Superresolution imaging across scales
Supplementary Information

Esley Torres García[1,*], Roberto Rodríguez Morales[2], Paul Hernández Herrera[3], Rocco D'Antuono[4], Patrick Hartigan[5], Remy Fernand Ávila Foucat[6], W. Luis Mochán Backal[7], Guillermina González-Rosendo[8], Christopher D. Wood[9], Adan Guerrero[9,10,*], and Federico Lecumberry[1,*]

[1]Facultad de Ingeniería, Universidad de la República, Montevideo, Uruguay.
[2]Instituto de Cibernética, Matemática y Física, Ciudad de la Habana, Cuba.
[3]Facultad de Ciencias, Universidad Autónoma de San Luis Potosí, San Luis Potosi, SLP, Mexico.
[4]Crick Advanced Light Microscopy Facility, London, UK.
[5]Physics and Astronomy Dept., Rice University, 6100 S. Main, Houston, TX 77005-1892.
[6]Centro de Física Aplicada y Tecnología Avanzada, Universidad Nacional Autónoma de México, Boulevard Juriquilla 3001, 76230, Querétaro, Qro., Mexico.
[7]Instituto de Ciencias Físicas, Universidad Nacional Autónoma de México, Av. Universidad S/N, Col. Chamilpa, 62210, Cuernavaca, Morelos, Mexico.
[8]Facultad de Química, Universidad Nacional Autónoma de México, Mexico.
[9]Laboratorio Nacional de Microscopía Avanzada, Instituto de Biotecnología, Universidad Nacional Autónoma de México, Cuernavaca, Morelos, Mexico.
[10]Centro Internacional de Ciencias A.C. Avenida Universidad 1001 Campus UAEM-UNAM. C.P. 62131. Cuernavaca, Mor. Mexico.
[*]Correspondence should be addressed to: adan.guerrero@ibt.unam.mx.

Overview

This supplementary document presents a comprehensive theoretical framework demonstrating how the MeanShift (MS) method can surpass the classical Sparrow resolution limit. While the main text focuses on the conceptual application and experimental validation of overcoming classical resolution barriers, this text delivers the analytical proofs required to validate those gains. To ensure a structured and clear derivation, the theoretical framework is systematically decomposed into interconnected mathematical steps, ranging from fundamental symmetry properties to discrete and continuous data models. Before detailing each formal proof, it is instructive to map out the global architecture of these derivations to understand how individual properties chain together to challenge traditional optical limits.

The analysis is structured in three stages and appendix. First, Proposition 1, the derivation of the Sparrow limit using Gaussian emitter models establishes that two sources are at the limit when separated by 2σ. Second, Theorem 1 extends the results to continuous data, proving that MS surpasses the Sparrow limit for all $h > 0$. A sequence of lemmas—rooted in properties of symmetry, differentiability, and the behavior of even functions—supports the theoretical claims. The curvature at the midpoint between emitters, reflected in the second derivative of the intensity profile, serves as a central analytical tool for determining resolvability. Third, Theorem 2 shows that for discrete data, resolution beyond this limit is achieved when the kernel radius satisfies $h < 1.915\sigma$. And fourth, an appendix provides specific information and technical details for both the main manuscript and this supplementary document, structured so as not to overwhelm the reader. Figures, simulations, and flowcharts illustrate the logical progression from fundamental principles to practical resolution gains. Together, these results offer a rigorous foundation for understanding how MS-based methods can analytically overcome classical resolution barriers in both discrete and continuous imaging scenarios.

Figure 1 presents a structured overview of the theoretical framework developed to analyze and surpass the Sparrow resolution limit. This visual overview reinforces the logical flow of lemmas and theorems supporting our main results. The diagram is divided into three main stages: determining the emitter distance that satisfies the Sparrow condition, analyzing data at the Sparrow limit in discrete form, and finally extending the analysis to the continuous domain. The first stage uses Claims 1–2, which establish properties of symmetry in even functions, to prove Proposition 1, showing that two Gaussian emitters are at the Sparrow limit when separated by 2σ. The second stage transitions to continuous data, where Lemmas 3–6 capture kernel properties at the origin, and Lemmas 7-8 focuses on the vanishing derivative of a coefficient function. The final stage addresses the behavior of discrete data at this limit, relying on symmetry (Lemmas 9–10) and derivative behavior (Lemmas 11–12) to support Theorem 2, which characterizes minimum separation using discrete samples.

Contents

1 Emitters' distance to satisfy Sparrow limit

Before extending the capabilities of the MS method to super-resolution regimes, it is essential to establish a rigorous baseline using classical resolution criteria. This first stage develops the analytical groundwork by formally defining the standard Sparrow limit under ideal Gaussian conditions. The following section introduces the fundamental symmetry properties, lemmas, and propositions that govern emitter interaction at this critical boundary.

1.1 Lemmas and theorems for proof of emitters at the Sparrow limit

It is well established in classical optics that, for incoherent Gaussian point-spread functions, the Sparrow resolution limit occurs at a separation of 2σ[1]. Although explicit derivations are uncommon in the literature, we present here a complete analytical demonstration based on fundamental properties of even functions and their derivatives. The following two claims are used in this section to highlight the key symmetry properties of an even function. Claim 1 states that if an even function is differentiable, its derivative at zero is necessarily zero at the origin. Claim 2 confirms that the sum of two identically even functions centered at symmetric positions from the origin $F(x) = f(x-a) + f(x+a)$ is also an even function, allowing us to deduce that the stationary point occurs at $x = 0$. These claims simplify the proof of Proposition 1 by reducing the problem to analyzing the sign of the second derivative at the origin, which directly determines whether the central point is a maximum, minimum, or inflection point.

