The MLE is a reliable source: sharp performance guarantees for localization problems

Nathanaël Munier, Emmanuel Soubies, Pierre Weiss

The MLE is a reliable source: sharp performance guarantees for localization problems

Nathanaël Munier[†], **Emmanuel Soubies**[‡], **and Pierre Weiss**[†]

[†] University of Toulouse, CNRS, CBI, IMT, France
[‡] University of Toulouse, CNRS, IRIT, France

E-mail: `nathanael.munier@ens-rennes.fr`

October 2022

Abstract. Single source localization from low-pass filtered measurements is ubiquitous in optics, wireless communications and sound processing. We analyse the performance of the maximum likelihood estimator (MLE) in this context with additive white Gaussian noise. We derive necessary conditions and sufficient conditions on the maximum admissible noise level to reach a given precision with high probability. The two conditions match closely, with a discrepancy related to the conditioning of a noiseless cost function. They tightly surround the Cramér-Rao lower bound for low noise levels. However, they are significantly more precise to describe the performance of the MLE for larger levels. As an outcome, we obtain a new criterion for the design of point spread functions in single molecule microscopy.

Contents

1. Introduction

Many measurement devices can be modeled by a convolution with the impulse response h of the system followed by a sampling step at some locations $z_1, \dots, z_M \in \mathbb{R}^D$. A signal $u : \mathbb{R}^D \to \mathbb{R}$ with $D \in \mathbb{N}$ therefore yields a measurement vector $y \in \mathbb{R}^M$ of the form

$$y_m = (u \star h)(z_m) + \varepsilon_m, \tag{1}$$

where $\varepsilon = (\varepsilon_1, \dots, \varepsilon_M)$ is a perturbation modeling noise on the system. A situation of major interest in applications is that of single source localization. This corresponds to assuming that u reads

$$u = \delta_{\bar{x}}, \tag{2}$$

where $\bar{x} \in \mathbb{R}^D$ is the source location. Assuming that $h \in C^0(\mathbb{R}^D)$, the model (1) yields

$$y_m \stackrel{\text{def}}{=} h(z_m - \bar{x}) + \varepsilon_m. \tag{3}$$

The aim of this paper is to analyze the performance of maximum likelihood estimators to recover $\bar{x}$ from the sole knowledge of the kernel h and the vector y. We work under what is possibly the simplest possible setting, by assuming that the noise is white and Gaussian, i.e., $\varepsilon \sim \mathcal{N}(0, \sigma^2 \mathrm{Id})$. In that case, the maximum likelihood estimator reads*

$$\hat{x} \in \operatorname*{argmin}_{x \in \mathbb{R}^D} \ell_\varepsilon(x) \quad \text{with} \quad \ell_\varepsilon(x) \stackrel{\text{def}}{=} \frac{1}{2} \left\| h(Z - x) - h(Z - \bar{x}) - \varepsilon \right\|_2^2. \tag{4}$$

The computation of $\hat{x}$ requires solving a non-convex optimization problem. This can be done, for instance, using nonlinear programming techniques. The experiments led in this paper were conducted using multiple Newton methods initialized with a fine grid. The final location with the minimal value was then used to approximate $\hat{x}$. This setting is obviously idealized. More realistic situations could include multiple sources with unknown weights and suffering from different sources of noise. Yet, it is at the heart of many engineering issues and we tackle a really basic, yet seemingly unexplored problem.

* Note: the true negative log-likelihood would have an additional normalization constant and be multiplied by $1/\sigma^2$. We chose to discard those terms for later simplifications.

Applications Our main motivation is related to single molecule localization microscopy [3, 19]. This technology was awarded the 2014 Nobel prize in chemistry. It made it possible to break the Abbe diffraction limit so as to reach nanometric resolution. It consists in sequentially activating and localizing fluorescent molecules with a sub-pixel accuracy. There, the setting where only a few scattered sources are activated at a time and are well "separated" is of importance in fixed-cell imaging. This motivates the analysis of the considered single-source model. It is worth mentioning that similar issues appear for the localization of sounds (microphone), stars (radio-telescope) or phones (wireless communications).

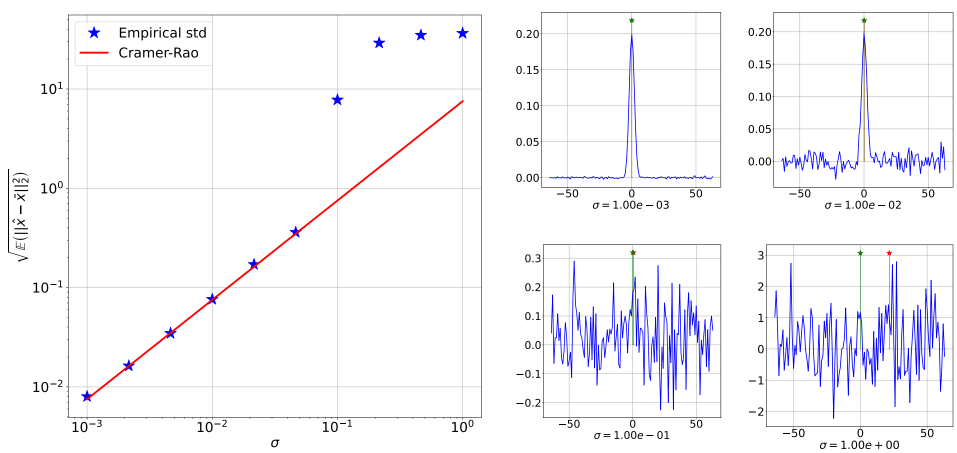


Figure 1: **A 1D localization experiment.** *Left:* the empirical standard deviation using the MLE (blue stars) coincides with the square root of the Cramér-Rao lower-bound (see (7) for its expression) for low noise regimes. In high noise regimes, we observe a significant discrepancy. *Right:* realizations of the vector y for different noise levels. The green and red bars correspond to the true location $\bar{x}$ and the estimated one $\hat{x}$ respectively. Note that y-axis is different in the four graphs on the right.

The Cramér-Rao lower-bound and its limits It has been the subject of numerous theoretical and applied studies in the past 50 years. In particular, the intrinsic performance limits of localization algorithms have been studied using the celebrated Cramér-Rao lower-bound [22]. This bound provides a theoretical limit on the best precision achievable in average with respect to the noise realizations. In optics, it is now used massively to characterize the performance of optical systems [4, 29, 21, 11] as well as a baseline to estimate the quality of algorithms [25]. It can also serve as an optimization criterion to design new efficient point spread functions [15, 26].

In order to motivate our study, let us start with a numerical experiment. In Fig. 1, we compare the empirical precision of the maximum likelihood estimator to the Cramér-Rao lower-bound. We observe that both coincide for low noise regimes, and then significantly deviate. This illustrates the fact that the Cramér-Rao bound is somewhat insufficient to explain the performance of the MLE.

Contributions and outline The main outcomes of our study are a set of *necessary and sufficient conditions* to reach a given localization accuracy r with a given probability.

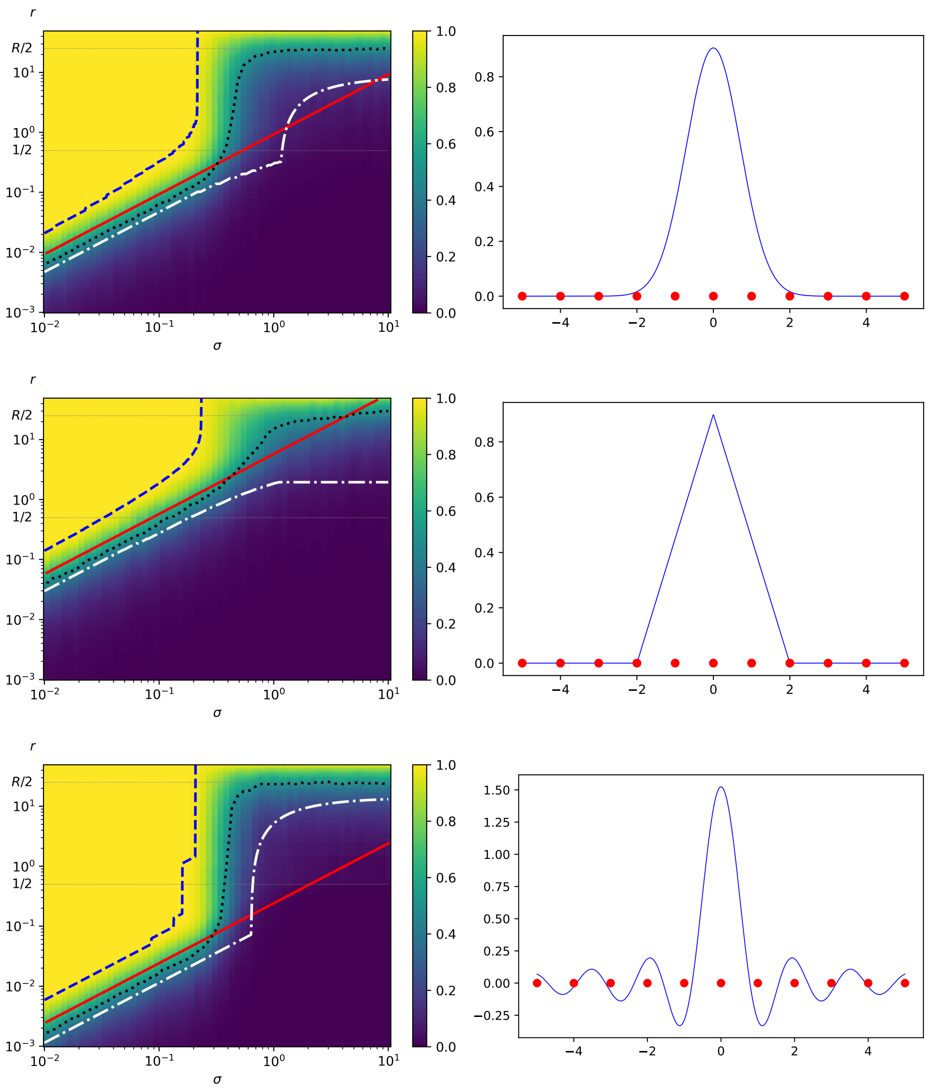


Figure 2: **Phase transitions for single source localization with the MLE.** *Left:* the heat map indicates the empirical probability of the event $\mathcal{E}_r = [\|\hat{x} - \bar{x}\|_2 \leq r]$. We observe a phase transition phenomenon: for a given r and a sufficiently low σ, the probability of success is overwhelming. Above a certain threshold (i.e., when $\sigma \geq \sigma_0$), it decays rapidly to a low probability. The transition occurs close to the square root of the Cramér-Rao lower-bound (red line corresponding to $r = \sqrt{\mathrm{CR}(\sigma)}$) in the low noise regime. The dotted black curve corresponds to an empirical probability of 0.5 (the level line 0.5 of the phase transition diagram). The blue (resp. white) dashed curve corresponds to our theoretical upper (resp. lower) bound on this 0.5 level line. They closely circumscribe the phase transition and predict the behavior for large noise levels as well. *Right:* The convolution kernels h and the sampling points corresponding to the experiments on the left. From top to bottom: Gaussian kernel, the reference kernel of Example 4.1, and a (sub-sampled) cardinal sine kernel.

Fig. 2 illustrates our findings. The heat map in this graph reflects the probability of localizing a Dirac mass with a precision r for a standard deviation σ. A clear *phase transition* appears: above a certain threshold, the probability of detection becomes overwhelming. The phase transition clearly happens around the Cramér-Rao bound in the low noise (or high precision) regime and then significantly deviates. The upper and lower bounds we obtain in this paper clearly circumscribe the phase transition*. Our results based on concentration inequalities also explain this transition behavior from a theoretical standpoint.

Our main conclusions are as follows:

(i) In general it is not true that the MLE attains the Cramér-Rao lower-bound, even asymptotically (i.e., $\sigma \to 0$). However, it becomes true under explicit identifiability hypotheses.

(ii) We provide explicit necessary and sufficient conditions for the global minimizer $\hat{x}$ to satisfy $\|\bar{x} - \hat{x}\|_2 \leq r$ in Section 4. The two conditions match for a specific kernel, showing the tightness of our bounds.

(iii) In addition, these conditions also match the Cramér-Rao bound asymptotically when $\sigma, r \to 0$. For an arbitrary r, the bounds clearly indicate that Cramér-Rao is not accurate anymore and shed new light on the geometry of the localization problem.

(iv) We believe that this work opens new avenues for the field of PSF engineering. If we can choose the kernel h, it suggests new optimization criteria to obtain the "best" possible performance.

Notation Throughout the paper, $\hat{x} \in \mathbb{R}^D$ denotes the maximum likelihood estimator and $\bar{x} \in \mathbb{R}^D$ the true location to be estimated. Assuming that h is of class C^2, we let $\partial_d h$ denote the d-th partial derivative of the impulse response (PSF) h. Moreover, we let $h'(x) = (\partial_1 h(x), \ldots, \partial_D h(x))^T$ denote the gradient of h at $x \in \mathbb{R}^D$. Similarly, we let $h''(x) \in \mathbb{R}^{D\times D}$ denote the Hessian of h at $x \in \mathbb{R}^D$. For a matrix $A \in \mathbb{R}^{M\times N}$, we let $\lambda_{\min}(A)$ (resp. $\lambda_{\max}(A)$) denote its smallest (resp. largest) singular value, $\|A\|_{2\to 2}$ denote its spectral norm and $\|A\|_F = \sqrt{\mathrm{Tr}(A^*A)}$ denote its Frobenius norm. Given two functions a, b we will write $a \lesssim b$ if there exists a constant $c > 0$ not depending on a and b such that for all x, $a(x) \leq c\, b(x)$.

