On strong basins of attractions for non-convex sparse spike estimation: upper and lower bounds

Yann Traonmilin, Jean-François Aujol, Pierre-Jean Bénard, Arthur Leclaire

On strong basins of attractions for non-convex sparse spike estimation: upper and lower bounds

Yann Traonmilin[1*], Jean-François Aujol[1], Pierre-Jean Bénard[1] and Arthur Leclaire[1]

[1]Univ. Bordeaux, Bordeaux INP, CNRS, IMB, UMR 5251, F-33400, Talence, France.

*Corresponding author(s). E-mail(s): yann.traonmilin@math.u-bordeaux.fr;
Contributing authors: jean-francois.aujol@math.u-bordeaux.fr;
pierre-jean.benard@math.u-bordeaux.fr; arthur.leclaire@math.u-bordeaux.fr;

Abstract

In this article, we study the size of strong basins of attractions for the non-convex sparse spike estimation problem. We first extend previous results to obtain a lower bound on the size of sets where gradient descent converges with a linear rate to the minimum of the non-convex objective functional. We then give an upper bound that shows that the dependency of the lower bound with respect to the number of measurements reflects well the true size of basins of attraction for random Gaussian Fourier measurements. These theoretical results are confirmed by experiments.

1 Introduction

In the space $\mathcal{M}$ of finite signed measures over $\mathbb{R}^d$, we aim at recovering a sum of spikes $x_0 = \sum_{i=1}^k a_i\delta_{t_i}$ (where δ_{t_i} is the Dirac measure at position $t_i \in \mathbb{R}^d$, i.e. a spike of amplitude 1) from measurements

$$y = Ax_0 + e. \tag{1}$$

The operator A is a linear observation operator defined by the duality products $(Ax_0)_l = \langle x_0, \alpha_l\rangle$ where $(\alpha_l)_{l=1}^m$ is a collection of m functions, typically sinusoids for Fourier measurements. The vector $y \in \mathbb{C}^m$ contains the m noisy measurements and e is an observation noise of finite energy. This problem, called the off-the-grid super-resolution problem [1], has seen many theoretical developments in the last years [2–6] with a wide range of applications such as microscopy [7] and audio signal processing [8].

Under a restricted isometry property (RIP) on A the following minimization estimates x_0 [9] in a stable manner:

$$x^* \in \underset{x\in\Sigma_{k,\epsilon}}{\operatorname{argmin}} \|Ax - y\|_2^2, \tag{2}$$

where $\Sigma_{k,\epsilon}$ is a low-dimensional set modeling a separation constraint on the k Dirac measures. The model $\Sigma_{k,\epsilon}$ can be parametrized by a set $\Theta_{k,\epsilon} \subset \mathbb{R}^D$ using the parametrization function ϕ defined by

$$\phi(a_1, \ldots, a_k, t_1, \ldots, t_k) := \sum_{i=1}^k a_i\delta_{t_i}, \tag{3}$$

with $a = (a_1, \ldots, a_k) \in \mathbb{R}^k$ and $(t_1, \ldots, t_k) \in (\mathbb{R}^d)^k$, i.e. $\Sigma_{k,\epsilon} = \phi(\Theta_{k,\epsilon})$ with

$$\Theta_{k,\epsilon} := \{\theta = (a, t_1, \dots, t_k) \in \mathbb{R}^{k(d+1)}, \\ a \in \mathbb{R}^k, t_i \in \mathbb{R}^d, \\ \forall i \neq j, \|t_i - t_j\|_2 > \epsilon, \|t_i\|_2 \leq R\}. \quad (4)$$

The process of calculating a solution of Problem (2) is called the ideal decoder. Using the function ϕ, it can be smoothly parametrized as a minimization in $\mathbb{R}^{k(d+1)}$: we rewrite (2) as

$$\theta^* \in \underset{\theta \in \Theta_{k,\epsilon}}{\operatorname{argmin}}\, g(\theta), \quad \text{with } g(\theta) = \|A\phi(\theta) - y\|_2^2. \quad (5)$$

In [10, 11], explicit basins of attraction of θ^* are given, that is, sets where one can choose to initialize the gradient descent with fixed step size to have convergence to θ^*. These studies of the geometry of the minimizers of the non-convex problems led to fast practical algorithms for sparse spike recovery [12, 13]. The size of basins of attractions is systematically linked with the number of measurements through restricted isometry constants of A. It was shown that their size is increasing (possibly not strictly) with respect to the number of measurements in a random measurement setting. This result was given in a very general setting requiring mostly a smooth parametrization function ϕ and an open set $\Theta_{k,\epsilon}$.

However, the sizes of basins of attractions given in [11] are only lower bounds, i.e. they lead to a sufficent condition for convergence. The question whether larger basins of attraction possibly exist remains open. Moreover, basins of attraction were defined as sets where only convergence of gradient descent (with fixed step size) is guaranteed, which led to typical sublinear convergence rates within such basins.

Contributions

In this article, we study the tightness of previous results on the size of basins of attraction for non-convex off-the-grid super-resolution.

- We begin by showing in Section 3, for the estimation of spike positions, that the sizes of basins of attractions given in [10, 11] can be slightly decreased to obtain strong basins of attraction, that is, basins of attraction leading to convergence of gradient descent with fixed step size with a linear convergence rate. The proof relies on a typical local strong convexity argument.
- In Section 4, in the context of random Gaussian Fourier measurements, we then give an upper bound on the sizes of sets where g is strongly convex: we show on the example of the recovery of one spike (i.e. the most favorable case) that the asymptotic behaviour of the size of basins of attraction where g is strongly convex (with respect to the number of measurements) matches the lower bound up to a constant. Given the chain of arguments (which are possibly not tight) leading to lower bounds on sizes of basins of attraction, it is remarkable that we can get such a match. Our result is obtained by a direct study of the Hessian of g using concentration inequalities (the fundamental tool also involved in RIP-based recovery guarantees) which is also of independent interest.
- In Section 5, we perform synthetic experiments, which confirm the results obtained in the theoretical study.

2 Tools

In this section, we recall some definitions and results that are required for our study. RIP-based theory requires the introduction of a norm that is appropriate for measuring objects in $\Sigma_{k,\epsilon}$.

Definition 2.1 (Kernel, scalar product and norm)**.** *For finite signed measures over $\mathbb{R}^d$, the Hilbert structure induced by a kernel h (a smooth function from $\mathbb{R}^d \times \mathbb{R}^d$ to $\mathbb{R}$) is defined by the following scalar product between two finite signed measures π_1, π_2 on $\mathbb{R}^d$:*

$$\langle \pi_1, \pi_2 \rangle_h = \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} h(t_1, t_2)\, \mathrm{d}\pi_1(t_1)\, \mathrm{d}\pi_2(t_2). \quad (6)$$

We can consequently define

$$\|\pi_1\|_h^2 = \langle \pi_1, \pi_1 \rangle_h. \quad (7)$$

Remark that we have the relation

$$\|\pi_1 + \pi_2\|_h^2 = \|\pi_1\|_h^2 + 2\langle \pi_1, \pi_2 \rangle_h + \|\pi_2\|_h^2. \quad (8)$$

Measuring distances with the help of $\|\cdot\|_h$ can be viewed as measuring distances at a given "resolution" adjusted with the kernel h. Typically we use a Gaussian kernel $h(t_1,t_2) = e^{-\frac{\|t_1-t_2\|_2^2}{2\lambda^2}}$. The kernel accuracy (i.e. its ability to distinguish two spikes) increases with respect to λ.

We now define the Restricted Isometry Property (RIP). Recall that the secant set of the model set Σ is $\Sigma - \Sigma := \{x - y : x \in \Sigma, y \in \Sigma\}$.

Definition 2.2 (Restricted Isometry Property). *A has the RIP on $\Sigma - \Sigma$ with respect to $\|\cdot\|_h$ with constant γ if for all $x \in \Sigma - \Sigma$,*

$$(1-\gamma)\|x\|_h^2 \leq \|Ax\|_2^2 \leq (1+\gamma)\|x\|_h^2. \quad (9)$$

As mentioned in the introduction, this property implies the success of the non-convex decoder (2). It is also guaranteed that weighted Gaussian random Fourier measurements (i.e. Fourier measurements performed on frequencies drawn independently according to a Gaussian distribution) have this property if the number of measurements is large enough (see Section 3 for the explicit bound on m). This property originates from the fact that Gaussian random Fourier measurements have a mean kernel that is Gaussian in the following sense : for $x \in \Sigma - \Sigma$, $\mathbb{E}(\|Ax\|_2^2) \propto \|x\|_h^2$ (see e.g. Lemma 4.1).

To calculate derivatives of g, we need to weakly differentiate Dirac measures.

Definition 2.3 (Directional derivatives of Dirac measures). *Let $v \in \mathbb{R}^d$ with $\|v\|_2 = 1$. The distribution $\delta'_{t_0,v}$ is defined by $\langle \delta'_{t_0,v}, f\rangle = -f'_v(t_0)$. for $f \in \mathcal{C}^1(\mathbb{R}^d)$, where $f'_v(t_0)$ is the derivative of f at t_0 in the direction v (we also write $\partial_v f(t_0) := f'_v(t_0)$). It is the limit of $\nu_\eta = -\frac{\delta_{t_0+\eta v}-\delta_{t_0}}{\eta}$ for $\eta \to 0^+$ in the distributional sense: for all $h \in \mathcal{C}^1(\mathbb{R}^d)$, $\int_{\mathbb{R}} h(t)\,\mathrm{d}\nu_\eta(t) \to_{\eta\to 0^+} \langle \delta'_{t_0,v}, h\rangle$. We will also use the notation $\partial_v \delta_{t_0} = \delta'_{t_0,v}$.*

Similarly, the distribution $\delta''_{t_0,v}$ is defined by $\langle \delta''_{t_0,v}, f\rangle = f''_v(t_0)$ for $f \in \mathcal{C}^2(\mathbb{R}^d)$ (the second order derivative in direction v). The distribution δ''_{t_0,v_1,v_2} is defined by $\langle \delta''_{t_0,v_1,v_2}, f\rangle = f''_{v_1,v_2}(t_0)$ for $f \in \mathcal{C}^2(\mathbb{R}^d)$ where $f''_{v_1,v_2}(t_0)$ is the second-order derivative of f at t_0 in successive directions v_1 and v_2. We will also use the notation $\partial_v^2 \delta_{t_0} = \delta''_{t_0,v}$.

With this definition, we can calculate directional derivatives of elements of Σ with respect to their positions. For fixed amplitudes $a_1,\ldots,a_k$ and $t = (t_1,\ldots,t_k) \in \mathbb{R}^{kd}$ consider $\phi(t) = \sum_{i=1}^k a_i\delta_{t_i}$. For $u = (u_1,\ldots,u_k) \in \mathbb{R}^{kd}$, we have $\partial_u\phi(t) = \sum_{i=1}^k a_i \partial_{u_i}\delta_{t_i}$.

The secant set of $\Sigma_{k,\epsilon}$ is composed of sums of dipoles.

Definition 2.4 ((ϵ-)Dipole, separation). *An ϵ-dipole (noted dipole for simplicity) is a measure $\pi = a_1\delta_{t_1} - a_2\delta_{t_2}$ where $\|t_1 - t_2\|_2 \leq \epsilon$. Two dipoles $\pi_1 = a_1\delta_{t_1} - a_2\delta_{t_2}$ and $\pi_2 = a_3\delta_{t_3} - a_4\delta_{t_4}$ are ϵ-separated if their supports are strictly ϵ-separated (with respect to the ℓ^2-norm on $\mathbb{R}^d$), i.e. if $\|t_1-t_3\|_2 > \epsilon$, $\|t_2-t_3\|_2 > \epsilon$, $\|t_1-t_4\|_2 > \epsilon$ and $\|t_2-t_4\|_2 > \epsilon$.*

We can define additionally generalized dipoles that can be constructed as limits of dipoles.

Definition 2.5 (Generalized dipole). *A generalized dipole ν is either a dipole or a distribution of order 1 of the form $a_1\delta_t + a_2\delta'_{t,v}$. Two generalized dipoles are ϵ-separated if their support are strictly ϵ-separated (with respect to the ℓ^2-norm on $\mathbb{R}^d$).*

We recall the following Lemma from [10, Lemma 2.4]:

Lemma 2.1. *Suppose that for all two ϵ-separated dipoles π_1,π_2, $\langle\pi_1,\pi_2\rangle_h \leq \mu_h\|\pi_1\|_h\|\pi_2\|_h$ (where μ_h is called the mutual coherence). Then for k ϵ-separated generalized dipoles $\nu_1,\ldots,\nu_k$ such that $\max_i\|\nu_i\|_h > 0$, we have*

$$1-(k-1)\mu_h \leq \frac{\|\sum_{i=1}^k \nu_i\|_h^2}{\sum_{i=1}^k \|\nu_i\|_h^2} \leq 1+(k-1)\mu_h. \quad (10)$$

We use the following definition of subexponential random variable (taken from [14] which uses explicit constants compared to [15]).

