



HAL
open science

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

Nathanaël Munier, Emmanuel Soubies, Pierre Weiss

► **To cite this version:**

Nathanaël Munier, Emmanuel Soubies, Pierre Weiss. The MLE is a reliable source: sharp performance guarantees for localization problems. 2022. hal-03811028v1

HAL Id: hal-03811028

<https://hal.science/hal-03811028v1>

Preprint submitted on 11 Oct 2022 (v1), last revised 21 Nov 2023 (v2)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

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 for larger levels. An outcome is new optimization criteria for the design of point spread functions in single molecule microscopy.

Contents

1	Introduction	2
2	A brief tour of existing performance bounds	6
3	Preliminary facts	6
3.1	Existence of minimizers	6
3.2	Identifiability	7
4	Main results	7
4.1	Technical assumptions	8
4.2	The main results	9
4.3	An informal proof	10
4.4	Interpretations and simplifications	11
4.5	Notable consequences	13
4.6	Analysis for well sampled bandlimited kernels	14
5	Acknowledgments	16
A	Derivatives of ℓ_ε	16

B Proof of Theorem 2.1	16
B.1 Establishing the Cramér-Rao lower-bound	16
B.2 Necessary and sufficient condition for attainment	17
C Proof of Theorem 4.1	18
C.1 An intermediary result	18
C.1.1 The general strategy	19
C.1.2 Bounding the suprema	19
C.1.3 Concluding the proof of Theorem C.1	22
C.2 Concluding the proof of Theorem 4.1	23
D Proof of Theorem 4.2	24
E Proof of Theorem 4.3	26

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 $u : \mathbb{R}^D \rightarrow \mathbb{R}$ 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., $\varepsilon \sim \mathcal{N}(0, \sigma^2 \text{Id})$. In that case, the maximum likelihood estimator reads*

$$\hat{x} \in \underset{x \in \mathbb{R}^D}{\text{argmin}} \ell_\varepsilon(x) \quad \text{with} \quad \ell_\varepsilon(x) \stackrel{\text{def}}{=} \frac{1}{2} \sum_{m=1}^M (h(z_m - x) - h(z_m - \bar{x}) - \varepsilon_m)^2. \tag{4}$$

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, 12]. 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. Similar issues appear for the localization of sounds (microphone), stars (radio-telescope) or phones (wireless communications).

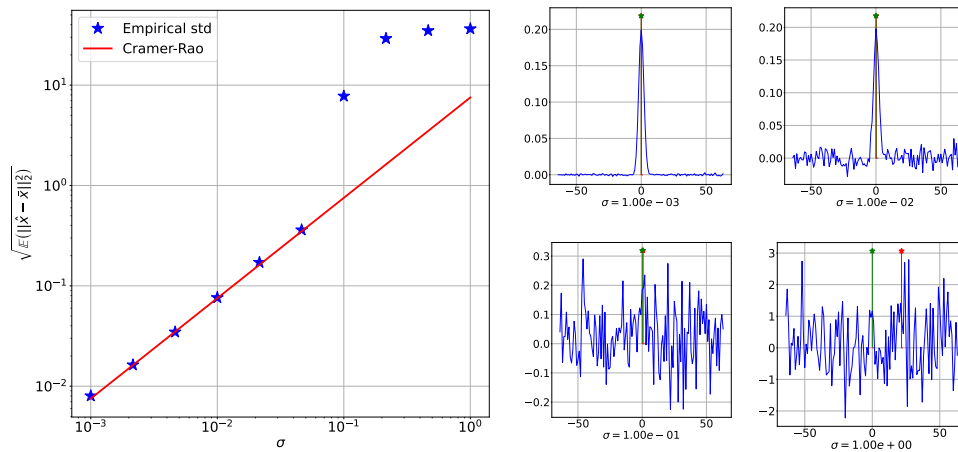


Figure 1: **A 1D localization experiment.** *Left:* the empirical standard deviation using the MLE (blue stars) coincides with the Cramér-Rao lower-bound for low noise regimes. In high noise regimes, we observe a significant discrepancy. *Right:* realizations of the vector y for different noise levels.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

Existing performance guarantees 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 [15]. 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, 20, 14, 7] as well as a baseline to estimate the quality of algorithms [17]. It can also serve as an optimization criterion to design new efficient point spread functions [9, 18].

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. 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