Lemma 1 (Even function derivative). *Let $f(x)$ be a differentiable even function, then $f'(0) = 0$.*

Lemma 2 (Sum of Shifted Even Functions). *Let $f(x)$ be an even function; then the function $F(x) = f(x-a) + f(x+a)$ is also even.*

Proposition 1 (Sparrow limit for Gaussian emitters). *The Sparrow resolution limit for two Gaussian point sources is reached when the distance between their centers is 2σ, where σ is the standard deviation of the Gaussian profile.*

Proof. Let G_a be the sum of two Gaussian functions centered at $\pm a$ defined as:

$$G_a(x) = g_{-a}(x) + g_a(x) = e^{-\frac{1}{2}\frac{(x-a)^2}{\sigma^2}} + e^{-\frac{1}{2}\frac{(x+a)^2}{\sigma^2}} \tag{1}$$

Since $g(x)$ is an even function, by lemmas 1 and 2, then $G'(0) = 0$. This satisfies the first derivative condition for a stationary point at $x = 0$, valid for all $a \in \mathbb{R}$.

We now analyze the second derivative $G''_a(x)$, specifically at $x = 0$. By computing:

$$G''_a(x) = g''_{-a}(x) + g''_a(x) = \left(\frac{(x-a)^2}{\sigma^4} - \frac{1}{\sigma^2}\right) e^{-\frac{1}{2}\frac{(x-a)^2}{\sigma^2}} + \left(\frac{(x+a)^2}{\sigma^4} - \frac{1}{\sigma^2}\right) e^{-\frac{1}{2}\frac{(x+a)^2}{\sigma^2}}.$$

Evaluating at $x = 0$, we obtain:

$$\begin{aligned} G''_a(0) &= \left(\frac{(-a)^2}{\sigma^4} - \frac{1}{\sigma^2}\right) e^{-\frac{1}{2}\frac{(-a)^2}{\sigma^2}} + \left(\frac{(a)^2}{\sigma^4} - \frac{1}{\sigma^2}\right) e^{-\frac{1}{2}\frac{(a)^2}{\sigma^2}} \\ &= \frac{2}{\sigma^2}\left(\frac{a^2}{\sigma^2} - 1\right) e^{-\frac{1}{2}\frac{(a)^2}{\sigma^2}}. \end{aligned} \tag{2}$$

From equation 2, since $e^{-\frac{1}{2}\frac{(a)^2}{\sigma^2}} > 0\ \forall a \in \mathbb{R}$, the sign of $G''_a(0)$ depends only on the term $\frac{a^2}{\sigma^2} - 1$. So, we analyze three cases:

- Case 1: $|a| = \sigma \Rightarrow \frac{(a)^2}{\sigma^2} = 1 \Rightarrow G''_a(0) = 0$,
- Case 2: $|a| > \sigma \Rightarrow \frac{(a)^2}{\sigma^2} > 1 \Rightarrow G''_a(0) > 0$
- Case 3: $|a| < \sigma \Rightarrow \frac{(a)^2}{\sigma^2} < 1 \Rightarrow G''_a(0) < 0$.

These cases imply the following classification at $x = 0$:

- a inflextion point when $|a| = \sigma$ (Sparrow condition).
- a local minimum when $|a| > \sigma$ (emitters are resolved),
- and a local maximum when $|a| < \sigma$ (emitters are unresolved).

Thus, the critical distance at which the two Gaussian sources transition from unresolved to resolved corresponds when $|a| = \sigma$. This confirms that the Sparrow limit is reached when a distance of 2σ separates the emitters. $\square$

Figure 2a illustrates the combined intensity profiles of two emitters positioned at varying distances. Three cases are shown: emitters are resolved ($|a| > \sigma$, yellow curve), emitters are at the Sparrow limit ($a = \pm\sigma$, blue curve), and emitters are unresolved ($|a| < \sigma$, red). The Sparrow limit is defined as the separation at which the second derivative of the intensity at the midpoint vanishes, corresponding to a distance of 2σ between emitters. Figure 2b shows the sign of the second derivative at the origin, $G_a''(0)$, for these different separations when varying the emitters distance.

2 MS theory in 1-dimensional continuous case

Following the formal derivation of the classical Sparrow limit under static Gaussian assumptions in the previous section, we now extend our theoretical framework to investigate the behavior of the *MS* method when applied to continuous data models. While traditional optical criteria dictate a rigid resolution boundary at a separation of 2σ, this section demonstrates that the *MS* algorithmic framework analytically surpasses this limit under continuous formulations for any kernel bandwidth $h > 0$.

To establish this capability with mathematical rigor, we introduce a systematic sequence of lemmas rooted in the fundamental properties of symmetry, differentiability, and the behavior of even functions under continuous integration. Central to this analytical progression is the tracking of the intensity profile's curvature at the exact midpoint between emitters, captured via its second-derivative behavior. These foundational lemmas directly support Theorem 1, providing the formal proof that continuous *MS* localization effectively dissolves classical resolution barriers.