To keep notation concise, we introduce the set of sampling points $Z = (z_1, \ldots, z_M)$ and use the following shorthand notation

$$
\begin{aligned}
h(Z - x) &\stackrel{\text{def}}{=} (h(z_1 - x), \ldots, h(z_M - x)) \in \mathbb{R}^M \\
h'(Z - x) &\stackrel{\text{def}}{=} (h'(z_1 - x), \ldots, h'(z_M - x)) \in \mathbb{R}^{D\times M} \\
h''(Z - x) &\stackrel{\text{def}}{=} (h''(z_1 - x), \ldots, h''(z_M - x)) \in \mathbb{R}^{D\times D\times M}.
\end{aligned}
$$

We let $L^2(\mathbb{R}^D)$ denote the set of squared integrable functions. For $f, g \in L^2(\mathbb{R}^D)$, we let

$$
\langle f, g\rangle_{L^2(\mathbb{R}^D)} \stackrel{\text{def}}{=} \int f(x)\bar{g}(x)\, dx \tag{5}
$$

* To generate the curves in Figure 2, we manually chose a value for the universal constant c appearing in Theorems 4.1 and 4.2, which is not explicitly known. The same constant has been used for all kernels. As such, the bounds represented in Figure 2 have to be interpreted qualitatively (global shape) rather than quantitatively.

denote the usual scalar product on $L^2(\mathbb{R}^D)$. For two vectors $u, v \in \mathbb{C}^N$, we set

$$\langle u, v\rangle \stackrel{\text{def}}{=} \sum_{1\leq n\leq N} u_n \bar{v}_n. \tag{6}$$

2. A brief tour of existing performance bounds

Cramér-Rao lower-bound and variants In our context, applying Theorem 3.1 in [16] yields the following result.

Theorem 2.1 (The Cramér-Rao lower-bound). *If $\varepsilon \sim \mathcal{N}(0, \sigma^2 \mathrm{Id})$ and $h \in C^1(\mathbb{R}^D)$, then any unbiased estimator $\hat{x}$ of $\bar{x}$ satisfies:*

$$\mathbb{E}\left[\|\bar{x} - \hat{x}\|_2^2\right] \geq \mathrm{CR}_\sigma \quad \textit{with} \quad \mathrm{CR}_\sigma \stackrel{\text{def}}{=} \frac{\sigma^2 D^2}{\|h'(Z - \bar{x})\|_F^2}. \tag{7}$$

In addition, there exists an estimator achieving the above lower-bound if and only if the dimension $D = 1$ and h is an affine non-constant function.

Proof. The proof is post-poned to Appendix B. □

While the first part of the theorem is well known (e.g. [21]), we found no reference stating the second. It shows that the Cramér-Rao bound fails to describe the performance of localization algorithms in general. Yet, the experiment in Fig. 1 shows a close match with the experimental points in the low noise regime.

This phenomenon was explained in [20, 23]. There, it was shown that under some technical assumptions, the MLE provides a performance asymptotically similar to the Cramér-Rao lower-bound for $\sigma \to 0$. Hence, the Cramér-Rao lower-bound is indeed a useful tool for small noise levels, but fails to describe the best possible performance for arbitrary noise levels.

Therefore, various authors proposed improved bounds depending nonlinearly on σ^2, especially in the context of array processing [2, 10, 32]. For instance, we refer the reader to the excellent summary in [32, Fig.3] for more details.

The theoretical bounds obtained in these works have revealed insufficiently precise for some practical applications. This led researchers to derive more heuristic but tighter approximations of the mean square error. The general idea is to describe the curvature of the log-likelihood beyond the origin to reach a more global description [30, 1, 20].

In this condensed description of a rich field, we see that existing results either describe the tightest possible performance limits through lower-bounds, or describe more heuristic approximations of the MLE variance. To the best of our knowledge, deriving theoretical upper-bounds remains an open research area that is at the heart of the present work. One of the authors recently conducted a similar study in [12], for the case of blind inverse problems with unknown weights. However, the proof was suboptimal and did not allow us to reach the Cramér-Rao bound asymptotically when $\sigma \to 0$, contrary to the present work.

Beurling-LASSO framework To resolve superpositions of Dirac masses, a lot of works have considered the so-called Beurling LASSO (BLASSO) problem [6, 7, 8, 9, 13, 14, 28]. In this framework, the sources are defined as sparse measures and sparsity is promoted by penalizing the total variation norm. This framework now possesses a rich theory. These include the study of the asymptotic small noise regime [9, 13, 14] where it has been proved that BLASSO produces a solution with exactly the same number of spikes as the target. Beyond this asymptotic noise regime, the authors in [6, 28] exploited the fine structure of the Gaussian white noise to control the effects of the noise and uncover near minimax rates of prediction. The improvement of [6] lies in the joint estimation of the signal and the noise level, making the estimator adaptive to unknown noise levels. More recently, the authors in [8] considered a general parameterized model (continuous dictionaries) with correlated Gaussian noise. In this context, they proved that under a minimal separation condition on the true parameters, rates of prediction similar to those attained by the Lasso (linear regression) are achieved.

The simplified model we consider in this paper (single source with fixed amplitude) can be seen as a special case of these works, in particular of [8]. Yet, in our setting the source location can be multidimensional. Moreover, the proposed analysis sheds new light on the problem at least in two important ways.

- We provide both upper and lower bounds on the localization error which are shown to match closely and surround the Cramér-Rao lower bound for low noise levels. In particular, we manage to discard some logarithmic factors asymptotically.
- Our bounds highlight geometrical features of the kernel h that should be controlled to limit localization errors.

These results are important from a practical point of view. For example in optics, the current practice for characterizing the performance of optical systems, estimating the quality of algorithms or designing new point spread functions is systematically based on the Cramér-Rao lower-bound. Our results show why and when this methodology might be meaningful and highlight other criteria that must be taken into account jointly.

3. Preliminary facts

In this section, we derive a few basic, yet partly surprising results.

3.1. Existence of minimizers

Before studying the variance of the MLE, it is important to check that it is well defined. In fact, the answer is negative.

Proposition 3.1. *For any positive kernel h vanishing at infinity, and for any noise level $\sigma > 0$, the probability of non existence of the MLE is non zero.*

Proof. Take a positive kernel h vanishing at infinity. In that case, $\mathbb{P}(y_m < 0, \forall 1 \leq m \leq M) > 0$ for all noise levels σ. If all the coordinates of y are negative, it is easy to see that (4) has no minimizer, since $\ell_\epsilon(x) > \lim_{x' \to +\infty} \ell_\epsilon(x')$ for any $x \in \mathbb{R}^D$. $\square$

One way to avoid this problem is to add box constraints on the location x. In that case, the minimizer would exist since we would minimize a continuous function over a compact set. However, in such a scenario, the estimator would then end up on the domain boundary. This simple example highlights a fundamental difficulty in the localization problem: outliers can significantly increase the variance of the MLE.

It will therefore be central to control the probability of $\hat{x}$ being an outlier. Similar observations have already been formulated in [30, 1].

3.2. Identifiability

A second key issue is the identifiability. To illustrate it, let us consider a problem on the real line, i.e., $D = 1$. Consider the family of kernels $h_n(x) \overset{\text{def}}{=} \sin(2\pi n x)$ and set $z_m = m/M$. In that case, the negative log-likelihood function ℓ_ε is periodical and therefore possesses an infinite number of global minimizers. Even restricted on the interval $[0, 1]$, the function possesses at least n global minimizers, therefore the MLE is not well defined again.

The Cramér-Rao bound in that case yields a somewhat contradictory result. Indeed, the bound (7) behaves as $\frac{\sigma^2}{n^2}$ and tends to 0 as $n \to \infty$. By minimizing the Cramér-Rao lower-bound, we would opt for a very oscillatory kernel h, while the problem gets less and less identifiable as n grows.

This simple example highlights the fact that the Cramér-Rao bound only provides a local information. In what follows, we will derive global localization results, requiring more stringent conditions on h.

4. Main results

In this section, we control the probability $\mathbb{P}\left(\|\hat{x} - \bar{x}\|_2 \leq r\right)$ for any given radius $r > 0$. We define the estimator $\hat{x}$ as

$$\hat{x} \in \underset{x \in \Omega}{\operatorname{argmin}}\, \ell_\varepsilon(x) \quad \text{with} \quad \ell_\varepsilon(x) = \frac{1}{2}\left\|h(Z - x) - h(Z - \bar{x}) - \varepsilon\right\|_2^2, \tag{8}$$

where $\Omega \subseteq \mathcal{B}_R \overset{\text{def}}{=} \{x \in \mathbb{R}^D, \|x - \bar{x}\|_2 \leq R\}$ is included in a ball of radius R centered at $\bar{x}$. We prefer minimizing on a compact set Ω rather than $\mathbb{R}^D$ for the following reasons:

(i) In the actual practice, the minimization is usually performed over a compact domain over which the source is assumed to lie. Such a priori knowledge is easily accessible in many applications.

(ii) It ensures the existence of a minimizer (minimization of a continuous function over a compact set).

(iii) It simplifies the presentation significantly.

We could take $R = +\infty$, recovering problem (4), by adding a decay assumption on h. We prefer skipping this aspect for conciseness.

4.1. Technical assumption and geometrical quantities

First of all, let us recall that we work under the assumption $h \in C^2(\mathbb{R}^D)$ which implies the existence of the following two quantities.

- Boundedness:

$$\Lambda \overset{\text{def}}{=} \sup_{x, x' \in \mathcal{B}_R} \|h(Z - x) - h(Z - x')\|, \tag{9}$$

- Lipschitz continuity:

$$L \stackrel{\text{def}}{=} \sup_{x,x' \in \mathcal{B}_R} \frac{\|h(Z-x) - h(Z-x')\|}{\|x-x'\|}. \tag{10}$$

Moreover, in what follows, we will let

$$R_L \stackrel{\text{def}}{=} \frac{\Lambda}{L}.$$

Remark 4.1. *The Lipschitz continuity over $\mathcal{B}_R$ implies that $\Lambda \leq 2LR$. Hence under the stated assumptions, we get $R_L \leq 2R$. However, the radius R could be large, in which case, it can be highly preferable to work with R_L than R.*

In the appendix, we derive a localization result that holds for arbitrary kernels h (Theorem D.1), but which is quite hard to grasp. To present our main results, we will work under the following additional assumption, which significantly eases the presentation and will allow us to draw parallels with the Cramér-Rao lower bound.

Assumption 4.1. *$\bar{x}$ is the unique global minimizer of the noiseless function ℓ_0 and $\ell_0''(\bar{x}) \succ 0$.*

Assumption 4.1 is quite weak. To illustrate it, we can consider a simplified model without sampling, leading to the following result.

Proposition 4.1 (Generality of Assumption 4.1)**.** *Assume that $y = h \star \delta_{\bar{x}} + \varepsilon$, with h belonging to the Sobolev space $W^{1,2}(\mathbb{R}^D) \stackrel{\text{def}}{=} \{f \in L^2(\mathbb{R}^D) \ ; \ f' \in L^2(\mathbb{R}^D)\}$. Then Assumption 4.1 is equivalent to h being non-periodic (there exists no non-zero $v \in \mathbb{R}^D$ such that $h(x+v) = h(x)$ for all $x \in \mathbb{R}^D$).*

Proof. The proof is post-poned to Appendix C. □

In the discrete case however, the study should be handled on a case by case basis since it depends on a complex interplay between the measurement grid and the PSF h.

Under Assumption 4.1, there exists $\mu > 0$ and a radius $R_\mu > 0$ such that

$$\ell_0(x) \geq \begin{cases} \frac{\mu^2}{2}\|x-\bar{x}\|_2^2 & \text{for } x \in \mathcal{B}_{R_\mu}, \\ \frac{\mu^2}{2} R_\mu^2 & \text{for } x \in \mathcal{B}_R \setminus \mathcal{B}_{R_\mu}. \end{cases} \tag{11}$$

The quantities Λ, μ, R_μ and L are related to simple geometrical properties of h (and thus ℓ_0). The strong convexity parameter μ denotes the local curvature of ℓ_0 around $\bar{x}$. Hence, by picking a small radius R_μ, μ can be associated to the minimal eigenvalue of $\ell_0''(\bar{x})$. The parameter L can be associated to the maximal curvature of ℓ_0 on $\mathbb{R}^D$. They are illustrated in Fig. 3 (left) and their behavior for different 1D kernels h is illustrated in Fig. 4.

A refined analysis of these quantities can be carried out case by case as illustrated in the central part of Figure 4. In this example, different choices of values of μ and R_μ are available depending on the sought for precision guarantees.

Example 4.1 (Reference kernel)**.** *We define a simple kernel h_{ref} for which our geometrical quantities are trivial. It will help us later for interpreting our results. It corresponds to the hat function in the second row of Fig. 2. Set $D = 1$ and*

$$h_{\text{ref}} \stackrel{\text{def}}{=} (1-|x|)_+, \quad M \in \mathbb{N}, \quad Z = (-M, -M+1, \ldots, M-1, M), \quad \bar{x} = 0. \tag{12}$$

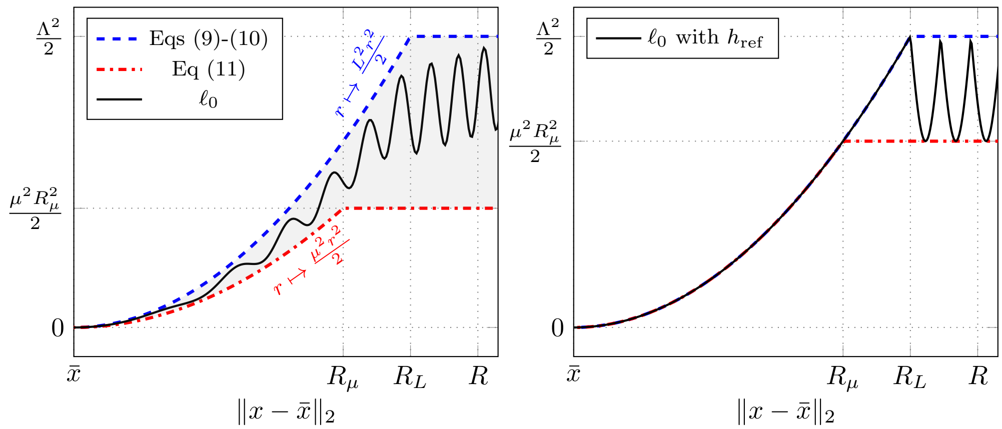


Figure 3: *Left:* Illustration of quantities Λ, L, μ and R_μ. The noiseless cost function ℓ_0 is sandwiched between two quadratic functions near $\bar{x}$ and between two constants far away from the origin. *Right:* The specific case h_{ref} where $\mu = L$.