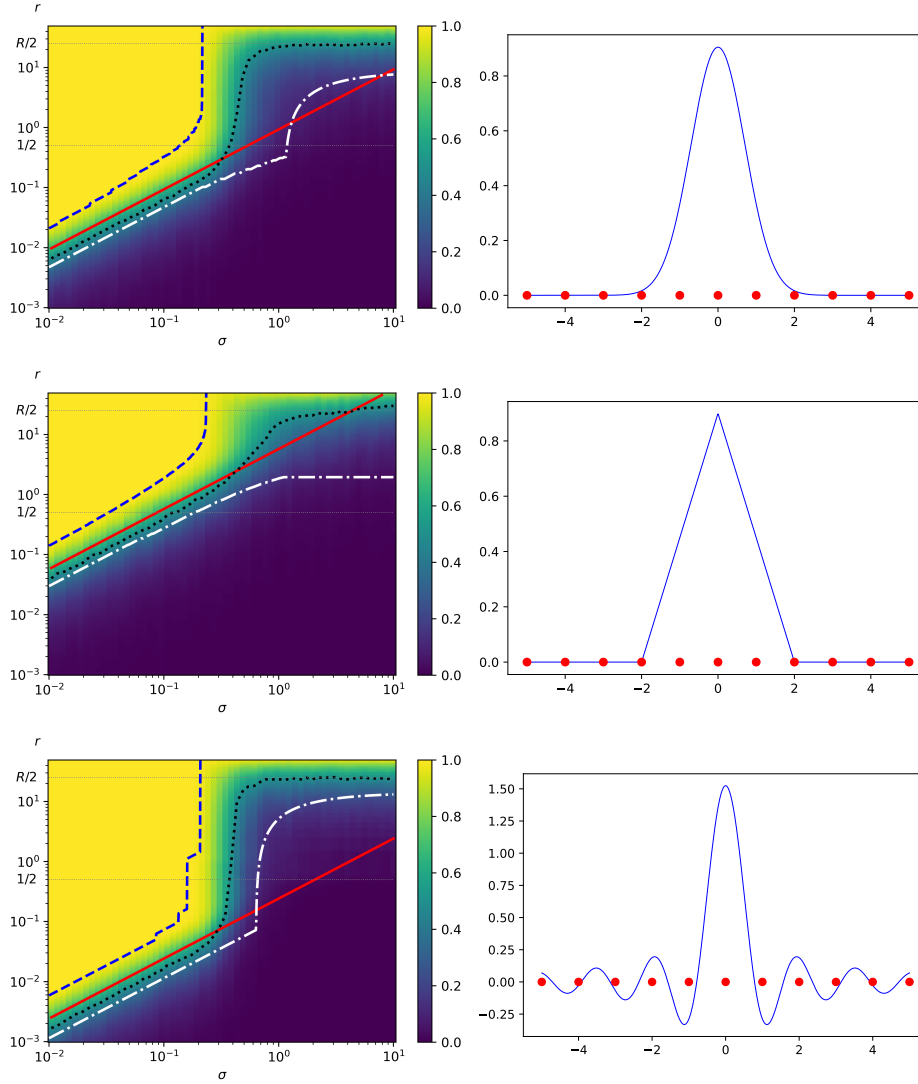


Figure 2: **Phase transitions for single source localization with the MLE.** *Left:* the heat map indicates the empirical probability of the event $\mathcal{E}_r = [\|\hat{x} - \bar{x}\|_2 \leq r]$. We observe a phase transition phenomenon: for a given r and a sufficiently low σ , the probability of success is overwhelming. Below a threshold, it decays rapidly to a low probability. The transition occurs close to the Cramér-Rao lower-bound (in red) 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.

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. 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 $\sigma, r \rightarrow 0$. 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. We let $\partial_d h$ denote the d -th partial derivative of the impulse response (PSF) h . Moreover, we let $h'(x) = (\partial_1 h(x), \dots, \partial_D h(x))$ 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, $\|A\|_{2 \rightarrow 2}$ denote its spectral norm and $\|A\|_F = \sqrt{\text{Tr}(A^*A)}$ denote its Frobenius norm. Given two quantities $a, b \in \mathbb{R}_+$, we will write $a \lesssim b$ if there exists a constant $c > 0$ not depending on a and b such that $a \leq cb$.

To keep notation concise, we introduce the set of sampling points $Z = (z_1, \dots, z_m)$ and use shorthand notation of the form

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

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

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

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

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

* The universal constants appearing in our bounds have been tuned by hand to reproduce the experimental curves. The same constants have been used for all kernels.

2. A brief tour of existing performance bounds

In our context, applying Theorem 3.1 in [10] yields the following result.

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

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

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. [14]), 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 [13, 16]. There, it was shown that under some technical assumptions, the MLE provides a performance asymptotically similar to the Cramér-Rao lower-bound for $\sigma \rightarrow 0$. 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, 6, 23]. For instance, we refer the reader to the excellent summary in [23, 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 [21, 1, 13].

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 [8], 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 $\sigma \rightarrow 0$, contrarily to the present work.

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. *There exist kernels h such that the probability of non existence of the MLE is non zero for any noise level $\sigma > 0$.*

Proof. Take a nonnegative 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 the cost function decreases as $|x| \rightarrow \infty$. \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, the estimator would then end up on the domain boundary. This simple example highlights a fundamental difficulty in the localization problem: the MLE can result in outliers that significantly increase the variance of the estimator. It will therefore be central to control the probability of \hat{x} being an outlier. Similar observations have already been formulated in [21, 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 $h_n(x) \stackrel{\text{def}}{=} \sin(2\pi nx)$ 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 $n \rightarrow \infty$. 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 $\mathbb{P}(\|\hat{x} - \bar{x}\|_2 \leq r)$ for any given radius $r > 0$. We define the estimator \hat{x} as

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

where $\Omega \subseteq \mathcal{B}_R \stackrel{\text{def}}{=} \{x \in \mathbb{R}^D, \|x - \bar{x}\|_2 \leq R\}$ 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,
- (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 assumptions

Our main result will depend on the following two natural assumptions.