2.1 MS theory for Continuous 1D Data Models

Let $x \in X \subset \mathbb{R}$ (X spatial space) and a differentiable function $f : Y \subset \mathbb{R} \to \mathbb{R}$ (Y range space). A punctured neighborhood of x with fixed radius $h > 0$ is denoted by $B \subset \mathbb{R}$ and is defined as:

$$B = B_h^*(x) = \{t \in [x-h, x+h] \backslash \{x\}\}.$$

Let k a differentiable profile over B defined as a function of f, given by:

$$k(x,t) = k\left(\left|\frac{f(t)-f(x)}{m(x,t)}\right|^2\right),$$

where $m(x) = \max_{t \in B} |f(t) - f(x)|$ and $t \in B$. k must satisfy the following conditions:

- non-negative
- non-increasing
- piece-wise continuous
- differentiable
- integrable.

The weights associated with the *MS* are defined as:

$$q(x,t) = \left(\frac{k(x,t)}{K(x)}\right),$$

where $K(x) = \int_{t \in B} k(x,t)\, dt$. Using these weights, *MS* formula becomes:

$$MS_h(f,x) = \int_B q(x,t) f(t) dt - f(x), \tag{3}$$

The main idea in the lemmas that follow is to exploit the property of parity of f, and they will be used to support the proof of Theorem 1.

2.2 Lemmas and theorems for continuous data

We now extend the MS formulation to the continuous case by analyzing the kernel's behavior around the origin. Also, assume for the rest of the section that function f is the even two-emitter intensity from Proposition 1, as in equation 1.

Lemma 3 (Kernel positive integral)**.** *Let k be the profile of a kernel defined over a neighborhood $B_h(x)$ with radius $h > 0$ and centered at a point x. Suppose that f is not constant within $B_h(x)$. Then the integral of $k(x,t)$ over the neighborhood is strictly positive:*

$$\int_{x-h}^{x+h} k(x,t)\, dt > 0, \ \ \forall x \in \mathbb{R}.$$

Proof. Since f is not constant in the interval $B_h(x)$, there exists at least one point $t \in B_h(x)$ such that $f(x) \neq f(t)$. By definition of the kernel function k (see kernel definition in previous Subsection 2.1), k is non-negative this implies that $k(x,t) > 0$ at that point. Furthermore, because $h > 0$, the interval $(x-h, x+h)$ has positive length. Therefore, the integral of $k(x,t) > 0$ in this interval is strictly positive. □

Lemma 4 (Eveness of kernel and *MS* coefficients)**.** *Let f be an even function, then $k(0,t)$ and $q(0,t)$ are even functions.*

Proof.

$$k(\pm x, \pm t) = k\left(\left|\frac{f(\pm t) - f(\pm x)}{\max\limits_{t \in B}\{|f(\pm t) - f(\pm x)|\}}\right|^2\right) = k\left(\left|\frac{f(t) - f(x)}{\max\limits_{t \in B}\{|f(t) - f(x)|\}}\right|^2\right) = k(x,t).$$

Thus the kernel is symmetric for both arguments, and in particular this holds when $x = 0$ we obtain $k(0,t) = k(0,-t)$.

For q are the following steps: $q(0,t) = \dfrac{k(0,t)}{K(0)}$ where $K(0) = \displaystyle\int_{-h}^{h} k(0,t)\, dt$.

Now

$$q(0,-t) = \frac{k(0,-t)}{K(0)} = \frac{k(0,t)}{K(0)} = q(0,t),$$

thus $q(0,t)$ is even. □

Lemma 5 (First derivative of kernel integral)**.** *If $x = 0$ then* $\left[\int_{x-h}^{x+h} k(x,t)\, dt\right]' = 0$

Proof. We apply the Fundamental Theorem of Calculus:

$$\left[\int_{x-h}^{x+h} k(x,t)\, dt\right]' = k(x,t)\Big|_{t=x-h}^{t=x+h} + \int_{x-h}^{x+h} \frac{\partial k(x,t)}{\partial x} dt.$$

Evaluating in $x = 0$ the first term and applying Lemma 4, we obtain:

$$k(0, 0+h) - k(0, 0-h) = k(0,h) - k(0,-h) = k(0,h) - k(0,h) = 0.$$

The last term becomes zero because $\dfrac{\partial k(x,t)}{\partial x}$ is an odd function of t (from the even symmetry of k in t at $x = 0$, see Lemma 4). □

Lemma 6 (First derivative of coefficients for *MS*)**.** *If $x = 0$, then* $\displaystyle\int_{x-h}^{x+h} \frac{\partial q(x,t)}{\partial x} f(t)\, dt = 0.$

Proof. To evaluate the integral, we first compute $\dfrac{\partial q(x,t)}{\partial x}$. By definition,

$$q(x,t) = \frac{k(x,t)}{K(x)}, \quad \text{where} \quad K(x) = \int k(x,t)\, dt.$$

Assuming that the kernel $k(x,t)$ is symmetric with respect to $x = 0$, it follows that $\dfrac{\partial q(x,t)}{\partial x}$ is an odd function in t when $x = 0$. Furthermore, under typical conditions, the function $f(t)$ is symmetric (for instance, when representing Gaussian point sources centered at $t = 0$, such as in the Sparrow configuration). The product $\dfrac{\partial q(x,t)}{\partial x} f(t)$ is therefore an odd function with respect to $t = 0$.