Definition 2.6. *We say that a random variable is subexponential X with parameter $a > 0$ if $\mathbb{E}(X) = 0$ and for all $|s| \leq 1/a$,*

$$\mathbb{E}[e^{sX}] \leq e^{s^2a^2/2}. \quad (11)$$

*In that case, we denote $X \sim$ **subE**(a).*

If X is Gaussian, one can see that $X^2 - \mathbb{E}(X^2)$ is subexponential [14, 15]. Also, the following Bernstein concentration inequality holds for subexponential variables.

Theorem 2.1. *Consider $\mathcal{X} = (X_l)_{l=1}^m$ with $X_l \sim$ **subE**(a) i.i.d and $\eta > 0$. Let $\bar{\mathcal{X}} = \frac{1}{m}\sum_{l=1}^m X_l$ then*

$$\max\left(\mathbb{P}\left(\bar{\mathcal{X}} > \eta\right), \mathbb{P}\left(\bar{\mathcal{X}} < -\eta\right)\right) \leq e^{-\frac{m}{2}\min(\frac{\eta^2}{a^2}, \frac{\eta}{a})}. \tag{12}$$

As far as we know, there is no unified version of this theorem, and we give a proof of this result in Appendix A for the sake of completeness.

3 Lower bound of the size of strong basins of attraction

In this section we study strong basins of attraction by slightly modifying the result of [11]. As our goal is to study sufficient and *necessary* conditions on basins of attraction with respect to the number of measurements, we focus on the estimation of the positions as the behaviour with respect to amplitudes in x_0 is already well understood: the size of basins of attraction necessarily depends on the ratio $\frac{a_{min}}{a_{max}}$ between minimal and maximal amplitudes [10] and the gradient descent can be adaptively preconditioned to reduce this dependency [16] (shown in the deterministic low-pass filtering case).

From now on, we assume the amplitudes to be known and fixed at 1 and we consider

$$\tilde{\Sigma}_{k,\epsilon} := \{\phi(\theta) = \sum_{i=1}^{k} \delta_{t_i} \,:\, \theta = (t_1, \dots, t_k) \in \mathbb{R}^{kd}, \\ t_i \in \mathbb{R}^d, \forall i \neq l, \|t_i - t_l\|_2 > \epsilon, \|t_i\|_2 \leq R\} \tag{13}$$

where we write, by abuse of notation $\phi(t_1, \dots, t_k) = \phi(1, \dots, 1, t_1, \dots, t_k)$. We define accordingly $\tilde{\Theta}_{k,\epsilon} = \phi^{-1}(\tilde{\Sigma}_{k,\epsilon})$. We further simplify the analysis by considering the noiseless case $e = 0$.

We will show that a certain set is a strong basin of attraction by ensuring strong convexity within this set.

Definition 3.1. *Let $\Lambda \subset \mathbb{R}^D$ be convex. Let $f \in \mathcal{C}^2(\Lambda)$ and $\xi > 0$. We say that f is ξ-strongly convex if for all $\theta \in \Lambda$, the Hessian H_θ of f at θ satisfies*

$$H_\theta - \xi I \succcurlyeq 0. \tag{14}$$

Such a Λ will be called basin of strong convexity.

For a given step size $\tau > 0$, consider the gradient descent with fixed step size:

$$\theta_{n+1} = \theta_n - \tau \nabla g(\theta_n). \tag{15}$$

Previous work gave lower bounds on the size of g-basins of attraction [11].

Definition 3.2 (g-basin of attraction)**.** *We say that a set $\Lambda \subset \mathbb{R}^d$ is a g-basin of attraction of $\theta^* \in \Lambda$ if there exists $\tau_0 > 0$ such that for any $\tau \in (0, \tau_0]$, if $\theta_0 \in \Lambda$, then the sequence $g(\theta_n)$ with θ_n defined by (15) converges to $g(\theta^*)$ with $\theta_n \in \Lambda$.*

The strong convexity permits to show convergence of iterates of the gradient descent with a linear convergence rate, as recalled below.

Definition 3.3 (Linear convergence rate)**.** *We have a (at least) linear convergence rate of the iterates (θ_n) if there exist constants $n_0 \in \mathbb{N}$, $0 \leq r < 1$ and $C \geq 0$ such that for all $n \geq n_0$*

$$\|\theta_n - \theta^*\|_2^2 \leq C r^n, \tag{16}$$

with θ^* a minimizer of g.

If a convex set Λ is a g-basin of attraction where g is ξ-strongly convex, then it is a strong basin of attraction.

Definition 3.4 (Strong basin of attraction)**.** *We say that a set $\Lambda \subset \mathbb{R}^d$ is a strong basin of attraction of $\theta^* \in \Lambda$ if there exists $\tau_0 > 0$ such that for any $\tau \in (0, \tau_0]$, if $\theta_0 \in \Lambda$, then the sequence θ_n with θ_n defined by (15) converges to θ^* **with a linear convergence rate** with $\theta_n \in \Lambda$.*

Note that the stability of iterates in Λ must be ensured to obtain the convergence with a strong convexity argument.

Recall that $(Ax)_l = \langle x, \alpha_l \rangle$. As we will consider Fourier measurements, the α_l are infinitely differentiable sinusoids and the objective function g is a C^2 function.

We recall the calculation of the gradient and the Hessian of g from [11].

Proposition 3.1. *For any $\theta \in \mathbb{R}^{kd}$, and $u \in \mathbb{R}^{kd}$, we have*

$$\partial_u g(\theta) = 2\mathcal{R}e\langle A\partial_u \phi(\theta), A\phi(\theta) - y\rangle. \quad (17)$$

Proposition 3.2. *For any $\theta \in \mathbb{R}^{kd}$ and $u \in \mathbb{R}^{kd}$, the Hessian H_θ of g at θ verifies the following equality:*

$$u^T H_\theta u = 2\|A\partial_u\phi(\theta)\|_2^2 + 2\mathcal{R}e\langle A\partial_u^2\phi(\theta), A\phi(\theta) - y\rangle. \quad (18)$$

We recall the size of basins of attraction in our case as given by Corollary 2.1 of [11] with the specific notations of the present article and in the noiseless case. Basins of attraction with a simple ℓ^2-ball shape were considered: for $\beta > 0$, we define the sets

$$\Lambda_\beta := \{\theta \in \mathbb{R}^{kd} : \|\theta - \theta^*\|_2 < \beta\}. \quad (19)$$

Theorem 3.1. *Suppose A has the RIP on $\tilde{\Sigma}_{k,\epsilon} - \tilde{\Sigma}_{k,\epsilon}$ with constant γ and $e = 0$. Let $\theta^* = (t_1, \dots, t_k) \in \tilde{\Theta}_{k,\epsilon}$ be a solution of the constrained Problem (5). Let $\beta_1 > 0$ such that*

1. *$\theta \in \Lambda_{2\beta_1}$ implies $\phi(\theta) \in \Sigma$ (local stability of the model set) and for all $\tilde{\theta} \in \tilde{\Theta}_{k,\epsilon}$ such that $\phi(\tilde{\theta}) = \phi(\theta^*)$ we have $\|\tilde{\theta} - \theta\|_2 > \|\theta^* - \theta\|_2$ (uniqueness of the projection of θ on the set of minimizers);*
2. *there is $C_{\phi,\theta^*} > 0$ such that*
$$\forall \theta \in \Lambda_{2\beta_1}, \quad \|\phi(\theta) - \phi(\theta^*)\|_h \le C_{\phi,\theta^*}\|\theta - \theta^*\|_2 \quad (20)$$
(local control of $\|\cdot\|_h$);
3. *the first-order derivatives of $A\phi$ are uniformly bounded on $\phi^{-1}(\phi(\theta^*))$:*
$$M_1 := \sup_{\theta \in \phi^{-1}(\phi(\theta^*))} \sup_{\|u\|_2 = 1} \|A\partial_u\phi(\theta)\|_2 < +\infty; \quad (21)$$
4. *the second-order derivatives of $A\phi$ are uniformly bounded on $\Lambda_{2\beta_1}$:*
$$M_2 := \sup_{\theta \in \Lambda_{2\beta_1}} \sup_{\|u\|_2=1, \|v\|_2=1} \|A\partial_v\partial_u\phi(\theta)\|_2 < +\infty; \quad (22)$$
5. *we have*
$$\beta_2 := \frac{(1-\gamma)}{C_{\phi,\theta^*}\sqrt{1+\gamma}} \inf_{\theta \in \Lambda_{\beta_1}} \left(\frac{\|\partial_{\theta^*-\theta}\phi(\theta)\|_h^2}{\|A\partial_{\theta^*-\theta}^2\phi(\theta)\|_2} \right) > 0. \quad (23)$$

Then $\Lambda_{\min(\beta_1,\beta_2)}$ is a g-basin of attraction of θ^*.

We can now give explicit lower bounds of the size of strong basins of attraction for $\tilde{\Sigma}_{k,\epsilon}$ and weighted Fourier measurements for which a RIP was shown in [9] (This ensures that the hypotheses of the following Theorem can be satisfied, see dicussion after the Theorem).

Theorem 3.2. *For $l = 1, \dots, m$, let $\omega_l \in \mathbb{R}^d$. Suppose*

$$\alpha_l(s) = \frac{1}{\sqrt{m}} c(\omega_l) e^{-j\langle \omega_l, s\rangle} \quad (24)$$

with $c(\omega_l) = \left(1 + \frac{\lambda^2\|\omega_l\|_2^2}{d}\right)^{-1}$. Consider $\|\cdot\|_h$ with $h(s_1, s_2) = e^{-\frac{\|s_1 - s_2\|_2^2}{2\lambda^2}}$ and suppose that it has mutual coherence μ_h.

Suppose A has the RIP on $\tilde{\Sigma}_{k,\epsilon/2} - \tilde{\Sigma}_{k,\epsilon/2}$ with constant γ and $e = 0$.

Let $\theta^* = (t_1, \dots, t_k) \in \tilde{\Theta}_{k,\epsilon}$ be a solution of the constrained Problem (5). Then

- *g has L-Lipschitz gradient on $\Lambda_{\epsilon/2} := \{\theta : \|\theta - \theta^*\|_2 < \epsilon/2\}$ with $L < +\infty$.*
- *Consider $\Lambda_{\beta_{spikes}} := \{\theta : \|\theta - \theta^*\|_2 < \beta_{spikes}\}$ where*
$$\beta_{spikes} := \min\left(\frac{\epsilon}{4}, \frac{(1-\gamma)(\lambda(1-(k-1)\mu_h) - \lambda^3\xi/2)}{\sqrt{1+\gamma}\, d\sqrt{(1+(k-1)\mu_h)}}\right), \quad (25)$$
for some $\xi < \min(2(1-(k-1)\mu_h)/\lambda^2, L/2)$. Then $\Lambda_{\beta_{spikes}}$ is a strong basin of attraction of θ^* with linear convergence rate $r = 1 - \tau\xi$ for step size $\tau \le \frac{1}{L}$.

- *We have that*

$$\|\theta_n - \theta^*\|_2^2 \leq \|\theta_0 - \theta^*\|_2^2 r^n, \tag{26}$$

and

$$\|\phi(\theta_n) - \phi(\theta^*)\|_h^2 \leq \frac{g(\theta_0) - g(\theta^*)}{1-\gamma} r^n. \tag{27}$$

Proof See Appendix B. □

The size of basins given by Theorem 3.2 is just obtained by expliciting constants from Theorem 3.1 and including the strong convexity constant ξ. The bigger this constant, the smaller the size of the strong basin. The smaller this constant is, the closer it is to the size of basins of attraction given by Theorem 3.1.

We now give the relation between this theorem and the number of measurements m. It has been shown in [9] that there is a universal constant C such that for any $0 < p < 1$, the operator A (with $\alpha_l(s) = \frac{1}{\sqrt{m}} c(\omega_l) e^{j\langle \omega_l, s\rangle}$, $c(\omega) = \frac{1}{1+\|\omega\|_2^2/d}$ and ω_l randomly drawn with a density proportional to $\frac{1}{c(\omega)^2} e^{-\|\omega\|_2^2/(2\lambda^2)}$) has RIP with constant γ with probability $1-p$ with a choice

$$m = C\frac{1}{\gamma^2}\Big(k^2 d \cdot (1 + \log(kd) + \log(\frac{R}{\epsilon})) + \log\frac{1}{\gamma} + k\log(\frac{1}{p})\Big). \tag{28}$$

Hence for a fixed p, for $m \to \infty$

$$\gamma = \Omega\left(\frac{1}{\sqrt{m}}\right), \tag{29}$$

where $\Omega(v_m)$ refers to a sequence u_m such that $u_m = \mathcal{O}(v_m)$ and $v_m = \mathcal{O}(u_m)$ when $m \to \infty$. Using Theorem 3.2, this means that given a probability $1-p$, we guarantee a size of strong basin of attraction

$$\min\left(\frac{\epsilon}{4}, C_0\left(1 - \Omega\Big(\frac{1}{\sqrt{m}}\Big)\right)\right) \tag{30}$$

where C_0 is a constant that does not depend on m.