Assumption 4.1 (Boundedness and regularity). *We assume that*

- *Boundedness:*

$$\forall x, x' \in \mathcal{B}_R, \quad \|h(Z - x) - h(Z - x')\|_2 \leq \Lambda < +\infty. \quad (9)$$

- *Lipschitz continuity:*

$$\forall x, x' \in \mathcal{B}_R, \quad \|h(Z - x) - h(Z - x')\|_2 \leq L \|x - x'\|_2 \quad (10)$$

In what follows, we will let

$$R_L \stackrel{\text{def}}{=} \frac{\Lambda}{L}. \quad (11)$$

In the appendix, we derive a localization result that holds for arbitrary kernels h (Theorem C.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 4.2 (Cost function behavior). *We assume that there exists $\mu > 0$ and a radius $R_\mu > 0$ such that*

$$\ell_0(x) \geq \begin{cases} \frac{\mu^2}{2} \|x - \bar{x}\|_2^2 & \text{for } x \in \mathcal{B}_{R_\mu}, \\ \frac{\mu^2}{2} R_\mu^2 & \text{for } x \in \mathcal{B}_R \setminus \mathcal{B}_{R_\mu}. \end{cases} \quad (12)$$

Assumptions 4.1 and 4.2 are illustrated in Fig. 3 (left). The following properties are standard facts:

- We have $\mu \leq L$ since $\frac{\mu^2}{2} \|x - \bar{x}\|_2^2 \leq \ell_0(x) \leq \frac{L^2}{2} \|x - \bar{x}\|_2^2$ under Assumption 4.1.
- The parameter μ^2 can be chosen at least as large as the strong convexity parameter of ℓ_0 (the minimal eigenvalue of ℓ_0') in the ball \mathcal{B}_{R_μ} .
- The quadratic growth in Assumption 4.2 implies the uniqueness of a global minimizer.

Example 4.1 (Reference kernel). *We define a simple kernel h^{ref} for which the quantities involved in Assumptions 4.1 and 4.2 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*

$$h^{\text{ref}} \stackrel{\text{def}}{=} (|x| - 1)_+, \quad M \in 2\mathbb{N} + 1, \quad (z_m = m)_{m \in \{-M/2, \dots, M/2\}}, \quad \bar{x} = 0. \quad (13)$$

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

$$\forall x \in [-M/2 + 1, M/2 - 1], \quad \ell_0^{\text{ref}}(x) = \begin{cases} x^2 & \text{if } |x| \leq 1, \\ 1 - \{x\} + \{x\}^2 & \text{otherwise} \end{cases} \quad (14)$$

where $\{x\} \stackrel{\text{def}}{=} x - \lfloor x \rfloor$ is the fractional part of x . The graph of ℓ_0^{ref} is depicted in Fig. 3 (right). Some elementary calculations yield:

$$L^{\text{ref}} = \mu^{\text{ref}} = \Lambda^{\text{ref}} = \sqrt{2}, \quad R_\mu^{\text{ref}} = \frac{\sqrt{3}}{2}, \quad R_L^{\text{ref}} = 1. \quad (15)$$

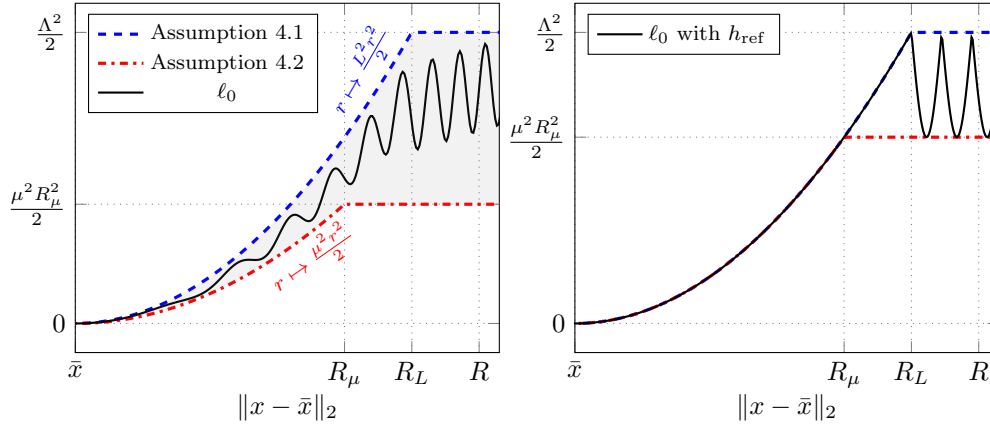


Figure 3: *Left*: Illustration of Assumptions 4.1 and 4.2. The noiseless cost function ℓ_0 is sandwiched between two quadratic functions near \bar{x} and between two constant far away from the origin. *Right*: The specific case h_{ref} where $\mu = L$.