For this choice of h and Z, the noiseless function ℓ_0 has a simple analytical formula:

$$\forall x \in [-M, M], \quad \ell_0^{\text{ref}}(x) = \begin{cases} x^2 & \text{if } |x| \leq 1, \\ 1 - \{x\} + \{x\}^2 & \text{if } 1 \leq |x| \leq M. \end{cases} \tag{13}$$

where $\{x\} \stackrel{\text{def}}{=} x - \lfloor x \rfloor$ is the fractional part of x. To get this result, we exploit the fact that h_{ref} is even and continuous, and so is ℓ_0^{ref}. Moreover, h_{ref} is linear on each interval $[m, m+1]$ with $-M \leq m \leq M-1$, thus ℓ_0^{ref} is quadratic on each of these intervals. The remaining computation is the interpolation from known values of ℓ_0^{ref} at points $(t/2)_{t=-2M}^{2M}$. The graph of ℓ_0^{ref} is depicted in Fig. 3 (right). Some elementary calculations yield:

$$L^{\text{ref}} = \mu^{\text{ref}} = \Lambda^{\text{ref}} = \sqrt{2}, \quad R_\mu^{\text{ref}} = \frac{\sqrt{3}}{2}, \quad R_L^{\text{ref}} = 1. \tag{14}$$

4.2. The main results

We are now ready to present our main results.

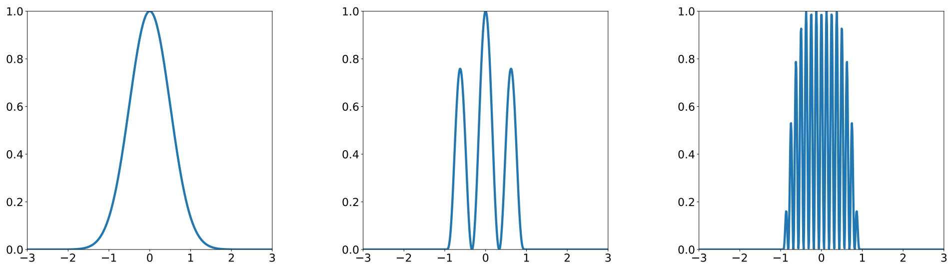
Theorem 4.1 (Sufficient conditions)**.** *Assume that $\varepsilon \sim \mathcal{N}(0, \sigma^2 \text{Id})$ and that the Assumption 4.1 is satisfied. Set $\rho > 0$, a radius $0 < r \leq R$ and $I = \lceil \log_2(1 + \log_2(R_\mu/r)) \rceil$. Consider the two conditions below, where $c > 0$ is a universal constant*:*

$$\frac{4}{\mu^2}\sigma L \left(c\sqrt{D} + \sqrt{\rho^2 + 2\ln(I)} \right) \leq r \tag{Cond$_1$}$$

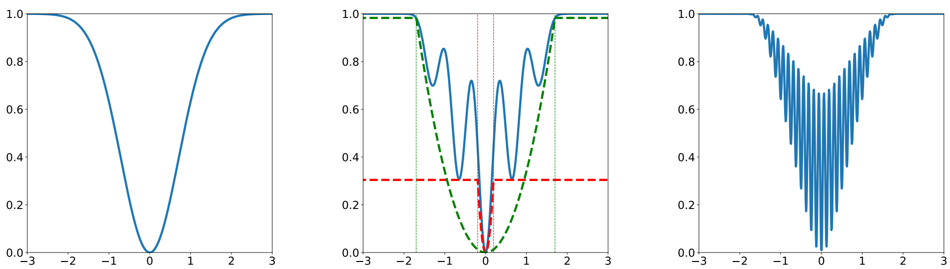
$$\frac{2}{\mu^2}\sigma L \left(c\sqrt{D}\sqrt{\ln\left(\frac{3R}{R_L}\right)} + \sqrt{\rho^2 + 2\ln(I)} \right) \leq \frac{R_\mu^2}{R_L}. \tag{Cond$_2$}$$

Then under either of the following conditions

* The proof of the result heavily relies on bounding the supremum of a Gaussian process. To the best of our knowledge, it is currently out of reach to control the multiplicative constants precisely.



(a) $h = \exp(-(\cdot)^2/(2\sigma^2))$ (b) $h = \cos(\omega_1 \cdot) \cdot \text{bump}$ (c) $h = \cos(\omega_2 \cdot) \cdot \text{bump}$



(d) ℓ_0 (e) ℓ_0 (f) ℓ_0

Figure 4: Different kernels h (top) and the corresponding functions ℓ_0 (bottom) for $\bar{x} = 0$. Left: a Gaussian with standard deviation 0.5. Middle and right: cosines with two different frequencies multiplied by a bump function. For the slowly oscillating cosine associated to the frequency ω_1, we show that different choices of μ and R_μ in Assumption 4.1 are possible with the green and red dashed curves. We can choose a large value of μ with a small radius R_μ (see the red dashed lower bound). As an effect, the bounds we obtain in Theorem 4.1 will ensure that a highly accuracte estimate can be obtained for low noise regimes. However, they will suffer from a high probability of false detection with higher noise levels. On the other hand, we can choose a smaller value of μ (see the green dashed lower bound) with a larger radius R_μ. As a result, our theorem provides sharper estimates for high noise regimes. The Cramér-Rao lower bound on its side is based on the curvature of ℓ_0 at $\bar{x}$ only. It therefore accomodates only with small noise levels and leads to an overoptimistic conclusion for the mean squared error of the estimator. The function on the right illustrates a pathological kernel for the Cramér-Rao bound: the bound converges to zero due to the high curvature at the origin, but there is a very high probability of false detection since shifted kernels correlate highly with h. Our theorems allow us to use a more precise lower enveloppe that provides sharp guarantees on the estimator location.

- $r \leq \frac{R_\mu}{2}$, $R \geq \frac{R_L}{2}$, Cond_1 *and* Cond_2.
- $r \leq \frac{R_\mu}{2}$, $R \leq \frac{R_L}{2}$ *and* Cond_1.
- $r \geq \frac{R_\mu}{2}$ *and* Cond_2.

the following inequality holds

$$\mathbb{P}\left(\|\bar{x} - \hat{x}\|_2 \leq r\right) \geq 1 - \exp\left(-\frac{\rho^2}{2}\right). \tag{15}$$

The complete proof is given in Appendix D. We will provide some insights in the next section. For now, let us present a similar result showing that the conditions in Theorem 4.1 are not only sufficient but also nearly *necessary*.

Theorem 4.2 (Necessary conditions)**.** *Under the same assumptions as in Theorem 4.1, $R > R_\mu$ and*

$$\|h(Z-x) - h(Z-x')\|_2 \geq \mu \min\left(\|x - x'\|_2, R_\mu\right) \quad \forall x, x' \in \mathcal{B}_R. \tag{16}$$

- *We have $\mathbb{P}\left(\hat{x} \notin \mathcal{B}_r\right) \geq 1 - \exp(-\rho_1^2/2) - \Phi(\rho_2)$, where Φ is the cumulative distribution function of the normal distribution, under the condition*

$$r < \frac{\left(\frac{\sqrt{\ln 2}}{8}\mu\sqrt{D-1} - \rho_2 L\right)}{L\left(c\sqrt{D} + \rho_1\right)} \min\left(R_\mu,\ \sigma\frac{\sqrt{\ln 2}}{4}\frac{\mu}{L^2}\sqrt{D-1}\right). \tag{Cond_1'}$$

- *We have $\mathbb{P}\left(\hat{x} \notin \mathcal{B}_r\right) \geq 1 - \exp(-\rho_1^2/2) - \exp(-\rho_2^2/2)$ under the condition*

$$\sigma\sqrt{\ln\left(\frac{R}{R_\mu}\right)} > \frac{R_L^2 L^2}{\sqrt{D}R_\mu\mu} + \frac{2\sigma[\rho_2 R_L L + Lrc\sqrt{D} + \rho_1]}{\sqrt{D}R_\mu\mu}. \tag{Cond_2'}$$

The complete proof is given in Appendix E.

4.3. An informal proof

In this section, we provide the essential ingredients behind Theorems 4.1 and 4.2 since they shed light on the problem's geometry. The starting point of these proofs is the following decomposition

$$\ell_\varepsilon = \ell_0 - \Delta_\varepsilon + \frac{1}{2}\|\varepsilon\|_2^2 \quad \text{where} \quad \Delta_\varepsilon(x) \stackrel{\text{def}}{=} \langle h(Z-x) - h(Z-\bar{x}), \varepsilon\rangle,\ \forall x \in \mathbb{R}^D.$$

The term $\frac{1}{2}\|\varepsilon\|_2^2$ is constant with respect to x and does not change the location of the minimizer. The term Δ_ε is a centered random Gaussian process with

$$\Delta_\varepsilon(x) \sim \mathcal{N}(0, 2\sigma^2\ell_0(x)) \tag{17}$$

Intuitions behind Theorem 4.1 A sufficient condition for success ($\hat{x} \in \mathcal{B}_r$) is that

$$0 = \ell_0(\bar{x}) - \Delta_\varepsilon(\bar{x}) < \inf_{x \in \mathcal{B}_R \setminus \mathcal{B}_r} \ell_0(x) - \Delta_\varepsilon(x), \tag{18}$$

as illustrated on Fig. 5 (left). Then, given that

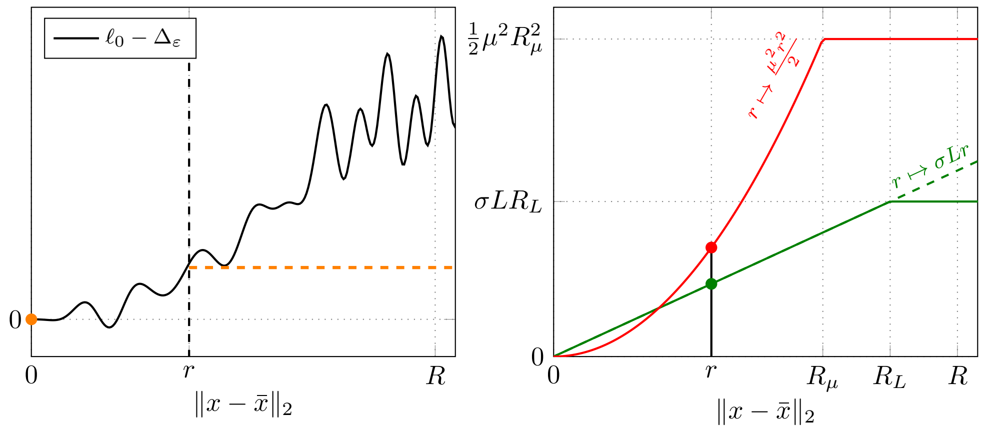


Figure 5: **Intuition behind Theorem 4.1.** *Left:* Success ($\hat{x} \in \mathcal{B}_r$) occurs if there exists $x \in \mathcal{B}_r$ (e.g., $x = \bar{x}$) such that $\ell_0(x) - \Delta_\varepsilon(x)$ (orange point) is lower than the infimum of $\ell_0 - \Delta_\varepsilon$ over $\mathcal{B}_R \setminus \mathcal{B}_r$ (orange dashed line). *Right:* The green curve is an upper-bound of the typical amplitude of Δ_ε. The red curve is a lower-bound of ℓ_0. An (informal) sufficient condition for $\hat{x} \in \mathcal{B}_r$ is that the red curve dominates the green one at r.

- $\Delta_\varepsilon(x) \lesssim \sigma L \min(\|x - \bar{x}\|_2, R_L)$ with high probability (from (17) and regularity of h (9) and (10)),
- $\ell_0(x) \geq \frac{\mu^2}{2} \min(\|x - \bar{x}\|_2^2, R_\mu^2)$ (from Assumption 4.1),

we can simplify (18) as

$$\sigma L \min(\|x - \bar{x}\|_2, R_L) \lesssim \frac{\mu^2}{2} \min(\|x - \bar{x}\|_2^2, R_\mu^2), \quad \text{for all} \quad \|x - \bar{x}\|_2 \geq r.$$

This is illustrated in Fig. 5 (right) with the quadratic curve being higher than the linear one for all $\|x - \bar{x}\|_2 \geq r$. This condition can be decomposed as $2\sigma L/\mu^2 < r$ and $2\sigma L/\mu^2 < R_\mu^2/R_L$ which correspond in essence to Cond_1 and Cond_2 in Theorem 4.1. The difference lies in additive logarithmic terms which appear, since the probability should not be controlled pointwise, but uniformly in $\mathcal{B}_R \setminus \mathcal{B}_r$. This uniform control is handled using discretization techniques combined with results on the suprema of random processes [27].

Intuitions behind Theorem 4.2 A sufficient condition for failure ($\hat{x} \notin \mathcal{B}_r$) is that there exists $x' \in \mathcal{B}_R \setminus \mathcal{B}_r$ with $t = \|x' - \bar{x}\|_2$ such that

$$0 \underset{r \text{ small}}{\approx} \inf_{x \in \mathcal{B}_r} \ell_0(x) - \Delta_\varepsilon(x) > \ell_0(x') - \Delta_\varepsilon(x'). \tag{19}$$

This condition is illustrated on Fig. 6 (left). Then, given that

- $\Delta_\varepsilon(x') \gtrsim \sigma\mu \min(t, R_\mu)$ with probability close to 1/2 (from (17) and since Δ_ε is symmetric),
- $\ell_0(x') \leq \frac{L^2}{2} \min(t^2, R_L^2)$ (from Assumption 4.1),

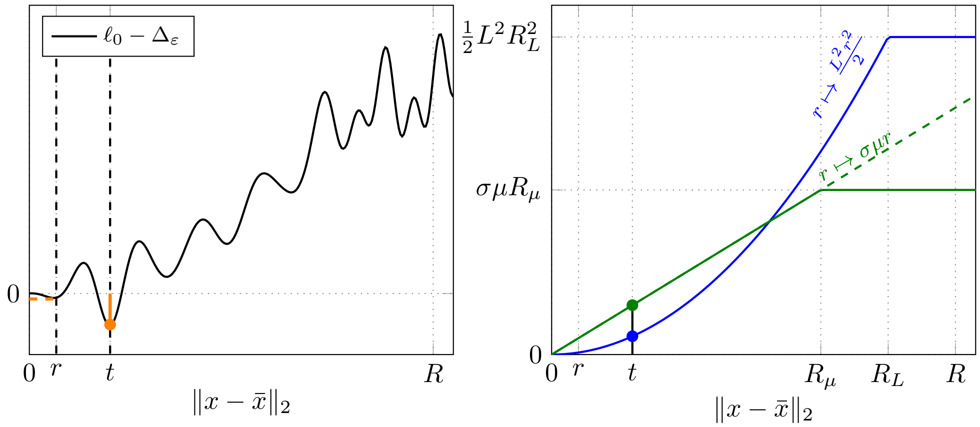


Figure 6: **Intuition behind Theorem 4.2.** *Left:* Failure ($\hat{x} \notin \mathcal{B}_r$) occurs if there exists $x_t \in \mathcal{B}_R \setminus \mathcal{B}_r$ (with $t = \|x_t - \bar{x}\|_2$) such that $\ell_0(x_t) - \Delta_\varepsilon(x_t)$ (orange point) is lower than the infimum of $\ell_0 - \Delta_\varepsilon$ over $\mathcal{B}_r$ (orange dashed line). *Right:* The green curve is a lower-bound of the typical amplitude of Δ_ε. The blue curve is an upper-bound of ℓ_0. An (informal) sufficient condition for $\hat{x} \notin \mathcal{B}_r$ is that the green curve dominates the blue one at t.

we can simplify (19) as

$$\sigma\mu \min(t, R_\mu) \gtrsim \frac{L^2}{2} \min\left(t^2, R_L^2\right).$$

This is illustrated on Fig. 6 (right) with the quadratic curve being lower than the linear one at $\|x' - \bar{x}\|_2 = t$. Taking $t = 2r$, we get the two conditions $r < \sigma\mu/L^2$ and $\sigma > L^2R_L^2/(2\mu R_\mu)$ which correspond to Cond_1' and Cond_2' in Theorem 4.2 (for small r, ρ_1, and ρ_2). A more careful analysis involves the additive terms appearing in the theorem.