Integrating an odd function over a symmetric interval yields zero. Hence, for $x = 0$,

$$\int_{x-h}^{x+h} \frac{\partial q(x,t)}{\partial x} f(t)\, dt = 0.$$

□

Lemma 7 (Kernel first and second derivatives at $x = 0$)**.** *If $x = 0$, then:*

- $\left.\dfrac{\partial k(x,t)}{\partial x}\right|_{x=0} = 0$, *and*
- $\left.\dfrac{\partial^2 k(x,t)}{\partial x^2}\right|_{x=0} = 0.$

Proof. We are interested in the second derivative with respect to x of the kernelized function $k(x,t)$ evaluated at $x = 0$. Let us define:

- $u(x,t) = \left(\frac{f(t)-f(x)}{m(x)}\right)^2$
- $k(x,t) = k(u(x,t))$

Our goal is to compute $\dfrac{\partial^2 k(x,t)}{\partial x^2}$ at $x = 0$. By the chain rule,

$$\frac{\partial k(x,t)}{\partial x} = k'(u(x,t)) \cdot \frac{\partial u(x,t)}{\partial x}$$

Let us compute $\frac{\partial u(x,t)}{\partial x}$:

$$u(x,t) = \frac{(f(t) - f(x))^2}{m(x)^2}.$$

So,

$$\frac{\partial u(x,t)}{\partial x} = \frac{2(f(t) - f(x))(-f'(x))}{m(x)^2} - \frac{2(f(t) - f(x))^2 m'(x)}{m(x)^3}.$$

At $x = 0$, since f is even, $f'(0) = 0$. If we also assume $m(x)$ is locally constant near $x = 0$ (because of the flatness by the Sparrow condition, see Figure 2), then $m'(0) = 0$. Thus,

$$\left.\frac{\partial u(x,t)}{\partial x}\right|_{x=0} = 0,$$

and consequently, the first derivative of $k(x,t)$ with respect to x also vanishes at $x = 0$:

$$\left.\frac{\partial k(x,t)}{\partial x}\right|_{x=0} = 0.$$

Applying the chain rule and product rule,

$$\frac{\partial^2 k(x,t)}{\partial x^2} = k''(u(x,t)) \left(\frac{\partial u(x,t)}{\partial x}\right)^2 + k'(u(x,t)) \frac{\partial^2 u(x,t)}{\partial x^2}$$

At $x = 0$, the first term vanishes because $\dfrac{\partial u}{\partial x}(0,t) = 0$, so:

$$\left.\frac{\partial^2 k(x,t)}{\partial x^2}\right|_{x=0} = k'(u(0,t)) \left.\frac{\partial^2 u(x,t)}{\partial x^2}\right|_{x=0}$$

Now, compute $\dfrac{\partial^2 u(x,t)}{\partial x^2}$ at $x = 0$:

$$\frac{\partial^2 u(x,t)}{\partial x^2} = \frac{2f''(x)(f(t) - f(x))}{m(x)^2} + (\text{terms involving } m'(x), m''(x))$$

Assuming $m(x)$ is constant near $x = 0$, the terms involving $m'(x)$ and $m''(x)$ vanish, so:

$$\left.\frac{\partial^2 u(x,t)}{\partial x^2}\right|_{x=0} = \frac{2f''(0)(f(t) - f(0))}{m(0)^2}.$$

Now recalling that f satisfy the Sparrow condition ($f''(0)=0$), therefore,

$$\begin{aligned}\left.\frac{\partial^2 k(x,t)}{\partial x^2}\right|_{x=0} &= k'(u(0,t))\cdot\frac{2f''(0)(f(t)-f(0))}{m(0)^2}\\ &= f''(0)\cdot k'(u(0,t))\cdot\frac{2(f(t)-f(0))}{m(0)^2}.\end{aligned}$$

So

$$\left.\frac{\partial^2 k(x,t)}{\partial x^2}\right|_{x=0}=0.$$

This result is critical in simplifying higher-order derivatives of the MS operator at the origin.

□

Lemma 8 (MS of second derivative coefficient)**.** *If* $x=0$ *then* $\left.\int_{x-h}^{x+h}\frac{\partial^2 q(x,t)}{\partial x^2}f(t)\,dt\right|_{x=0}=0.$

Proof. Applying similar steps than lemma 6 and the second result of Lemma 7, we it is easy to obtain:

$$\left.\frac{\partial^2 q(x,t)}{\partial x^2}\right|_{x=0}=\frac{1}{K(0)}\cdot\left.\frac{\partial^2 k(x,t)}{\partial x^2}\right|_{x=0}=0,$$

where $K(0)=\int_{-h}^{h}k(0,s)\,ds$ is constant. Thus,

$$\left.\int_{x-h}^{x+h}\frac{\partial^2 q(x,t)}{\partial x^2}f(t)\,dt\right|_{x=0}=0.$$

□

Theorem 1 (Continuous Domain Resolution Enhancement)**.** *Let two point sources be described by Gaussian distribution that satisfy the Sparrow condition. Then the MS operator has a local maximum at the midpoint ($x=0$) for any spatial neighborhood radius $h>0$.*

Proof. We aim to show that the Mean Shift function has a local maximum at the midpoint ($x=0$) between two Gaussian sources. This will be proven by verifying that:

1. the first derivative of MS at $x=0$ is zero, and
2. the second derivative of MS is negative for all $h>0$.

We begin by computing the first derivative using the Leibniz rule for differentiation under the integral:

$$MS_h'(f,x) = \left[\int_{x-h}^{x+h} q(x,t)f(t)dt - f(x)\right]' \tag{4}$$

$$\tag{5}$$

which yields:

$$MS_h'(f,x) = q(x,x+h)f(x+h)\,[x+h]' - q(x,x-h)f(x-h)\,[x+h]' + \int_{x-h}^{x+h}\frac{\partial q(x,t)}{\partial x}f(t)\,dt - f'(x). \tag{6}$$

By lemmas on kernel symmetry and derivatives, at $x=0$:

- $q(0,t)$ is even in t (Lemma 4),

- $\int\limits_{x-h}^{x+h} \frac{\partial q(x,t)}{\partial x} f(t)\, dt = 0$, (Lemma 6)
- f is even, so $f'(0) = 0$ (Lemma 1)

Thus:

$$\begin{aligned} MS_h'(f,0) &= q(0,h)f(h) - q(0,-h)f(-h) - f'(0) \\ &= q(0,h)f(h) - q(0,h)f(h) - 0 \\ &= 0 \end{aligned}$$

Therefore, the first derivative vanishes at $x = 0$.