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 conditions 4.1, 4.2 are 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*:*

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

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

Then under either of the following conditions

- $r \leq \frac{R_\mu}{2}$, $R \geq \frac{R_L}{2}$, 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

$$\mathbb{P}(\|\bar{x} - \hat{x}\|_2 \leq r) \geq 1 - \exp\left(-\frac{\rho^2}{2}\right). \quad (16)$$

The complete proof is given in Appendix C. 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*.

* 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.

Theorem 4.2 (Necessary conditions). *Under the same assumptions as in Theorem 4.1:*

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

$$r < \frac{\rho_2 \mu}{2L(c\sqrt{D} + \rho_1)} \cdot \min\left(R_\mu, R_L, \frac{\sigma \rho_2 \mu}{L^2}\right). \quad (\text{Cond}'_1)$$

The proof shows that the term $c\sqrt{D}$ can be discarded if we accept to lower the probability of failure to 1/4 (rather than 1/2 in this form).

- Assume that $R > R_\mu$ and that

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

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

$$\sigma \sqrt{\ln\left(\frac{R}{R_\mu}\right)} > \frac{R_L^2 L^2}{\sqrt{D} R_\mu \mu} + \frac{2\sigma[\rho_2 R_L L + Lrc\sqrt{D} + \rho_1]}{\sqrt{D} R_\mu \mu}. \quad (\text{Cond}'_2)$$

The complete proof is given in Appendix D.

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

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

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

$$\Delta_\varepsilon(x) \sim \mathcal{N}(0, \sigma^2 \|h(Z - x) - h(Z - \bar{x})\|_2^2). \quad (17)$$

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

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

as illustrated on Fig. 4 (left). Then, given that

- $\Delta_\varepsilon(x) \lesssim \sigma L \min(\|x - \bar{x}\|_2, R_L)$ with high probability (from (17)),
- $\ell_0(x) \geq \frac{\mu^2}{2} \min(\|x - \bar{x}\|_2^2, R_\mu^2)$ (from Assumption 4.2),

we can simplify (18) as

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

This is illustrated in Fig. 4 (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 rather deep results on the suprema of random processes [19].

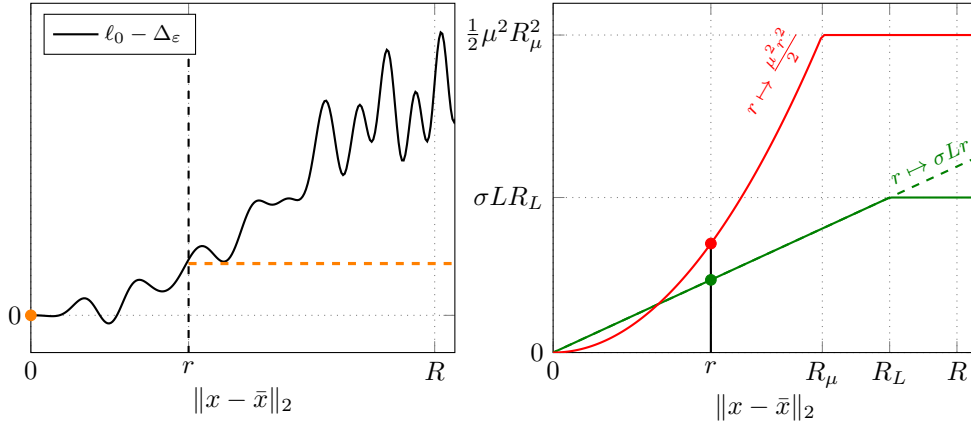


Figure 4: **Intuition behind Theorem 4.1.** *Left:* Success ($\hat{x} \in \mathcal{B}_r$) occurs if there exists $x \in \mathcal{B}_r$ (e.g., $x = \bar{x}$) such that $\ell_0(x) - \Delta_\varepsilon(x)$ (orange point) is lower than the infimum of $\ell_0 - \Delta_\varepsilon$ over $\mathcal{B}_R \setminus \mathcal{B}_r$ (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 .

Intuitions behind Theorem 4.2 A sufficient condition for failure ($\hat{x} \notin \mathcal{B}_r$) is that there exists $x_t \in \mathcal{B}_R \setminus \mathcal{B}_r$ with $t = \|x_t - \bar{x}\|_2$ such that

$$0 \approx \inf_{r \text{ small}} \inf_{x \in \mathcal{B}_r} \ell_0(x) - \Delta_\varepsilon(x) > \ell_0(x_t) - \Delta_\varepsilon(x_t), \quad (19)$$

This condition is illustrated on Fig. 5 (left). Then, given that

- $\Delta_\varepsilon(x_t) \gtrsim \sigma\mu \min(t, R_\mu)$ with probability close to 1/2 (from (17) and since Δ_ε is symmetric),
- $\ell_0(x_t) \leq \frac{L^2}{2} \min(t^2, R_L^2)$ (from Assumption 4.2),

we can simplify (19) as

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

This is illustrated on Fig. 5 (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 $\sigma > L^2 R_L^2 / (2\mu R_\mu)$ 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:

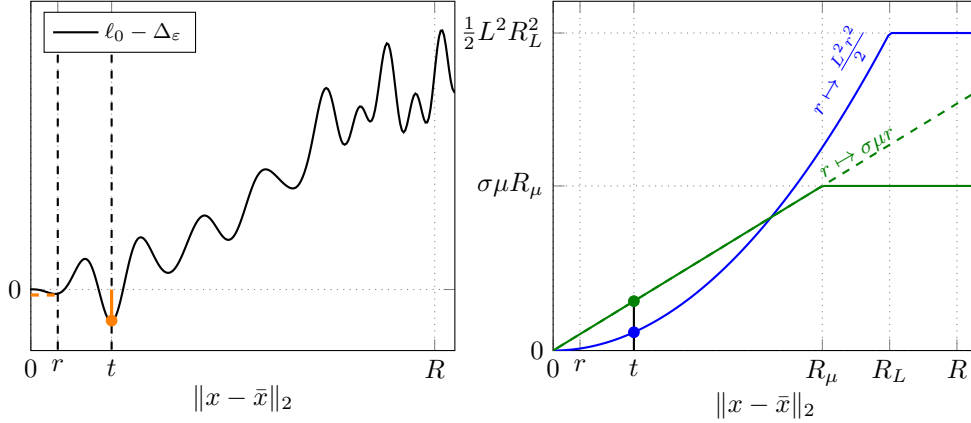


Figure 5: **Intuition behind Theorem 4.2.** *Left:* Failure ($\hat{x} \notin \mathcal{B}_r$) occurs if there exists $x_t \in \mathcal{B}_R \setminus \mathcal{B}_r$ (with $t = \|x_t - \bar{x}\|_2$) such that $\ell_0(x_t) - \Delta_\varepsilon(x_t)$ (orange point) is lower than the infimum of $\ell_0 - \Delta_\varepsilon$ 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 .

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

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

- The Lipschitz constant of the gradient of ℓ_0 :

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

- 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

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

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

In most applications, only the left term in the minimum above plays a role. 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

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

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 provide an intuitive explanation of this, let us recall that:

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

Under our sole Lipschitz assumption, the two random variables $\Delta_\varepsilon(x)$ and $\Delta_\varepsilon(x')$ could be essentially independent for $\|x - x'\|_2 \gtrsim \frac{1}{L}$. A ball of radius R , can be packed with $(R/R_L)^D$ balls of radius R_L . Letting x_n denote the centers of these balls, we would therefore have

$$\mathbb{E} \left[\sup_{x \in \mathcal{B}_R} \Delta_\varepsilon(x) \right] \geq \mathbb{E} \left[\max_n \Delta_\varepsilon(x_n) \right] \gtrsim \sqrt{\ln(R/R_L)^D}.$$

The term $\ln(I)$ We do not know if the term $\ln(I) = \ln(\lceil \log_2(1 + \log_2(R_\mu/r)) \rceil)$ is an artefact of the proof, or if it is actually needed. What is sure, is that it can be safely considered as a moderate constant. 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. Assuming that ℓ_0 is μ^2 -strongly convex on \mathcal{B}_{R_μ} implies the quadratic growth in 12 and we have

$$\mu^2 \leq \inf_{x \in \mathcal{B}_{R_\mu}} \lambda_{\min}(\ell_0''(x)) \leq \sup_{x \in \mathcal{B}_R} \lambda_{\max}(\ell_0''(x)) \leq L^2 \quad (25)$$

where $\ell_0''(x)$ is the Hessian of ℓ_0 at x . Taking $R_\mu \rightarrow 0$ (which implies $r \rightarrow 0$ and $\sigma \rightarrow 0$ from Cond_2), we obtain

$$\mu^2 \leq \lambda_{\min}(\ell_0''(\bar{x})) \leq \lambda_{\max}(\ell_0''(\bar{x})) \leq L^2. \quad (26)$$

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

$$\mu^2 \leq \lambda_{\min}(\ell_0''(\bar{x})) \leq \frac{\|h'(Z - \bar{x})\|_F^2}{D} \leq \lambda_{\max}(\ell_0''(\bar{x})) \leq L^2 \quad (27)$$

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

$$\|A\|_F^2 = \sum_{d=1}^D \sigma_d^2(A) = \sum_{d=1}^D \lambda_d(AA^T) \quad (28)$$

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,

$$r \gtrsim \frac{\sqrt{D}\sigma L}{\mu^2} \gtrsim \frac{D \cdot \sigma}{\|h'(Z - \bar{x})\|_F} = \sqrt{\text{CR}_\sigma}. \quad (29)$$

This is in line with the asymptotic analysis existing in the literature [13, 16].

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. [18, 9]. 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:

$$\mu(r) \stackrel{\text{def}}{=} \inf_{x \in \mathcal{B}_r} \|h'(Z - x)\|_F. \quad (30)$$

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 very 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. Hence, unlocalized kernels could yield interesting results, in the low density regime, where only scattered sources are present.