4.4. Interpretations and simplifications

In this section, we aim at explaining the different ingredients from the theorems above.

The key geometric features The proposed analysis emphasizes the role of a few key geometrical quantities:

- The radius R_μ of quadratic growth and its associated parameter μ

$$\mu^2 \geq \inf_{x \in \mathcal{B}_{R_\mu}} \lambda_{\min}(\ell_0''(x)). \tag{20}$$

- The Lipschitz constant of the gradient of ℓ_0:

$$L^2 \leq \sup_{x \in \mathcal{B}_R} \lambda_{\max}(\ell_0''(x)). \tag{21}$$

- The local conditioning of ℓ_0 around $\bar{x}$: $\kappa \stackrel{\text{def}}{=} \frac{L}{\mu} \geq 1$.

- The square root of the quotient between the upper and lower bound of ℓ_0 far from $\bar{x}$: $\theta \stackrel{\text{def}}{=} \frac{R_L L}{R_\mu \mu} \geq 1$.

The two last theorems can be summarized informally as follows. For sufficiently small r, we can get $\hat{x} \in \mathcal{B}_r$ with large probability under the following

$$\bullet \text{ Sufficient condition:} \quad \sigma \lesssim \frac{\mu}{\sqrt{D}} \min\left(r\kappa^{-1}, R_\mu \theta^{-1} \ln\left(\frac{R}{R_L}\right)^{-1/2}\right) \tag{22}$$

$$\bullet \text{ Necessary condition:} \quad \sigma \lesssim \frac{L}{\sqrt{D}} \min\left(r\kappa^{2}, R_L \theta \ln\left(\frac{R}{R_\mu}\right)^{-1/2}\right). \tag{23}$$

Only the left term in the minimum above plays a role in the control of the local error around the true location. The term $\ln(R)^{-1/2}$ comes from the fact that as the radius R increases, the probability of false detection far away from $\bar{x}$ increases. However, it does so at a very moderate rate.

Tightness The two conditions differ mostly from the conditioning factors κ and θ. They become *equivalent up to multiplicative factors* for the reference kernel h_{ref}. Indeed, set $R = M/2 - 1$, which amounts to looking for the source only around the sampled points. For this kernel, they both read

$$\sigma\sqrt{D} \lesssim \min\left(r, \sqrt{\ln(M)}\right).$$

This proves the tightness of the theorem.

The term $\sqrt{D\ln(R/R_L)}$ This term is not a proof artefact and needs to be accounted for. To illustrate this, let us consider the simple function $h(x) = (1 - L\|x\|_\infty)_+$. We assume for simplicity that $L = 1/K$ for some integer $K \in \mathbb{N}$ and that the sampling set is the Euclidean grid $Z = [\![1, M^{1/D}]\!]^D$, where we assume that $M^{1/D}$ is an integer. The function h is L-Lipschitz continuous, has a support $[-K, K]^D$ and satisfies $R_L \propto \frac{1}{L}$. By definition, $\Delta_\varepsilon(x) = \langle h(Z - x), \varepsilon\rangle - \langle h(Z - \bar{x}), \varepsilon\rangle$. Since h is compactly supported, notice that the random variables $\langle h(Z - x), \varepsilon\rangle$ and $\langle h(Z - x'), \varepsilon\rangle$ are *independent* for $|x - x'| \geq 2K$. In addition, for $x, x' \in Z \cap (2K\mathbb{N})^D$, they are independent and *identically distributed Gaussian* random variables.

We can pack $O(RL)^D$ balls of radius K in a ball of radius R. To find a lower-bound on the supremum of $\Delta_\varepsilon(x)$, we can therefore use Sudakov's inequality:

$$X \sim \mathcal{N}(0, \mathrm{Id}_N) \Rightarrow \mathbb{E}\left[\max_{1\leq n\leq N} X_n\right] \gtrsim \frac{\sqrt{2\ln(N)}}{2}. \tag{24}$$

Applied to our setting this yields:

$$\begin{aligned}
\mathbb{E}\left[\sup_{x\in\mathcal{B}_R} \Delta_\varepsilon(x)\right] &\geq \mathbb{E}\left[\max_{x\in Z\cap(2K\mathbb{N})^D} \Delta_\varepsilon(x)\right] \\
&= \mathbb{E}\left[\max_{x\in Z\cap(2K\mathbb{N})^D} \langle h(Z - x), \varepsilon\rangle\right] \qquad \gtrsim \sqrt{D\ln(R/R_L)}.
\end{aligned}$$

The term $\ln(I)$ The term $\ln(I) = \ln(\lceil \log_2(1+\log_2(R_\mu/r))\rceil)$ is likely an artefact of the proof. It appears when we derive a tail bound (see (D.8)) for the term $\sup_{x\in\mathcal{B}_r} \Delta_\varepsilon(x)$. In [6, 8], the authors propose a different proof strategy based on the Rice theorem, which is more elegant, but restricted to the 1D case. We did not find a way to extend the proof to the multivariate case. In 1D, the results coincide, apart from constants and the logarithmic term $\ln(I)$. It can be safely neglected in practice. For instance $\ln(\lceil \log_2(1 + \log_2(10^{100000}))\rceil) < 3$!

4.5. Notable consequences

Theorem 4.1 has a few interesting consequences.

Phase transition An important consequence of Theorem 4.1 is a phase transition behavior. Whenever (Cond_1) and (Cond_2) are satisfied with a sufficiently high value of ρ (say $\rho = 3$), it becomes very unlikely to see the global minimizer $\hat{x}$ escaping from the ball $\mathcal{B}(\bar{x}, r)$. In applications, we would typically set a small value of r (e.g. one tenth of a pixel in single molecule localization) and the theorem tells that whenever the condition is satisfied, the estimator will nearly always succeed.

Relationship to Cramér-Rao The condition (Cond_1) is strongly connected to the Cramér-Rao lower-bound. Using assumptions (10) and (11), we obtain

$$\mu^2 \leq \lambda_{\min}(\ell_0''(\bar{x})) \leq \lambda_{\max}(\ell_0''(\bar{x})) \leq L^2. \tag{25}$$

Given that $\ell_0''(\bar{x}) = \sum_{m=1}^M h'(z_m - \bar{x})\ h'(z_m - \bar{x})^T$, we get

$$\mu^2 \leq \lambda_{\min}\left(\ell_0''(\bar{x})\right) \leq \frac{\|h'(Z-\bar{x})\|_F^2}{D} \leq \lambda_{\max}\left(\ell_0''(\bar{x})\right) \leq L^2 \tag{26}$$

since for any matrix $A \in \mathbb{R}^{D\times M}$ with $D \leq M$

$$\|A\|_F^2 = \sum_{d=1}^D \sigma_d^2(A) = \sum_{d=1}^D \lambda_d(AA^T) \tag{27}$$

where $\sigma_d(A)$ and $\lambda_d(A)$ are respectively the singular values and the eigenvalues of A. Finally, from (Cond_1) we get that asymptotically,

$$r \gtrsim \frac{\sqrt{D}\sigma L}{\mu^2} \gtrsim \frac{D\cdot\sigma}{\|h'(Z-\bar{x})\|_F} = \sqrt{\mathrm{CR}_\sigma}. \tag{28}$$

This is in line with the asymptotic analysis existing in the literature [20, 23].

PSF engineering - How to optimize a kernel? A few authors proposed to optimize the point spread function of optical systems by maximizing $\|h'(Z-\bar{x})\|_F$, motivated by the Cramér-Rao bound, see e.g. [26, 15]. The second part of Theorem 2.1, shows that this might not be enough. Our results highlight that other facts must be taken into account.

Let us assume that we wish to obtain a localization precision of order r. Looking only at the sufficient condition, the most important factors are then:

- The quadratic growth parameter μ in $\mathcal{B}_R$. A good upper-bound for this term is:

$$\mu(r) \stackrel{\text{def}}{=} \inf_{x\in\mathcal{B}_r} \|h'(Z-x)\|_F. \tag{29}$$

  This term is essentially equivalent to the Cramér-Rao bound, except that it needs to be controlled uniformly in $\mathcal{B}_r$.

- We also need to ensure that $\ell_0(x)$ is sufficiently large for all $x \in \mathcal{B}_R \setminus \mathcal{B}_r$. This condition is there to ensure the identifiability of the problem. For example, this condition discards the pathological sine kernel discussed in Section 3.2. This would not be taken into account using the Cramér-Rao bound only.
- Finally, taking R too large increases the probability of false positives at the slow rate $\sqrt{\ln(R)}$. This means that cameras with large field of views can be safely used without increasing the false detection rate significantly.

Notice that the factors mentioned above do not include any support size or decay rate constraints, which are usually added for PSF engineering [26]. Hence, unlocalized kernels could yield interesting results, in the low density regime, where only scattered sources are present. Such unlocalized PSFs are physically realistic [33, 17] and our results suggest that they may be of interest in the context of SMLM for instance. To our knowledge, this has not been investigated to date.

4.6. Analysis for well sampled bandlimited kernels

In this section, we analyze the theorems in the specific case of well sampled bandlimited kernels. This case is of major importance in optics applications. Let us introduce the Fourier transform defined for all $f \in L^2(\mathbb{R}^D)$ by

$$\mathcal{F}(f)(\omega) \stackrel{\text{def}}{=} \frac{1}{(2\pi)^{D/2}} \int_{\mathbb{R}^D} f(x) \exp(-i\langle x, \omega\rangle)\, dx. \tag{30}$$

Let us recall the Plancherel formula $\|f\|_{L^2(\mathbb{R}^D)} = \|\mathcal{F}(f)\|_{L^2(\mathbb{R}^D)}$.

Definition 4.1 (Paley-Wiener space)**.** *For $W > 0$ we say that $f \in L^2(\mathbb{R}^D)$ is W-bandlimited if $\operatorname{supp}(\mathcal{F}(f)) \subseteq [-W, W]^D$. We let $\mathcal{PW}(W)$ denote the set of W-bandlimited functions in $L^2(\mathbb{R}^D)$.*

A bandlimited function belongs to $C^\infty(\mathbb{R}^D)$. Let us recall the following fundamental result.

Theorem 4.3 (A variant of Whittaker's theorem [18])**.** *Assume that $f, g \in \mathcal{PW}(\pi)$. Set a sampling step $\tau \leq 1$. Then for all $x \in \mathbb{R}^D$, we have:*

$$\langle f, g\rangle_{L^2(\mathbb{R}^D)} = \left(\frac{\tau}{2\pi}\right)^D \sum_{z\in\tau\mathbb{Z}^D} f(z-x)g(z-x). \tag{31}$$

In particular, taking $f = g$, we get

$$\|f\|^2_{L^2(\mathbb{R}^D)} = \left(\frac{\tau}{2\pi}\right)^D \sum_{z\in\tau\mathbb{Z}^D} f(z-x)^2. \tag{32}$$

Proof. The proof is given in Appendix F for completeness. □

Now assume that $h \in \mathcal{PW}(\pi)$ and that it is sampled on $Z = [-R, R]^D \cap \tau\mathbb{Z}^D$, where $\tau < 1$ is the grid size ($1/\tau$ is the oversampling factor). By Theorem 4.3, we have for all x

$$\|h(Z-x)\|_2^2 = \sum_{z\in Z} h^2(z-x) = \left(\frac{2\pi}{\tau}\right)^D \|h\|^2_{L^2(\mathbb{R}^D)} - \sum_{z\in\tau\mathbb{Z}^D\setminus Z} h^2(z-x)$$
$$\approx \left(\frac{2\pi}{\tau}\right)^D \|h\|^2_{L^2(\mathbb{R}^D)}$$

for sufficiently large R and decaying h. Similarly, we have for all x, x'

$$\|h'(Z-x)\|_2^2 \approx \left(\frac{2\pi}{\tau}\right)^D \|h'\|^2_{L^2(\mathbb{R}^D)}$$
$$\langle h(Z-x), h(Z-x')\rangle \approx \left(\frac{2\pi}{\tau}\right)^D \langle h(\cdot - x), h(\cdot - x')\rangle_{L^2(\mathbb{R}^D)}$$

using the fact that $h \in \mathcal{PW}(\pi) \Rightarrow h' \in \mathcal{PW}(\pi)$. We then get that

$$\ell_0(x) \approx \left(\frac{2\pi}{\tau}\right)^D \left(\|h\|^2_{L^2(\mathbb{R}^D)} - \langle h(\cdot - x), h(\cdot - \bar{x})\rangle_{L^2(\mathbb{R}^D)}\right).$$

Now let Λ, L, and μ be defined according to equations (9) to (11), and 4.1 for $x \mapsto \ell_0(x)\tau^D$ so that they are independent of τ. Then, the conditions (22) and (23) read

$$\sigma \lesssim \frac{\mu}{\tau^{D/2}} \min\left(\frac{r\kappa^{-1}}{\sqrt{D}}, \frac{R_\mu\theta^{-1}}{\sqrt{D}} \ln\left(\frac{R}{R_L}\right)^{-1/2}\right) \tag{33}$$
$$\sigma \lesssim \frac{L}{\tau^{D/2}} \min\left(r\kappa^2, \frac{R_L\theta}{\sqrt{D}} \ln\left(\frac{R}{R_\mu}\right)^{-1/2}\right). \tag{34}$$

where κ, θ, R_L, and R_μ are independent of τ.

