Gaussian Mixture Model with unknown diagonal covariances via continuous sparse regularization

Romane Giard[(1)], Yohann De Castro[(1,3)] and Clément Marteau[(2)]

[1] Centrale Lyon, CNRS UMR 5208, Institut Camille Jordan, Écully, France.
[2] Univ. Claude Bernard, CNRS UMR 5208, Institut Camille Jordan, Villeurbanne, France.
[3] Institut Universitaire de France (IUF)

version as of September 4, 2025

Abstract

This paper addresses the statistical estimation of Gaussian Mixture Models (GMMs) with unknown diagonal covariances from independent and identically distributed samples. We employ the Beurling-LASSO (BLASSO), a convex optimization framework that promotes sparsity in the space of measures, to simultaneously estimate the number of components and their parameters.

Our main contribution extends the BLASSO methodology to multivariate GMMs with component-specific unknown diagonal covariance matrices—a significantly more flexible setting than previous approaches requiring known and identical covariances. We establish non-asymptotic recovery guarantees with nearly parametric convergence rates for component means, diagonal covariances, and weights, as well as for density prediction.

A key theoretical contribution is the identification of an explicit separation condition on mixture components that enables the construction of non-degenerate dual certificates—essential tools for establishing statistical guarantees for the BLASSO. Our analysis leverages the Fisher-Rao geometry of the statistical model and introduces a novel semi-distance adapted to our framework, providing new insights into the interplay between component separation, parameter space geometry, and achievable statistical recovery.

Corresponding author: Romane Giard
E-mail address: romane.giard@ec-lyon.fr

1 Introduction

Gaussian Mixture Models (GMMs) are a cornerstone of statistical modeling, offering a flexible and powerful framework for representing complex data distributions. They find widespread application in diverse fields, including clustering and density estimation [McLachlan and Peel, 2000; Bouveyron et al., 2019], image processing [Houdard et al., 2018], bioinformatics and economics [Fruhwirth-Schnatter et al., 2019, Section 8.4]. Despite their ubiquity, the estimation of GMMs, particularly when the number of components and their covariance structures are unknown, presents significant statistical and computational challenges.

The predominant method for GMM estimation is the Expectation-Maximization (EM) algorithm, which iteratively maximizes the log-likelihood. While the EM algorithm guarantees a non-decreasing likelihood and converges to a stationary point under some conditions [McLachlan and Krishnan, 1997; Wu, 1983; Balakrishnan et al., 2017], it faces significant practical challenges. The log-likelihood function is non-convex and multi-modal, making the algorithm sensitive to initialization and prone to converging to local maxima rather than the global optimum. Crucially, it requires the number of mixture components to be specified in advance. We aim to overcome these limitations.

This paper introduces a novel approach to the estimation of multivariate GMMs with unknown diagonal covariances, leveraging the Beurling-LASSO (BLASSO) methodology [De Castro and Gamboa, 2012; Candès and Fernandez-Granda, 2014; Duval and Peyré, 2015; Boyer et al., 2017; Poon, 2019; De Castro et al., 2021a]. The BLASSO offers a convex optimization framework in the space of measures. It promotes sparsity, thereby offering a principled way to simultaneously estimating the number of components, their respective weights, means, and diagonal covariance matrices. It has not yet been applied to settings where each mixture component is allowed to have its own covariance structure. This extension makes it possible to handle significantly more relevant models. For instance, in clustering applications, assuming identical covariances across components yields Voronoi partitions (linear decision boundaries), as in k-means. In contrast, allowing distinct covariance matrices leads to boundaries defined by quadratic equations [Murphy, 2012, Section 4.2].

A pivotal contribution of our research is the identification of a separation condition for the mixture components. This condition is instrumental in the construction of so-called *non-degenerate dual certificates* [Candès and Fernandez-Granda, 2014; Duval and Peyré, 2015], which play a key role in establishing estimation guarantees. Our analysis is grounded in the Fisher-Rao geometry of the statistical model, providing theoretical insights into the intricate relationships between component separation, the underlying geometry of the parameter space, and the achievable statistical recovery.

1.1 Continuous sparse regression for Gaussian Mixture Models

We denote by $\mathcal{N}(t, C)$ a multivariate Gaussian distribution in dimension $d \in \mathbb{N}^*$, with mean $t \in \mathbb{R}^d$ and covariance matrix $C \in \mathbb{S}^d_{++}$, where $\mathbb{S}^d_{++}$ denotes the space of positive-definite symmetric matrices of size $d \times d$. Given $u = (u_1, \dots, u_d) \in (\mathbb{R}^*_+)^d$ a vector with positive entries, we denote by $\operatorname{diag}((u_1)^2, \dots, (u_d)^2)$ the $d \times d$ diagonal matrix with diagonal entries u_k^2, where each u_k is interpreted as a marginal standard deviation. The density of $\mathcal{N}\left(t, \operatorname{diag}((u_1)^2, \dots, (u_d)^2)\right)$ is denoted by $\varphi_{(t,u)}$.

In this paper, we observe a n-sample $X_1, \dots, X_n \in \mathbb{R}^d$ drawn from a Gaussian Mixture distribution with diagonal covariances, defined by:

$$X_i \overset{i.i.d.}{\sim} \sum_{j=1}^{s} a_j^0 \varphi_{(t_j^0, u_j^0)} =: f^0 \quad \text{with} \quad \sum_{j=1}^{s} a_j^0 = 1 \quad \text{and} \quad a_j^0 > 0 \tag{1}$$

where $t_j^0 \in \mathbb{R}^d$, $u_j^0 = (u_{j,k}^0)_{k=1,\dots,d} \in (\mathbb{R}^*_+)^d$. Our goal is to estimate the number of *components* s (also called the sparsity index), the *weights* a_j^0 and the *location* parameters (t_j^0, u_j^0) of each mixture component, indexed by $j \in \{1, \dots, s\}$, from the observations $X_1, \dots, X_n$.

Remark 1.1. As a matter of fact, the Gaussian Mixture Model with unknown number of components, weights, means and covariance matrices:

$$\left\{ \sum_{j=1}^{p} a_j \mathcal{N}(t_j, C_j) \,:\, p \geq 1\,,\; a_j > 0\,,\; t_j \in \mathbb{R}^d\,,\; C_j \in \mathbb{S}^d_{++}\,,\; \forall i \neq j\,,\; (t_j, C_j) \neq (t_i, C_i)\,,\; \sum_{j=1}^{p} a_j = 1 \right\}$$

is identifiable, *i.e.*, if two distributions of this model are equal, then they have the same number of components, and the same components (up to a permutation) with the same associated weights. We refer to [Teicher, 1961] for the proof in dimension $d = 1$, and [Yakowitz and Spragins, 1968] for the generalization in dimension $d \geq 1$. Under mild assumptions, an even stronger result holds for *continuous* mixtures of Gaussian distributions, defined by a density on the space of mean and covariance matrix [Bruni and Koch, 1985].

Our estimation strategy relies on the Beurling-LASSO (BLASSO). This framework, introduced in [De Castro and Gamboa, 2012; Candès and Fernandez-Granda, 2014], has been successfully applied to various statistical estimation problems, particularly in the context of sparse signal recovery [Duval and Peyré, 2015] and compressed sensing [Poon et al., 2023]. The BLASSO approach is characterized by its ability to promote sparsity in the space of measures, enabling the recovery of discrete measures from continuous data. This framework has been extended to various settings, including the estimation of Gaussian Mixture Models with known covariances [De Castro et al., 2021a; Poon et al., 2023]. Algorithmic implementations are discussed in [Chizat, 2022] and [De Castro et al., 2023]. One key modeling idea of this approach is to lift the parameter space onto the space of measures according to the embedding

$$(a_j, t_j, u_j)_{j=1}^p \mapsto \mu = \sum_{j=1}^{p} a_j \delta_{(t_j, u_j)}\,,$$

where $(t_j, u_j)_{j=1}^p$ are referred to as the *particles* of μ. Any discrete probability measure on $\mathbb{R}^d \times (0, +\infty)^d$ with finite support, of size p for any $p \geq 1$, describes a set of parameters. We stress that the parameter p, the number of components, is free and not prescribed. The law of the n-sample $(X_i)_{i=1}^n$ (see (1)) is unambiguously described by the target parameters $(a_j^0, t_j^0, u_j^0)_{j=1}^s$ (by identifiability of the model, see Remark 1.1), and hence unambiguously represented by the so-called *target measure* $\mu^0 \coloneqq \sum_{j=1}^s a_j^0 \delta_{(t_j^0, u_j^0)}$.

Our estimator will be defined as a solution of a minimization problem over the space of measures, constructed from the observations $X_1, \dots, X_n$. It is designed to estimate the target measure μ^0. In particular, we want our estimator to put a mass close to a_j^0 around the particle (t_j^0, u_j^0). We will consider a convex loss of the form

$$F_{n,\tau}(\mu) + \kappa R(\mu) \tag{2}$$

where $F_{n,\tau}(\mu)$ is a *data fidelity* term, $R(\mu)$ is a *regularization* term enforcing sparsity, and $\kappa > 0$ a tuning parameter. The data fidelity term compares a predicted density encoded by μ with an empirical approximation of the target density. This approximation is obtained from the empirical distribution of $(X_i)_{i=1}^n$ by convolution with a Gaussian kernel of covariance $\tau^2 \mathrm{Id}_d$, depending on a *smoothing parameter* τ. This approach is standard in kernel density estimation [Tsybakov, 2008], but the main difference here is that τ will not be necessarily chosen to match some nonparametric rate. Its calibration will sometimes answer to an alternative purpose. The regularization term $R(\mu)$ is the total variation (TV) norm, analog of the ℓ^1-norm for measures, and aims to concentrate the mass of our estimator in a few regions. We refer to Section 2 and ($\mathcal{P}_\kappa$) for a complete description of the BLASSO procedure.

1.2 Contributions

Our primary contribution is the extension of the BLASSO framework to estimate GMMs with unknown diagonal covariances. This setting introduces a key challenge: the associated kernel (see Section 5) is not translation-invariant, departing from many standard BLASSO applications. We address this by:

- Introducing a reparametrization of the measures to work with a normalized kernel—an object essential for the theoretical analysis of the BLASSO, describing the correlation between 2 location parameters (t, u) and (t', u').
- Establishing recovery guarantees through the construction of non-degenerate dual certificates. This involves proving a modified version of the *local positive curvature* assumption (LPC) from [Poon et al., 2023, Assumption 1] for our specific non translation-invariant kernel.
- Using a semi-distance naturally aligned with the kernel-induced geometry. This allows us to formulate more tractable conditions for the certificate construction, relaxing the reliance on the Fisher-Rao distance used in related works.
- Establishing conditions on the separation between the components of the mixture (*i.e.*, the particles of μ^0) with respect to the semi-distance. These conditions depend on bounds on the variance, the sparsity s, the dimension d, and the smoothing parameter τ. Higher sparsity levels or looser bounds on the variances necessitate a wider minimal separation between the components of μ^0. See (17) for a precise condition.

We also provide novel prediction guarantees for the target density, achieving nearly parametric rates of convergence in different regularization regimes.

Informal results on error bounds We define μ^0_ω as a reparametrized version of the target measure μ^0,

$$\mu^0_\omega = \sum_{j=1}^{s} \omega^0_j \, \delta_{(t^0_j, u^0_j)} \quad \text{with} \quad \omega^0_j = W(x^0_j)\, a^0_j \,,$$

where W is a positive function that will be specified later. According to the procedure displayed in Section 2, we estimate μ^0_ω rather than μ^0. Our estimator is given by the argument minimum of a function J_W of the form (2) over nonnegative measures with support restricted to a compact set $\mathcal{X} \subset \mathbb{R}^d \times [u_{\min}, +\infty)^d$ (see ($\mathcal{P}_\kappa$) for a formalized version). Here, $u_{\min}$ denotes a lower bound on the diagonal elements of the covariance matrix.

Most of our results remain valid for approximate solutions, that is for any measure μ satisfying $J_W(\mu) \leq J_W(\mu^0_\omega)$, and not only the exact argument minimum solution of J_W. Accordingly, we will use the notation $\mu^\star_{n,\omega}$ when an exact solution is required, and $\hat{\mu}_{n,\omega}$ when an approximate solution suffices.

We evaluate the estimator by comparing its mass against the mass of μ^0_ω within "near regions" $\mathcal{X}^{near}_j(r_e)$ centered on the true parameters (t^0_j, u^0_j), and in the complementary "far region" within $\mathcal{X}$. Near regions correspond to balls of radius r_e for a semi-distance $\mathfrak{d}$ (given by (11) below). Figure 1 illustrates this partitioning. The near region $\mathcal{X}^{near}_j(r_e)$ depends only on its radius r_e, x^0_j, the dimension d, and the smoothing parameter τ.

Our estimator satisfies the following properties, given in expected value over $X_1, \dots, X_n$.

Theorem 1.1 (Recovery guarantees for the estimation of μ^0_ω, informal result)**.** *Assume that the particles (t^0_j, u^0_j) of μ^0 are sufficiently separated, where the minimal separation constraint only depends on the dimension d, the sparsity index s, bounds on the variance and choice of a smoothing parameter $\tau \leq u_{\min}$. Choosing as regularization parameter $\kappa = \frac{\sqrt{2}}{(2\pi)^{d/4}\tau^{d/2}\sqrt{n}}$, for any r_e such that $0 < r_e \leq r$ with r a fixed constant depending on d, it holds that, omitting the dependence on d,*

$$\mathbb{E}\left[|\mu^0_\omega(\mathcal{X}^{near}_j(r_e)) - \hat{\mu}_{n,\omega}(\mathcal{X}^{near}_j(r_e))|\right] \lesssim \frac{s}{r_e^2\sqrt{n}\tau^{d/2}} \quad \forall\, j = 1, \dots, s \tag{3}$$

and

$$\mathbb{E}\left[\left|\hat{\mu}_{n,\omega}\Big(\mathcal{X} \setminus \bigcup_j \mathcal{X}^{near}_j(r)\Big)\right|\right] \lesssim \frac{s}{\sqrt{n}\tau^{d/2}} \,.$$

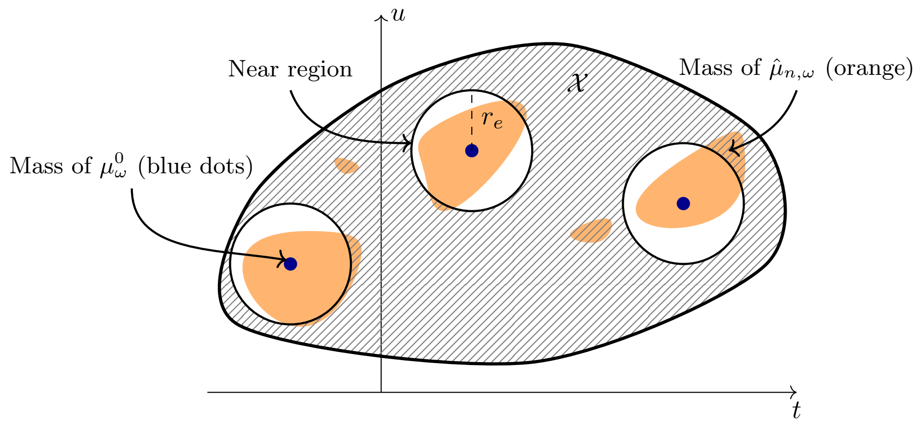

Figure 1: Schematic representation for Gaussian mixture models in dimension $d = 1$. Both parameters u (standard deviation) and t (mean) are in dimension 1, resulting in a 2-dimensional plot in location space (t, u). The discs represent the near regions, shown schematically: these regions correspond to balls defined with respect to a semi-distance, not the Euclidean distance. The hatched area corresponds to the far region.

We refer to Theorems 3.1 and 5.1, Proposition 3.1 with Lemma 5.4 for precise statements and related discussions. Our estimator allows us to locate the particles of the reparametrized target measure on fixed regions with a parametric convergence rate. The radius r_e can be chosen as desired (sufficiently small), leading to degraded bounds (the rate depends on r_e^{-2}). We give in Corollary 3.1 a variation of the previous result, for a direct estimator of μ^0—and with decreasing size of regions.

Remark 1.2. Note that in the above result, κ does not depend on s (unknown in practice). We call it the *s-agnostic* choice. Making the *s-dependent* choice $\kappa = \frac{\sqrt{2}}{(2\pi)^{d/4}\tau^{d/2}\sqrt{sn}}$, we obtain a rate of $\frac{\sqrt{s}}{r_e^2\sqrt{n}\tau^{d/2}}$ for the bound (3) (better dependence on s). See Remark 3.6.

We provide in the same time recovery guarantees for the prediction of the target density f^0. We achieve an almost parametric rate (up to a logarithmic factor) for the prediction of the target density, in two distinct regimes: under small regularization (Proposition 4.1) and with larger regularization (Theorem 4.1) when the assumption of Theorem 1.1 is verified. We present an informal version of these results.

Theorem 1.2 (Recovery guarantees for the prediction of f^0, informal results)**.** *Let $\tau = \frac{\sqrt{2}u_{\min}}{\sqrt{\ln n}}$. We construct an estimator $\hat{h}_n$ of f^0 from $\hat{\mu}_{n,\omega}$. We consider two regimes.*

- *Assume that the particles (t_j^0, u_j^0) of μ^0 are sufficiently separated. Choosing $\kappa = \frac{\sqrt{2}(\ln n)^{d/4}}{(2\pi)^{d/4}(2u_{\min}^2)^{d/4}\sqrt{n}}$, omitting the dependence on d and on bounds on the variance,*

$$\mathbb{E}\left[\left\|\hat{h}_n - f^0\right\|^2_{L^2(\mathbb{R}^d)}\right] \lesssim \frac{s(\ln n)^{d/2}}{n}.$$

- *Without separation assumptions on $(t_j^0, u_j^0)_j$, choosing $\kappa = \frac{4(\ln n)^{d/2}}{(2\pi)^{d/2}(2u_{\min}^2)^{d/2}n}$, omitting the dependence on d and on bounds on the variance,*

$$\mathbb{E}\left[\left\|\hat{h}_n - f^0\right\|^2_{L^2(\mathbb{R}^d)}\right] \lesssim \frac{(\ln n)^{d/2}}{n}.$$

The regularization parameter κ is allowed to depend on n. Its value has some importance on the performances of the procedure. It is calibrated according to the objective to be achieved: estimation, prediction or both. Under some conditions on μ^0, large regularization provides good estimation and prediction. Small regularization allows the BLASSO to focus on minimizing the data fidelity term, yielding good prediction without any separation condition on the particles of μ^0. For theoretical purposes, we will also briefly discuss an s-dependent choice of κ, which leads to improved bounds in the large regularization regime (the bound is of order $\frac{(\ln n)^{d/2}}{n}$ for s reasonably small, see Remark 4.4 below).

We finally derive an alternative result for $\mu^\star_{n,\omega}$, an exact solution to $(\mathcal{P}_\kappa)$, under some specific conditions. We provide below an informal result, formalized in Corollary 6.1.

Theorem 1.3 (Sparsity of the estimator for a large sample size, informal result)**.** *Let $0 < \tau \leq u_{\min}$ (fixed). Assume that the particles of μ^0 are sufficiently separated, where the minimal separation constraint only depends on d, s, τ and bounds on the variance. Choosing $\kappa = \frac{\alpha\sqrt{\ln n}}{(2\pi)^{d/4}\tau^{d/2}\sqrt{n}}$ for any $\alpha > 0$, there exists $n_0 \in \mathbb{N}$ depending on $\mu^0, \mathcal{X}, \tau, \alpha$, such that if the sample size n verifies $n \geq n_0$, then $\mu^\star_{n,\omega}$ is s-sparse with probability greater than $1 - C_\Gamma n^{-\frac{\gamma_0^2\alpha^2}{C_\Gamma^2}}$, with C_Γ a universal positive constant and γ_0 depending on $\mu^0, \mathcal{X}, \tau$. Moreover, writing $\mu^\star_{n,\omega} = \sum_{j=1}^s \omega^\star_j \delta_{x^\star_j}$, omitting constants depending on $\mu^0, \mathcal{X}, \tau$ we have for all $j = 1, \ldots, s$*

$$|\omega^0_j - \omega^\star_j| \lesssim \alpha\sqrt{\frac{\ln n}{n}} \quad \textit{and} \quad \mathfrak{d}((t^\star_j, u^\star_j), (t^0_j, u^0_j))^2 \lesssim \alpha\sqrt{\frac{\ln n}{n}}\,, \tag{4}$$

where $\mathfrak{d}(\bullet,\bullet)$ is the semi-distance between location parameters defined in (11) below.

The result displayed here provides a different flavor on the BLASSO performances. First, this bound holds provided the sample size is large enough (and under separation conditions that are slightly stronger than those required in Theorem 1.1). In such case, we first establish that the measure $\mu^\star_{n,\omega}$ has, with high probability, exactly the same sparsity index s than the target μ^0. Moreover, the bound (4) provides theoretical guarantees on the estimation of the mixture parameters themselves instead of a control on far and near regions (Theorem 1.1).

Outline Section 2 introduces the statistical framework and the BLASSO estimators $\hat{\mu}_{n,\omega}$, $\mu^\star_{n,\omega}$ used for recovering Gaussian Mixture Models. In Section 3, we establish recovery guarantees for the estimation of the target measure, relying on the existence of non-degenerate dual certificates. Section 4 focuses on prediction guarantees for the density f^0, providing rates of convergence under different regularization regimes. In Section 5, we construct non-degenerate dual certificates by analyzing the kernel properties and deriving sufficient conditions on μ^0. Finally, Section 6 demonstrates that, for sufficiently large sample sizes and under sufficient separation between components, the estimator $\mu^\star_{n,\omega}$ is, with high probability, a discrete measure. Section 7 proposes a concluding discussion including possible extensions, algorithmic issues and open problems. The main results are summarized in Tables 1 and 2 with corresponding assumptions, choices of smoothing parameter and regularization. Table 3 summarizes notation used throughout the paper. All proofs and technical results are gathered in the appendix at the end of the paper. To facilitate verification, we also provide a notebook (the associated Zenodo repository can be found at [Giard, 2025]) implementing our calculations using a symbolic computation library.

2 Statistical modeling

Our approach operates in the space of Radon measures, targeting the recovery of a sparse (*i.e.* discrete) measure. Each particle of this discrete measure encodes the parameters of a Gaussian component, and its associated mass corresponds to the component's proportion in the mixture. In this section, we introduce the necessary notation, describe the statistical model, and present the BLASSO estimator.

2.1 Radon measures

We first start with some definitions allowing for a rigorous introduction of the BLASSO principle.

For $A \subset \mathbb{R}^p$ locally compact, we denote $\mathcal{C}(A)$ the set of all continuous functions from A to $\mathbb{R}$ and $\mathcal{C}_0(A)$ the set of continuous functions that vanish at infinity, *i.e.*

$$\mathcal{C}_0(A) \coloneqq \Big\{ f \in \mathcal{C}(A) \ : \ \forall \varepsilon > 0\,,\ \exists C \text{ compact s.t. } |f| \leq \varepsilon \text{ on } A \setminus C \Big\}\,.$$

When A is compact, $\mathcal{C}_0(A) = \mathcal{C}(A)$.

Definition 2.1 (Radon measure on $A \subset \mathbb{R}^p$)**.** Let $A \subset \mathbb{R}^p$ be locally compact. The space of (real-valued) Radon measures $\mathcal{M}(A)$ is defined as the dual of $(\mathcal{C}_0(A), \|\bullet\|_\infty)$. Its dual norm is the total variation norm:

$$\|\mu\|_{\mathrm{TV}} = \sup_{\substack{\eta \in \mathcal{C}_0(A) \\ \|\eta\|_\infty \leq 1}} \int_A \eta(x) \mathrm{d}\mu(x).$$

We denote $\mathcal{M}(A)^+$ the set of nonnegative measures: $\mu \in \mathcal{M}(A)^+$ if for all $\eta \in \mathcal{C}_0(A)$ such that $\eta \geq 0$, $\int_A \eta \,\mathrm{d}\mu \geq 0$. A discrete mixture measure, or a sparse measure, is a measure that can be expressed as a finite weighted sum of Dirac measures:

$$\mu = \sum_{j=1}^s a_j \delta_{x_j}$$

where $s \geq 1$, $a_1, \ldots, a_s \in \mathbb{R}$, $x_1, \ldots, x_s \in A$. Remark that $\|\mu\|_{\mathrm{TV}} = \sum_{j=1}^s |a_j|$ when the $(x_j)_j$ are distinct.

Other details about the functional framework can be found in Appendix A.

2.2 Model and estimator

Model Our aim is to recover the target measure measure encoding the Gaussian Mixture distribution:

$$\mu^0 = \sum_{j=1}^s a_j^0 \delta_{x_j^0} \quad \text{where} \quad a_1^0, \ldots a_s^0 > 0\,,\ \sum_{j=1}^s a_j^0 = 1\,.$$

We assume that the location parameters $(x_1^0, \ldots, x_s^0)$ are distinct points, each of the form $x_j^0 = (t_j^0, u_j^0)$ with $t_j^0 = (t_{j,k}^0)_{k=1}^d \in \mathbb{R}^d$ and $u_j^0 = (u_{j,k}^0)_{k=1}^d \in [u_{\min}, +\infty)^d$ where $u_{\min}$ is some positive lower bound on $(u_{j,k}^0)_k$. The parameters t_j^0 and u_j^0 represent respectively the mean and square root of the diagonal covariance of a Gaussian component with weight a_j^0.

We denote φ the standard Gaussian density in $\mathbb{R}$, and $\sigma \coloneqq \mathcal{F}[\varphi] = e^{-\frac{\cdot^2}{2}}$ its associate Fourier transform. We can rewrite $f^0 = \Phi\mu^0$ where $\Phi : \mathcal{M}(\mathbb{R}^d \times [u_{\min}, +\infty)^d) \longrightarrow L^2(\mathbb{R}^d)$ is the linear operator defined by

$$\Phi\mu : z \in \mathbb{R}^d \mapsto \int_{\mathbb{R}^d \times [u_{\min}, +\infty)^d} \prod_{k=1}^d \frac{1}{u_k} \varphi\left(\frac{z_k - t_k}{u_k}\right) \mathrm{d}\mu(t, u) \quad \forall\, \mu \in \mathcal{M}(\mathbb{R}^d \times [u_{\min}, +\infty)^d)\,.$$

Our aim is to recover μ^0 from a n-sample $X_1, \ldots, X_n \overset{i.i.d.}{\sim} f^0$ (see (1)), considering that the weights $\{a_j^0\}_j$, the location parameters $\{x_j^0\}_j$ and the sparsity index s are unknown. The empirical density associated with our sample $X_1, \ldots, X_n$ is defined as

$$\hat{f}_n \coloneqq \frac{1}{n} \sum_{i=1}^n \delta_{X_i}\,.$$

This empirical measure is smoothed, hence allowing for a comparison with any prediction $\Phi\mu$. We will use a Gaussian convolution with smoothing parameter $\tau > 0$, introducing

$$\lambda : z \in \mathbb{R}^d \mapsto \frac{e^{-\frac{\|z\|_2^2}{2\tau^2}}}{(2\pi\tau^2)^{d/2}} \quad \text{and} \quad \Lambda \coloneqq \mathcal{F}[\lambda] = e^{-\frac{\tau^2 \|\cdot\|_2^2}{2}}\,.$$

Then, we set

$$L \circ\ \hat{f}_n = \lambda * \hat{f}_n \quad \text{and} \quad L \circ f = \lambda * f \quad \forall f \in L^2(\mathbb{R}^d)\,.$$

The term $L \circ\ \hat{f}_n$ is related to Kernel Density Estimation (KDE, [Tsybakov, 2008]). The choice of τ is important if we want to estimate f^0 in addition to μ^0, and will be discussed in Section 4. Our setting differs in this respect from super-resolution [Candès and Fernandez-Granda, 2014], where the analogous parameter λ_c is imposed by the experimental conditions (namely, the frequency cut-off in that paper).

Hilbert space for the data fidelity term The Hilbert space $\mathbb{L}$ in which we compare the observation and the prediction is the RKHS associated with λ. We define it using Mercer's theorem:

$$\mathbb{L} \coloneqq \left\{ f : \mathbb{R}^d \to \mathbb{R}\ :\ f \in L^2(\mathbb{R}^d) \text{ s.t. } \|f\|_{\mathbb{L}}^2 = \int_{\mathbb{R}^d} \frac{|\mathcal{F}[f](\xi)|^2}{\mathcal{F}[\lambda](\xi)} \mathrm{d}\xi < +\infty \right\},$$

with dot product

$$\forall f, g \in \mathbb{L}, \quad \langle f, g \rangle_{\mathbb{L}} = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \frac{\overline{\mathcal{F}[f](\xi)} \mathcal{F}[g](\xi)}{\mathcal{F}[\lambda](\xi)} \mathrm{d}\xi\,. \tag{5}$$

Problem We will search the target measure over the space of nonnegative Radon measures $\mathcal{M}(\mathcal{X})^+$ where $\mathcal{X}$ is a compact set of $\mathbb{R}^d \times [u_{\min}, +\infty)^d$. We will restrict the possible values of the covariance even further later.

The BLASSO problem constructed in [De Castro et al., 2021a] can be generalized to our setting. We may consider the following optimization problem:

$$\underset{\mu \in \mathcal{M}(\mathcal{X})^+}{\arg\min}\ J(\mu) \quad \text{where} \quad J(\mu) \coloneqq \frac{1}{2} \left\| L \circ \hat{f}_n - L \circ \Phi\mu \right\|_{\mathbb{L}}^2 + \kappa \|\mu\|_{\mathrm{TV}}\ , \quad \kappa > 0\,. \tag{$\bar{\mathcal{P}}$}$$

The data fidelity term $\frac{1}{2} \left\| L \circ \hat{f}_n - L \circ \Phi\mu \right\|_{\mathbb{L}}^2$ compares smooth versions of the empirical density to any candidate for the predicted density function $\Phi\mu$. The regularization $\kappa \|\mu\|_{\mathrm{TV}}$ promotes the sparsity of the solution. The choice of κ will be discussed in Section 3: a balance must be found between the data fidelity term and the regularization term.

However, we do not work with the loss J defined by ($\bar{\mathcal{P}}$). Indeed, the model we consider is more complex than that of [De Castro et al., 2021a]. Specifically, the addition of diagonal covariances in the parametrization of μ renders the problem ($\bar{\mathcal{P}}$) unsuitable for the use of standard proof techniques from the BLASSO literature [Poon, 2019]. In fact, a key object is the associated kernel $\bar{K}$ describing the correlation between 2 features, *i.e.*

$$\bar{K}(x, x') \coloneqq \langle L \circ \Phi\delta_x, L \circ \Phi\delta'_x \rangle_{\mathbb{L}}$$

for $x, x' \in \mathcal{X}$. This kernel is not normalized, namely $\bar{K}(x, x)$ is not constant (it depends on u), which causes issues in the investigation of recovery guarantees for a solution of ($\bar{\mathcal{P}}$).

Reparametrization To address this, we reparametrize the problem to work with a normalized kernel. This is achieved by introducing a positive weighting function $W(x)$. First notice that if $W : x \in \mathcal{X} \mapsto W(x) \in \mathbb{R}_+^*$ is continuous, then as $\mathcal{X}$ is compact, $\frac{1}{W}$ is continuous and bounded on $\mathcal{X}$. Hence we can reparametrize the weights of a measure by $\frac{1}{W}$. We will work with

$$W(x) \coloneqq \prod_{k=1}^{d} (2\pi)^{-1/4}(2u_k^2 + \tau^2)^{-1/4} \quad \forall x = ((t_1, \dots, t_d), (u_1, \dots, u_d)) \in \mathbb{R}^d \times [u_{\min}, +\infty)^d . \tag{6}$$

This function allows us to renormalize the kernel (see Section 5). The BLASSO problem we consider from now on is therefore

$$\min_{\mu \in \mathcal{M}(\mathcal{X})^+} J_W(\mu) \quad \text{where} \quad J_W(\mu) \coloneqq \frac{1}{2} \left\| L \circ \hat{f}_n - L \circ \Phi\left(\frac{\mu}{W}\right) \right\|_{\mathbb{L}}^2 + \kappa \|\mu\|_{\mathrm{TV}} , \tag{$\mathcal{P}_\kappa$}$$

with $\kappa > 0$ and where we define, for all $\mu \in \mathcal{M}(\mathbb{R}^d \times [u_{\min}, +\infty)^d)$, $\frac{\mu}{W}$ as the measure such that

$$\forall \eta \in \mathcal{C}_0(\mathbb{R}^d \times [u_{\min}, +\infty)^d) , \quad \int_{\mathbb{R}^d \times [u_{\min}, +\infty)^d} \eta \, \mathrm{d}\left(\frac{\mu}{W}\right) = \int_{\mathbb{R}^d \times [u_{\min}, +\infty)^d} \frac{1}{W(x)} \eta(x) \, \mathrm{d}\mu(x) .$$

We can show that the optimization problem ($\mathcal{P}_\kappa$) has a solution (see the appendix, Proposition B.1).

Estimator According to the problem ($\mathcal{P}_\kappa$), we consider an estimator $\mu^\star_{n,\omega}$ defined as

$$\mu^\star_{n,\omega} \in \operatorname*{arg\,min}_{\mu \in \mathcal{M}(\mathcal{X})^+} J_W(\mu) , \tag{7}$$

Most of our results also hold for an *approximate solution*, and we will use the notation $\hat{\mu}_{n,\omega}$ for our estimator when it suffices that

$$\hat{\mu}_{n,\omega} \in \left\{ \mu \in \mathcal{M}(\mathcal{X})^+ \ : \ J_W(\mu) \leq J_W(\mu^0_\omega) \right\} , \tag{8}$$

where μ^0_ω is the weighted measure defined as

$$\mu^0_\omega = \sum_{j=1}^{s} \omega^0_j \delta_{x^0_j} \quad \text{with} \quad \omega^0_j = W(x^0_j) a^0_j \quad \forall j \in \{1, \dots, s\}. \tag{9}$$

Remark that $\mu^\star_{n,\omega}$ satisfies the relaxed condition (8). This flexibility allows us to obtain the guarantees presented in Sections 3 and 4 without requiring the exact minimization of J_W—this is computationally advantageous. While algorithmic details are beyond the scope of this paper, they will be briefly discussed in Section 7 (see also Chizat, 2022; De Castro et al., 2023). In contrast to other sections, the results of Section 6 specifically require an *exact solution* $\mu^\star_{n,\omega}$.

Remark 2.1. Historically, the BLASSO method has been applied to recover a sparse target measure that is not necessarily a probability measure. As μ^0 is a probability measure in our setting, one may instead consider the constrained problem

$$\operatorname*{arg\,min}_{\mu \in \mathcal{M}(\mathcal{X})^+, \, \|\mu\|_{\mathrm{TV}} = 1} \frac{1}{2} \left\| L \circ \hat{f}_n - L \circ \Phi\mu \right\|_{\mathbb{L}}^2 . \tag{$\mathcal{P}_C$}$$

Such an approach deserves some comments. First, the unconstrained programs ($\mathcal{P}_\kappa$) and ($\bar{\mathcal{P}}$) belong to a family of convex potentials studied in optimization on the space of measures. A line of work, *e.g.* [Chizat, 2022; De Castro et al., 2023], shows convergence of various gradient descent strategies (particle, Wasserstein gradient flows, stochastic gradient), while such results do not exist for the constrained program ($\mathcal{P}_C$) to the best of our knowledge. Second, for the program that we deal with (see ($\mathcal{P}_\kappa$)), the reparametrized target μ^0_ω does not verify $\|\mu^0_\omega\|_{\mathrm{TV}} = 1$, and enforcing the constraint $\left\| \frac{\hat{\mu}_{n,\omega}}{W} \right\|_{\mathrm{TV}} = 1$ is an obstacle to establishing guarantees.

3 Estimation guarantees

In this section, we give guarantees for the recovery of μ^0_ω and μ^0 using the estimator $\hat{\mu}_{n,\omega}$ introduced in (8). We provide bounds on the difference in mass assigned by the estimator and the target measure over relevant regions of the space. Our guarantees hold under the existence of so-called non-degenerate *dual certificates*, which are the key objects for analyzing the properties of the BLASSO estimator [Candès and Fernandez-Granda, 2014; Duval and Peyré, 2015; De Castro et al., 2021a; Poon et al., 2023]. In a general context, they are functions interpolating the signs of the particles of the sparse target measure, with some prescribed smoothness and shape constraints. We investigate the construction of such objects in Section 5.

3.1 Non-degenerate dual certificates

In this section, we define non-degenerate certificates and introduce their properties, closely related to the optimization problem at hand. These certificates are linked to the *feature map*, which is defined as the linear operator

$$\Psi : \mathcal{M}(\mathbb{R}^d \times [u_{\min}, +\infty)^d) \to \mathbb{L}\,,\ \mu \mapsto L \circ \Phi \frac{\mu}{W}\,.$$

Remark that the loss function of $(\mathcal{P}_\kappa)$ can be rewritten as $J_W : \mu \in \mathcal{M}(\mathcal{X})^+ \mapsto \frac{1}{2}\left\| L \circ \hat{f}_n - \Psi\mu \right\|^2_{\mathbb{L}} + \kappa\, \|\mu\|_{\mathrm{TV}}$. The adjoint operator of Ψ restricted to $\mathcal{M}(\mathcal{X})$ verifies

$$\langle \Psi\mu, p\rangle_{\mathbb{L}} = \langle \Psi^* p, \mu\rangle_{\mathcal{C}(\mathcal{X}),\mathcal{M}(\mathcal{X})} = \int \Psi^* p \,\mathrm{d}\mu \quad \forall\, p \in \mathbb{L} \tag{10}$$

for all $\mu \in \mathcal{M}(\mathcal{X})$. In particular, for all $x \in \mathcal{X}$, $[\Psi^* p]\,(x) = \langle p, \Psi\delta_x\rangle_{\mathbb{L}}$.

Dual certificates are continuous functions of the form $\Psi^* p$ for some $p \in \mathbb{L}$. Our estimation guarantees are based on controls of these certificates on regions defined by μ^0 and some appropriate distance. In [De Castro et al., 2021a] the Euclidean distance is used. The latter is however not adapted to the geometry of our problem, as we deal with unknown covariances. In particular, the associated kernel is not translation-invariant, see Section 5. [Poon et al., 2023] work alternatively with the Fisher-Rao distance, which is however impractical to use in our context. In fact, interactions between the kernel and the Fisher-Rao appear to be quite intricate to manage. A key aspect of our contribution is to introduce greater flexibility in the control of certificates. We relax the assumptions from [Poon et al., 2023] which are challenging to verify in our context, by expressing controls using a specific distance on regions defined by a different semi-distance. More specifically, we use the semi-distance $\mathscr{d}$ defined by

$$\mathscr{d}(x,x')^2 = \sum_{k=1}^{d}\left(\frac{(t_k - t'_k)^2}{{u'_k}^2 + u_k^2 + \tau^2} + \ln\left(\frac{{u'_k}^2 + u_k^2 + \tau^2}{\sqrt{2u_k^2 + \tau^2}\sqrt{2{u'_k}^2 + \tau^2}}\right)\right) \quad \forall\, x, x' \in \mathbb{R}^d \times [u_{\min}, +\infty)^d\,. \tag{11}$$

This semi-distance is symmetric, nonnegative, verifies $\mathscr{d}(x,x') = 0 \iff x = x'$ for all $x, x' \in \mathbb{R}^d \times [u_{\min}, +\infty)^d$, but the triangle inequality does not hold. The connection between the semi-distance $\mathscr{d}$ and the optimization problem $(\mathcal{P}_\kappa)$ will be made precise in Section 5.

Controls of the certificates also depend on some distance $\mathfrak{d}_{\mathfrak{g}}$ on $\mathcal{X}$, left unspecified in this section. We will see in Section 5 that the Fisher-Rao distance is well-suited for this problem. To establish the existence of non-degenerate dual certificates, we ensure compatibility between the semi-distance and $\mathfrak{d}_{\mathfrak{g}}$ (see Section 5). Up to the semi-distance $\mathscr{d}$, the elements displayed in this section are quite generic in the BLASSO literature and have been discussed in various contexts [Duval and Peyré, 2015; De Castro et al., 2021a; Poon, 2019].

First, we introduce the so-called *far* and *near* regions.

Definition 3.1 (Near and far regions)**.** For $r > 0$, we define the near regions associated with each particle x^0_j, $j \in \{1, \dots, s\}$ as

$$\mathcal{X}^{near}_j(r) \coloneqq \{x \in \mathcal{X} \ :\ \mathscr{d}(x, x^0_j) \le r\}$$

and the far region

$$\mathcal{X}^{far}(r) \coloneqq \mathcal{X} \setminus \mathcal{X}^{near}(r) \quad \text{where} \quad \mathcal{X}^{near}(r) \coloneqq \bigcup_j \mathcal{X}^{near}_j(r)\,.$$

The jth near region identifies the points in $\mathcal{X}$ that are close to x^0_j, the jth particle of μ^0. To ensure accurate recovery, we require the existence of a (global) dual certificate η that satisfies non-degeneracy conditions. These conditions are adapted below from [Poon et al., 2023, Definition 2], with modifications to account for the definition of near and far regions, all based on the semi-distance $\mathscr{d}$ rather than the Fisher-Rao distance.

Definition 3.2 (Global non-degenerate certificate)**.** Let $\varepsilon_0, \varepsilon_2, r \in \mathbb{R}_+^*$. A function $\eta \in \mathrm{Im}(\Psi^*)$ is a $(\varepsilon_0, \varepsilon_2, r)$-non-degenerate certificate associated with the measure μ^0 if

1. $\eta(x_j^0) = 1$ for all $j = 1, \ldots, s$,
2. $|\eta(x)| \leq 1 - \varepsilon_0$ for all $x \in \mathcal{X}^{far}(r)$,
3. $\eta(x) \leq 1 - \varepsilon_2 \mathfrak{d}_{\mathfrak{g}}(x, x_j^0)^2$ for all $x \in \mathcal{X}_j^{near}(r)$, $j \in \{1, \ldots, s\}$.

A global non-degenerate certificate is thus a function whose regularity is prescribed (it must be in $\mathrm{Im}(\Psi^*)$), which interpolates (item 1 of Definition 3.2) and localizes (items 2 and 3) the particles of the target μ^0.

We will make use of additional non-degenerate certificates, localizing the jth particle of μ^0.

Definition 3.3 (Local non-degenerate certificates)**.** Let $\tilde{\varepsilon}_0, \tilde{\varepsilon}_2, r \in \mathbb{R}_+^*$. For $j \in \{1, \ldots, s\}$, $\eta_j \in \mathrm{Im}(\Psi^*)$ is a $(\tilde{\varepsilon}_0, \tilde{\varepsilon}_2, r)$-non-degenerate certificate for the jth near region if

1. $\eta_j(x_j^0) = 1$ and $\eta_j(x_i^0) = 0$ for all $i \in \{1, \ldots, s\}$ such that $i \neq j$,
2. $|\eta_j(x)| \leq 1 - \tilde{\varepsilon}_0$ for all $x \in \mathcal{X}^{far}(r)$,
3. $|1 - \eta_j(x)| \leq \tilde{\varepsilon}_2 \mathfrak{d}_{\mathfrak{g}}(x, x_j^0)^2$ for all $x \in \mathcal{X}_j^{near}(r)$,
4. $|\eta_j(x)| \leq \tilde{\varepsilon}_2 \mathfrak{d}_{\mathfrak{g}}(x, x_i^0)^2$ for all $x \in \mathcal{X}_i^{near}(r)$, for all $i \in \{1, \ldots, s\}$ such that $i \neq j$.

Note that the near and far regions along with Ψ^* depend on the parameter τ appearing in λ. The distance $\mathfrak{d}_{\mathfrak{g}}$ will also depend on it. For the estimation of μ^0, the choice of τ is related via $(x_1^0, \ldots, x_s^0)$ to the possibility of constructing non-degenerate certificates (see Section 5), and influences the convergence rate (see Theorem 3.1). For the estimation of the density f^0, this choice is more critical: we suggest possible values for τ in Section 4.

3.2 Error bounds

We extend here the results of [De Castro et al., 2021a] to our setting, where both means and covariances of the mixture model are unknown. Providing estimation guarantees requires to bound the error we made by approximating $L \circ f^0$ from our random observations (see Lemma 3.1). We give results taken in expected value over $X_1, \ldots, X_n \overset{i.i.d.}{\sim} f^0$.

Lemma 3.1 (Control of the noise. [De Castro et al., 2021b, Lemma 3])**.** *We define the so-called noise term* Γ_n *as*

$$\Gamma_n \coloneqq L \circ \hat{f}_n - L \circ \Phi\mu^0 \,. \tag{12}$$

We have

$$\mathbb{E}\left[\|\Gamma_n\|_{\mathbb{L}}^2\right] \leq \frac{4\int_{\mathbb{R}^d} \Lambda}{(2\pi)^d n} = \frac{4}{(2\pi)^{d/2}\tau^d n} =: \rho_n^2 \,.$$

Moreover, universal constants $C_\Gamma, \tilde{C}_\Gamma > 0$ *exist such that*

$$\mathbb{E}\left[\|\Gamma_n\|_{\mathbb{L}}^4\right] \leq \tilde{C}_\Gamma \rho_n^4$$

and

$$\forall \rho > 0\,, \quad \|\Gamma_n\|_{\mathbb{L}}^2 \leq \rho \frac{C_\Gamma^2 \int_{\mathbb{R}} \Lambda}{n(2\pi)^d} = \rho \frac{C_\Gamma^2}{n\tau^d (2\pi)^{d/2}} \quad \text{with probability greater than } 1 - C_\Gamma e^{-\rho}\,.$$

The proof is given in Appendix C. The constants $C_\Gamma, \tilde{C}_\Gamma$ in the previous lemma can be made explicit (see [Houdré and Reynaud-Bouret, 2003, Theorem 3.1]).

The control of any estimator $\hat{\mu}_{n,\omega}$ satisfying (8) also involves the non-degenerate dual certificates introduced in Definitions 3.2 and 3.3 above. Our recovery guarantees are based on the following assumption.

Assumption 1. *There exists* $\eta = \Psi^* p$ *a global* $(\varepsilon_0, \varepsilon_2, r)$*-non-degenerate certificate associated with* μ^0*, and local* $(\tilde{\varepsilon}_0, \tilde{\varepsilon}_2, r)$*-non-degenerate certificates* $\eta_j = \Psi^* p_j$ $(j \in \{1, \ldots, s\})$ *for each near region. Moreover, there exists* $c_p > 0$ *such that* $\|p\|_{\mathbb{L}}^2 \leq c_p s$ *and* $\|p_j\|_{\mathbb{L}}^2 \leq c_p$ *for* $j = 1, \ldots, s$.

The dependence on s for $\|p\|_L^2$ is quite natural (related to our construction of certificates, see Proposition I.1 in the appendix) and already appears in the literature, *c.f.* [Poon, 2019]. In Section 5, we show that Assumption 1 holds under some conditions on $\mathcal{X}$ and μ^0 by explicitly constructing certificates. In particular, we consider $\mathcal{X} \subset \mathbb{R}^d \times [u_{\min}, u_{\max}]^d$ and we impose a minimal separation between the particles of μ^0 depending on $s, u_{\min}, u_{\max}, d, \tau$. We provide fixed values for $r, \varepsilon_i, \tilde{\varepsilon}_i, c_p$, depending only on the dimension d. For a precise statement, see Theorem 5.1 below.