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

$$\mathcal{F}(f)(\omega) \stackrel{\text{def}}{=} \frac{1}{(2\pi)^{D/2}} \int_{\mathbb{R}^D} f(x) \exp(-i\langle x, \omega \rangle) dx. \quad (31)$$

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 $\text{supp}(\mathcal{F}(f)) \subseteq [-W, W]^D$. 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). 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:

$$\langle f, g \rangle_{L^2(\mathbb{R}^D)} = \left(\frac{\tau}{2\pi}\right)^D \sum_{z \in \tau\mathbb{Z}^D} f(z-x)g(z-x). \quad (32)$$

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

$$\|f\|_{L^2(\mathbb{R}^D)}^2 = \left(\frac{\tau}{2\pi}\right)^D \sum_{z \in \tau\mathbb{Z}^D} f(z-x)^2. \quad (33)$$

Proof. The proof is given in Appendix E for completeness. \square

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

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

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

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

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

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

Now let Λ , L , and μ be defined according to Assumptions 4.1 and 4.2 for $x \mapsto \ell_0(x)\tau^D$ so that they are independent of τ . Then, the conditions (22) and (23) read

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

$$\sigma \lesssim \frac{L}{\tau^{D/2}} \min \left(r\kappa^2, \frac{R_L\theta}{\sqrt{D}} \ln \left(\frac{R}{R_\mu} \right)^{-1/2} \right). \quad (35)$$

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

Hence, for well sampled bandlimited kernels, we have that:

- 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 (34) and (35)).

5. Acknowledgments

This work was supported by the ANR Micro-Blind (grant ANR-21-CE48-0008) 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.

Appendices

A. Derivatives of ℓ_ε

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

$$\ell_\varepsilon(x) = \frac{1}{2} \sum_{m=1}^M (h(z_m - x) - h(z_m - \bar{x}) - \varepsilon_m)^2 \quad (\text{A.1})$$

$$\ell'_\varepsilon(x) = - \sum_{m=1}^M (h(z_m - x) - h(z_m - \bar{x}) - \varepsilon_m) h'(z_m - x) \quad (\text{A.2})$$

$$\ell''_\varepsilon(x) = \sum_{m=1}^M (h(z_m - x) - h(z_m - \bar{x}) - \varepsilon_m) h''(z_m - x) + h'(z_m - x) h'(z_m - x)^T \quad (\text{A.3})$$

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

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

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

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

B. Proof of Theorem 2.1

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

By [10, (3.20)], we know that the covariance matrix of any unbiased estimator \hat{x} satisfies

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

In our case, we have

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

Hence

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

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

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

Summing over d yields

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

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 [10, 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*

$$\frac{\partial \ln p(y|\bar{x})}{\partial \bar{x}_d} = I(\bar{x}) \cdot (g(y) - \bar{x}), \quad (\text{B.5})$$

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

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

where we have introduced the following notations

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

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

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

Differentiating the equality with respect to y yields

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

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

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

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

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

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

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

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

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

which implies that $I(\bar{x})$ does not depend 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

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

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

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

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 $g(y) = \frac{\beta}{a} + \frac{1}{M} \sum_{m=1}^M h(z_m - \frac{y_m}{a})$ are appropriate. \square

C. Proof of Theorem 4.1

C.1. An intermediary result

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

Theorem C.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*

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

where c is a universal constant. Set $\rho > 0$. Under Assumptions 4.1 and 4.2. We have $\hat{x} \in \mathcal{B}_r$ with probability larger than $1 - \exp(-\rho^2/2)$ under the I conditions $1 \leq i \leq I$:

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

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

C.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 $\Delta_\varepsilon(x) \stackrel{\text{def}}{=} \langle h(Z - x) - h(Z - \bar{x}), \varepsilon \rangle$. We partition the domain Ω as

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

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

$$\begin{aligned} [\|\hat{x} - \bar{x}\|_2 \leq r] &\Leftrightarrow \left[\inf_{x \in \mathcal{B}_r} \ell_\varepsilon(x) < \inf_{x \in \mathcal{B}_r^c} \ell_\varepsilon(x) \right] \\ &\Leftrightarrow \left[\inf_{x \in \mathcal{B}_r} \ell_0(x) - \Delta_\varepsilon(x) < \inf_{x \in \mathcal{B}_r^c} \ell_0(x) - \Delta_\varepsilon(x) \right] \\ &\Leftrightarrow \left[\inf_{x \in \mathcal{B}_r} \ell_0(x) - \Delta_\varepsilon(x) < \inf_{x \in \Omega_i} \ell_0(x) - \Delta_\varepsilon(x), \forall 1 \leq i \leq I \right] \\ &\Leftrightarrow \left[\underbrace{\ell_0(\bar{x}) - \Delta_\varepsilon(\bar{x})}_0 < \inf_{x \in \Omega_i} \ell_0(x) - \sup_{x \in \Omega_i} \Delta_\varepsilon(x), \forall 1 \leq i \leq I \right] \\ &\Leftrightarrow \left[\inf_{x \in \Omega_i} \ell_0(x) > \sup_{x \in \Omega_i} \Delta_\varepsilon(x), \forall 1 \leq i \leq I \right]. \end{aligned} \quad (\text{C.3})$$

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 C.3 and the one before to be close. The interest of C.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.

C.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 [19, 5, 22] for an in depth treatment of this topic. In our specific case, we obtain the following result.