Note that a recent work [17] shows that the RIP for Gaussian measurements without weighting, i.e. $c(\omega) = 1$ but only in a regime where γ is not too close to 0, which is exactly the regime that we study in the next section, i.e. we cannot use these simpler measurements without weights for our analysis.

Remark 3.1. *The dependency of C_0 on $1/d$ might be avoided under a stronger RIP assumption (as observed in the experimental Section 5). This dependency comes from the upper bound calculated on $\|A\partial_u^2\phi(\theta)\|$. It can be shown [10] that the RIP on $\Sigma_{k,\epsilon} - \Sigma_{k,\epsilon}$ permits to use a RIP on $\partial_u\phi(\theta)$. Similarly a RIP on $(\Sigma_{k,\epsilon} - \Sigma_{k,\epsilon}) - (\Sigma_{k,\epsilon} - \Sigma_{k,\epsilon})$ would imply a RIP on $\partial_u^2\phi(\theta)$ leading to the following bound:*

$$\|A\partial_u^2\phi(\theta)\|_2 \leq \sqrt{1+\gamma}\sqrt{1+(k-1)\mu_h}, \tag{31}$$

hence removing the dependency on d. We conjecture that such a RIP is verified for random Gaussian Fourier measurements under a similar bound on the number of measurements: from [9], it suffices to verify that the kernel has a mutual coherence property with respect to ϵ-separated quadripoles and that the normalized set $\{x/\|x\|_h : x \in (\Sigma_{k,\epsilon} - \Sigma_{k,\epsilon}) - (\Sigma_{k,\epsilon} - \Sigma_{k,\epsilon})\}$ has finite covering numbers.

In the case of random Gaussian measurements we further conjecture that the dependency on d might be avoided with high probability with the same RIP assumption, since we could use a concentration argument to bound $\|A\partial_u^2\phi(\theta)\|_2$. We verify numerically that this is the case in Section C.

We obtain in the next section an upper bound of the size of the basin that have the same growth with respect to m.

4 An upper bound of the size of basins of strong convexity

Let us first make a remark about the mean kernel h. The random operator A is designed to approximate a mean Gaussian kernel when the number of measurements grows, as shown in the following Lemma.

Lemma 4.1. *Assume that $\alpha_l(t) = \frac{1}{\sqrt{m}} c(\omega_l) e^{-j\langle \omega_l, t\rangle}$ where $c(\omega_l)$ is a positive function of ω_l, and where the $(\omega_l)_{l=1,\dots,m}$ are* *i.i.d. random variables in* $\mathbb{R}^d$ *with density* $p(\omega) = \frac{D_c(\lambda)}{c(\omega)^2} e^{-\frac{\lambda^2 \|\omega\|_2^2}{2}}$*, and* $D_c(\lambda)$ *is a normalizing constant. Then, for any* $t \in \mathbb{R}^d$,

$$\mathbb{E}\left(\|A\delta_t - A\delta_0\|_2^2\right) = 2D_c(\lambda)\frac{(2\pi)^{\frac{d}{2}}}{\lambda^d}\left(1 - e^{-\frac{\|t\|_2^2}{2\lambda^2}}\right) = \frac{(2\pi)^{\frac{d}{2}}}{\lambda^d} D_c(\lambda)\|\delta_t - \delta_0\|_h^2. \tag{32}$$

Proof See Section D. □

The function $t \mapsto 1 - e^{-\|t\|_2^2/(2\lambda^2)}$ has a unique global minimum on $\mathbb{R}^d$. However it is convex only for $\|t\|_2 \leq \lambda$. As our proof of the upper bound of strong basins of attraction is based on a strong convexity argument, we notice that even in the ideal mean case of the recovery of one Dirac measure, there is a fundamental (pessimistic) limit that we cannot overcome. However, it must be noted that in the case of multiple Dirac measures the minimal separation ϵ between positions is also a fundamental upper limit for the size of the basin of attraction (as shown by the shape of the bound from Theorem 3.2). This separation also forces the choice of a kernel (i.e. a choice of λ) such that $\|\delta_t - \delta_0\|_h^2$ is small for $\|t\|_2 > \epsilon$. It has also been shown practically in [18] that λ^2 must not be too small (i.e. frequencies chosen too high). This can potentially be explained by the fact that g becomes very flat almost everywhere except in a very tight neighborhood of the global optimum. Hence, λ is generally chosen as a multiple of ϵ, and Figure 1 shows that studying local (strong) convexity cannot explain the situation for t between λ and ϵ, although the size of the set where convexity is obtained is still a multiple of ϵ (which is an upper bound on the size of basins of attraction).

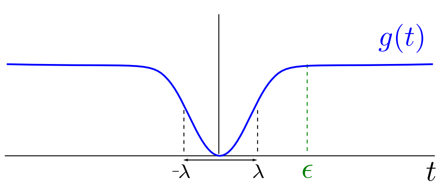


Fig. 1: A representation of $g(t) = 2(1 - e^{-\frac{t^2}{2\lambda^2}})$ for the case of one Dirac measure. The parameter λ is generally chosen so that it is a multiple of ϵ. The function is convex on $[-\lambda, \lambda]$.

In order to find a meaningful upper bound of strong basins of attraction with respect to the number of measurements m, these considerations suggest that only the behaviour in sets smaller than the minimum separation matters, i.e. sets where the interaction between Dirac masses is small.

Considering the same operator A as in the previous section, we study the case of the recovery of 1 spike (i.e. the most favorable case for the size of strong basins of attractions). In this case we have that the parameter $\theta = t \in \mathbb{R}^d$ is the position of the spike.

As the problem is invariant by translation, we consider the recovery of $x_0 = \phi(\theta^*) = \phi(0) = \delta_0$ (in the noiseless case). We start with the expression of the Hessian.

Lemma 4.2. *Let* $\phi(\theta) = \phi(t) = \delta_t$, $\phi(\theta^*) = \delta_0$ *and* $\alpha_l(s) = \frac{1}{\sqrt{m}} c(\omega_l) e^{j\langle \omega_l, s\rangle}$. *For* $\|u\|_2 = 1$,

$$\frac{1}{2} u^T H_\theta u = \frac{1}{m}\sum_{l=1}^m c(\omega_l)^2 |\langle \omega_l, u\rangle|^2 \cos(\langle \omega_l, t\rangle). \tag{33}$$

Proof From Proposition 3.2, we have

$$u^T H_\theta u = 2\|A\partial_u \phi(\theta)\|_2^2 + 2\mathcal{R}e\langle A\partial_u^2\phi(\theta), A(\phi(\theta) - \phi(\theta_0))\rangle. \tag{34}$$

The first term is

$$\begin{aligned}\|A\partial_u\phi(\theta)\|_2^2 &= \sum_{l=1}^m |\langle \partial_u \delta_t, \alpha_l\rangle|^2 \\ &= \sum_{l=1}^m |\partial_u \alpha_l(t)|^2 \\ &= \frac{1}{m}\sum_{l=1}^m |\langle \omega_l, u\rangle e^{-j\langle \omega_l, t\rangle}|^2 c(\omega_l)^2 \\ &= \frac{1}{m}\sum_{l=1}^m |\langle \omega_l, u\rangle|^2 c(\omega_l)^2.\end{aligned} \tag{35}$$

The second term is

$$\begin{aligned}&\mathcal{R}e\langle \partial_u^2 A\phi(\theta), A(\phi(\theta) - \phi(\theta_*))\rangle \\ &= \sum_{l=1}^m c(\omega_l)^2 \mathcal{R}e((\overline{\partial_u^2 \alpha_l(t)})(\alpha_l(t) - \alpha_l(0))) \\ &= \frac{1}{m}\sum_{l=1}^m c(\omega_l)^2 \mathcal{R}e\Big(-|\langle \omega_l, u\rangle|^2 e^{-j\langle \omega_l, t\rangle}(e^{j\langle \omega_l, t\rangle} - 1)\Big) \\ &= -\frac{1}{m}\sum_{l=1}^m c(\omega_l)^2 |\langle \omega_l, u\rangle|^2 (1 - \cos(\langle \omega_l, t\rangle)).\end{aligned} \tag{36}$$

We get

$$\frac{1}{2}u^T H_\theta u = \frac{1}{m}\sum_{l=1}^{m} c(\omega_l)^2 |\langle \omega_l, u\rangle|^2 (1 - 1 + \cos(\langle \omega_l, t\rangle))$$
$$= \frac{1}{m}\sum_{l=1}^{m} c(\omega_l)^2 |\langle \omega_l, u\rangle|^2 \cos(\langle \omega_l, t\rangle). \tag{37}$$

□

The remarkable phenomenom in this calculation of the Hessian is that the whole term $\|A\partial_u \phi(\theta)\|_2^2$ is compensated by the second term. Hence no lower bound is needed on this term unlike the proof of Theorem 3.2. Consequently, Theorem 3.2 might not be very tight when applied to the favorable case of one spike. We will see in the following how well the qualitative behaviour of the size of basins is described by Theorem 3.2.

Take $h \in \mathbb{R}_+, u \in \mathbb{R}^d$ such that $\theta = t = hu$ and $\|u\|_2 = 1$. We define

$$f(u,h) := \frac{1}{2}u^T H_\theta u$$
$$= \frac{1}{m}\sum_{l=1}^{m} c(\omega_l)^2 |\langle \omega_l, u\rangle|^2 \cos(\langle \omega_l, u\rangle h)$$
$$= \frac{1}{m}\sum_{l=1}^{m} X_l \tag{38}$$

with

$$X_l := X_l(u,h) := c(\omega_l)^2 |\langle \omega_l, u\rangle|^2 \cos(\langle \omega_l, u\rangle h). \tag{39}$$

Remark 4.1. *Let us assume just for a moment that $c(\omega_l) = 1$ (which is not the case in the rest of the paper). Since $\|u\|_2 = 1$, $c(\omega_l) = 1$, we have*

$$f(u,h) = \frac{1}{m}\sum_{l=1}^{m} Z_l^2 \cos(hZ_l) = \frac{1}{m}\sum_{l=1}^{m} X_l, \tag{40}$$

where $Z_l := \langle \omega_l, u\rangle$ follows a Gaussian distribution $\mathcal{N}(0, 1/\lambda^2)$ and $X_l = Z_l^2 \cos(hZ_l)$.

We will use a concentration inequality on $f(u,h)$ in order to give an upper bound for the size of strong basins of attraction of δ_0 with high probability. We need to calculate its mean with respect to the distribution of ω.

Lemma 4.3. *Let $Y = Z^2\cos(hZ)$ with $Z \sim \mathcal{N}(0, 1/\lambda^2)$. Then*

$$\mathbb{E}(Y) = F(h) := \frac{(\lambda^2 - h^2)e^{-\frac{h^2}{2\lambda^2}}}{\lambda^4}. \tag{41}$$

Proof See Section D. □

We deduce the mean of the considered random variable with the weighting scheme included.

Lemma 4.4. *Consider the random variable $X = c^2(\omega_l)|\langle \omega_l, u\rangle|^2 \cos(h\langle \omega_l, u\rangle)$ where $\|u\|_2 = 1$ and ω_l is a random variable with density $p(\omega) = \frac{D_c(\lambda)}{c(\omega)^2}e^{-\lambda^2\|\omega\|_2^2}$. Then*

$$\mathbb{E}(X) = \frac{D_c(\lambda)}{D(\lambda)}F(h) = \frac{D_c(\lambda)}{D(\lambda)}\frac{(\lambda^2 - h^2)e^{-\frac{h^2}{2\lambda^2}}}{\lambda^4}, \tag{42}$$

where $D_c(\lambda)^{-1} = \int_{\mathbb{R}^d} e^{-\frac{\lambda^2}{2}\|\omega\|_2^2} c(\omega)^{-2} d\omega$ and $D(\lambda)^{-1} = \int_{\mathbb{R}^d} e^{-\frac{\lambda^2}{2}\|\omega\|_2^2} d\omega$.

Proof See Section D. □

The behaviour of the expected value of the second derivative is the same with or without the weights $c(\omega)$ (up to a constant). In the mean case (approximated by our random observations), we do not have convexity over all the domain.

To use the concentration inequality we will use the fact that the Hessian is a sum of subexponential random variables.

Lemma 4.5. *Let $X = c^2(\omega)|\langle \omega, u\rangle^2|\cos(h\langle \omega, u\rangle)$ where $\|u\|_2 = 1$ and ω has density $p(\omega) = \frac{D_c(\lambda)}{c(\omega)^2}e^{-\lambda^2\|\omega\|_2^2/2}$ and $0 < c(\omega) \leq 1$. Then $X - \mathbb{E}(X) \in \mathbf{subE}(8/\lambda^2)$.*

Proof See Section D. □

The hypothesis on the weights of this Lemma is verified for $c(\omega) = \frac{1}{1+\frac{\lambda^2}{d}\|\omega\|_2^2}$. The application of the Bernstein inequality for subexponential variables leads to the following concentration inequality.

