 2 + \frac{2}{3}\), \(\chi\) has \(2d + 1\) roots and at most \(2d\) distinct roots set around the poles of \(G\) in the same way than in the case \(p \leq 1\) in the proof of (iii).

The same happens when we remove the smallest element \(y_0\) of \(S_0\). Similarly \(S_0 \setminus \{y_0\}\) is concentrated on its own Choquet boundary.

Now we prove that \(S_0\) is not concentrated on its own Choquet boundary. If \(p|y_0|^{p-2}\) is a single root of \(\chi\), we select \(y'_1, \ldots, y'_{d+1} \in S_0\) such that 0 is in their convex hull. By Proposition
if $y \in S_0$ and $X := p|y|^{p-2}$, then

$$y = G(X) \sum_{i=1}^{d+1} \frac{\lambda_i}{X - a_i} y_i \quad \text{with} \quad G(X) = \left( \sum_{i=1}^{d+1} \frac{\lambda_i}{X - a_i} \right)^{-1}.$$  \hspace{1cm} (3.18)

Case 1: We assume that $y_1' = y_0$. Then we apply (3.18) to $X := p|y|^{p-2}$ the second smallest zero of $\chi$ which is strictly smaller than the first pole by Theorem 2.17 (which also means that $G(X) \geq 0$): $y := G(X) \sum_{i=0}^{d} \frac{\lambda_i}{a_i - X} y_i \in S_0$, or written otherwise:

$$\frac{\lambda_0 G(X)}{X - a_0} y_0 = G(X) \sum_{i=2}^{d+1} \frac{\lambda_i}{X - a_i} y_i - y$$

$G$ has its first zero at $a_0$ which is smaller than its first pole which is between $a_1$ and $a_2$ strictly, so that $G(X) > 0$. This gives the result, rewriting the barycenter equation, we get:

$$y_0 = \sum_{i=2}^{d+1} \frac{\lambda_i (X - a_0)}{\lambda_0 (X - a_i)} y_i + \frac{G(X)}{\lambda_0} y$$

Therefore, $y_0 \in \text{conv}(S_0 \setminus \{y_0\})$.

Case 2: Now we assume that $y_0' \neq y_0$. We write the barycenter equation for $X = p|y_0|^{p-2}$, we get:

$$y_0 = \sum_{i=0}^{d} \frac{\lambda_i G(X)}{X - a_i} y_i' \quad \text{with} \quad G(X) = \left( \sum_{i=0}^{d} \frac{\lambda_i}{X - a_i} \right)^{-1}.$$  

Then for any $i$, $\frac{\lambda_i G(X)}{X - a_i} > 0$ as all the $\frac{\lambda_i}{X - a_i}$ have the same sign. Therefore $y_0 \in \text{conv}(S_0 \setminus \{y_0\})$.  

\hspace{1cm} $\blacksquare$

4 Numerical experiment

In the particular example $c(X, Y) = |X - Y|^p$, the computations are easy as the important unknown parameter $\lambda = p|y|^{p-2}$ is one-dimensional. We coded a solver that generates random $y_1, \ldots, y_{d+1} \in \mathbb{R}^d$ and determines the missing $y_{d+2}, \ldots, y_k$, with $k = 2d$ if $p \leq 1$, and $k = 2d + 1$ if $p > 1$ such that $\{y_1, \ldots, y_k\} = \{c_x(0, Y) = A(Y)\}$ for some $A \in \text{Aff}_d$, see Theorem 2.17. (As we chose randomly these vectors, we are in a non-degenerate case with probability 1). Theorem 2.23 only covers the case in which $p < 2 - \frac{2}{5}$ or $p > 2 + \frac{2}{3}$, however the numerical experiment seems to show that the result of this theorem still holds for all $2 \neq p > 1$. Figures 2, 3, 4, 5, and 6 show configurations ($S_0$, on the left) for $p = 1.9$ and $p = 2.1$ in which the result of the theorem holds, and the graphs of $2 \log \left| \frac{y(\lambda)}{p} \right|$ compared to $(p - 2) \log(\lambda)$ (on the right). The intersections are in bijection with the points in $S_0$ because of the non-degeneracy. We begin with Figures 2 and 3 in two dimensions.
Figure 2: $S_0$ for $d = 2$ and $p = 1.9$.