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

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

For $t \geq 0$, we have

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

with

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

Proof. First notice that the random process Δ_ε is Gaussian since $\varepsilon \sim \mathcal{N}(0, \sigma^2 \text{Id})$. It can be written alternatively as $\sigma \Delta_\epsilon$ with $\epsilon \sim \mathcal{N}(0, \text{Id})$. 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

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

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

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

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 $\mathcal{D} \subset \mathbb{R}^D$

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

Under Assumption 4.1, we have

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

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

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

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 Assumption 4.1. Let us introduce the pseudo-metric $d : \mathbb{R}^D \times \mathbb{R}^D \rightarrow \mathbb{R}$ defined by

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

$$= \sqrt{\sigma^2 \|h(Z - x) - h(Z - x')\|_2^2} = \sigma \|h(Z - x) - h(Z - x')\|_2 \quad (\text{C.10})$$

It will be used in conjunction with the following tool.

Definition C.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 C.2 (Dudley's inequality). *The following inequality holds:*

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

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

Lemma C.2. *Under assumption 4.1, we have*

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

for some universal constant c .

Proof. In what follows, c is a universal constant that may change from one line to the other. By assumption, we have

$$d(x, x') \leq \sigma \cdot \min(L \|x - x'\|_2, \Lambda) \quad (\text{C.13})$$

and we can upper bound E_i as

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

Using (C.13) and [22, Cor. 4.2.13], we obtain

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

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

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

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

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

where we use the inequality $\operatorname{erfc}(z) < \frac{\exp(-z^2)}{\sqrt{\pi}z}$ 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:

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

□

□

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

$$\theta_i \geq E_i, \quad \forall 1 \leq i \leq I. \quad (\text{C.16})$$

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

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

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

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

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

C.2. Concluding the proof of Theorem 4.1

Proof. From Theorem C.1, the inequality (16) is valid if

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

Assumption 4.2, allows us to get

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

Let us assume that $R > \frac{R_L}{2}$. Without loss of generality, we can also assume that $R_\mu \leq \frac{R_\mu}{2}$, since the inequalities in (12) are still valid when replacing \mathcal{B}_{R_μ} by $\mathcal{B}_{R'}$ with $R' \leq R_\mu$. In this setting, the success condition \mathcal{C}_I reads

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

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

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

This can be rewritten as:

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

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

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

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. \square

D. 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 (C.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

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

Proof of the part related to (Cond'₁) Now, set $t \in (r, R_\mu]$ and take an arbitrary point x with $\|x - \bar{x}\|_2 = t$. At this x , we have $\ell_0(x) \leq \min\left(\frac{L^2 t^2}{2}, \frac{\Lambda^2}{2}\right)$ under Assumption 4.1. In addition, we have $\Delta_\varepsilon(x) \sim \mathcal{N}(0, \sigma^2 \|h(Z - x) - h(Z - \bar{x})\|_2^2)$ with $\|h(Z - x) - h(Z - \bar{x})\|_2^2 \geq \frac{1}{2}\mu^2 t^2$ under Assumption 4.2. The previous inequality implies that

$$\mathbb{P}(\Delta_\varepsilon(x) \geq \rho_2 \sigma \mu t) \geq 1 - \Phi(\rho_2/\sqrt{2}),$$

where Φ is the cumulative distribution function of the normal distribution. Define the event

$$\mathcal{E}_2 \stackrel{\text{def}}{=} \left[\inf_{x \in \mathcal{B}_r^c} \ell_0(x) - \Delta_\varepsilon(x) \leq \min\left(\frac{L^2 t^2}{2}, \frac{\Lambda^2}{2}\right) - \rho_2 \sigma \mu t \right].$$

We have $\mathbb{P}(\mathcal{E}_2) \geq 1 - \Phi(\rho_2/\sqrt{2})$ and

$$\begin{aligned} \mathbb{P}(\mathcal{E}_1 \cap \mathcal{E}_2) &= 1 - \mathbb{P}(\mathcal{E}_1^c \cup \mathcal{E}_2^c) \geq 1 - \mathbb{P}(\mathcal{E}_1^c) - \mathbb{P}(\mathcal{E}_2^c) = \mathbb{P}(\mathcal{E}_1) + \mathbb{P}(\mathcal{E}_2) - 1 \\ &\geq 1 - \exp(-\rho_1^2/2) + 1 - \Phi(\rho_2/\sqrt{2}) - 1 = 1 - \exp(-\rho_1^2/2) - \Phi(\rho_2/\sqrt{2}). \end{aligned}$$

We get $\hat{x} \notin \mathcal{B}_r$ under the event $\mathcal{E}_1 \cap \mathcal{E}_2$ together with the condition

$$\exists t \in (r, R_\mu], \sigma \left(\rho_2 \mu t - Lr \left(c\sqrt{D} + \rho_1 \right) \right) > \min\left(\frac{L^2 t^2}{2}, \frac{\Lambda^2}{2}\right). \quad (\text{D.1})$$