We now compute the second derivative:

$$\begin{aligned} MS_h''(f,x) &= \left[\int\limits_{t\in B} q(x,t)f(t)\, dt - f(x) \right]'' \\ &= \left[q(x,x+h)f(x+h) - q(x,x-h)f(x-h) + \int\limits_{x-h}^{x+h} \frac{\partial q(x,t)}{\partial x} f(t)\, dt - f'(x) \right]' \\ &= [q(x,x+h)f(x+h)]' - [q(x,x-h)f(x-h)]' + \left[\int\limits_{x-h}^{x+h} \frac{\partial q(x,t)}{\partial x} f(t)\, dt \right]' - f''(x) \end{aligned}$$

By using integration by parts and applying the rule of the expression in parenthesis becomes:

$$\left[\int\limits_{x-h}^{x+h} \frac{\partial q(x,t)}{\partial x} f(t)\, dt \right]' = f(x+h)\frac{\partial q(x,x+h)}{\partial x} - f(x-h)\frac{\partial q(x,x-h)}{\partial x} - \int\limits_{x-h}^{x+h} \frac{\partial^2 q(x,t)}{\partial x^2} f(t)\, dt \tag{7}$$

If $x = 0$ and applying lemma 8, and also note that if $\left.\frac{\partial k(x,t)}{\partial x}\right|_{x=0} = 0$ (by lemma 7) then $\left.\frac{\partial^2 q(x,t)}{\partial x^2}\right|_{x=0} = 0$. Consequently, the expression 7 becomes:

$$\begin{aligned} \left[\int\limits_{x-h}^{x+h} \frac{\partial q(x,t)}{\partial x} f(t)\, dt \right]'_{x=0} &= f(h) \left.\frac{\partial q(x,x+h)}{\partial x}\right|_{x=0} - f(-h) \left.\frac{\partial q(x,x-h)}{\partial x}\right|_{x=0} - 0 \\ &= 0. \end{aligned}$$

Then we have that $f''(0) = 0$ by Sparrow condition, $q'(0,h) = 0$ (by Lemma 6), and also, $q'(0,t)$ and f' are odd functions. So, applying all of these on $MS_h''(f,0)$ we get:

$$\begin{aligned} MS_h''(f,0) &= q'(0,h)f(h) + q(0,h)f'(h) - q'(0,-h)f(-h) - q(0,-h)f'(-h) - f''(0) \\ &= 0f(h) + q(0,h)f'(h) - 0f(-h) - q(0,-h)f'(-h) - 0 \\ &= q(0,h)f'(h) - q(0,-h)f'(-h) \\ &= q(0,h)f'(h) - q(0,h)(-f'(h)) \\ &= 2q(0,h)f'(h). \end{aligned}$$

Now, we substitute f by G defined in Proposition 1 as a sum of Gaussian-sources. By lemma 6 we have that $q'(0,t) = 0$ for all $t \in B$, in particular for $t = \pm h$, then

$$MS''_h(G,0) = 2q(h)G'(h) = 2q(h)[g'_{-\sigma}h + g'_{\sigma}h].$$

Now, since q_h are weighted coefficients and thus always positive, we focus solely on the expression within the brackets.

$$\begin{aligned}
g'_{-\sigma}h + g'_{\sigma}h &= -\left(\frac{-\sigma+h}{\sigma}\right)e^{-\frac{1}{2}\left(\frac{-\sigma+h}{\sigma}\right)^2} - \left(\frac{\sigma+h}{\sigma}\right)e^{-\frac{1}{2}\left(\frac{\sigma+h}{\sigma}\right)^2} \\
&= -\left(-1+\frac{h}{\sigma}\right)e^{-\frac{1}{2}\left(-1+\frac{h}{\sigma}\right)^2} - \left(1+\frac{h}{\sigma}\right)e^{-\frac{1}{2}\left(1+\frac{h}{\sigma}\right)^2} \\
&= -\left[\left(-1+\frac{h}{\sigma}\right)e^{-\frac{1}{2}\left(-1+\frac{h}{\sigma}\right)^2} + \left(1+\frac{h}{\sigma}\right)e^{-\frac{1}{2}\left(1+\frac{h}{\sigma}\right)^2}\right].
\end{aligned}$$

Let $b = \dfrac{h}{\sigma}$ then

$$g'_{-\sigma}h + g'_{\sigma}h = -\left[(-1+b)\,e^{-\frac{1}{2}(-1+b)^2} + (1+b)\,e^{-\frac{1}{2}(1+b)^2}\right].$$

Working with the exponential expressions we get a common exponential factor:

$e^{-\frac{1}{2}(-1+b)^2} = e^{-\frac{1}{2}\left(-1-2b+b^2\right)} = e^{-\frac{1}{2}\left(1+b^2\right)}e^{b}$ and $e^{-\frac{1}{2}(1+b)^2} = e^{-\frac{1}{2}\left(1+2b+b^2\right)} = e^{-\frac{1}{2}\left(1+b^2\right)}e^{-b}$. So

$$\begin{aligned}
g'_{-\sigma}h + g'_{\sigma}h &= -\left[(-1+b)\,e^{-\frac{1}{2}\left(1+b^2\right)}e^{b} + (1+b)\,e^{-\frac{1}{2}\left(1+b^2\right)}e^{-b}\right] \\
&= -e^{-\frac{1}{2}\left(1+b^2\right)}\left[(-1+b)\,e^{b} + (1+b)\,e^{-b}\right]
\end{aligned}$$

Now we must prove that the expression in brackets is positive for $b > 0$.

$$\begin{aligned}
(-1+b)\,e^{b} + (1+b)\,e^{-b} &> 0 \\
-\left(e^{b} - e^{-b}\right) + b\left(e^{b} + e^{-b}\right) &> 0 \\
b &> \frac{e^{b} - e^{-b}}{e^{b} + e^{-b}} \\
0 &> \tanh(b) - b
\end{aligned}$$

Let $u(x) = \tanh(x) - x$ then

$$u'(x) = 1 - \tanh^2(x) - 1 = -\tanh^2(x) < 0$$

This means that u is strictly decreasing for all $x \in \mathbb{R}$, and therefore also injective. Since $u(0) = 0$, given its decreasing monotony and injectivity, then $u(x) < 0$ for all $x > 0$. So, $0 > \tanh(x) - x$ when $x > 0$.

Figure 3 illustrates the behavior of the function $u(x) = \tanh(x) - x$ and its derivative $u'(x)$ over the interval $x \in [-4, 4]$. The plot shows that $u(x)$ is a strictly decreasing function, as confirmed by the negative values of its derivative $u'(x) = -\tanh^2(x)$ throughout the domain. The function $u(x)$ intersects the horizontal axis at $x = 0$, and its decreasing trend implies that $u(x) < 0$ for all $x > 0$ and $u(x) > 0$ for all $x < 0$. This graphical representation supports the analytical conclusion that $\tanh(x) < x$ when $x > 0$, a property used in the proof of the negativity of the second derivative of the Mean Shift function at the midpoint between two Gaussian sources.

Since $b = \dfrac{h}{\sigma}$, then b is always positive (note that h is a radius and σ a standard deviation of Gaussian function) and satisfy the previous condition, then

$$b = \frac{h}{\sigma} > 0.$$

Since $q(h) > 0$ and $G'(h)$ for all h then always $MS''_h(f,0) < 0$. □

Figure 4 shows how $MS''(0)$ varies relative to h and σ for continuous data. In the context of the theoretical proof, the negativity of this function is a crucial indicator of the presence of a local maximum. Then a maximum is detected, which implies that MS can resolve two sources beyond the Sparrow limit (see Figure 1, in the main document for clarification). Some details are of great importance to highlight in Figure 4, such as the following:

1. Yellow Regions: These areas indicate that $MS''(0)$ is approximately zero. When the second derivative is close to zero, the intensity profile near the midpoint between the two emitters is relatively flat. This corresponds to scenarios where the MS algorithm provides little to no resolution enhancement because it doesn't significantly alter the shape of the intensity profile. Essentially, although values are negative, the maximum created by the MS is barely noticeable.

2. Green Regions: The green regions represent values of $MS''(0)$ that are moderately negative. This means that the Mean Shift algorithm is beginning to introduce some curvature in the intensity profile between the emitters. This indicates a moderate level of resolution enhancement. While the two sources aren't visible resolved, the profile is changing in a way that makes them more distinguishable.

3. Blue Regions: These areas correspond to strongly negative values of $MS''(0)$. A strongly negative second derivative indicates a well-defined curvature, meaning that the MS algorithm has successfully created a distinct peak or valley between the two emitters. This signifies that the MS has effectively enhanced resolution, clearly separating the two point sources. The darkest blue regions represent good combinations of h and σ where MS achieves a discernible curvature and clear resolution.

4. Red line: The red line has a slope of 1.9149. This line indicates an optimal relationship between h and σ when applying the MS. This optimal founding has direct relationship with Theorem 2 of discrete data.

Figure 4 provides strong visual evidence supporting Theorem 1, which states that for continuous one-dimensional data, the Mean Shift method always surpasses the Sparrow limit for any neighborhood with radius $h > 0$. The persistent negative values of $MS''(0)$ across a wide range of h and σ confirm that MS consistently generates a local maximum at $x = 0$, allowing it to resolve two sources that would otherwise be indistinguishable under the Sparrow criterion.

3 MS in 1-dimensional case for discrete data

In this section, we demonstrate how the Sparrow resolution limit can be surpassed by applying *MS* to 1-dimensional two-point sources. The proof is presented using 1-dimensional discrete data. We chose this formulation to simplify the analysis while retaining core features of the problem. On the other hand, discrete data is considered because of the pixeled nature of digital images, making this approach directly relevant for real applications (see Main text, Practical Experiments section).

First, we introduce the expression for *MS* and explain the concept of the neighborhood where *MS* is applied. We also establish conditions for the weights relative to each neighbor. Later, a sequence of lemmas (from Lemma 9 to Lemma 12 is presented to support Theorem 2. Notably, this theorem establishes a sufficient condition to surpass the Sparrow limit using MS under discrete sampling. However, this result is limited in scope; we examine other situations where the limit is surpassed and it does not satisfy the given Theorem 2.