Now, we have all the ingredients to provide our first results concerning the performances of our estimator. We present a control of the mass of $\hat{\mu}_{n,\omega}$ on the near and far regions (Definition 3.1) in the next theorem, whose proof is displayed in Appendix D.

Theorem 3.1 (Estimation error)**.** *Assume that Assumption 1 holds. Setting* $\kappa = \frac{\rho_n}{\sqrt{c_p}}$ *in* ($\mathcal{P}_\kappa$)*, we have the following controls on* $\hat{\mu}_{n,\omega}$*.*

1. *Control over the mass of the estimator on the far region:*

$$\mathbb{E}\left[\hat{\mu}_{n,\omega}(\mathcal{X}^{far}(r))\right] \leq \frac{\sqrt{c_p}}{2\varepsilon_0}\rho_n(1+\sqrt{s})^2 .$$

2. *Accuracy of the mass reconstruction: for any* $j \in \{1, \dots, s\}$*,*

$$\mathbb{E}\left[\left|\omega_j^0 - \hat{\mu}_{n,\omega}(\mathcal{X}_j^{near}(r))\right|\right] \leq 2\sqrt{c_p}\rho_n(1+\sqrt{s}) + \max\left\{\frac{1-\tilde{\varepsilon}_0}{\varepsilon_0}, \frac{\tilde{\varepsilon}_2}{\varepsilon_2}\right\}\frac{\sqrt{c_p}}{2}\rho_n(1+\sqrt{s})^2 .$$

3. *Stability of the mass:*

$$-2\sqrt{c_p s}\rho_n\left(1+\sqrt{s}\right) \leq \mathbb{E}\left[\|\hat{\mu}_{n,\omega}\|_{\mathrm{TV}}\right] - \left\|\mu_\omega^0\right\|_{\mathrm{TV}} \leq \frac{\sqrt{c_p}}{2}\rho_n .$$

We present below several remarks on this result and its proof.

Remark 3.1 (A Bregman divergence approach)*.* The non-degenerate certificates from Assumption 1 allow us to derive recovery guarantees from the control of the so-called Bregman divergence, defined for $\eta \in \mathcal{C}(\mathcal{X})$ by

$$D_\eta(\hat{\mu}_{n,\omega}, \mu_\omega^0) = \|\hat{\mu}_{n,\omega}\|_{\mathrm{TV}} - \left\|\mu_\omega^0\right\|_{\mathrm{TV}} - \int_{\mathcal{X}} \eta \,\mathrm{d}\left(\hat{\mu}_{n,\omega} - \mu_\omega^0\right) . \tag{13}$$

In particular, if η is a global non-degenerate certificate, we get $\int_{\mathcal{X}} \eta \,\mathrm{d}\mu_\omega^0 = \left\|\mu_\omega^0\right\|_{\mathrm{TV}}$. In such a case, the Bregman divergence is nonnegative: $D_\eta(\hat{\mu}_{n,\omega}, \mu_\omega^0) = \int_{\mathcal{X}}(1-\eta)\,\mathrm{d}\hat{\mu}_{n,\omega} \geq 0$. The proof of Theorem 3.1 is based on lower and upper bounds on the Bregman divergence with η from Assumption 1. First, using that $\eta = \Psi^* p$ with the Cauchy-Schwarz inequality, we get

$$\left|\int_{\mathcal{X}} \eta \,\mathrm{d}(\hat{\mu}_{n,\omega} - \mu_\omega^0)\right| = \left|\left\langle p, L \circ \Phi(\hat{\mu}_{n,\omega} - \mu_\omega^0)\right\rangle_{\mathbb{L}}\right| \leq \|p\|_{\mathbb{L}} \left\|L \circ \Phi(\hat{\mu}_{n,\omega} - \mu_\omega^0)\right\|_{\mathbb{L}} .$$

We control this quantity by using the inequality $J_W(\hat{\mu}_{n,\omega}) \leq J_W(\mu_\omega^0)$ (recall (8)), along with the control on the noise term established in Lemma 3.1. This leads to the following upper bound on the Bregman divergence:

$$\mathbb{E}\left[D_\eta\left(\hat{\mu}_{n,\omega}, \mu_\omega^0\right)\right] \leq \frac{\sqrt{c_p}}{2}\rho_n(1+\sqrt{s})^2 .$$

In the same time, since η satisfies Definition 3.2, we get

$$D_\eta\left(\hat{\mu}_{n,\omega}, \mu_\omega^0\right) = \int (1-\eta)\,\mathrm{d}\hat{\mu}_{n,\omega} \geq \varepsilon_0 \hat{\mu}_{n,\omega}(\mathcal{X}^{far}(r)) + \varepsilon_2 \sum_{j=1}^{s} \int_{\mathcal{X}_j^{near}(r)} \mathfrak{d}_{\mathfrak{g}}(x, x_j^0)^2 \,\mathrm{d}\hat{\mu}_{n,\omega}(x)$$

from which we deduce the control of the estimator on the far region. The local dual certificates enable us to retrieve the control on the near regions, using again the lower bound on the Bregman divergence.

Remark 3.2 (Choice of κ, dependence on s)*.* The regularization parameter κ is chosen as $\frac{\rho_n}{\sqrt{c_p}}$ in Theorem 3.1. In particular, it does not depend on the sparsity index s of μ^0, unknown in practice—we call this choice *s-agnostic*. As a result, the user does not require prior knowledge of the target measure to select the regularization parameter: the proposed value can be directly used to solve the BLASSO numerically. The *s-dependent* choice $\kappa = \frac{\rho_n}{\sqrt{s c_p}}$ results in better rates for the estimation (*i.e.*, for the bounds presented in Theorem 3.1)—linear on $\sqrt{s}$ rather than on s (see Equation (33) in Appendix). In all cases, the estimation error depends on s; recovering a mixture with a larger number of components incurs a higher estimation cost. To conclude this discussion, we stress that these possible choices for κ are proposed according to theoretical considerations. For practical applications, this regularization parameter can be calibrated via cross-validation.

Remark 3.3 (Soft-thresholding effect). It is known that the LASSO estimator has a soft-thresholding effect, *e.g.* [Friedman et al., 2007]. It suggests that the ℓ_1-regularization is biased in the sense that each of the s weight components is under-estimated by an additive factor proportional to the regularization term κ. Having s components, we expect this bias to be of the order of $s\kappa$. Our result in (3) (Theorem 3.1) is aligned with this comment since

$$\left|\mathbb{E}\left[\|\hat{\mu}_{n,\omega}\|_{\mathrm{TV}}\right] - \|\mu_\omega^0\|_{\mathrm{TV}}\right| = \mathcal{O}(s\kappa)\,,$$

up to a constant that may depend on c_p.

Remark 3.4 (Choice of τ). Although this section focuses on the estimation of μ^0, an additional objective that can be pursued in parallel is the estimation of the associated density f^0—a task we refer to as *prediction*. This prediction objective influences the choice of the smoothing parameter. As we will see in Section 4, when $\mathcal{X} \subset \mathbb{R}^d \times [u_{\min}, +\infty)^d$ the choice $\tau = \sqrt{2}\frac{u_{\min}}{\sqrt{\ln n}}$ results in almost parametric convergence rates for both the estimation of f^0 (Theorem 4.1) and the estimation of μ_ω^0. In fact, under Assumption 1, Theorem 3.1 entails that with $\tau = \sqrt{2}\frac{u_{\min}}{\sqrt{\ln n}}$, setting $\kappa = \frac{\rho_n}{\sqrt{c_p}}$, keeping only the dependence on n and s we have

$$\mathbb{E}\left[\left|\omega_j^0 - \hat{\mu}_{n,\omega}(\mathcal{X}_j^{near}(r))\right|\right] \lesssim \frac{s(\ln n)^{d/4}}{\sqrt{n}} \quad \forall\, j = 1, \ldots, s\,.$$

If we do not need to predict the density f^0, we can alternatively take $0 < \tau \le u_{\min}$ fixed. This leads to a rate of $\frac{s}{\sqrt{n}}$ for the estimation. Note that Assumption 1 is more difficult to check when τ is large, *c.f.* Theorem 5.1. Also note that the near and far regions depend on τ through the semi-distance $\mathcal{d}$.

3.3 Effective near regions

Theorem 3.1 describes the performances of the BLASSO estimator in our setting. It provides control over the proximity between $\hat{\mu}_{n,\omega}$ and μ_ω^0 on associated far and near regions. However, our initial target is μ^0 instead of μ_ω^0. Using Theorem 3.1 and (9), we can easily give a bound for $\mathbb{E}\left[\left|a_j^0 - \frac{\hat{\mu}_{n,\omega}}{W(x_j^0)}(\mathcal{X}_j^{near}(r))\right|\right]$, but we cannot access $W(x_j^0)$ without knowledge of μ^0. In this context, a fair estimator of μ^0 is the renormalized measure $\frac{\hat{\mu}_{n,\omega}}{W}$. Providing recovery guarantees for $\frac{\hat{\mu}_{n,\omega}}{W}$ requires to control W on the near regions.

Another restrictive aspect of Theorem 3.1 is that we consider controls on regions with a fixed radius r (which is directly related to the dual certificate). We would like to locate the mass of $\hat{\mu}_{n,\omega}$ more precisely.

However, we can overcome these initial limitations by providing controls of the estimator on

$$\mathcal{X}_j^{near}(r_e) \coloneqq \{x \in \mathcal{X} \;:\; \mathcal{d}(x, x_j^0) \le r_e\} \tag{14}$$

for $r_e \le r$. Such regions are called effective near regions, and have been introduced by [De Castro et al., 2025]. To extend controls on $\mathcal{X}_j^{near}(r)$ to controls on $\mathcal{X}_j^{near}(r_e)$, the choice of $\mathfrak{d}_\mathfrak{g}$ plays an important role. It should be indeed compatible with our semi-distance $\mathcal{d}$ in a sense which is made precise in Proposition 3.1 below.

Proposition 3.1. *Let $r > 0$. Assume that Assumption 1 holds. Assume that there exists $\tilde{\varepsilon}_3 > 0$ such that for all $0 < r_e \le r$, for all $j \in \{1 \ldots, s\}$, we have*

$$\mathfrak{d}_\mathfrak{g}(x_j^0, x)^2 \ge \frac{r_e^2}{\tilde{\varepsilon}_3} \quad \forall\, x \in \mathcal{X}_j^{near}(r) \setminus \mathcal{X}_j^{near}(r_e)\,. \tag{15}$$

Then, for $\kappa = \frac{\rho_n}{\sqrt{c_p}}$ and for any $0 < r_e \le r$,

$$\mathbb{E}\left[|\omega_j^0 - \hat{\mu}_{n,\omega}(\mathcal{X}_j^{near}(r_e))|\right] \le 2\sqrt{c_p}\rho_n(1+\sqrt{s}) + \max\left\{\frac{1-\tilde{\varepsilon}_0}{\varepsilon_0}, \frac{1}{\varepsilon_2}\left(\frac{\tilde{\varepsilon}_3}{r_e^2} + \tilde{\varepsilon}_2\right)\right\}\rho_n\frac{\sqrt{c_p}}{2}(1+\sqrt{s})^2\,. \tag{16}$$

The proof is similar to that of Theorem 3.1, and is presented in Appendix F.1.

We check the assumption (15) in Lemma H.5 (in Appendix), for $\mathfrak{d}_\mathfrak{g}$ chosen as the Fisher-Rao distance. For the specific choice of r which is made in Lemma 5.4 (depending on d) it can be shown that $\tilde{\varepsilon}_3$ can be chosen as a constant not depending on d. Note that, as we decrease the size of near region r_e towards zero, the bound (16) grows as $\frac{s\rho_n}{r_e^2}$, omitting the dependence on c_p, ε_0, ε_2, $\tilde{\varepsilon}_0$, $\tilde{\varepsilon}_2$ and $\tilde{\varepsilon}_3$.

Remark 3.5. We can make choices for r_e that depend on n in Proposition 3.1. We give the associated bound for $\mathbb{E}\left[|\omega_j^0 - \hat{\mu}_{n,\omega}(\mathcal{X}_j^{near}(r_e))|\right]$ (see (16)), that holds under the assumptions of Proposition 3.1 and with n large enough such that $r_e \le r$. We only keep the dependence on n, τ and s. Let $\alpha > 0$.

- For $r_e = (\ln n)^{-\alpha}$,

$$\mathbb{E}\left[|\omega_j^0 - \hat{\mu}_{n,\omega}(\mathcal{X}_j^{near}((\ln n)^{-\alpha}))|\right] \lesssim \frac{s(\ln n)^{2\alpha}}{\sqrt{n}\tau^{d/2}} .$$

- For $r_e = n^{-\alpha}$, $\alpha < \frac{1}{4}$,

$$\mathbb{E}\left[|\omega_j^0 - \hat{\mu}_{n,\omega}(\mathcal{X}_j^{near}(n^{-\alpha}))|\right] \lesssim \frac{s}{n^{\frac{1}{2}-2\alpha}\tau^{d/2}} .$$

These choices provide an overview of possible convergence rates, and show the trade-off between localization and control of the mass.

The above proposition provides, as Theorem 3.1, guarantees on the estimation on μ_ω^0 rather than on μ^0. However, by combining Proposition 3.1 and a control of the function W on the effective regions, we can derive guarantees for the estimation of μ^0 by $\hat{\mu}_n \coloneqq \frac{\hat{\mu}_{n,\omega}}{W}$, as displayed in the following corollary.

Corollary 3.1. *We work under the same assumption as Proposition 3.1. We choose $\kappa = \frac{\rho_n}{\sqrt{c_p}}$ and $r_e = n^{-1/6}$. Assume that $0 < r_e \le r$. Omitting the dependence on d, c_p, ε_i, $\tilde{\varepsilon}_i$, r we have*

$$\mathbb{E}\left[\left|a_j^0 - \frac{\hat{\mu}_{n,\omega}}{W}(\mathcal{X}_j^{near}(n^{-1/6}))\right|\right] \lesssim \left(s\tau^{-d/2}W(x_j^0)^{-1} + a_j^0\right) n^{-1/6} .$$

The proof can be found in Appendix F.2. The difficulty that appears here is that providing control on $\frac{\hat{\mu}_{n,\omega}}{W}$ requires local control of the function W, which slightly deteriorates the bound for the estimation of the weights $(a_j^0)_{j=1}^s$. The rate $n^{-1/6}$ provides the best compromise for the size of the effective near regions r_e when dealing with this upper bound.

Remark 3.6. Making the s-dependent choice of regularization $\kappa = \frac{\rho_n}{\sqrt{c_p s}}$, a modified version of Proposition 3.1 yields, keeping only the dependence on r_e, τ, s, n,

$$\mathbb{E}\left[|\omega_j^0 - \hat{\mu}_{n,\omega}(\mathcal{X}_j^{near}(r_e))|\right] \lesssim \frac{\sqrt{s}}{\tau^{d/2}\sqrt{nr_e^2}} ,$$

c.f. (38) in Appendix. With this choice of κ, the result of Corollary 3.1 becomes

$$\mathbb{E}\left[\left|a_j^0 - \frac{\hat{\mu}_{n,\omega}}{W}(\mathcal{X}_j^{near}(n^{-1/6}))\right|\right] \lesssim \left(\sqrt{s}\tau^{-d/2}W(x_j^0)^{-1} + a_j^0\right) n^{-1/6} ,$$

see (42) in Appendix.

Remark that all the results given in this section are controls in expected value. Controls with high probability could also be given, using Lemma 3.1. Also note that the target μ^0 is fixed (does not vary with n).

4 Prediction

In Section 3, we provided controls for the estimator $\hat{\mu}_{n,\omega}$. In this section, we look at the prediction $\Phi\hat{\mu}_n = \Phi\frac{\hat{\mu}_{n,\omega}}{W}$ made by the BLASSO of the target density $\Phi\mu^0 = f^0$, and we derive bounds for the so-called prediction error $\left\|\Phi(\hat{\mu}_n - \mu^0)\right\|_{L^2}^2$. This is a question that has been quite overlooked in the BLASSO literature. We can however mention [Butucea et al., 2024] where the prediction error is investigated in a different context and for a slightly different problem.

We establish in this section almost parametric rates for the bound on the prediction error, in two distinct regimes characterized by the value of the regularization parameter κ. With small regularization, no assumptions on μ^0 are needed (Section 4.1). With stronger regularization, we show in Section 4.2 that the BLASSO achieves a good prediction when dual certificates exist (namely when Assumption 1 holds).

4.1 Prediction under small regularization

We can achieve good prediction results with small regularization, using control of the low frequencies of the estimated density (using Lemma 3.1) and of its high frequencies, resulting from an appropriate choice of τ. These controls require an upper bound on the variance. We hence assume from now on that $\mathcal{X} \subset \mathbb{R}^d \times [u_{\min}, u_{\max}]^d$.

Proposition 4.1 (Prediction error under small regularization)**.** *Assume that $\mathcal{X} \subset \mathbb{R}^d \times [u_{\min}, u_{\max}]^d$. Choosing $\tau = \frac{\sqrt{2}u_{\min}}{\sqrt{\ln n}}$ and $\kappa = \rho_n^2$, keeping only the dependence on n we have*

$$\mathbb{E}\left[\left\|\Phi(\hat{\mu}_n - \mu^0)\right\|_{L^2}^2\right] \lesssim \frac{(\ln n)^{d/2}}{n} .$$

We do not make any assumption on the existence of dual certificates in the above proposition. The proof is given in Appendix G.1. This result shows that in the small regularization regime corresponding to $\kappa = \rho_n^2$, we obtain a quasi-parametric rate (up to a log factor) for the estimation of the density f^0.

Remark 4.1 (Choice of τ, bounds on the variances). Keeping τ fixed does not lead to a good prediction rate: to control the high frequencies of the predicted density with Lemma G.1 in Appendix, we need to lower bound the variances of $\mathcal{X}$ with a parameter not depending on n, while τ must decrease as n grows. In Proposition 4.1, we also use an upper bound on the variances in $\mathcal{X}$ to control $\mathbb{E}\left[\|\hat{\mu}_n\|_{\mathrm{TV}}^2\right]$.

Remark 4.2 (Comparison with Kernel Density Estimation). Note that $L \circ \hat{f}_n$ is itself a kernel density estimator of the density (see [Tsybakov, 2008]). However, the resulting rate of convergence deteriorates very quickly with the dimension: it can be shown that with $\mathcal{X} \subset \mathbb{R}^d \times [u_{\min}, +\infty)^d$, setting $\tau = \frac{1}{\sqrt{\ln n} n^{\frac{1}{4+d}}}$, omitting the dependence on d we have

$$\mathbb{E}\left[\left\|L \circ \hat{f}_n - \Phi\mu^0\right\|_{L^2}^2\right] \lesssim \frac{(\ln n)^{d/2}}{n^{\frac{4}{d+4}} u_{\min}^{d+4}} .$$

The proof is in Appendix G.2. This bound can be explained by the fact that the Gaussian kernel is not appropriate for the estimation of so-called super-smooth target densities. Nevertheless, as displayed in Proposition 4.1, an almost parametric rate is obtained for $\Phi\hat{\mu}_n$, which is based on a regularized version of $L \circ \hat{f}_n$.

Choice of regularization Proposition 4.1 does not take advantage on the fact that μ^0 is a discrete measure. This is not surprising, as the considered regularization is very small. The choice $\kappa = \rho_n^2$ is not suited for the estimation of the target measure μ^0 (see Section 3). In fact, the result displayed in Proposition 4.1 does not take into account the trade-off we would like to have between making a good prediction and guaranteeing a good estimation of the target measure μ^0. In particular, choosing $\kappa = \rho_n^2$, we have no control over the proximity between $\left\|\mu_\omega^0\right\|_{\mathrm{TV}}$ and $\|\hat{\mu}_{n,\omega}\|_{\mathrm{TV}}$ as n grows: we only know that $\mathbb{E}\left[\|\hat{\mu}_{n,\omega}\|_{\mathrm{TV}}\right] \leq \left\|\mu_\omega^0\right\|_{\mathrm{TV}} + \frac{1}{2}$ (as $\kappa\mathbb{E}\left[\|\hat{\mu}_{n,\omega}\|_{\mathrm{TV}}\right] \leq \frac{\rho_n^2}{2} + \kappa\left\|\mu_\omega^0\right\|_{\mathrm{TV}}$), which is enough to control the high frequencies of $\Phi\hat{\mu}_n$. For an estimation purpose, we need more regularization (*i.e.* a larger κ). This is exactly what has been done in Theorem 3.1 where $\kappa = \frac{\rho_n}{\sqrt{c_p}}$.

Remark 4.3. This property of the BLASSO estimator, *i.e.* obtaining a good prediction rate under small regularization, highlights a strong difference from the LASSO framework. In the LASSO setting [Tibshirani, 1996; Tibshirani, 2023], regularization plays a greater role in controlling the prediction error: to get a parametric rate, we need to use bounds on the estimation error.

4.2 Prediction under large regularization

Provided that non-degenerate dual certificates exist, one can obtain appropriate bounds for both estimation and prediction, with κ chosen as in Theorem 3.1. The following theorem presents the prediction bound in this regime.

Theorem 4.1. *Assume that $\mathcal{X} \subset \mathbb{R}^d \times [u_{\min}, u_{\max}]^d$. Let $\tau^2 = \frac{2u_{\min}^2}{\ln n}$. If Assumption 1 holds, choosing $\kappa = \frac{\rho_n}{\sqrt{c_p}}$, keeping only the dependence on s and n we have*

$$\mathbb{E}\left[\left\|\Phi\hat{\mu}_n - \Phi\mu^0\right\|_{L^2(\mathbb{R}^d)}^2\right] \lesssim \frac{s(\ln n)^{d/2}}{n} .$$

The proof is given in Appendix G.3. The upper bound on the prediction error displayed here is slightly larger than in Proposition 4.1. In particular, this bound depends linearly on the sparsity index s. We refer for instance to [Butucea et al., 2024] for a similar bound in a different context. We recall that this specific choice for κ allows us to control the estimation performances of the estimator (see Theorem 3.1).

Remark 4.4. Similarly to the estimation task (see Remark 3.2), a choice of regularization depending on s leads to a better rate. With $\kappa = \frac{\rho}{\sqrt{c_p s}}$, we get $\mathbb{E}\left[\left\|\Phi\hat{\mu}_n - \Phi\mu^0\right\|_{L^2(\mathbb{R}^d)}^2\right] \lesssim \frac{(\ln n)^{d/2}}{n}$ for $s = \mathcal{O}(n(\ln n)^{d/2})$. We refer to Equation (45) in Appendix G.3.

5 Dual certificates

The proofs of Theorem 3.1 and Proposition 3.1 rely on the existence of non-degenerate dual certificates associated with μ^0 (Assumption 1). In this section, we explain how to construct such objects and the assumptions needed.

Connection with previous works In the general framework of BLASSO, the objective is to recover a sparse target measure from the observed signal. A key analytical tool in this setting is the dual certificate—a smooth function associated with the underlying feature map of the problem at hand. These certificates identify the locations of the target particles by interpolating the signs of the target measure.

Dual certificates can be traced back to super-resolution [Candès and Fernandez-Granda, 2014] and minimal extrapolation [De Castro and Gamboa, 2012]. They are related to the dual solutions of the BLASSO when the observation is noiseless, *c.f.* [Duval and Peyré, 2015] (in our framework, the noiseless observation corresponds to $\mathbb{E}\left[L \circ \hat{f}_n\right] = L \circ f^0$). Our particular construction of dual certificates is inspired from the BLASSO literature (*e.g.*, "vanishing derivatives pre-certificate" in [Duval and Peyré, 2015], "limit certificate" in [Poon et al., 2023]). Although our setting differs—we observe a sample drawn from some mixture distribution—the construction of the dual certificates follows the same principles: they depend solely on the target measure μ^0 and the feature map Ψ, and not on the observed data nor the regularization. Let us mention that other constructions exist for translation invariant kernels (*e.g.*, *pivot certificates* in [De Castro et al., 2021a; De Castro et al., 2025]).

The existence of certificates requires assumptions on μ^0, mainly on the separation between its particles. In [Poon et al., 2023], precise conditions are proposed, based on controls on the kernel. In this section, we adapt and check these assumptions.

Contributions Our main contribution is the construction of dual certificates for a BLASSO problem involving a kernel that is not translation invariant.

In Section 2, we have adapted the data fitting term to obtain a normalized kernel, and introduced reparametrized measures (see ($\mathcal{P}_\kappa$)). It is however challenging to check that K_{norm} satisfies the Local Positive Curvature assumption [Poon et al., 2023, Assumption 1] with the Fisher-Rao distance. We therefore consider regions for controls defined instead by the semi-distance $\mathcal{d}$ introduced in (11). Even so, controls of the certificates are expressed with the Fisher-Rao distance, since they are carried out using Taylor expansions along the Fisher-Rao geodesics.

A technical contribution is the derivation of bounds and local controls for the Riemannian derivatives of the normalized Gaussian kernel. Additionally, we establish the compatibility between the Fisher–Rao metric and the semi-distance. It is essential to control the Fisher–Rao geodesics within the balls with respect to $\mathcal{d}$ in order to make use of the local bounds on the kernel. Combining these elements, we derive sufficient conditions on the target μ^0 that guarantee the existence of non-degenerate dual certificates. The next subsections are dedicated to the proof of the following theorem.

Theorem 5.1 (Main result: existence of certificates under a minimal separation)**.** *Assume that $\mathcal{X}$ is a compact set of $\mathbb{R}^d \times [u_{\min}, u_{\max}]^d$. Let $s \geq 2$, $0 < \tau \leq u_{\min}$ and $\{x_j^0\}_{j=1}^s \subset \mathcal{X}$. We set $\mathfrak{d}_{\mathfrak{g}}$ as the Fisher-Rao distance (associated with the metric $\mathfrak{g}$ defined by (20)). If*

$$\min_{i \neq j} \mathcal{d}(x_i^0, x_j^0) \geq \max\left\{ \frac{\sqrt{u_{\max}^2 + \frac{0.3025^2}{d}(2u_{\max}^2 + \tau^2)}}{u_{\min}} \left(\Delta + \frac{0.3025}{\sqrt{d}}\right), 2\frac{u_{\max}}{u_{\min}}\Delta \right\} + \sqrt{d \ln\left(\frac{u_{\max}^2}{u_{\min}^2}\right)} \tag{17}$$

where $\Delta = 2\sqrt{11.9 + 3\ln(d + 6.62) + \ln(s-1)}$, then Assumption 1 holds with $r = \frac{0.3025}{\sqrt{d}}$, $(\varepsilon_0, \varepsilon_2) = \left(\frac{0.03911}{d}, 0.06158\right)$, $(\tilde{\varepsilon}_0, \tilde{\varepsilon}_2) = \left(\frac{0.03911}{d}, \frac{\sqrt{4d^2+10d}}{2} + 0.004106\right)$ and $c_p = 2$.

This result stems from Theorems 5.2 and 5.3 below. The calculation of the parameters is detailed in [Giard, 2025, Section VII.3].

Theorem 5.1 ensures the existence of non-degenerate dual certificates when the particles $\{x_j^0\}_{j=1}^s$ are sufficiently separated. This separation, involving both means and covariances of the Gaussian components, is expressed via the semi-distance $\mathcal{d}$, rather than the Fisher-Rao distance.

Note also that we give explicit constants, but a less constructive approach could be considered, based on the continuity of K_{norm} and its derivatives.

5.1 Geometrical framework

Kernel Building on the work of [Duval and Peyré, 2015] and [Poon et al., 2023], we consider dual certificates of the form

$$\eta_{\alpha,\beta} = \sum_{j=1}^{s} \alpha_j K_{\text{norm}}(x_j^0, \bullet) + \sum_{j=1}^{s} \beta_j^T \nabla_1 K_{\text{norm}}(x_j^0, \bullet) \tag{18}$$

where $\alpha_j \in \mathbb{R}$, $\beta_j \in \mathbb{R}^{2d}$ for all $j \in \{1, \ldots, s\}$, and K_{norm} is the real-valued kernel defined, for all $x, x' \in \mathbb{R}^d \times [u_{\min}, +\infty)^d$, by

$$K_{\text{norm}}(x, x') = \langle \Psi\delta_x, \Psi\delta_{x'} \rangle_{\mathbb{L}} = \prod_{k=1}^{d} (2u_k^2 + \tau^2)^{1/4} (2{u'_k}^2 + \tau^2)^{1/4} \frac{e^{-\frac{(t_k - t'_k)^2}{2(u_k^2 + {u'_k}^2 + \tau^2)}}}{(u_k^2 + {u'_k}^2 + \tau^2)^{1/2}} . \tag{19}$$

The kernel K_{norm} is normalized according to the definition of W (involved in Ψ). It is indeed easy to check that $K_{\text{norm}}(x, x) = 1$ for all $x \in \mathbb{R}^d \times [u_{\min}, +\infty)^d$.

The gradient of K_{norm} with respect to its first variable is written as $\nabla_1 K_{\text{norm}}$ ($\nabla_2 K_{\text{norm}}$ denotes the gradient with respect to the second variable).

Remark 5.1 (Smoothness of the feature map). Note that K_{norm} is C^∞ on $(\mathbb{R}^d \times [u_{\min}, +\infty)^d)^2$. By iteratively applying [Christmann and Steinwart, 2008, Lemma 4.34], it comes that

$$(x \mapsto \Psi\delta_x) \in \mathcal{C}^\infty(\mathbb{R}^d \times [u_{\min}, +\infty)^d, \mathbb{L})$$

and that $\langle \partial_1 \Psi\delta_x, \partial_2 \Psi\delta_{x'} \rangle_{\mathbb{L}} = \partial_1 \partial_2 K_{\text{norm}}(x, x')$, where ∂_1 (resp. ∂_2) denotes here any derivative w.r.t. to x (resp. x').

The next lemma shows the relevance of constructing η of the form (18): such functions belong to $\text{Im}(\Psi^*)$.

Lemma 5.1. *Let $\alpha \in \mathbb{R}^s$, $\beta \in \mathbb{R}^{2d \times s}$. The function $\eta_{\alpha,\beta}$ introduced in (18) verifies, for all $x \in \mathbb{R}^d \times [u_{\min}, +\infty)^d$,*

$$\eta_{\alpha,\beta}(x) = \langle p_{\alpha,\beta}, \Psi\delta_x \rangle_{\mathbb{L}} \quad \textit{with} \quad p_{\alpha,\beta} = \sum_{j=1}^{s} \alpha_j \Psi\delta_{x_j^0} + \sum_{j=1}^{s} \beta_j \nabla_x \left(\Psi\delta_{x_j^0} \right) \in \mathbb{L} .$$

Furthermore, $\eta_{\alpha,\beta} \in \mathcal{C}^\infty(\mathbb{R}^d \times [u_{\min}, +\infty)^d)$ and in particular, $\eta_{\alpha,\beta}|_{\mathcal{X}} \in \text{Im}(\Psi^*)$.

The proof is an immediate consequence of Remark 5.1 and of the definition of K_{norm} in (19).

The construction of a non-degenerate (global) dual certificate follows several steps: first we construct a function η of the form (18), choosing α_j, β_j such that $\eta(x_j^0) = 1$ and $\nabla\eta(x_j^0) = 0$. This amounts to solve some linear system (see (63) in Appendix). Then we use Taylor expansions on geodesics associated with the distance $\mathfrak{d}_{\mathfrak{g}}$ to control η on the near and far regions, in the spirit of [Poon et al., 2023]. The Fisher-Rao distance appears to be well suited for this purpose.

Fisher-Rao metric We work in a Riemannian geometry framework, and we use the Fisher-Rao metric induced by K_{norm} (*c.f.* [Poon et al., 2023, Lemma 1]), defined for $x \in \mathbb{R}^d \times [u_{\min}, +\infty)^d$ by

$$\mathfrak{g}_x \coloneqq \nabla_1 \nabla_2 K_{\text{norm}}(x, x) = \text{diag}\left(\frac{1}{2u_1^2 + \tau^2}, \ldots, \frac{1}{2u_d^2 + \tau^2}, \frac{2u_1^2}{(2u_1^2 + \tau^2)^2}, \ldots, \frac{2u_d^2}{(2u_d^2 + \tau^2)^2} \right) . \tag{20}$$

We will use the associated norm, defined by

$$\|v\|_x = \sqrt{v^T \mathfrak{g}_x v} \quad \forall\, v \in \mathbb{R}^{2d},\ x \in \mathbb{R}^d \times [u_{\min}, +\infty)^d .$$

Details about this metric are provided in Appendix H, together with a description of the associated geodesics. We recall that a geodesic for the metric $\mathfrak{g}$ between $x, x' \in \mathbb{R}^d \times [u_{\min}, +\infty)^d$ is a piecewise continuously differentiable function $\gamma : [0, 1] \to \mathbb{R}^d \times [u_{\min}, +\infty)^d$ such that $\gamma(0) = x, \gamma(1) = x'$, minimizing the quantity $\int_0^1 \|\dot{\gamma}(y)\|_{\gamma(y)}\, \mathrm{d}y$. In dimension $d = 1$, our Fisher-Rao geodesics share the same paths as those of the Poincaré half-plane model (*c.f.* Lemma H.2 in Appendix): they are portions of straight lines parallel to $\{t = 0\}$ and semicircles whose origin is on $\{u = 0\}$. The notation $\mathfrak{d}_{\mathfrak{g}}$ refers to the associated distance. It is defined by $\mathfrak{d}_{\mathfrak{g}}(x, x') = \int_0^1 \|\dot{\gamma}(y)\|_{\gamma(y)}\, \mathrm{d}y$ with γ the geodesic between x, x'. This Fisher-Rao distance is not practical to define near and far regions since it is not directly linked to the correlation between 2 features calculated with the kernel. This motivates the use of another distance on $\mathcal{X}$ to express the recovery results: we use the semi-distance defined by

$$d(x, x') = \sqrt{-2 \ln(K_{\text{norm}}(x, x'))} , \tag{21}$$

whose expression is given by (11).

Riemannian derivatives Following the work of [Poon et al., 2023], we will show that the dual certificates we construct are non-degenenerate (see Definitions 3.2 and 3.3) by controlling the Riemannian derivatives of K_{norm}. For details about this framework, see [Poon et al., 2023, p.263].

Riemannian derivatives involve the Christoffel symbols associated with $\mathfrak{g}$. We will use the notation $\Gamma^{t_k}, \Gamma^{u_k}$, and refer to (46) in Appendix H.1 for a precise definition.

Definition 5.1. Let $\psi \in \mathcal{C}^2(\mathbb{R}^d \times [u_{\min}, +\infty]^d)$. Let $x, x' \in \mathbb{R}^d \times [u_{\min}, +\infty]^d$.
Riemannian Hessian: we define

$$H^{\mathfrak{g}}\psi(x) = \nabla^2\psi(x) - \sum_{k=1}^{d} \Gamma^{t_k}\partial_{t_k}\psi(x) - \sum_{k=1}^{d} \Gamma^{u_k}\partial_{u_k}\psi(x) .$$

Covariant derivatives: Let $v, v' \in \mathbb{R}^{2d}$. We define

$$D_0[\psi](x) = \psi(x) , \quad D_1[\psi](x)[v] = v^T\nabla\psi(x) , \quad D_2[\psi](x)[v, v'] = v^T H^{\mathfrak{g}}\psi(x)v' .$$

Operator norms: For $j \in \{0, 1, 2\}$, we define the operator norm

$$\|D_j[\psi](x)\|_x \coloneqq \sup_{\substack{V=[v_1,\dots,v_j]\in(\mathbb{R}^{2d})^j \\ \forall l=1,\dots,j\,,\ \|v_l\|_x\le 1}} D_j[\psi](x)[V] .$$

Kernel derivatives and associated operators: Let $i, j \in \{0, 1, 2\}$. We define the covariant derivative of the kernel of order i with respect to the first variable x and of order j with respect to the second variable x' by

$$[Q]K^{(ij)}_{\text{norm}}(x, x')[V] = \langle D_i[\Psi](x)[Q], D_j[\Psi](x')[V]\rangle_{\mathbb{L}} \quad \forall Q \in (\mathbb{R}^{2d})^i , V \in (\mathbb{R}^{2d})^j .$$

The associated operator norm is

$$\left\|K^{(ij)}_{\text{norm}}(x, x')\right\|_{x,x'} \coloneqq \sup_{\substack{Q=[Q_1,\dots,Q_i]\in(\mathbb{R}^{2d})^i\,,\ V=[V_1,\dots,V_j]\in(\mathbb{R}^{2d})^j \\ \forall l\ \|Q_l\|_x\,,\ \|V_l\|_x\le 1}} [Q]K^{(ij)}_{\text{norm}}(x, x')[V] .$$

Simplified expressions are given in Appendix (Lemma J.1).

5.2 Construction of dual certificates

In what follows, we give some results allowing us to adapt and check the hypothesis of [Poon et al., 2023, Theorem 2], which shows that the (global) non-degenerate certificate exists under some conditions on the kernel. The corresponding proof can be adapted (see for instance [Poon et al., 2023, Section 6.7]) to show the existence of additional certificates for the near regions, under the same conditions.

Local positive curvature assumption The following definition allows us to adapt [Poon et al., 2023, Assumption 1] to our framework. In particular, we require some smoothness and structural properties for the kernel that enable the construction of dual certificates, as stated below.

Definition 5.2 (Kernel of local positive curvature with parameters s, Δ, r, $\bar{\varepsilon}_0$ and $\bar{\varepsilon}_2$)**.** Let K be a real-valued normalized kernel of positive type, in $C^3((\mathbb{R}^d \times [u_{\min}, +\infty)^d)^2)$. It is said to satisfy the local positive curvature assumption (LPC) if the following holds:

1. For all $i, j \in 0, 1, 2$ with $i + j \le 3$,
$$\sup_{x,x'\in\mathbb{R}^d\times[u_{\min},+\infty)^d} \left\|K^{(ij)}(x, x')\right\|_{x,x'} \le B_{ij} < +\infty .$$
For $i = 0, 1, 2$, we denote $B_i \coloneqq 1 + B_{0i} + B_{1i}$.
2. There exists $r > 0$ such that K has strictly positive curvature constants $\bar{\varepsilon}_0$ and $\bar{\varepsilon}_2$ with
$$K(x, x') \le 1 - \bar{\varepsilon}_0 , \quad \forall x, x' \in \mathbb{R}^d \times [u_{\min}, +\infty)^d \text{ s.t. } \mathfrak{d}(x, x') \ge r ,$$
$$-K^{(02)}(x, x')[v, v] \ge \bar{\varepsilon}_2\|v\|^2_{x'} , \quad \forall v \in \mathbb{R}^{2d} , \quad \forall x, x' \in \mathbb{R}^d \times [u_{\min}, +\infty)^d \text{ s.t. } \mathfrak{d}(x, x') < r .$$
3. There exist $s \ge 2$, $\Delta(s) > 0$ such that for all $\{x_l\}_{l=1}^s \in \mathcal{S}_{\Delta(s)}$,
$$\sum_{l=2}^{s} \left\|K^{(ij)}(x_1, x_l)\right\|_{x_1,x_l} \le \frac{1}{64} \min\left(\frac{\bar{\varepsilon}_0(r)}{B_0}, \frac{\bar{\varepsilon}_2(r)}{B_2}\right) \quad \forall (i, j) \in \{0, 1\} \times \{0, 1, 2\} ,$$
where $\mathcal{S}_\Delta \coloneqq \left\{\{x_l\}_{l=1}^s \subset \mathbb{R}^d \times [u_{\min}, +\infty)^d \,:\, \min_{m\ne l} \mathfrak{d}(x_m, x_l) \ge \Delta\right\}$.

Such kernel is said to verify the LPC with parameters s, $\Delta(s)$, r, $\bar{\varepsilon}_0(r)$ and $\bar{\varepsilon}_2(r)$.

Remark 5.2. This definition differs on some points from [Poon et al., 2023, Assumption 1]. First, since we solve ($\mathcal{P}_\kappa$) for nonnegative measures, we do not have to control the negative part of the certificates, and thus do not require $r < B_{02}^{-1/2}$. Moreover, we use controls on regions defined by the semi-distance d instead of $\mathfrak{d}_\mathfrak{g}$. We need to provide bounds on the kernel on $\mathbb{R}^d \times [u_{\min}, +\infty)^d$ and not $\mathcal{X}$, because a Fisher-Rao geodesic between 2 points of $\mathcal{X}$ could be outside $\mathcal{X}$. The use of a semi-distance, which better describes the correlation between 2 features (as expressed by the kernel), also leads to difficulties. We need to ensure that d is compatible with the Fisher-Rao distance. This is detailed in the paragraph below.

Compatibility between the Fisher-Rao metric and the semi-distance For $x \in \mathbb{R}^d \times [u_{\min}, +\infty)^d$ and $R > 0$, we denote

$$B_d(x, R) \coloneqq \{x' \in \mathbb{R}^d \times [u_{\min}, +\infty)^d \ : \ d(x, x') \leq R\}$$

and

$$\mathring{B}_d(x, R) \coloneqq \{x' \in \mathbb{R}^d \times [u_{\min}, +\infty)^d \ : \ d(x, x') < R\}\,.$$

For $A \subset \mathbb{R}^d \times [u_{\min}, +\infty)^d$, we define the set containing all points from geodesics connecting 2 points of A,

$$\mathcal{G}(A) \coloneqq \{\gamma(y) \ : \ \gamma \text{ is a Fisher-Rao geodesic}\,,\ y \in [0,1]\,,\ \gamma(0), \gamma(1) \in A\}\,.$$

Lemma 5.2 (Fisher-Rao geodesics remain within balls w.r.t. the semi-distance)**.** *Let $r > 0$ and $x_0 \in \mathbb{R}^d \times [u_{\min}, +\infty)^d$. Then*

$$\mathcal{G}\left(B_d(x_0, r)\right) \subset B_d(x_0, r)\,.$$

The proof is detailed in Appendix H.3.1. This lemma allows us to use the controls we have on the balls $B_d(x_j^0, r)$ on the paths of the geodesics between 2 points in a ball. It entails that d and $\mathfrak{d}_\mathfrak{g}$ are in some sense compatible.

To control the certificates, we also require that certain balls (with respect to the semi-distance) around the particles be disjoint. We restrict the possible values for the variances to establish this property: we need upper and lower bounds to establish a "pseudo-quasi triangle inequality" for d (see Appendix H.3.2).

Lemma 5.3 (Separation of the balls w.r.t. the semi-distance)**.** *Let $r, \Delta > 0$. Assume that $\mathcal{X} \subset \mathbb{R}^d \times [u_{\min}, u_{\max}]^d$. If*

$$\min_{i \neq j} d(x_i^0, x_j^0) \geq \max\left\{\frac{\sqrt{u_{\max}^2 + r^2(2u_{\max}^2 + \tau^2)}}{u_{\min}}(\Delta + r)\,,\ 2\frac{u_{\max}}{u_{\min}}\Delta\right\} + \sqrt{d \ln\left(\frac{u_{\max}^2}{u_{\min}^2}\right)} \eqqcolon \Delta_\tau\,,$$

then the balls $\mathring{B}_d(x_j^0, \Delta) \cap \mathcal{X}$ are disjoint, and for all $j \neq i$, $\mathcal{G}(B_d(x_j^0, r) \cap \mathcal{X})$ does not intersect $\mathring{B}_d(x_i^0, \Delta)$.

The proof is given in Appendix H.3.2. We stress that this lemma leads to the separation condition in Theorem 5.1. It is crucial to manage the kernel on the far and near regions according to the local positive curvature assumption.

Non-degeneration of certificates under a minimal separation Using the compatibility between the semi-distance and the Fisher-Rao metric, and controls on K_{norm}, we can show the existence of non-degenerate certificates under some condition on the minimal separation between the particles of μ^0. This is what the following theorem entails.

Theorem 5.2. *Assume that $\mathcal{X} \subset \mathbb{R}^d \times [u_{\min}, u_{\max}]^d$ and that K_{norm} satisfies the LPC (Definition 5.2) with parameters $s, \Delta, r, \bar{\varepsilon}_0, \bar{\varepsilon}_2$.*

If $\{x_j^0\}_{j=1}^s \subset \mathcal{X}$ satisfies $\min_{i \neq j} d(x_i^0, x_j^0) \geq \Delta_\tau$ (defined in Lemma 5.3), then there exists a $\left(\varepsilon_0 = \frac{7}{8}\bar{\varepsilon}_0, \varepsilon_2 = \frac{15}{32}\bar{\varepsilon}_2, r\right)$-global non-degenerate certificate η of the form (18).

Under the same assumptions, for all $j = 1, \dots, s$ there exists a certificate η_j for the jth near region of the form (18), of parameters $\left(\tilde{\varepsilon}_0 = \frac{7}{8}\bar{\varepsilon}_0, \tilde{\varepsilon}_2 = \frac{B_{02} + \bar{\varepsilon}_2/16}{2}, r\right)$.

Moreover, $\eta|_{\mathcal{X}} = \Psi^* p$ and $\eta_j|_{\mathcal{X}} = \Psi^* p_j$ where $\|p\|_{\mathbb{L}} \leq \sqrt{2s}$ and $\|p_j\|_{\mathbb{L}} \leq \sqrt{2}$.

The proof is based on an adaptation of [Poon et al., 2023, Theorem 2], and is given in Appendix I.

It remains to prove that K_{norm} satisfies the LPC—this is the purpose of the following theorem, whose proof is provided in Appendix J. We give general bounds for K_{norm} in dimension $d \geq 1$, and we provide a tighter constant for $\Delta(s)$ in the case $d = 1$.

Theorem 5.3. *Let $s \geq 2$, $d \geq 1$. Assume that $\mathcal{X} \subset \mathbb{R}^d \times [u_{\min}, u_{\max}]^d$ and that $\tau \leq u_{\min}$. Then K_{norm} satisfies the LPC with parameters s, $r = \frac{0.3025}{\sqrt{d}}$, $\bar{\varepsilon}_2(r) = 0.13139$, $\bar{\varepsilon}_0(r) = \frac{0.0894}{2d}$, and*

$$\Delta(s) = 2\sqrt{11.9 + 3\ln(d + 6.62) + \ln(s-1)}\,.$$

Moreover, we can take $B_{02} = \sqrt{4d^2 + 10d}$ (item 1 of Definition 5.2).

Remark 5.3. When $d = 1$, the minimal separation consition can be improved: K_{norm} satisfies the LPC with the same parameters $r, \bar{\varepsilon}_0, \bar{\varepsilon}_2$ as in Theorem 5.3, with $\Delta(s) = 2\sqrt{13.88 + \ln(s-1)}$ (Appendix J).