By setting $\rho_2 \mu t = 2Lr \left(c\sqrt{D} + \rho_1 \right)$, we get failure with probability larger than $1 - \exp(-\rho_1^2/2) - \Phi(\rho_2/\sqrt{2})$ under the conditions:

$$t > r \quad \text{and} \quad t \leq R_\mu \quad \text{and} \quad \frac{L^2 t^2}{2} \leq \frac{\Lambda^2}{2} \quad \text{and} \quad (\text{D.1}) \Leftrightarrow \sigma \frac{\rho_2 \mu t}{2} > \frac{L^2 t^2}{2}.$$

These conditions can still be rewritten in terms of the radius r and the parameters ρ_1 and ρ_2 as

$$\frac{2L(c\sqrt{D} + \rho_1)}{\rho_2\mu} > 1$$

$$r < \frac{\rho_2\mu}{2L(c\sqrt{D} + \rho_1)} \cdot \min\left(R_\mu, \frac{\Lambda}{L}, \frac{\sigma\rho_2\mu}{L^2}\right).$$

A sufficient condition for the first inequality reads $\rho_2 \leq \rho_1$ since $L \geq \mu^*$. For instance, set $\rho_1 = 3$ and $\rho_2 = \sqrt{2}$. We then get $\mathbb{P}(\mathcal{E}_1 \cap \mathcal{E}_2) \geq 0.14$. This shows that for a sufficiently small radius r , a necessary condition for $\hat{x} \in \mathcal{B}$ with probability larger than 0.86 is

$$r > \frac{\sigma\mu^2}{L^3(2c\sqrt{D} + 6)}. \quad (\text{D.2})$$

In the case $\mu \propto L$, this yields a condition of the form $r \lesssim \frac{\sigma}{L}$, which is in essence equivalent to the Cramér-Rao bound, but in a non asymptotic setting.

Proof of the part related to (Cond'₂) Here, we assume that $R > R_\mu$ and that the growth condition is globalized:

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

Hence, we get using [22, Cor. 4.2.13]:

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

Using Sudakov's inequality

cite[Thm 7.4.1]vershynin2018high 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

$$\mathbb{E} \left[\sup_{x \in \mathcal{B}_R} \Delta_\varepsilon(x) \right] \geq \frac{\sqrt{D}}{2} R_\mu\sigma\mu \sqrt{\ln \left(\frac{R}{R_\mu} \right)}. \quad (\text{D.3})$$

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

$$\mathbb{P} \left(\mathbb{E} \left[\sup_{x \in \mathcal{B}_R} \Delta_\varepsilon(x) \right] - \sup_{x \in \mathcal{B}_R} \Delta_\varepsilon(x) \geq t \right) \leq \exp \left(-\frac{t^2}{2\sigma^2\Lambda^2} \right). \quad (\text{D.4})$$

Define the event

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

* In fact, this condition is also satisfied under weak conditions such as $\rho_2 \leq 2c\sqrt{D}$ and the analysis shows that $c \geq 1$.

Combining the previous results, we obtain

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

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

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

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

E. Proof of Theorem 4.3

The equality (32) is a consequence of standard results in sampling theory. Set $W > 0$ and let s_W denote the function whose 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 $(s_W(\cdot - n/W))_{n \in \mathbb{Z}^D}$ is an orthonormal basis of $\mathcal{PW}(\pi W)$ (see e.g. [11, 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)

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

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:

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

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

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

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] Dan Chazan, Moshe Zakai, and Jacob Ziv. Improved lower bounds on signal parameter estimation. *IEEE transactions on Information Theory*, 21(1):90–93, 1975.
- [7] 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.
- [8] Valentin Debarnot and Pierre Weiss. Blind inverse problems with isolated spikes. *Information and Inference: A Journal of the IMA*, 2022.
- [9] 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.
- [10] Steven Kay. Fundamentals of statistical signal processing: estimation theory. *Technometrics*, 37(4):465–466, 1993.
- [11] Stephane Mallat. *A wavelet tour of signal processing*. Academic Press, 3 edition, 2008.
- [12] William E Moerner and Lothar Kador. Optical detection and spectroscopy of single molecules in a solid. *Physical review letters*, 62(21):2535, 1989.
- [13] 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.
- [14] Raimund J Ober, Sripad Ram, and E Sally Ward. Localization accuracy in single-molecule microscopy. *Biophysical journal*, 86(2):1185–1200, 2004.
- [15] Calyampudi Radhakrishna Rao. *Linear statistical inference and its applications*. Wiley New York, 1973.
- [16] 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.
- [17] 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.
- [18] 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.
- [19] Michel Talagrand. Upper and lower bounds for stochastic processes. decomposition theorems. *Ergebnisse der Mathematik und ihrer Grenzgebiete*, 2021.
- [20] 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.
- [21] Harry L Van Trees. *Detection, estimation, and modulation theory, part I: detection, estimation, and linear modulation theory*. John Wiley & Sons, 2004.
- [22] Roman Vershynin. *High-dimensional probability: An introduction with applications in data science*. Cambridge university press, 2018.
- [23] Ehud Weinstein and A Weiss. Fundamental limitations in passive time-delay estimation—part ii: Wide-band systems. *IEEE transactions on acoustics, speech, and signal processing*, 32(5):1064–1078, 1984.