Hence, for well sampled bandlimited kernels, we have the following qualitative behaviour:

- The performance does not depend on $\bar{x}$ (as $\|h'(Z - \cdot)\|_2$ is nearly constant).
- Over-sampling with $\tau < 1$ leads to improved performances (from (33) and (34)).

5. Acknowledgments

This work was supported by the ANR Micro-Blind (grant ANR-21-CE48-0008), the University Research School EUR-MINT (State support managed by the National Research Agency for Future Investments program bearing the reference ANR-18-EURE-0023) and by the ANR LabEx CIMI (grant ANR-11-LABX-0040) within the French State Programme "Investissements d'Avenir". P. Weiss acknowledges the support of AI Interdisciplinary Institute ANITI funding, through the French "Investing for the Future— PIA3" program under the Grant Agreement ANR-19-PI3A-0004. The authors thank Yaw Tze Bong and Yann Delaporte for a preliminary study of this problem during a summer internship. They are grateful to the anonymous reviewers who really helped improving the overall presentation and guided us to relevant references.

Appendices

A. Derivatives of ℓ_ε

Let us start with a set of identities that will be used continuously throughout the proofs. We have

$$\ell_\varepsilon(x) = \frac{1}{2}\left\|h(Z-x)-y\right\|_2^2, \tag{A.1}$$

$$\ell'_\varepsilon(x) = -h'(Z-x)(h(Z-x)-y), \tag{A.2}$$

$$\ell''_\varepsilon(x) = h''(Z-x)(h(Z-x)-y) + h'(Z-x)h'(Z-x)^T. \tag{A.3}$$

Thus evaluating at $x = \bar{x}$ gives

$$\ell_\varepsilon(\bar{x}) = \frac{\|\varepsilon\|^2}{2}, \tag{A.4}$$

$$\ell'_\varepsilon(\bar{x}) = \sum_{m=1}^{M} \varepsilon_m\ h'(z_m-\bar{x}), \tag{A.5}$$

$$\ell''_\varepsilon(\bar{x}) = \sum_{m=1}^{M} \left(h'(z_m-\bar{x})\ h'(z_m-\bar{x})^T - \varepsilon_m\ h''(z_m-\bar{x})\right). \tag{A.6}$$

B. Proof of Theorem 2.1

B.1. Establishing the Cramér-Rao lower-bound

By [16, (3.20)], we know that the covariance matrix of any unbiased estimator $\hat{x}$ satisfies

$$\mathrm{cov}(\hat{x}) \succcurlyeq I^{-1}(\bar{x}) \quad \text{with} \quad [I(x)]_{d,d'} \overset{\text{def}}{=} \mathbb{E}\left[\frac{\partial \ln p(y|x)}{\partial x_d}\cdot\frac{\partial \ln p(y|x)}{\partial x_{d'}}\right]. \tag{B.1}$$

In our case, we have

$$p(y|x) \propto \exp\left(-\frac{\|y-h(Z-x)\|_2^2}{2\sigma^2}\right). \tag{B.2}$$

Hence

$$\frac{\partial \ln p(y|x)}{\partial x_d} = \frac{1}{\sigma^2}\sum_{m=1}^{M}(h(z_m-x)-y_m)\cdot h'_d(z_m-x), \tag{B.3}$$

where we use the notation $h'_d = \partial_d h$ in this proof. This yields for any unbiased estimator $\hat{x}$

$$\begin{aligned}
\mathbb{E}\left[|\hat{x}_d-\bar{x}_d|_2^2\right] &\geq \mathbb{E}\left[\frac{\partial \ln p(y|\bar{x})}{\partial x_d}\cdot\frac{\partial \ln p(y|\bar{x})}{\partial x_d}\right]^{-1} \\
&= \sigma^4\cdot\mathbb{E}\left[\left(\sum_{m=1}^{M}-\varepsilon_m h'_d(z_m-\bar{x})\right)\left(\sum_{m=1}^{M}-\varepsilon_m h'_d(z_m-\bar{x})\right)\right]^{-1} \\
&= \sigma^4\cdot\mathbb{E}\left[\sum_{m=1}^{M}\varepsilon_m^2 h'_d(z_m-\bar{x})^2\right]^{-1} = \frac{\sigma^2}{\|h'_d(Z-\bar{x})\|_2^2}.
\end{aligned}$$

Summing over d yields

$$\mathbb{E}\left[\|\hat{x}-\bar{x}\|_2^2\right] \geq \sum_{d=1}^{D} \frac{\sigma^2}{\|h'_d(Z-\bar{x})\|_2^2} \geq \frac{D^2\sigma^2}{\|h'(Z-\bar{x})\|_F^2}, \tag{B.4}$$

where the last inequality is obtained by the fact that the arithmetic mean is greater than the harmonic one (i.e., $\frac{1}{D}\sum_{d=1}^{D} a_d \geq D(\sum_{d=1}^{D} \frac{1}{a_d})^{-1}$).

B.2. Necessary and sufficient condition for attainment

Now, let us prove that the Cramér-Rao bound can be attained only in dimension $D = 1$ and for affine kernels h. To this end, we recall the second part of [16, Thm. 3.2], adapted to our setting.

Proposition B.1. *An unbiased estimator $\hat{x}$ of $\bar{x}$ that attains the Cramér-Rao bound exists if and only if*

$$\nabla_{\bar{x}} \ln p(y|\bar{x}) = I(\bar{x}) \cdot (g(y) - \bar{x}), \tag{B.5}$$

for all y, $\bar{x}$ and some D-dimensional functions g and $D \times D$ matrix I. The optimal estimator is then $\hat{x} = g(y)$ and the covariance matrix is $I(\bar{x})^{-1}$.

Corollary B.1 (Necessary conditions for attaining Cramér-Rao)**.** *In our case, there exists an unbiased estimator $\hat{x}$ of $\bar{x}$ that attains the Cramér-Rao bound if and only if $D = 1$ and h is an affine non constant function.*

Proof. Firstly, assume that such an estimator that attains the Cramér-Rao bound exists. Let's show that h should be affine. In our case, the condition from the general case reads

$$\frac{1}{\sigma^2} J\tilde{h}(\bar{x})^T \cdot (h(Z-\bar{x}) - y) = I(\bar{x}) \cdot (g(y) - \bar{x}) \tag{B.6}$$

where we have introduced the following notations

$$\begin{array}{rcccl} \tilde{h} & : & \mathbb{R}^D & \longrightarrow & \mathbb{R}^M \\ & & \bar{x} & \longmapsto & h(Z-\bar{x}) \stackrel{\text{def}}{=} (h(z_m - \bar{x}))_m \end{array}$$

and for a function $f : \mathbb{R}^D \longrightarrow \mathbb{R}^I$ we have fixed the Jacobian of f to be

$$Jf(x) \stackrel{\text{def}}{=} \left(\frac{\partial f_i}{\partial x_d}\right)_{i,d} = (\nabla f_1(x) \ \ldots \ \nabla f_I(x))^T \in \mathbb{R}^{I \times D}.$$

Differentiating the equality with respect to y yields

$$-\frac{1}{\sigma^2} J\tilde{h}(\bar{x})^T = I(\bar{x}) \cdot Jg(y). \tag{B.7}$$

The left-hand side does not depend on y. Hence, $Jg(y)$ is constant, meaning that

$$g(y) = A \cdot y + b \tag{B.8}$$

for some $A \in \mathbb{R}^{D \times M}$ and $b \in \mathbb{R}^D$. This simplifies equation (B.7) as

$$J\tilde{h}(\bar{x})^T = -\sigma^2 I(\bar{x}) \cdot A. \tag{B.9}$$

Now let us rewrite (B.6) by replacing $g(y)$ and $J\tilde{h}(\bar{x})^T$:

$$-I(\bar{x}) \cdot A \cdot (h(Z - \bar{x}) - y) = I(\bar{x}) \cdot (A \cdot y + b - \bar{x}).$$

Since from the general case, $I(\bar{x})^{-1}$ is well defined and is the covariance matrix, one has that $I(\bar{x})$ is invertible. The previous equality simplifies as

$$-A \cdot h(Z - \bar{x}) = b - \bar{x}. \tag{B.10}$$

Differentiating this expression with respect to $\bar{x}$ and injecting (B.9) gives

$$-\sigma^2 I(\bar{x}) \cdot AA^T = \mathrm{Id}_D \tag{B.11}$$

which implies that $I(\bar{x})$ does not depends on $\bar{x}$. With (B.9), we finally get that $\tilde{h}$ (and thus h) is affine.

To complete the proof, it remains to show that $D = 1$ and that h cannot be constant. As h is affine, let us set $h : x \mapsto \langle a, x\rangle + \beta$ with $a \in \mathbb{R}^D$ and $\beta \in \mathbb{R}$. Then

$$J\tilde{h}(\bar{x})^T = [a \ a \ \cdots \ a] \in \mathbb{R}^{D\times M}. \tag{B.12}$$

We then deduce two facts from the invertibility of $I(\bar{x})$. First, from (B.11) AA^T is also invertible. Second, from (B.9) there exists $c \in \mathbb{R}^D$ such that

$$A = c \cdot 1^T \in \mathbb{R}^{D\times M} \quad \text{and} \quad AA^T = c \cdot 1^T \cdot 1 \cdot c^T = M \cdot cc^T.$$

Clearly, these two properties can be satisfied at the same time only if $D = 1$ and $a \neq 0$. This shows that $D = 1$ and h is affine non constant.

Conversely, any affine non constant function h in the case $D = 1$ attains Cramér-Rao bound since $I(\bar{x}) = Ma^2$ and $g(y) = \frac{\beta}{a} + \frac{1}{M}\sum_{m=1}^{M} h\left(z_m - \frac{y_m}{a}\right)$ are appropriate. □

C. Proof of Proposition 4.1

Proof. The function ℓ_0 and its derivatives read for any $v \in \mathbb{R}^D$

$$\begin{aligned}
\ell_0(x) &= \frac{1}{2}\int_{\mathbb{R}^D} \left(h(z-x) - h(z-\bar{x})\right)^2 \mathrm{d}z \\
\ell_0'(x)v &= -\int_{\mathbb{R}^D} \left(h(z-x) - h(z-\bar{x})\right) \cdot \langle h'(z-x), v\rangle \ \mathrm{d}z \\
\ell_0''(\bar{x})(v,v) &= \int_{\mathbb{R}^D} \langle h'(z), v\rangle^2 \ \mathrm{d}z.
\end{aligned}$$

If x is another minimizer of ℓ_0, then $\ell_0(x) = 0$ and thus $h(z - x) = h(z - \bar{x})$ for almost every $z \in \mathbb{R}^D$. This is equivalent to $h(z) = h(z + (x - \bar{x}))$ which means that h is $(x - \bar{x})$-periodic.

Let's show that $\ell_0''(\bar{x}) \succ 0 \iff h \neq 0$. We proceed by contraposition

$$\begin{aligned}
\ell_0''(\bar{x}) = 0 &\iff \ell_0''(\bar{x})(v,v) = 0, \ \forall v \in \mathbb{R}^D \\
&\iff \langle h'(z), v\rangle = 0, \ \forall v, z \in \mathbb{R}^D \\
&\iff h' = 0 \overset{h\in L^2(\mathbb{R}^D)}{\iff} h = 0.
\end{aligned}$$

This concludes the discussion in the continuous case. In the discrete case however, the study is more involved as it depends on the measurement grid and the PSF h. It should be handled on a case by case basis. □

D. Proof of Theorem 4.1

D.1. An intermediary result

For $r \geq 0$, let $\mathcal{B}_r \overset{\text{def}}{=} \{x \in \mathbb{R}^D, \|x - \bar{x}\|_2 \leq r\}$.

Theorem D.1 (A general result)**.** *Let $(r_i)_{0 \leq i \leq I}$ denote an increasing sequence with $r_0 = r$ and $r_I = R$. Define $\Omega_i = \mathcal{B}_{r_i} \setminus \mathcal{B}_{r_{i-1}}$ and*

$$\bar{E}_i \overset{\text{def}}{=} \begin{cases} c\sigma\sqrt{D}\Lambda \frac{r_i}{R_L} & \text{if } r_i \leq \frac{R_L}{2} \\ c\sigma\sqrt{D}\Lambda \sqrt{\ln\left(\frac{3r_i}{R_L}\right)} & \text{otherwise} \end{cases} \quad \text{and} \quad \bar{\Lambda}_i \overset{\text{def}}{=} \Lambda \min\left(1, \frac{r_i}{R_L}\right). \tag{D.1}$$

where c is a universal constant. Set $\rho > 0$. From (9), (10) *and under Assumption 4.1. We have $\hat{x} \in \mathcal{B}_r$ with probability larger than $1 - \exp(-\rho^2/2)$ under the I conditions $1 \leq i \leq I$:*

$$\bar{E}_i + \sqrt{\rho^2 + 2\ln(I)}\sigma\bar{\Lambda}_i \leq \inf_{x \in \Omega_i} \ell_0(x). \tag{$\mathcal{C}_i$}$$

Proof. The proof of this result is quite long and technical.