Lemma 4.6. *Consider $\mathcal{X} = (X_l)_{l=1}^m$ with i.i.d variables $X_l = c(\omega_l)^2|\langle \omega_l, u\rangle|^2 \cos(h\langle \omega_l, u\rangle)$ with*

ω_l *with density* $p(\omega) = \frac{D_c(\lambda)}{c(\omega)^2} e^{-\frac{\lambda^2}{2}\|\omega\|_2^2}$ *and* $0 < c(\omega_l) \leq 1$*. Let* $\bar{\mathcal{X}} = \frac{1}{m}\sum_{l=1}^m X_l$*. Then for any* $\eta > 0$,

$$\max\left(\mathbb{P}\left(\bar{\mathcal{X}} - \mathbb{E}X_l > \eta\right), \mathbb{P}\left(\bar{\mathcal{X}} - \mathbb{E}X_l < -\eta\right)\right) \leq e^{-\frac{m}{2}\min(\eta^2\frac{\lambda^4}{8^2}, \eta\frac{\lambda^2}{8})}. \tag{43}$$

Proof From Lemma 4.5, $X - \mathbb{E}X \in \mathbf{subE}(8/\lambda^2)$. Applying Theorem 2.1 with $a = 8/\lambda^2$ yields the result. □

Consider $\Gamma(E, \xi)$ the set of functions $\mathcal{C}^2$ and ξ-strongly convex on a convex set E. We can now give an upper bound on the size of sets where g is strongly convex.

Theorem 4.1 (Upper bound of basins of strong convexity)**.** *Let* u *such that* $\|u\|_2 = 1$*. Given* $\xi > 0$*, there is* $h > 0$ *and* $m_0 \in \mathbb{N}$ *such that for any* $m \geq m_0$

$$h^2 \leq \lambda^2\left(1 - \frac{D(\lambda)}{D_c(\lambda)}\frac{8\sqrt{2}}{\sqrt{m}}\right) \tag{44}$$

and

$$\mathbb{P}\left(g \in \Gamma([0, uh], \xi)\right) \leq \frac{1}{e} \approx 0.37 \quad . \tag{45}$$

Proof See Section D. □

This lemma states that g cannot be strongly convex on basins of attraction of size larger that $\sqrt{\lambda^2\left(1 - \frac{D(\lambda)}{D_c(\lambda)}\frac{8\sqrt{2}}{\sqrt{m}}\right)} = \lambda\left(1 - O(\frac{1}{\sqrt{m}})\right)$ with high probability, i.e. with a probability close to 1. Asymptotically, this matches the rate of the lower bound given by Theorem 3.2. We observe in the next experimental section that this behaviour is still a good approximation of the size of basins for small m.

For large m, the most important comparison is the asymptotic behaviour of our bounds on the size of strong basins of attraction. The upper bound converges to λ as expected. The lower bound from Theorem 3.2 converges to $\min\left(\epsilon/4, \frac{\lambda(1-(k-1)\mu_h - \lambda^3\xi/2)}{d\sqrt{1+(k-1)\mu_h}}\right) \approx \frac{\lambda}{d}$ for small ξ and μ_h and sufficiently small λ. Indeed the parameter λ is typically chosen to be smaller that ϵ, which in turn guarantees a small mutual coherence μ_h of the kernel. As mentioned in the previous section, the dependency on d might be an artifact of a suboptimal bound in the proof of Theorem 3.2 for the particular case of random Fourier sampling (as shown by experiments in the next section).

In terms of tightness of our bounds with respect to m, the only question left is whether the necessary condition of order $\lambda\sqrt{1 - O(1/\sqrt{m})}$ can be improved to $\tilde{\lambda}\sqrt{1 - O(1/\sqrt{m})}$ with $\tilde{\lambda} < \lambda$ (the previous comparison with Theorem 3.2 suggests it cannot). In the case of one spike, we show that the upper bound we gave is indeed tight in this sense. To give a lower bound, we use an ϵ-net argument to show strong convexity with high probability in the relevant set.

Lemma 4.7 (Lower bound of basins of strong convexity)**.** *There exists a constant* C *such that, given an arbitrary high probability* $1 - p_0$*, we can find* h *such that*

$$h \geq \sqrt{\lambda^2 - \frac{C\sqrt{-\log(p_0)}}{\sqrt{m}}}. \tag{46}$$

and

$$\mathbb{P}(g \in \Gamma(\Lambda_h)) \geq 1 - p_0. \tag{47}$$

Proof See Section D. □

We summarize the different bounds obtained by our results in Figure 2.

5 Experiments

In this section, for random Gaussian Fourier measurements, we provide experiments on the size of basins of attraction with respect to the different parameters of the problem: the number of measurements, the dimension d of the support, the choice of frequencies (and induced kernel) and the minimum separation between spikes. We observe that the experiments match the theoretical results of the previous Sections.

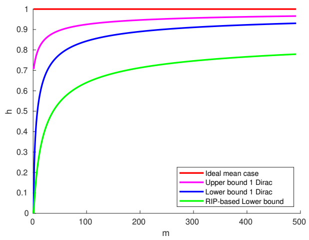


Fig. 2: A schematic representation of the shape of the curves obtained by our theorems for $\lambda = 1$. Red: $h = 1$ (exact bound mean case), Magenta: $h = \sqrt{1 - 1.5/\sqrt{m}}$ (one spike upper bound), Blue: $h = \sqrt{1 - 3/\sqrt{m}}$ (one spike lower bound), Green: $h = 0.9 \cdot (1 - 3/\sqrt{m})$ (ideal RIP-based lower bound). The chosen constants are for illustration purpose only and are not originated from our Theorems. True sizes of basins of attraction are investigated experimentally in the next Section.

5.1 Method and parameters for experiments

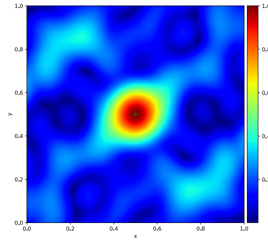
Given a ground truth $x_0 = \phi(\theta^*)$ and a set of parameters, we repeat the following experiment 10 000 times: we perform a gradient descent with line search (for faster practical convergence) and with a random initialization θ_{init} and we record whether the descent has converged to θ^* or not with threshold 0.01 on the localization error $\|\theta_{nit_{max}} - \theta^*\|_2$ after a fixed number of iterations (500). We calculate the size of the basin of attraction as the smallest distance β such that 99.8% of experiments that did not converge to θ^* were initialized outside of Λ_β, i.e. initialization in Λ_β yields convergence with a high probability.

The frequencies of the Gaussian measurements follow a normal distribution $\mathcal{N}(0, \frac{1}{\lambda^2} I)$. To illustrate how well the measurements approximate the mean Gaussian kernel as their number increases, we show the back-projection of the measurements (i.e. the signal $z(t) = \sum_{l=1}^m y_l \alpha_l(t)$) of one spike on a grid in Figure 3.

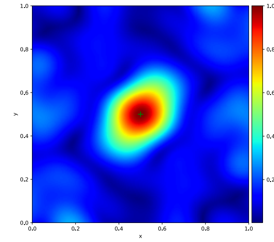
5.1.1 Effect of the dimension

To study the effect of the dimension d, we use the signal $x_0 = \delta_0$ and take d from $\{1, 2, 3\}$. We initialize randomly θ_{init} with

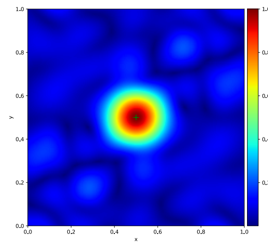
$$\theta_{init} \in \{t \in \mathbb{R}^d, 0.01 \leq \|t\|_2 \leq 0.5\}. \tag{48}$$



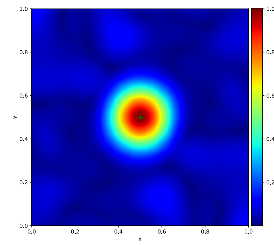
(a) $m = 32$



(b) $m = 64$



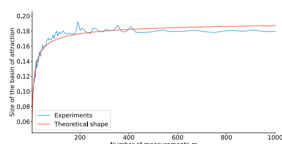
(c) $m = 128$



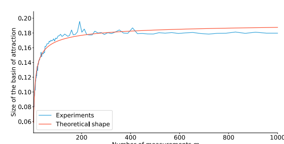
(d) $m = 512$

Fig. 3: Back-projection of the signal $x_0 = \delta_t$ observed with m random Fourier measurements. We observe that the measurement operator converges to a Gaussian kernel as m increases.

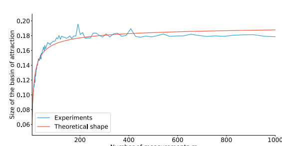
We obtain the results shown in Figure 4.



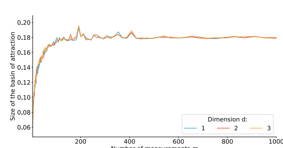
(a) $d = 1$ and theoretical curve



(b) $d = 2$ and theoretical curve



(c) $d = 3$ and theoretical curve



(d) $d = 1, 2, 3$

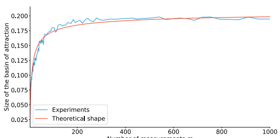
Fig. 4: Size of the basin of attraction with respect to the number of measurements m with various dimensions d. The theoretical curve is obtained by a least squares fitting of the function $C_1\sqrt{1 - C_2/\sqrt{m}}$ parametrized by C_1 and C_2.

We observe that the size of the basins of attraction does not depend on d as shown in

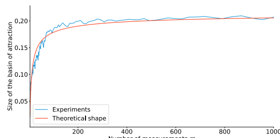
Section 4. This also shows the suboptimality (due to generality) of Theorem 3.2 with respect to d.

5.1.2 Effect of the number of spikes

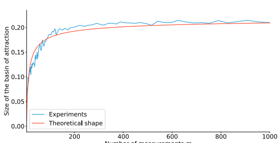
We verify in this section that the behaviour with respect to m does not change with respect to the number of spikes k. We use the signal $x_0 = \sum_{i=1}^{k} \delta_{t_i}$ with $\{t_1, \ldots, t_k\}$ the k-th roots of the unit circle of radius $r = 1$ in order to set a constant distance between spikes. We perform the experiment for $k \in \{2, 3, 4, 5\}$. We initialize a signal $x_{init} = \sum_{i=1}^{k} \delta_{s_i}$ such that for all $i = 1, \ldots, k$, $0.01 \leq \|t_i - s_i\|_2 \leq 0.5$. To evaluate the distance between x_0 and the estimated x^*, we take $\max_{i=1,\ldots,k} \|t_i^* - t_i\|_2$. We obtain the results shown in Figure 5. We observe that the size of basins behave similarly for various k. There is a small mismatch for small m as expected given the asymptotic nature of our result.



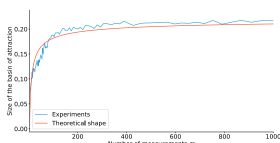
(a) $k = 2$ and theoretical curve



(b) $k = 3$ and theoretical curve



(c) $k = 4$ and theoretical curve



(d) $k = 5$ and theoretical curve

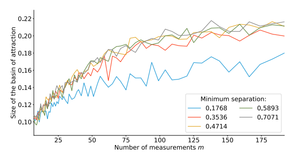
Fig. 5: Size of the basin of attraction with respect to the number of measurements m with different number of spikes k. The theoretical curve is obtain by a least squares fitting of the function $C_1\sqrt{1 - C_2/\sqrt{m}}$ parametrized by C_1 and C_2.

5.1.3 Effect of the minimum separation and of the choice of measurements

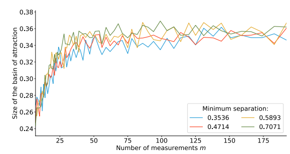
We show in Figure 6 the effect of the minimum separation between spikes on the size of basins of attraction, for 5 spikes with a given configuration. We change the separation parameter by simply rescaling the positions of the spikes.

As expected the size of basins of attraction increases with respect to the minimal separation, but only up to a limit: this experiment shows the impact of the choice of λ (the parameter that sets the choice of the distribution of random frequencies). In order to reach a size of basin of attraction close to the minimum separation, it should be set accordingly.

This is confirmed by Figure 7, where we repeat experiments with different λ with a fixed minimum separation $\epsilon = 0.5$. This confirms our theoretical results: the size of basins is increasing with respect to λ, up to a limit corresponding to the minimum separation.



(a) $\lambda = 0.1$



(b) $\lambda = 0.2$

Fig. 6: Size of the basin of attraction with respect to the minimum separation ϵ.

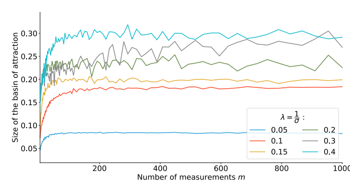


Fig. 7: Size of the basin of attraction with respect to the choice of the resolution parameter λ. Minimum separation is $\epsilon = 0.5$.

6 Conclusion

In this article, we gave an asymptotic upper bound and lower bounds on the size of basins of strong convexity (and also are strong basins of attractions) for the non-convex sparse spike estimation problem with random Fourier measurements with respect to the number of measurements. These two bounds have a similar shape that matches non-asymptotic experiments (similarly as phase transitions for conventional sparse recovery in finite dimension). This work highlights that the choice of measured frequencies is critical to obtain basins of attraction as large as possible.