Remark 5.4. Note that the assumption on the separation between particles given in Theorem 5.2 depends on τ through the semi-distance. We write $\mathfrak{d} = \mathfrak{d}_\tau$ to emphasize this dependency. At first sight, the separation condition thus appears to depend on n when choosing $\tau^2 = \frac{2u_{\min}^2}{\ln n}$. However, as $\tau \in \mathbb{R}^+ \mapsto \mathfrak{d}_\tau(x, x')$ is decreasing and $\tau \in \mathbb{R}^+ \mapsto \Delta_\tau$ is increasing, if $\{x_j^0\}_{j=1}^s$ satisfies $\min_{i \neq j} \mathfrak{d}_{\tau_1}(x_i^0, x_j^0) \geq \Delta_{\tau_1}$ for some $\tau_1 > 0$, then it also satisfies $\min_{i \neq j} \mathfrak{d}_{\tau_2}(x_i^0, x_j^0) \geq \Delta_{\tau_2}$ for $0 < \tau_2 \leq \tau_1$. This motivates the introduction of the following assumption:

$$\min_{x_i^0 \neq x_j^0} \mathfrak{d}_{\tau_{\max}}(x_i^0, x_j^0) \geq \Delta_{\tau_{\max}} \quad \text{and} \quad \forall j = 1, \ldots, s\,,\ x_j^0 \in \mathcal{X} \subset \mathbb{R}^d \times [u_{\min}, u_{\max}]^d\,, \quad \tau_{\max} \leq u_{\min}\,. \tag{22}$$

The recovery guarantees are not expressed using this assumption, but it is useful when thinking about the convergence of the solution towards μ_ω^0 as $n \to \infty$. The target is fixed, but for a good prediction τ must decrease with the number of observations. The constraint (22) provides in this context a condition on μ^0 that does not depend on n.

Controls on the effective near regions Theorem 5.3 sets the size of near regions we consider for the certificates. We can lower-bound the Fisher-Rao distance with the semi-distance on these regions, allowing us to extend the control of the estimator to near regions of smaller size (Proposition 3.1). This result is key to provide recovery guarantees for the proximity of μ^0 and the renormalized estimator $\frac{\hat{\mu}_{n,\omega}}{W}$ (*c.f.* Corollary 3.1). Its proof can be found in Appendix H.4.

Lemma 5.4. *Assume that $x, x_0 \in \mathbb{R}^d \times [u_{\min}, +\infty)^d$. If $r_e \leq \mathfrak{d}(x, x_0) \leq r = \frac{0.3025}{\sqrt{d}}$, then $\mathfrak{d}_{\mathfrak{g}}(x, x_0)^2 \geq \frac{r_e^2}{\tilde{\varepsilon}_3}$ with $\tilde{\varepsilon}_3 = 2.84$.*

6 Sparsity of the solution under large sample sizes

The certificates constructed in Section 5.2 satisfy the requirements of Assumption 1. Handling such certificates allows us to obtain the non-asymptotic recovery guarantees presented in Section 3. These results are stated in terms of control of the mass in the far and near regions. Nevertheless, they say nothing about the sparsity of the estimator.

In this section, we present results of a different nature. Our estimator is here an exact solution $\mu_{n,\omega}^\star$, verifying (7). Following [Duval and Peyré, 2015], we show that for large sample sizes and with high probability, $\mu_{n,\omega}^\star$ is sparse and has exactly 1 particle in each near region. We stress that in the following, $0 < \tau \leq u_{\min}$ is fixed (does not decrease as $n \to \infty$).

Non Degenerate Source Condition We present here the key tool of our analysis, based again on non-degenerate dual certificates (Definition 3.2), with additional properties. We show the existence of a global non-degenerate certificate η_{NDSC} verifying $\eta_{\text{NDSC}} > -1$. It is said to satisfy the Non Degenerate Source Condition (NDSC).

Lemma 6.1 (Non-degeneration of η_{NDSC})**.** *Let $s \geq 2$ and assume that $\mathcal{X} \subset \mathbb{R}^d \times [u_{\min}, u_{\max}]^d$. There exist $r_{NDSC} > 0$ depending only on d, and $\Delta_{NDSC} > 0$ depending only on $u_{\min}, u_{\max}, d, s, \tau$ such that if $\{x_j^0\}_{j=1}^s$ satisfies $\min_{x_i^0 \neq x_j^0} \mathfrak{d}(x_i^0, x_j^0) \geq \Delta_{NDSC}$, then there exists a $(\varepsilon_{0,NDSC}, \varepsilon_{2,NDSC}, r_{NDSC})$-non-degenerate certificate of the form (18) (of parameters only depending on d), that we denote by η_{NDSC}, verifying*

$$|\eta_{NDSC}| \leq 1\,, \quad |\eta_{NDSC}(x)| = 1 \iff x \in \{x_j^0\}_j\,, \quad \nabla^2 \eta_{NDSC}(x_j^0) \prec 0 \quad \forall\, j = 1, \ldots, s\,. \tag{NDSC}$$

It is said to satisfy the Non Degenerate Source Condition (NDSC).

Under the same assumptions, for all $j = 1, \ldots, s$ there exists a certificate $\eta_{j,NDSC}$ for the jth near region of the form (18), of parameters $(\tilde{\varepsilon}_{0,NDSC}, \tilde{\varepsilon}_{2,NDSC}, r_{NDSC})$ only depending on d.

Moreover, $\eta_{NDSC}|_{\mathcal{X}} = \Psi^* p_{NDSC}$ and $\eta_{j,NDSC}|_{\mathcal{X}} = \Psi^* p_{j,NDSC}$ where $\|p_{NDSC}\|_{\mathbb{L}} \leq \sqrt{2s}$ and $\|p_{j,NDSC}\|_{\mathbb{L}} \leq \sqrt{2}$.

The proof is in Appendix K.1. We construct η_{NDSC} in the same way as in Theorem 5.2. The main difference with Theorem 5.1 is the condition $\eta_{\text{NDSC}} > -1$, which is verified under a stronger separation condition on $\{x_j^0\}_{j=1}^s$.

The certificate η_{NDSC} from Lemma 6.1 is—as the global dual certificate constructed in Theorem 5.1—the vanishing derivative pre-certificate of [Duval and Peyré, 2015, Section 4]. We can then show (see Lemma K.2 in Appendix) that η_{NDSC} corresponds to the minimal norm certificate [Duval and Peyré, 2015, Proposition 7]: we have $\eta_{\text{NDSC}}|_{\mathcal{X}} = \Psi^* p_{0,0}$ where

$$p_{0,0} \coloneqq \underset{p\in\mathbb{L}}{\arg\min} \left\{ \|p\|_{\mathbb{L}} \;:\; \Psi^* p \in \partial \left\|\mu^0\right\|_{\text{TV}} \right\} , \tag{$p_{0,0}$}$$

with $\partial \|\cdot\|_{\text{TV}}$ denoting the subdifferential of the TV-norm. This certificate allows us to control the sparsity index of $\mu^\star_{n,\omega}$ for large sample sizes, leading to the following theorem.

Theorem 6.1. *Under the assumptions of Lemma 6.1, there exists $\kappa_0 > 0$ and $\gamma_0 > 0$ (depending on $\mathcal{X}$, τ and μ^0) such that for all $\kappa \leq \kappa_0$ and if $\|\Gamma_n\|_{\mathbb{L}} \leq \gamma_0\kappa$, then $\mu^\star_{n,\omega}$ is s-sparse and has exactly 1 particle in each $\mathcal{X}_j^{near}(r_{\text{NDSC}})$.*

The proof can be found in Appendix K.2.

This result should be seen in conjunction with Lemma 3.1. The latter indicates that the assumption $\|\Gamma_n\|_{\mathbb{L}} \leq \gamma_0\kappa_0$ is verified with high probability for n large enough. We detail this in the following corollary.

Corollary 6.1. *Let $c_\kappa > 0$. Assume that the conditions given in Lemma 6.1 hold. Let $n \geq \frac{c_\kappa^2}{\kappa_0^2 (2\pi)^{d/2}\tau^d}$ and $\kappa = \frac{c_\kappa}{(2\pi)^{d/4}\tau^{d/2}\sqrt{n}}$. With probability greater than $1 - C_\Gamma e^{-\left(\frac{\gamma_0 c_\kappa}{C_\Gamma}\right)^2}$,*

$$\mu^\star_{n,\omega} = \sum_{j=1}^{s} \omega_j^\star \delta_{x_j^\star} \quad \textit{with} \quad \omega_j^\star > 0\,, \quad x_j^\star \in \mathcal{X}_j^{near}(r_{\text{NDSC}}) \quad \forall\, j = 1,\dots,s\,.$$

Moreover,

$$|\omega_j^0 - \omega_j^\star| \leq c_0 \frac{c_\kappa}{\sqrt{n}} \tag{23}$$

where $c_0 > 0$ depends on $\mathcal{X}$, τ and μ^0. In addition, if c_κ is chosen as $c_{\kappa,n} = \underset{n\to+\infty}{o}(\sqrt{n})$, there exists $n_0 \in \mathbb{N}$ depending on $(c_{\kappa,n})_n, \mathcal{X}, \tau, \mu^0$ such that for all $n \geq n_0$,

$$d(x_j^\star, x_j^0)^2 \leq \tilde{c}_0 \frac{c_{\kappa,n}}{\sqrt{n}}$$

with probability at least $1 - C_\Gamma e^{-\left(\frac{\gamma_0 c_{\kappa,n}}{C_\Gamma}\right)^2}$, where $\tilde{c}_0$ depends on $\mathcal{X}$, τ and μ^0.

The proof is given in Appendix K.3. The result is established for a generic value of c_κ, but different specific choices can be considered. For instance, choosing $c_{\kappa,n} = \alpha\sqrt{\ln n}$ with $\alpha > 0$, there exists $n_{0,\alpha} \in \mathbb{N}$ depending on $\mathcal{X}, \tau, \mu^0, \alpha$ such that for all $n \geq n_{0,\alpha}$ and with probability at least $1 - C_\Gamma n^{-\frac{\gamma_0^2\alpha^2}{C_\Gamma^2}}$, $\mu^\star_{n,\omega} = \sum_{j=1}^s \omega_j^\star \delta_{x_j^\star}$ where for all $j = 1,\dots,s$,

$$|\omega_j^0 - \omega_j^\star| \leq c_0 \alpha \frac{\sqrt{\ln n}}{\sqrt{n}} \quad \text{and} \quad d(x_j^\star, x_j^0)^2 \leq \tilde{c}_0 \alpha \frac{\sqrt{\ln n}}{\sqrt{n}}\,. \tag{24}$$

Compared to Theorem 3.1, Inequality (24) yields more classical results in parametric estimation. We indeed obtain parametric rates of convergence (up to a logarithmic factor) for the estimation of both weight and location parameters. Moreover, the estimator has a sparsity index exactly matching that of the target measure μ^0. These bounds hold under slightly more restrictive conditions and for sufficiently large sample sizes. We also stress that these results are obtained under a specific tuning of the parameters κ and τ, which differs from that used in the previous sections.

7 Discussion and open problems

In this paper, we have described some theoretical properties of the BLASSO in the framework of Gaussian mixtures. In particular, we have considered the case where each underlying component has an unknown diagonal covariance. At this stage, further improvements and extensions are still possible and are discussed below.

Algorithmic considerations Recall that Theorems 3.1 and 4.1 apply with any $\hat{\mu}_{n,\omega}$ that verifies $J_W(\hat{\mu}_{n,\omega}) \leq J_W(\mu^0_\omega)$: our estimator is not necessarily the solution of $(\mathcal{P}_\kappa)$. So we can restrict $(\mathcal{P}_\kappa)$ to any subset of $\mathcal{M}(\mathcal{X})^+$ containing μ^0_ω.

This remark is interesting from an algorithmic point of view. We do not know how to solve $(\mathcal{P}_\kappa)$ numerically on the entire measure space $\mathcal{M}(\mathcal{X})^+$, due to the absence of a parametrization of this space. In Hardy, 2023, a version of the BLASSO over the space of K-sparse measures (discrete measures with less than K particles) is considered. This problem, although not convex, is closer to a realistic algorithmic framework, as we can parametrize the space of K-sparse measures. If $K \geq s$, μ^0_ω belongs to this space.

Several algorithms have been proposed to solve the BLASSO on sparse measures, such as the sliding Frank-Wolfe algorithm [Denoyelle et al., 2019] or the Conic Particle Gradient Descent (CPGD, see [Chizat, 2022]). Adapting the latter to our framework could form the core of a future work.

Non-diagonal covariance matrices In this contribution, we are only dealing with the case where the covariances of the mixture are diagonal. Note that we can easily extend our model and results to the case where all covariances share the same (known) orientation, *i.e.* we can diagonalize them in the same basis. However, we are not yet able to process general covariances with varying or unknown orientations. Indeed, the Fisher-Rao metric seems impractical to work with, mainly because it might be tricky to prove the compatibility between the associated distance and the semi-distance, and obtain the desired controls on the kernel. A possible outcome could be to consider an alternative metric.

Minimal separation condition Our results are established under a separation condition for the mixture components. In [De Castro et al., 2021a], the separation between the particles can go to zero as n grows, because a translation-invariant kernel is used. Such kernels have been investigated in [De Castro et al., 2025] and this analysis enables the construction of certificates based on a "pivot" kernel, whose decay can increase when dealing with closer particles. This is not the case here: our kernel is not practical to build a pivot, and its decay does not allow us to consider $\Delta \to 0$ (for $(t,u) \neq (t',u') \in \mathbb{R} \times [u_{\min}, +\infty)$, $e^{-\frac{(t-t')^2}{2(u^2+u'^2+\tau^2)}}$ cannot be as close to 0 as wanted by changing τ). Moreover, the use of the semi-distance $\mathscr{d}$ imposes a larger minimal separation.

Statistical learning of Gaussian Mixtures Gaussian mixtures have been at the core of several investigations and it is not possible to provide a complete state of the art in a single paragraph. Different tasks can be considered, ranging from clustering [Chen and Zhang, 2024], testing [Donoho and Jin, 2004] or estimating the component parameters. For the latter, different settings and related assumptions have been considered. Estimating the mixture parameters is an hard task [Anandkumar et al., 2012; Doss et al., 2023], even in the case where the component covariances are known. Our approach, based on convex optimization on the space of measures, leads to specific bounds related to the mass of the estimator on far and near regions. A complete comparison with alternative approaches may require additional investigations that are outside the scope of this paper.

Technical side notes We have used the semi-distance $\mathscr{d}$ to define the near and far regions, and the Fisher-Rao metric to control the certificates with Taylor expansions. These are somewhat arbitrary or improvable choices. The semi-distance $\mathscr{d}$ seems appropriate to control the kernel, because it measures precisely how 2 points are spaced for the kernel. But the downside is that it does not satisfy the triangle inequality. As a consequence, we need to take Δ_τ much larger than Δ. The Fisher-Rao metric allows us to retrieve quite easily global bounds for the kernel (*i.e.* evaluations of the bounds B_{ij}, see item 1 of Definition 5.2), but we could have used the Euclidean metric—although it leads to controls that depend on bounds on the variance.

Summary of the main results

The target measure is $\mu^0 = \sum_{j=1}^s a^0_j \delta_{(t^0_j, u^0_j)} \in \mathcal{M}(\mathcal{X})$. We assume in the following that $\mathcal{X}$ is a compact set of $\mathbb{R}^d \times [u_{\min}, u_{\max}]^d$, and that $n \geq 2$, $s \geq 2$, $0 < \tau \leq u_{\min}$. The results displayed in the following table require more assumptions, that we explicit here. Keep in mind that $\mathscr{d}$ depends on τ.

Assumption 2.

$$\min_{i \neq j} \mathscr{d}(x^0_i, x^0_j) \geq \max\left\{ \frac{\sqrt{u^2_{\max} + \frac{0.3025^2}{d}(2u^2_{\max} + \tau^2)}}{u_{\min}} \left(\Delta + \frac{0.3025}{\sqrt{d}}\right), 2\frac{u_{\max}}{u_{\min}}\Delta \right\} + \sqrt{d \ln\left(\frac{u^2_{\max}}{u^2_{\min}}\right)}$$

where $\Delta = 2\sqrt{11.9 + 3\ln(d + 6.62) + \ln(s-1)}$.

Assumption 3.

$$\min_{i\neq j} \mathfrak{d}(x_i^0, x_j^0) \geq \Delta_{\text{NDSC}}$$

(see Lemma 6.1—Δ_{NDSC} depends on $s, \tau, d, u_{\min}, u_{\max}$).

Some constants are omitted in the bounds, this is expressed using $\lesssim_c$ when the dependence on c is not taken into account.

Table 1: Recovery guarantees, κ not depending on s

	Parameters	Asm.	Bounds
Estimation (Prop. 3.1, Thm. 5.1, Lem. 5.4, Cor. 3.1)	$\kappa = \frac{\sqrt{2}}{(2\pi\tau^2)^{d/4}\sqrt{n}}$	2	For $0 < r_e \leq \frac{0.3025}{\sqrt{d}}$, $\mathbb{E}\left[\left\lvert\hat{\mu}_{n,\omega}(\mathcal{X}_j^{\text{near}}(r_e)) - \mu_\omega^0(\mathcal{X}_j^{\text{near}}(r_e))\right\rvert\right] \lesssim_d \frac{s}{r_e^2\sqrt{n}\tau^{d/2}}$ For $n^{-1/6} \leq \frac{0.3025}{\sqrt{d}}$, $\mathbb{E}\left[\left\lvert a_j^0 - \frac{\hat{\mu}_{n,\omega}}{W}(\mathcal{X}_j^{\text{near}}(n^{-1/6}))\right\rvert\right] \lesssim_d \left(\frac{s}{\tau^{d/2}W(x_j^0)} + a_j^0\right) n^{-1/6}$
Sparsity (Cor. 6.1)	$c_{\kappa,n} > 0$, $c_{\kappa,n} = o(\sqrt{n})$ $\kappa = \frac{c_{\kappa,n}}{(2\pi\tau^2)^{d/4}\sqrt{n}}$	3	For $n \geq n_0$ (depends on $\mu^0, \mathcal{X}, \tau, (c_{\kappa,n})_n$), with prob. $\geq 1 - C_\Gamma e^{-\left(\frac{\gamma_0 c_{\kappa,n}}{C_\Gamma}\right)^2}$: $\mu_{n,\omega}^\star = \sum_{j=1}^s \omega_j^\star \delta_{x_j^\star}$ and $\mathfrak{d}(x_j^\star, x_j^0)^2 \lesssim_{\mu^0,\mathcal{X},\tau} \frac{c_{\kappa,n}}{\sqrt{n}}$, $\lvert\omega_j^0 - \omega_j^\star\rvert \lesssim_{\mu^0,\mathcal{X},\tau} \frac{c_{\kappa,n}}{\sqrt{n}}$
Prediction with small regularization (Prop. 4.1)	$\tau = \frac{\sqrt{2}u_{\min}}{\sqrt{\ln n}}$ $\kappa = \frac{4(\ln n)^{d/2}}{(4\pi u_{\min 2})^{d/2} n}$	none	$\mathbb{E}\left[\lVert\Phi\frac{\hat{\mu}_{n,\omega}}{W} - \Phi\mu^0\rVert_{L^2}^2\right] \lesssim_{d,u_{\min},u_{\max}} \frac{(\ln n)^{d/2}}{n}$
Prediction with a good estimator (Thm. 4.1)	$\tau = \frac{\sqrt{2}u_{\min}}{\sqrt{\ln n}}$ $\kappa = \frac{\sqrt{2}(\ln n)^{d/2}}{(4\pi u\min^2)^{d/4}\sqrt{n}}$	2	$\mathbb{E}\left[\lVert\Phi\frac{\hat{\mu}_{n,\omega}}{W} - \Phi\mu^0\rVert_{L^2}^2\right] \lesssim_{d,u_{\min},u_{\max}} \frac{s(\ln n)^{d/2}}{n}$

Table 2: Recovery guarantees, κ depending on s

	Parameters	Asm.	Bounds
Estimation, v2 (Rk. 3.6, Thm. 5.1, Lem. 5.4)	$\kappa = \frac{\sqrt{2}}{(2\pi\tau^2)^{d/4}\sqrt{sn}}$	2	For $0 < r_e \leq \frac{0.3025}{\sqrt{d}}$, $\mathbb{E}\left[\left\lvert\hat{\mu}_{n,\omega}(\mathcal{X}_j^{\text{near}}(r_e)) - \mu_\omega^0(\mathcal{X}_j^{\text{near}}(r_e))\right\rvert\right] \lesssim_d \frac{\sqrt{s}}{r_e^2\sqrt{n}\tau^{d/2}}$ For $n^{-1/6} \leq \frac{0.3025}{\sqrt{d}}$, $\mathbb{E}\left[\left\lvert a_j^0 - \frac{\hat{\mu}_{n,\omega}}{W}(\mathcal{X}_j^{\text{near}}(n^{-1/6}))\right\rvert\right] \lesssim_d \left(\frac{\sqrt{s}}{\tau^{d/2}W(x_j^0)} + a_j^0\right) n^{-1/6}$
Prediction with a good estimator, v2 (Rk. 4.4)	$\tau = \frac{\sqrt{2}u_{\min}}{\sqrt{\ln n}}$ $\kappa = \frac{\sqrt{2}}{(2\pi\tau^2)^{d/4}\sqrt{sn}}$	2	If $s = \mathcal{O}(n(\ln n)^{d/2})$, $\mathbb{E}\left[\lVert\Phi\frac{\hat{\mu}_{n,\omega}}{W} - \Phi\mu^0\rVert_{L^2}^2\right] \lesssim_{d,u_{\min},u_{\max}} \frac{(\ln n)^{d/2}}{n}$

Notation

Table 3: Table of notations

Global notation	
$\mathcal{X}$	compact of $\mathbb{R}^d \times [u_{\min}, +\infty)^d$ with $u_{\min}, u_{\max} > 0$
$\mathcal{M}(\mathcal{X})^+$	nonnegative Radon measures on $\mathcal{X}$
W	reparametrization function. For $x = ((t_1, \dots, t_d), (u_1, \dots, u_d)) \in \mathbb{R}^d \times [u_{\min}, +\infty)^d$, $W(x) = \prod_{k=1}^d (2\pi)^{-1/4}(2u_k^2 + \tau^2)^{-1/4}$
μ^0, μ^0_ω	resp. the target probability measure $\sum_{j=1}^s a_j^0 \delta_{x_j^0}$ where $x_j^0 = (t_j^0, u_j^0) \in \mathcal{X}$; its reparametrized version $W\mu^0 = \sum_{j=1}^s \omega_j^0 \delta_{x_j^0}$
φ, σ, Φ	resp. the function $z \in \mathbb{R} \mapsto \frac{e^{-\frac{z^2}{2}}}{\sqrt{2\pi}}$; its Fourier transform; the operator $\mu \mapsto \left(z \in \mathbb{R}^d \mapsto \int_{\mathbb{R}^d \times [u_{\min}, +\infty)^d} \prod_{k=1}^d \frac{1}{u_k} \varphi\left(\frac{z_k - t_k}{u_k}\right) \mathrm{d}\mu(t, u) \right)$
$X_1, \dots, X_n$	i.i.d. observations in $\mathbb{R}^d$, drawn from $f^0 = \Phi\mu^0$
$\mathbb{E}$	expected value w.r.t. $X_1, \dots, X_n$
$\hat{f}_n$	empirical density $\frac{1}{n}\sum_{i=1}^n \delta_{X_i}$
κ	regularization constant
$\hat{\mu}_{n,\omega}, \hat{\mu}_n, \mu^\star_{n,\omega}$	resp. a measure of $\mathcal{M}(\mathcal{X})^+$ such that $J_W(\hat{\mu}_{n,\omega}) \leq J_W(\mu^0_\omega)$ (see ($\mathcal{P}_\kappa$)); the measure $\frac{\hat{\mu}_{n,\omega}}{W}$; an exact solution of ($\mathcal{P}_\kappa$)
Γ_n, ρ_n^2	resp. the noise $L \circ \hat{f}_n - L \circ f^0$; a bound on $\mathbb{E}\left[\|\Gamma_n\|_{\mathbb{L}}^2\right]$ equal to $\frac{4}{(2\pi)^{d/2}\tau^d n}$
Kernel, differential geometry	
τ, λ, Λ, L	resp. the smoothing parameter τ; the function $z \in \mathbb{R}^d \mapsto \frac{e^{-\frac{\|z\|_2^2}{2\tau^2}}}{(2\pi\tau^2)^{d/2}}$; its Fourier transform; the operator $f \mapsto \lambda * f$
$\mathbb{L}$	RKHS associated with λ (scalar product given by (5))
Ψ, K_{norm}	resp. the feature map $\mu \in \mathcal{M}(\mathbb{R}^d \times [u_{\min}, +\infty)^d) \mapsto L \circ \Phi \frac{\mu}{W}$; the normalized kernel $(x, x') \in \mathbb{R}^d \times [u_{\min}, +\infty)^d \mapsto \langle \Psi\delta_x, \Psi\delta_{x'} \rangle_{\mathbb{L}}$
$\mathfrak{d}$	semi-distance $(x, x') \mapsto \sqrt{-2\ln(K_{\text{norm}}(x, x'))}$ (expression given by (11))
$\mathcal{X}_j^{near}(r), \mathcal{X}^{far}(r)$	resp. the jth near region of radius r, $\{x \in \mathcal{X} : \mathfrak{d}(x, x_j^0) \leq r\}$; the far region $\mathcal{X} \setminus \bigcup_{j=1}^s \mathcal{X}_j^{near}(r)$
η	dual certificate for μ^0, of the form $\Psi^* p$ with $p \in \mathbb{L}$
$D_\eta(\hat{\mu}_{n,\omega}, \mu^0_\omega)$	Bregman divergence $\|\hat{\mu}_{n,\omega}\|_{\text{TV}} - \|\mu^0_\omega\|_{\text{TV}} - \int_{\mathcal{X}} \eta \, \mathrm{d}\left(\hat{\mu}_{n,\omega} - \mu^0_\omega\right)$
$\mathfrak{g}, \mathfrak{d}_{\mathfrak{g}}$	resp. the Fisher-Rao metric (see (20)); the associated distance
$\gamma, \tilde{\gamma}$	geodesic for the Fisher-Rao metric (parametrized on $[0, 1]$ and by arc-length respectively)

References

Anandkumar A., Hsu D., and Kakade S. M. (2012). "A method of moments for mixture models and hidden Markov models". *Conference on learning theory*. JMLR Workshop and Conference Proceedings, pp. 33–1.

Arcones M. A. and Giné E. (1993). "Limit Theorems for U-Processes". *The Annals of Probability* 21.3, pp. 1494–1542. DOI: 10.1214/aop/1176989128.

Balakrishnan S., Wainwright M. J., and Yu B. (2017). "Statistical guarantees for the EM algorithm: From population to sample-based analysis". *The Annals of Statistics* 45.1, pp. 77–120. DOI: 10.1214/16-AOS1435.

Beardon A. F. (1983). *The geometry of discrete groups*. Springer. DOI: 10.1007/978-1-4612-1146-4.

Bouveyron C., Celeux G., Murphy T. B., and Raftery A. E. (2019). *Model-based clustering and classification for data science: with applications in R*. Cambridge series in statistical and probabilistic mathematics. Cambridge University Press. DOI: 10.1017/9781108644181.

Boyer C., De Castro Y., and Salmon J. (2017). "Adapting to unknown noise level in sparse deconvolution". *Information and Inference: A Journal of the IMA* 6.3, pp. 310–348. DOI: 10.1093/imaiai/iaw024.

Bruni C. and Koch G. (1985). "Identifiability of Continuous Mixtures of Unknown Gaussian Distributions". *The Annals of Probability* 13.4, pp. 1341–1357. DOI: 10.1214/aop/1176992817.

Butucea C., Delmas J.-F., Dutfoy A., and Hardy C. (2024). "Off-the-grid prediction and testing for linear combination of translated features". *Electronic Journal of Statistics* 18.2, pp. 2931–2973. DOI: 10.1214/24-EJS2267.

Candès E. J. and Fernandez-Granda C. (2014). "Towards a Mathematical Theory of Super-resolution". *Communications on Pure and Applied Mathematics* 67. DOI: 10.1002/cpa.21455.

Chen X. and Zhang A. Y. (2024). "Achieving optimal clustering in Gaussian mixture models with anisotropic covariance structures". *Advances in Neural Information Processing Systems* 37, pp. 113698–113741.

Chizat L. (2022). "Sparse optimization on measures with over-parameterized gradient descent". *Mathematical Programming* 194.1, pp. 487–532. DOI: 10.1007/s10107-021-01636-z.

Christmann A. and Steinwart I. (2008). *Support vector machines*. Springer. DOI: 10.1007/978-0-387-77242-4.

De Castro Y., Gadat S., and Marteau C. (2023). *FastPart: Over-Parameterized Stochastic Gradient Descent for Sparse optimisation on Measures*. arXiv: 2312.05993 [math.OC].

De Castro Y., Gadat S., Marteau C., and Maugis C. (2021a). "SuperMix: Sparse Regularization for Mixtures". *Annals of Statistics* 49.3, pp. 1779–1809. DOI: 10.1214/20-AOS2022.

De Castro Y., Gadat S., Marteau C., and Maugis-Rabusseau C. (2021b). "Supplement to "SuperMix: Sparse regularization for mixtures"". *Annals of Statistics* 49.3. DOI: 10.1214/20-AOS2022SUPP.

De Castro Y. and Gamboa F. (2012). "Exact reconstruction using Beurling minimal extrapolation". *Journal of Mathematical Analysis and Applications* 395.1, pp. 336–354. DOI: 10.1016/j.jmaa.2012.05.011.

De Castro Y., Gribonval R., and Jouvin N. (2025). *Effective regions and kernels in continuous sparse regularisation, with application to sketched mixtures*. arXiv: 2507.08444 [math.ST].

Denoyelle Q., Duval V., Peyré G., and Soubies E. (2019). "The sliding Frank–Wolfe algorithm and its application to super-resolution microscopy". *Inverse Problems* 36.1, p. 014001. DOI: 10.1088/1361-6420/ab2a29.

Donoho D. and Jin J. (2004). "Higher criticism for detecting sparse heterogeneous mixtures". *The Annals of Statistics* 32.3, pp. 962–994. DOI: 10.1214/009053604000000265.

Doss N., Wu Y., Yang P., and Zhou H. H. (2023). "Optimal estimation of high-dimensional Gaussian location mixtures". *The Annals of Statistics* 51.1, pp. 62–95. DOI: 10.1214/22-AOS2207.

Duval V. and Peyré G. (2015). "Exact Support Recovery for Sparse Spikes Deconvolution". *Foundations of Computational Mathematics* 15.5, pp. 1315–1355. DOI: 10.1007/s10208-014-9228-6.

Friedman J., Hastie T., Höfling H., and Tibshirani R. (2007). "Pathwise coordinate optimization". *The Annals of Applied Statistics* 1.2, pp. 302–332. DOI: 10.1214/07-AOAS131.

Fruhwirth-Schnatter S., Celeux G., and Robert C. P. (2019). *Handbook of mixture analysis*. Chapman & Hall/CRC Press. DOI: 10.1201/9780429055911.

Giard R. (2025). *giardr/blasso_gmm_sympy: Symbolic computation for the BLASSO applied to GMMs with diagonal covariances*. Zenodo. Version v1.1. DOI: 10.5281/zenodo.16942355.

Hardy C. (2023). "Estimation and testing of mixtures of Hilbert-valued features issued from a continuous dictionary". PhD thesis. École des Ponts ParisTech.

Houdard A., Bouveyron C., and Delon J. (2018). "High-Dimensional Mixture Models for Unsupervised Image Denoising (HDMI)". *SIAM Journal on Imaging Sciences* 11.4, pp. 2815–2846. DOI: 10.1137/17M1135694.

Houdré C. and Reynaud-Bouret P. (2003). "Exponential Inequalities, with Constants, for U-statistics of Order Two". *Stochastic Inequalities and Applications*. Basel: Birkhäuser Basel, pp. 55–69. DOI: 10.1007/978-3-0348-8069-5_5.

McLachlan G. J. and Krishnan T. (1997). *The EM algorithm and extensions*. John Wiley & Sons. DOI: 10.1002/9780470191613.

McLachlan G. J. and Peel D. (2000). *Finite mixture models*. Vol. 299. Probability and Statistics – Applied Probability and Statistics Section. Wiley. DOI: 10.1002/0471721182.

Murphy K. P. (2012). *Machine learning : a probabilistic perspective*. MIT Press. ISBN: 9780262018029.

Poon C. (2019). *An introduction to sparse spikes recovery via the BLASSO [Lecture Notes]*. URL: https://cmhsp2.github.io/files/teaching/sparsity/blasso.pdf.

Poon C., Keriven N., and Peyré G. (2023). "The geometry of off-the-grid compressed sensing". *Foundations of Computational Mathematics* 23, pp. 241–327. DOI: 10.1007/s10208-021-09545-5.

Rudin W. (1991). *Functional Analysis*. International series in pure and applied mathematics. McGraw-Hill. ISBN: 9780070542365.

Stahl S. (1993). *The Poincaré half-plane: A gateway to modern geometry*. Jones & Bartlett Learning. ISBN: 086720298X.

Teicher H. (1961). "Identifiability of Mixtures". *The Annals of Mathematical Statistics* 32.1, pp. 244–248. DOI: 10.1214/aoms/1177705155.

Tibshirani R. (1996). "Regression Shrinkage and Selection via the Lasso". *Journal of the Royal Statistical Society. Series B (Methodological)* 58.1, pp. 267–288. DOI: 10.1111/j.2517-6161.1996.tb02080.x.

Tibshirani R. (2023). *High-Dimensional Regression: Lasso [Lecture Notes]*. URL: https://www.stat.berkeley.edu/~ryantibs/statlearn-s23/lectures/lasso.pdf.

Tsybakov A. B. (2008). *Introduction to Nonparametric Estimation*. 1st. Springer Publishing Company, Incorporated. DOI: 10.1007/b13794.

Wu C. F. J. (1983). "On the Convergence Properties of the EM Algorithm". *The Annals of Statistics* 11.1, pp. 95–103. DOI: 10.1214/aos/1176346060.

Yakowitz S. J. and Spragins J. D. (1968). "On the Identifiability of Finite Mixtures". *The Annals of Mathematical Statistics* 39.1, pp. 209–214. DOI: 10.1214/aoms/1177698520.

Gaussian Mixture Model with unknown diagonal covariances via continuous sparse regularization

—

Appendix

A Functional framework

The BLASSO operates on the space of Radon measures. In this section, we provide definitions of the operators on measures used throughout the paper.

Let $A \subset \mathbb{R}^d$. In the following, $\mathcal{L}^\infty(A)$ is the set of bounded functions from A to $\mathbb{R}$; $L^1(A)$ (resp. $L^2(A)$) the set of functions whose absolute value (resp. square) have a finite integral. The notions of convolution and Fourier transform can be extended to measures.

Definition A.1 (Convolution $g * \mu$)**.** Let $A \subset \mathbb{R}^p$. The convolution between $g \in \mathcal{L}^\infty(A)$ and $\mu \in \mathcal{M}(A)$ is defined by

$$g * \mu \coloneqq \int_A g(\bullet - x)\mathrm{d}\mu(x)\,.$$

Definition A.2 (Fourier transform over $L^1(A)$ and $\mathcal{M}(A)$)**.** Let $A \subset \mathbb{R}^p$. We use the Fourier transform defined for $g \in L^1(A)$ by

$$\mathcal{F}\left[g\right](\xi) \coloneqq \int_A e^{-i\langle \xi, x\rangle} g(x)\,\mathrm{d}x\,.$$

We also use its extension to $\mathcal{M}(A)$:

$$\mathcal{F}\left[\mu\right](\xi) \coloneqq \int_A e^{-i\langle \xi, x\rangle}\,\mathrm{d}\mu(x) \quad \forall\, \mu \in \mathcal{M}(A)\,.$$

The standard properties of convolution and Fourier transform apply, such as $\mathcal{F}\left[g * \mu\right] = \mathcal{F}\left[g\right]\mathcal{F}\left[\mu\right]$ for all $g \in L^1(A) \cap \mathcal{L}^\infty(A)$ and all $\mu \in \mathcal{M}(A)$ (see, for example, [Rudin, 1991, Part 2]).

B Existence of a solution to the BLASSO

Proposition B.1. *The problem* ($\mathcal{P}_\kappa$) *has a solution.*

Proof. J_W is lower semi-continuous on $\mathcal{M}(\mathcal{X})^+$ for the weak* convergence: The TV norm is lower semi-continuous for the weak* convergence. It then suffices to show that $\mu \mapsto L \circ \Phi \frac{\mu}{W}$ is weak* to weak continuous. Let $\mu_j \overset{*}{\rightharpoonup} \mu$ in $\mathcal{M}(\mathcal{X})$. We want to show that for $f \in \mathbb{L}$,

$$\left\langle L \circ \Phi \frac{\mu_j}{W}, f \right\rangle_{\mathbb{L}} \to \left\langle L \circ \Phi \frac{\mu}{W}, f \right\rangle_{\mathbb{L}}\,.$$

Using [Christmann and Steinwart, 2008, Lemma 4.29], as

$$K_{\mathrm{norm}} : (x, x') \in (\mathbb{R}^d \times [u_{\min}, +\infty)^d)^2 \mapsto \left\langle L \circ \Phi \frac{\delta_x}{W}, L \circ \Phi \frac{\delta_{x'}}{W} \right\rangle_{\mathbb{L}}$$

is continuous (see (19)), we have

$$\left(x \mapsto L \circ \Phi \frac{\delta_x}{W}\right) \in \mathcal{C}(\mathcal{X}, \mathbb{L})\,.$$

Hence

$$\left(x \mapsto \left\langle L \circ \Phi \frac{\delta_x}{W}, f \right\rangle_{\mathbb{L}}\right) \in \mathcal{C}(\mathcal{X})\,.$$

As $\mathcal{M}(\mathcal{X}) = \mathcal{C}(\mathcal{X})^*$, it comes

$$\int \left\langle L \circ \Phi \frac{\delta_x}{W}, f \right\rangle_{\mathbb{L}} \mathrm{d}\mu_j(x) \to \int \left\langle L \circ \Phi \frac{\delta_x}{W}, f \right\rangle_{\mathbb{L}} \mathrm{d}\mu(x)\,,$$

which concludes this part of the proof.
Conclusion: The result follows noticing that $J_W(\mu) \underset{\|\mu\|_{\mathrm{TV}}\to\infty}{\longrightarrow} \infty$: we can restrict the problem to a closed ball of $\mathcal{M}(\mathcal{X})^+$. By the Banach–Alaoglu theorem, this ball is weakly* compact (a ball of $\mathcal{M}(\mathcal{X})$ is weakly* compact, and $\mathcal{M}(\mathcal{X})^+$ is a weakly* closed subspace of $M(\mathcal{X})$). We deduce the existence of a minimizer of J_W on $\mathcal{M}(\mathcal{X})^+$. □

C Proof of Lemma 3.1

Expected value of $\|\Gamma_n\|_{\mathbb{L}}^2$: Note that, according to the definition of Γ_n,

$$\Gamma_n = \frac{1}{n}\sum_{i=1}^n \left(L\circ\delta_{X_i} - L\circ f^0\right) .$$

For all $h\in\mathbb{L}$, for all $i\in\{1,\dots,n\}$, we have $\mathbb{E}\left[\langle L\circ\delta_{X_i}, h\rangle_{\mathbb{L}}\right] = \left\langle L\circ f^0, h\right\rangle_{\mathbb{L}}$. This entails that $\mathbb{E}\left[\langle\Gamma_n, h\rangle_{\mathbb{L}}\right]=0$. Defining

$$Z_i \coloneqq L\circ\delta_{X_i} - L\circ f^0$$

(where we recall that $f^0=\Phi\mu^0$), observe that $Z_1,\dots,Z_n$ are i.i.d. Since for all $i\in\{1,\dots,n\}$,

$$\|Z_i\|_{\mathbb{L}}^2 = \frac{1}{(2\pi)^d}\int_{\mathbb{R}^d}\Lambda|\mathcal{F}\left[\delta_{X_i}-f^0\right]|^2 \le \frac{4\int_{\mathbb{R}^d}\Lambda}{(2\pi)^d} ,$$

we have

$$\begin{aligned}\|\Gamma_n\|_{\mathbb{L}}^2 &= \frac{1}{n^2}\sum_{i=1}^n\|Z_i\|_{\mathbb{L}}^2 + \frac{1}{n^2}\sum_{i\neq j}\langle Z_i, Z_j\rangle_{\mathbb{L}} ,\\ &\le \frac{4\int_{\mathbb{R}^d}\Lambda}{(2\pi)^d n} + \frac{1}{n^2}\sum_{i\neq j}\langle Z_i, Z_j\rangle_{\mathbb{L}} .\end{aligned}$$

We deduce that $\mathbb{E}\left[\|\Gamma_n\|_{\mathbb{L}}^2\right] \le \frac{4\int_{\mathbb{R}^d}\Lambda}{(2\pi)^d n}$ using that $\mathbb{E}[Z_i]=0$ for all $i\in\{1,\dots,n\}$.
Control in probability: The control in probability of $\|\Gamma_n\|_{\mathbb{L}}^2$ comes from [De Castro et al., 2021b, Lemma 3], and stems from results on U-processes (see [Arcones and Giné, 1993, Proposition 2.3]). In particular, we have

$$\forall\rho>0\,,\quad \mathbb{P}\left(\|\Gamma_n\|_{\mathbb{L}}^2 > \rho\frac{C_\Gamma^2}{n\tau^d(2\pi)^{d/2}}\right) \le C_\Gamma e^{-\rho}\,, \tag{25}$$

for some positive constant $C_\Gamma>0$.
Expected value of $\|\Gamma_n\|_{\mathbb{L}}^4$: Using (25), it comes that

$$\begin{aligned}\mathbb{E}\left[\|\Gamma_n\|_{\mathbb{L}}^4\right] &= \int_0^\infty \mathbb{P}\left(\|\Gamma_n\|_{\mathbb{L}}^4 > x\right)\mathrm{d}x\,,\\ &= \int_0^\infty \mathbb{P}\left(\|\Gamma_n\|_{\mathbb{L}}^2 > \sqrt{x}\right)\mathrm{d}x\,,\\ &\le \int_0^\infty C_\Gamma e^{-\sqrt{x}\frac{n\tau^d(2\pi)^{d/2}}{C_\Gamma^2}}\mathrm{d}x\,,\end{aligned}$$

so

$$\mathbb{E}\left[\|\Gamma_n\|_{\mathbb{L}}^4\right] \le C_\Gamma\frac{C_\Gamma^4}{n^2\tau^{2d}(2\pi)^{2d}}\int_0^\infty e^{-\sqrt{x}}\mathrm{d}x = \frac{2C_\Gamma^5}{n^2\tau^{2d}(2\pi)^{2d}} .$$

This proves the desired result.

D Proof of Theorem 3.1

The following proof is standard when dealing with the BLASSO procedure. Its main arguments can be found in [De Castro et al., 2021a]; we adapt their proof to our setting. It is based on bounds on the Bregman divergence introduced in (13). With W defined by (6), we denote in the following $\hat{\mu}_n = \frac{\hat{\mu}_{n,\omega}}{W}$. We also recall from (9) that

$\mu^0 = \frac{\mu^0_\omega}{W}$.