D.1.1. The general strategy Let $\mathcal{B}_r = \{x \in \mathbb{R}^D, \|x - \bar{x}\|_2 \leq r\}$. We can decompose ℓ_ε as $\ell_\varepsilon = \ell_0 - \Delta_\varepsilon + \frac{1}{2}\|\varepsilon\|_2^2$, where $\Delta_\varepsilon(x) \overset{\text{def}}{=} \langle h(Z - x) - h(Z - \bar{x}), \varepsilon\rangle$. Without loss of generality, we consider here that $\Omega = \mathcal{B}_R$. We partition the domain Ω, as

$$\Omega = \bigsqcup_{0 \leq i \leq I} \Omega_i \text{ with } \Omega_i = \mathcal{B}_{r_i} \setminus \mathcal{B}_{r_{i-1}} \text{ for } i \geq 1, \tag{D.2}$$

with $\Omega_0 = \mathcal{B}_r$, $(r_i)_{0 \leq i \leq I}$ an increasing sequence with $r_0 = r$ and $r_I = R$. Now, remark that

$$\begin{aligned}
[\|\hat{x} - \bar{x}\|_2 \leq r] &\Leftrightarrow \left[\inf_{x \in \mathcal{B}_r} \ell_\varepsilon(x) < \inf_{x \in \mathcal{B}_r^c} \ell_\varepsilon(x)\right] \\
&\Leftrightarrow \left[\inf_{x \in \mathcal{B}_r} \ell_0(x) - \Delta_\varepsilon(x) < \inf_{x \in \mathcal{B}_r^c} \ell_0(x) - \Delta_\varepsilon(x)\right] \\
&\Leftrightarrow \left[\inf_{x \in \mathcal{B}_r} \ell_0(x) - \Delta_\varepsilon(x) < \inf_{x \in \Omega_i} \ell_0(x) - \Delta_\varepsilon(x),\ \forall 1 \leq i \leq I\right] \\
&\Leftarrow \left[\underbrace{\ell_0(\bar{x}) - \Delta_\varepsilon(\bar{x})}_{0} < \inf_{x \in \Omega_i} \ell_0(x) - \sup_{x \in \Omega_i} \Delta_\varepsilon(x),\ \forall 1 \leq i \leq I\right] \qquad \text{(D.3)} \\
&\Leftrightarrow \left[\inf_{x \in \Omega_i} l_0(x) > \sup_{x \in \Omega_i} \Delta_\varepsilon(x),\ \forall 1 \leq i \leq I\right].
\end{aligned}$$

Here ^{c} denote the relative complement with respect to $\mathcal{B}_R$. The reason we partition the domain in concentric annuli is for the event at line D.3 and the one before to be close. The interest of D.3 is that we only need to control the supremum of the *centered* process Δ_ε instead on the non centered process $\ell_0 - \Delta_\varepsilon$. The sequence of radii $(r_i)_{0 \leq i \leq I}$ will be optimized at the end of the proof.

D.1.2. Bounding the suprema The main technical difficulty is to find probabilistic bounds on the supremum $\sup_{x\in\Omega_i}\Delta_\varepsilon(x)$. To this end, we will use a combination of Gaussian concentration results and Dudley's type inequality. We refer to the three excellent monographs [27, 5, 31] for an in depth treatment of this topic. In our specific case, we obtain the following result.

Lemma D.1 (Expectation and tail bounds for the supremum)**.** *Let*

$$E_i \stackrel{\text{def}}{=} \mathbb{E}\left[\sup_{x\in\Omega_i}\Delta_\varepsilon(x)\right]. \tag{D.4}$$

For $t \geq 0$, we have

$$\mathbb{P}\left(\sup_{x\in\Omega_i}\Delta_\varepsilon(x)\geq t\right)\leq \exp\left(-\frac{(t-E_i)^2}{2\sigma^2L^2\min(R_L^2,r_i^2)}\right). \tag{D.5}$$

with

$$E_i \leq c\cdot\sigma\cdot L\cdot\sqrt{D}\cdot\begin{cases} r_i & \text{if } r_i\leq\frac{R_L}{2},\\ R_L\cdot\sqrt{\ln\left(\frac{3r_i}{R_L}\right)} & \text{otherwise.}\end{cases} \tag{D.6}$$

Proof. First notice that the random process Δ_ε is Gaussian since $\varepsilon\sim\mathcal{N}(0,\sigma^2\mathrm{Id})$. It can be written alternatively as $\sigma\Delta_\epsilon$ with $\epsilon\sim\mathcal{N}(0,\mathrm{Id})$. Let $\mathcal{D}\subset\mathbb{R}^D$ denote a domain and define the mapping $f_\mathcal{D}:\epsilon\mapsto\sup_{x\in\mathcal{D}}\sigma\Delta_\epsilon(x)$. Let us define

$$D_h(x_1,x_2)\stackrel{\text{def}}{=} h(Z-x_1)-h(Z-x_2) \quad\text{and}\quad \Lambda_\mathcal{D}\stackrel{\text{def}}{=}\sup_{x\in\mathcal{D}}\|D_h(x,\bar{x})\|_2. \tag{D.7}$$

We have for $\delta\in\mathbb{R}^M$

$$f_\mathcal{D}(\epsilon+\delta)=\sigma\sup_{x\in\mathcal{D}}\langle D_h(x,\bar{x}),\epsilon+\delta\rangle\leq\sigma\sup_{x\in\mathcal{D}}\langle D_h(x,\bar{x}),\epsilon\rangle+\sigma\Lambda_\mathcal{D}\|\delta\|_2,$$
$$f_\mathcal{D}(\epsilon+\delta)=\sigma\sup_{x\in\mathcal{D}}\langle D_h(x,\bar{x}),\epsilon+\delta\rangle\geq\sigma\sup_{x\in\mathcal{D}}\langle D_h(x,\bar{x}),\epsilon\rangle-\sigma\Lambda_\mathcal{D}\|\delta\|_2.$$

Hence $|f_\mathcal{D}(\epsilon+\delta)-f_\mathcal{D}(\epsilon)|\leq\sigma\Lambda_\mathcal{D}\|\delta\|_2$ and we can conclude that $f_\mathcal{D}$ is Lipschitz continuous with constant $\sigma\Lambda_\mathcal{D}$. Using [5, Theorem 5.6], we therefore get that, for any $u>0$ and $\mathcal{D}\subset\mathbb{R}^D$,

$$\mathbb{P}\left(\sup_{x\in\mathcal{D}}\Delta_\varepsilon(x)\geq\mathbb{E}\left[\sup_{x\in\mathcal{D}}\Delta_\varepsilon(x)\right]+u\right)\leq e^{-\frac{u^2}{2\sigma^2\Lambda_\mathcal{D}^2}}. \tag{D.8}$$

From (9) and (10), we have

$$\|D_h(x,\bar{x})\|_2\leq\min(\Lambda,L\|x-\bar{x}\|_2).$$

Hence for $\mathcal{D}=\Omega_i$, we obtain

$$\Lambda_{\Omega_i}\leq\min(\Lambda,Lr_i).$$

The remaining technical difficulty is to find an upper-bound for $\mathbb{E}\left[\sup_{x\in\Omega_i}\Delta_\varepsilon(x)\right]$. which is a hard problem in general. In this work, we will use Dudley's inequality together with (9) and (10). Let us introduce the pseudo-metric $d : \mathbb{R}^D \times \mathbb{R}^D \to \mathbb{R}$ defined by

$$d(x,x') \stackrel{\text{def}}{=} \sqrt{\mathbb{E}\left[(\Delta_\varepsilon(x) - \Delta_\varepsilon(x'))^2\right]} = \sqrt{\mathbb{E}\left[(\langle h(Z-x) - h(Z-x'), \varepsilon\rangle)^2\right]} \tag{D.9}$$

$$= \sqrt{\sigma^2 \left\| h(Z-x) - h(Z-x') \right\|_2^2} = \sigma \left\| h(Z-x) - h(Z-x') \right\|_2 . \tag{D.10}$$

It will be used in conjunction with the following tool.

Definition D.1 (Covering number)**.** *The covering number $\mathcal{N}(\delta, \mathcal{S}, d)$ is defined as the minimal number of balls of radius δ in the pseudo-metric d needed to cover $\mathcal{S}$.*

We are ready to present Dudley's inequality (see e.g. [5, Cor. 13.2]).

Theorem D.2 (Dudley's inequality)**.** *The following inequality holds:*

$$\mathbb{E}\left[\sup_{x\in\mathcal{S}} \Delta_\varepsilon(x)\right] \le c\int_0^\delta \sqrt{\ln(\mathcal{N}(u,\mathcal{S},d))}\, du, \tag{D.11}$$

where c is a universal constant and $\delta > 0$ can be taken as the smallest number such that $\mathcal{N}(u, \mathcal{S}, d) \ge 1$.

Lemma D.2. *We have for some universal constant c,*

$$E_i \le c \cdot \sigma \cdot L \cdot \sqrt{D} \cdot \begin{cases} r_i & \text{if } r_i \le \frac{R_L}{2}, \\ R_L \cdot \sqrt{\ln\left(\frac{3r_i}{R_L}\right)} & \text{otherwise.} \end{cases} \tag{D.12}$$

Proof. In what follows, c is a universal constant that may change from one line to the other. From (9) and (10), we have

$$d(x,x') \le \sigma \cdot \min\left(L\|x - x'\|_2, \Lambda\right) \tag{D.13}$$

and we can upper bound E_i as

$$E_i = \mathbb{E}\left[\sup_{x\in\Omega_i} \Delta_\varepsilon(x)\right] \le \mathbb{E}\left[\sup_{x\in\mathcal{B}_{r_i}} \Delta_\varepsilon(x)\right].$$

Using (D.13) and [31, Cor. 4.2.13], we obtain

$$\mathcal{N}(u, \mathcal{B}_{r_i}, d) \le \begin{cases} 1 & \text{if } u \ge \sigma \cdot \min(2Lr_i, \Lambda), \\ \left(\frac{3r_i\sigma L}{u}\right)^D & \text{otherwise.} \end{cases} \tag{D.14}$$

Hence, in the case $2Lr_i \le \Lambda$, we obtain using Theorem D.2:

$$\begin{aligned} E_i &\le c\int_0^{2r_i\sigma L} \sqrt{\ln(\mathcal{N}(u,\mathcal{B}_{r_i},d))}\, du \le c\int_0^{3r_i\sigma L} \sqrt{\ln(\mathcal{N}(u,\mathcal{B}_{r_i},d))}\, du \\ &\le c\sqrt{D}\int_0^{3r_i\sigma L} \sqrt{\ln\left(\frac{3r_i\sigma L}{u}\right)}\, du = c \cdot r_i \cdot \sigma \cdot L \cdot \sqrt{D}. \end{aligned}$$

In the case $2Lr_i > \Lambda$, we get

$$
\begin{aligned}
E_i &\leq c \int_0^{\sigma\Lambda} \sqrt{\ln(\mathcal{N}(u, \mathcal{B}_{r_i}, d))}\, du \\
&\leq c\sqrt{D} \int_0^{\sigma\Lambda} \sqrt{\ln\left(\frac{3r_i\sigma L}{u}\right)}\, du \\
&= c\sigma\Lambda\sqrt{D} \int_0^{1} \sqrt{\ln\left(\frac{3r_i L}{\Lambda v}\right)}\, dv \\
&= c\sigma\Lambda\sqrt{D} \left(\sqrt{\ln\left(\frac{3r_i L}{\Lambda}\right)} + \frac{\sqrt{\pi}}{2}\frac{3r_i L}{\Lambda}\mathrm{erfc}\left(\sqrt{\frac{3r_i L}{\Lambda}}\right)\right) \\
&\leq c\sigma\Lambda\sqrt{D} \left(\sqrt{\ln\left(\frac{3r_i L}{\Lambda}\right)} + \frac{1}{2}\sqrt{\frac{3r_i L}{\Lambda}}\exp\left(-\frac{3r_i L}{\Lambda}\right)\right)
\end{aligned}
$$

where we use the inequality $\mathrm{erfc}(z) < \frac{\exp(-z^2)}{\sqrt{\pi}z}$ for $z > 0$ to get the last inequality. Finally, since $\frac{1}{2}\sqrt{z}e^{-z} \leq \sqrt{\ln(z)}$ for $z \geq 3/2$, the condition $2r_i > R_L$ implies the simplification:

$$
\mathbb{E}\left[\sup_{x\in\mathcal{B}_{r_i}} \Delta_\varepsilon(x)\right] \leq c\cdot\sigma\cdot\Lambda\cdot\sqrt{D}\cdot\sqrt{\ln\left(\frac{3r_i}{R_L}\right)}. \tag{D.15}
$$

□

□

D.1.3. Concluding the proof of Theorem D.1 Let $\theta_i \overset{\text{def}}{=} \inf_{x\in\Omega_i} \ell_0(x)$. Now, assume that

$$
\theta_i \geq E_i,\ \forall 1 \leq i \leq I. \tag{D.16}
$$

Then, the probability of the event $\hat{x} \in \mathcal{B}_r$ can be bounded above as follows:

$$
\begin{aligned}
\mathbb{P}\left(\hat{x} \in \mathcal{B}_r\right) &\geq \mathbb{P}\left(\bigcap_{1\leq i\leq I}\left[\sup_{x\in\Omega_i} \Delta_\varepsilon(x) - \theta_i < 0\right]\right) \\
&\geq 1 - \mathbb{P}\left(\bigcup_{1\leq i\leq I}\left[\sup_{x\in\Omega_i} \Delta_\varepsilon(x) - \theta_i \geq 0\right]\right) \\
&\geq 1 - \sum_{i=1}^{I}\mathbb{P}\left(\sup_{x\in\Omega_i} \Delta_\varepsilon(x) - \theta_i \geq 0\right) \\
&\geq 1 - \sum_{i=1}^{I}\exp\left(-\frac{(\theta_i - E_i)^2}{2\sigma^2 L^2 \min(R_L, r_i)^2}\right).
\end{aligned}
$$

Set $\rho \geq 0$ and $\rho' = \sqrt{\rho^2 + 2\ln(I)}$. Under the conditions

$$
E_i + \rho'\sigma L \min(R_L, r_i) \leq \theta_i, \tag{D.17}
$$

we get that the probability of success is higher than $1 - \exp(-\rho^2/2)$. □

D.2. Concluding the proof of Theorem 4.1

Proof. From Theorem D.1, the inequality (15) is valid if

$$\bar{E}_i + \sqrt{\rho^2 + 2\ln(I)}\sigma\bar{\Lambda}_i \leq \inf_{x\in\Omega_i} \ell_0(x) \quad \forall 1 \leq i \leq I. \tag{D.18}$$

Assumption 4.1 allows us to get

$$\inf_{x\in\Omega_i} \ell_0(x) \geq \min\left(\frac{\mu^2 r_{i-1}^2}{2}, \frac{\mu^2 R_\mu^2}{2}\right).$$

Let us assume that $R > \frac{R_L}{2}$. Without loss of generality, we can also assume that $R_\mu \leq \frac{R_L}{2}$, since the inequalities in (11) are still valid when replacing $\mathcal{B}_{R_\mu}$ by $\mathcal{B}_{R'}$ with $R' \leq R_\mu$. In this setting, the success condition $(\mathcal{C}_I)$ reads

$$c\sigma\sqrt{D}\Lambda\sqrt{\ln\left(\frac{3R}{R_L}\right)} + \sigma\sqrt{\rho^2 + 2\ln(I)}\Lambda \leq \frac{\mu^2 R_\mu^2}{2}. \tag{D.19}$$