Future work could be a non-asymptotic study of such bounds to confirm the observed behaviour in our experiments. While we cannot expect much from a direct extension to k spikes of our sufficient condition by the direct analysis of the Hessian in the case of one spike and random Fourier measurements, such a result could be useful for the study of measurement operators where the RIP is not well suited (such as structured operators).

Acknowledgments

Y. Traonmilin acknowledges the support of the French Agence Nationale de la Recherche (ANR) under reference ANR-20-CE40-0001 EFFIREG. Experiments presented in this paper were carried out using the PlaFRIM experimental testbed, supported by Inria, CNRS (LABRI and IMB), Université de Bordeaux, Bordeaux INP and Conseil Régional d'Aquitaine (see https://www.plafrim.fr).

A Proof of Bernstein inequality for subexponential variables (Theorem 2.1)

In this section, we prove Theorem 2.1. For $l = 1, \dots, m$, let $X_l \sim \mathbf{subE}(a)$ i.i.d and let $\eta > 0$. Let $\bar{\mathcal{X}} = \frac{1}{m}\sum_{l=1}^m X_l$.

First notice that, by symmetry, $-X_l \sim \mathbf{subE}(a)$ so we only need to prove

$$\mathbb{P}\left(\bar{\mathcal{X}} > \eta\right) \leq e^{-\frac{m}{2}\min(\frac{\eta^2}{a^2},\frac{\eta}{a})}. \tag{49}$$

Using Chernoff inequality and the fact that the X_l are i.i.d, we get that

$$\begin{aligned} \forall s \geq 0, \quad \mathbb{P}\left(\bar{\mathcal{X}} > \eta\right) &\leq e^{-s\eta}\mathbb{E}\left(e^{s\bar{\mathcal{X}}}\right) \\ &= e^{-s\eta}\prod_{l=1}^m \mathbb{E}\left(e^{\frac{s}{m}X_l}\right). \end{aligned} \tag{50}$$

Now, using that $X_l \sim \mathbf{subE}(a)$, we get that for $\frac{s}{m} \in [0, \frac{1}{a}]$,

$$\mathbb{P}\left(\bar{\mathcal{X}} > \eta\right) \leq e^{-s\eta}\left(e^{\frac{s^2}{m}\frac{a^2}{2}}\right) \leq e^{-s\eta + \frac{s^2a^2}{2m}}. \tag{51}$$

Optimizing this bound consists in finding the minimum of $f(s) = \frac{s^2}{m}\frac{a^2}{2} - s\eta$ on $[0, \frac{m}{a}]$. It is clear that f admits a global minimum on $\mathbb{R}$ attained at $s_* = \frac{\eta m}{a^2}$ and

$$f(s_*) = -\frac{\eta^2 m}{2a^2}. \tag{52}$$

Then we just have to distinguish the cases $s_* \leq \frac{m}{a}$ or $> \frac{m}{a}$. Notice that

$$\begin{aligned} s_* \leq \frac{m}{a} \quad &\iff \quad \frac{\eta}{a^2} \leq \frac{1}{a} \quad \iff \quad \frac{\eta^2}{a^2} \leq \frac{\eta}{a} \\ &\iff \quad \frac{\eta}{a} \leq 1. \end{aligned} \tag{53}$$

Therefore, if $\frac{s_*}{m} \leq \frac{1}{a}$,

$$\min_{[-\frac{m}{a},\frac{m}{a}]} f = f(s_*) = -\frac{m}{2}\frac{\eta^2}{a^2} = -\frac{m}{2}\min\left(\frac{\eta^2}{a^2},\frac{\eta}{a}\right), \tag{54}$$

and if $\frac{s_*}{m} > \frac{1}{a}$, since $f(0) = 0$, we have

$$\begin{aligned} \min_{[-\frac{m}{a},\frac{m}{a}]} f = f\left(\frac{m}{a}\right) &= \frac{m}{2} - \frac{m}{a}\eta \\ &\leq -\frac{m}{2}\frac{\eta}{a} \\ &= -\frac{m}{2}\min\left(\frac{\eta^2}{a^2},\frac{\eta}{a}\right). \end{aligned} \tag{55}$$

Gathering both cases, we get (49).

B Proof of Theorem 3.2

Note that the main argument is a standard proof of convergence under a strong convexity assumption. However, we must verify that each step of the argument is valid in our setting. We also give explicit constants.

We have that

$$g(\theta) = \sum_{l=1}^m \left|\sum_{i=1}^k e^{-j\langle\omega_l,t_i\rangle} - y_l\right|^2 \tag{56}$$

i.e. g is a sum of m $\mathcal{C}^\infty$ functions. It is thus $\mathcal{C}^\infty$.

First note that in [10], the hypotheses of Theorem 3.1 were shown to be fulfilled in our case with $\beta_1 = \epsilon/4$, because $\phi(\Lambda_{2\beta_1}) \subset \Sigma_{k,\epsilon}$, and for some constants C_{ϕ,θ^*} (that will be made explicit below) and

$$\beta_2 = \frac{(1-\gamma)}{C_{\phi,\theta^*}\sqrt{1+\gamma}} \inf_{\theta\in\Lambda_{\beta_1}} \left(\frac{\|\partial_{\theta^*-\theta}\phi(\theta)\|_h^2}{\|A\partial^2_{\theta^*-\theta}\phi(\theta)\|_2} \right) > 0. \quad (57)$$

Under these hypotheses and with $\beta = \min(\beta_1, \beta_2)$, it is guaranteed in [11, Proof of Theorem 2.8] that g has L-Lipschitz gradient on $\Lambda_{2\beta_1}$ with $L < +\infty$. It is also guaranteed that for a step size $\tau \le \tau_0 = \frac{1}{L}$, the iterates (θ_n) of the gradient descent (15) initialized in Λ_β satisfy:

- the value of the function is decreasing:

$$\begin{aligned} g(\theta_{n+1}) - g(\theta_n) &\le -\tau(1-\frac{\tau L}{2})\|\nabla g(\theta_n)\|_2^2 \\ &\le -\frac{\tau}{2}\|\nabla g(\theta_n)\|_2^2; \end{aligned} \quad (58)$$

- we have stability of the iterates: $\theta_n \in \Lambda_\beta$.

Consider any $\theta \in \Lambda_\beta$ and let $u \in \mathbb{R}^d$ be such that $\|u\|_2 = 1$. Using the extension of the RIP to generalized dipoles (directional derivatives of Dirac measures, see [11, Lemma 2.1]) and Cauchy-Schwarz inequality, we have

$$\begin{aligned} u^T H_\theta u =& 2\|A\partial_u\phi(\theta)\|_2^2 \\ &+ 2\langle A\partial_u^2\phi(\theta), A(\phi(\theta)-\phi(\theta^*))\rangle \\ \ge& 2(1-\gamma)\|\partial_u\phi(\theta)\|_h^2 \\ &- 2\|A\partial_u^2\phi(\theta)\|_2(\sqrt{1+\gamma}C_{\phi,\theta^*}\beta). \end{aligned} \quad (59)$$

We deduce that $u^T H_\theta u - \xi \ge 0$ as soon as

$$\beta \le \frac{(1-\gamma)}{C_{\phi,\theta^*}\sqrt{1+\gamma}} \inf_{\theta\in\Lambda_\beta} \inf_{\|u\|_2=1} \left(\frac{\|\partial_u\phi(\theta)\|_h^2 - \xi/2}{\|A\partial_u^2\phi(\theta)\|_2} \right). \quad (60)$$

If this last inequality is satisfied, then ξ-strong convexity is guaranteed on Λ_β. Using $g(\theta^*) - g(\theta_n) \ge -\langle \nabla g(\theta_n), \theta_n - \theta^*\rangle + \frac{\xi}{2}\|\theta_n - \theta^*\|_2^2$ (thanks to strong convexity), we have

$$\begin{aligned} \|\theta_{n+1}-\theta^*\|_2^2 =& \|\theta_n-\theta^*\|_2^2 - 2\tau\langle\nabla g(\theta_n), \theta_n-\theta^*\rangle \\ &+ \tau^2\|\nabla g(\theta_n)\|_2^2 \\ \le& \|\theta_n-\theta^*\|_2^2 - \tau\xi\|\theta_n-\theta^*\|_2^2 \\ &+ 2\tau(g(\theta^*)-g(\theta^n)) + \tau^2\|\nabla g(\theta_n)\|_2^2. \end{aligned} \quad (61)$$

Using (58) for $\tau = \frac{1}{L}$ and $g(\theta^*) \le g(\theta_{n+1})$, we get

$$\|\nabla g(\theta_n)\|_2^2 \le 2L(g(\theta_n) - g(\theta^*)). \quad (62)$$

For $\tau \le \tau_0 = \frac{1}{L}$ we obtain,

$$\begin{aligned} &\|\theta_{n+1}-\theta^*\|_2^2 \\ &\le (1-\tau\xi)\|\theta_n-\theta^*\|_2^2 + 2\tau(g(\theta^*)-g(\theta^n)) \\ &\quad + 2\tau^2 L(g(\theta_n)-g(\theta^*)) \\ &= (1-\tau\xi)\|\theta_n-\theta^*\|_2^2 + 2\tau(1-\tau L)(g(\theta^*)-g(\theta^n)) \\ &\le (1-\tau\xi)\|\theta_n-\theta^*\|_2^2. \end{aligned} \quad (63)$$

This gives the convergence of θ_n with a linear rate.

$$\|\theta_n-\theta^*\|_2^2 \le (1-\tau\xi)^n \|\theta_0-\theta^*\|_2^2. \quad (64)$$

We deduce that we can take β_{spikes} as a lower bound of the right-hand side of (60) to obtain a strong basin of attraction. To make the constants explicit, we need to calculate three bounds.

- Bound 1: Let $\theta = (t_1, \ldots, t_k) \in \tilde{\Theta}_{k,\epsilon}$ and let $u = (u_1, \ldots, u_k) \in \mathbb{R}^{kd}$ such that $\|u\|_2^2 = \sum_{r=1}^k \|u_r\|_2^2 = 1$. Using $\|\delta'_{t_r,u_r}\|_h^2 = \|u_r\|_2^2|\rho''(0)|$ with $\rho(s) = e^{-\frac{s^2}{2\lambda^2}}$ from [10, Lemma 2.1] and Lemma 2.1, we have

$$\begin{aligned} \|\partial_u\phi(z)\|_h^2 &= \|\sum_{r=1}^k \delta'_{t_r,u_r}\|_h^2 \\ &\ge (1-(k-1)\mu_h)\sum_{r=1}^k \|\delta'_{t_r,u_r}\|_h^2 \\ &= (1-(k-1)\mu_h)\sum_{r=1}^k |\rho''(0)|\|u_r\|_2^2 \\ &= (1-(k-1)\mu_h)|\rho''(0)|. \end{aligned} \quad (65)$$

We have $\rho''(s) = \frac{e^{-\frac{s^2}{2\lambda^2}}}{\lambda^2}(\frac{s^2}{\lambda^2} - 1)$ and $|\rho''(0)| = 1/\lambda^2$ which yields

$$\inf_{\|u\|=1, z\in\tilde{\Theta}_{k,\epsilon}} \|\partial_u \phi(z)\|_h^2 \geq \frac{1}{\lambda^2}(1-(k-1)\mu_h). \tag{66}$$

- Bound 2: Let $\theta = (t_1, \dots, t_k) \in \tilde{\Theta}_{k,\epsilon}$ and let $u = (u_1, \dots, u_k) \in \mathbb{R}^{kd}$ such that $\|u\|_2^2 = \sum_{r=1}^k \|u_k\|_2^2 = 1$. We have

$$\begin{aligned} \|A\partial_u^2 \phi(\theta)\|_2^2 &= \sum_{l=1}^m \Big| \sum_{i=1}^k \partial_{u_i}^2 \alpha_l(t_i) \Big|^2 \\ &\leq \sum_{l=1}^m \left(\sum_{i=1}^k |\partial_{u_i}^2 \alpha_l(t_i)| \right)^2 . \end{aligned} \tag{67}$$

We calculate

$$\partial_{u_i}^2 \alpha_l(t_i) = -\frac{1}{\sqrt{m}} c(\omega_l) |\langle u_i, \omega_l \rangle|^2 e^{-j\langle \omega_l, t_i \rangle}. \tag{68}$$

Using the Cauchy-Schwarz inequality, we get

$$\begin{aligned} \|A\partial_u^2 \phi(\theta)\|_2^2 &\leq \frac{1}{m} \sum_{l=1}^m c(\omega_l)^2 \left(\sum_{i=1}^k |\langle u_i, \omega_l \rangle|^2 \right)^2 \\ &\leq \frac{1}{m} \sum_{l=1}^m c(\omega_l)^2 \left(\sum_{i=1}^k \|u_i\|_2^2 \|\omega_l\|_2^2 \right)^2 \\ &= \frac{1}{m} \sum_{l=1}^m c(\omega_l)^2 \|\omega_l\|_2^4 \left(\sum_{i=1}^k \|u_i\|_2^2 \right)^2 \\ &= \frac{1}{m} \sum_{l=1}^m \left(1 + \frac{\lambda^2 \|\omega_l\|_2^2}{d} \right)^{-2} \|\omega_l\|_2^4. \end{aligned} \tag{69}$$

A direct study of the function $v \to v^4/(1 + \lambda^2 v^2/d)^2$ shows that it is non-decreasing on $[0, +\infty)$ and bounded by its limit d^2/λ^4 when $v \to \infty$. We obtain

$$\|A\partial_u^2 \phi(\theta)\|_2 \leq \frac{d}{\lambda^2}. \tag{70}$$

- Bound 3: Let $(s_1, ..., s_k) \in \Lambda_\beta$. Using Lemma 2.1 and $e^{-x} \geq 1 - x$, we get

$$\begin{aligned} &\left\| \sum_{i=1}^k \delta_{s_i} - \sum_{i=1}^k \delta_{t_i} \right\|_h^2 \\ &\leq (1 + (k-1)\mu_h) \sum_{i=1}^k \|\delta_{s_i} - \delta_{t_i}\|_h^2 \\ &= 2(1 + (k-1)\mu_h) \sum_{i=1}^k \left(1 - e^{-\frac{\|s_i - t_i\|_2^2}{2\lambda^2}} \right) \\ &\leq (1 + (k-1)\mu_h) \frac{1}{\lambda^2} \sum_{i=1}^k \|s_i - t_i\|_2^2. \end{aligned} \tag{71}$$

Therefore we can use the explicit constant $C_{\phi,\theta^*} = \frac{1}{\lambda}\sqrt{(1 + (k-1)\mu_h)}$ in (20).