Figure 3: $S_0$ for $d = 2$ and $p = 2.1$.

Now Figures 4 and 5 in three dimensions.

Figure 4: $S_0$ for $d = 3$ and $p = 1.9$. 
Finally, Figure 6 shows two experiments in which $|S_0|$ contains exactly 17 elements for $d = 8$.

Figure 6: $S_0$ for $d = 8$, $p = 1.9$ on the left and $p = 2.1$ on the right.

References

[1] Gerald Beer. A polish topology for the closed subsets of a polish space. *Proceedings of the American Mathematical Society*, 113(4):1123–1133, 1991.

[2] Mathias Beiglböck, Pierre Henry-Labordère, and Friedrich Penkner. Model-independent bounds for option prices: a mass transport approach. *Finance and Stochastics*, 17(3):477–501, 2013.

[3] Mathias Beiglböck, Pierre Henry-Labordere, and Nizar Touzi. Monotone martingale transport plans and skorohod embedding. *preprint*, 2015.

[4] Mathias Beiglböck and Nicolas Juillet. On a problem of optimal transport under marginal martingale constraints. *The Annals of Probability*, 44(1):42–106, 2016.

[5] Alexander MG Cox and Jan Oblój. Robust pricing and hedging of double no-touch options. *Finance and Stochastics*, 15(3):573–605, 2011.

[6] Hadrien De March. Quasi-sure duality for multi-dimensional martingale optimal transport. *arXiv preprint arXiv:1805.01757*, 2018.
[7] Hadrien De March and Nizar Touzi. Irreducible convex paving for decomposition of multidimensional martingale transport plans. *arXiv preprint arXiv:1702.08298*, 2017.

[8] Alfred Galichon, Pierre Henry-Labordere, and Nizar Touzi. A stochastic control approach to no-arbitrage bounds given marginals, with an application to lookback options. *The Annals of Applied Probability*, 24(1):312–336, 2014.

[9] Nassif Ghoussoub, Young-Heon Kim, and Tongseok Lim. Structure of optimal martingale transport plans in general dimensions. *arXiv preprint arXiv:1508.01806*, 2015.

[10] Joe Harris. *Algebraic geometry: a first course*, volume 133. Springer Science & Business Media, 2013.

[11] Robin Hartshorne. *Algebraic geometry*, volume 52. Springer Science & Business Media, 2013.

[12] Pierre Henry-Labordère and Nizar Touzi. An explicit martingale version of the one-dimensional brenier theorem. *Finance and Stochastics*, 20(3):635–668, 2016.

[13] Jean-Baptiste Hiriart-Urruty and Claude Lemaréchal. *Convex analysis and minimization algorithms I: Fundamentals*, volume 305. Springer science & business media, 2013.

[14] David Hobson. The skorokhod embedding problem and model-independent bounds for option prices. In *Paris-Princeton Lectures on Mathematical Finance 2010*, pages 267–318. Springer, 2011.

[15] David Hobson and Martin Klimmek. Robust price bounds for the forward starting straddle. *Finance and Stochastics*, 19(1):189–214, 2015.

[16] David Hobson and Anthony Neuberger. Robust bounds for forward start options. *Mathematical Finance*, 22(1):31–56, 2012.

[17] David G Hobson. Robust hedging of the lookback option. *Finance and Stochastics*, 2(4):329–347, 1998.

[18] GJO Jameson. Counting zeros of generalised polynomials: Descartes’ rule of signs and laguerre’s extensions. *The Mathematical Gazette*, 90(518):223–234, 2006.

[19] Tongseok Lim. Optimal martingale transport between radially symmetric marginals in general dimensions. *arXiv preprint arXiv:1412.3530*, 2014.

[20] Tongseok Lim. Multi-martingale optimal transport. *arXiv preprint arXiv:1611.01496*, 2016.

[21] Ludger Rüschendorf. Fréchet-bounds and their applications. *Advances in probability distributions with given marginals*, pages 151–187, 1991.

[22] Volker Strassen. The existence of probability measures with given marginals. *The Annals of Mathematical Statistics*, pages 423–439, 1965.

[23] Danila A Zaev. On the monge–kantorovich problem with additional linear constraints. *Mathematical Notes*, 98(5-6):725–741, 2015.