Upper bound on the Bregman divergence: Since $J_W(\hat{\mu}_{n,\omega}) \leq J_W(\mu^0_\omega)$, we get

$$\left\| L \circ \hat{f}_n - L \circ \Phi \hat{\mu}_n \right\|_{\mathbb{L}}^2 + 2\kappa \left\| \hat{\mu}_{n,\omega} \right\|_{\mathrm{TV}} \leq \left\| L \circ \hat{f}_n - L \circ \Phi \mu^0 \right\|_{\mathbb{L}}^2 + 2\kappa \left\| \mu^0_\omega \right\|_{\mathrm{TV}} \tag{26}$$

which can be rewritten as

$$\left\| L \circ \hat{f}_n - L \circ \Phi \hat{\mu}_n \right\|_{\mathbb{L}}^2 + 2\kappa D_\eta \left(\hat{\mu}_{n,\omega}, \mu^0_\omega \right) + 2\kappa \int_{\mathcal{X}} \eta \, \mathrm{d}(\hat{\mu}_{n,\omega} - \mu^0_\omega) \leq \left\| \Gamma_n \right\|_{\mathbb{L}}^2 , \tag{27}$$

for any dual certificate η satisfying the requirements of Assumption 1. As $\eta = \Psi^* p$, recalling (10) we have

$$\int_{\mathcal{X}} \eta \, \mathrm{d}(\hat{\mu}_{n,\omega} - \mu^0_\omega) = \left\langle p, L \circ \Phi(\hat{\mu}_n - \mu^0) \right\rangle_{\mathbb{L}} .$$

So using Cauchy-Schwarz inequality, (27) leads to

$$\left\| L \hat{f}_n - L \circ \Phi \hat{\mu}_n \right\|_{\mathbb{L}}^2 + 2\kappa D_\eta \left(\hat{\mu}_{n,\omega}, \mu^0_\omega \right) - 2\kappa \left\| p \right\|_{\mathbb{L}} \left\| L \circ \Phi \mu^0 - L \circ \Phi \hat{\mu}_n \right\|_{\mathbb{L}} \leq \left\| \Gamma_n \right\|_{\mathbb{L}}^2 .$$

The triangle inequality

$$\left\| L \circ \Phi \mu^0 - L \circ \Phi \hat{\mu}_n \right\|_{\mathbb{L}} \leq \left\| \Gamma_n \right\|_{\mathbb{L}} + \left\| L \hat{f}_n - L \circ \Phi \hat{\mu}_n \right\|_{\mathbb{L}}$$

then leads to

$$\left(\left\| L \circ \hat{f}_n - L \circ \Phi \hat{\mu}_n \right\|_{\mathbb{L}} - \kappa \left\| p \right\|_{\mathbb{L}} \right)^2 + 2\kappa D_\eta \left(\hat{\mu}_{n,\omega}, \mu^0_\omega \right) \leq \left(\left\| \Gamma_n \right\|_{\mathbb{L}} + \kappa \left\| p \right\|_{\mathbb{L}} \right)^2 .$$

As the Bregman divergence is positive, we deduce from the previous inequality that

$$D_\eta \left(\hat{\mu}_{n,\omega}, \mu^0_\omega \right) \leq \frac{\left\| \Gamma_n \right\|_{\mathbb{L}}^2}{2\kappa} + \frac{\kappa}{2} \left\| p \right\|_{\mathbb{L}}^2 + \left\| \Gamma_n \right\|_{\mathbb{L}} \left\| p \right\|_{\mathbb{L}} \quad \text{and} \quad \left\| L \circ \hat{f}_n - L \circ \Phi \hat{\mu}_n \right\|_{\mathbb{L}} \leq \left\| \Gamma_n \right\|_{\mathbb{L}} + 2\kappa \left\| p \right\|_{\mathbb{L}} . \tag{28}$$

The bound in expected value on the Bregman divergence follows by applying Lemma 3.1 and using $\left\| p \right\|_{\mathbb{L}} \leq \sqrt{c_p s}$. We have

$$\mathbb{E} \left[D_\eta \left(\hat{\mu}_{n,\omega}, \mu^0_\omega \right) \right] \leq \frac{\rho_n^2}{2\kappa} + \kappa \frac{c_p}{2} s + \rho_n \sqrt{c_p s} . \tag{29}$$

Taking $\kappa = \frac{\rho_n}{\sqrt{c_p}}$ gives

$$\mathbb{E} \left[D_\eta \left(\hat{\mu}_{n,\omega}, \mu^0_\omega \right) \right] \leq \frac{\sqrt{c_p}}{2} \rho_n (1 + \sqrt{s})^2 . \tag{30}$$

Lower bound on the Bregman divergence: We use the controls of the non-degenerate certificate η on the near and far regions. Since $\hat{\mu}_{n,\omega}$ is nonnegative, according to Definition 3.2 we have

$$D_\eta \left(\hat{\mu}_{n,\omega}, \mu^0_\omega \right) = \int (1 - \eta) \, \mathrm{d}\hat{\mu}_{n,\omega} \geq \varepsilon_0 \hat{\mu}_{n,\omega}(\mathcal{X}^{far}(r)) + \varepsilon_2 \sum_{j=1}^{s} \int_{\mathcal{X}_j^{near}(r)} \mathfrak{d}_{\mathfrak{g}}(x, x_j^0)^2 \, \mathrm{d}\hat{\mu}_{n,\omega}(x) . \tag{31}$$

Combining (30) and (31), we deduce the bound for the far region (item 1 of Theorem 3.1). To control the mass of the estimator on the jth near region, we make use of the local non-degenerate certificate η_j (Definition 3.3). We have, for all $j = 1, \ldots, s$,

$$\begin{aligned}
|\omega_j^0 - \hat{\mu}_{n,\omega}(\mathcal{X}_j^{near}(r))| &= \left| \omega_j^0 - \int \eta_j \, \mathrm{d}\hat{\mu}_{n,\omega} + \int \eta_j \, \mathrm{d}\hat{\mu}_{n,\omega} - \int_{\mathcal{X}_j^{near}(r)} \mathrm{d}\hat{\mu}_{n,\omega} \right| , \\
&\leq \left| \int \eta_j \, \mathrm{d}(\mu^0_\omega - \hat{\mu}_{n,\omega}) \right| + \int_{\mathcal{X}^{near}(r) \setminus \mathcal{X}_j^{near}(r)} |\eta_j| \, \mathrm{d}\hat{\mu}_{n,\omega} + \int_{\mathcal{X}_j^{near}(r)} |1 - \eta_j| \, \mathrm{d}\hat{\mu}_{n,\omega} \\
&\quad + \int_{\mathcal{X}^{far}(r)} |\eta_j| \mathrm{d}\hat{\mu}_{n,\omega} , \\
&\leq | \left\langle p_j, L \circ \Phi \mu^0 - L \circ \Phi \hat{\mu}_n \right\rangle_{\mathbb{L}} | + \tilde{\varepsilon}_2 \sum_{l=1}^{s} \int_{\mathcal{X}_l^{near}(r)} \mathfrak{d}_{\mathfrak{g}}(x, x_l^0)^2 \, \mathrm{d}\hat{\mu}_{n,\omega}(x) \\
&\quad + (1 - \tilde{\varepsilon}_0) \hat{\mu}_{n,\omega}(\mathcal{X}^{far}(r)) , \\
&\leq \left\| p_j \right\|_{\mathbb{L}} \left(2 \left\| \Gamma_n \right\|_{\mathbb{L}} + 2\kappa \left\| p \right\|_{\mathbb{L}} \right) + \max \left\{ \frac{1 - \tilde{\varepsilon}_0}{\varepsilon_0}, \frac{\tilde{\varepsilon}_2}{\varepsilon_2} \right\} D_\eta \left(\hat{\mu}_{n,\omega}, \mu^0_\omega \right) ,
\end{aligned} \tag{32}$$

where for the last inequality, we have used (31), and

$$|\langle p_j, L\circ\Phi\mu^0 - L\circ\Phi\hat{\mu}_n\rangle_{\mathbb{L}}| \le \|p_j\|_{\mathbb{L}}\left(\|\Gamma_n\|_{\mathbb{L}} + \left\|L\circ\hat{f}_n - L\circ\Phi\hat{\mu}_n\right\|_{\mathbb{L}}\right) \le \|p_j\|_{\mathbb{L}}\left(2\|\Gamma_n\|_{\mathbb{L}} + 2\kappa\|p\|_{\mathbb{L}}\right)$$

together with (28). As $\|p_j\|_{\mathbb{L}} \le \sqrt{c_p}$, $\|p\|_{\mathbb{L}} \le \sqrt{c_p s}$ and $\kappa = \frac{\rho_n}{\sqrt{c_p}}$, we finally have

$$|\omega_j^0 - \hat{\mu}_{n,\omega}(\mathcal{X}_j^{near}(r))| \le 2\sqrt{c_p}(\|\Gamma_n\|_{\mathbb{L}} + \rho_n\sqrt{s}) + \max\left\{\frac{1-\tilde{\varepsilon}_0}{\varepsilon_0}, \frac{\tilde{\varepsilon}_2}{\varepsilon_2}\right\} D_\eta\left(\hat{\mu}_{n,\omega}, \mu_\omega^0\right) .$$

The result in expected value follows from $\mathbb{E}\left[\|\Gamma_n\|_{\mathbb{L}}\right] \le \rho_n$ (Lemma 3.1 used with Jensen's inequality) and (30).

For the stability of the mass (item 3), we use a similar but simpler reasoning. As $\eta \le 1$ and $\hat{\mu}_{n,\omega}$ is nonnegative, we have

$$\begin{aligned}
\|\mu_\omega^0\|_{\mathrm{TV}} - \|\hat{\mu}_{n,\omega}\|_{\mathrm{TV}} &= \int \eta\,(\mathrm{d}\mu_\omega^0 - \mathrm{d}\hat{\mu}_{n,\omega}) + \int(\eta-1)\,\mathrm{d}\hat{\mu}_{n,\omega}\,,\\
&\le \int \eta\,(\mathrm{d}\mu_\omega^0 - \mathrm{d}\hat{\mu}_{n,\omega})\,,\\
&= \langle p, L\circ\Phi\mu^0 - L\circ\Phi\hat{\mu}_n\rangle_{\mathbb{L}}\,,\\
&\le \|p\|_{\mathbb{L}}\left(2\|\Gamma_n\|_{\mathbb{L}} + 2\kappa\|p\|_{\mathbb{L}}\right).
\end{aligned}$$

Using (26), we also have $\|\hat{\mu}_{n,\omega}\|_{\mathrm{TV}} \le \|\mu_\omega^0\|_{\mathrm{TV}} + \frac{1}{2\kappa}\|\Gamma_n\|_{\mathbb{L}}^2$. We can conclude by taking the expectation in these inequalities and using Lemma 3.1.

With an s-dependent choice of regularization: Choosing $\kappa = \frac{\rho_n}{\sqrt{c_p s}}$, (29) gives $\mathbb{E}\left[D_\eta(\hat{\mu}_{n,\omega}, \mu_\omega^0)\right] \le 2\rho_n\sqrt{c_p s}$. It comes that

$$\mathbb{E}\left[|\omega_j^0 - \hat{\mu}_{n,\omega}(\mathcal{X}_j^{near}(r))|\right] \le 2\sqrt{c_p}\rho_n(1+\sqrt{s}) + \max\left\{\frac{1-\tilde{\varepsilon}_0}{\varepsilon_0}, \frac{\tilde{\varepsilon}_2}{\varepsilon_2}\right\} 2\rho_n\sqrt{c_p s}\,. \tag{33}$$

E Basic inequalities

The following lemma gives inequalities useful when dealing with the semi-distance d. We will use them in various proofs.

Lemma E.1 (Basic inequalities)**.** *Let $a, b \in \mathbb{R}_+^*$, $c \ge 1$. Then*

$$\frac{a^2+b^2}{2ab} \le c \iff a \in \left[b(c-\sqrt{c^2-1}), b(c+\sqrt{c^2-1})\right] \tag{34}$$

and

$$\frac{a^2+b^2}{2ab} \le c \implies \frac{|a^2-b^2|}{a^2+b^2} \le \sqrt{c^2-1} \tag{35}$$

along with

$$\frac{a^2+b^2}{2ab} \le c \implies \frac{2a^2}{a^2+b^2} \le c + \sqrt{c^2-1}\,. \tag{36}$$

Proof. Let $a, b \in \mathbb{R}_+^*$ and $c \ge 1$.

Proof of (34): We have

$$\frac{a^2+b^2}{2ab} \le c \iff a^2+b^2-2cab \le 0\,.$$

As $4c^2b^2 - 4b^2 = 4b^2(c^2-1) \ge 0$, the roots of this polynomial in a are

$$\frac{2cb \pm 2b\sqrt{c^2-1}}{2} = b(c \pm \sqrt{c^2-1})$$

from which we deduce (34).

Proof of (35): Using (34) and $2ab \le a^2+b^2$, we get

$$\begin{aligned}
\frac{|a^2-b^2|}{a^2+b^2} &\le \frac{|a-b||a+b|}{2ab}\,,\\
&\le \frac{\min\{a,b\}(1+c+\sqrt{c^2-1})\max\{a,b\}(1-(c-\sqrt{c^2-1}))}{2ab}\,,\\
&= \frac{(1+c+\sqrt{c^2-1})(1+\sqrt{c^2-1}-c)}{2}\,,\\
&= \sqrt{c^2-1}\,.
\end{aligned}$$

Proof of (36): Using again (34) with $2ab \leq a^2 + b^2$, we have

$$\frac{2a^2}{a^2+b^2} \leq \frac{2ab(c+\sqrt{c^2-1})}{2ab}, \\ \leq c + \sqrt{c^2-1}.$$

□

F Proofs related to the control of the estimator on the effective near regions

F.1 Proof of Proposition 3.1

The proof is similar to that of Theorem 3.1 in Section D. Let $j \in \{1, \ldots, s\}$ and η_j a corresponding local non-degenerate dual certificate satisfying the requirements of Assumption 1. According to Definition 3.3, for all $x \in \mathcal{X}_j^{near}(r)$, we have $|\eta_j| \leq 1 + \tilde{\varepsilon}_2 \mathfrak{d}_{\mathfrak{g}}(x, x_j^0)^2$. We deduce that

$$\begin{aligned}
|\omega_j^0 - \hat{\mu}_{n,\omega}(\mathcal{X}_j^{near}(r_e))| &= \left| \omega_j^0 - \int \eta_j \, \mathrm{d}\hat{\mu}_{n,\omega} + \int \eta_j \, \mathrm{d}\hat{\mu}_{n,\omega} - \int_{\mathcal{X}_j^{near}(r_e)} \mathrm{d}\hat{\mu}_{n,\omega} \right|, \\
&\leq \left| \int \eta_j \, \mathrm{d}(\mu_\omega^0 - \hat{\mu}_{n,\omega}) \right| + \int_{\mathcal{X}^{near}(r) \setminus \mathcal{X}_j^{near}(r)} |\eta_j| \, \mathrm{d}\hat{\mu}_{n,\omega} + \int_{\mathcal{X}_j^{near}(r) \setminus \mathcal{X}_j^{near}(r_e)} |\eta_j| \, \mathrm{d}\hat{\mu}_{n,\omega} \\
&\quad + \int_{\mathcal{X}_j^{near}(r_e)} |1 - \eta_j| \, \mathrm{d}\hat{\mu}_{n,\omega} + \int_{\mathcal{X}^{far}(r)} |\eta_j| \mathrm{d}\hat{\mu}_{n,\omega}, \\
&\leq \|p_j\|_{\mathbb{L}} \left(2 \|\Gamma_n\|_{\mathbb{L}} + 2\kappa \|p\|_{\mathbb{L}}\right) + \tilde{\varepsilon}_2 \sum_{l \neq j} \int_{\mathcal{X}_l^{near}(r)} \mathfrak{d}_{\mathfrak{g}}(x, x_l^0)^2 \, \mathrm{d}\hat{\mu}_{n,\omega} + (1 - \tilde{\varepsilon}_0)\hat{\mu}_{n,\omega}(\mathcal{X}^{far}(r)) \\
&\quad + \tilde{\varepsilon}_2 \int_{\mathcal{X}_j^{near}(r_e)} \mathfrak{d}_{\mathfrak{g}}(x, x_j^0)^2 \, \mathrm{d}\hat{\mu}_{n,\omega} + \int_{\mathcal{X}_j^{near}(r) \setminus \mathcal{X}_j^{near}(r_e)} (1 + \tilde{\varepsilon}_2 \mathfrak{d}_{\mathfrak{g}}(x, x_j^0)^2) \, \mathrm{d}\hat{\mu}_{n,\omega}.
\end{aligned}$$

From (15) we get $1 \leq \frac{\tilde{\varepsilon}_3}{r_e^2} \mathfrak{d}_{\mathfrak{g}}(x_j^0, x)^2$ for all $x \in \mathcal{X}_j^{near}(r) \setminus \mathcal{X}_j^{near}(r_e)$, so $1 + \tilde{\varepsilon}_2 \mathfrak{d}_{\mathfrak{g}}(x_j^0, x)^2 \leq \left(\frac{\tilde{\varepsilon}_3}{r_e^2} + \tilde{\varepsilon}_2 \right) \mathfrak{d}_{\mathfrak{g}}(x_j^0, x)^2$. Using again (31), we deduce that

$$|\omega_j^0 - \hat{\mu}_{n,\omega}(\mathcal{X}_j^{near}(r_e))| \leq \|p_j\|_{\mathbb{L}} \left(2 \|\Gamma_n\|_{\mathbb{L}} + 2\kappa \|p\|_{\mathbb{L}}\right) + \max \left\{ \frac{1 - \tilde{\varepsilon}_0}{\varepsilon_0}, \frac{1}{\varepsilon_2} \left(\frac{\tilde{\varepsilon}_3}{r_e^2} + \tilde{\varepsilon}_2 \right) \right\} D_\eta(\hat{\mu}_{n,\omega}, \mu_\omega^0). \quad (37)$$

We can conclude the proof using the controls on $\mathbb{E}\left[\|\Gamma_n\|_{\mathbb{L}}\right]$, $\|p_j\|_{\mathbb{L}}$, $\|p\|_{\mathbb{L}}$, $\mathbb{E}\left[D_\eta(\hat{\mu}_{n,\omega}, \mu_\omega^0)\right]$ stemming from our assumptions along with Lemma 3.1 and (30). Choosing $\kappa = \frac{\rho_n}{\sqrt{c_p}}$,

$$|\omega_j^0 - \hat{\mu}_{n,\omega}(\mathcal{X}_j^{near}(r_e))| \leq 2\sqrt{c_p}\rho_n(1 + \sqrt{s}) + \max \left\{ \frac{1 - \tilde{\varepsilon}_0}{\varepsilon_0}, \frac{1}{\varepsilon_2} \left(\frac{\tilde{\varepsilon}_3}{r_e^2} + \tilde{\varepsilon}_2 \right) \right\} \frac{\sqrt{c_p}}{2} \rho_n (1 + \sqrt{s})^2.$$

With an s-dependent choice of regularization: Choosing $\kappa = \frac{\rho_n}{\sqrt{c_p s}}$, using (37) and the control of $\mathbb{E}\left[D_\eta(\hat{\mu}_{n,\omega}, \mu_\omega^0)\right]$ stemming from (29) we get

$$\begin{aligned}
\mathbb{E}\left[|\omega_j^0 - \hat{\mu}_{n,\omega}(\mathcal{X}_j^{near}(r_e))|\right] &\leq 4\sqrt{c_p}\rho_n + \max \left\{ \frac{1 - \tilde{\varepsilon}_0}{\varepsilon_0}, \frac{1}{\varepsilon_2} \left(\frac{\tilde{\varepsilon}_3}{r_e^2} + \tilde{\varepsilon}_2 \right) \right\} 2\rho_n \sqrt{c_p s}, \\
&\lesssim \frac{\sqrt{s}}{\tau^{d/2} \sqrt{n} r_e^2}
\end{aligned} \quad (38)$$

keeping only the dependence on s, τ, r_e, n.

F.2 Proof of Corollary 3.1

Let $0 < r_e \leq r$. We first prove that

$$\left| \frac{\hat{\mu}_{n,\omega}}{W}(\mathcal{X}_j^{near}(r_e)) - a_j^0 \right| \leq (1 + H(r)r_e) W(x_j^0)^{-1} \left| \omega_j^0 - \hat{\mu}_{n,\omega}(\mathcal{X}_j^{near}(r_e)) \right| + a_j^0 H(r) r_e \quad (39)$$

where

$$H(r) = \frac{(e^{r^2} + \sqrt{e^{2r^2} - 1})^{d/2} - 1}{r}. \quad (40)$$

We use the triangle inequality

$$\left|a_j^0 - \frac{\hat\mu_{n,\omega}}{W}(\mathcal{X}_j^{near}(r_e))\right| \le \underbrace{\left|a_j^0 - \frac{\hat\mu_{n,\omega}(\mathcal{X}_j^{near}(r_e))}{W(x_j^0)}\right|}_{=:A} + \underbrace{\left|\frac{\hat\mu_{n,\omega}(\mathcal{X}_j^{near}(r_e))}{W(x_j^0)} - \frac{\hat\mu_{n,\omega}}{W}(\mathcal{X}_j^{near}(r_e))\right|}_{=:B}.$$

Control of B: We recall the definitions of effective near regions (14) and of W (see (6)). We have

$$\frac{\hat\mu_{n,\omega}}{W}(\mathcal{X}_j^{near}(r_e)) = \int_{\mathcal{X}_j^{near}(r_e)} \frac{W(x_j^0)}{W(x)}\,\mathrm{d}\frac{\hat\mu_{n,\omega}}{W(x_j^0)}(x)$$

so

$$\left|\frac{\hat\mu_{n,\omega}}{W}(\mathcal{X}_j^{near}(r_e)) - \frac{\hat\mu_{n,\omega}(\mathcal{X}_j^{near}(r_e))}{W(x_j^0)}\right| \le \max\left\{\sup_{x\in\mathcal{X}_j^{near}(r_e)} \frac{W(x_j^0)}{W(x)} - 1\,,\, 1 - \inf_{x\in\mathcal{X}_j^{near}(r_e)} \frac{W(x_j^0)}{W(x)}\right\} \times \frac{\hat\mu_{n,\omega}(\mathcal{X}_j^{near}(r_e))}{W(x_j^0)}.$$

For $x \in \mathcal{X}_j^{near}(r_e)$, as $\prod_{k=1}^d \frac{(u_{j,k}^0)^2+u_k^2+\tau^2}{\sqrt{2(u_{j,k}^0)^2+\tau^2}\sqrt{2u_k^2+\tau^2}} \le e^{r_e^2}$ and as each term of this product is greater than 1, we have $\frac{(u_{j,k}^0)^2+u_k^2+\tau^2}{\sqrt{2(u_{j,k}^0)^2+\tau^2}\sqrt{2u_k^2+\tau^2}} \le e^{r_e^2}$ for all $k = 1,\dots,d$. Using (34), this implies

$$\sqrt{u_k^2 + \frac{\tau^2}{2}} \in \left[\sqrt{(u_{j,k}^0)^2 + \frac{\tau^2}{2}(e^{r_e^2} - \sqrt{e^{2r_e^2}-1})}, \sqrt{(u_{j,k}^0)^2 + \frac{\tau^2}{2}(e^{r_e^2} + \sqrt{e^{2r_e^2}-1})}\right] \ \forall\, k = 1,\dots,d,$$

from which we deduce that

$$\begin{aligned}\max\left\{\sup_{x\in\mathcal{X}_j^{near}(r_e)} \frac{W(x_j^0)}{W(x)} - 1\,,\, 1 - \inf_{x\in\mathcal{X}_j^{near}(r_e)} \frac{W(x_j^0)}{W(x)}\right\} &\le \max\left\{(e^{r_e^2} + \sqrt{e^{2r_e^2}-1})^{d/2} - 1\,,\, 1 - (e^{r_e^2} - \sqrt{e^{2r_e^2}-1})^{d/2}\right\},\\ &= (e^{r_e^2} + \sqrt{e^{2r_e^2}-1})^{d/2} - 1\end{aligned}$$

where we used that $(e^{r_e^2} + \sqrt{e^{2r_e^2}-1})^{d/2} \ge 1$ to establish that

$$1 - (e^{r_e^2} - \sqrt{e^{2r_e^2}-1})^{d/2} = 1 - (e^{r_e^2} + \sqrt{e^{2r_e^2}-1})^{-d/2} \le (e^{r_e^2} + \sqrt{e^{2r_e^2}-1})^{d/2} - 1.$$

Hence

$$\left|\frac{\hat\mu_{n,\omega}}{W}(\mathcal{X}_j^{near}(r_e)) - \frac{\hat\mu_{n,\omega}(\mathcal{X}_j^{near}(r_e))}{W(x_j^0)}\right| \le \left((e^{r_e^2} + \sqrt{e^{2r_e^2}-1})^{d/2} - 1\right)\frac{\hat\mu_{n,\omega}(\mathcal{X}_j^{near}(r_e))}{W(x_j^0)}. \tag{41}$$

Control of A and proof of (39): As

$$\left|a_j^0 - \frac{\hat\mu_{n,\omega}(\mathcal{X}_j^{near}(r_e))}{W(x_j^0)}\right| = W(x_j^0)^{-1}\left|\omega_j^0 - \hat\mu_{n,\omega}(\mathcal{X}_j^{near}(r_e))\right|,$$

from (41) we get

$$\left|\frac{\hat\mu_{n,\omega}}{W}(\mathcal{X}_j^{near}(r_e)) - a_j^0\right| \le (e^{r_e^2}+\sqrt{e^{2r_e^2}-1})^{d/2}W(x_j^0)^{-1}|\omega_j^0 - \hat\mu_{n,\omega}(\mathcal{X}_j^{near}(r_e))| + a_j^0\left((e^{r_e^2}+\sqrt{e^{2r_e^2}-1})^{d/2} - 1\right).$$

We conclude the proof of (39) by noticing that $h : r_e \in \mathbb{R}^+ \mapsto (e^{r_e^2} + \sqrt{e^{2r_e^2}-1})^{d/2}$ is convex (for all $d \in \mathbb{N}^*$), hence for $r_e \le r$ we have

$$(e^{r_e^2} + \sqrt{e^{2r_e^2}-1})^{d/2} \le h(0) + \frac{h(r)-h(0)}{r}r_e = 1 + \frac{h(r)-1}{r}r_e.$$

Conclusion: Taking $r_e = n^{-\alpha}$ with $\alpha > 0$, (39) gives

$$\mathbb{E}\left[\left|a_j^0 - \frac{\hat\mu_{n,\omega}}{W}(\mathcal{X}_j^{near}(n^{-\alpha}))\right|\right] \le W(x_j^0)^{-1}\mathbb{E}\left[\left|\omega_j^0 - \hat\mu_{n,\omega}(\mathcal{X}_j^{near}(n^{-\alpha}))\right|\right]\left(1 + H(r)n^{-\alpha}\right) + a_j^0 H(r)n^{-\alpha}$$

where $H(r)$ is defined by (40). Proposition 3.1 gives $\mathbb{E}\left[\left|\omega_j^0 - \hat\mu_{n,\omega}(\mathcal{X}_j^{near}(n^{-\alpha}))\right|\right] \lesssim \frac{s}{\tau^{d/2}\sqrt{n r_e^2}} = \frac{s}{\tau^{d/2}n^{1/2-2\alpha}}$. We choose $r_e = n^{-1/6}$ to balance the terms.

With $\kappa = \frac{\rho_n}{\sqrt{c_p s}}$: Choosing $\kappa = \frac{\rho_n}{\sqrt{c_p s}}$, using (38) we get for $n^{-1/6} \le r$

$$\mathbb{E}\left[\left|a_j^0 - \frac{\hat\mu_{n,\omega}}{W}(\mathcal{X}_j^{near}(n^{-1/6}))\right|\right] \lesssim \left(W(x_j^0)^{-1}\sqrt{s}\tau^{-d/2} + a_j^0\right)n^{-1/6}. \tag{42}$$

G Proofs related to guarantees on the prediction

We will use the following lemma to go from controls of $\left\|L \circ \Phi(\hat{\mu}_n - \mu^0)\right\|_{\mathbb{L}}^2$ to controls on $\left\|\Phi(\hat{\mu}_n - \mu^0)\right\|_{L^2(\mathbb{R}^d)}^2$.

Lemma G.1 (Control of the high frequencies)**.** *Assume that $\mathcal{X} \subset \mathbb{R}^d \times [u_{\min}, +\infty)^d$. Let $\tau > 0$. We work with $\Lambda(\xi) = e^{-\frac{1}{2}\tau^2\|\xi\|_2^2}$. Let $\mu_1, \mu_2 \in \mathcal{M}(\mathcal{X})$. Then*

$$\|\Phi(\mu_1 - \mu_2)\|_{L^2}^2 \leq e\,\|L \circ \Phi(\mu_1 - \mu_2)\|_{\mathbb{L}}^2 + \frac{2}{(2\pi)^d}(\|\mu_1\|_{\mathrm{TV}}^2 + \|\mu_2\|_{\mathrm{TV}}^2)\frac{\tau^d d^{d/2}}{2^{d/2}u_{\min}^{2d}} e^{-2\frac{u_{\min}^2}{\tau^2}} .$$

Proof. Let $\mathrm{T} > 0$. First write

$$\|\Phi(\mu_1 - \mu_2)\|_{L^2}^2 = \frac{1}{(2\pi)^d}\int_{[-\frac{1}{\mathrm{T}},\frac{1}{\mathrm{T}}]^d} |\mathcal{F}[\Phi(\mu_1 - \mu_2)]|^2 + \frac{1}{(2\pi)^d}\int_{\mathbb{R}^d\setminus[-\frac{1}{\mathrm{T}},\frac{1}{\mathrm{T}}]^d} |\mathcal{F}[\Phi(\mu_1 - \mu_2)]|^2 .$$

Then, remark that

$$\begin{aligned}
\frac{1}{(2\pi)^d}\int_{[-\frac{1}{\mathrm{T}},\frac{1}{\mathrm{T}}]^d} |\mathcal{F}[\Phi(\mu_1 - \mu_2)]|^2 &= \frac{1}{(2\pi)^d}\int_{[-\frac{1}{\mathrm{T}},\frac{1}{\mathrm{T}}]^d} \frac{\Lambda}{\Lambda}|\mathcal{F}[\Phi(\mu_1 - \mu_2)]|^2 ,\\
&\leq e^{\frac{d\tau^2}{2\mathrm{T}^2}}\frac{1}{(2\pi)^d}\int_{[-\frac{1}{\mathrm{T}},\frac{1}{\mathrm{T}}]^d} \Lambda|\mathcal{F}[\Phi(\mu_1 - \mu_2)]|^2 ,\\
&\leq e^{\frac{d\tau^2}{2\mathrm{T}^2}}\,\|L \circ \Phi(\mu_1 - \mu_2)\|_{\mathbb{L}}^2 .
\end{aligned}$$

Concerning the high frequencies of $\Phi(\mu_1 - \mu_2)$, recall that $u_1, \ldots, u_d \geq u_{\min}$ for all $((t_1, \ldots, t_d), (u_1, \ldots, u_d)) \in \mathcal{X}$. Hence

$$\begin{aligned}
\frac{1}{(2\pi)^d}\int_{\mathbb{R}^d\setminus[-\frac{1}{\mathrm{T}},\frac{1}{\mathrm{T}}]^d} |\mathcal{F}[\Phi(\mu_1 - \mu_2)]|^2 &\leq \frac{2}{(2\pi)^d}(\|\mu_1\|_{\mathrm{TV}}^2 + \|\mu_2\|_{\mathrm{TV}}^2)\int_{\mathbb{R}^d\setminus[-\frac{1}{\mathrm{T}},\frac{1}{\mathrm{T}}]^d} e^{-u_{\min}^2\|\xi\|_2^2}\,\mathrm{d}\xi ,\\
&\leq \frac{2}{(2\pi)^d}(\|\mu_1\|_{\mathrm{TV}}^2 + \|\mu_2\|_{\mathrm{TV}}^2)\left(\frac{\mathrm{T}}{u_{\min}^2}e^{-\frac{u_{\min}^2}{\mathrm{T}^2}}\right)^d
\end{aligned}$$

using

$$\int_{\mathbb{R}\setminus[-\frac{1}{\mathrm{T}},\frac{1}{\mathrm{T}}]} e^{-u_{\min}^2 z^2}\,\mathrm{d}z = 2\frac{1}{\mathrm{T}}\int_{[1,+\infty)} e^{-u_{\min}^2\frac{z^2}{\mathrm{T}^2}}\,\mathrm{d}z \leq 2\frac{1}{\mathrm{T}}\int_{[1,+\infty)} z e^{-u_{\min}^2\frac{z^2}{\mathrm{T}^2}}\,\mathrm{d}z = \frac{\mathrm{T}}{u_{\min}^2}e^{-\frac{u_{\min}^2}{\mathrm{T}^2}} . \tag{43}$$

So

$$\|\Phi(\mu_1 - \mu_2)\|_{L^2(\mathbb{R}^d)}^2 \leq e^{\frac{d\tau^2}{2\mathrm{T}^2}}\,\|L \circ \Phi(\mu_1 - \mu_2)\|_{\mathbb{L}}^2 + \frac{2}{(2\pi)^d}(\|\mu_1\|_{\mathrm{TV}}^2 + \|\mu_2\|_{\mathrm{TV}}^2)\left(\frac{\mathrm{T}}{u_{\min}^2}e^{-\frac{u_{\min}^2}{\mathrm{T}^2}}\right)^d .$$

Taking $\mathrm{T} = \frac{\tau\sqrt{d}}{\sqrt{2}}$, we get

$$\|\Phi(\mu_1 - \mu_2)\|_{L^2}^2 \leq e\,\|L \circ \Phi(\mu_1 - \mu_2)\|_{\mathbb{L}}^2 + \frac{2}{(2\pi)^d}(\|\mu_1\|_{\mathrm{TV}}^2 + \|\mu_2\|_{\mathrm{TV}}^2)\frac{\tau^d d^{d/2}}{2^{d/2}u_{\min}^{2d}} e^{-2\frac{u_{\min}^2}{\tau^2}} .$$

□

G.1 Proof of Proposition 4.1

We do not make any assumption on the existence of dual certificates in this proof. From $J_W(\hat{\mu}_{n,\omega}) \leq J_W(\mu_\omega^0)$, we have

$$\begin{aligned}
\left\|L \circ \Phi(\hat{\mu}_n - \mu^0)\right\|_{\mathbb{L}}^2 &\leq 2\left\|L \circ \hat{f}_n - L \circ \Phi\hat{\mu}_n\right\|_{\mathbb{L}}^2 + 2\left\|L \circ \hat{f}_n - L \circ \Phi\mu^0\right\|_{\mathbb{L}}^2 ,\\
&\leq 4\,\|\Gamma_n\|_{\mathbb{L}}^2 + 4\kappa\left\|\mu_\omega^0\right\|_{\mathrm{TV}} .
\end{aligned}$$

Combining this inequality with Lemma G.1, we get

$$\left\|\Phi(\hat{\mu}_n - \mu^0)\right\|_{L^2}^2 \leq e\left(4\,\|\Gamma_n\|_{\mathbb{L}}^2 + 4\kappa\left\|\mu_\omega^0\right\|_{\mathrm{TV}}\right) + \frac{2}{(2\pi)^d}(\|\mu_1\|_{\mathrm{TV}}^2 + \|\mu_2\|_{\mathrm{TV}}^2)\frac{\tau^d d^{d/2}}{2^{d/2}u_{\min}^{2d}} e^{-2\frac{u_{\min}^2}{\tau^2}} .$$

Remark that for $\mu \in \mathcal{M}(\mathcal{X})$ we have $\left\|\frac{\mu}{W}\right\|_{\mathrm{TV}} \leq \|\mu\|_{\mathrm{TV}} \sup_{\mathcal{X}} \frac{1}{W} \leq (2\pi)^{d/4}(2u_{\max}^2+\tau^2)^{d/4}\|\mu\|_{\mathrm{TV}}$, and in the same way $\|\mu\|_{\mathrm{TV}} \leq (2\pi)^{-d/4}(2u_{\min}^2+\tau^2)^{-d/4}\left\|\frac{\mu}{W}\right\|_{\mathrm{TV}}$. As $J_W(\hat{\mu}_{n,\omega}) \leq J_W(\mu_\omega^0)$ implies that $\|\hat{\mu}_{n,\omega}\|_{\mathrm{TV}} \leq \frac{1}{2\kappa}\|\Gamma_n\|_{\mathbb{L}}^2 + \|\mu_\omega^0\|_{\mathrm{TV}}$, we deduce that

$$\begin{aligned}
\frac{2}{(2\pi)^d}(\|\hat{\mu}_n\|_{\mathrm{TV}}^2 + \|\mu^0\|_{\mathrm{TV}}^2) &\leq \frac{2}{(2\pi)^d}\left((2\pi)^{d/2}(2u_{\max}^2+\tau^2)^{d/2}\left(\frac{1}{2\kappa^2}\|\Gamma_n\|_{\mathbb{L}}^4 + 2\|\mu_\omega^0\|_{\mathrm{TV}}^2\right) + \|\mu^0\|_{\mathrm{TV}}^2\right), \\
&\leq \frac{(2\pi)^{-d/2}(2u_{\max}^2+\tau^2)^{d/2}}{\kappa^2}\|\Gamma_n\|_{\mathbb{L}}^4 + \frac{2}{(2\pi)^d}\left(2\left(\frac{2u_{\max}^2+\tau^2}{2u_{\min}^2+\tau^2}\right)^{d/2}+1\right)\|\mu^0\|_{\mathrm{TV}}^2, \\
&\leq \frac{(2\pi)^{-d/2}(2u_{\max}^2+\tau^2)^{d/2}}{\kappa^2}\|\Gamma_n\|_{\mathbb{L}}^4 + \frac{2}{(2\pi)^d}\left(2\left(\frac{u_{\max}}{u_{\min}}\right)^{d}+1\right)\|\mu^0\|_{\mathrm{TV}}^2 \qquad (44)
\end{aligned}$$

using that $\tau \in \mathbb{R}^+ \mapsto \frac{2u_{\max}^2+\tau^2}{2u_{\min}^2+\tau^2}$ is decreasing. Hence

$$\begin{aligned}
\|\Phi(\hat{\mu}_n - \mu^0)\|_{L^2}^2 &\leq e\left(4\|\Gamma_n\|_{\mathbb{L}}^2 + 4\kappa(2\pi)^{-d/4}(2u_{\min}^2+\tau^2)^{-d/4}\|\mu^0\|_{\mathrm{TV}}\right) \\
&\quad + \left(\frac{(2\pi)^{-d/2}(2u_{\max}^2+\tau^2)^{d/2}}{\kappa^2}\|\Gamma_n\|_{\mathbb{L}}^4 + \frac{2}{(2\pi)^d}\left(2\left(\frac{u_{\max}}{u_{\min}}\right)^{d}+1\right)\|\mu^0\|_{\mathrm{TV}}^2\right)\frac{\tau^d d^{d/2}}{2^{d/2}u_{\min}^{2d}}e^{-2\frac{u_{\min}^2}{\tau^2}}.
\end{aligned}$$

With Lemma 3.1, choosing $\tau = \frac{\sqrt{2}u_{\min}}{\sqrt{\ln n}}$ and $\kappa = \rho_n^2$ it comes

$$\begin{aligned}
\mathbb{E}\left[\|\Phi(\hat{\mu}_n - \mu^0)\|_{L^2}^2\right] &\leq 4e\rho_n^2\left(1 + (2\pi)^{-d/4}(\sqrt{2}u_{\min})^{-d/2}\left(1+\frac{1}{\ln n}\right)^{-d/4}\|\mu^0\|_{\mathrm{TV}}\right) \\
&\quad + \left(\tilde{C}_\Gamma \pi^{-d/2}\left(u_{\max}^2 + \frac{u_{\min}^2}{\ln n}\right)^{d/2} + 2(2\pi)^{-d}\|\mu^0\|_{\mathrm{TV}}^2\left(2\left(\frac{u_{\max}}{u_{\min}}\right)^{d}+1\right)\right)\frac{d^{d/2}}{(\ln n)^{d/2}u_{\min}^d n}, \\
&\lesssim \frac{(\ln n)^{d/2}}{n}.
\end{aligned}$$

G.2 Prediction with Kernel Density Estimation

Lemma G.2. *With $\mathcal{X} \subset \mathbb{R}^d \times [u_{\min}, +\infty)^d$, setting $\tau = \frac{1}{\sqrt{\ln n}\, n^{\frac{1}{4+d}}}$, omitting the dependence on d we have*

$$\mathbb{E}\left[\left\|L \circ \hat{f}_n - \Phi\mu^0\right\|_{L^2}^2\right] \lesssim \frac{(\ln n)^{d/2}}{n^{\frac{4}{d+4}} u_{\min}^{d+4}}.$$

Proof. Control of $\|L \circ \Phi\mu^0 - \Phi\mu^0\|_{L^2}^2$: As $|\mathcal{F}[\Phi\mu^0](\xi)|^2 \leq \|\mu^0\|_{\mathrm{TV}}^2 e^{-u_{\min}^2\|\xi\|_2^2}$ for $\xi \in \mathbb{R}^d$ and using (43), it comes that for $\mathrm{T} > 0$,

$$\begin{aligned}
\|L \circ \Phi\mu^0 - \Phi\mu^0\|_{L^2}^2 &= \frac{1}{(2\pi)^d}\int_{\mathbb{R}^d} |\Lambda(\xi) - 1|^2 |\mathcal{F}[\Phi\mu^0](\xi)|^2 \,\mathrm{d}\xi, \\
&\leq \frac{1}{(2\pi)^d}\int_{\mathbb{R}^d \setminus [-\frac{1}{\mathrm{T}}, \frac{1}{\mathrm{T}}]^d} |\mathcal{F}[\Phi\mu^0](\xi)|^2 \,\mathrm{d}\xi + \frac{1}{(2\pi)^d}\frac{d^2\tau^4}{4\mathrm{T}^4}\int_{[-\frac{1}{\mathrm{T}}, \frac{1}{\mathrm{T}}]^d} |\mathcal{F}[\Phi\mu^0](\xi)|^2 \,\mathrm{d}\xi, \\
&\leq \frac{\|\mu^0\|_{\mathrm{TV}}^2}{(2\pi)^d}\left(\frac{\mathrm{T}^d}{u_{\min}^{2d}}e^{-\frac{du_{\min}^2}{\mathrm{T}^2}} + \frac{2^d}{\mathrm{T}^d}\frac{\tau^4 d^2}{4\mathrm{T}^4}\right),
\end{aligned}$$

where we used that for $\xi \in \mathbb{R}^d$, $|\Lambda(\xi) - 1|^2 = |e^{-\frac{\tau^2}{2}\|\xi\|_2^2} - 1| \leq 1$ and

$$|\Lambda(\xi) - 1|^2 \leq |e^{-\frac{d\tau^2}{2\mathrm{T}^2}} - 1| \leq \frac{d^2\tau^4}{4\mathrm{T}^4} \quad \forall \xi \in \left[-\frac{1}{\mathrm{T}}, \frac{1}{\mathrm{T}}\right]^d.$$

Control of $\mathbb{E}\left[\left\|L \circ \hat{f}_n - \Phi\mu^0\right\|_{L^2}^2\right]$: Using Lemma 3.1, we have

$$\mathbb{E}\left[\left\|L \circ \hat{f}_n - \Phi\mu^0\right\|_{L^2}^2\right] \leq 2\rho_n^2 + 2\|L \circ \Phi\mu^0 - \Phi\mu^0\|_{L^2}^2 \leq \frac{8}{(2\pi)^{d/2}\tau^d n} + \frac{2\|\mu^0\|_{\mathrm{TV}}^2}{(2\pi)^d}\left(\frac{\mathrm{T}^d}{u_{\min}^{2d}}e^{-\frac{du_{\min}^2}{\mathrm{T}^2}} + \frac{2^d}{\mathrm{T}^d}\frac{\tau^4 d^2}{4\mathrm{T}^4}\right).$$

To balance these terms, we choose $\mathrm{T} = \frac{u_{\min}\sqrt{d}}{\sqrt{\ln n}}$ and $\tau = \frac{1}{\sqrt{\ln n} n^{\frac{1}{4+d}}}$. It comes

$$\mathbb{E}\left[\left\|L \circ \hat{f}_n - \Phi\mu^0\right\|_{L^2}^2\right] \leq \frac{8(\ln n)^{d/2}}{(2\pi)^{d/2} n^{\frac{4}{d+4}}} + \frac{2\left\|\mu^0\right\|_{\mathrm{TV}}^2}{(2\pi)^d}\left(\frac{d^{d/2}}{u_{\min}^d (\ln n)^{d/2} n} + \frac{2^d (\ln n)^{d/2}}{4d^{d/2} u_{\min}^{d+4} n^{\frac{4}{d+4}}}\right).$$

□

G.3 Proof of Theorem 4.1

Equation (28) gives

$$\left\|L \circ \Phi(\hat{\mu}_n - \mu^0)\right\|_{\mathbb{L}} \leq 2\left\|\Gamma_n\right\|_{\mathbb{L}} + 2\kappa\left\|p\right\|_{\mathbb{L}}$$

from which we deduce, using Lemma 3.1 and $\|p\|_{\mathbb{L}} \leq \sqrt{c_p s}$, that

$$\mathbb{E}\left[\left\|L \circ \Phi(\hat{\mu}_n - \mu^0)\right\|_{\mathbb{L}}^2\right] \leq 4(\rho_n + \kappa\sqrt{c_p s})^2 .$$

To go from a control of $\mathbb{E}\left[\left\|L \circ \Phi(\hat{\mu}_n - \mu^0)\right\|_{\mathbb{L}}^2\right]$ to a bound on $\mathbb{E}\left[\left\|\Phi\hat{\mu}_n - \Phi\mu^0\right\|_{L^2(\mathbb{R}^d)}^2\right]$, we use (44) with Lemmas G.1 and 3.1. We get

$$\begin{aligned}
\mathbb{E}\left[\left\|\Phi\hat{\mu}_n - \Phi\mu^0\right\|_{L^2(\mathbb{R}^d)}^2\right] &\leq e\mathbb{E}\left[\left\|L \circ \Phi(\hat{\mu}_n - \mu^0)\right\|_{\mathbb{L}}^2\right] + \frac{2}{(2\pi)^d}\left(\mathbb{E}\left[\left\|\hat{\mu}_n\right\|_{\mathrm{TV}}^2\right] + \left\|\mu^0\right\|_{\mathrm{TV}}^2\right)\frac{\tau^d d^{d/2}}{2^{d/2} u_{\min}^{2d}} e^{-2\frac{u_{\min}^2}{\tau^2}}, \\
&\leq 4e(\rho_n + \kappa\sqrt{c_p s})^2 \\
&\quad + \left(\frac{(2\pi)^{d/2}(2u_{\max}^2 + \tau^2)^{d/2}\tilde{C}_\Gamma \rho_n^4}{\kappa^2} + 2\left(2\left(\frac{u_{\max}}{u_{\min}}\right)^d + 1\right)\left\|\mu^0\right\|_{\mathrm{TV}}^2\right)\frac{(2\pi)^{-d}\tau^d d^{d/2}}{2^{d/2} u_{\min}^{2d}} e^{-2\frac{u_{\min}^2}{\tau^2}}, \\
&\leq 4e(\rho_n + \kappa\sqrt{c_p s})^2 \\
&\quad + \left(4\left(\frac{u_{\max}^2}{u_{\min}^2} + \frac{1}{\ln n}\right)^{d/2} \tilde{C}_\Gamma \frac{(\ln n)^{d/2}\rho_n^2}{n\kappa^2} + 2\left(2\left(\frac{u_{\max}}{u_{\min}}\right)^d + 1\right)\left\|\mu^0\right\|_{\mathrm{TV}}^2\right)\frac{(2\pi)^{-d} d^{d/2}}{u_{\min}^d (\ln n)^{d/2} n}.
\end{aligned}$$

With the choice $\kappa = \frac{\rho_n}{\sqrt{c_p}}$, it comes

$$\begin{aligned}
\mathbb{E}\left[\left\|\Phi\hat{\mu}_n - \Phi\mu^0\right\|_{L^2(\mathbb{R}^d)}^2\right] &\leq 4e\rho_n^2(1 + \sqrt{s})^2 \\
&\quad + \left(4\left(\frac{u_{\max}^2}{u_{\min}^2} + \frac{1}{\ln n}\right)^{d/2} \tilde{C}_\Gamma c_p \frac{(\ln n)^{d/2}}{n} + 2\left\|\mu^0\right\|_{\mathrm{TV}}^2\left(2\left(\frac{u_{\max}}{u_{\min}}\right)^d + 1\right)\right)\frac{(2\pi)^{-d} d^{d/2}}{u_{\min}^d (\ln n)^{d/2} n}.
\end{aligned}$$

This concludes the proof of Theorem 4.1. With the choice $\kappa = \frac{\rho_n}{\sqrt{c_p s}}$, we get

$$\begin{aligned}
\mathbb{E}\left[\left\|\Phi\hat{\mu}_n - \Phi\mu^0\right\|_{L^2(\mathbb{R}^d)}^2\right] &\leq 16e\rho_n^2 + \left(4\left(\frac{u_{\max}^2}{u_{\min}^2} + \frac{1}{\ln n}\right)^{d/2} \tilde{C}_\Gamma c_p \frac{s(\ln n)^{d/2}}{n} + 2\left\|\mu^0\right\|_{\mathrm{TV}}^2\left(2\left(\frac{u_{\max}}{u_{\min}}\right)^d + 1\right)\right)\frac{(2\pi)^{-d} d^{d/2}}{u_{\min}^d (\ln n)^{d/2} n}, \\
&\lesssim \left(\frac{s}{n(\ln n)^{d/2}} + 1\right)\frac{(\ln n)^{d/2}}{n} \qquad (45)
\end{aligned}$$

keeping only the dependence on n and s.

H Properties of the Fisher-Rao metric

We present properties associated with the Fisher-Rao metric $\mathfrak{g}$, defined at point $x \in \mathbb{R}^d \times [u_{\min}, +\infty)^d$ by $\mathfrak{g}_x = \nabla_1\nabla_2 K_{\mathrm{norm}}(x, x)$ (see also (20)). Note that this metric depends on the smoothing parameter $\tau > 0$ through K_{norm}. We recall the definition of the Riemannian norm: for $v \in \mathbb{R}^{2d}$ and $x \in \mathbb{R}^d \times [u_{\min}, +\infty)^d$, we define $\|v\|_x = \sqrt{v^T \mathfrak{g}_x v}$.