We finally obtain that ξ-strong convexity on Λ_β is ensured for

$$\begin{aligned} \beta &\leq \frac{(1-\gamma)((1-(k-1)\mu_h)/\lambda^2 - \xi/2)}{\sqrt{1+\gamma}(d/\lambda^2)(\sqrt{(1+(k-1)\mu_h)}/\lambda)} \\ &= \frac{(1-\gamma)(\lambda(1-(k-1)\mu_h) - \lambda^3\xi/2)}{\sqrt{1+\gamma}\, d\sqrt{(1+(k-1)\mu_h)}}. \end{aligned} \tag{72}$$

Rate of convergence of $g(\theta_n) - g(\theta^*)$.
With strong convexity, we have that

$$\begin{aligned} &g(\theta_n) - g(\theta^*) \\ &\leq \langle \nabla g(\theta_n), \theta_n - \theta^* \rangle - \frac{\xi}{2}\|\theta_n - \theta^*\|_2^2 \\ &= -\frac{\xi}{2}\Big(-2\langle \frac{1}{\xi}\nabla g(\theta_n), \theta_n - \theta^* \rangle + \|\theta_n - \theta^*\|_2^2 \Big) \\ &= -\frac{\xi}{2}\Big(\|\theta_n - \theta^* - \frac{1}{\xi}\nabla g(\theta_n)\|_2^2 - \frac{1}{\xi^2}\|\nabla g(\theta_n)\|_2^2 \Big) \\ &= \frac{1}{2\xi}\|\nabla g(\theta_n)\|_2^2 - \frac{\xi}{2}\|\theta_n - \theta^* - \frac{1}{\xi}\nabla g(\theta_n)\|_2^2 \\ &\leq \frac{1}{2\xi}\|\nabla g(\theta_n)\|_2^2. \end{aligned} \tag{73}$$

We deduce from this inequality and from (58) that

$$\begin{aligned} g(\theta_{n+1}) - g(\theta^*) &\leq g(\theta_n) - g(\theta^*) - \frac{\tau}{2}\|\nabla g(\theta_n)\|_2^2 \\ &\leq (1 - \xi\tau)\,(g(\theta_n) - g(\theta^*)). \end{aligned} \tag{74}$$

Rate of convergence of $\phi(\theta_n)$. As we are in the noiseless case $g(\theta^*) = 0$, and using the RIP,

$$\|\phi(\theta_n) - \phi(\theta^*)\|_h^2 \leq \frac{1}{1-\gamma}\|A\phi(\theta_n) - A\phi(\theta^*)\|_2^2 = \frac{1}{1-\gamma}(g(\theta_n) - g(\theta^*)). \tag{75}$$

C Numerical bounds of $\|A\partial_u^2\phi(\theta)\|_2$ for random weighted Fourier measurements

For a given direction u and $k = 1$, we sample $m = 10000$ frequencies according to the probability density function $\propto \frac{1}{c(\omega)^2}e^{-\frac{\lambda^2\|\omega\|_2^2}{2}}$ (and perform 10000 experiments for each parameter set). We then evaluate $\|A\partial_u^2\phi(\theta)\|$ in (70) and take the maximum value. We repeat this process for different d and we obtain the diagram shown in Figure 8. These experiments suggest that $\|A\partial_u^2\phi(\theta)\|_2$ can be upper bounded with a constant that does not depend on d.

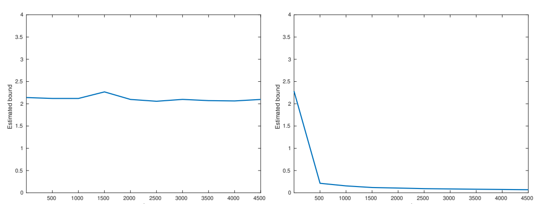


Fig. 8: Evaluation of $\|A\partial_u^2\phi(\theta)\|_2$ for $k = 1$ and different d. Left: $c(\omega) = 1$. Right: $c(\omega) = 1/(1 + \lambda^2\|\omega\|_2^2/d)$. We observe that $\|A\partial_u^2\phi(\theta)\|_2$ can be bounded independently of d.

D Proofs for Section 4

Proof of Lemma 4.1 For any signed measure x, we have

$$\begin{aligned}\|Ax\|_2^2 &= \sum_{l=1}^m |\langle x, \alpha_l\rangle|^2 \\ &= \sum_{l=1}^m \overline{\left(\int_{\mathbb{R}^d}\alpha_l(s)\,\mathrm{d}x(s)\right)}\int_{\mathbb{R}^d}\alpha_l(u)\,\mathrm{d}x(u) \\ &= \sum_{l=1}^m \int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\overline{\alpha_l(s)}\alpha_l(u)\,\mathrm{d}x(s)\,\mathrm{d}x(u) \\ &= \frac{1}{m}\sum_{l=1}^m c(\omega_l)^2\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}e^{-j\langle\omega_l,u-s\rangle}\,\mathrm{d}x(s)\,\mathrm{d}x(u).\end{aligned} \tag{76}$$

Thus, for $x = \delta_t - \delta_0$,

$$\|A(\delta_t - \delta_0)\|_2^2 = \frac{1}{m}\sum_{l=1}^m c(\omega_l)^2(2 - e^{-j\langle\omega_l,t\rangle} - e^{j\langle\omega_l,t\rangle}). \tag{77}$$

Also, for any $t \in \mathbb{R}^d$, and any $l = 1, \dots, m$, the Fourier transform of the Gaussian gives

$$\begin{aligned}\mathbb{E}(c(\omega_l)^2e^{-j\langle\omega_l,t\rangle}) &= D_c(\lambda)\int_{\mathbb{R}^d}e^{-j\langle\omega_l,t\rangle}e^{-\frac{\lambda^2\|\omega_l\|_2^2}{2}}\,\mathrm{d}\omega_l \\ &= D_c(\lambda)\frac{(2\pi)^{\frac{d}{2}}}{\lambda^d}e^{-\frac{\|t\|^2}{2\lambda^2}},\end{aligned} \tag{78}$$

and in particular, for $t = 0$,

$$\mathbb{E}(c(\omega_l)^2) = D_c(\lambda)\int_{\mathbb{R}^d}e^{-\frac{\lambda^2\|\omega_l\|_2^2}{2}}\,\mathrm{d}\omega_l = D_c(\lambda)\frac{(2\pi)^{\frac{d}{2}}}{\lambda^d}, \tag{79}$$

which gives the first part of the formula. Since $h(s,u) = e^{-\frac{1}{2\lambda^2}\|s-u\|_2^2}$, the second part follows immediately:

$$\begin{aligned}\|\delta_t - \delta_0\|_h^2 &= \int_{\mathbb{R}^d}\int_{\mathbb{R}^d}h(s,u)\,\mathrm{d}(\delta_t-\delta_0)(s)\,\mathrm{d}(\delta_t-\delta_0)(u). \\ &= 2\left(1 - e^{-\frac{\|t\|_2^2}{2\lambda^2}}\right).\end{aligned} \tag{80}$$

□

Proof of Lemma 4.3 We have

$$\begin{aligned}\mathbb{E}(Y) &= \mathbb{E}(Z^2\cos(hZ)) \\ &= \frac{\lambda}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}z^2\cos(hz)e^{-\frac{\lambda^2}{2}z^2}dz \\ &= \frac{1}{h^3}\frac{\lambda}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}z^2\cos(z)e^{-\frac{\lambda^2}{2h^2}z^2}dz,\end{aligned} \tag{81}$$

where the last equality is obtained with the change of variable $hz \to z$. This integral can be computed with the Fourier tranform of the Gaussian

$$\forall s \in \mathbb{R}, \quad \int_{-\infty}^{+\infty} e^{isz} e^{-a^2 z^2} dz = \frac{\sqrt{\pi}}{a} e^{-\frac{s^2}{4a^2}}. \tag{82}$$

By differentiating with respect to s twice under the integral, we get

$$\begin{aligned}\forall s \in \mathbb{R}, \quad & -\int_{-\infty}^{+\infty} z^2 e^{isz} e^{-a^2 z^2} dz \\ & = \frac{\sqrt{\pi}}{a} \left(\frac{s^2}{4a^4} - \frac{1}{2a^2} \right) e^{-\frac{s^2}{4a^2}}.\end{aligned} \tag{83}$$

and thus, with $s = 1$, taking the real part, we obtain

$$\int_{-\infty}^{+\infty} z^2 \cos(z) e^{-a^2 z^2} dz = \frac{\sqrt{\pi}(2a^2 - 1) e^{-1/(4a^2)}}{4a^5}. \tag{84}$$

Hence

$$\begin{aligned}\mathbb{E}(Y) &= \frac{1}{h^3} \frac{\lambda}{\sqrt{2\pi}} \frac{\sqrt{\pi}(\frac{2\lambda^2}{2h^2} - 1) e^{-\frac{2h^2}{4\lambda^2}}}{4\lambda^5/(4\sqrt{2}h^5)} \\ &= h^2 \frac{(\frac{\lambda^2}{h^2} - 1) e^{-\frac{h^2}{2\lambda^2}}}{\lambda^4} \\ &= \frac{(\lambda^2 - h^2) e^{-\frac{h^2}{2\lambda^2}}}{\lambda^4}.\end{aligned} \tag{85}$$

□

Proof of Lemma 4.4 Using the definition of X, we have

$$\begin{aligned}&\mathbb{E}(X) \\ &= D_c(\lambda) \int_{\mathbb{R}^d} c(\omega)^2 |\langle \omega, u \rangle|^2 \cos(\langle \omega, u \rangle h) \frac{e^{-\frac{\lambda^2}{2}\|\omega\|_2^2}}{c(\omega)^2} d\omega \\ &= D_c(\lambda) \int_{\mathbb{R}^d} |\langle \omega, u \rangle|^2 \cos(\langle \omega, u \rangle h) e^{-\frac{\lambda^2}{2}\|\omega\|_2^2} d\omega.\end{aligned} \tag{86}$$

Hence

$$\mathbb{E}(X) = \frac{D_c(\lambda)}{D(\lambda)} \mathbb{E}(Z^2 \cos(hZ)). \tag{87}$$

The result is directly deduced from Lemma 4.3.

□

The variations of the calculated mean $F(h)$ in Lemma 4.3 with respect to h will be useful for the proof of Theorem 4.1.

Lemma D.1. *Let* $F(h) := \frac{(\lambda^2 - h^2) e^{-\frac{h^2}{2\lambda^2}}}{\lambda^4}$. *Then, for* $h \geq 0$

- $F(0) = 1/\lambda^2 = \max_{h \geq 0} F(h)$;
- $F(h) = 0$ *implies* $h^2 = \lambda^2$;
- F *is decreasing on* $[0, \sqrt{3}\lambda]$.