3.1 Theory for 1-dimensional case for discrete data

Let x, x_i be points in $X \subset \mathbb{R}$ (X spatial space) with indexes i from the discrete set $N = [-n,n] \backslash \{0\} \subset \mathbb{Z}$, $n \in \mathbb{N}$ and f a differentiable function $f : Y \subset \mathbb{R} \to \mathbb{R}$ (Y range space). We define in X, a punctured neighborhood B centered at x with fixed radius $h > 0$, as:

$$B = B_h^*(x) = \left\{ x_i \in X | x_i = x + i\Delta x,\ i \in N \text{ and } \Delta x = \frac{h}{n} \right\}.$$

Note that, since h and n are greater than 0, then $\Delta x > 0$ represents the minimal separation among the x_i elements. Thus, the points $x_i \in B$ are the neighbors of x, and they are uniformly and symmetrically distributed around x by a distance $i\Delta x$ as shown in Figure 5.

The *MS* at x over the function f, denoted by $MS_h(f,x)$, measures the difference between the weighted average of the function f over the neighborhood B (known as the "sample mean") and the function f evaluated at the center x. The *MS* formula is given by expression 8:

$$MS_h(f,x) = \sum_{x_i \in B} \left(\frac{k_i(x)}{\sum\limits_{x_j \in B} k_j(x)} \right) f(x_i) - f(x) \tag{8}$$

where $k_i(x)$ is a positive 1D differentiable kernel profile $k : \mathbb{R} \to \mathbb{R}^+$, normalize by the sum over the neighborhood B.

The kernel depends on the maximum deviation between $f(x)$ and its neighbors $f(x_i)$, express as:

$$m_h(x) = \max_{i \in N} \{ |f(x_i) - f(x)| \},$$

and we define:

$$k_i(x) = k\left(\left| \frac{f(x_i) - f(x)}{m_h(x)} \right|^2 \right)$$

ensuring that $k_i(x)$ maps the values into the interval $[0,1]$.

Equally to the continuous case (see Subsection 2.1), the kernel k must satisfy:

- non-negativity
- non-increasing behavior
- piece-wise continuity
- differentiability
- integrability.

Common choices for k are Gaussian, Epanecknikov, Uniform[2,3].

Since the weighted coefficient $q_i(f,x)$ depends of f and x, the normalized weights are:

$$q_i = \frac{k_i(x)}{\sum\limits_j k_j(x)},$$

and the MS formula becomes:

$$MS_h(f,x) = \sum_i q_i(x) f(x_i) - f(x), \tag{9}$$

Furthermore, the q_i coefficients represent a convex linear combination of the neighbors' values of $f(x_i)$, meaning:

$$0 < q_i \leq 1, \quad \sum_i q_i = 1.$$

To compute MS at x, we assume the existence of at least one neighborhood x_i of x for all neighborhoods B. Moreover, since we operate over symmetric intervals, there always exist two neighbors in B. Finally, the following properties hold for all $i, j \in N$:

- $|i\Delta x| \leq h$ for all $i \in \mathbb{N}$, (10)
- $n\Delta x = h$, (11)
- $0 < i < j \Rightarrow q_i > q_j$, (12)
- $i < j < 0 \Rightarrow q_i < q_j$. (13)

These properties follow directly from the spatial decay of the normalized weights q_i with respect to the distance between the evaluation point x and its neighboring samples $x_i = x + i\Delta x$. Since the kernel $k_i(x)$ is assumed to be a non-increasing function of $|x - x_i|$, closer points contribute more strongly to the computation of $MS_h(f,x)$, leading to $q_i > q_j$ whenever $0 < i < j$. The neighborhood size is bounded by the radius h, such that $|i\Delta x| \leq h$ and $n\Delta x = h$, ensuring symmetric sampling around x. Consequently, the coefficients q_i decrease monotonically with increasing index i, reflecting the attenuation of the kernel with distance and preserving the symmetry of the weights respect the central point.

For simplicity, we adopt the compact notation 9 for the formula of MS in the rest of this section.

3.2 Lemmas and Theorem 2 for Discrete Data

We now establish key properties of the MS operator for even, differentiable functions.

Lemma 9 (MS symmetric coeffitients)**.** *Let $f(x)$ be an even function and $h > 0$ then coefficient q_i satisfies the following properties:*

- $q_{-i}(-x) = q_i(x)$,
- $q_{-i}(x) = q_i(-x)$.

Proof. For the first property, assume f is even. Then:

$$\begin{aligned} f(-x + (-i\Delta x)) - f(-x) &= f(-x - i\Delta x) - f(-x) \\ &= f(-(x + i\Delta x)) - f(-x) \\ &= f(x + i\Delta x) - f(x). \end{aligned}$$

Consequently,

$$q_{-i}(-x) = \frac{k\left(\left|\frac{f(-x+(-i\Delta x))-f(-x)}{\max\limits_{i\in N}\{|f(-x+(-i\Delta x))-f(-x)|\}}\right|^2\right)}{\sum\limits_i k\left(\left|\frac{f(-x+(-i\Delta x))-f(-x)}{\max\limits_{i\in N}\{|f(-x+(-i\Delta x))-f(-x)|\}}\right|^2\right)} = \frac{k\left(\left|\frac{f(x+i\Delta x)-f(x)}{m_h(x)}\right|^2\right)}{\sum\limits_i k\left(\left|\frac{f(x+i\Delta x)-f(x)}{\max\limits_{i\in N}\{|f(x+i\Delta x)-f(x)}\right|^2\right)} = q_i(x).$$

This proves the first property.