H.1 Christoffel symbols

The non-zero Christoffel symbols associated with $\mathfrak{g}$ are

$$\begin{aligned}
{\Gamma^{t_k}}_{u_k t_k} = {\Gamma^{t_k}}_{t_k u_k} &= \frac{-2u_k}{2u_k^2 + \tau^2}, \\
{\Gamma^{u_k}}_{t_k t_k} &= \frac{1}{u_k}, \\
{\Gamma^{u_k}}_{u_k u_k} &= \frac{\tau^2 - 2u_k^2}{u_k(2u_k^2 + \tau^2)}
\end{aligned}$$

with $k = 1, \dots, d$ (see [Giard, 2025, Section I.1]). We define

$$\Gamma^{t_k} = \begin{pmatrix} ({\Gamma^{t_k}}_{t_l t_m})_{1\le l,m\le d} & ({\Gamma^{t_k}}_{t_l u_m})_{1\le l,m\le d} \\ ({\Gamma^{t_k}}_{u_l t_m})_{1\le l,m\le d} & ({\Gamma^{t_k}}_{u_l u_m})_{1\le l,m\le d} \end{pmatrix} \quad \text{and} \quad \Gamma^{u_k} = \begin{pmatrix} ({\Gamma^{u_k}}_{t_l t_m})_{1\le l,m\le d} & ({\Gamma^{u_k}}_{t_l u_m})_{1\le l,m\le d} \\ ({\Gamma^{u_k}}_{u_l t_m})_{1\le l,m\le d} & ({\Gamma^{u_k}}_{u_l u_m})_{1\le l,m\le d} \end{pmatrix}. \tag{46}$$

H.2 Geodesics and geodesic distance

The next lemmas provide the parametrization of the geodesics associated with the Fisher-Rao metric. We denote $\tilde{\gamma}$ a geodesic parametrized by arc length connecting the points $x = \tilde{\gamma}(0), x' = \tilde{\gamma}(l) \in \mathbb{R}^d \times [u_{\min}, +\infty)^d$. The parameter l is the geodesic distance between x and x', denoted by $\mathfrak{d}_{\mathfrak{g}}(x, x')$. We also denote $\gamma : y \in [0,1] \mapsto \tilde{\gamma}(ly)$ (it is the geodesic such that $x = \gamma(0)$, $x' = \gamma(1)$) and $\dot{\gamma}$ its derivative.

We do not use the formula of the geodesic distance $\mathfrak{d}_{\mathfrak{g}}$ in this paper, but we give it in the next lemmas for information.

Lemma H.1 (Geodesics of the Poincaré half-plane model)**.** *The Poincaré half-plane is* $\{x = (t,u) \in \mathbb{R} \times \mathbb{R}_+^*\}$, *on which we consider the metric defined by* $\mathfrak{h}_x = \begin{pmatrix} \frac{1}{u^2} & 0 \\ 0 & \frac{1}{u^2} \end{pmatrix}$ *for all* $x = (t,u) \in \mathbb{R} \times \mathbb{R}_+^*$. *The associated norm is defined by* $\|v\|_x = \sqrt{v^T \mathfrak{h}_x v}$ *for* $v \in \mathbb{R}^2$.

The Poincaré geodesics are circular arcs whose origin is on the axis $\{u = 0\}$ *and straight vertical lines (parallel to* $\{t = 0\}$*).*

A Poincaré geodesic parametrized by arc-length, denoted by $\tilde{h} = (\tilde{h}_t, \tilde{h}_u)$*, is of the form*

$$\tilde{h} : y \in [0,l] \mapsto \left(\frac{\tanh(C_2 + y)}{C_1} + C_3, \frac{1}{\cosh(C_2 + y)|C_1|} \right)$$

(semicircle) or

$$\tilde{h} : y \in [0,l] \mapsto (C_3, |C_1| e^{y}) \quad \text{or} \quad \tilde{h} : y \in [0,l] \mapsto (C_3, |C_1| e^{-y})$$

(straight line), where $C_1 \in \mathbb{R}^*$, $C_2, C_3 \in \mathbb{R}$, $l \in \mathbb{R}^+$.

Moreover, writing $\mathfrak{d}_{\mathfrak{h}}$ *the Poincaré distance,*

$$\mathfrak{d}_{\mathfrak{h}}(x, x') = \ln \left(\frac{\sqrt{(t-t')^2 + (u+u')^2} + \sqrt{(t-t')^2 + (u-u')^2}}{\sqrt{(t-t')^2 + (u+u')^2} - \sqrt{(t-t')^2 + (u-u')^2}} \right) \quad \forall\, x, x' \in \mathbb{R} \times \mathbb{R}_+^*.$$

Proof. The fact that the Poincaré geodesics are semicircles whose origin is on the axis $\{u = 0\}$ and straight lines parallel to $\{t = 0\}$ is well-known (see for instance [Stahl, 1993, Theorem 4.2.1]). The formula for $\mathfrak{d}_{\mathfrak{h}}$ can be found in [Beardon, 1983, Theorem 7.2.1].

We can check that the parametrizations given for the geodesics verify the geodesic equations

$$\begin{cases} \frac{(\dot{\tilde{h}}_t)^2 + (\dot{\tilde{h}}_u)^2}{\tilde{h}_u^2} = 1 \\ \ddot{\tilde{h}}_t - 2\frac{\dot{\tilde{h}}_t \dot{\tilde{h}}_u}{\tilde{h}_u} = 0 \\ \ddot{\tilde{h}}_u - \frac{(\dot{\tilde{h}}_u)^2}{\tilde{h}_u} + \frac{(\dot{\tilde{h}}_t)^2}{\tilde{h}_u} = 0 \end{cases},$$

as done in [Giard, 2025, Section I.2]. We found all the geodesics, because all the portions of Poincaré semicircles and straight lines can be obtained with appropriate choices of C_1, C_2, C_3, l. □

Lemma H.2 (Geodesics for $d = 1$)**.** *Let* $x, x' \in \mathbb{R} \times [u_{\min}, +\infty)$.

- *If* $t = t'$*, the geodesic is of the form*

$$\tilde{\gamma}(y) = \left(c_3, \sqrt{\frac{c_1^2}{2} e^{\sqrt{8}y} - \frac{\tau^2}{2}} \right) \quad \forall\, y \in [0,l] \quad \text{or} \quad \tilde{\gamma}(y) = \left(c_3, \sqrt{\frac{c_1^2}{2} e^{-\sqrt{8}y} - \frac{\tau^2}{2}} \right) \quad \forall\, y \in [0,l] \tag{47}$$

  where $c_1, c_3 \in \mathbb{R}$. *It is a portion of a straight line parallel to* $\{t = 0\}$.

- *If* $t \neq t'$*, the geodesic is of the form*

$$\tilde{\gamma}(y) = \left(c_3 + \frac{\sqrt{2}\tanh\left(\frac{c_2}{2} + \sqrt{2}y\right)}{2c_1}, \sqrt{-\frac{\tau^2}{2} + \frac{1 - \tanh^2\left(\frac{c_2}{2} + \sqrt{2}y\right)}{2c_1^2}} \right) \quad \forall\, y \in [0,l] \tag{48}$$

where $c_1 \neq 0$, $c_2, c_3 \in \mathbb{R}$. It is a portion of a semicircle with center $(c_3, 0)$ and radius

$$\sqrt{\frac{1}{2c_1^2} - \frac{\tau^2}{2}} = \sqrt{\frac{1}{4}\left(\frac{-(t-t')^2 + u^2 - u'^2}{t'-t}\right)^2 + u^2}\,. \tag{49}$$

Moreover, the Fisher-Rao distance between x and x' is

$$\mathfrak{d}_{\mathfrak{g}}(x,x') = \sqrt{2}\ln\left(\frac{\sqrt{(t-t')^2 + \left(\sqrt{u^2+\frac{\tau^2}{2}} - \sqrt{u'^2+\frac{\tau^2}{2}}\right)^2} + \sqrt{(t-t')^2 + \left(\sqrt{u^2+\frac{\tau^2}{2}} + \sqrt{u'^2+\frac{\tau^2}{2}}\right)^2}}{\sqrt{2}(2u^2+\tau^2)^{1/4}(2u'^2+\tau^2)^{1/4}}\right).$$

Proof. Link with the Poincaré half-plane model: Recall that the variational formulation of a (Fisher-Rao) geodesic $\gamma = (\gamma_t, \gamma_u)$ connecting x, x' is

$$\inf_{\gamma(0)=x,\gamma(1)=x'} \int_0^1 \sqrt{\dot\gamma_t(y)^2 \frac{1}{2\gamma_u(y)^2+\tau^2} + \dot\gamma_u(y)^2\frac{2\gamma_u(y)^2}{(2\gamma_u(y)^2+\tau^2)^2}}\,\mathrm{d}y\,.$$

We use the change of variable $h = (h_t, h_u) = \left(\gamma_t, \sqrt{\gamma_u^2 + \frac{\tau^2}{2}}\right)$. Noticing that $(\dot h_u)^2 = \frac{\dot\gamma_u(y)^2\gamma_u(y)^2}{\gamma_u^2+\frac{\tau^2}{2}}$, it comes that the variational formulation is equivalent to the problem

$$\inf\left\{\frac{1}{\sqrt{2}}\int_0^1 \sqrt{(\dot h_t(y))^2\frac{1}{h_u(y)^2} + (\dot h_u(y))^2\frac{1}{h_u(y)^2}}\,\mathrm{d}y \;:\; h(0) = \left(t, \sqrt{u^2+\frac{\tau^2}{2}}\right), h(1) = \left(t', \sqrt{u'^2+\frac{\tau^2}{2}}\right)\right\},$$

and we recognize the Poincaré metric tensor ($\frac{\mathrm{d}t^2+\mathrm{d}u^2}{u^2}$) in this formulation. So h is the geodesic for the Poincaré half-plane metric connecting $\left(t, \sqrt{u^2+\frac{\tau^2}{2}}\right)$ and $\left(t', \sqrt{u'^2+\frac{\tau^2}{2}}\right)$.

Geodesic distance: In particular, using the formula for $\mathfrak{d}_{\mathfrak{h}}$ in Lemma H.1, we have

$$\begin{aligned}
\mathfrak{d}_{\mathfrak{g}}(x,x') &= \frac{1}{\sqrt{2}}\mathfrak{d}_{\mathfrak{h}}\left(\left(t, \sqrt{u^2+\frac{\tau^2}{2}}\right), \left(t', \sqrt{u'^2+\frac{\tau^2}{2}}\right)\right),\\
&= \frac{1}{\sqrt{2}}\ln\left(\frac{\sqrt{(t-t')^2 + \left(\sqrt{u^2+\frac{\tau^2}{2}} + \sqrt{u'^2+\frac{\tau^2}{2}}\right)^2} + \sqrt{(t-t')^2 + \left(\sqrt{u^2+\frac{\tau^2}{2}} - \sqrt{u'^2+\frac{\tau^2}{2}}\right)^2}}{\sqrt{(t-t')^2 + \left(\sqrt{u^2+\frac{\tau^2}{2}} + \sqrt{u'^2+\frac{\tau^2}{2}}\right)^2} - \sqrt{(t-t')^2 + \left(\sqrt{u^2+\frac{\tau^2}{2}} - \sqrt{u'^2+\frac{\tau^2}{2}}\right)^2}}\right),\\
&= \sqrt{2}\ln\left(\frac{\sqrt{(t-t')^2 + \left(\sqrt{u^2+\frac{\tau^2}{2}} - \sqrt{u'^2+\frac{\tau^2}{2}}\right)^2} + \sqrt{(t-t')^2 + \left(\sqrt{u^2+\frac{\tau^2}{2}} + \sqrt{u'^2+\frac{\tau^2}{2}}\right)^2}}{\sqrt{2}(2u^2+\tau^2)^{1/4}(2u'^2+\tau^2)^{1/4}}\right).
\end{aligned}$$

Parametrization by arc-length: We saw that $\left(\gamma_t, \sqrt{\gamma_u^2+\frac{\tau^2}{2}}\right)$ is a geodesic for the Poincaré half-plane model. Lemma H.1 gives its parametrization by arc-length, $\tilde h$. For the $\mathfrak{h}$-norm, $\left\|\dot{\tilde h}_t(y), \dot{\tilde h}_u(y)\right\|_{\tilde h(y)} = 1$. We deduce that for the $\mathfrak{g}$-norm, writing $g(y) = \left(\tilde h_t(y), \sqrt{\tilde h_u(y)^2 - \frac{\tau^2}{2}}\right)$, we have $\|\dot g_t(y), \dot g_u(y)\|_{g(y)} = \frac{1}{\sqrt{2}}$. So defining $\tilde\gamma(y) = g(\sqrt{2}y)$, it comes that $\left\|\dot{\tilde\gamma}_t(y), \dot{\tilde\gamma}_u(y)\right\|_{\gamma(y)} = 1$ for the $\mathfrak{g}$-norm: $\tilde\gamma$ is a geodesic parametrized by arc-length. The geodesics we provide are of this form.

Radius of the semicircle connecting x and x': The parametrization of the semicircle gives

$$(\tilde\gamma_t - c_3)^2 + \tilde\gamma_u^2 = \frac{1}{2c_1^2} - \frac{\tau^2}{2}$$

which is the square of the radius of the semicircle. We also have

$$\frac{1}{2c_1^2} - \frac{\tau^2}{2} = (t-c_3)^2 + u^2 = (t'-c_3)^2 + u'^2\,,$$

from which we deduce that $c_3 = \frac{t'^2+u'^2-(t^2+u^2)}{2(t'-t)}$ along with $\frac{1}{2c_1^2} - \frac{\tau^2}{2} = \frac{1}{4}\left(\frac{-(t-t')^2+u^2-u'^2}{t'-t}\right)^2 + u^2$. ☐

We can extend this result to higher dimensions. By abuse of notation, $\mathfrak{d}_{\mathfrak{g}}(x_k, x'_k)$ will refer to the Fisher-Rao distance in dimension 1 between x_k and x'_k, for all $k \in \{1, \ldots, d\}$. The notation $\mathfrak{g}_{x_k}$ follows the same principle.

Lemma H.3 (Geodesics in dimension $d \geq 1$)**.** *Let $x, x' \in \mathbb{R}^d \times [u_{\min}, +\infty)^d$. For $k = 1, \ldots, d$, we denote $\bar{\gamma}_k$ the geodesic in dimension 1, parametrized by arc length connecting $x_k = (t_k, u_k)$ and $x'_k = (t'_k, u'_k)$. The geodesic connecting x and x' is of the form*

$$\tilde{\gamma} = (\tilde{\gamma}_{t_1}, \ldots, \tilde{\gamma}_{t_d}, \tilde{\gamma}_{u_1}, \ldots, \tilde{\gamma}_{u_d}) \quad \textit{where} \quad (\tilde{\gamma}_{t_k}(y), \tilde{\gamma}_{u_k}(y)) = \bar{\gamma}_k(\sqrt{g_k} y) \quad \textit{with} \quad g_k \geq 0 \quad \textit{and} \quad \sum_{i=1}^{d} g_i = 1\,.$$

Moreover,

$$\mathfrak{d}_{\mathfrak{g}}(x, x') = \sqrt{\sum_{k=1}^{d} \mathfrak{d}_{\mathfrak{g}}(x_k, x'_k)^2}\,.$$

Proof. For all $k \in \{1, \ldots, d\}$, we denote $\mathfrak{d}_k = \mathfrak{d}_{\mathfrak{g}}(x_k, x'_k)^2$ and $\tilde{\gamma}_k(y) = \bar{\gamma}_k\left(\frac{\sqrt{\mathfrak{d}_k}}{\sqrt{\sum_j \mathfrak{d}_j}} y\right)$. Then, for all $v \in \mathbb{R}^{2d}$ and $x \in \mathbb{R}^d \times [u_{\min}, +\infty)^d$, we have

$$\|v\|_x^2 = v^T \mathfrak{g}_x v = \sum_{k=1}^{d} (v_k, v_{k+d}) \mathfrak{g}_{x_k} (v_k, v_{k+d})^T,$$

where (v_k, v_{k+d}) represents the k-th component of v in $\mathbb{R}^d \times [u_{\min}, +\infty)^d$. Using this, the geodesic $\tilde{\gamma}$ connecting x and x' can be expressed as:

$$\tilde{\gamma} = (\tilde{\gamma}_{t_1}, \ldots, \tilde{\gamma}_{t_d}, \tilde{\gamma}_{u_1}, \ldots, \tilde{\gamma}_{u_d}),$$

where each $(\tilde{\gamma}_{t_k}, \tilde{\gamma}_{u_k})$ corresponds to the geodesic $\tilde{\gamma}_k$ in dimension 1 connecting $x_k = (t_k, u_k)$ and $x'_k = (t'_k, u'_k)$. This ensures that $\tilde{\gamma}$ is the geodesic connecting x and x' in $\mathbb{R}^d \times [u_{\min}, +\infty)^d$.

☐

H.3 Compatibility with the semi-distance

H.3.1 Proof of Lemma 5.2

Let $r > 0$ and $x_0 \in \mathbb{R}^d \times [u_{\min}, +\infty)^d$. Let γ be a geodesic between $x_0 = \gamma(0)$ and $x = \gamma(1) \in \mathbb{R}^d \times [u_{\min}, +\infty)^d$ for the metric $\mathfrak{g}$.

Lower bound on the variance We write $\gamma = (\gamma_{t_1}, \ldots, \gamma_{t_d}, \gamma_{u_1}, \ldots, \gamma_{u_d})$. Recall that $(\gamma_{t_k}, \gamma_{u_k})$ is a portion of a straight line parallel to $\{t_k = 0\}$ or of a semicircle with center on $\{u_k = 0\}$. As $x_0, x_1 \in \mathbb{R}^d \times [u_{\min}, +\infty)^d$, we deduce that $u_{\min} \leq \gamma_{u_k}(y)$ for all $k = 1, \ldots, d$, $y \in [0, 1]$.

The function $y \in [0,1] \mapsto \mathfrak{d}(x_0, \gamma(y))$ is non-decreasing We can reduce the problem to the case $d = 1$. Indeed, for each $k \in \{1, \ldots, d\}$, $\gamma_k := (\gamma_{t_k}, \gamma_{u_k})$ is the geodesic in dimension 1 connecting $x_{0,k} = \gamma_k(0)$ and $x_k = \gamma_k(1)$ (see Lemma H.3). Furthermore, $\mathfrak{d}(x_0, \gamma(y))^2 = \sum_{k=1}^d \mathfrak{d}(x_{0,k}, \gamma_k(y))^2$, where, by abuse of notation, $\mathfrak{d}(x_{0,k}, \gamma_k(y))$ denotes the semi-distance in dimension 1 between $x_{0,k}$ and $\gamma_k(y)$. Hence, if the function $y \in [0,1] \mapsto \mathfrak{d}(x_{0,k}, \gamma_k(y))$ is increasing for all k, then $y \in [0,1] \mapsto \mathfrak{d}(x_0, \gamma(y))$ is also increasing.

Until the end of the proof, we therefore concentrate our attention on the case $d = 1$. Let $x_0, x \in \mathbb{R} \times [u_{\min}, +\infty)$, and let γ and $\tilde{\gamma}$ be the geodesic connecting x_0 and x, parametrized by $[0, 1]$ and by arc-length, respectively. Proving that $y \in [0,1] \mapsto \mathfrak{d}(x_0, \gamma(y))$ is increasing is equivalent to proving that $h : y \in [0, l] \mapsto \mathfrak{d}(x_0, \tilde{\gamma}(y))^2$ is increasing, where $l = \mathfrak{d}_{\mathfrak{g}}(x_0, x)$. We will consider alternatively the cases where the geodesic γ is a straight line and a semicircle. It suffices to show that $h' \geq 0$ in both cases.

Proof for straight lines, $d = 1$: See [Giard, 2025, Section II.2]. Using the form of the geodesic given in Lemma H.2, we can show that if $\tilde{\gamma}$ is a straight line,

$$h(y) = \ln(\cosh(\sqrt{2}y))\,.$$

We obtain this formula by deducing from $\tilde{\gamma}(0) = (t_0, u_0)$ in (47) that $c_1^2 = 2u_0^2 + \tau^2$ (the two formulas of (47) give the same result). This function is non-decreasing on $\mathbb{R}^+$.

Proof for semicircles, $d = 1$: See [Giard, 2025, Section II.1]. If $\tilde{\gamma}$ is a semicircle, then

$$h(y) = \ln\left(\frac{\cosh^2\left(\frac{c_2}{2}\right) + \cosh^2\left(\frac{c_2}{2} + \sqrt{2}y\right)}{2\cosh\left(\frac{c_2}{2}\right)\cosh\left(\frac{c_2}{2} + \sqrt{2}y\right)}\right) + \frac{\sinh^2\left(\sqrt{2}y\right)}{\cosh^2\left(\frac{c_2}{2}\right) + \cosh^2\left(\frac{c_2}{2} + \sqrt{2}y\right)}\,.$$

We obtain this formula by deducing from $\tilde{\gamma}(0) = (t_0, u_0)$ in (48) that $c_3 = t_0 - \frac{\sqrt{2}\tanh\left(\frac{c_2}{2}\right)}{2c_1}$ and $c_1 = \pm\frac{1}{\sqrt{\tau^2+2u_0^2\cosh\left(\frac{c_2}{2}\right)}}$. Let $y \geq 0$. We have $h(y) = A(y) + B(y)$ with

$$A(y) = \frac{\sinh^2\left(\sqrt{2}y\right)}{\cosh^2\left(\frac{c_2}{2}\right) + \cosh^2\left(\frac{c_2}{2} + \sqrt{2}y\right)}.$$

and

$$B(y) = \ln\left(\frac{\cosh^2\left(\frac{c_2}{2}\right) + \cosh^2\left(\frac{c_2}{2} + \sqrt{2}y\right)}{2\cosh\left(\frac{c_2}{2}\right)\cosh\left(\frac{c_2}{2} + \sqrt{2}y\right)}\right).$$

Then,

$$A'(y) = \frac{\sqrt{2}\sinh(\sqrt{2}y)\left(\cosh(c_2 + \sqrt{2}y) + \cosh(\sqrt{2}y)(2 + \cosh(c_2))\right)}{\left(\cosh^2\left(\frac{c_2}{2}\right) + \cosh^2\left(\frac{c_2}{2} + \sqrt{2}y\right)\right)^2} \geq 0$$

and

$$B'(y) = \frac{\sqrt{2}\sinh\left(\frac{c_2}{2} + \sqrt{2}y\right)\left(\cosh^2\left(\frac{c_2}{2} + \sqrt{2}y\right) - \cosh^2\left(\frac{c_2}{2}\right)\right)}{\cosh\left(\frac{c_2}{2} + \sqrt{2}y\right)\left(\cosh^2\left(\frac{c_2}{2}\right) + \cosh^2\left(\frac{c_2}{2} + \sqrt{2}y\right)\right)}.$$

If $c_2 \geq 0$, then as $y \geq 0$, $\cosh^2\left(\frac{c_2}{2} + \sqrt{2}y\right) \geq \cosh^2\left(\frac{c_2}{2}\right)$ so $B' \geq 0$, hence $d(x_0, \tilde{\gamma}(y))^2$ is increasing on $\mathbb{R}^+$. We now deal with the case $c_2 < 0$. Remark that $B'(y) > 0$ if and only if $y \notin \left[\frac{-c_2}{2\sqrt{2}}, \frac{-c_2}{\sqrt{2}}\right]$. We want to show that $A' + B' \geq 0$ on the interval $\left[\frac{-c_2}{2\sqrt{2}}, \frac{-c_2}{\sqrt{2}}\right]$. First, using that for $y \in \left[\frac{-c_2}{2\sqrt{2}}, \frac{-c_2}{\sqrt{2}}\right]$, we have

$$\cosh^2\left(\frac{c_2}{2} + \sqrt{2}y\right) \leq \cosh^2\left(\frac{c_2}{2}\right), \quad \sinh(\sqrt{2}y) \geq \sinh\left(\frac{-c_2}{2}\right),$$

$$\cosh(\sqrt{2}y) \geq \cosh\left(\frac{c_2}{2}\right) \quad \text{and} \quad 2 + \cosh(c_2) \geq 2\cosh^2\left(\frac{c_2}{2}\right),$$

we obtain a lower bound for A' on $\left[\frac{-c_2}{2\sqrt{2}}, \frac{-c_2}{\sqrt{2}}\right]$:

$$\begin{aligned}
A'(y) &\geq \frac{\sqrt{2}\sinh\left(\frac{-c_2}{2}\right)\left(\cosh(c_2 + \sqrt{2}y) + \cosh(\sqrt{2}y)(2 + \cosh(c_2))\right)}{2\cosh^2\left(\frac{c_2}{2}\right)\left(\cosh^2\left(\frac{c_2}{2}\right) + \cosh^2\left(\frac{c_2}{2} + \sqrt{2}y\right)\right)}, \\
&\geq \frac{\sqrt{2}\sinh\left(\frac{-c_2}{2}\right)\cosh\left(\frac{c_2}{2}\right)(2 + \cosh(c_2))}{2\cosh^2\left(\frac{c_2}{2}\right)\left(\cosh^2\left(\frac{c_2}{2}\right) + \cosh^2\left(\frac{c_2}{2} + \sqrt{2}y\right)\right)}, \\
&\geq \frac{\sqrt{2}\sinh\left(\frac{-c_2}{2}\right)\cosh\left(\frac{c_2}{2}\right)}{\cosh^2\left(\frac{c_2}{2}\right) + \cosh^2\left(\frac{c_2}{2} + \sqrt{2}y\right)}.
\end{aligned}$$

It follows that

$$A'(y)+B'(y) \geq \frac{\sqrt{2}\cosh\left(\frac{c_2}{2} + \sqrt{2}y\right)\sinh\left(\frac{-c_2}{2}\right)\cosh\left(\frac{c_2}{2}\right) + \sqrt{2}\sinh\left(\frac{c_2}{2} + \sqrt{2}y\right)\left(\cosh^2\left(\frac{c_2}{2} + \sqrt{2}y\right) - \cosh^2\left(\frac{c_2}{2}\right)\right)}{\left(\cosh^2\left(\frac{c_2}{2}\right) + \cosh^2\left(\frac{c_2}{2} + \sqrt{2}y\right)\right)\cosh\left(\frac{c_2}{2} + \sqrt{2}y\right)}.$$

The positivity of $A' + B'$ follows, since

$$\begin{aligned}
&\cosh\left(\frac{c_2}{2} + \sqrt{2}y\right)\sinh\left(\frac{-c_2}{2}\right)\cosh\left(\frac{c_2}{2}\right) + \sinh\left(\frac{c_2}{2} + \sqrt{2}y\right)\left(\cosh^2\left(\frac{c_2}{2} + \sqrt{2}y\right) - \cosh^2\left(\frac{c_2}{2}\right)\right) \\
&= \frac{1}{4}\left(\sinh\left(\frac{3c_2}{2} + 3\sqrt{2}y\right) - 2\sinh\left(\frac{3c_2}{2} + \sqrt{2}y\right) - \sinh\left(\frac{c_2}{2} + \sqrt{2}y\right)\right), \\
&\geq \frac{1}{2}\left(\sinh\left(\frac{c_2}{2} + \sqrt{2}y\right) - \sinh\left(\frac{3c_2}{2} + \sqrt{2}y\right)\right), \\
&\geq 0,
\end{aligned}$$

where we used

$$\begin{aligned}
\sinh\left(\frac{3c_2}{2} + 3\sqrt{2}y\right) &= \sinh\left(\frac{c_2}{2} + \sqrt{2}y\right)\cosh(c_2 + 2\sqrt{2}y) + \cosh\left(\frac{c_2}{2} + \sqrt{2}y\right)\sinh(c_2 + 2\sqrt{2}y), \\
&\geq \sinh\left(\frac{c_2}{2} + \sqrt{2}y\right) + \sinh(c_2 + 2\sqrt{2}y),
\end{aligned}$$

$$\sinh\left(c_2 + 2\sqrt{2}y\right) = 2\sinh\left(\frac{c_2}{2} + \sqrt{2}y\right)\cosh\left(\frac{c_2}{2} + \sqrt{2}y\right) \geq 2\sinh\left(\frac{c_2}{2} + \sqrt{2}y\right)$$

and $c_2 < 0$ together with the fact that sinh is increasing.

We have just proved the following statement: if $d(x_0, x) \leq r$, then

$$\forall\, y \in [0,1], \quad \gamma(y) \in \{x' \in \mathbb{R}^d \times [u_{\min}, +\infty)^d \ :\ d(x_0, x') \leq r\}\,.$$

This concludes the proof of Lemma 5.2.

H.3.2 Proof of Lemma 5.3

Let $r, \Delta > 0$. Assume that $\mathcal{X} \subset \mathbb{R}^d \times [u_{\min}, u_{\max}]^d$. The semi-distance d is defined by (11).
Control of the variance of the geodesics: Let $x_j^0 \in \mathcal{X}$. For $x \in B_d(x_j^0, r) \cap \mathbb{R}^d \times [u_{\min}, u_{\max}]^d$, we have $|t_k - t_{j,k}^0| \leq r\sqrt{2u_{\max}^2 + \tau^2}$ for all $k \in \{1, \dots, d\}$. We deduce that

$$B_d(x_j^0, r) \cap (\mathbb{R}^d \times [u_{\min}, u_{\max}]^d) \subset \left(\bigtimes_{k=1}^{d} [t_{j,k}^0 \pm r\sqrt{2u_{\max}^2 + \tau^2}]\right) \times [u_{\min}, u_{\max}]^d$$

where $\bigtimes_{k=1}^d$ denotes the d-ary Cartesian product.

For any $k \in \{1, \dots, d\}$, the semicircle $\gamma_k = (\gamma_{t_k}, \gamma_{u_k})$ (geodesic in dimension 1) connecting the points $(t_{j,k}^0 - r\sqrt{2u_{\max}^2 + \tau^2}, u_{\max})$ and $(t_{j,k}^0 + r\sqrt{2u_{\max}^2 + \tau^2}, u_{\max})$ satisfies $\gamma_{u_k}(y)^2 \leq u_{\max}^2 + r^2(2u_{\max}^2 + \tau^2)$. This follows because the square of the radius of the semicircle is $u_{\max}^2 + r^2(2u_{\max}^2 + \tau^2)$ (see Lemma H.2). This is the geodesic achieving the largest variance, *i.e.* all the portions of semicircles between points of the rectangle $[t_{j,k}^0 - r\sqrt{2u_{\max}^2 + \tau^2}\,,\, t_{j,k}^0 + r\sqrt{2u_{\max}^2 + \tau^2}] \times [u_{\min}, u_{\max}]$ are below this geodesic. To see why it is the case, one can note that 2 geodesics in our model have at most 1 intersection point, or their union is a geodesic. So a geodesic of $[t_{j,k}^0 - r\sqrt{2u_{\max}^2 + \tau^2}\,,\, t_{j,k}^0 + r\sqrt{2u_{\max}^2 + \tau^2}] \times [u_{\min}, u_{\max}]$ cannot cross two times the semicircle connecting $(t_{j,k}^0 - r\sqrt{2u_{\max}^2 + \tau^2}, u_{\max})$ and $(t_{j,k}^0 + r\sqrt{2u_{\max}^2 + \tau^2}, u_{\max})$.

Using Lemma H.3, we deduce that

$$\mathcal{G}(B_d(x_j^0, r) \cap (\mathbb{R}^d \times [u_{\min}, u_{\max}]^d)) \subset \mathbb{R}^d \times [u_{\min}, \sqrt{u_{\max}^2 + r^2(2u_{\max}^2 + \tau^2)}]^d\,.$$

Pseudo-quasi triangle inequality: Let $x, x', \tilde{x} \in \mathbb{R}^d \times [u_{\min}, \tilde{u}_{\max}]^d$, where $\tilde{u}_{\max}$ is not necessarily equal to $u_{\max}$. We have

$$\frac{u_k^2 + u_k'^2 + \tau^2}{\sqrt{2u_k^2 + \tau^2}\sqrt{2u_k'^2 + \tau^2}} \leq \frac{2\tilde{u}_{\max}^2 + \tau^2}{2u_{\min}^2 + \tau^2} \leq \frac{\tilde{u}_{\max}^2}{u_{\min}^2}$$

using that $\tau \in \mathbb{R}^+ \mapsto \frac{2\tilde{u}_{\max}^2 + \tau^2}{2u_{\min}^2 + \tau^2}$ is decreasing (because $u_{\min} \leq \tilde{u}_{\max}$). Moreover, using the triangle inequality for the ℓ^2-norm in dimension d and that $u_k^2 + \tilde{u}_k^2 + \tau^2$ and $u_k'^2 + \tilde{u}_k^2 + \tau^2$ are smaller than $\frac{\tilde{u}_{\max}^2}{u_{\min}^2}(u_k^2 + u_k'^2 + \tau^2)$, we get

$$\begin{aligned}
\sqrt{\sum_{k=1}^d \frac{(t_k - t_k')^2}{u_k^2 + u_k'^2 + \tau^2}} &\leq \sqrt{\sum_{k=1}^d \frac{(t_k - \tilde{t}_k)^2}{u_k^2 + u_k'^2 + \tau^2}} + \sqrt{\sum_{k=1}^d \frac{(t_k' - \tilde{t}_k)^2}{u_k^2 + u_k'^2 + \tau^2}}\,, \\
&\leq \frac{\tilde{u}_{\max}}{u_{\min}}\left(\sqrt{\sum_{k=1}^d \frac{(t_k - \tilde{t}_k)^2}{u_k^2 + \tilde{u}_k^2 + \tau^2}} + \sqrt{\sum_{k=1}^d \frac{(t_k' - \tilde{t}_k)^2}{u_k'^2 + \tilde{u}_k^2 + \tau^2}}\right)\,, \\
&\leq \frac{\tilde{u}_{\max}}{u_{\min}}\left(d(x, \tilde{x}) + d(x', \tilde{x})\right)\,.
\end{aligned}$$

Hence

$$d(x, x') \leq \frac{\tilde{u}_{\max}}{u_{\min}}\left(d(x, \tilde{x}) + d(x', \tilde{x})\right) + \sqrt{d\ln\left(\frac{\tilde{u}_{\max}^2}{u_{\min}^2}\right)}\,.$$

So $d(x, \tilde{x}) \geq \frac{u_{\min}}{\tilde{u}_{\max}}\left(d(x, x') - \sqrt{d\ln\left(\frac{\tilde{u}_{\max}^2}{u_{\min}^2}\right)}\right) - d(x', \tilde{x})$.

Conclusion: Let $x_j^0, x_i^0 \in \mathcal{X}$. Firstly, replacing $\tilde{u}_{\max}$ by $u_{\max}$ in the pseudo-quasi triangle inequality, for $\tilde{x} \in B_d(x_j^0, \Delta) \cap \mathcal{X}$, if $\frac{u_{\min}}{u_{\max}}\left(d(x_i^0, x_j^0) - \sqrt{d\ln\left(\frac{u_{\max}^2}{u_{\min}^2}\right)}\right) - d(\tilde{x}, x_j^0) > \Delta$ we have $d(x_i^0, \tilde{x}) > \Delta$. Note also that $d(\tilde{x}, x_j^0) < \Delta$, hence it suffices that $d(x_j^0, x_i^0) \geq 2\frac{u_{\max}}{u_{\min}}\Delta + \sqrt{d\ln\left(\frac{u_{\max}^2}{u_{\min}^2}\right)}$ for the open balls $\mathring{B}_d(x_j^0, \Delta) \cap \mathcal{X}$ to be disjoint.

Secondly, for $\tilde{x} \in \mathcal{G}(B_{\mathscr{d}}(x_j^0, r) \cap \mathcal{X})$, taking $\tilde{u}_{\max} = \sqrt{u_{\max}^2 + r^2(2u_{\max}^2 + \tau^2)}$ we have $\mathscr{d}(x_i^0, \tilde{x}) \geq \Delta$ as soon as $\frac{u_{\min}}{\tilde{u}_{\max}} \left(\mathscr{d}(x_i^0, x_j^0) - \sqrt{d \ln \left(\frac{\tilde{u}_{\max}^2}{u_{\min}^2} \right)} \right) \geq \Delta + r$. We used that $\mathcal{G}(B_{\mathscr{d}}(x_j, r)) \subset B_{\mathscr{d}}(x_j, r)$, so $\mathscr{d}(\tilde{x}, x_j) \leq r$ (see Lemma 5.2). This condition can be rewritten as $\mathscr{d}(x_i^0, x_j^0) \geq \frac{\tilde{u}_{\max}}{u_{\min}}(\Delta + r) + \sqrt{d \ln \left(\frac{u_{\max}^2}{u_{\min}^2} \right)}$. We proved that if

$$\min_{i \neq j} \mathscr{d}(x_i^0, x_j^0) \geq \max \left\{ \frac{\sqrt{u_{\max}^2 + r^2(2u_{\max}^2 + \tau^2)}}{u_{\min}}(\Delta + r) \, , \, 2\frac{u_{\max}}{u_{\min}}\Delta \right\} + \sqrt{d \ln \left(\frac{u_{\max}^2}{u_{\min}^2} \right)} =: \Delta_\tau \, ,$$

then the open balls $\mathring{B}_{\mathscr{d}}(x_j^0, \Delta) \cap \mathcal{X}$ are disjoint, and for all $j \neq i$, the ball $\mathcal{G}(B_{\mathscr{d}}(x_j^0, r) \cap \mathcal{X})$ does not intersect $\mathring{B}_{\mathscr{d}}(x_i^0, \Delta)$.

H.3.3 Local control with the semi-distance

Lemma H.4 (Local control of $\mathfrak{d}_{\mathfrak{g}}$ with the semi-distance, lower bound, $d = 1$). *Let $x_0, x \in \mathbb{R} \times [u_{\min}, +\infty)$. If $\mathscr{d}(x_0, x) \leq r$, then*

$$\mathfrak{d}_{\mathfrak{g}}(x_0, x)^2 \geq \frac{\mathscr{d}(x_0, x)^2}{\tilde{\varepsilon}_3}$$

for $\tilde{\varepsilon}_3 \geq 1 + \frac{1}{R(r)}$ with $R(r)$ defined by (57) below.

Proof. Let $x_0 = (t_0, u_0), x = (t, u) \in \mathbb{R} \times [u_{\min}, +\infty)$ such that $r_e := \mathscr{d}(x_0, x) \leq r$. Without loss of generality, we assume that $u_0 \leq u$.

We examine the case where u is "far" from u_0, and then the case where t is far from t_0 and u, u_0 are close.

Bounds on $B_{\mathscr{d}}(x_0, r)$, exhibition of 2 cases: Since $\mathscr{d}(x_0, x)^2 = \frac{(t_0 - t)^2}{u^2 + u_0^2 + \tau^2} + \ln \left(\frac{u^2 + u_0^2 + \tau^2}{\sqrt{2u^2 + \tau^2}\sqrt{2u_0^2 + \tau}} \right) = r_e^2$, we have that for all $0 < w < 1$,

$$\frac{(t_0 - t)^2}{u^2 + u_0^2 + \tau^2} \geq w r_e^2 \tag{50}$$

(first case) or

$$\frac{u^2 + u_0^2 + \tau^2}{\sqrt{2u^2 + \tau^2}\sqrt{2u_0^2 + \tau^2}} \geq e^{(1-w)r_e^2}$$

(second case). The second case can be rewritten as

$$\sqrt{u^2 + \frac{\tau^2}{2}} \geq \sqrt{u_0^2 + \frac{\tau^2}{2}}(e^{(1-w)r_e^2} + \sqrt{e^{2(1-w)r_e^2} - 1}) \tag{51}$$

(see (34)).

Assuming the first case holds and the second case does not: Provided that (50) holds, then $t_0 \neq t$ and the geodesic connecting x and x_0 is a semicircle (see Lemma H.2). We suppose additionally that we are not in the second case, namely that

$$\sqrt{u^2 + \frac{\tau^2}{2}} \leq \sqrt{u_0^2 + \frac{\tau^2}{2}}(e^{(1-w)r_e^2} + \sqrt{e^{2(1-w)r_e^2} - 1}) \, .$$

In particular, using (35) we have

$$\frac{|u_0^2 - u^2|}{u_0^2 + u^2 + \tau^2} \leq \sqrt{e^{2(1-w)r_e^2} - 1} \, . \tag{52}$$

Denoting l the arc length (*i.e.* $l = \mathfrak{d}_{\mathfrak{g}}(x, x_0)$), using (48) and (49) we have

$$\frac{|\tanh(\frac{c_2}{2} + \sqrt{2}l) - \tanh(\frac{c_2}{2})|}{\sqrt{2}} = |c_1||t - t_0| \tag{53}$$

for some c_2 and c_1 verifying

$$\frac{1}{2c_1^2} - \frac{\tau^2}{2} = \frac{1}{4}\left(\frac{-(t - t_0)^2 + u_0^2 - u^2}{t - t_0} \right)^2 + u_0^2 \, .$$

According to (50) and (52),

$$\begin{aligned}\frac{1}{|c_1||t-t_0|} &= \sqrt{\frac{1}{2}\left(-1+\frac{u_0^2-u^2}{(t-t_0)^2}\right)^2+\frac{2u_0^2+\tau^2}{(t-t_0)^2}}\,,\\ &= \sqrt{\frac{1}{2}+\frac{1}{2}\frac{(u_0^2-u^2)^2}{(t-t_0)^4}+\frac{u_0^2+u^2+\tau^2}{(t-t_0)^2}}\,,\\ &\le \sqrt{\frac{1}{2}+\frac{e^{2(1-w)r_e^2}-1}{2w^2r_e^4}+\frac{1}{wr_e^2}}\,,\\ &\le \frac{1}{\sqrt{w}r_e}\sqrt{\frac{r^2}{2}+\frac{1-w}{w}\frac{e^{2r^2}-1}{2r^2}+1}\end{aligned} \tag{54}$$

where we used that $y\mapsto e^{2y}-1$ is convex along with $(1-w)r_e^2\le r^2$, and $wr_e^2\le r^2$.

So as tanh is 1-Lipschitz,

$$\begin{aligned}\mathfrak{d}_{\mathfrak{g}}(x,x_0) = l &\ge |c_1||t-t_0|\,,\\ &\ge \frac{r_e\sqrt{w}}{\sqrt{1+\frac{r^2}{2}+\frac{1-w}{w}\frac{e^{2r^2}-1}{2r^2}}}\,.\end{aligned} \tag{55}$$

Assuming the second case holds: If (51) holds, we must consider the 2 possible geodesics (*c.f.* Lemma H.2). If $t=t_0$, the geodesic is of the form $\left(c_3,\ \sqrt{-\frac{\tau^2}{2}+\frac{c_1^2}{2}e^{\sqrt{8}y}}\right)$. As $\tilde{\gamma}(0)=x_0$ we have $c_1^2=2u_0^2+\tau^2$. Since $\tilde{\gamma}_u(l)=u$, we have $e^{\sqrt{8}l}=\frac{2u^2+\tau^2}{2u_0^2+\tau^2}$ from which we deduce that

$$\mathfrak{d}_{\mathfrak{g}}(x_0,x) = l = \frac{1}{\sqrt{2}}\ln\left(\frac{\sqrt{u^2+\frac{\tau^2}{2}}}{\sqrt{u_0^2+\frac{\tau^2}{2}}}\right) \ge \frac{1}{\sqrt{2}}\ln\left(e^{(1-w)r_e^2}+\sqrt{e^{2(1-w)r_e^2}-1}\right)\,.$$

If $t\neq t_0$, $\tilde{\gamma}_u$ is of the form $\sqrt{-\frac{\tau^2}{2}+\frac{1}{2\cosh^2\left(\frac{c_2}{2}+\sqrt{2}y\right)c_1^2}}$. Using that $\tilde{\gamma}_u(0)=u_0$ and $\tilde{\gamma}_u(l)=u$, we get $\frac{\cosh^2\left(\frac{c_2}{2}\right)}{\cosh^2\left(\frac{c_2}{2}+\sqrt{2}l\right)}=\frac{2u^2+\tau^2}{2u_0^2+\tau^2}$, from which we deduce using (51) that

$$\frac{\cosh\left(\frac{c_2}{2}\right)}{\cosh\left(\frac{c_2}{2}+\sqrt{2}l\right)} \ge e^{(1-w)r_e^2}+\sqrt{e^{2(1-w)r_e^2}-1}\,.$$

As

$$\frac{\cosh\left(\frac{c_2}{2}\right)}{\cosh\left(\frac{c_2}{2}+\sqrt{2}l\right)} = \cosh(\sqrt{2}l)-\frac{\sinh\left(\frac{c_2}{2}\right)\sinh(\sqrt{2}l)}{\cosh\left(\frac{c_2}{2}\right)} \le \cosh(\sqrt{2}l)+|\sinh(\sqrt{2}l)| = e^{|\sqrt{2}l|}\,,$$

it comes that $e^{\sqrt{2}l}\ge e^{(1-w)r_e^2}+\sqrt{e^{2(1-w)r_e^2}-1}$ leading again to

$$l \ge \frac{1}{\sqrt{2}}\ln\left(e^{(1-w)r_e^2}+\sqrt{e^{2(1-w)r_e^2}-1}\right)\,.$$

As $\frac{1}{\sqrt{2}}\ln\left(e^{y^2}+\sqrt{e^{2y^2}-1}\right)\ge y$ for all $y\in\mathbb{R}$, we get for this second case

$$\mathfrak{d}_{\mathfrak{g}}(x,x_0) \ge \sqrt{1-w}r_e\,. \tag{56}$$

Conclusion: We now choose $0<w<1$ depending only on r to balance the bounds of the two cases (namely, (55) and (56)). Writing $X=\frac{1-w}{w}$, we want to pick w such that

$$\begin{aligned}\sqrt{1-w} &= \frac{\sqrt{w}}{\sqrt{1+\frac{r^2}{2}+\frac{1-w}{w}\frac{e^{2r^2}-1}{2r^2}}}\\ \iff \frac{e^{2r^2}-1}{2r^2}X^2 &+ \left(1+\frac{r^2}{2}\right)X-1=0\,.\end{aligned}$$

This polynomial has exactly 1 positive root

$$R(r)=\frac{-\left(1+\frac{r^2}{2}\right)+\sqrt{\left(1+\frac{r^2}{2}\right)^2+2\frac{e^{2r^2}-1}{r^2}}}{\frac{e^{2r^2}-1}{r^2}}\,. \tag{57}$$

We can take $w_r=\frac{1}{R(r)+1}$. Then

$$\mathfrak{d}_{\mathfrak{g}}(x_0,x)\geq\sqrt{1-w_r}r_e=\sqrt{\frac{R(r)}{1+R(r)}}r_e\,.$$

□

Lemma H.5 (Local control of $\mathfrak{d}_{\mathfrak{g}}$ with the semi-distance, lower bound, $d\geq 1$)**.** *Let $x_0,x\in\mathbb{R}^d\times[u_{\min},+\infty)^d$. If $d(x_0,x)\leq r$, then*

$$\mathfrak{d}_{\mathfrak{g}}(x_0,x)^2\geq\frac{d(x_0,x)^2}{\tilde{\varepsilon}_3}$$

for $\tilde{\varepsilon}_3\geq 1+\frac{1}{R(r)}$ with $R(r)$ defined by (57).