Proof We have $\lambda^4 F'(h) = -2h e^{-\frac{h^2}{2\lambda^2}} - \frac{h}{\lambda^2}(\lambda^2 - h^2) e^{-\frac{h^2}{2\lambda^2}}$. Hence, $F'(h) = 0$ implies $h(-2 - 1 + h^2/\lambda^2) = 0$ i.e. $h = 0$ or $h^2 = 3\lambda^2$. We conclude with a standard study of the function F. □

Proof of Lemma 4.5 Inspired by [14], we write

$$\begin{aligned}\mathbb{E} e^{s(X - \mathbb{E}[X])} =& 1 + \mathbb{E}(s(X - \mathbb{E}[X])) \\ & + \sum_{r=2}^{+\infty} \frac{s^r \mathbb{E}((X - \mathbb{E}[X])^r)}{r!}\end{aligned} \tag{88}$$

Since $|X - \mathbb{E}[X]| \leq |X| + \mathbb{E}[|X|]$, the triangle inequality (for the norm $(\mathbb{E}|\cdot|^r)^{\frac{1}{r}}$) and the Jensen inequality give

$$\begin{aligned}\mathbb{E}[(X - \mathbb{E}[X])^r]^{\frac{1}{r}} &\leq \mathbb{E}[|X - \mathbb{E}[X]|^r]^{\frac{1}{r}} \\ &\leq \mathbb{E}[|X|^r]^{\frac{1}{r}} + [(\mathbb{E}[|X|])^r]^{\frac{1}{r}} \\ &\leq 2\mathbb{E}(|X|^r)^{\frac{1}{r}}.\end{aligned} \tag{89}$$

We deduce

$$\mathbb{E} e^{s(X - \mathbb{E}[X])} \leq 1 + \sum_{r=2}^{+\infty} \frac{|s|^r 2^r \mathbb{E}(|X|^r)}{r!}. \tag{90}$$

Since $c(\omega) \leq 1$, we bound

$$\begin{aligned}\mathbb{E}(|X|^r) = & \int_{\mathbb{R}^d} \Big[(c(\omega))^{2r} |\langle \omega_l, u \rangle|^{2r} |\cos(h\langle \omega, u \rangle)|^r \\ & \frac{D_c(\lambda)}{(c(\omega))^2} e^{-\frac{\lambda^2}{2}\|\omega\|_2^2} \Big] d\omega \\ =& D_c(\lambda) \int_{\mathbb{R}^d} \Big[(c(\omega))^{2(r-1)} |\langle \omega_l, u \rangle|^{2r} \\ & |\cos(h\langle \omega, u \rangle)|^{2r} e^{-\frac{\lambda^2}{2}\|\omega\|_2^2} \Big] d\omega \\ \leq & \frac{D_c(\lambda)}{D(\lambda)} \int_{\mathbb{R}^d} |\langle \omega_l, u \rangle|^{2r} D(\lambda) e^{-\frac{\lambda^2}{2}\|\omega\|_2^2} d\omega.\end{aligned} \tag{91}$$

Let $Z \sim \mathcal{N}(0, 1/\lambda^2)$, the previous inequality implies

$$\mathbb{E}(|X|^r) \leq \frac{D_c(\lambda)}{D(\lambda)} \mathbb{E}(|Z|^{2r}). \tag{92}$$

As $c(\omega) \leq 1$, we have $(c(\omega))^{-2} \geq 1$ and

$$\begin{aligned}\frac{D_c(\lambda)}{D(\lambda)} &= \frac{\int_{\mathbb{R}^d} e^{-\frac{\lambda^2}{2}\|\omega\|_2^2} d\omega}{\int_{\mathbb{R}^d} e^{-\frac{\lambda^2}{2}\|\omega\|_2^2} (c(\omega))^{-2} d\omega} \\ &\leq \frac{\int_{\mathbb{R}^d} e^{-\frac{\lambda^2}{2}\|\omega\|_2^2} d\omega}{\int_{\mathbb{R}^d} e^{-\frac{\lambda^2}{2}\|\omega\|_2^2} d\omega (\inf_\omega (c(\omega))^{-2})} \leq 1.\end{aligned} \tag{93}$$

We deduce $\mathbb{E}(|X|^r) \leq \mathbb{E}(|Z|^{2r})$ and

$$\mathbb{E}e^{s(X-\mathbb{E}[X])} \leq 1 + \sum_{r=2}^{+\infty} \frac{|s|^r 2^r (\mathbb{E}(|Z|^{2r}))}{r!} \tag{94}$$

We have that $\mathbb{E}(|Z|^{2r}) = \frac{1}{\sqrt{\pi}} \left(\frac{2}{\lambda^2}\right)^r \Gamma((2r+1)/2) \leq \frac{1}{\sqrt{\pi}} r! \left(\frac{2}{\lambda^2}\right)^r$. For $4|s|/\lambda^2 \leq 1/2$ (i.e. $|s| \leq \lambda^2/8$), we have

$$\begin{aligned}
\mathbb{E}e^{s(X-\mathbb{E}[X])} &\leq 1 + \frac{1}{\sqrt{\pi}} \sum_{r=2}^{+\infty} |s|^r 2^r \left(\frac{2}{\lambda^2}\right)^r \\
&= 1 + \frac{1}{\sqrt{\pi}} \sum_{r=2}^{+\infty} \left(\frac{4|s|}{\lambda^2}\right)^r \\
&= 1 + \frac{1}{\sqrt{\pi}} \left(\frac{4|s|}{\lambda^2}\right)^2 \sum_{r=0}^{+\infty} \left(\frac{4|s|}{\lambda^2}\right)^r \\
&\leq 1 + \frac{1}{\sqrt{\pi}} \left(\frac{4s}{\lambda^2}\right)^2 \frac{1}{1-1/2} \\
&= 1 + \frac{32}{\sqrt{\pi}} \left(\frac{s}{\lambda^2}\right)^2 \\
&\leq e^{\frac{64}{\sqrt{\pi}} \frac{s^2}{2\lambda^4}} \leq e^{\frac{s^2}{2}\left(\frac{8}{\lambda^2}\right)^2}.
\end{aligned} \tag{95}$$

This implies that $X \sim \mathbf{subE}(8/\lambda^2)$ (Definition 2.6 with $a = 8/\lambda^2$). □

Proof of Theorem 4.1 For $0 \leq \eta \leq \frac{1}{\lambda^2} \frac{D_c(\lambda)}{D(\lambda)} = F(0)$, Lemma D.1 gives a unique $0 \leq h_\eta \leq \lambda$ such that

$$\frac{(\lambda^2 - h_\eta^2) e^{-\frac{h_\eta^2}{2\lambda^2}}}{\lambda^4} \frac{D_c(\lambda)}{D(\lambda)} = \eta \quad \text{i.e.} \quad \mathbb{E}(X_l(u, h_\eta)) = \eta. \tag{96}$$

Then, with Lemma 4.6, for such (η, h_η), we have

$$\begin{aligned}
\mathbb{P}\Big(\frac{1}{m}\sum_{l=1}^m X_l > 2\eta\Big) &= \mathbb{P}\Big(\frac{1}{m}\sum_{l=1}^m X_l - \mathbb{E}(X_l) > \eta\Big) \\
&\leq p(m, \lambda, \eta) := e^{-\frac{m}{2}\min(\eta^2 \frac{\lambda^4}{8^2}, \eta\frac{\lambda^2}{8})}.
\end{aligned} \tag{97}$$

Since $c(\omega) \leq 1$ (see proof of Lemma 4.5), we have $\frac{D_c(\lambda)}{D(\lambda)} \leq 1$ and thus it holds that $\eta\lambda^2 \leq 1$. Hence $\frac{\eta\lambda^2}{8} \leq 1$ and $\eta^2 \frac{\lambda^4}{8^2} \leq \frac{\eta\lambda^2}{8}$. We thus get that $\min(\eta^2 \frac{\lambda^4}{8^2}, \eta\frac{\lambda^2}{8}) = \eta^2 \frac{\lambda^4}{8^2}$.

Taking $0 < \eta \leq F(0)$ and m_0 such that $\frac{m_0}{2}\eta^2\frac{\lambda^4}{8^2} \geq 1$, we get $p(m_0, \lambda, \eta) \leq 1/e$. Hence this probabilistic control is guaranteed for any $m \geq m_0$, which satisfies (by expliciting the value of η^2)

$$\begin{aligned}
&\frac{m}{2}\left(\frac{(\lambda^2 - h_\eta^2)e^{-h_\eta^2/(2\lambda^2)}}{\lambda^4} \frac{\lambda^2}{8} \frac{D_c(\lambda)}{D(\lambda)}\right)^2 \geq 1 \\
&\text{i.e.} \quad \left((\lambda^2 - h_\eta^2)e^{-h_\eta^2/(2\lambda^2)} \frac{D_c(\lambda)}{D(\lambda)}\right)^2 \geq \frac{2 \cdot 8^2 \lambda^4}{m} \\
&\text{i.e.} \quad (\lambda^2 - h_\eta^2)e^{-h_\eta^2/(2\lambda^2)} \geq \frac{D(\lambda)}{D_c(\lambda)} \frac{\sqrt{2}\cdot 8\lambda^2}{\sqrt{m}}.
\end{aligned} \tag{98}$$

This in turn implies

$$\begin{aligned}
&\lambda^2 - h_\eta^2 \geq \frac{D(\lambda)}{D_c(\lambda)} \frac{\sqrt{2}\cdot 8\lambda^2}{\sqrt{m}} \\
&\text{i.e.} \quad h_\eta^2 \leq \lambda^2 \left(1 - \frac{D(\lambda)}{D_c(\lambda)} \frac{8\sqrt{2}}{\sqrt{m}}\right).
\end{aligned} \tag{99}$$

and

$$\mathbb{P}\left(\frac{1}{m}\sum_{l=1}^m X_l > 2\eta\right) \leq 1/e. \tag{100}$$

For any sufficiently small $\xi = 4\eta > 0$, recalling that $\frac{1}{m}\sum_{l=1}^m X_l = \frac{1}{2}u^T H_\theta u$ with $\theta = hu$, we can find h such that

$$h^2 \leq \lambda^2 \left(1 - \frac{D(\lambda)}{D_c(\lambda)} \frac{8\sqrt{2}}{\sqrt{m}}\right) \tag{101}$$

and

$$\begin{aligned}
\mathbb{P}(g \in \Gamma([0, uh], \xi)) &\leq \mathbb{P}\left(\frac{1}{2}u^T H_{hu} u > \frac{1}{2}\xi\right) \\
&= \mathbb{P}\left(\frac{1}{m}\sum_{l=1}^m X_l > 2\eta\right) \leq 1/e.
\end{aligned} \tag{102}$$

□

We need the following Lemma to give a lower bound on sets Λ where g is strongly convex.

Lemma D.2. *Let* $f(u,h) = \frac{1}{m}\sum_{l=1}^m c(\omega_l)^2 |\langle \omega_l, u\rangle|^2 \cos(\langle \omega_l, u\rangle h)$. *Then*

$$\begin{aligned}
&\sup_{\|u\|=1, 0\leq h \leq h_{max}} \|\nabla f(u,h)\|_2 \\
&\leq \frac{d\sqrt{d}}{2\lambda^2} \sqrt{\left(1 + \frac{h_{max} 3\sqrt{3d}}{8\lambda}\right)^2 + \left(\frac{3\sqrt{3}}{8\lambda}\right)^2} \\
&=: \varrho(d, \lambda, h_{max})
\end{aligned} \tag{103}$$

Proof For $\|u\|_2 = 1$, with Cauchy-Schwarz inequality,

$$\Big|\frac{\partial f(u,z)}{\partial z}\Big| = \Big|\frac{1}{m}\sum_{l=1}^{m} c(\omega_l)^2 |\langle \omega_l, u\rangle|^2 \langle \omega_l, u\rangle sin(\langle \omega_l, u\rangle z)\Big| \leq \max_{l\in\{1,\dots,m\}} \frac{|\langle \omega_l, u\rangle|^3}{(1+\lambda^2\|\omega_l\|_2^2/d)^2} \leq \max_{l\in\{1,\dots,m\}} \frac{\|\omega_l\|^3}{(1+\lambda^2\|\omega_l\|_2^2/d)^2}. \tag{104}$$

The derivative of $v \to v^3/(1+\lambda^2 v^2/d)^2$ is zeroed for v such that

$$\begin{aligned} 3v^2(1+\lambda^2v^2/d)^2 - 2v^3(2v\lambda^2/d)(1+v^2\lambda^2/d) &= 0 \\ v^2(1+v^2\lambda^2/d)(3(1+v^2\lambda^2/d) - 4v^2\lambda^2/d) &= 0 \\ v^2(1+v^2\lambda^2/d)(3-v^2\lambda^2/d) &= 0 \end{aligned} \tag{105}$$

i.e. $v^2 = 3d/\lambda^2$ or $v = 0$. Hence its maximum value is $\frac{3d\sqrt{3d}}{16\lambda^3}$ and

$$\Big|\frac{\partial f(u,z)}{\partial z}\Big| \leq \frac{3d\sqrt{3d}}{16\lambda^3}. \tag{106}$$

We bound the other partial derivatives.