Now, let us set $r_{I-1} = R_\mu$. For $1 \leq i \leq I-1$, the conditions $(\mathcal{C}_i)$ read:

$$\sigma r_i L\left(c\sqrt{D} + \sqrt{\rho^2 + 2\ln(I)}\right) \leq \frac{\mu^2 r_{i-1}^2}{2}. \tag{D.20}$$

This can be rewritten as:

$$\frac{2}{\mu^2}\sigma L\left(c\sqrt{D} + \sqrt{\rho^2 + 2\ln(I)}\right) \leq \frac{r_{i-1}^2}{r_i}. \tag{D.21}$$

Now, by setting $r_i = 2^{2^i-1}r$, we get $\frac{r_{i-1}^2}{r_i} = \frac{r}{2}$ for all i. In addition $2^{2^i-1}r \geq R_\mu$ for $i \geq \lceil \log_2(\log_2(2R_\mu/r))\rceil$. Hence, we can set $I = \lceil \log_2(1 + \log_2(R_\mu/r))\rceil$. This value is larger than 2 for $r < \frac{R_\mu}{2}$. Hence, under this condition, we get $\hat{x} \in \mathcal{B}_r$ given that

$$\begin{aligned}
&\frac{4}{\mu^2}\sigma L\left(c\sqrt{D} + \sqrt{\rho^2 + 2\ln(I)}\right) \leq r \\
&\frac{2}{\mu^2}\sigma\Lambda\left(c\sqrt{D}\sqrt{\ln\left(\frac{3R}{R_L}\right)} + \sqrt{\rho^2 + 2\ln(I)}\right) \leq R_\mu^2.
\end{aligned}$$

In the case $r \geq \frac{R_\mu}{2}$, we can set $I = 1$, $\Omega_1 = \mathcal{B}_{R_\mu}$ and only the second condition is sufficient for success:

$$\frac{2}{\mu^2}\sigma\Lambda\left(c\sqrt{D}\sqrt{\ln\left(\frac{3R}{R_L}\right)} + \rho\right) \leq R_\mu^2.$$

The case $R \leq \frac{R_L}{2}$ can be treated as previously, but only Cond_1 matters, since the other one is automatically verified. □

E. Proof of Theorem 4.2

Proof. First notice that

$$\inf_{x\in\mathcal{B}_r} \ell_0(x) - \Delta_\varepsilon(x) \geq \inf_{x\in\mathcal{B}_r} \ell_0(x) - \sup_{x\in\mathcal{B}_r} \Delta_\varepsilon(x) \geq -\sup_{x\in\mathcal{B}_r} \Delta_\varepsilon(x).$$

Using Lemma (D.1), we obtain

$$E_r \stackrel{\text{def}}{=} \mathbb{E}\left[\sup_{x\in\mathcal{B}_r} \Delta_\varepsilon(x)\right] \leq cr\sigma L\sqrt{D},$$

where c is a universal constant and for $\rho_1 > 0$

$$\mathbb{P}\left(\sup_{x\in\mathcal{B}_r} \Delta_\varepsilon(x) \geq E_r + \rho_1\sigma Lr\right) \leq \exp(-\rho_1^2/2).$$

Define the event

$$\mathcal{E}_1 \stackrel{\text{def}}{=} \left[\inf_{x\in\mathcal{B}_r} \ell_0(x) - \Delta_\varepsilon(x) \geq -\sigma Lr\left(c\sqrt{D}+\rho_1\right)\right].$$

The previous inequalities imply that

$$\mathbb{P}\left(\mathcal{E}_1\right) \geq 1 - \exp(-\rho_1^2/2).$$

Proof of the part related to (Cond$_1'$) Now, set $t \in (r, R_\mu]$ and take an arbitrary point x with $\|x-\bar{x}\|_2 = t$. Let's consider the following inequalities

$$\inf_{x\in\mathcal{S}_t} \ell_\varepsilon(x) = \inf_{x\in\mathcal{S}_t} \ell_0(x) - \Delta_\varepsilon(x) \leq \ell_0(\tilde{x}) - \Delta_\varepsilon(\tilde{x}) \leq \sup_{x\in\mathcal{S}_t} \ell_0(x) + \inf_{x\in\mathcal{S}_t} -\Delta_\varepsilon(x) \quad \text{(E.1)}$$

where $\tilde{x} \stackrel{\text{def}}{=} \operatorname{argmin}_{x\in\mathcal{S}_t} -\Delta(x)$. It only remains to impose the two following inequalities

$$\sup_{x\in\mathcal{S}_t} \ell_0(x) + \inf_{x\in\mathcal{S}_t} -\Delta_\varepsilon(x) < -\sigma Lr\left(c\sqrt{D}+\rho_1\right) \leq \inf_{x\in\mathcal{B}_r} \ell_\varepsilon(x) \quad \text{(E.2)}$$

in order to get the result $\hat{x} \notin \mathcal{B}_r$. Indeed, this inequality provides the relevant relation $\inf_{\mathcal{S}_t} \ell_\varepsilon < \inf_{\mathcal{B}_r} \ell_\varepsilon$. The second inequality is controlled by the event $\mathcal{E}_1$, whereas we will control the first inequality with the event

$$\mathcal{E}_2 \stackrel{\text{def}}{=} \left[\inf_{x\in\mathcal{S}_t} -\Delta_\varepsilon(x) < -\sigma Lr\left(c\sqrt{D}+\rho_1\right) - \sup_{x\in\mathcal{S}_t} \ell_0(x)\right]. \quad \text{(E.3)}$$

Since the law of the random variable Δ_ε is even, the event $\mathcal{E}_2'$ defined as

$$\mathcal{E}_2' \stackrel{\text{def}}{=} \left[\sup_{x\in\mathcal{S}_t} \Delta_\varepsilon(x) > \sigma Lr\left(c\sqrt{D}+\rho_1\right) + \sup_{x\in\mathcal{S}_t} \ell_0(x)\right] \quad \text{(E.4)}$$

has the same probability than $\mathcal{E}_2$.

Applying the same reasoning as the one for Lemma D.1 to the random variable $\sup_{x\in\mathcal{S}_t} \Delta_\varepsilon(x)$, implies that

$$\mathbb{P}\left(\sup_{x\in\mathcal{S}_t} \Delta_\varepsilon(x) > \mathbb{E}\left[\sup_{x\in\mathcal{S}_t} \Delta_\varepsilon(x)\right] - u\right) \geq 1 - \exp\left(-\frac{u^2}{2\sigma^2(Lt)^2}\right) \quad \text{(E.5)}$$

for any $u > 0$. Setting $u = \mathbb{E}\left[\sup_{x\in\mathcal{S}_t} \Delta_\varepsilon(x)\right] - \sup_{x\in\mathcal{S}_t} \ell_0(x) - \sigma Lr\left(c\sqrt{D} + \rho_1\right)$, this allows us to get

$$\mathbb{P}\left(\mathcal{E}_2\right) = \mathbb{P}\left(\mathcal{E}_2'\right) \geq 1 - e^{-\rho_2^2/2} \tag{E.6}$$

under the condition

$$\rho_2 \leq \frac{u}{\sigma Lt}$$

which rewrites as

$$\rho_2 \sigma Lt \leq \mathbb{E}\left[\sup_{x\in\mathcal{S}_t} \Delta_\varepsilon(x)\right] - \sup_{x\in\mathcal{S}_t} \ell_0(x) - \sigma Lr\left(c\sqrt{D} + \rho_1\right). \tag{E.7}$$

The next step is to find a lower bound for the expectation $\mathbb{E}\left[\sup_{x\in\mathcal{S}_t} \Delta_\varepsilon(x)\right]$ given in (E.22). This is accomplished thanks to the following lemmas:

Lemma E.1. *For $D \geq 2$, let $\mathcal{B}_1, \mathcal{B}_2$ be two intersecting balls of $\mathbb{R}^D$ of respective radius $0 \leq r_1 \leq r_2$. Denoting $\mathcal{S}_1, \mathcal{S}_2$ their associated spheres, we define two sphere caps as*

$$\mathcal{C}_1 = \mathcal{S}_1 \cap \mathcal{B}_2 \quad \textit{and} \quad \mathcal{C}_2 = \mathcal{S}_2 \cap \mathcal{B}_1. \tag{E.8}$$

Then the sphere cap $\mathcal{C}_1$ has a greater surface area than $\mathcal{C}_2$, i.e.

$$\lambda_{D-1}(\mathcal{C}_1) \geq \lambda_{D-1}(\mathcal{C}_2). \tag{E.9}$$

Proof. Denote ϕ_1 (resp. ϕ_2) the polar angle of the sphere cap $\mathcal{C}_1$ (resp. $\mathcal{C}_2$). A simple formula to compute the surface area of a sphere cap can be found in [24] for example. Denoting $A_D(r)$ (resp. $A_D^{\mathrm{cap}}(r,\phi)$) the surface area of a sphere in $\mathbb{R}^D$ of radius r (resp. of a sphere cap in $\mathbb{R}^D$ of radius r and polar angle ϕ), one has the following formulae

$$A_D(r) = \frac{2\pi^{\frac{D-1}{2}}}{\Gamma(\frac{D-1}{2})} r^{D-1}, \tag{E.10}$$

$$A_D^{\mathrm{cap}}(r,\phi) = \int_0^\phi A_{D-1}(r\sin\theta) r \mathrm{d}\theta = \int_0^{r\sin\phi} \frac{1}{\sqrt{1-\frac{s^2}{r^2}}} A_{D-1}(s)\mathrm{d}s. \tag{E.11}$$

The last equality is obtained thanks to the change of variables $s = r\sin\theta$. Because the two balls are intersecting each other, the radius a of the base is common to the two sphere caps. Indeed, one has

$$r_1 \sin\phi_1 = a = r_2 \sin\phi_2. \tag{E.12}$$

Finally, since $r_1 \leq r_2$, for any $s < a$, the inequality $\frac{1}{\sqrt{1-\frac{s^2}{r_2^2}}} \leq \frac{1}{\sqrt{1-\frac{s^2}{r_1^2}}}$ allows to find the relation

$$A_D^{\mathrm{cap}}(r_2,\phi_2) \leq A_D^{\mathrm{cap}}(r_1,\phi_1) \tag{E.13}$$

which concludes the proof. □

Lemma E.2. *For any $0 < u \leq t\mu\sigma$, noting d the pseudo-metric defined in* (D.9)*, one has*

$$\mathcal{N}(u, \mathcal{S}_t, d) \geq \left(\frac{t\mu\sigma}{u}\right)^{D-1}. \tag{E.14}$$

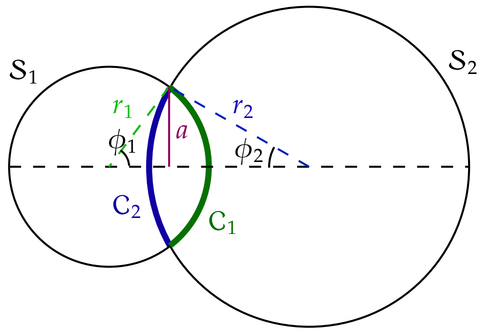


Figure E1: Example of two sphere caps $\mathcal{C}_1$ and $\mathcal{C}_2$. The representation gives their intersection on a plane passing through the centers of the spheres.

Proof of Lemma E.2. From Assumption (16), the problem is reduced to the Euclidean metric as

$$\mathcal{N}(u, \mathcal{S}_t, d) \geq \mathcal{N}\left(\frac{u}{\mu\sigma}, \mathcal{S}_t, d_{\text{eucli}}\right). \tag{E.15}$$

From definition of the covering number, let's take balls $\mathcal{B}_1, \ldots, \mathcal{B}_N$ of radius $\tilde{u} \overset{\text{def}}{=} \frac{u}{\mu\sigma} < t$ covering the sphere $\mathcal{S}_t$ with $N = \mathcal{N}(\tilde{u}, \mathcal{S}_t, d_{\text{eucli}})$. The result will come by considering the surface area of the caps $\mathcal{C}_n \overset{\text{def}}{=} \mathcal{S}_t \cap \mathcal{B}_n$. Indeed, since the balls $\mathcal{B}_1, \ldots, \mathcal{B}_N$ cover $\mathcal{S}_t$ we have $\mathcal{S}_t \subseteq \bigcup_{n=1}^N \mathcal{C}_n$. Then, taking the Lebesgues measure leads to

$$\lambda_{D-1}(\mathcal{S}_t) \leq \sum_{n=1}^N \lambda_{D-1}(\mathcal{C}_n). \tag{E.16}$$

We set ϕ_n the polar angle of the sphere cap $\mathcal{C}_n$. With the notations introduced in the proof of Lemma E.1, we can reformulate the previous inequality as

$$A_D(t) \leq \sum_{n=1}^N A_D^{\text{cap}}(t, \phi_n). \tag{E.17}$$

Moreover, we get the following bounds from Lemma E.1

$$A_D^{\text{cap}}(t, \phi_n) \leq A_D^{\text{cap}}(\tilde{u}, \psi_n) \leq A_D(\tilde{u}) \tag{E.18}$$

where ψ_n is the polar angle of the counter-part sphere cap $\mathcal{B}_t \cap \mathcal{S}_n$. This implies that

$$A_D(t) \leq N \cdot A_D(\tilde{u}). \tag{E.19}$$

This equation combined with the area formula (E.10) concludes since

$$\mathcal{N}(\tilde{u}, \mathcal{S}_t, d_{\text{eucli}}) \geq \frac{A_D(t)}{A_D(\tilde{u})} = \left(\frac{t}{\tilde{u}}\right)^{D-1}. \tag{E.20}$$

□

Going back to the necessary condition (E.7), using Sudakov's inequality [31, Thm 7.4.1] or [5, Thm 13.4], one gets thanks to lemma E.2

$$\mathbb{E}\left[\sup_{x\in\mathcal{S}_t}\Delta_\varepsilon(x)\right] \geq \frac{\sqrt{D-1}}{2}\sup_{v\in[0,t\mu\sigma]} v\sqrt{\ln\left(\frac{t\mu\sigma}{v}\right)}. \tag{E.21}$$