Proof. Let $x_0,x\in\mathbb{R}^d\times[u_{\min},+\infty)^d$ such that $d(x_0,x)\leq r$. For all $k\in\{1,\dots,d\}$, $d(x_{0,k},x_k)\leq r$. Lemma H.4 gives $\mathfrak{d}_{\mathfrak{g}}(x_k,x_{0,k})^2\geq\frac{d(x_{0,k},x_k)^2}{1+\frac{1}{R(r)}}$. We conclude using that

$$\mathfrak{d}_{\mathfrak{g}}(x,x_0)^2=\sum_{k=1}^{d}\mathfrak{d}_{\mathfrak{g}}(x_k,x_{0,k})^2\geq\sum_{k=1}^{d}\frac{d(x_{0,k},x_k)^2}{1+\frac{1}{R(r)}}=\frac{d(x_0,x)^2}{1+\frac{1}{R(r)}}\,.$$

□

Lemma H.6 (Local control of $\mathfrak{d}_{\mathfrak{g}}$ with the semi-distance, upper bound, $d=1$)**.** *Let $x,x_0\in\mathbb{R}\times[u_{\min},+\infty)$. If $d(x,x_0)<\sqrt{2}$, then $\mathfrak{d}_{\mathfrak{g}}(x,x_0)^2\leq F(d(x,x_0))$ (F defined by (62) below).*

Proof. Let $x,x_0\in\mathbb{R}\times[u_{\min},+\infty)$. Without loss of generality, we can assume that $u_0\leq u$. We denote $l=\mathfrak{d}_{\mathfrak{g}}(x,x_0)$ and $r_e=d(x,x_0)$. The proof is close to that of Lemma H.4.
<u>First case:</u> If $t\neq t_0$, the geodesic connecting x and x_0 is a semicircle. Recalling (53) and (54), we have

$$\frac{|\tanh(\frac{c_2}{2}+\sqrt{2}l)-\tanh(\frac{c_2}{2})|}{\sqrt{2}|c_1|}=|t-t_0|$$

for some c_2 and c_1 verifying

$$\begin{aligned}\frac{1}{|c_1||t-t_0|}&=\sqrt{\frac{1}{2}+\frac{1}{2}\frac{(u_0^2-u^2)^2}{(t-t_0)^4}+\frac{u_0^2+u^2+\tau^2}{(t-t_0)^2}}\,,\\&\geq\frac{1}{r_e}\,.\end{aligned}$$

Hence

$$\left|\tanh\left(\frac{c_2}{2}+\sqrt{2}l\right)-\tanh\left(\frac{c_2}{2}\right)\right|\leq\sqrt{2}r_e\,. \tag{58}$$

Furthermore, using $\frac{u_0^2+u^2+\tau^2}{\sqrt{2u^2+\tau^2}\sqrt{2u_0^2+\tau^2}}\leq e^{r_e^2}$ together with (34) and recalling that $u_0\leq u$, we have

$$1\leq\frac{\cosh\left(\frac{c_2}{2}\right)}{\cosh\left(\frac{c_2}{2}+\sqrt{2}l\right)}=\sqrt{\frac{2u^2+\tau^2}{2u_0^2+\tau^2}}\leq e^{r_e^2}+\sqrt{e^{2r_e^2}-1}\,. \tag{59}$$

Using (59) along with the equality $\frac{\cosh(A+B)}{\cosh(A)}=\cosh(B)+\tanh(A)\sinh(B)$, we deduce that

$$\cosh\left(\sqrt{2}l\right)+\tanh\left(\frac{c_2}{2}\right)\sinh\left(\sqrt{2}l\right)\leq 1$$

and that

$$\cosh\left(\sqrt{2}l\right)-\tanh\left(\frac{c_2}{2}+\sqrt{2}l\right)\sinh\left(\sqrt{2}l\right)\leq e^{r_e^2}+\sqrt{e^{2r_e^2}-1}\,.$$

It comes

$$2\cosh\left(\sqrt{2}l\right)+\left(\tanh\left(\frac{c_2}{2}\right)-\tanh\left(\frac{c_2}{2}+\sqrt{2}l\right)\right)\sinh\left(\sqrt{2}l\right)\leq e^{r_e^2}+\sqrt{e^{2r_e^2}-1}+1\,,$$

so

$$2\cosh\left(\sqrt{2}l\right) - \sinh\left(\sqrt{2}l\right)\sqrt{2}r_e \leq e^{r_e^2} + \sqrt{e^{2r_e^2}-1} + 1$$

where we have used (58). As $\cosh \geq 0$ and $\frac{\sinh(\sqrt{2}l)}{\cosh(\sqrt{2}l)} \leq 1$, we get

$$\cosh(\sqrt{2}l)\left(1 - \frac{1}{\sqrt{2}}r_e\right) \leq \frac{1}{2}\left(e^{r_e^2} + \sqrt{e^{2r_e^2}-1} + 1\right).$$

For $r_e < \sqrt{2}$ we get

$$l^2 \leq \frac{1}{2}\operatorname{arcosh}\left(\frac{\frac{1}{2}\left(e^{r_e^2} + \sqrt{e^{2r_e^2}-1} + 1\right)}{1 - \frac{1}{\sqrt{2}}r_e}\right)^2. \tag{60}$$

Second case: If $t_0 = t$, using (47) we get $e^{\sqrt{8}l} = \frac{2u^2+\tau^2}{2u_0^2+\tau^2} \leq (e^{r_e^2} + \sqrt{e^{2r_e^2}-1})^2$ so

$$l^2 \leq \frac{1}{2}\ln(e^{r_e^2} + \sqrt{e^{2r_e^2}-1})^2 = \frac{1}{2}\operatorname{arcosh}(e^{r_e^2})^2. \tag{61}$$

Comparison of the bounds found, conclusion: Let $0 \leq r_e < \sqrt{2}$. As $1 - \frac{1}{\sqrt{2}}r_e \leq 1$ and $\sqrt{e^{2r_e^2}-1} \geq e^{r_e^2} - 1$, the bound found for the first case (see (60)) is larger than the one found for the second case (see (61)).

So if $d(x, x_0) < \sqrt{2}$, $\mathfrak{d}_{\mathfrak{g}}(x, x_0)^2 \leq F(d(x, x_0))$ with

$$F : y \in \mathbb{R}^+ \mapsto \frac{1}{2}\operatorname{arcosh}\left(\frac{\frac{1}{2}\left(e^{y^2} + \sqrt{e^{2y^2}-1} + 1\right)}{1 - \frac{1}{\sqrt{2}}y}\right)^2. \tag{62}$$

□

Lemma H.7 (Local control of $\mathfrak{d}_{\mathfrak{g}}$ with the semi-distance, upper bound, $d \geq 1$)**.** *Let $x, x_0 \in \mathbb{R}^d \times [u_{\min}, +\infty)^d$. If $d(x, x_0) < \sqrt{2}$, then $\mathfrak{d}_{\mathfrak{g}}(x, x_0)^2 \leq dF(d(x, x_0))$ (F defined by (62)).*

Proof. Let $x, x_0 \in \mathbb{R}^d \times [u_{\min}, +\infty)^d$ such that $d(x, x_0) = r_e < \sqrt{2}$. As for all $k = 1, \dots, d$, $d(x_k, x_{0,k}) \leq d(x, x_0)$, by Lemma H.6 it comes that

$$\begin{aligned}\mathfrak{d}_{\mathfrak{g}}(x, x_0)^2 &= \sum_{k=1}^{d} \mathfrak{d}_{\mathfrak{g}}(x_k, x_{0,k})^2 \\ &\leq \sum_{k=1}^{d} F(d(x_k, x_{0,k})), \\ &\leq dF(d(x, x_0))\end{aligned}$$

where we used that F is non-decreasing. □

H.4 Proof of Lemma 5.4

We use Lemma H.5. It comes that we can choose $\tilde{\varepsilon}_3 = 1 + \frac{1}{R\left(\frac{0.3025}{\sqrt{d}}\right)}$ with R defined by (57) in the appendix. We aim to find a more interpretable parameter. Remark that $r \in \mathbb{R}^+ \mapsto \frac{e^{2r^2}-1}{r^2}$ is increasing, along with $\frac{e^{2r^2}-1}{r^2} \geq 2$. We deduce that for all $0 < r \leq 0.3025$, $R(r) \geq \frac{-\left(1+\frac{0.3025^2}{2}\right)+\sqrt{5}}{\frac{e^{2\times 0.3025^2}-1}{0.3025^2}}$. So $1 + \frac{1}{R\left(\frac{0.3025}{\sqrt{d}}\right)} \leq 2.84$ (see [Giard, 2025, Section III]).

I Proof of Theorem 5.2

Construction of the certificates by solving a linear system We give explicit formulas for η, η_j of Theorem 5.2. These certificates are of the form (18). In the following, we write $\eta = \eta_{\alpha,\beta}$ and $\eta_j = \eta_{\alpha^j,\beta^j}$. Recalling Definitions 3.2 and 3.3, we want that for all $j = 1, \dots, s$, $\eta_{\alpha,\beta}(x_j^0) = 1$ and $\nabla\eta_{\alpha,\beta}(x_j^0) = 0_{2d}$. We also want $\eta_{\alpha^j,\beta^j}(x_j^0) = 1$, $\eta_{\alpha^j,\beta^j}(x_l^0) = 0$ for all $l \neq j$, and $\nabla\eta_{\alpha^j,\beta^j}(x_l^0) = 0_{2d}$.

These constraints translate to linear systems. Writing e_j the vector of size s containing a 1 at position j, and zeros elsewhere, along with $1_s = \sum_{j=1}^{s} e_j$, we want to solve

$$\Upsilon \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \begin{pmatrix} 1_s \\ 0_{s2d} \end{pmatrix} =: \boldsymbol{u}_s \quad \text{and} \quad \Upsilon \begin{pmatrix} \alpha^j \\ \beta^j \end{pmatrix} = \begin{pmatrix} e_j \\ 0_{s2d} \end{pmatrix} =: \boldsymbol{u}_s^j \quad \forall\, j = 1, \ldots, s$$
$$\text{where} \quad \Upsilon = \begin{pmatrix} (K_{\text{norm}}(x_i^0, x_j^0))_{i,j=1,\ldots,s} & (\nabla_1 K_{\text{norm}}(x_i^0, x_j^0))_{i,j=1,\ldots,s}^T \\ (\nabla_2 K_{\text{norm}}(x_i^0, x_j^0))_{i,j=1,\ldots,s} & (\nabla_1 \nabla_2 K_{\text{norm}}(x_i^0, x_j^0))_{i,j=1,\ldots,s} \end{pmatrix} \in \mathbb{R}^{s(1+2d)\times s(1+2d)} . \tag{63}$$

The following lemma entails that these linear systems admit a solution provided that a minimal separation condition holds.

Lemma I.1. *Let $\{x_j^0\}_{j=1}^s \subset \mathbb{R}^d \times [u_{\min}, +\infty)^d$. Assume that K_{norm} verifies Definition 5.2. If $\min_{i\neq j} \mathfrak{d}(x_i^0, x_j^0) \geq \Delta$, then Υ (see (63)) is invertible.*

Moreover, $\begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \Upsilon^{-1}\boldsymbol{u}_s$ and $\begin{pmatrix} \alpha^j \\ \beta^j \end{pmatrix} = \Upsilon^{-1}\boldsymbol{u}_s^j$ are well-defined and verify, for all $j = 1, \ldots, s$,

$$\begin{aligned} & \|\alpha\|_\infty \vee \left\|\alpha^j\right\|_\infty \leq \frac{1}{1-2h} , \\ & \max_{l=1,\ldots,s} \|\beta_l\|_{x_l^0} \vee \max_{l=1,\ldots,s} \left\|\beta_l^j\right\|_{x_l^0} \leq 4h \\ \textit{and} \quad & \|\alpha - 1_s\|_\infty \vee |\alpha_j^j - 1| \vee \max_{l\neq j} |\alpha_l^j| \leq \frac{2h}{1-2h} , \end{aligned} \tag{64}$$

where $h = \frac{1}{64} \min\left(\frac{\bar{\varepsilon}_0(r)}{B_0}, \frac{\bar{\varepsilon}_2(r)}{B_2}\right)$.

The proof can be found in [Poon et al., 2023, pp. 269-270]. This proof explicitly requires to handle normalized kernels. This motivates the introduction of W in $(\mathcal{P}_\kappa)$.

Controlling the certificates via the kernel We provide controls on the certificates we constructed in Lemma I.1 on the far and near regions, in order to show that they are non-degenerate (Definitions 3.2 and 3.3). To do so, we use Taylor expansions on Fisher-Rao geodesics along with bounds on the kernel stemming from the LPC (Definition 5.2).

Lemma I.2. *Let $r > 0$ and $x_0 \in \mathcal{X} \subset \mathbb{R}^d \times [u_{\min}, +\infty)^d$. We define*

$$\mathcal{X}_0^{near}(r) \coloneqq \{x' \in \mathcal{X} \;:\; \mathfrak{d}(x', x_0) \leq r\} .$$

Assume that there exists $\bar{\varepsilon}_2 > 0$ such that, for all $x \in \mathcal{G}(\mathcal{X}_0^{near}(r))$ and $v \in \mathbb{R}^{2d}$,

$$-K_{\text{norm}}^{(02)}(x_0, x)[v, v] \geq \bar{\varepsilon}_2 \|v\|_x^2 \quad \textit{and} \quad \left\|K_{\text{norm}}^{(02)}(x_0, x)\right\|_x \leq B_{02} .$$

Let $\eta : \mathbb{R}^d \times [u_{\min}, +\infty)^d \to \mathbb{R}$ be a $\mathcal{C}^2$ function. Then the following holds:

(i) If $\eta(x_0) = 0$, $\nabla\eta(x_0) = 0$ and $\|D_2[\eta](x)\|_x \leq \delta$ for all $x \in \mathcal{G}(\mathcal{X}_0^{near}(r))$, then $|\eta(x)| \leq \delta \mathfrak{d}_{\mathfrak{g}}(x, x_0)^2$ for all $x \in \mathcal{X}_0^{near}(r)$.

(ii) Let $a \in \{-1, 1\}$. If $\eta(x_0) = a$, $\nabla\eta(x_0) = 0$ and $\left\|aD_2[\eta](x) - K_{\text{norm}}^{(02)}(x_0, x)\right\|_x \leq \delta$ for all $x \in \mathcal{G}(\mathcal{X}_0^{near}(r))$, for some $\delta < \bar{\varepsilon}_2$, then

$$1 - \frac{B_{02} + \delta}{2} \mathfrak{d}_{\mathfrak{g}}(x, x_0)^2 \leq a\eta(x) \leq 1 - \frac{\bar{\varepsilon}_2 - \delta}{2} \mathfrak{d}_{\mathfrak{g}}(x, x_0)^2 \quad \forall\, x \in \mathcal{X}_0^{near}(r) .$$

Proof. This proof is based on [Poon et al., 2023, Lemma 2]. We only show (ii), as the proof for (i) is similar. Let $x \in \mathcal{X}_0^{near}(r)$. We denote by γ the geodesic for the metric $\mathfrak{g}$ between x_0 and x, parametrized between 0 and 1. We refer to Section H.2 for a description of the geodesic properties and related notations. By definition, $\gamma(y) \in \mathcal{G}(\mathcal{X}_0^{near}(r))$ for all $y \in [0, 1]$. Hence, for all $y \in [0, 1]$,

$$\left\|aD_2[\eta](\gamma(y)) - K_{\text{norm}}^{(02)}(x_0, \gamma(y))\right\|_{\gamma(y)} \leq \delta \quad \text{and} \quad -K_{\text{norm}}^{(02)}(x_0, \gamma(y))[\dot{\gamma}(y), \dot{\gamma}(y)] \geq \bar{\varepsilon}_2 \|\dot{\gamma}(y)\|_{\gamma(y)}^2 ,$$

from which we deduce that

$$aD_2[\eta](\gamma(y))[\dot\gamma(y),\dot\gamma(y)] \le (-\bar\varepsilon_2+\delta)\left\|\dot\gamma(y)\right\|^2_{\gamma(y)} .$$

By a Taylor expansion, we get

$$\begin{aligned} a\eta(x) &= a\eta(x_0) + a\nabla\eta(x_0)^T\dot\gamma(0) + \int_0^1 (1-y)aD_2[\eta](\gamma(y))[\dot\gamma(y),\dot\gamma(y)]\,\mathrm{d}y\,,\\ &= 1 + \int_0^1 (1-y)aD_2[\eta](\gamma(y))[\dot\gamma(y),\dot\gamma(y)]\,\mathrm{d}y\,,\\ &\le 1-(\bar\varepsilon_2-\delta)\int_0^1 (1-y)\left\|\dot\gamma(y)\right\|^2_{\gamma(y)}\,\mathrm{d}y\,,\\ &= 1-\frac{\bar\varepsilon_2-\delta}{2}\mathfrak{d}_{\mathfrak{g}}(x,x_0)^2\,. \end{aligned}$$

Using the same reasoning, we also have $a\eta(x)\ge 1-\frac{B_{02}+\delta}{2}\mathfrak{d}_{\mathfrak{g}}(x,x_0)^2$.

Note that we do not have (unlike in [Poon et al., 2023, Lemma 2]) the bound $a\eta(x)\ge -1+\frac{\bar\varepsilon_2-\delta}{2}\mathfrak{d}_{\mathfrak{g}}(x,x_0)^2$. This is not a problem in our framework: as we work with nonnegative measures, we do not need to control the negative part of the certificate. □

Theorem I.1. *Let $\{x_j^0\}_{j=1}^s\subset\mathcal{X}\subset\mathbb{R}^d\times[u_{\min},u_{\max}]^d$. Assume that K_{norm} verifies Definition 5.2. If $\min_{i\ne j}\mathscr{d}(x_i^0,x_j^0)\ge\Delta_\tau$ (defined in Lemma 5.3), then the certificates constructed in Lemma I.1 are non-degenerate. The global certificate $\eta_{\alpha,\beta}$ is $(\frac78\bar\varepsilon_0,\frac{15}{32}\bar\varepsilon_2,r)$-non-degenerate and the local certificate η_{α^j,β^j} is $(\frac78\bar\varepsilon_0,\frac{B_{02}+\bar\varepsilon_2/16}{2},r)$-non-degenerate.*

Proof. This proof is largely based on [Poon et al., 2023, pp. 270-241]. First remark that Lemmas I.1 and 5.3 hold under these assumptions. We denote $h=\frac{1}{64}\min\left(\frac{\bar\varepsilon_0(r)}{B_0},\frac{\bar\varepsilon_2(r)}{B_2}\right)$.

Control on the far region: Let $x\in\mathcal{X}^{far}(r)$. As the open balls $\mathring{B}_{\mathscr{d}}(x_i^0,\Delta)\cap\mathcal{X}$ are disjoint (Lemma 5.3), there exists at most one index j such that $\mathscr{d}(x,x_j^0)<\Delta$ and for all $i\ne j$, $\mathscr{d}(x,x_i^0)\ge\Delta$. So using bounds displayed in Lemma I.1 (see (64)),

$$\begin{aligned} |\eta_{\alpha,\beta}(x)| &= \left|\alpha_jK_{\text{norm}}(x_j^0,x)+\sum_{j\ne i}\alpha_iK_{\text{norm}}(x_i^0,x)+\beta_j^TK^{(10)}_{\text{norm}}(x_j^0,x)+\sum_{j\ne i}\beta_i^TK^{(10)}_{\text{norm}}(x_i^0,x)\right|\,,\\ &\le \|\alpha\|_\infty\left(|K_{\text{norm}}(x_j^0,x)|+\sum_{j\ne i}|K_{\text{norm}}(x_i^0,x)|\right)+\max_i\|\beta_i\|_{x_i^0}\left(\left\|K^{(10)}_{\text{norm}}(x_j^0,x)\right\|_{x_j^0}+\sum_{j\ne i}\left\|K^{(10)}_{\text{norm}}(x_i^0,x)\right\|_{x_i^0}\right)\,,\\ &\le \frac{1}{1-2h}(1-\bar\varepsilon_0+h)+4h(B_{10}+h)\,,\\ &\le 1-\frac{\bar\varepsilon_0-3h}{1-2h}+4h(B_{10}+h)\,,\\ &\le 1-\bar\varepsilon_0+3h+4h(B_{10}+h)\,,\\ &\le 1-\bar\varepsilon_0+3\frac{\bar\varepsilon_0}{64}+4\frac{\bar\varepsilon_0}{64}+4\frac{\bar\varepsilon_0}{64^2}\,,\\ &\le 1-\frac78\bar\varepsilon_0\,. \end{aligned}$$

We can apply the same reasoning to show that $|\eta_{\alpha^j,\beta^j}(x)|\le 1-\frac78\bar\varepsilon_0$.

Controls on the near regions: Let $x\in\mathcal{G}(\mathcal{X}_j^{near}(r))$. We have

$$D_2[\eta_{\alpha,\beta}](x)=K^{(02)}_{\text{norm}}(x_j^0,x)+(\alpha_j-1)K^{(02)}_{\text{norm}}(x_j^0,x)+\sum_{i\ne j}\alpha_iK^{(02)}_{\text{norm}}(x_i^0,x)+[\beta_j]K^{(12)}_{\text{norm}}(x_j^0,x)+\sum_{i\ne j}[\beta_i]K^{(12)}_{\text{norm}}(x_i^0,x)\,.$$

As $\mathcal{G}(B_{\mathscr{d}}(x_j^0,r)\cap\mathcal{X})$ is disjoint from $\mathring{B}_{\mathscr{d}}(x_i^0,\Delta)$ for $i\ne j$ (Lemma 5.3), we have $\mathscr{d}(x,x_i^0)\ge\Delta$ for all $i\ne j$. Using Lemma I.1, it comes

$$\begin{aligned} \left\|D_2[\eta_{\alpha,\beta}](x)-K^{(02)}_{\text{norm}}(x_j^0,x)\right\|_x &\le |\alpha_j-1|B_{02}+\|\alpha\|_\infty h+\max_i\|\beta_i\|_{x_i^0}(B_{12}+h)\,,\\ &\le \frac{2h}{1-2h}B_{02}+\frac{h}{1-2h}+4h(B_{12}+h)\,,\\ &\le \frac{\bar\varepsilon_2}{16}\,. \end{aligned}$$

Using Lemma I.2 with $\delta = \frac{\bar{\varepsilon}_2}{16}$, we deduce that for all $x \in \mathcal{X}_j^{near}(r)$, $\eta_{\alpha,\beta}(x) \leq 1 - \frac{15}{32}\bar{\varepsilon}_2 \mathfrak{d}_{\mathfrak{g}}(x, x_j^0)^2$.

With the same reasoning, for all $x \in \mathcal{G}(\mathcal{X}_j^{near}(r))$ we can obtain $\left\| D_2[\eta_{\alpha^j,\beta^j}](x) - K_{\text{norm}}^{(02)}(x_j^0, x) \right\|_x \leq \frac{\bar{\varepsilon}_2}{16}$. So from Lemma I.2, for all $x \in \mathcal{X}_j^{near}(r)$ we have $|1 - \eta_{\alpha^j,\beta^j}(x)| \leq \frac{B_{02}+\bar{\varepsilon}_2/16}{2} \mathfrak{d}_{\mathfrak{g}}(x, x_j^0)^2$. We used that $B_{02} \geq \bar{\varepsilon}_2$.

We also have, for $i \neq j$, and $x \in \mathcal{G}(\mathcal{X}_i^{near}(r))$,

$$
\begin{aligned}
\left\| D_2[\eta_{\alpha^j,\beta^j}](x) \right\|_x &= \left\| \alpha_i^j K_{\text{norm}}^{(02)}(x_i^0, x) + \sum_{l \neq i} \alpha_l^j K_{\text{norm}}^{(02)}(x_l^0, x) + [\beta_i^j] K_{\text{norm}}^{(12)}(x_i^0, x) + \sum_{l \neq i} [\beta_l^j] K_{\text{norm}}^{(12)}(x_l^0, x) \right\|_x , \\
&\leq |\alpha_i^j| B_{02} + \left\| \alpha^j \right\|_\infty h + \max_i \left\| \beta_i^j \right\|_{x_i^0} (B_{12} + h) , \\
&\leq \frac{2h}{1-2h} B_{02} + \frac{h}{1-2h} + 4h(B_{12} + h) , \\
&\leq \frac{\bar{\varepsilon}_2}{16} .
\end{aligned}
$$

Lemma I.2 ensures that for all $x \in \mathcal{X}_i^{near}(r)$, $|\eta_{\alpha^j,\beta^j}(x)| \leq \frac{\bar{\varepsilon}_2}{16} \mathfrak{d}_{\mathfrak{g}}(x, x_i^0)^2$. This concludes the proof. □

Norm of the certificates We use bounds on the norm of the certificates to get the controls in estimation and in prediction (c_p in Assumption 1). The construction of certificates presented in Lemma I.1 allows us to obtain the bounds displayed in the following proposition.

Proposition I.1. *The certificates constructed in Lemma I.1 verify, for all $x \in \mathbb{R}^d \times [u_{\min}, +\infty)^d$, for all $j \in \{1, \dots, s\}$,*

$$
\eta_{\alpha,\beta}(x) = \langle \Psi\delta_x, p_{\alpha,\beta} \rangle_{\mathbb{L}} \quad \textit{and} \quad \eta_{\alpha^j,\beta^j}(x) = \left\langle \Psi\delta_x, p_{\alpha^j,\beta^j} \right\rangle_{\mathbb{L}} \quad \textit{with} \quad \| p_{\alpha,\beta} \|_{\mathbb{L}} \leq \sqrt{2s} , \quad \left\| p_{\alpha^j,\beta^j} \right\|_{\mathbb{L}} \leq \sqrt{2} .
$$

Proof. From Lemma 5.1, for all $j \in \{1, \dots, s\}$,

$$
p_{\alpha,\beta} = \sum_{i=1}^{s} \alpha_i \Psi \delta_{x_i^0} + \sum_{i=1}^{s} \beta_i \nabla_x \left(\Psi \delta_{x_i^0} \right)
$$

and

$$
p_{\alpha^j,\beta^j} = \sum_{i=1}^{s} \alpha_i^j \Psi \delta_{x_i^0} + \sum_{i=1}^{s} \beta_i^j \nabla_x \left(\Psi \delta_{x_i^0} \right) .
$$

So recalling (63) and defining

$$
D_{\mathfrak{g}} \coloneqq \begin{pmatrix} \mathrm{Id}_s & & & \\ & \mathfrak{g}_{x_1^0}^{-\frac{1}{2}} & & \\ & & \ddots & \\ & & & \mathfrak{g}_{x_s^0}^{-\frac{1}{2}} \end{pmatrix} \in \mathbb{R}^{s(d+1) \times s(d+1)} , \quad \tilde{\Upsilon} \coloneqq D_{\mathfrak{g}} \Upsilon D_{\mathfrak{g}} ,
$$

using the results of [Poon et al., 2023, p. 268] we have $\| p_{\alpha,\beta} \|_{\mathbb{L}}^2 = \begin{pmatrix} \alpha \\ \beta \end{pmatrix}^T \Upsilon \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \boldsymbol{u}_s^T \tilde{\Upsilon}^{-1} \boldsymbol{u}_s$. We can apply [Poon et al., 2023, Lemma 3] that gives $\left\| \tilde{\Upsilon}^{-1} \right\|_2 \leq 2$. So $\| p_{\alpha,\beta} \|_{\mathbb{L}}^2 \leq \left\| \tilde{\Upsilon}^{-1} \right\|_2 \| \boldsymbol{u}_s \|_2^2 \leq 2s$. We repeat the same reasoning to bound $\left\| p_{\alpha^j,\beta^j} \right\|_{\mathbb{L}}$: we have $\left\| p_{\alpha^j,\beta^j} \right\|_{\mathbb{L}}^2 = (\boldsymbol{u}_s^j)^T \tilde{\Upsilon}^{-1} \boldsymbol{u}_s^j \leq \left\| \tilde{\Upsilon}^{-1} \right\|_2 \| \boldsymbol{u}_s^j \|_2^2 \leq 2$. □

The combination of Theorem I.1 and Proposition I.1 concludes the proof of Theorem 5.2.

J Proof of Theorem 5.3

Lemma J.1. *Let $x, x' \in \mathbb{R}^d \times [u_{\min}, +\infty)^d$. We can establish that*

$$\begin{aligned}
\left\|K^{(00)}_{\text{norm}}(x,x')\right\|_{x,x'} &= |K_{\text{norm}}(x,x')|\,,\\
\left\|K^{(10)}_{\text{norm}}(x,x')\right\|_{x,x'} &= \left\|\mathfrak{g}_x^{-1/2}\nabla_1 K_{\text{norm}}(x,x')\right\|_2\,,\\
\left\|K^{(11)}_{\text{norm}}(x,x')\right\|_{x,x'} &= \left\|\mathfrak{g}_x^{-1/2}\nabla_1\nabla_2 K_{\text{norm}}(x,x')\mathfrak{g}_{x'}^{-1/2}\right\|_2\,,\\
\left\|K^{(02)}_{\text{norm}}(x,x')\right\|_{x,x'} &= \left\|\mathfrak{g}_{x'}^{-1/2}H_2^{\mathfrak{g}} K_{\text{norm}}(x,x')\mathfrak{g}_{x'}^{-1/2}\right\|_2\,,\\
\left\|K^{(12)}_{\text{norm}}(x,x')\right\|_{x,x'} &\le \sqrt{2d}\max_{k=1,\dots d}\Big\{\left\|\mathfrak{g}_{t_k t_k}^{-1/2}\mathfrak{g}_{x'}^{-1/2}\partial_{t_k}H_2^{\mathfrak{g}}K_{\text{norm}}(x,x')\mathfrak{g}_{x'}^{-1/2}\right\|_2\,,\\
&\qquad\qquad \left\|\mathfrak{g}_{u_k u_k}^{-1/2}\mathfrak{g}_{x'}^{-1/2}\partial_{u_k}H_2^{\mathfrak{g}}K_{\text{norm}}(x,x')\mathfrak{g}_{x'}^{-1/2}\right\|_2\Big\}\,.
\end{aligned}$$

Proof. To get the simplified expressions for the operator norms (Definition 5.1), we use that $\|v\|_x^2 := v^T\mathfrak{g}_x v$. The result for $\left\|K^{(10)}_{\text{norm}}(x,x')\right\|_{x,x'}$, $\left\|K^{(11)}_{\text{norm}}(x,x')\right\|_{x,x'}$, $\left\|K^{(02)}_{\text{norm}}(x,x')\right\|_{x,x'}$ is stated by [Poon et al., 2023, Equation (27)].

To deal with $\left\|K^{(12)}_{\text{norm}}(x,x')\right\|_{x,x'}$, we also use that our metric is diagonal. Denoting $\tilde{q} = \sqrt{\mathfrak{g}_x}q$, $\tilde{V}_1 = \sqrt{\mathfrak{g}_{x'}}V_1$, $\tilde{V}_2 = \sqrt{\mathfrak{g}_{x'}}V_2$, we have

$$\begin{aligned}
\left\|K^{(12)}_{\text{norm}}(x,x')\right\|_{x,x'} &= \sup_{\substack{\|V_1\|_{x'},\|V_2\|_{x'}\le 1\\ \|q=(q_1,\dots q_{2d})\|_x\le 1}} \sum_{k=1}^d q_k V_1^T\partial_{t_k}H_2^{\mathfrak{g}}K_{\text{norm}}(x,x')V_2 + \sum_{k=1}^d q_{k+d}V_1^T\partial_{u_k}H_2^{\mathfrak{g}}K_{\text{norm}}(x,x')V_2\,,\\
&= \sup_{\|\tilde{V}_1\|_2\,,\,\|\tilde{V}_2\|_2\le 1\,,\,\|\tilde{q}\|_2\le 1}\Bigg(\sum_{k=1}^d \tilde{q}_k\tilde{V}_1^T\mathfrak{g}_{t_k t_k}^{-1/2}\mathfrak{g}_{x'}^{-1/2}\partial_{t_k}H_2^{\mathfrak{g}}K_{\text{norm}}(x,x')\mathfrak{g}_{x'}^{-1/2}\tilde{V}_2\\
&\qquad + \sum_{k=1}^d \tilde{q}_{k+d}\tilde{V}_1^T\mathfrak{g}_{u_k u_k}^{-1/2}\mathfrak{g}_{x'}^{-1/2}\partial_{u_k}H_2^{\mathfrak{g}}K_{\text{norm}}(x,x')\mathfrak{g}_{x'}^{-1/2}\tilde{V}_2\Bigg)\,,\\
&\le \sqrt{2d}\max_{k=1,\dots d}\Bigg\{\sup_{\|\tilde{V}_1\|_2,\|\tilde{V}_2\|_2\le 1}\left|\tilde{V}_1^T\mathfrak{g}_{t_k t_k}^{-1/2}\mathfrak{g}_{x'}^{-1/2}\partial_{t_k}H_2^{\mathfrak{g}}K_{\text{norm}}(x,x')\mathfrak{g}_{x'}^{-1/2}\tilde{V}_2\right|\,,\\
&\qquad \sup_{\|\tilde{V}_1\|_2,\|\tilde{V}_2\|_2\le 1}\left|\tilde{V}_1^T\mathfrak{g}_{u_k u_k}^{-1/2}\mathfrak{g}_{x'}^{-1/2}\partial_{u_k}H_2^{\mathfrak{g}}K_{\text{norm}}(x,x')\mathfrak{g}_{x'}^{-1/2}\tilde{V}_2\right|\Bigg\}\,,
\end{aligned}$$

giving the desired result. □

Global controls To check the first condition of Definition 5.2, we give global controls for the Riemannian derivatives of the kernel. We take advantage of the metric used, which allows us to have uniform bounds on $\mathbb{R}^d \times [u_{\min}, +\infty)^d$.

Lemma J.2 (Global controls). *For $(i,j) \in \{0,1\}\times\{0,1,2\}$ we have*

$$\sup_{x,x'\in\mathbb{R}^d\times[u_{\min},+\infty)^d}\left\|K^{(ij)}_{\text{norm}}(x,x')\right\|_{x,x'} \le B_{ij}$$

with

$$\begin{aligned}
B_{00} &= 1\,,\\
B_{10} = B_{01} &= \sqrt{2d}\,,\\
B_{11} &= 2d\,,\\
B_{02} = B_{20} &= \sqrt{4d^2+10d}\,,\\
B_{12} = B_{21} &= \sqrt{2d}B_{02}\,.
\end{aligned}$$

Proof. The bounds are based on the expressions given in Lemma J.1. We make use of the relation between the Frobenius norm and the 2-norm:

$$\forall\, M = (m_{ij})_{\substack{1\le i\le n\\ 1\le j\le m}} \in \mathbb{R}^{n\times m}\,, \quad \|M\|_2 \le \sqrt{\sum_{\substack{1\le i\le n\\ 1\le j\le m}} m_{ij}^2} \le \sqrt{nm \max_{i,j} m_{ij}^2}\,.$$

Let $x, x' \in \mathbb{R}^d \times [u_{\min}, +\infty)^d$.

<u>B_{00}</u>: Using Cauchy-Schwarz and since the kernel is normalized, we have

$$|K_{\mathrm{norm}}(x,x')| = |\,\langle \Psi\delta_x, \Psi\delta_{x'}\rangle_{\mathbb{L}}\,| \le \sup_{x\in\mathbb{R}^d\times[u_{\min},+\infty)^d} \|\Psi\delta_x\|_{\mathbb{L}}^2 = \sup_{x\in\mathbb{R}^d\times[u_{\min},+\infty)^d} K_{\mathrm{norm}}(x,x) = 1\,.$$

<u>B_{10}</u>: We use that $\mathfrak{g}_x$ is diagonal, and name its diagonal elements by $\mathrm{diag}(\mathfrak{g}_{t_1t_1},\dots,\mathfrak{g}_{t_dt_d},\mathfrak{g}_{u_1u_1},\dots,\mathfrak{g}_{u_du_d}) = \mathfrak{g}_x$. Remark that $\nabla_1 K_{\mathrm{norm}}(x,x') = \langle \nabla\Psi\delta_x, \Psi\delta_{x'}\rangle_{\mathbb{L}}$ along with, for all $k\in\{1,\dots d\}$, $\|\partial_{t_k}\Psi\delta_x\|_{\mathbb{L}} = \mathfrak{g}_{t_kt_k}^{1/2}$ and $\|\partial_{u_k}\Psi\delta_x\|_{\mathbb{L}} = \mathfrak{g}_{u_ku_k}^{1/2}$. We use Cauchy-Schwarz to write

$$\begin{aligned}
\left\|\mathfrak{g}_x^{-1/2}\nabla_1 K_{\mathrm{norm}}(x,x')\right\|_2 &= \sqrt{\sum_{k=1}^d \mathfrak{g}_{t_kt_k}^{-1}\,\langle\partial_{t_k}\Psi\delta_x,\Psi\delta_{x'}\rangle_{\mathbb{L}}^2 + \sum_{k=1}^d \mathfrak{g}_{u_ku_k}^{-1}\,\langle\partial_{u_k}\Psi\delta_x,\Psi\delta_{x'}\rangle_{\mathbb{L}}^2}\,,\\
&\le \sqrt{\sum_{k=1}^d \|\Psi\delta_{x'}\|_{\mathbb{L}}^2 + \sum_{k=1}^d \|\Psi\delta_{x'}\|_{\mathbb{L}}^2}\,,\\
&\le \sqrt{2d}\,.
\end{aligned}$$

<u>B_{11}</u>: We use the same reasoning. With $\mathfrak{g}_x^{-1/2}\nabla_1\nabla_2 K_{\mathrm{norm}}(x,x')\mathfrak{g}_{x'}^{-1/2} =: (m_{ij})_{1\le i,j\le 2d}$, we have m_{ij} of the form $\mathfrak{g}_{b_kb_k}^{-1/2}\mathfrak{g}_{b_lb_l}^{-1/2}\,\langle\partial_{b_k}\Psi\delta_x,\partial_{b_l}\Psi\delta_{x'}\rangle_{\mathbb{L}}$ where b_m stands for u_m or t_m, $1\le m\le d$. So

$$m_{ij}^2 \le \mathfrak{g}_{b_kb_k}^{-1}\mathfrak{g}_{b_lb_l}^{-1}\,\|\partial_{b_k}\Psi\delta_x\|_{\mathbb{L}}^2\,\|\partial_{b_l}\Psi\delta_{x'}\|_{\mathbb{L}}^2 \le 1\,,$$

from which we conclude that $\|M\|_2 \le \sqrt{4d^2} = 2d$.

<u>B_{02}</u>: Here, we denote $(m_{ij})_{1\le i,j\le 2d} = \mathfrak{g}_{x'}^{-1/2} H_2^{\mathfrak{g}} K_{\mathrm{norm}}(x,x')\mathfrak{g}_{x'}^{-1/2}$. Recall that

$$H_2^{\mathfrak{g}}K_{\mathrm{norm}}(x,x') = \nabla_2^2 K_{\mathrm{norm}}(x,x') - \sum_{k=1}^d \Gamma^{t'_k}\partial_{t'_k}K_{\mathrm{norm}}(x,x') - \sum_{k=1}^d \Gamma^{u'_k}\partial_{u'_k}K_{\mathrm{norm}}(x,x')\,.$$

Using that $\Gamma^{b_k}{}_{b_lb_m} \ne 0$ implies that $b_k = b_l = t_k, b_m = u_k$ or $b_k = b_m = t_k, b_l = u_k$ or $b_k = u_k, b_l = b_m = t_k$ or $b_k = b_l = b_m = u_k$, we can treat the $4d$ terms $m_{ii}, m_{i2i}, m_{2ii}, m_{2i2i}$ ($i\in\{1,\dots d\}$) separately from the rest of the matrix.

The other terms are of the form

$$\mathfrak{g}_{b'_kb'_k}^{-1/2}\mathfrak{g}_{b'_lb'_l}^{-1/2}\partial_{b'_kb'_l}K_{\mathrm{norm}}(x,x')$$

where $k\ne l$ (*i.e.* the derivatives are associated with different dimensions). By abuse of notation, we denote $K_{\mathrm{norm}}(x_m,x'_m)$ an evaluation of the kernel in dimension $d=1$ at points $x_m = (t_m,u_m), x'_m = (t'_m,u'_m)$, and we consider in a similar way $\Psi\delta_{x_m}$. Remark that $\mathfrak{g}_{b_mb_m} = \|\partial_{b_m}\Psi\delta_{x_m}\|_{\mathbb{L}}^2$. As $K_{\mathrm{norm}}(x,x') = \prod_{m=1}^d K_{\mathrm{norm}}(x_m,x'_m)$, for $k\ne l$ we have

$$\partial_{b'_kb'_l}K_{\mathrm{norm}}(x,x') = \partial_{b'_k}K_{\mathrm{norm}}(x_k,x'_k)\partial_{b'_l}K_{\mathrm{norm}}(x_l,x'_l)\prod_{m\notin\{k,l\}}K_{\mathrm{norm}}(x_m,x'_m)\,. \tag{65}$$

So using Cauchy-Schwarz, for m_{ij} such that $i\ne j$ and $i\ne 2j$, $j\ne 2i$,

$$m_{ij}^2 \le \mathfrak{g}_{b'_kb'_k}^{-1}\mathfrak{g}_{b'_lb'_l}^{-1}\left(\partial_{b'_k}K_{\mathrm{norm}}(x_k,x'_k)\right)^2\left(\partial_{b'_l}K_{\mathrm{norm}}(x_l,x'_l)\right)^2\prod_{m\notin\{k,l\}}K_{\mathrm{norm}}(x_m,x'_m)^2 \le 1\,.$$

This quantity bounds $(2d)^2 - 4d$ squared coefficients of M.

The terms m_{ii}, m_{i2i} and m_{2ii}, m_{2i2i} ($i\in\{1,\dots d\}$) are of the form

$$\mathfrak{g}_{t'_kt'_k}^{-1}\left(\partial_{t'_kt'_k}K_{\mathrm{norm}}(x,x') - \Gamma^{u'_k}{}_{t'_kt'_k}\partial_{u'_k}K_{\mathrm{norm}}(x,x')\right) \tag{66}$$

$$\text{or}\quad \mathfrak{g}_{t'_kt'_k}^{-1/2}\mathfrak{g}_{u'_ku'_k}^{-1/2}\left(\partial_{t'_ku'_k}K_{\mathrm{norm}}(x,x') - \Gamma^{t'_k}{}_{t'_ku'_k}\partial_{t'_k}K_{\mathrm{norm}}(x,x')\right) \tag{67}$$

$$\text{or}\quad \mathfrak{g}_{u'_ku'_k}^{-1}\left(\partial_{u'_ku'_k}K_{\mathrm{norm}}(x,x') - \Gamma^{u'_k}{}_{u'_ku'_k}\partial_{u'_k}K_{\mathrm{norm}}(x,x')\right)\,. \tag{68}$$

These forms concern respectively d, $2d$, d coefficients. We can again reduce the problem to dimension 1 using the decomposition of the kernel:

$$\begin{aligned}
&\partial_{t'_k t'_k} K_{\text{norm}}(x,x') - \Gamma^{u'_k}{}_{t'_k t'_k}\partial_{u'_k} K_{\text{norm}}(x,x') = \left(\partial_{t'_k t'_k} K_{\text{norm}}(x_k,x'_k) - \Gamma^{u'_k}{}_{t'_k t'_k}\partial_{u'_k} K_{\text{norm}}(x_k,x'_k)\right)\prod_{l\neq k} K_{\text{norm}}(x_l,x'_l)\,,\\
&\partial_{t'_k u'_k} K_{\text{norm}}(x,x') - \Gamma^{t'_k}{}_{t'_k u'_k}\partial_{t'_k} K_{\text{norm}}(x,x') = \left(\partial_{t'_k u'_k} K_{\text{norm}}(x_k,x'_k) - \Gamma^{t'_k}{}_{t'_k u'_k}\partial_{t'_k} K_{\text{norm}}(x_k,x'_k)\right)\prod_{l\neq k} K_{\text{norm}}(x_l,x'_l)\,,\\
&\partial_{u'_k u'_k} K_{\text{norm}}(x,x') - \Gamma^{u'_k}{}_{u'_k u'_k}\partial_{u'_k} K_{\text{norm}}(x,x') = \left(\partial_{u'_k u'_k} K_{\text{norm}}(x_k,x'_k) - \Gamma^{u'_k}{}_{u'_k u'_k}\partial_{u'_k} K_{\text{norm}}(x_k,x'_k)\right)\prod_{l\neq k} K_{\text{norm}}(x_l,x'_l)\,.
\end{aligned} \tag{69}$$

So for the first form (66),

$$\begin{aligned}
m_{ii}^2 &= \mathfrak{g}_{t'_k t'_k}^{-2}\left\langle \Psi\delta_x, \partial_{t'_k t'_k}\Psi\delta_{x'} - \Gamma^{u'_k}{}_{t'_k t'_k}\partial_{u'_k}\Psi\delta_{x'}\right\rangle_{\mathbb{L}}^2\,,\\
&\leq \mathfrak{g}_{t'_k t'_k}^{-2}\left\|\partial_{t'_k t'_k}\Psi\delta_{x'_k} - \Gamma^{u'_k}{}_{t'_k t'_k}\partial_{u'_k}\Psi\delta_{x'_k}\right\|_{\mathbb{L}}^2\,,\\
&= \mathfrak{g}_{t'_k t'_k}^{-2}\left(\partial_{t_k t_k t'_k t'_k} K_{\text{norm}}(x'_k,x'_k) + \Gamma^{u'_k}{}^2_{t'_k t'_k}\partial_{u_k u'_k} K_{\text{norm}}(x'_k,x'_k) - 2\Gamma^{u'_k}{}_{t'_k t'_k}\partial_{t_k t_k u'_k} K_{\text{norm}}(x'_k,x'_k)\right)\,,\\
&= 1
\end{aligned}$$

where $\partial_{t_k} K_{\text{norm}}$ (resp. $\partial_{t'_k} K_{\text{norm}}$) denotes a derivative w.r.t. the first (resp. the second) variable of the kernel. We can calculate this quantity, which is constant equal to 1 (see [Giard, 2025, Section IV]). In a similar way, for the second form (67) we have

$$\begin{aligned}
m_{i2i}^2, m_{2ii}^2 &\leq \mathfrak{g}_{t'_k t'_k}^{-1}\mathfrak{g}_{u'_k u'_k}^{-1}\left\|\partial_{t'_k u'_k}\Psi\delta_{x'_k} - \Gamma^{u'_k}{}_{t'_k u'_k}\partial_{t'_k}\Psi\delta_{x'_k}\right\|_{\mathbb{L}}^2\,,\\
&= \mathfrak{g}_{t'_k t'_k}^{-1}\mathfrak{g}_{u'_k u'_k}^{-1}\left(\partial_{t_k u_k t'_k u'_k} K_{\text{norm}}(x'_k,x'_k) + \Gamma^{t'_k}{}^2_{t'_k u'_k}\partial_{t_k t'_k} K_{\text{norm}}(x'_k,x'_k) - 2\Gamma^{t'_k}{}_{t'_k u'_k}\partial_{t_k u_k t'_k} K_{\text{norm}}(x'_k,x'_k)\right)\,,\\
&= 3\,.
\end{aligned}$$

For the third one (68),

$$\begin{aligned}
m_{2i2i}^2 &\leq \mathfrak{g}_{u'_k u'_k}^{-2}\left(\partial_{u_k u_k u'_k u'_k} K_{\text{norm}}(x'_k,x'_k) + \Gamma^{u'_k}{}^2_{u'_k u'_k}\partial_{u_k u'_k} K_{\text{norm}}(x'_k,x'_k) - 2\Gamma^{u'_k}{}_{u'_k u'_k}\partial_{u_k u_k u'_k} K_{\text{norm}}(x'_k,x'_k)\right)\,,\\
&= 7\,.
\end{aligned}$$

Hence $\|M\|_2 \leq \sqrt{4d^2 - 4d + d + 2\times 3d + 7d} = \sqrt{4d^2 + 10d}$.