$$\begin{aligned} &\Big|\frac{\partial f(u,z)}{\partial u_i}\Big| \\ &= \Big|\frac{1}{m}\sum_{l=1}^{m} c(\omega_l)^2 \Big(2\omega_{l,i}\langle \omega_l, u\rangle \cos(\langle \omega_l, u\rangle z) \\ &\quad - |\langle \omega_l, u\rangle|^2 \omega_{l,i} z \sin(\langle \omega_l, u\rangle z)\Big)\Big| \\ &\leq \frac{1}{m}\sum_{l=1}^{m} c(\omega_l)^2 \left(|2\omega_{l,i}\langle \omega_l, u\rangle| + |\langle \omega_l, u\rangle|^2 |\omega_{l,i} z|\right) \\ &\leq \frac{1}{m}\sum_{l=1}^{m} c(\omega_l)^2 \left(2\|\omega_l\|_2^2 + z\|\omega_l\|_2^3\right) \\ &\leq \frac{1}{m}\sum_{l=1}^{m} \left(c(\omega_l)^2 2\|\omega_l\|_2^2 + z\frac{3d\sqrt{3d}}{16\lambda^3}\right). \end{aligned} \tag{107}$$

The derivative of $v \to v^2/(1+v^2\lambda^2/d)^2$ is zeroed for v such that

$$\begin{aligned} 2v(1+v^2\lambda^2/d)^2 - 2v^2(2v\lambda^2/d)(1+v^2\lambda^2/d) &= 0 \\ v(1+v^2\lambda^2/d)(2(1+v^2\lambda^2/d) - 4v^2\lambda^2/d) &= 0 \\ 2v(1+v^2\lambda^2/d)(1-v^2\lambda^2/d) &= 0 \end{aligned} \tag{108}$$

i.e. $v^2 = d/\lambda^2$ or $v = 0$. Hence its maximum value of the function is $d/(4\lambda^2)$ and, for $z \leq h_{max}$,

$$\Big|\frac{\partial f(u,z)}{\partial u_i}\Big| \leq \frac{d}{2\lambda^2} + \frac{h_{max}3d\sqrt{3d}}{16\lambda^3}. \tag{109}$$

Finally,

$$\begin{aligned} &\|\nabla f(u,z)\|_2 \\ &\leq \sqrt{d\left(\frac{d}{2\lambda^2} + \frac{h_{max}3d\sqrt{3d}}{16\lambda^3}\right)^2 + \left(\frac{3d\sqrt{3d}}{16\lambda^3}\right)^2} \\ &= \frac{d\sqrt{d}}{2\lambda^2}\sqrt{\left(1 + \frac{h_{max}3\sqrt{3d}}{8\lambda}\right)^2 + \left(\frac{3\sqrt{3}}{8\lambda}\right)^2}. \end{aligned} \tag{110}$$

□

Proof of Lemma 4.7 For $0 \leq \eta \leq \frac{1}{\lambda^2}\frac{D_c(\lambda)}{D(\lambda)} = F(0)$, Lemma D.1 gives a unique $0 \leq h_\eta \leq \lambda$ such that

$$\frac{(\lambda^2 - h_\eta^2)e^{-h_\eta^2/(2\lambda^2)}}{\lambda^4}\frac{D_c(\lambda)}{D(\lambda)} = \eta \quad \text{i.e.} \quad \mathbb{E}(X_l) = \eta. \tag{111}$$

From Lemma 4.6, we have, for any $\|u\|_2 = 1$ and $h \leq h_\eta$,

$$\mathbb{P}\left(\frac{1}{m}\sum_{l=1}^{m} X_l < -\eta + F(h)\right) \leq p(m,\lambda,\eta) := e^{-\frac{m}{2}\min(\eta^2\frac{\lambda^4}{8^2}, \eta\frac{\lambda^2}{8})}. \tag{112}$$

Using Lemma D.1, $F(h) \geq F(h_\eta)$, hence $\frac{1}{m}\sum_{l=1}^{m} X_l < -\eta + F(h_\eta) = \eta$ implies $\frac{1}{m}\sum_{l=1}^{m} X_l < -\eta + F(h)$ and

$$\mathbb{P}\left(\frac{1}{m}\sum_{l=1}^{m} X_l < \eta\right) \leq \mathbb{P}\left(\frac{1}{m}\sum_{l=1}^{m} X_l < -\eta + F(h)\right) \leq p(m,\lambda,\eta). \tag{113}$$

Recall that, for differentiable g, we have

$$|g(x+h) - g(x)| \leq \sup_z \|\nabla g(x)\|_2 \cdot \|h\|_2. \tag{114}$$

We now construct $(u_i, h_i)_{i\in I}$ such that u_i, h_i are the centers of balls of radius ζ covering $B(1)\times[0,h_\eta] \subset \mathbb{R}^{d+1}$. The ℓ^2-ball $B(1)$ is covered by $(3/\zeta_1)^d$ balls of radius ζ_1, the set $[0,h_\eta]$ is covered by h_η/ζ_2 balls of radius ζ_2.

Let us set $\zeta_1 = \zeta_2 = \zeta/\sqrt{2}$. Taking $(u_i,h_i)_{i\in I}$ as the cross-product of the two coverings, we have that for any $(u,h) \in B(1)\times[0,h_\eta]$, there exists u_i, h_i such that $\|(u,h) - (u_i,h_i)\|_2^2 \leq \zeta_1^2 + \zeta_2^2 = \zeta^2$. We thus get a ζ-covering with $|I| = (3\sqrt{2}/\zeta)^d\sqrt{2}h_\eta/\zeta = 3^d\sqrt{2}^{d+1}h_\eta/\zeta^{d+1}$.

Suppose $\forall i \in I, f(u_i,h_i) \geq \eta$. We have for all $h \leq h_\eta$, $|f(u,h) - f(u_i,h_i)| \leq \varrho(d,\lambda,h_\eta)\zeta$ (with $\varrho(d,\lambda,h_\eta)$ given by Lemma D.2). This implies $f(u,h) \geq \eta -$

$\varrho(d,\lambda,h_\eta)\zeta$. We deduce, using a union bound, that

$$\begin{aligned}
&\mathbb{P}(\forall u/\|u\|_2=1, \forall h \leq h_\eta, f(u,h) \geq \eta - \varrho(d,\lambda,h_\eta)\zeta) \\
&\geq \mathbb{P}(\forall u_i, \forall h_i, f(u_i,h_i) \geq \eta) \\
&= 1 - \mathbb{P}(\exists u_i, h_i, f(u_i,h_i) < \eta) \\
&\geq 1 - \sum_{i \in I} \mathbb{P}(f(u_i,h_i) < \eta) \\
&\geq 1 - |I| p(m,\lambda,\eta) \\
&= 1 - \frac{3^d \sqrt{2}^{d+1} h_\eta}{\zeta^{d+1}} p(m,\lambda,\eta)
\end{aligned} \tag{115}$$

Take ζ such that $\varrho(d,\lambda,h_\eta)\zeta = \eta/2$. Using the fact that $h_\eta \leq \lambda$, we obtain

$$\begin{aligned}
&\mathbb{P}(\forall u, \forall h \leq h_\eta, f(u,h) \geq \eta/2) \\
&\geq 1 - 3^d(2\sqrt{2})^{d+1}(\varrho(d,\lambda,h_\eta))^{d+1} h_\eta p(m,\lambda,\eta)/\eta^{d+1} \\
&\geq 1 - 3^d(2\sqrt{2})^{d+1}(\varrho(d,\lambda,\lambda))^{d+1} \lambda p(m,\lambda,\eta)/\eta^{d+1}.
\end{aligned} \tag{116}$$

Taking $\xi = \eta$, as $f(u,h) = \frac{1}{2}u^T H_{hu} u$, we have just showed that ξ-strong convexity is obtained on $[0,h_\xi]$ with probability $1-p_0$ where $p_0 = 3^d(2\sqrt{2})^{d+1}(\varrho(d,\lambda,\lambda))^{d+1}\lambda p(m,\lambda,\eta)/\xi^{d+1}$ and for h_ξ such that

$$\frac{(\lambda^2 - h_\xi^2)e^{-h_\xi^2/(2\lambda^2)}}{\lambda^4} \frac{D_c(\lambda)}{D(\lambda)} \geq \xi. \tag{117}$$

As we consider $h_\xi^2 \leq \lambda^2$, this is also guaranteed for $h \leq h_\xi$

$$\frac{(\lambda^2 - h^2)e^{-1/2}}{\lambda^4} \frac{D_c(\lambda)}{D(\lambda)} = \xi. \tag{118}$$

For a fixed ξ sufficiently small, we have

$$\begin{aligned}
&-\log(p_0) \\
&= -\log(p(m,\lambda,\xi)) - \log\left(3^d(2\sqrt{2})^{d+1}(\varrho(d,\lambda,\lambda))^{d+1}\lambda/\xi^{d+1}\right) \\
&= \frac{m}{2}\xi^2\lambda^4/8^2 - \log\left(3^d(2\sqrt{2})^{d+1}(\varrho(d,\lambda,\lambda))^{d+1}\lambda/\xi^{d+1}\right) \\
&= m\xi^2\lambda^4/128 + \tilde{\varrho}(d,\lambda) + (d+1)\log(\xi)
\end{aligned} \tag{119}$$

where $\tilde{\varrho}(d,\lambda) = -\log\left(3^d(2\sqrt{2})^{d+1}(\varrho(d,\lambda,\lambda))^{d+1}\lambda\right)$. For large m, using (118), we have

$$\begin{aligned}
-\log(p_0) &= \Omega(m\xi^2\lambda^4) \\
&= \Omega(m(\lambda^2 - h^2)^2).
\end{aligned} \tag{120}$$

This means there is a constant C such that an arbitrary small p_0 is guaranteed with a size of basin h such that

$$h \geq \sqrt{\lambda^2 - \frac{C\sqrt{-\log(p_0)}}{\sqrt{m}}}. \tag{121}$$

□

References

[1] Candès, E.J., Fernandez-Granda, C.: Towards a mathematical theory of super-resolution. Comm. Pure Appl. Math. **67**(6), 906–956 (2014)

[2] Candès, E.J., Fernandez-Granda, C.: Super-resolution from noisy data. . Fourier Anal. Appl. **19**(6), 1229–1254 (2013)

[3] Bhaskar, B.N., Tang, G., Recht, B.: Atomic norm denoising with applications to line spectral estimation. IEEE Trans. Signal Process. **61**(23), 5987–5999 (2013)

[4] Tang, G., Bhaskar, B.N., Shah, P., Recht, B.: Compressed sensing off the grid. IEEE Trans. Inform. Theory **59**(11), 7465–7490 (2013)

[5] De Castro, Y., Gamboa, F., Henrion, D., Lasserre, J.-B.: Exact solutions to super resolution on semi-algebraic domains in higher dimensions. IEEE Trans. Inform. Theory **63**(1), 621–630 (2016)

[6] Duval, V., Peyré, G.: Exact support recovery for sparse spikes deconvolution. Found. Comput. Math. **15**(5), 1315–1355 (2015)

[7] Denoyelle, Q., Duval, V., Peyré, G., Soubies, E.: The sliding frank–wolfe algorithm and its application to super-resolution microscopy. Inverse Problems **36**(1), 014001 (2019)

[8] Peic Tukuljac, H., Deleforge, A., Gribonval, R.: Mulan: A blind and off-grid method for multichannel echo retrieval. Advances in Neural Information Processing Systems **31** (2018)

[9] Gribonval, R., Blanchard, G., Keriven, N., Traonmilin, Y.: Statistical learning guarantees for compressive clustering and compressive mixture modeling. Math. Stat. Learn., In press **3**(2), 165–257 (2021)

[10] Traonmilin, Y., Aujol, J.-F.: The basins of attraction of the global minimizers of the non-convex sparse spike estimation problem. Inverse Problems **36**(4), 045003 (2020)

[11] Traonmilin, Y., Aujol, J.-F., Leclaire, A.: The basins of attraction of the global minimizers of non-convex inverse problems with low-dimensional models in infinite dimension. Information and Inference: A Journal of the IMA (2022) https://arxiv.org/abs/https://academic.oup.com/imaiai/advance-article-pdf/doi/10.1093/imaiai/iaac011/43469874/iaac011.pdf. https://doi.org/10.1093/imaiai/iaac011. iaac011

[12] Traonmilin, Y., Aujol, J.-F., Leclaire, A.: Projected gradient descent for non-convex sparse spike estimation. IEEE Signal Processing Letters **27**, 1110–1114 (2020)

[13] Bénard, P.-J., Traonmilin, Y., Aujol, J.-F.: Fast off-the-grid sparse recovery with over-parametrized projected gradient descent. arXiv preprint arXiv:2202.13757 (2022)

[14] Rigollet, P., Hütter, J.-C.: High dimensional statistics. Lecture notes for course 18S997 **813**(814), 46 (2019)

[15] Vershynin, R.: High-dimensional Probability: An Introduction with Applications in Data Science vol. 47. Cambridge university press, ??? (2018)

[16] Da Costa, M.F., Chi, Y.: Local geometry of nonconvex spike deconvolution from low-pass measurements (2022). https://doi.org/10.48550/ARXIV.2208.10073

[17] Belhadji, A., Gribonval, R.: Revisiting RIP guarantees for sketching operators on mixture models. Preprint https://hal.science/hal-03872878 (2022). https://hal.science/hal-03872878

[18] Chatalic, A., Gribonval, R., Keriven, N.: Large-scale high-dimensional clustering with fast sketching. In: 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 4714–4718 (2018). IEEE