Taking $v = \frac{t}{2}\mu\sigma$, we get

$$\mathbb{E}\left[\sup_{x\in\mathcal{S}_t}\Delta_\varepsilon(x)\right] \geq \frac{\sqrt{\ln 2}}{4}t\mu\sigma\sqrt{D-1}. \tag{E.22}$$

Injecting this inequality in (E.7), we have

$$\rho_2\sigma L t \leq \frac{\sqrt{\ln 2}}{4}t\mu\sigma\sqrt{D-1} - \frac{L^2}{2}t^2 - \sigma L r\left(c\sqrt{D}+\rho_1\right). \tag{E.23}$$

Denoting $t = \alpha r$ with $\alpha \geq 1$, this simplifies as

$$\sigma\left(\frac{\sqrt{\ln 2}}{4}\alpha\frac{\mu}{L}\sqrt{D-1} - \rho_2\alpha - \left(c\sqrt{D}+\rho_1\right)\right) \geq \frac{L\alpha^2}{2}r. \tag{E.24}$$

We fix the value of α (and thus also the value of t) such that

$$\rho_2\alpha + \left(c\sqrt{D}+\rho_1\right) = \frac{\sqrt{\ln 2}}{8}\alpha\frac{\mu}{L}\sqrt{D-1}. \tag{E.25}$$

Moreover we check that the condition $\alpha \geq 1$ is well verified since it is equivalent to the statement (for small ρ_2)

$$c\sqrt{D}+\rho_1+\rho_2 \geq \frac{\sqrt{\ln 2}}{8}\frac{\mu}{L}\sqrt{D-1} \tag{E.26}$$

which is true since $c > 1$ and $\mu \leq L$. Thanks to this setting, the necessary condition (E.24) becomes

$$\sigma\frac{\sqrt{\ln 2}}{8}\mu\sqrt{D-1} \geq \frac{L^2\alpha}{2}r. \tag{E.27}$$

By adding the condition $t = \alpha r \leq R_\mu$, we finally get

$$r \leq \frac{1}{\alpha}\min\left(R_\mu,\ \sigma\frac{\sqrt{\ln 2}}{4}\frac{\mu}{L^2}\sqrt{D-1}\right). \tag{E.28}$$

Plugging the definition of α allows to conclude that (Cond_1') is a necessary condition.

Proof of the part related to (Cond_2') Here, we assume that $R > R_\mu$ and that the growth condition is globalized:

$$\|h(Z-x)-h(Z-x')\|_2 \geq \mu\min\left(\|x-x'\|_2, R_\mu\right) \quad \forall x,x' \in \mathcal{B}_R.$$

Hence, we get using [31, Cor. 4.2.13]:

$$\mathcal{N}(u,\mathcal{B}_R,d) \geq \left(\frac{R\mu\sigma}{u}\right)^D \quad \text{for} \quad u \leq R_\mu\sigma\mu.$$

Using Sudakov's inequality [31, Thm 7.4.1] or [5, Thm 13.4], we obtain:

$$\mathbb{E}\left[\sup_{x\in\mathcal{B}_R}\Delta_\varepsilon(x)\right] \geq \frac{\sqrt{D}}{2}\sup_{u\in[0,R_\mu\sigma\mu]} u\cdot\sqrt{\ln\left(\frac{R\mu\sigma}{u}\right)}.$$

Taking $u = R_\mu\sigma\mu$, we get

$$\mathbb{E}\left[\sup_{x\in\mathcal{B}_R}\Delta_\varepsilon(x)\right] \geq \frac{\sqrt{D}}{2}R_\mu\sigma\mu\sqrt{\ln\left(\frac{R}{R_\mu}\right)}. \tag{E.29}$$

Applying the same reasoning as the one for Lemma D.1 to the random variable $-\sup_{x\in\mathcal{B}_R}\Delta_\varepsilon(x)$, implies that

$$\mathbb{P}\left(\mathbb{E}\left[\sup_{x\in\mathcal{B}_R}\Delta_\varepsilon(x)\right] - \sup_{x\in\mathcal{B}_R}\Delta_\varepsilon(x) \geq t\right) \leq \exp\left(-\frac{t^2}{2\sigma^2\Lambda^2}\right). \tag{E.30}$$

Define the event

$$\mathcal{E}_2 \stackrel{\text{def}}{=} \left[\sup_{x\in\mathcal{B}_R}\Delta_\varepsilon(x) \geq \frac{\sqrt{D}}{2}R_\mu\sigma\mu\sqrt{\ln\left(\frac{R}{R_\mu}\right)} - \rho_2\sigma\Lambda\right].$$

Combining the previous results, we obtain

$$\mathbb{P}\left(\mathcal{E}_2\right) \geq 1 - \exp(-\rho_2^2/2).$$

The event $\mathcal{E}_1 \cap \mathcal{E}_2$ happens with probability larger than $1 - \exp(-\rho_1^2/2) - \exp(-\rho_2^2/2)$. In addition, this event together with the condition

$$\frac{\Lambda^2}{2} - \frac{\sqrt{D}}{2}R_\mu\sigma\mu\sqrt{\ln\left(\frac{R}{R_\mu}\right)} + \rho_2\sigma\Lambda < -\sigma Lr\left(c\sqrt{D} + \rho_1\right)$$

implies that $\hat{x} \notin \mathcal{B}_r$. □

F. Proof of Theorem 4.3

The equality (31) is a consequence of standard results in sampling theory. Set $W > 0$ and let s_W denote the function which Fourier transform equals to $\mathcal{F}(s) = \frac{1}{(2\pi W)^{D/2}}\mathbb{1}_{[-\pi W,\pi W]^D}$. A direct calculation shows that it corresponds to a scaled and tensorized cardinal sine. It is well known that the family $\left(s_W(\cdot - n/W)\right)_{n\in\mathbb{Z}^D}$ is an orthonormal basis of $\mathcal{PW}(\pi W)$ (see e.g. [18, Thm. 3.5] in 1D, and use the fact that the tensor product of an orthogonal basis is still an orthogonal basis). In addition, we have the identity for any $f_W \in \mathcal{PW}(\pi W)$ (obtained by direct calculation again)

$$\langle f_W, s_W(\cdot - n/W)\rangle_{L^2(\mathbb{R}^D)} = f_W(n/W)\cdot(2\pi W)^{-D/2}. \tag{F.1}$$

Combining the two results yields

$$\begin{aligned} f_W &= \sum_{n\in\mathbb{Z}^D}\langle f_W, s_W(\cdot - n/W)\rangle_{L^2(\mathbb{R}^D)} s_W(\cdot - n/W) \\ &= \frac{1}{(2\pi W)^{D/2}}\sum_{n\in\mathbb{Z}^D} f_W(n/W)s_W(\cdot - n/W), \end{aligned}$$

which is nothing but Shannon-Whittaker theorem. Using again the fact that $(s_W(\cdot - n/W))_{n\in\mathbb{Z}^D}$ is an orthonormal basis yields:

$$\langle f_W, g_W \rangle_{L^2(\mathbb{R}^D)} = \frac{1}{(2\pi W)^D} \sum_{n\in\mathbb{Z}^D} f_W(n/W) g_W(n/W) \tag{F.2}$$

for $f_W, g_W \in \mathcal{PW}(\pi W)$.

We have $\mathcal{PW}(\pi) \subset \mathcal{PW}(\pi/\tau)$. Hence, we can apply the previous identity with $W = \frac{1}{\tau}$. This yields

$$\langle f, g \rangle_{L^2(\mathbb{R}^D)} = \frac{\tau^D}{(2\pi)^D} \sum_{n\in\mathbb{Z}^D} f(\tau n) g(\tau n). \tag{F.3}$$

The result is still valid by shifting f and g by x, resulting in the claimed result.

References

[1] Fredrik Athley. Threshold region performance of maximum likelihood direction of arrival estimators. *IEEE Transactions on Signal Processing*, 53(4):1359–1373, 2005.

[2] Edward W Barankin. Locally best unbiased estimates. *The Annals of Mathematical Statistics*, 20(4):477–501, 1949.

[3] Eric Betzig, George H Patterson, Rachid Sougrat, O Wolf Lindwasser, Scott Olenych, Juan S Bonifacino, Michael W Davidson, Jennifer Lippincott-Schwartz, and Harald F Hess. Imaging intracellular fluorescent proteins at nanometer resolution. *Science*, 313(5793):1642–1645, 2006.

[4] Norman Bobroff. Position measurement with a resolution and noise-limited instrument. *Review of Scientific Instruments*, 57(6):1152–1157, 1986.

[5] Stéphane Boucheron, Gábor Lugosi, and Pascal Massart. *Concentration inequalities: A nonasymptotic theory of independence*. Oxford university press, 2013.

[6] Claire Boyer, Yohann De Castro, and Joseph Salmon. Adapting to unknown noise level in sparse deconvolution. *Information and Inference: A Journal of the IMA*, 6(3):310–348, 2017.

[7] Kristian Bredies and Hanna Katriina Pikkarainen. Inverse problems in spaces of measures. *ESAIM: Control, Optimisation and Calculus of Variations*, 19(1):190–218, 2013.

[8] Cristina Butucea, Jean-François Delmas, Anne Dutfoy, and Clément Hardy. Off-the-grid learning of sparse mixtures from a continuous dictionary. *arXiv preprint arXiv:2207.00171*, 2022.

[9] Emmanuel J Candès and Carlos Fernandez-Granda. Towards a mathematical theory of super-resolution. *Communications on pure and applied Mathematics*, 67(6):906–956, 2014.

[10] Dan Chazan, Moshe Zakai, and Jacob Ziv. Improved lower bounds on signal parameter estimation. *IEEE transactions on Information Theory*, 21(1):90–93, 1975.

[11] A. Marie d'Avigneau, Sumeetpal S. Singh, and Raimund J. Ober. Limits of accuracy for parameter estimation and localization in single-molecule microscopy via sequential monte carlo methods. *SIAM Journal on Imaging Sciences*, 15(1):139–171, January 2022.

[12] Valentin Debarnot and Pierre Weiss. Blind inverse problems with isolated spikes. *Information and Inference: A Journal of the IMA*, 2022.

[13] Quentin Denoyelle, Vincent Duval, and Gabriel Peyré. Support recovery for sparse super-resolution of positive measures. *Journal of Fourier Analysis and Applications*, 23:1153–1194, 2017.

[14] Vincent Duval and Gabriel Peyré. Exact support recovery for sparse spikes deconvolution. *Foundations of Computational Mathematics*, 15(5):1315–1355, 2015.

[15] Corinne Fournier, Loïc Denis, and Thierry Fournel. On the single point resolution of on-axis digital holography. *Journal of the Optical Society of America A*, 27(8):1856–1862, 2010.

[16] Steven Kay. Fundamentals of statistical signal processing: estimation theory. *Technometrics*, 37(4):465–466, 1993.

[17] Fanglin Linda Liu, Grace Kuo, Nick Antipa, Kyrollos Yanny, and Laura Waller. Fourier diffuserscope: single-shot 3d fourier light field microscopy with a diffuser. *Optics Express*, 28(20):28969–28986, 2020.

[18] Stéphane Mallat. *A wavelet tour of signal processing.* Elsevier, 1999.

[19] William E Moerner and Lothar Kador. Optical detection and spectroscopy of single molecules in a solid. *Physical review letters*, 62(21):2535, 1989.

[20] Eran Naftali and Nicholas C Makris. Necessary conditions for a maximum likelihood estimate to become asymptotically unbiased and attain the cramer–rao lower bound. part i. general approach with an application to time-delay and doppler shift estimation. *The Journal of the Acoustical Society of America*, 110(4):1917–1930, 2001.

[21] Raimund J Ober, Sripad Ram, and E Sally Ward. Localization accuracy in single-molecule microscopy. *Biophysical journal*, 86(2):1185–1200, 2004.

[22] Calyampudi Radhakrishna Rao. *Linear statistical inference and its applications.* Wiley New York, 1973.

[23] Alexandre Renaux, Philippe Forster, Eric Chaumette, and Pascal Larzabal. On the high-snr conditional maximum-likelihood estimator full statistical characterization. *IEEE Transactions on Signal Processing*, 54(12):4840–4843, 2006.

[24] Li S. Concise formulas for the area and volume of a hyperspherical cap. *Asian Journal of Mathematics & Statistics*, 2011.

[25] Daniel Sage, Thanh-An Pham, Hazen Babcock, Tomas Lukes, Thomas Pengo, Jerry Chao, Ramraj Velmurugan, Alex Herbert, Anurag Agrawal, Silvia Colabrese, Ann Wheeler, Anna Archetti, Bernd Rieger, Raimund Ober, Guy M. Hagen, Jean-Baptiste Sibarita, Jonas Ries, Ricardo Henriques, Michael Unser, and Seamus Holden. Super-resolution fight club: assessment of 2d and 3d single-molecule localization microscopy software. *Nature Methods*, 16(5):387–395, 2019.

[26] Yoav Shechtman, Steffen J. Sahl, Adam S. Backer, and W. E. Moerner. Optimal point spread function design for 3d imaging. *Physical Review Letters*, 113(13):133902, 2014.

[27] Michel Talagrand. Upper and lower bounds for stochastic processes. decomposition theorems. *Ergebnisse der Mathematik und ihrer Grenzgebiete*, 2021.

[28] Gongguo Tang, Badri Narayan Bhaskar, and Benjamin Recht. Near minimax line spectral estimation. *IEEE Transactions on Information Theory*, 61(1):499–512, 2014.

[29] Russell E. Thompson, Daniel R. Larson, and Watt W. Webb. Precise nanometer localization analysis for individual fluorescent probes. *Biophysical Journal*, 82(5):2775–2783, 2002.

[30] Harry L Van Trees. *Detection, estimation, and modulation theory, part I: detection, estimation, and linear modulation theory.* John Wiley & Sons, 2004.

[31] Roman Vershynin. *High-dimensional probability: An introduction with applications in data science.* Cambridge university press, 2018.

[32] A. Weiss and E. Weinstein. Fundamental limitations in passive time delay estimation–part i: Narrow-band systems. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 31(2):472–486, 1983.

[33] Kyrollos Yanny, Nick Antipa, William Liberti, Sam Dehaeck, Kristina Monakhova, Fanglin Linda Liu, Konlin Shen, Ren Ng, and Laura Waller. Miniscope3d: optimized single-shot miniature 3d fluorescence microscopy. *Light: Science & Applications*, 9(1):171, 2020.