$\underline{B_{12}}$: Lemma J.1 gives

$$\left\|K_{\text{norm}}^{(12)}(x,x')\right\|_{x,x'} \leq \sqrt{2d}\max_{t_k,u_k,\,k=1,\dots,d}\left\{\left\|\mathfrak{g}_{t_k t_k}^{-1/2}\mathfrak{g}_{x'}^{-1/2}\partial_{t_k}H_2^{\mathfrak{g}}K_{\text{norm}}(x,x')\mathfrak{g}_{x'}^{-1/2}\right\|_2, \left\|\mathfrak{g}_{u_k u_k}^{-1/2}\mathfrak{g}_{x'}^{-1/2}\partial_{u_k}H_2^{\mathfrak{g}}K_{\text{norm}}(x,x')\mathfrak{g}_{x'}^{-1/2}\right\|_2\right\}\,.$$

For any $k \in \{1,\dots,d\}$, $b_k = t_k$ or $b_k = u_k$, the coefficients of the matrix $M_k = \mathfrak{g}_{b_k b_k}^{-1/2}\mathfrak{g}_{x'}^{-1/2}\partial_{b_k}H_2^{\mathfrak{g}}K_{\text{norm}}(x,x')\mathfrak{g}_{x'}^{-1/2}$ are of the form

$$\begin{aligned}
&\mathfrak{g}_{b_m b_m}^{-1/2}\mathfrak{g}_{b'_l b'_l}^{-1/2}\mathfrak{g}_{b'_k b'_k}^{-1/2}\partial_{b_m b'_l b'_k} K_{\text{norm}}(x,x') \quad \text{where } k \neq l\\
\text{or}\quad &\mathfrak{g}_{b_m b_m}^{-1/2}\mathfrak{g}_{t'_k t'_k}^{-1}\left(\partial_{b_m t'_k t'_k} K_{\text{norm}}(x,x') - \Gamma^{u'_k}{}_{t'_k t'_k}\partial_{b_m u'_k} K_{\text{norm}}(x,x')\right)\\
\text{or}\quad &\mathfrak{g}_{b_m b_m}^{-1/2}\mathfrak{g}_{t'_k t'_k}^{-1/2}\mathfrak{g}_{u'_k u'_k}^{-1/2}\left(\partial_{b_m t'_k u'_k} K_{\text{norm}}(x,x') - \Gamma^{t'_k}{}_{t'_k u'_k}\partial_{b_m t'_k} K_{\text{norm}}(x,x')\right)\\
\text{or}\quad &\mathfrak{g}_{b_m b_m}^{-1/2}\mathfrak{g}_{u'_k u'_k}^{-1}\left(\partial_{b_m u'_k u'_k} K_{\text{norm}}(x,x') - \Gamma^{u'_k}{}_{u'_k u'_k}\partial_{b_m u'_k} K_{\text{norm}}(x,x')\right)\,.
\end{aligned}$$

These forms correspond respectively to $(2d)^2 - 4d$, d, $2d$, d coefficients. We can bound them with the same arguments as before. We obtain $\|M_k\|_2 \leq B_{02}$, so $\left\|K_{\text{norm}}^{(12)}(x,x')\right\|_{x,x'} \leq \sqrt{2d}B_{02}$. □

Controls when $\mathfrak{d}(x,x')$ is small

Lemma J.3 (Curvature constants in dimension $d = 1$)**.** *Let $r > 0$. Let $x, x' \in \mathbb{R}\times[u_{\min},+\infty)$. If $\mathfrak{d}(x,x') \geq r$, then $K_{\text{norm}}(x,x') \leq 1 - \bar{\varepsilon}_0(r)$ where $\bar{\varepsilon}_0(r) \leq 1 - e^{-r^2/2}$. Moreover, if $\mathfrak{d}(x,x') \leq r$ where $r \leq 0.32$, then*

$$-v^T H_2^{\mathfrak{g}} K_{\text{norm}}(x,x')v \geq \bar{\varepsilon}_2(r)\|v\|_{x'}^2 \quad \forall v \in \mathbb{R}^2$$

where $\bar{\varepsilon}_2(r) \leq e^{-r^2/2}|G(r)|$ ($G(r)$ defined by (76) below).

Proof. Obtaining $\bar{\varepsilon}_0$: By definition of the semi-distance (see (21)), $\textit{d}(x, x') \geq r$ implies that $K_{\text{norm}}(x, x') \leq e^{-r^2/2}$. Obtaining $\bar{\varepsilon}_2$, general overview: We use that

$$-v^T H_2^{\mathfrak{g}} K_{\text{norm}}(x, x')v \geq \bar{\varepsilon}_2 \|v\|_{x'}^2 \quad \forall v \in \mathbb{R}^2 \qquad \Longleftrightarrow \qquad -v^T \mathfrak{g}_{x'}^{-1/2} H_2^{\mathfrak{g}} K_{\text{norm}}(x, x')\mathfrak{g}_{x'}^{-1/2} v \geq \bar{\varepsilon}_2 \|v\|_2^2 \quad \forall v \in \mathbb{R}^2 .$$

Defining $\tilde{H}^{02}(x, x') \coloneqq K_{\text{norm}}(x, x')^{-1} \mathfrak{g}_{x'}^{-1/2} H_2^{\mathfrak{g}} K_{\text{norm}}(x, x')\mathfrak{g}_{x'}^{-1/2}$, we have (see [Giard, 2025, Section V.1])

$$\tilde{H}^{02}(x, x') = \begin{pmatrix} \tilde{H}^{02}_{t't'}(x, x') & \tilde{H}^{02}_{t'u'}(x, x') \\ \tilde{H}^{02}_{t'u'}(x, x') & \tilde{H}^{02}_{u'u'}(x, x') \end{pmatrix} \quad \text{where}$$

$$\begin{aligned}
\tilde{H}^{02}_{t't'}(x, x') &= -1 , \\
\tilde{H}^{02}_{t'u'}(x, x') &= \frac{(t - t')^3(2u'^2 + \tau^2)^{3/2}}{\sqrt{2}(u^2 + u'^2 + \tau^2)^3} - \frac{3(t - t')(2u'^2 + \tau^2)^{1/2}(u'^2 - u^2)}{\sqrt{2}(u^2 + u'^2 + \tau^2)^2} , \\
\tilde{H}^{02}_{u'u'}(x, x') &= \frac{(t - t')^4(2u'^2 + \tau^2)^2}{2(u^2 + u'^2 + \tau^2)^4} + \frac{3(t - t')^2(2u'^2 + \tau^2)(u^2 - u'^2)}{(u^2 + u'^2 + \tau^2)^3} + \frac{(u^2 - u'^2)^2}{2(u^2 + u'^2 + \tau^2)^2} - \frac{(2u^2 + \tau^2)(2u'^2 + \tau^2)}{(u^2 + u'^2 + \tau^2)^2} .
\end{aligned} \tag{70}$$

If the maximal eigenvalue λ of $\tilde{H}^{02}(x, x')$ is smaller than some $c < 0$, as $K_{\text{norm}}(x, x') \geq e^{-r^2/2}$,

$$-v^T H_2^{\mathfrak{g}} K_{\text{norm}}(x, x')v \geq -\lambda K_{\text{norm}}(x, x') \|v\|_{\mathbb{L}}^2 \geq -ce^{-r^2/2} \|v\|_{\mathbb{L}}^2$$

and we can take $\bar{\varepsilon}_2(r) = -e^{-r^2/2}c$. It remains to bound λ. We use that

$$\lambda \leq \max\{\tilde{H}^{02}_{t't'}(x, x'), \tilde{H}^{02}_{u'u'}(x, x')\} + |\tilde{H}^{02}_{t'u'}(x, x')| .$$

We will provide control of the three terms in the right-hand side of the previous inequality.
Basic inequalities: Using that $\textit{d}(x, x') \leq r$, we have

$$\frac{|t - t'|}{\sqrt{u^2 + u'^2 + \tau^2}} \leq r \tag{71}$$

and

$$\frac{u^2 + u'^2 + \tau^2}{\sqrt{2u^2 + \tau^2}\sqrt{2u'^2 + \tau^2}} \leq e^{r^2} .$$

Using (35), we get

$$\frac{|u^2 - u'^2|}{u^2 + u'^2 + \tau^2} \leq \sqrt{e^{2r^2} - 1} . \tag{72}$$

From (36) we also have

$$\frac{2u'^2 + \tau^2}{u^2 + u'^2 + \tau^2} \leq e^{r^2} + \sqrt{e^{2r^2} - 1} . \tag{73}$$

Control of $\tilde{H}^{02}_{t'u'}$: Using (71), (72) and (73), we get from (70) that

$$\begin{aligned}
|\tilde{H}^{02}_{t'u'}(x, x')| &\leq r^3 \frac{(2u'^2 + \tau^2)^{3/2}}{\sqrt{2}(u^2 + u'^2 + \tau^2)^{3/2}} + 3r \frac{(2u'^2 + \tau^2)^{1/2}|u^2 - u'^2|}{\sqrt{2}(u^2 + u'^2 + \tau^2)^{3/2}} , \\
&\leq \frac{1}{\sqrt{2}} r^3 (e^{r^2} + \sqrt{e^{2r^2} - 1})^{3/2} + \frac{3}{\sqrt{2}} r \sqrt{e^{2r^2} - 1} \sqrt{e^{r^2} + \sqrt{e^{2r^2} - 1}} .
\end{aligned} \tag{74}$$

Control of $\tilde{H}^{02}_{u'u'}$: As $\frac{\sqrt{2u^2+\tau^2}\sqrt{2u'^2+\tau^2}}{u^2+u'^2+\tau^2} \geq e^{-r^2}$,

$$\begin{aligned}
\tilde{H}^{02}_{u'u'}(x, x') &\leq r^4 \frac{(2u'^2 + \tau^2)^2}{2(u^2 + u'^2 + \tau^2)^2} + 3r^2 \frac{(2u'^2 + \tau^2)|u^2 - u'^2|}{(u^2 + u'^2 + \tau^2)^2} + \frac{(u^2 - u'^2)^2}{2(u^2 + u'^2 + \tau^2)^2} - \frac{(2u^2 + \tau^2)(2u'^2 + \tau^2)}{(u^2 + u'^2 + \tau^2)^2} , \\
&\leq \frac{1}{2} r^4 (e^{r^2} + \sqrt{e^{2r^2} - 1})^2 + 3r^2 \sqrt{e^{2r^2} - 1}(e^{r^2} + \sqrt{e^{2r^2} - 1}) + \frac{1}{2}(e^{2r^2} - 1) - e^{-2r^2}
\end{aligned} \tag{75}$$

where we have used (71), (72) and (73) again.
Conclusion: The previous bound is greater than $\tilde{H}^{02}_{t't'}(x, x') = -1$ (because $-e^{-2r^2} \geq -1$). It comes that

$$\begin{aligned}
\lambda \leq\ & \frac{1}{\sqrt{2}} r^3 (e^{r^2} + \sqrt{e^{2r^2} - 1})^{3/2} + \frac{3}{\sqrt{2}} r \sqrt{e^{2r^2} - 1} \sqrt{e^{r^2} + \sqrt{e^{2r^2} - 1}} \\
&+ \frac{1}{2} r^4 (e^{r^2} + \sqrt{e^{2r^2} - 1})^2 + 3r^2 \sqrt{e^{2r^2} - 1}(e^{r^2} + \sqrt{e^{2r^2} - 1}) + \frac{1}{2}(e^{2r^2} - 1) - e^{-2r^2} =: G(r) .
\end{aligned} \tag{76}$$

The function G is non-decreasing on $\mathbb{R}^+$ as a sum of non-decreasing functions. It is negative for $r \leq 0.32$ (see [Giard, 2025, Section V.2]). Then $\bar{\varepsilon}_2(r)$ can be chosen as $-e^{-r^2/2}G(r) = e^{-r^2/2}|G(r)|$ (or smaller) for $r \in [0, 0.32]$. ☐

Lemma J.4 (Curvature constants in dimension $d \geq 1$)**.** *Let $r \geq 0$. Let $x, x' \in \mathbb{R}^d \times [u_{\min}, +\infty)^d$. If $d(x,x') \geq r$, then $K_{\text{norm}}(x,x') \leq 1 - \bar{\varepsilon}_0(r)$ where $\bar{\varepsilon}_0(r) \leq 1 - e^{-\frac{r^2}{2}}$. Moreover, if $d(x,x') \leq r$ where $r = \frac{r_0}{\sqrt{d}}$ with $0 < r_0 \leq 0.32$, then*

$$-vH_2^{\mathfrak{g}}K_{\text{norm}}(x,x')v \geq \bar{\varepsilon}_2(r)\,\|v\|_{x'}^2 \quad \forall v \in \mathbb{R}^2$$

where $\bar{\varepsilon}_2(r) \leq e^{-\frac{r_0^2}{2d}}|G(r_0)|$ ($G(r_0)$ defined by (76)).

Proof. To get $\bar{\varepsilon}_0(r)$, remark again that $d(x,x') \geq r$ implies that $K_{\text{norm}}(x,x') \leq e^{-r^2/2}$ (see (21)).
Reduction to dimension 1: For $\bar{\varepsilon}_2(r)$, we use the same reasoning as in Lemma J.3.
We denote $\tilde{H}^{02}(x,x') := K_{\text{norm}}(x,x')^{-1}\mathfrak{g}_{x'}^{-1/2}H_2^{\mathfrak{g}}K_{\text{norm}}(x,x')\mathfrak{g}_{x'}^{-1/2}$. Its maximum eigenvalue λ is smaller than $\max_{1\leq i\leq 2d}\{\tilde{H}^{02}_{ii}(x,x') + \sum_{j\neq i}|\tilde{H}^{02}_{ij}(x,x')|\}$. We denote $\tilde{H}^{02}_{ij}(x,x') = \tilde{H}^{02}_{b'_k b'_l}(x,x')$ to specify that this coefficient corresponds to the derivatives w.r.t. b'_k, b'_l where $1 \leq k,l \leq d$ (b can be u or t).

We then remark that we can reduce the problem to dimension 1. First, according to (21), $d(x,x') \leq r$ implies that $d(x_k, x'_k) \leq r$ for all $k = 1, \ldots, d$. Using (65), we get for $l \neq k$

$$\begin{aligned}|\tilde{H}^{02}_{b'_k b'_l}(x,x')| &\leq |\mathfrak{g}^{-1/2}_{b'_k b'_k}\mathfrak{g}^{-1/2}_{b'_l b'_l}K_{\text{norm}}(x_k,x'_k)^{-1}K_{\text{norm}}(x_l,x'_l)^{-1}\partial_{b'_k}K_{\text{norm}}(x_k,x'_k)\partial_{b'_l}K_{\text{norm}}(x_l,x'_l)|\,,\\ &= |\mathfrak{g}^{-1/2}_{b'_k b'_k}K_{\text{norm}}(x_k,x'_k)^{-1}\partial_{b'_k}K_{\text{norm}}(x_k,x'_k)||\mathfrak{g}^{-1/2}_{b'_l b'_l}K_{\text{norm}}(x_l,x'_l)^{-1}\partial_{b'_l}K_{\text{norm}}(x_l,x'_l)|\,. \end{aligned} \tag{77}$$

For $l = k$, using (69) we have

$$\tilde{H}^{02}_{t'_k t'_k}(x,x') \leq \tilde{H}^{02}_{t'_k t'_k}(x_k,x'_k) = -1\,, \quad \tilde{H}^{02}_{u'_k u'_k}(x,x') \leq \tilde{H}^{02}_{u'_k u'_k}(x_k,x'_k)\,, \quad |\tilde{H}^{02}_{t'_k u'_k}(x,x')| \leq |\tilde{H}^{02}_{t'_k u'_k}(x_k,x'_k)|\,. \tag{78}$$

Bounds for $\tilde{H}^{02}_{b'_k b'_k}(x,x')$: We use (78). From (74) and (75), we have

$$\tilde{H}^{02}_{t'_k t'_k}(x,x') \vee \tilde{H}^{02}_{u'_k u'_k}(x,x') \leq \frac{1}{2}r^4(e^{r^2}+\sqrt{e^{2r^2}-1})^2 + 3r^2\sqrt{e^{2r^2}-1}(e^{r^2}+\sqrt{e^{2r^2}-1}) + \frac{1}{2}(e^{r^2}-1) - e^{-2r^2}$$

and

$$|\tilde{H}^{02}_{t'_k u'_k}(x,x')| \leq \frac{1}{\sqrt{2}}r^3(e^{r^2}+\sqrt{e^{2r^2}-1})^{3/2} + \frac{3}{\sqrt{2}}r\sqrt{e^{2r^2}-1}\sqrt{e^{r^2}+\sqrt{e^{2r^2}-1}}\,.$$

Bounds for $\tilde{H}^{02}_{b'_k b'_l}(x,x')$, $l \neq k$: See [Giard, 2025, Section V.3]. We calculate

$$\mathfrak{g}_{x_k}^{-1/2}K_{\text{norm}}(x_k,x'_k)^{-1}\nabla_1 K_{\text{norm}}(x_k,x'_k) = \begin{pmatrix} -\frac{(t_k-t'_k)\sqrt{2u_k^2+\tau^2}}{u_k^2+u_k'^2+\tau^2} \\ \frac{(t_k-t'_k)^2(2u_k^2+\tau^2)}{\sqrt{2}(u_k^2+u_k'^2+\tau^2)^2} + \frac{u_k'^2-u_k^2}{\sqrt{2}(u_k^2+u_k'^2+\tau^2)} \end{pmatrix}.$$

With (71), (72) and (73), it follows using (77) that for $l \neq k$,

$$\begin{aligned}|\tilde{H}^{02}_{t'_k t'_l}(x,x')| &\leq \frac{|t'_k-t_k|\sqrt{2u_k'^2+\tau^2}}{u_k^2+u_k'^2+\tau^2}\frac{|t_l-t'_l|\sqrt{2u_l'^2+\tau^2}}{u_l^2+u_l'^2+\tau^2}\,,\\ &\leq r^2(e^{r^2}+\sqrt{e^{2r^2}-1})\end{aligned}$$

and

$$\begin{aligned}|\tilde{H}^{02}_{u'_k t'_l}(x,x')| &\leq \frac{|t'_l-t_l|\sqrt{2u_l'^2+\tau^2}}{u_l^2+u_l'^2+\tau^2}\left(\frac{(t_k-t'_k)^2(2u_k'^2+\tau^2)}{\sqrt{2}(u_k^2+u_k'^2+\tau^2)^2} + \frac{|u_k^2-u_k'^2|}{\sqrt{2}(u_k^2+u_k'^2+\tau^2)}\right),\\ &\leq r\sqrt{e^{r^2}+\sqrt{e^{2r^2}-1}} \times \frac{1}{\sqrt{2}}\left(r^2(e^{r^2}+\sqrt{e^{2r^2}-1}) + \sqrt{e^{2r^2}-1}\right)\end{aligned}$$

along with

$$|\tilde{H}^{02}_{u'_k u'_l}(x,x')| \leq \frac{1}{2}\left(r^2(e^{r^2}+\sqrt{e^{2r^2}-1}) + \sqrt{e^{2r^2}-1}\right)^2.$$

Conclusion: The bound found for $|\tilde{H}^{02}_{u'_k u'_l}(x,x')|$ is greater than the one found for $|\tilde{H}^{02}_{t'_k t'_l}(x,x')|$. In fact, using that $e^{r^2} \geq 1 + r^2$,

$$\begin{aligned}
\frac{1}{2}\left(r^2(e^{r^2} + \sqrt{e^{2r^2}-1}) + \sqrt{e^{2r^2}-1}\right)^2 &\geq r^2(e^{r^2} + \sqrt{e^{2r^2}-1})\sqrt{e^{2r^2}-1} + \frac{1}{2}(e^{2r^2}-1)\,,\\
&\geq r^2\sqrt{e^{2r^2}-1} + \frac{1}{2}(2r^2+1)(e^{2r^2}-1)\,,\\
&\geq r^2\sqrt{e^{2r^2}-1} + \frac{1}{2}(2r^2+1)(e^{r^2}(1+r^2)-1)\,,\\
&\geq r^2\sqrt{e^{2r^2}-1} + r^2e^{r^2} + \frac{1}{2}(r^2+1)e^{r^2} - \frac{1}{2}(2r^2+1)\,,\\
&\geq r^2\sqrt{e^{2r^2}-1} + r^2e^{r^2}\,.
\end{aligned}$$

Hence

$$\begin{aligned}
\lambda &\leq \frac{1}{2}r^4(e^{r^2} + \sqrt{e^{2r^2}-1})^2 + 3r^2\sqrt{e^{2r^2}-1}(e^{r^2} + \sqrt{e^{2r^2}-1}) + \frac{1}{2}(e^{2r^2}-1) - e^{-2r^2}\\
&\quad + (d-1)\frac{1}{2}\left(r^2(e^{r^2} + \sqrt{e^{2r^2}-1}) + \sqrt{e^{2r^2}-1}\right)^2\\
&\quad + (d-1)r\sqrt{e^{r^2} + \sqrt{e^{2r^2}-1}} \times \frac{1}{\sqrt{2}}\left(r^2(e^{r^2} + \sqrt{e^{2r^2}-1}) + \sqrt{e^{2r^2}-1}\right)\\
&\quad + \frac{1}{\sqrt{2}}r^3(e^{r^2} + \sqrt{e^{2r^2}-1})^{3/2} + \frac{3}{\sqrt{2}}r\sqrt{e^{2r^2}-1}\sqrt{e^{r^2} + \sqrt{e^{2r^2}-1}}\,,\\
&= -e^{-2r^2} + 2r^2\sqrt{e^{2r^2}-1}(e^{r^2} + \sqrt{e^{2r^2}-1}) + \frac{d}{2}\left(r^2(e^{r^2} + \sqrt{e^{2r^2}-1}) + \sqrt{e^{2r^2}-1}\right)^2\\
&\quad + \frac{dr}{\sqrt{2}}\sqrt{e^{r^2} + \sqrt{e^{2r^2}-1}}\left(r^2(e^{r^2} + \sqrt{e^{2r^2}-1}) + \sqrt{e^{2r^2}-1}\right)\\
&\quad + \sqrt{2}r\sqrt{e^{2r^2}-1}\sqrt{e^{r^2} + \sqrt{e^{2r^2}-1}} =: G_d(r)\,.
\end{aligned}$$

Furthermore, $r \in \mathbb{R}^+ \mapsto \sqrt{e^{2r^2}-1}$ is convex. In fact, for $r > 0$, using that $e^{2r^2} \geq 1 + 2r^2$, we have

$$\frac{\partial^2}{\partial r^2}\sqrt{e^{2r^2}-1} = \frac{2e^{2r^2}(2r^2e^{2r^2} - 4r^2 + e^{2r^2} - 1)}{(e^{2r^2}-1)^{3/2}} \geq \frac{2e^{2r^2}(2r^2 - 4r^2 + 1 + 2r^2 - 1)}{(e^{2r^2}-1)^{3/2}} \geq 0\,.$$

Let $r_0 > 0$. As $r := \frac{r_0}{\sqrt{d}} \leq r_0$, $\sqrt{e^{2r^2}-1} \leq \frac{\sqrt{e^{2r_0^2}-1}}{\sqrt{d}}$ and we deduce that $G_d(r) \leq G_1(r_0) = G(r_0)$ (see (76).) If $r_0 \leq 0.32$, this quantity is negative (proof of Lemma J.3). We can take $\bar{\varepsilon}_2\left(\frac{r_0}{\sqrt{d}}\right) = -e^{-\frac{r_0^2}{2d}}G(r_0)$. The choice of the dependence on d for $r = \frac{r_0}{\sqrt{d}}$ is intended to compensate for the term $d(e^{2r^2}-1)$ appearing in $G_d(r)$. □

Controls when $\mathfrak{d}(x,x')$ is large The constraint $\tau \leq u_{\min}$ is used in the following lemma.

Lemma J.5 (Bounds under a large separation, in dimension $d = 1$)**.** *Let $\Delta > 0$. Let $x, x' \in \mathbb{R} \times [u_{\min}, +\infty)$. Assume that $\tau \leq u_{\min}$. If $\mathfrak{d}(x,x') \geq \Delta$, then*

$$\max_{(i,j)\in\{0,1\}\times\{0,1,2\}} \left\|K^{(ij)}_{\text{norm}}(x,x')\right\|_{x,x'} \leq \sqrt{2}\,153.05e^{-\frac{\Delta^2}{4}}\,.$$

Proof. The calculations in this proof are presented in [Giard, 2025, Sections VI.1 and VI.2]. Let $x, x' \in \mathbb{R} \times [u_{\min}, +\infty)$ such that $\mathfrak{d}(x,x') \geq \Delta$. The following bounds on the operator norms are based on the simplified expressions from Lemma J.1. We will use the inequalities

$$\frac{\sqrt{2u^2+\tau^2}\sqrt{2u'^2+\tau^2}}{u'^2+u^2+\tau^2} \vee \frac{|u'^2-u^2|}{u'^2+u^2+\tau^2} \leq 1\,, \quad \frac{2u^2+\tau^2}{u'^2+u^2+\tau^2} \leq 2\,, \quad \frac{2u^2+\tau^2}{u^2} \leq 3\,, \tag{79}$$

which holds since $\tau \leq u_{\min}$. As it holds $y^q e^{-\frac{y^2}{4}} \leq \left(\frac{2q}{e}\right)^{q/2}$ for all $y \geq 0$, $q \geq 1$, we also have that for all $x, x' \in \mathbb{R} \times [u_{\min}, +\infty)$,

$$\begin{aligned}
\forall q \geq 1\,, \quad \frac{|t-t'|^q}{(u'^2+u^2+\tau^2)^{\frac{q}{2}}}\sqrt{K_{\text{norm}}(x,x')} &\leq \left(\frac{2q}{e}\right)^{q/2}\frac{(2u^2+\tau^2)^{1/8}(2u'^2+\tau^2)^{1/8}}{(u'^2+u^2+\tau^2)^{\frac{1}{4}}}\,,\\
&\leq \left(\frac{2q}{e}\right)^{q/2}\,.
\end{aligned} \tag{80}$$

We will also use that $\sqrt{K_{\text{norm}}(x,x')} \le e^{-\Delta^2/4}$ for $d(x,x') \ge \Delta$ (see (21)).

In what follows, we bound the 2-norm of a matrix $M = (m_{ij})_{ij}$ by its Frobenius norm $\sqrt{\sum_{i,j} m_{ij}^2}$.

$\underline{\left\| K^{(00)}_{\text{norm}}(x,x') \right\|}$: We have $|K_{\text{norm}}(x,x')| \le e^{-\Delta^2/2}$, which is an immediate consequence of (21).

$\underline{\left\| K^{(10)}_{\text{norm}}(x,x') \right\|_x}$: First,

$$\begin{pmatrix} H^{10}_t(x,x') \\ H^{10}_u(x,x') \end{pmatrix} = H^{10}(x,x') \coloneqq \sqrt{K_{\text{norm}}(x,x')}^{-1} \mathfrak{g}_x^{-1/2} \nabla_1 K_{\text{norm}}(x,x') ,$$
$$= \sqrt{K_{\text{norm}}(x,x')} \begin{pmatrix} \frac{(t-t')\sqrt{2u^2+\tau^2}}{u^2+u'^2+\tau^2} \\ \frac{(t-t')^2(2u^2+\tau^2)}{\sqrt{2}(u^2+u'^2+\tau^2)^2} + \frac{u'^2-u^2}{\sqrt{2}(u^2+u'^2+\tau^2)} \end{pmatrix} .$$

Using (80) and (79) along with $\sqrt{K_{\text{norm}}(x,x')} \le 1$, it comes

$$\begin{aligned} &|H^{10}_t(x,x')| \le \frac{2}{\sqrt{e}} \\ \text{and} \quad &|H^{10}_u(x,x')| \le \left(\frac{4\sqrt{2}}{e} + \frac{1}{\sqrt{2}} \right) , \end{aligned} \tag{81}$$

hence $\left\| H^{10}(x,x') \right\|_2 \le \sqrt{\frac{4}{e} + \left(\frac{4\sqrt{2}}{e} + \frac{1}{\sqrt{2}} \right)^2}$ and

$$\left\| K^{(10)}_{\text{norm}}(x,x') \right\|_x = \sqrt{K_{\text{norm}}(x,x')} \left\| H^{10}(x,x') \right\|_2 \le e^{-\Delta^2/4} \sqrt{\frac{4}{e} + \left(\frac{4\sqrt{2}}{e} + \frac{1}{\sqrt{2}} \right)^2} .$$

$\underline{\left\| K^{(11)}_{\text{norm}}(x,x') \right\|_{x,x'}}$: We have

$$H^{11}(x,x') \coloneqq \sqrt{K_{\text{norm}}(x,x')}^{-1} \mathfrak{g}_x^{-1/2} \nabla_1 \nabla_2 K_{\text{norm}}(x,x') \mathfrak{g}_{x'}^{-1/2} = \sqrt{K_{\text{norm}}(x,x')} \begin{pmatrix} \tilde{H}^{11}_{tt'}(x,x') & \tilde{H}^{11}_{tu'}(x,x') \\ \tilde{H}^{11}_{tu'}(x,x') & \tilde{H}^{11}_{uu'}(x,x') \end{pmatrix}$$

where

$$\begin{aligned} \tilde{H}^{11}_{tt'}(x,x') &= \frac{\sqrt{2u^2+\tau^2}\sqrt{2u'^2+\tau^2}}{u^2+u'^2+\tau^2} - \frac{(t-t')^2\sqrt{2u^2+\tau^2}\sqrt{2u'^2+\tau^2}}{(u^2+u'^2+\tau^2)^2} , \\ \tilde{H}^{11}_{tu'}(x,x') &= \frac{-(t-t')^3 u'(2u^2+\tau^2)\sqrt{2u'^2+\tau^2}}{\sqrt{2}u(u^2+u'^2+\tau^2)^3} + \frac{\sqrt{2}(t-t')u'(2u^2+\tau^2)\sqrt{2u'^2+\tau^2}}{u(u^2+u'^2+\tau^2)^2} \\ &\quad + \frac{(t-t')u'(u'^2-u^2)(2u^2+\tau^2)}{\sqrt{2}u\sqrt{2u'^2+\tau^2}(u^2+u'^2+\tau^2)^2} , \\ \tilde{H}^{11}_{tu'}(x,x') &= \tilde{H}^{11}_{tu'}(x',x) , \\ \tilde{H}^{11}_{uu'}(x,x') &= \frac{(t-t')^4(2u^2+\tau^2)(2u'^2+\tau^2)}{2(u^2+u'^2+\tau^2)^4} + \frac{-2(t-t')^2(2u^2+\tau^2)(2u'^2+\tau^2)}{(u^2+u'^2+\tau^2)^3} + \frac{(t-t')^2(u'^2-u^2)^2}{(u^2+u'^2+\tau^2)^3} \\ &\quad + \frac{(2u^2+\tau^2)(2u'^2+\tau^2)}{(u^2+u'^2+\tau^2)^2} - \frac{(u^2-u'^2)^2}{2(u^2+u'^2+\tau^2)^2} . \end{aligned}$$

The constraint $\tau \le u_{\min}$ is used here, to control $\frac{\sqrt{2u^2+\tau^2}}{u}$ in $\tilde{H}^{11}_{tu'}, \tilde{H}^{11}_{tu'}$. Using again (80), (79) and the normalization $\sqrt{K_{\text{norm}}(x,x')} \le 1$, it comes

$$\begin{aligned} |H^{11}_{tt'}(x,x')| &\le \left(1 + \frac{4}{e} \right) , \\ |H^{11}_{ut'}(x,x')| \vee |H^{11}_{tu'}(x,x')| &\le \left(\frac{\sqrt{3}}{\sqrt{2}} \left(\frac{6}{e} \right)^{3/2} + \frac{2\sqrt{3}}{\sqrt{e}} + \frac{\sqrt{3}}{\sqrt{e}} \right) , \\ |H^{11}_{uu'}(x,x')| &\le \left(\frac{1}{2} \left(\frac{8}{e} \right)^2 + \frac{8}{e} + \frac{4}{e} + 1 + \frac{1}{2} \right) . \end{aligned} \tag{82}$$

Hence $\left\|H^{(11)}(x,x')\right\|_2 \le \sqrt{\left(1+\frac{4}{e}\right)^2 + 2\left(\frac{\sqrt{3}}{\sqrt{2}}\left(\frac{6}{e}\right)^{3/2} + \frac{3\sqrt{3}}{\sqrt{e}}\right)^2 + \left(\frac{32}{e^2}+\frac{12}{e}+\frac{3}{2}\right)^2}$ and

$$\left\|K^{(11)}_{\text{norm}}(x,x')\right\|_{x,x'} \le e^{-\Delta^2/4}\sqrt{\left(1+\frac{4}{e}\right)^2 + 2\left(\frac{\sqrt{3}}{\sqrt{2}}\left(\frac{6}{e}\right)^{3/2} + \frac{3\sqrt{3}}{\sqrt{e}}\right)^2 + \left(\frac{32}{e^2}+\frac{12}{e}+\frac{3}{2}\right)^2}\,.$$

$\left\|K^{(02)}_{\text{norm}}(x,x')\right\|_{x'}$: We use (70) to get the expression of

$$H^{02}(x,x') \coloneqq \sqrt{K_{\text{norm}}(x,x')}\tilde{H}^{02}(x,x') = \sqrt{K_{\text{norm}}(x,x')}^{-1}\mathfrak{g}_{x'}^{-1/2}H_2^{\mathfrak{g}}K_{\text{norm}}(x,x')\mathfrak{g}_{x'}^{-1/2}\,.$$

Using the same techniques as before, we find $\left\|H^{02}(x,x')\right\|_2 \le \sqrt{1+2\left(3\frac{\sqrt{2}}{\sqrt{e}}+2\left(\frac{6}{e}\right)^{3/2}\right)^2 + \left(\frac{128}{e^2}+\frac{24}{e}+\frac{3}{2}\right)^2}$ so

$$\left\|K^{(02)}_{\text{norm}}(x,x')\right\|_{x'} \le e^{-\Delta^2/4}\sqrt{1+2\left(3\frac{\sqrt{2}}{\sqrt{e}}+2\left(\frac{6}{e}\right)^{3/2}\right)^2 + \left(\frac{128}{e^2}+\frac{24}{e}+\frac{3}{2}\right)^2}\,.$$

$\left\|K^{(12)}_{\text{norm}}(x,x')\right\|_{x,x'}$: We denote

$$H^{12,1}(x,x') = \sqrt{K_{\text{norm}}(x,x')}^{-1}\mathfrak{g}_{tt}^{-1/2}\mathfrak{g}_{x'}^{-1/2}\partial_t H_2^{\mathfrak{g}}K_{\text{norm}}(x,x')\mathfrak{g}_{x'}^{-1/2}$$

and

$$H^{12,2}(x,x') = \sqrt{K_{\text{norm}}(x,x')}^{-1}\mathfrak{g}_{uu}^{-1/2}\mathfrak{g}_{x'}^{-1/2}\partial_u H_2^{\mathfrak{g}}K_{\text{norm}}(x,x')\mathfrak{g}_{x'}^{-1/2}\,.$$

We find $\tilde{H}^{12,1} = \sqrt{K_{\text{norm}}(x,x')}\begin{pmatrix}\tilde{H}^{12,1}_{t't'}(x,x') & \tilde{H}^{12,1}_{t'u'}(x,x') \\ \tilde{H}^{12,1}_{t'u'}(x,x') & \tilde{H}^{12,1}_{u'u'}(x,x')\end{pmatrix}$ with

$$\begin{aligned}
\tilde{H}^{12,1}_{t't'}(x,x') &= \frac{(t-t')\sqrt{2u^2+\tau^2}}{u^2+u'^2+\tau^2}\,,\\
\tilde{H}^{12,1}_{t'u'}(x,x') &= \frac{3(t-t')^2\sqrt{2u^2+\tau^2}(2u'^2+\tau^2)^{3/2}}{\sqrt{2}(u^2+u'^2+\tau^2)^3} - \frac{(t-t')^4\sqrt{2u^2+\tau^2}(2u'^2+\tau^2)^{3/2}}{\sqrt{2}(u^2+u'^2+\tau^2)^4}\\
&\quad + \frac{3(t-t')^2\sqrt{2u^2+\tau^2}\sqrt{2u'^2+\tau^2}(u'^2-u^2)}{\sqrt{2}(u^2+u'^2+\tau^2)^3} + \frac{3\sqrt{2u^2+\tau^2}\sqrt{2u'^2+\tau^2}(u^2-u'^2)}{\sqrt{2}(u^2+u'^2+\tau^2)^2}
\end{aligned}$$

and

$$\begin{aligned}
\tilde{H}^{12,1}_{u'u'}(x,x') &= \frac{-(t-t')^5(2u'^2+\tau^2)^2\sqrt{2u^2+\tau^2}}{2(u^2+u'^2+\tau^2)^5} + \frac{2(t-t')^3(2u'^2+\tau^2)^2\sqrt{2u^2+\tau^2}}{(u^2+u'^2+\tau^2)^4}\\
&\quad + \frac{3(t-t')^3(u'^2-u^2)(2u'^2+\tau^2)\sqrt{2u^2+\tau^2}}{(u^2+u'^2+\tau^2)^4} - \frac{(t-t')\sqrt{2u^2+\tau^2}(u^2-u'^2)^2}{2(u^2+u'^2+\tau^2)^3}\\
&\quad + \frac{(t-t')(2u'^2+\tau^2)(2u^2+\tau^2)^{3/2}}{(u^2+u'^2+\tau^2)^3} + \frac{6(t-t')(2u'^2+\tau^2)(u^2-u'^2)\sqrt{2u^2+\tau^2}}{(u^2+u'^2+\tau^2)^3}\,.
\end{aligned}$$

Using the same ideas as before to bound the coefficients, we get

$$\left\|H^{12,1}(x,x')\right\|_2 \le \sqrt{\frac{4}{e}+2\left(\frac{18\sqrt{2}}{e}+\sqrt{2}\frac{64}{e^2}+\frac{3}{\sqrt{2}}\right)^2 + \left(\sqrt{2}\left(\frac{10}{e}\right)^{5/2}+7\sqrt{2}\left(\frac{6}{e}\right)^{3/2}+\frac{15}{\sqrt{e}}\right)^2}\,.$$

We also have $\tilde{H}^{12,2} = \sqrt{K_{\text{norm}}(x,x')} \begin{pmatrix} \tilde{H}^{12,1}_{t't'}(x,x') & \tilde{H}^{12,1}_{t'u'}(x,x') \\ \tilde{H}^{12,1}_{t'u'}(x,x') & \tilde{H}^{12,1}_{u'u'}(x,x') \end{pmatrix}$ with

$$\begin{aligned}
\tilde{H}^{12,1}_{t't'}(x,x') &= \frac{-(t-t')^2(2u^2+\tau^2)}{\sqrt{2}(u^2+u'^2+\tau^2)^2} + \frac{u^2-u'^2}{\sqrt{2}(u^2+u'^2+\tau^2)}\,, \\
\tilde{H}^{12,1}_{t'u'}(x,x') &= \frac{(t-t')^5(2u'^2+\tau^2)^{3/2}(2u^2+\tau^2)}{2(u^2+u'^2+\tau^2)^5} + \frac{(t-t')^3(2u'^2+\tau^2)^{1/2}(2u^2+\tau^2)(u^2-u'^2)}{(u^2+u'^2+\tau^2)^4} \\
&\quad - \frac{4(t-t')^3(2u'^2+\tau^2)^{3/2}(2u^2+\tau^2)}{(u^2+u'^2+\tau^2)^4} + \frac{(t-t')^3(2u'^2+\tau^2)^{1/2}}{(u^2+u'^2+\tau^2)^2} \\
&\quad + \frac{3(t-t')(2u'^2+\tau^2)^{1/2}(2u^2+\tau^2)^2}{(u^2+u'^2+\tau^2)^3} + \frac{21(t-t')(2u'^2+\tau^2)^{1/2}(2u^2+\tau^2)(u'^2-u^2)}{2(u^2+u'^2+\tau^2)^3} \\
&\quad + \frac{3(t-t')(2u'^2+\tau^2)^{1/2}(u^2-u'^2)}{2(u^2+u'^2+\tau^2)^2}
\end{aligned}$$

and

$$\begin{aligned}
\tilde{H}^{12,1}_{u'u'}(x,x') =& \frac{\sqrt{2}(t-t')^6(2u'^2+\tau^2)^2(2u^2+\tau^2)}{4(u^2+u'^2+\tau^2)^6} + \frac{5\sqrt{2}(t-t')^4(2u'^2+\tau^2)(2u^2+\tau^2)(u^2-u'^2)}{4(u^2+u'^2+\tau^2)^5} \\
&- \frac{2\sqrt{2}(t-t')^4(2u'^2+\tau^2)^2(2u^2+\tau^2)}{(u^2+u'^2+\tau^2)^5} + \frac{\sqrt{2}(t-t')^4(2u'^2+\tau^2)(u^2-u'^2)^2}{2(u^2+u'^2+\tau^2)^5} \\
&- \frac{5\sqrt{2}(t-t')^2(2u'^2+\tau^2)(u^2-u'^2)^2}{4(u^2+u'^2+\tau^2)^4} + \frac{5\sqrt{2}(t-t')^2(2u^2+\tau^2)(2u'^2+\tau^2)^2}{2(u^2+u'^2+\tau^2)^4} \\
&+ \frac{\sqrt{2}(t-t')^2(u^2-u'^2)^3}{2(u^2+u'^2+\tau^2)^4} - \frac{7\sqrt{2}(t-t')^2(2u^2+\tau^2)(2u'^2+\tau^2)(u^2-u'^2)}{(u^2+u'^2+\tau^2)^4} \\
&+ \frac{7\sqrt{2}(2u'^2+\tau^2)(2u^2+\tau^2)(u^2-u'^2)}{2(u^2+u'^2+\tau^2)^3} - \frac{\sqrt{2}(u^2-u'^2)^3}{4(u^2+u'^2+\tau^2)^3}\,.
\end{aligned}$$

We get

$$\left\|H^{12,2}(x,x')\right\|_2 \le \sqrt{\begin{aligned}&\left(\frac{4\sqrt{2}}{e}+1\right)^2 + 2\left(\frac{1}{\sqrt{2}}\left(\frac{10}{e}\right)^{5/2} + 6\sqrt{2}\left(\frac{6}{e}\right)^{3/2} + \frac{36}{\sqrt{e}}\right)^2 \\ &+ \left(\frac{432\sqrt{2}}{e^3} + \frac{272\sqrt{2}}{e^2} + \frac{60\sqrt{2}}{e} + \frac{15\sqrt{2}}{4}\right)^2\end{aligned}} \le 153.05\,.$$

As this bound is greater than the one found for $\left\|H^{12,1}(x,x')\right\|_2$, we deduce that

$$\left\|K^{(12)}_{\text{norm}}(x,x')\right\|_{x,x'} \le \sqrt{2}\,153.05 e^{-\Delta^2/4}\,.$$

This bound is the largest we obtained for $\left\|K^{(ij)}_{\text{norm}}(x,x')\right\|_{x,x'}$, hence it is an upper bound for all the derivatives investigated by this lemma. □

Lemma J.6 (Bounds under a large separation, in dimension $d \ge 1$)**.** *Let $\Delta > 0$. Let $x, x' \in \mathbb{R}^d \times [u_{\min}, +\infty)^d$. Assume that $\tau \le u_{\min}$. If $\mathfrak{d}(x,x') \ge \Delta$, then*

$$\max_{(i,j)\in\{0,1\}\times\{0,1,2\}} \left\|K^{(ij)}_{\text{norm}}(x,x')\right\|_{x,x'} \le \sqrt{2d}(170.5 + 25.78d)e^{-\Delta^2/4}\,.$$

Proof. Let $x, x' \in \mathbb{R}^d \times [u_{\min}, +\infty)^d$ such that $\mathfrak{d}(x,x') \ge \Delta$. We use the decomposition of the kernel $K_{\text{norm}}(x,x') = \prod_{k=1}^d K_{\text{norm}}(x_k, x'_k)$ to reuse the results proven in dimension 1 (see Lemma J.5). As in Lemma J.4, we use the expressions for the operator norms given in Lemma J.1 and we bound the 2-norm of a matrix $M = (m_{ij})_{ij}$ by $\sqrt{\sum_{i,j} m_{ij}^2}$. The calculations in this proof are presented in [Giard, 2025, Section VI.3].

$\underline{\left\|K^{(00)}_{\text{norm}}(x,x')\right\|}$: According to (21), we still have $|K^{(00)}_{\text{norm}}(x,x')| \le e^{-\Delta^2/2}$.

$\underline{\left\|K^{(10)}_{\text{norm}}(x,x')\right\|_x}$: We recall that the notation $\partial_{b_k} K_{\text{norm}}$ (resp. $\partial_{b'_k} K_{\text{norm}}$) refers to a derivative w.r.t. to the first (resp. the second) variable of the kernel, where b_k is t_k or u_k. The decomposition of the kernel gives $\partial_{b_k} K_{\text{norm}}(x,x') = \partial_{b_k} K_{\text{norm}}(x_k,x'_k)\prod_{l\neq k}^d K_{\text{norm}}(x_l,x'_l)$. Using (80) for $K_{\text{norm}}(x_k,x'_k)$, we get that

$$\forall\, q\geq 1\,, \qquad \frac{|t_k-t'_k|^q}{(u'^2_k+u_k^2+\tau^2)^{\frac{q}{2}}}\sqrt{K_{\text{norm}}(x_k,x'_k)} \leq \left(\frac{2q}{e}\right)^{q/2}\,.$$

We have

$$|\mathfrak{g}_{b_kb_k}^{-1/2}\partial_{b_k}K_{\text{norm}}(x,x')| \leq \sqrt{\prod_{l=1}^d K_{\text{norm}}(x_l,x'_l)}\left|\sqrt{K_{\text{norm}}(x_k,x'_k)}^{-1}\mathfrak{g}_{b_kb_k}^{-1/2}\partial_{b_k}K_{\text{norm}}(x_k,x'_k)\right|\,.$$

The term in the right-hand side has already been dealt with in Lemma J.5: (81) gives a bound for

$$\left|\sqrt{K_{\text{norm}}(x_k,x'_k)}^{-1}\mathfrak{g}_{b_kb_k}^{-1/2}\partial_{b_k}K_{\text{norm}}(x_k,x'_k)\right| = |H^{10}_{b_k}(x_k,x'_k)|\,,$$

that we denote by $B_{H^{10}_{b_k}}$ (note that this bound does not depend on Δ). So

$$|\mathfrak{g}_{b_kb_k}^{-1/2}\partial_{b_k}K_{\text{norm}}(x,x')| \leq e^{-\Delta^2/4}B_{H^{10}_{b_k}}\,.$$

Hence $\left\|K^{(10)}_{\text{norm}}(x,x')\right\|_x \leq \sqrt{d}\sqrt{B^2_{H^{10}_{t_k}}+B^2_{H^{10}_{u_k}}}e^{-\Delta^2/4} \leq 3.05\sqrt{d}e^{-\Delta^2/4}$.

$\left\|K^{(11)}_{\text{norm}}(x,x')\right\|_{x,x'}$: We use in the same way as for $B_{H^{(10)}_{b_k}}$ the notation $B_{H^{(11)}_{b_k,b'_k}}$, denoting the bound obtained in Lemma J.5 for $\mathfrak{g}_{x_k}^{-1/2}\partial_{b_k,b'_k}K_{\text{norm}}(x_k,x'_k)\mathfrak{g}_{x'_k}^{-1/2}\sqrt{K_{\text{norm}}(x_k,x'_k)}^{-1} = |H^{11}_{b_kb'_k}(x_k,x'_k)|$ (see (82)). For $k\neq l$,

$$\begin{aligned}
|\mathfrak{g}_{b_kb_k}^{-1/2}\partial_{b_kb'_l}K_{\text{norm}}(x_k,x'_k)\mathfrak{g}_{b'_lb'_l}^{-1/2}| &\leq \sqrt{\prod_{m=1}^d K_{\text{norm}}(x_m,x'_m)}\left|\sqrt{K_{\text{norm}}(x_k,x'_k)}^{-1}\mathfrak{g}_{b_kb_k}^{-1/2}\partial_{b_k}K_{\text{norm}}(x_k,x'_k)\right| \\
&\quad\times\left|\sqrt{K_{\text{norm}}(x_l,x'_l)}^{-1}\mathfrak{g}_{b'_lb'_l}^{-1/2}\partial_{b'_l}K_{\text{norm}}(x_l,x'_l)\right|\,, \\
&\leq e^{-\Delta^2/4}B_{H^{10}_{b_k}}B_{H^{01}_{b'_l}} = e^{-\Delta^2/4}B_{H^{10}_{b_k}}B_{H^{10}_{b_l}}\,.
\end{aligned}$$

For $k=l$, we have in the same way

$$\begin{aligned}
|\mathfrak{g}_{t_kt_k}^{-1/2}\partial_{t_kt'_k}K_{\text{norm}}(x_k,x'_k)\mathfrak{g}_{t'_kt'_k}^{-1/2}| &\leq e^{-\Delta^2/4}B_{H^{11}_{t_k,t'_k}}\,, \\
|\mathfrak{g}_{u_ku_k}^{-1/2}\partial_{u_kt'_k}K_{\text{norm}}(x_k,x'_k)\mathfrak{g}_{t'_kt'_k}^{-1/2}| &\leq e^{-\Delta^2/4}B_{H^{11}_{u_k,t'_k}}\,, \\
|\mathfrak{g}_{u_ku_k}^{-1/2}\partial_{u_ku'_k}K_{\text{norm}}(x_k,x'_k)\mathfrak{g}_{u'_ku'_k}^{-1/2}| &\leq e^{-\Delta^2/4}B_{H^{11}_{u_k,u'_k}}\,.
\end{aligned}$$

So

$$\begin{aligned}
\left\|K^{(11)}_{\text{norm}}(x,x')\right\|_{x,x'} &\leq e^{-\Delta^2/4}\sqrt{\begin{aligned}&d(d-1)B^4_{H^{10}_{t_k}}+d(d-1)B^4_{H^{10}_{u_k}}+2d(d-1)B^2_{H^{10}_{t_k}}B^2_{H^{10}_{u_k}}\\ &\quad+2dB^2_{H^{11}_{t_k,u'_k}}+dB^2_{H^{11}_{t_k,t'_k}}+dB^2_{H^{11}_{u_k,u'_k}}\end{aligned}}\,, \\
&\leq (9.25d+6.95)e^{-\Delta^2/4}\,.
\end{aligned}$$

$\left\|K^{(02)}_{\text{norm}}(x,x')\right\|_{x'}$: With the same arguments, the terms of $\mathfrak{g}_{x'}^{-1/2}H_2^{\mathfrak{g}}K_{\text{norm}}(x,x')\mathfrak{g}_{x'}^{-1/2}$ corresponding to derivatives taken in b'_k,b'_l with $k\neq l$ can be bounded by $e^{-\Delta^2/4}B_{H^{10}_{b_k}}B_{H^{10}_{b'_l}}$. We bound the terms corresponding to $k=l$ with $e^{-\Delta^2/4}B_{H^{02}_{t'_kt'_k}}$ or $e^{-\Delta^2/4}B_{H^{02}_{t'_ku'_k}}$ or $e^{-\Delta^2/4}B_{H^{02}_{u'_ku'_k}}$, where $B_{H^{02}_{b'_kb'_k}}$ denotes again the bound obtained in Lemma J.5. We get

$$\begin{aligned}
\left\|K^{(02)}_{\text{norm}}(x,x')\right\|_{x'} &\leq e^{-\Delta^2/4}\sqrt{\begin{aligned}&d(d-1)B^4_{H^{10}_{t_k}}+d(d-1)B^4_{H^{10}_{u_k}}+2d(d-1)B^2_{H^{10}_{t_k}}B^2_{H^{10}_{u_k}}\\ &\quad+2dB^2_{H^{02}_{t'_ku'_k}}+dB^2_{H^{02}_{t'_kt'_k}}+dB^2_{H^{02}_{u'_ku'_k}}\end{aligned}}\,, \\
&\leq (9.25d+45.9)e^{-\Delta^2/4}\,.
\end{aligned}$$

$\left\|K^{(12)}_{\mathrm{norm}}(x,x')\right\|_{x,x'}$: We denote $M_{b_k} = \mathfrak{g}_{b_k b_k}^{-1/2}\mathfrak{g}_{x'}^{-1/2}\partial_{b_k} H_2^{\mathfrak{g}} K_{\mathrm{norm}}(x,x')\mathfrak{g}_{x'}^{-1/2}$. The terms at positions b'_l, b'_m where $l \neq m$ and $l, m \neq k$ can be bounded by $e^{-\Delta^2/4} B_{H^{10}_{b_k}} B_{H^{10}_{b_l}} B_{H^{10}_{b_m}}$. The terms at positions b'_l, b'_k or b'_k, b'_l where $l \neq k$ can be bounded by $e^{-\Delta^2/4} B_{H^{11}_{b_k,t'_k}} B_{H^{10}_{b_l}}$ or $e^{-\Delta^2/4} B_{H^{11}_{b_k,u'_k}} B_{H^{10}_{b_l}}$. The terms at positions b'_l, b'_m where $l = m \neq k$ can be bounded by $e^{-\Delta^2/4} B_{H^{10}_{b_k}} B_{H^{02}_{b'_l b'_m}}$. The terms at positions b'_l, b'_m where $l = m = k$ can be bounded by $e^{-\Delta^2/4} B_{H^{12,b_k}_{t'_k,t'_k}}$ or $e^{-\Delta^2/4} B_{H^{12,b_k}_{t'_k,u'_k}}$ or $e^{-\Delta^2/4} B_{H^{12,b_k}_{u'_k,u'_k}}$. Writing $M_{b_k} = (m_{b'_l b'_m})_{b'_l, b'_m \in \{t'_1,\ldots,t'_d,u'_1,\ldots u'_d\}}$, we use the following bound for its 2-norm: $\|M_{b_k}\|_2 \leq \max_l \sqrt{m^2_{t'_l t'_l} + m^2_{t'_l u'_l} + m^2_{u'_l t'_l} + m^2_{u'_l u'_l}} + \sqrt{\sum_{l \neq m} m^2_{b'_k b'_l}}$. It comes that

$$\|M_{t_k}\|_2 \leq e^{-\Delta^2/4}\left(\max\left\{\sqrt{B^2_{H^{10}_{t_k}} B^2_{H^{02}_{t'_k t'_k}} + 2B^2_{H^{10}_{t_k}} B^2_{H^{02}_{t'_k u'_k}} + B^2_{H^{10}_{t_k}} B^2_{H^{02}_{u'_k u'_k}}}, \sqrt{2B^2_{H^{12,1}_{t'_k,u'_k}} + B^2_{H^{12,1}_{t'_k,t'_k}} + B^2_{H^{12,1}_{u'_k,u'_k}}}\right\}\right.$$
$$\left. + \sqrt{\begin{aligned}&(d^2-3d+2)B^2_{H^{10}_{t_k}}\left(B^4_{H^{10}_{t_k}} + 2B^2_{H^{10}_{t_k}} B^2_{H^{10}_{u_k}} + B^4_{H^{10}_{u_k}}\right)\\ &+ 2(d-1)\left(B^2_{H^{11}_{t_k t'_k}} B^2_{H^{10}_{t_k}} + B^2_{H^{11}_{t_k t'_k}} B^2_{H^{10}_{u_k}} + B^2_{H^{11}_{t_k u'_k}} B^2_{H^{10}_{t_k}} + B^2_{H^{11}_{t_k u'_k}} B^2_{H^{10}_{u_k}}\right)\end{aligned}}\right)$$

and

$$\|M_{u_k}\|_2 \leq e^{-\Delta^2/4}\left(\max\left\{\sqrt{B^2_{H^{10}_{u_k}} B^2_{H^{02}_{t'_k t'_k}} + 2B^2_{H^{10}_{u_k}} B^2_{H^{02}_{t'_k u'_k}} + B^2_{H^{10}_{u_k}} B^2_{H^{02}_{u'_k u'_k}}}, \sqrt{2B^2_{H^{12,2}_{t'_k,u'_k}} + B^2_{H^{12,2}_{t'_k,t'_k}} + B^2_{H^{12,2}_{u'_k,u'_k}}}\right\}\right.$$
$$\left. + \sqrt{\begin{aligned}&(d^2-3d+2)B^2_{H^{10}_{u_k}}\left(B^4_{H^{10}_{t_k}} + 2B^2_{H^{10}_{t_k}} B^2_{H^{10}_{u_k}} + B^4_{H^{10}_{u_k}}\right)\\ &+ 2(d-1)\left(B^2_{H^{11}_{u_k t'_k}} B^2_{H^{10}_{t_k}} + B^2_{H^{11}_{u_k t'_k}} B^2_{H^{10}_{u_k}} + B^2_{H^{11}_{u_k u'_k}} B^2_{H^{10}_{t_k}} + B^2_{H^{11}_{u_k u'_k}} B^2_{H^{10}_{u_k}}\right)\end{aligned}}\right),$$
$$\leq e^{-\Delta^2/4}(170.5 + 25.78d)\,.$$

As this last bound is greater than the previous one, $\left\|K^{(12)}_{\mathrm{norm}}(x,x')\right\|_{x,x'} \leq \sqrt{2d}(170.5 + 25.78d)e^{-\Delta^2/4}$. □

Choice of r, Δ for the LPC

Proposition J.1 (K_{norm} satisfies the LPC, $d = 1$)**.** *Let $s \geq 2$. Assume that $\mathcal{X} \subset \mathbb{R} \times [u_{\min}, u_{\max}]$ and that $\tau \leq u_{\min}$. Then K_{norm} satisfies the LPC (see Definition 5.2) with parameters s, $\Delta(s) = 2\sqrt{13.88 + \ln(s-1)}$, $r = 0.3025$, $\bar{\varepsilon}_2(0.3025) = 0.13139$, $\bar{\varepsilon}_0(0.3025) = 0.04472$.*

Proof. To establish that K_{norm} satisfies the LPC, we first determine the size of the near regions r, giving the constraint on the minimal separation Δ (see Definition 5.2).

We want to pick r such that $\bar{\varepsilon}_0(r), \bar{\varepsilon}_2(r)$ exist and $\frac{1}{64}\min\left\{\frac{\bar{\varepsilon}_0(r)}{B_0}, \frac{\bar{\varepsilon}_2(r)}{B_2}\right\}$ is maximal, using Lemmas J.3 and J.2. This allows us to get the smallest Δ (see item 3 of Definition 5.2). Graphically, we choose $r = 0.3025$ (see [Giard, 2025, Section VII.1]). We can take $\bar{\varepsilon}_2(0.3025) = 0.13139$ and $\bar{\varepsilon}_0(0.3025) = 0.04472$. Then

$$\frac{1}{64}\min\left(\frac{0.04472}{B_0}, \frac{0.13139}{B_2}\right) \geq 0.000204618 =: c_r$$

(see [Giard, 2025, Section VII.1]).

Let $s \geq 2$. The minimal separation $\Delta(s)$ must satisfy item 3 of Definition 5.2. Using Lemma J.5 , it suffices that

$$(s-1)\sqrt{2}\,153.05 e^{-\frac{\Delta(s)^2}{4}} \leq c_r\,,$$

i.e. that $\Delta(s) \geq 2\sqrt{c_\Delta + \ln(s-1)}$ where $c_\Delta = 13.88 \geq \ln\left(\frac{\sqrt{2}153.05}{c_r}\right)$ (see [Giard, 2025, Section VII.1]). □

Proposition J.2 (K_{norm} satisfies the LPC, $d \geq 1$)**.** *Let $s \geq 2$. Assume that $\mathcal{X} \subset \mathbb{R}^d \times [u_{\min}, u_{\max}]^d$ and that $\tau \leq u_{\min}$. Then K_{norm} satisfies the LPC with parameters s, $r = \frac{0.3025}{\sqrt{d}}$, $\bar{\varepsilon}_2(r) = 0.13139$, $\bar{\varepsilon}_0(r) = \frac{0.0894}{2d}$, $\Delta(s) = 2\sqrt{11.9 + 3\ln(d + 6.62) + \ln(s-1)}$.*

Proof. We take Proposition J.1 as a starting point. We make use of Lemmas J.2, J.4 and J.6. Setting $r = \frac{0.3025}{\sqrt{d}}$, we can take $\bar{\varepsilon}_2(r) = 0.13139$ (as in dimension $d = 1$) because $d \in \mathbb{N}^* \mapsto e^{-\frac{r_0^2}{2d}}|G(r_0)|$ with $r_0 = 0.3025$ is non-decreasing.

Furthermore, for $r \leq 0.3025$ we have $1 - e^{-r^2/2} \geq 0.977 \frac{r^2}{2}$. In fact, for $0 < c < 1$, denoting

$$\square_c : r \in \mathbb{R}^+ \mapsto 1 - e^{-r^2/2} - c\frac{r^2}{2},$$

we have $\frac{\partial}{\partial r}\square_c(r) = r(e^{-r^2/2} - c)$ so $\square_c$ is increasing then decreasing. Then remark that $\square_c(0) = 0$ and $\square_{0.977}(0.3025) \geq 0$. So we can take $\bar{\varepsilon}_0(r) = \frac{0.0894}{2d} \leq 0.977 \frac{0.3025^2}{2d}$ (see [Giard, 2025, Section VII.2]).

Moreover, $\frac{\bar{\varepsilon}_0(r)}{1+B_{00}+B_{10}} \leq \frac{\bar{\varepsilon}_2(r)}{1+B_{02}+B_{12}}$ (see [Giard, 2025, Section VII.2]). So

$$\frac{1}{64}\min\left(\frac{\bar{\varepsilon}_0(r)}{1+B_{00}+B_{10}}, \frac{\bar{\varepsilon}_2(r)}{1+B_{02}+B_{12}}\right) \geq \frac{1}{64}\frac{\varepsilon_0(r)}{1+B_{00}+B_{10}} \geq \frac{1}{64}\frac{0.0894}{2d(2+\sqrt{2d})} =: c_{d,r}.$$

The minimal separation should verify $(s-1)\sqrt{2d}(170.5 + 25.78d)e^{-\Delta(s)^2/4} \leq c_{d,r}$, *i.e.*

$$\Delta(s) \geq 2\sqrt{\ln\left(\frac{64}{0.0894}\right) + \ln\left((170.5 + 25.78d)2d(2+\sqrt{2d})\sqrt{2d}\right) + \ln(s-1)}.$$

It suffices that $\Delta(s) \geq 2\sqrt{11.9 + 3\ln(d + 6.62) + \ln(s-1)}$ (see [Giard, 2025, Section VII.2]). □

K Proofs of Section 6

K.1 Proof of Lemma 6.1

We construct η_{NDSC} in the same way as in Theorem 5.2, but we do not track the constants and dependence on d. We first determine r_{NDSC} such that η_{NDSC} can satisfy (NDSC). Then we prove that K_{norm} verifies the LPC (Definition 5.2) with $r = r_{\text{NDSC}}$ and a sufficiently large minimal separation, showing the existence of the non-degenerate certificate η_{NDSC}.

Choice of r_{NDSC}: Given $r_{\text{NDSC}} > 0$, according to Lemma I.2 it suffices that

$$1 - \frac{B_{02} + \bar{\varepsilon}_2(r_{\text{NDSC}})/16}{2}\mathfrak{d}_{\mathfrak{g}}(x, x_j^0)^2 > -1 \quad \forall x \in \mathcal{X}_j^{near}(r_{\text{NDSC}}) \tag{83}$$

to have $\eta_{\text{NDSC}} > -1$ (the existence of this certificate is proven later). Lemma J.2 gives $B_{02} = \sqrt{4d^2 + 10d}$. To bound $\mathfrak{d}_{\mathfrak{g}}(x, x_j^0)^2$, we use Lemma H.7. If $d\!\!\!/(x, x_j^0) < \sqrt{2}$, we have $\mathfrak{d}_{\mathfrak{g}}(x, x_j^0)^2 \leq dF(d\!\!\!/(x, x_j^0))$ where F is defined by (62) in the appendix. We remark that F is continuous on $[0, \sqrt{2}[$ and that $F(0) = 0$. Furthermore, if $r \leq \frac{0.32}{\sqrt{d}}$ we can take $\bar{\varepsilon}_2(r) = e^{-\frac{0.32^2}{2}}|G(0.32)|$ (see Lemma J.4). This quantity does not depend on r. So there exists $0 < r_{\text{NDSC}} \leq \frac{0.32}{\sqrt{d}}$ (that depends only on d) such that (83) is satisfied.

LPC and existence of η_{NDSC}: It remains to show the existence of η_{NDSC}, a global non-degenerate certificate of the form (18) satisfying Definition 3.2 where the near regions are of radius r_{NDSC}. We first prove the LPC with this choice of radius, choosing Δ large enough, and then make use of Theorem 5.2.

The curvature constants $\bar{\varepsilon}_0(r_{\text{NDSC}})$ and $\bar{\varepsilon}_2(r_{\text{NDSC}})$ are given by Lemma J.4. The separation $\Delta(s)$ should satisfy

$$(s-1)\sqrt{2}\,153.05e^{-\frac{\Delta(s)^2}{4}} \leq \frac{1}{64}\min\left(\frac{\bar{\varepsilon}_0(r_{\text{NDSC}})}{B_0}, \frac{\bar{\varepsilon}_2(r_{\text{NDSC}})}{B_2}\right).$$

Such a $\Delta(s)$ exists as $e^{-\frac{\Delta(s)^2}{4}} \underset{\Delta(s)\to+\infty}{\longrightarrow} 0$. It depends only on d (through $\bar{\varepsilon}_0(r_{\text{NDSC}})$ and $\bar{\varepsilon}_2(r_{\text{NDSC}})$) and s. Then K_{norm} satisfies the LPC (Definition 5.2) with these parameters.

Finally, Theorem 5.2 applies. The minimal separation Δ_{NDSC} follows from Lemma 5.3. It depends on d, $u_{\min}$, $u_{\max}$, τ, r_{NDSC}, $\Delta(s)$. Under this separation, we can construct a non-degenerate certificate η_{NDSC} verifying $\eta_{\text{NDSC}} > -1$. We can also construct local non-degenerate certificates $\eta_{j,\text{NDSC}}$ for $j = 1, \ldots, s$.

Negative definiteness of $\nabla^2\eta_{\text{NDSC}}(x_j^0)$: Let $j \in \{1, \ldots, s\}$. To show that $\nabla^2\eta_{\text{NDSC}}(x_j^0)$ is negative-definite, first remark that $\nabla^2\eta_{\text{NDSC}}(x_j^0) = D_2[\eta_{\text{NDSC}}](x_j^0)$ because $\nabla\eta_{\text{NDSC}}(x_j^0) = 0$ (equality of the Riemannian and Euclidean hessians, see Definition 5.1). Then, for $v \in \mathbb{R}^2$ such that $\|v\|_{x_j^0} = 1$, using the control on the near regions (proof of Theorem 5.2) we deduce that

$$D_2[\eta_{\text{NDSC}}](x_j^0)[v,v] \leq K_{\text{norm}}^{(02)}(x_j^0, x_j^0)[v,v] + \left\|D_2[\eta_{\text{NDSC}}](x_j^0) - K_{\text{norm}}^{(02)}(x_j^0, x_j^0)\right\|_{x_j^0} \leq -\bar{\varepsilon}_2 + \frac{\bar{\varepsilon}_2}{16} = -\frac{\bar{\varepsilon}_2}{15}.$$

So for all $v \in \mathbb{R}^{2d} \setminus \{0\}$,

$$v^T\nabla^2\eta_{\text{NDSC}}(x_j^0)v \leq -\|v\|_{x_j^0}\frac{\bar{\varepsilon}_2}{15} < 0.$$

K.2 Proof of Theorem 6.1

We begin with a preliminary result.

Lemma K.1. *Under the hypothesis of Lemma 6.1, G_{S_0} is full-rank, where*

$$G_{S_0} : (\alpha_1, \dots, \alpha_s), (\beta_1, \dots, \beta_s) \in \mathbb{R}^s \times \mathbb{R}^{2d\times s} \mapsto \sum_{j=1}^{s} \alpha_j \Psi\delta_{x_j^0} + \sum_{j=1}^{s} \beta_j^T \nabla_x(\Psi\delta_{x_j^0})\,. \tag{84}$$

Moreover, $B \coloneqq \max_{k=0,1,2} \sup_{x\in\mathcal{X}} \left\|\nabla_x^k \Psi(\delta_x)\right\|_{\mathbb{L},k} < \infty$ where

$$\begin{aligned}
\left\|\nabla_x^0(\Psi\delta_x)\right\|_{\mathbb{L},0} &= \|\Psi\delta_x\|_{\mathbb{L}}\,,\\
\left\|\nabla_x^1(\Psi\delta_x)\right\|_{\mathbb{L},1} &= \left\|\left((\|\partial_{t_k}(\Psi\delta_x)\|_{\mathbb{L}})_{k=1,\dots,d}, (\|\partial_{u_k}(\Psi\delta_x)\|_{\mathbb{L}})_{k=1,\dots,d}\right)\right\|_2\,,\\
\left\|\nabla_x^2(\Psi\delta_x)\right\|_{\mathbb{L},2} &= \left\|\begin{pmatrix} (\left\|\partial^2_{t_k t_l}(\Psi\delta_x)\right\|_{\mathbb{L}})_{1\le k,l\le d} & (\left\|\partial^2_{t_k u_l}(\Psi\delta_x)\right\|_{\mathbb{L}})_{1\le k,l\le d} \\ (\left\|\partial^2_{t_k u_l}(\Psi\delta_x)\right\|_{\mathbb{L}})_{1\le k,l\le d} & (\left\|\partial^2_{u_k u_l}(\Psi\delta_x)\right\|_{\mathbb{L}})_{1\le k,l\le d}\end{pmatrix}\right\|_2\,.
\end{aligned}$$

Proof. $\underline{G_{S_0}\text{ is full-rank:}}$ We used the invertibility of Υ (see (63)) to construct η_{NDSC}: we already know that

$$\sum_{j=1}^{s} \alpha_j K_{\text{norm}}(x_j^0, \bullet) + \sum_{j=1}^{s} \beta_j^T \nabla_1 K_{\text{norm}}(x_j^0, \bullet) = 0 \quad\Longrightarrow\quad \alpha_j = 0\,,\ \beta_j = 0_{\mathbb{R}^2}\ \forall j = 1, \dots s\,.$$

$\underline{B < \infty:}$ It can be deduced from the fact that $x \mapsto \Psi\delta_x$ is C^∞ on the compact $\mathcal{X}$ (Remark 5.1). □

Below we present the main steps of the proof of Theorem 6.1, adapted from [Duval and Peyré, 2015, Lemma 2, Proposition 7] and [Poon, 2019, Section 4.3]. We work under the assumptions of Lemma 6.1: there exists η_{NDSC} of the form (18) that verifies $|\eta_{\text{NDSC}}| \le 1$, $|\eta_{\text{NDSC}}(x)| = 1 \iff x \in \{x_j^0\}_j$ and $\nabla^2\eta_{\text{NDSC}}(x_j^0) \prec 0$ for all $j = 1, \dots, s$ (Non Degenerate Source Condition, or NDSC).

Dual, noisy and noiseless problems We denote $\mathcal{D}_{\kappa,b}$ the dual problem with regularization κ and noise b. More precisely:

$$\operatorname*{arg\,max}_{p\in\mathbb{L}, \|\Psi^* p\|_\infty \le 1} \left\langle \Psi\mu_\omega^0, p\right\rangle_{\mathbb{L}} \tag{$\mathcal{D}_{0,0}$}$$

is the dual problem of

$$\operatorname*{arg\,min}_{\mu\in\mathcal{M}(\mathcal{X})} \|\mu\|_{\text{TV}} \text{ such that } \Psi\mu = \Psi\mu_\omega^0\,. \tag{$\mathcal{P}_{0,0}$}$$

For $\kappa > 0$, the dual problem is

$$\operatorname*{arg\,min}_{p\in\mathbb{L}, \|\Psi^* p\|_\infty \le 1} \left\|\frac{y}{\kappa} - p\right\|_{\mathbb{L}}^2\,, \tag{$\mathcal{D}_{\kappa,b}$}$$

where we observe $y = \Psi\mu_\omega^0 + b$ (*i.e.* $b = \Gamma_n$). Its solution is unique. It is the dual problem of

$$\operatorname*{arg\,min}_{\mu\in\mathcal{M}(\mathcal{X})} \frac{1}{2}\|y - \Psi\mu\|_{\mathbb{L}}^2 + \kappa\|\mu\|_{\text{TV}}\,. \tag{$\mathcal{P}_{\kappa,b}$}$$

We use the notation $p_{\kappa,b}$ for the solution of the dual problem, and $\eta_{\kappa,b} = \Psi^* p_{\kappa,b}$. For $\kappa = 0$, we choose the particular dual solution defined by ($p_{0,0}$).

The dual and primal solutions are connected via the subdifferential of the TV-norm: we recall that for $\eta \in \mathcal{C}(\mathcal{X})$ and $\mu \in \mathcal{M}(\mathcal{X})$,

$$\eta \in \partial\|\mu\|_{\text{TV}} \iff \left(\|\eta\|_\infty \le 1\,, \text{supp}(\mu^-) \subset \{\eta = -1\}\,, \text{supp}(\mu^+) \subset \{\eta = 1\}\right)\,.$$

Strong duality ensures that for $\mu_{\kappa,b}$ a solution of the primal problem, $\eta_{\kappa,b} \in \partial\|\mu_{\kappa,b}\|_{\text{TV}}$. Moreover, for $\kappa > 0$ we have $p_{\kappa,b} = -\frac{1}{\kappa}(\Psi\mu_{\kappa,b} - \Psi\mu_\omega^0 - b)$.

Lemma K.2. *Under the assumptions of Lemma 6.1, μ_ω^0 is the unique solution of ($\mathcal{P}_{0,0}$). Moreover, $\eta_{0,0} \coloneqq \eta_{NDSC}|_{\mathcal{X}} = \Psi^* p_{0,0}$ where $p_{0,0}$ is the solution of ($\mathcal{D}_{0,0}$) with minimal norm,* i.e.

$$p_{0,0} = \operatorname*{arg\,min}_{p\in\mathbb{L}} \left\{\|p\|_{\mathbb{L}} \,:\, \Psi^* p \in \partial\left\|\mu_\omega^0\right\|_{\text{TV}}\right\}\,.$$

Proof. As $\eta_{\text{NDSC}}|_{\mathcal{X}} \in \partial \left\|\mu_\omega^0\right\|_{\text{TV}}$, we deduce that $p_{0,0}$ is a solution of $(\mathcal{D}_{0,0})$, linked to any solution $\mu_{0,0}$ of $(\mathcal{P}_{0,0})$ by $\Psi^* p_{0,0} = \eta_{\text{NDSC}}|_{\mathcal{X}} \in \partial \|\mu_{0,0}\|_{\text{TV}}$. So $\text{supp}(\mu_{0,0}) \subset \{|\eta_{\text{NDSC}}|_{\mathcal{X}}| = 1\} = \{x_j^0\}_{j=1}^s$. Using the injectivity of G_{S_0} (Lemma K.1), we have $\mu_{0,0} = \mu_\omega^0$.
Moreover, as η_{NDSC} (which is of the form (18)) verifies the NDSC, [Duval and Peyré, 2015, Proposition 7] shows that $p_{0,0}$ is of minimal norm. □

Lemma K.3 (Convergence of the dual solutions). *It holds that* $\|p_{\kappa,0} - p_{0,0}\|_{\mathbb{L}} \underset{\kappa\to 0}{\to} 0$.
Moreover, $\|p_{\kappa,b} - p_{\kappa,0}\|_{\mathbb{L}} \leq \frac{\|b\|_{\mathbb{L}}}{\kappa}$.

Proof. $\|p_{\kappa,0} - p_{0,0}\|_{\mathbb{L}} \underset{\kappa\to 0}{\to} 0$: Using that $p_{\kappa,0}$ is the solution of $(\mathcal{D}_{\kappa,b})$ for $b = 0$ and $p_{0,0}$ is a solution of $(\mathcal{D}_{0,0})$, writing $y = \Psi\mu_\omega^0$ we have

$$\langle y, p_{\kappa,0}\rangle_{\mathbb{L}} - \frac{\kappa}{2}\|p_{\kappa,0}\|^2 \geq \langle y, p_{0,0}\rangle_{\mathbb{L}} - \frac{\kappa}{2}\|p_{0,0}\|^2 \quad \text{and} \quad \langle y, p_{0,0}\rangle_{\mathbb{L}} \geq \langle y, p_{\kappa,0}\rangle_{\mathbb{L}} .$$

We deduce that $\|p_{\kappa,0}\|_{\mathbb{L}}$ is bounded by $\|p_{0,0}\|_{\mathbb{L}}$. Given $\kappa_n \to 0$, as the closed unit ball of a Hilbert space is weakly sequentially compact, we can extract $p_{\kappa_{n_k},0}$ that converges weakly towards $p^* \in \mathbb{L}$.
We show that $p^* = p_{0,0}$. We have $\langle y, p^*\rangle_{\mathbb{L}} \geq \langle y, p_{0,0}\rangle_{\mathbb{L}}$ so

$$\left\langle \mu_\omega^0, \Psi^* p^*\right\rangle \geq \left\langle \mu_\omega^0, \Psi^* p_{0,0}\right\rangle = \left\|\mu_\omega^0\right\|_{\text{TV}} .$$

We also have that Ψ^* is weakly continuous from $\mathbb{L}$ to $\mathcal{C}(\mathcal{X})$ so

$$\|\Psi^* p^*\|_\infty \leq \left\|\Psi^* p_{\kappa_{n_k},0}\right\|_\infty = 1 ,$$

hence $p^* \in \partial \left\|\mu_\omega^0\right\|_{\text{TV}}$. Finally,

$$\|p_{0,0}\|_{\mathbb{L}} \geq \liminf \left\|p_{\kappa_{n_k},0}\right\|_{\mathbb{L}} \geq \|p^*\|_{\mathbb{L}}$$

as if $h_n \rightharpoonup h$ in $\mathbb{L}$, then $\|h\|_{\mathbb{L}}^2 \leq \liminf \|h_n\|_{\mathbb{L}}^2$. Hence $p^* = p_{0,0}$.
Then $p_{\kappa_{n_k},0} \to p_{0,0}$ strongly because $\left\|p_{\kappa_{n_k},0}\right\|_{\mathbb{L}} \to \|p_{0,0}\|_{\mathbb{L}}$ and $p_{\kappa_{n_k},0} \rightharpoonup p_{0,0}$. Finally, $p_{\kappa,0} \underset{n_k\to\infty}{\longrightarrow} p_{0,0}$ strongly because any convergent subsequence of $(p_{\kappa,0})_\kappa$ has limit $p_{0,0}$.
$\|p_{\kappa,b} - p_{\kappa,0}\|_{\mathbb{L}} \leq \frac{\|b\|_{\mathbb{L}}}{\kappa}$: The mapping $P : \frac{y}{\kappa} \to p_{\kappa,b}$ is the projection onto a close convex set so it is non expansive: for $h_1, h_2 \in \mathbb{L}$,

$$\langle h_1 - P(h_1), P(h_2) - P(h_1)\rangle \leq 0 \quad \text{and} \quad \langle h_2 - P(h_2), P(h_1) - P(h_2)\rangle \leq 0 .$$

Subtracting the two and applying Cauchy-Schwarz leads to

$$\langle P(h_1) - P(h_2), P(h_1) - P(h_2)\rangle \leq \langle h_1 - h_2, P(h_1) - P(h_2)\rangle \leq \|h_1 - h_2\| \|P(h_1) - P(h_2)\| .$$

Hence

$$\|p_{\kappa,b} - p_{\kappa,0}\| \leq \left\|\frac{\Psi\mu_\omega^0 + b - \Psi\mu_\omega^0}{\kappa}\right\| = \frac{\|b\|}{\kappa} .$$

□

Proposition K.1 (Sparsity of the solution). *Under the assumptions of Lemma 6.1, there exists* $\tilde\kappa_0 > 0$, γ_0 *such that for all* $\kappa \leq \tilde\kappa_0$ *and* b *such that* $\|b\|_{\mathbb{L}} \leq \gamma_0\kappa$*, the solution* $\mu_{\kappa,b}$ *of* $(\mathcal{P}_{\kappa,b})$ *is a discrete measure and has at most 1 particle in each* $\mathcal{X}_j^{near}(r_{\text{NDSC}})$.

Proof. Using NDSC, the fact that $\mathcal{X}$ is compact and that $diam(\mathcal{X}_j^{near}(r_e)) \underset{r_e\to 0}{\to} 0$, we deduce that there exists $\varepsilon > 0$, $r_\varepsilon \in]0, r_{\text{NDSC}}]$ such that $|\eta_{0,0}(x)| < 1 - \varepsilon$ for all $x \notin \bigcup \mathcal{X}_j^{near}(r_\varepsilon)$ and $\nabla^2\eta_{0,0}(x) \prec -\varepsilon I$ for all $x \in \bigcup \mathcal{X}_j^{near}(r_\varepsilon)$.
In light of Lemma K.3, we can choose $\tilde\kappa_0$ such that for all $\kappa \leq \tilde\kappa_0$, $B\|p_{\kappa,0} - p_{0,0}\|_{\mathbb{L}} \leq \frac{\varepsilon}{4}$ (where B is defined in Lemma K.1). For $\|b\|_{\mathbb{L}} \leq \gamma_0\kappa$ with $\gamma_0 = \frac{\varepsilon}{4B}$, using Cauchy-Schwarz we get that for all $x \in \mathcal{X}$,

$$\begin{aligned}\left\|\nabla^k\eta_{\kappa,b}(x) - \nabla^k\eta_{0,0}(x)\right\|_{\mathbb{L},k} &\leq \|p_{\kappa,b} - p_{0,0}\|_{\mathbb{L}} \left\|\nabla_x^k\Psi(\delta_x)\right\|_{\mathbb{L},k} ,\\ &\leq B\left(\frac{\|b\|_{\mathbb{L}}}{\kappa} + \|p_{\kappa,0} - p_{0,0}\|_{\mathbb{L}}\right) ,\\ &\leq \frac{\varepsilon}{2} .\end{aligned}$$

So $|\eta_{\kappa,b}(x)| < 1$ for all $x \notin \bigcup \mathcal{X}_j^{near}(r_\varepsilon)$ and $\nabla^2\eta_{\kappa,b}(x)$ is negative-definite for all $x \in \bigcup \mathcal{X}_j^{near}(r_\varepsilon)$. Hence there exists at most 1 point in each $\mathcal{X}_j^{near}(r_\varepsilon)$ such that $\eta_{\kappa,b}(x) = 1$. As $\eta_{\kappa,b} \in \partial \|\mu_{\kappa,b}\|_{\text{TV}}$, we deduce that $\mu_{\kappa,b}$ is a discrete measure and that it has at most 1 particle in each $\mathcal{X}_j^{near}(r_\varepsilon)$. □

Lemma K.4. *Under the assumptions of Lemma 6.1, there exists $\kappa_0 > 0$, γ_0 such that for all $\kappa \leq \kappa_0$ and b such that $\|b\|_{\mathbb{L}} \leq \gamma_0 \kappa$, $\mu_{\kappa,b}$ is a discrete measure and has exactly 1 particle in each $\mathcal{X}_j^{near}(r_{\text{NDSC}})$.*

Proof. We take Proposition K.1 as a starting point. We show that $\mu_{\kappa,b}(\mathcal{X}_j^{near}(r_\varepsilon)) > 0$ for $\|b\|_{\mathbb{L}} \leq \gamma_0\kappa$ and $\kappa \leq \kappa_0$ with $\kappa_0 \leq \tilde{\kappa}_0$ small enough.

It suffices to show the weak* convergence of $\mu_{\kappa,b}$ towards μ_ω^0 as $\kappa \to 0$ and for $\|b\|_{\mathbb{L}} \leq \gamma_0\kappa$ (this result is also stated in [Duval and Peyré, 2015, Proposition 4]). We have

$$\frac{1}{2}\left\|\Psi\mu_\omega^0 + b - \Psi\mu_{\kappa,b}\right\|_{\mathbb{L}}^2 + \kappa\left\|\mu_{\kappa,b}\right\|_{\text{TV}} \leq \frac{1}{2}\|b\|_{\mathbb{L}}^2 + \kappa\left\|\mu_\omega^0\right\|_{\text{TV}}$$

since $\mu_{\kappa,b}$ minimizes

$$J_{\kappa,b} : \mu \mapsto \frac{1}{2}\left\|\Psi\mu_\omega^0 + b - \Psi\mu\right\|_{\mathbb{L}}^2 + \kappa\left\|\mu\right\|_{\text{TV}}\,.$$

Dividing by κ, it follows that $\|\mu_{\kappa,b}\|_{\text{TV}}$ is bounded by $\frac{\gamma_0^2}{2}\kappa + \left\|\mu_\omega^0\right\|_{\text{TV}}$. So we can extract μ_k a weak* convergent subsequence of $\mu_{\kappa,b}$ as $\kappa, b \to 0$ with $\|b\| \leq \gamma_0\kappa$. Let μ^* be its limit. By the lower semi-continuity of the TV norm for the weak* convergence, $\|\mu^*\|_{\text{TV}} \leq \liminf \|\mu_k\|_{\text{TV}} \leq \left\|\mu_\omega^0\right\|_{\text{TV}}$. To establish $\mu^* = \mu_\omega^0$, it remains to show that $\Psi\mu^* = \Psi\mu_\omega^0$ (because μ_ω^0 is the only solution of $(\mathcal{P}_{0,0})$, see Lemma K.2). Recall that $\|p_{\kappa,b} - p_{\kappa,0}\|_{\mathbb{L}} \leq \frac{\|b\|_{\mathbb{L}}}{\kappa}$ and that $p_{\kappa,b} = -\frac{1}{\kappa}(\Psi\mu_{\kappa,b} - \Psi\mu_\omega^0 - b)$, giving $\|\Psi\mu_{\kappa,b} - b - \Psi\hat{\mu}_{\kappa,0}\|_{\mathbb{L}} \leq \|b\|$. As $\left\|\Psi\mu_{\kappa,0} - \Psi\mu_\omega^0\right\|_{\mathbb{L}}^2 \leq 2\kappa\left\|\mu_\omega^0\right\|_{\text{TV}}$ (because $J_{\kappa,0}(\mu_{\kappa,0}) \leq J_{\kappa,0}(\mu_\omega^0)$), it comes

$$\begin{aligned}\left\|\Psi\mu_{\kappa,b} - \Psi\mu_\omega^0\right\|_{\mathbb{L}} &\leq \|\Psi\mu_{\kappa,b} - \Psi\hat{\mu}_{\kappa,0}\|_{\mathbb{L}} + \left\|\Psi\mu_{\kappa,0} - \Psi\mu_\omega^0\right\|_{\mathbb{L}}\,,\\ &\leq 2\|b\|_{\mathbb{L}} + \sqrt{2\kappa\left\|\mu_\omega^0\right\|_{\text{TV}}}\,.\end{aligned}$$

As Ψ is weak* to weak continuous from $\mathcal{M}(\mathcal{X})$ to $\mathbb{L}$, for all $p \in \mathbb{L}$ we have

$$|\left\langle p, \Psi(\mu_k - \mu_\omega^0)\right\rangle_{\mathbb{L}}| \to |\left\langle p, \Psi(\mu^* - \mu_\omega^0)\right\rangle_{\mathbb{L}}|\,.$$

Applying Cauchy-Schwarz, we have $|\left\langle p, \Psi(\mu_k - \mu_\omega^0)\right\rangle_{\mathbb{L}}| \to 0$ from which we conclude that $\left\langle p, \Psi(\mu^* - \mu_\omega^0)\right\rangle_{\mathbb{L}} = 0$. So $\Psi\mu^* = \Psi\mu_\omega^0$, and $\mu_{\kappa,b}$ converges towards μ_ω^0 as $\kappa \to 0$ and $\|b\|_{\mathbb{L}} \leq \gamma_0\kappa$ for the weak* topology. This concludes the proof. ☐

K.3 Proof of Corollary 6.1

If $n \geq \frac{c_\kappa^2}{\kappa_0^2(2\pi)^{d/2}\tau^d}$, choosing $\kappa = \frac{c_\kappa}{(2\pi)^{d/4}\tau^{d/2}\sqrt{n}}$ we have $\kappa \leq \kappa_0$. Moreover, it holds that $\|\Gamma_n\| \leq \gamma_0\kappa$ with probability greater than $1 - C_\Gamma e^{-\left(\frac{\gamma_0 c_\kappa}{C_\Gamma}\right)^2}$ (*c.f.* Lemma 3.1 where we took $\rho = \frac{\gamma_0^2 c_\kappa^2}{C_\Gamma^2}$). Theorem 6.1 applies: $\mu_{n,\omega}^\star = \sum_{j=1}^s \omega_j^\star \delta_{x_j^\star}$ where $\omega_j^\star > 0$ and $x_j^\star \in \mathcal{X}_j^{near}(r_{\text{NDSC}})$ for all $j = 1, \ldots, s$.

From (28), using $\|\Gamma_n\|_{\mathbb{L}} \leq \gamma_0\kappa$ and $\|p_{\text{NDSC}}\|_{\mathbb{L}} \leq \sqrt{2s}$ (see Lemma 6.1), we have

$$\begin{aligned}D_{\eta_{\text{NDSC}}}(\mu_{n,\omega}^\star, \mu_\omega^0) &\leq \frac{\kappa}{2}(\gamma_0 + \|p_{\text{NDSC}}\|_{\mathbb{L}})^2\,,\\ &\leq \frac{c_\kappa}{2(2\pi)^{d/4}\tau^{d/2}\sqrt{n}}(\gamma_0 + \sqrt{2s})^2\,. \end{aligned}\tag{85}$$

<u>Bound for $|\omega_j^0 - \omega_j^\star|$:</u> Recalling (32), as $\mu_{n,\omega}^\star(\mathcal{X}^{far}(r_{\text{NDSC}})) = 0$ we have

$$\begin{aligned}|\omega_j^0 - \omega_j^\star| &\leq \|p_{j,\text{NDSC}}\|_{\mathbb{L}}\left(2\|\Gamma_n\|_{\mathbb{L}} + 2\kappa\|p_{\text{NDSC}}\|_{\mathbb{L}}\right) + \frac{\tilde{\varepsilon}_{2,\text{NDSC}}}{\varepsilon_{2,\text{NDSC}}}D_{\eta_{\text{NDSC}}}\left(\mu_{n,\omega}^\star, \mu_\omega^0\right)\,,\\ &\leq 2\kappa\|p_{j,\text{NDSC}}\|_{\mathbb{L}}\left(\gamma_0 + \|p_{\text{NDSC}}\|_{\mathbb{L}}\right) + \frac{\tilde{\varepsilon}_{2,\text{NDSC}}}{\varepsilon_{2,\text{NDSC}}}D_{\eta_{\text{NDSC}}}\left(\mu_{n,\omega}^\star, \mu_\omega^0\right)\,,\\ &\leq \frac{c_\kappa}{(2\pi)^{d/4}\tau^{d/2}\sqrt{n}}\left(2\sqrt{2}(\gamma_0 + \sqrt{2s}) + \frac{1}{2}(\gamma_0 + \sqrt{2s})^2\frac{\tilde{\varepsilon}_{2,\text{NDSC}}}{\varepsilon_{2,\text{NDSC}}}\right)\,.\end{aligned}$$

The second factor of this product only depends on $\mathcal{X}$, τ and μ^0.
<u>Control of $\mathfrak{d}(x_j^\star, x_j^0)$:</u> Using (85) together with (31), we get

$$\sum_{j=1}^s \omega_j^\star \mathfrak{d}_{\mathfrak{g}}(x_j^\star, x_j^0)^2 \leq \frac{1}{\varepsilon_{2,\text{NDSC}}}\frac{c_\kappa}{2(2\pi)^{d/4}\tau^{d/2}\sqrt{n}}(\gamma_0 + \sqrt{2s})^2\,.$$

To control the proximity between x_j^0 and $x_j^\star$ with the semi-distance, we use Lemma H.5. It comes that

$$\sum_{j=1}^{s} \omega_j^\star \mathfrak{d}(x_j^\star, x_j^0)^2 \leq \frac{\tilde{\varepsilon}_{3,\text{NDSC}}}{\varepsilon_{2,\text{NDSC}}} \frac{c_\kappa}{2(2\pi)^{d/4}\tau^{d/2}\sqrt{n}}(\gamma_0 + \sqrt{2s})^2$$

where $\tilde{\varepsilon}_{3,\text{NDSC}}$ only depends on r_{NDSC}. As $\omega_j^* \geq \omega_j^0 - |\omega_j^0 - \omega_j^*|$, if $|\omega_j^0 - \omega_j^*| < \omega_j^0$ then

$$\begin{aligned} \mathfrak{d}(x_j^\star, x_j^0)^2 &\leq \frac{1}{\omega_j^0 - |\omega_j^0 - \omega_j^*|} \omega_j^* \mathfrak{d}(x_j^\star, x_j^0)^2 \,, \\ &\leq \frac{1}{\omega_j^0 - |\omega_j^0 - \omega_j^*|} \frac{\tilde{\varepsilon}_{3,\text{NDSC}}}{\varepsilon_{2,\text{NDSC}}} \frac{c_\kappa}{2(2\pi)^{d/4}\tau^{d/2}\sqrt{n}}(\gamma_0 + \sqrt{2s})^2 \,. \end{aligned}$$

Recalling (23) and the condition $\frac{n}{c_\kappa^2} \geq \frac{1}{\kappa_0^2 (2\pi)^{d/2}\tau^d}$, if c_κ is chosen as $c_{\kappa,n} = \underset{n\to+\infty}{o}(\sqrt{n})$, there exists $n_0 \in \mathbb{N}$ depending on $(c_{\kappa,n})_n$, μ^0, $\mathcal{X}$, τ and $\tilde{c}_0 > 0$ that depends on μ^0, $\mathcal{X}$, τ (through γ_0, ω_j^0, s *etc.*) such that for all $n \geq n_0$, $|\omega_j^0 - \omega_j^*| < \omega_j^0$ and $\frac{\tilde{\varepsilon}_{3,\text{NDSC}}}{\varepsilon_{2,\text{NDSC}}} \frac{(\gamma_0+\sqrt{2s})^2}{2(2\pi)^{d/4}\tau^{d/2}(\omega_j^0 - |\omega_j^0 - \omega_j^*|)} \leq \tilde{c}_0$ with probability at least $1 - C_\Gamma e^{-\left(\frac{\gamma_0 c_{\kappa,n}}{C_\Gamma}\right)^2}$. Then $\mathfrak{d}(x_j^\star, x_j^0)^2 \leq \tilde{c}_0 \frac{c_{\kappa,n}}{\sqrt{n}}$.