For the second property, we observe

$$q_{-i}(x) = q_{-i}(-(-x)) = q_i(-x),$$

using the first property in the last equality. □

See figure 6 for the distribution of q_i around 0 for an even function. We remark that if $x = 0$, these imply $q_{-i}(0) = q_i(0)$, which is used later.

Lemma 10 (MS eveness)**.** *Let $f(x)$ be a given even function and $h > 0$. Then the function $MS(x)$ is also even.*

Proof. To demonstrate that MS is even, we must prove that $MS(-x) = MS(x)$ for all values of x.

First, let us recall the definition of $MS(x)$:

$$MS(x) = \sum_i q_i(x)(f(x+i\Delta x) - f(x)) = q_{-n}(x)(f(x-n\Delta x)) + \ldots + q_n(x)(f(x+n\Delta x) - f(x)).$$

For $MS(-x)$, we have:

$$MS(-x) = \sum_i q_i(-x)(f(-x+i\Delta x) - f(-x)).$$

Since $f(x)$ is even, we can apply the property $f(-x) = f(x)$. Additionally, we can rewrite:

$$f(-x+i\Delta x) = f(-(x-i\Delta x)) = f(x-i\Delta x)$$

Furthermore, from Lemma 9 property 2, we know that $q_i(-x) = q_{-i}(x)$. Substituting these into our expression:

$$\begin{aligned}
MS(-x) &= \sum_i q_i(-x)(f(-x+i\Delta x) - f(-x)) \\
&= \sum_i q_i(-x)(f(x-i\Delta x) - f(x)) \\
&= \sum_i q_{-i}(x)(f(x-i\Delta x) - f(x))
\end{aligned}$$

By making the substitution $j = -i$ $(i = -n, \ldots, n)$, we can rewrite the summation:

$$\begin{aligned}
MS(-x) &= \sum_{-j} q_j(x)(f(x+j\Delta x) - f(x)) \\
&= q_{-(-n)}(x)(f(x-(-n)\Delta x)) + \ldots + q_{(-n)}(x)(f(x+(-n)\Delta x) - f(x)) \\
&= q_n(x)(f(x+n\Delta x)) + \ldots + q_{-n}(x)(f(x-n\Delta x) - f(x)) \\
&= q_{-n}(x)(f(x-n\Delta x)) + \ldots + q_n(x)(f(x+n\Delta x) - f(x)) \\
&= \sum_j q_j(x)(f(x+j\Delta x) - f(x)) \\
&= MS(x)
\end{aligned}$$

Therefore, $MS(-x) = MS(x)$, proving that $MS(x)$ is an even function. □

Lemma 11 (Derivative coeffitient simplification). *Let $f(x)$ be a differentiable even function and $h > 0$ then coefficient q_i satisfies the following property:*

$$\left[\sum_i q_i f\right]' = \sum_i q_i f'$$

Proof.

$$\left[\sum_i q_i(x) f(x_i)\right]' = \sum_i [q_i(x) f(x_i)]' = \sum_i q_i'(x) f(x_i) + \sum_i q_i(x) f'(x_i).$$

Now we will prove that the first summand is equal to zero.

$$\sum_i q_i'(x) f(x_i) = \sum_i \left(\frac{k_i(x)}{\sum_j k_j(x)}\right)' f(x_i).$$

Applying the derivative of the quotient we have:

$$\left(\frac{k_i(x)}{\sum_j k_j(x)}\right)' = \frac{k_i'(x)\left(\sum_j k_j(x)\right) - k_i'(x)\left(\sum_j k_j'(x)\right)}{\left(\sum_j k_j(x)\right)^2},$$

Therefore:

$$
\begin{aligned}
\sum_i q_i'(x) f(x_i) &= \sum_i \frac{k_i'(x)\left(\sum_j k_j(x)\right) - k_i(x)\left(\sum_j k_j'(x)\right)}{\left(\sum_j k_j(x)\right)^2} f(x_i) \\
&= \frac{1}{\left(\sum_j k_j(x)\right)^2} \sum_i \left(k_i'(x)\left(\sum_j k_j(x)\right) - k_i(x)\left(\sum_j k_j'(x)\right)\right) f(x_i) \\
&= \frac{1}{\left(\sum_j k_j(x)\right)^2} \left[\sum_i k_i'(x)\left(\sum_j k_j(x)\right) f(x_i) - \sum_i k_i(x)\left(\sum_j k_j'(x)\right) f(x_i)\right] \\
&= \frac{1}{\left(\sum_j k_j(x)\right)^2} \left[\sum_i k_i'(x) \sum_j k_j(x) f(x_i) - \sum_j k_j'(x) \sum_i k_i(x) f(x_i)\right]
\end{aligned}
$$

Since the indexes i and j move over the same set of values, they take the same of values (except in the order). Then, there exist an arrangement of j values in N such as:

$$k_i'(x)\sum_j k_j(x) = k_j'(x)\sum_i k_i(x), \ \ \forall i \in N,$$

and then

$$\sum_i k_i'(x)\sum_j k_j(x) - \sum_j k_j'(x)\sum_i k_i(x) = 0$$

Then the expression in parentheses is equal to zero and consequently:

$$\sum_i q_i'(x) f(x_i) = 0.$$

□

Lemma 12. *Let $f(x)$ be a differentiable even function. Then*

$$\frac{d^n MS}{dx^n}(f,x) = MS\left(\frac{d^n f}{dx^n}, x\right).$$

Proof. This proof uses the result of Lemma 11 and the Principle of Mathematical Induction. Note that for Lemma 11, there are no changes to the consecutive derivatives for the convex combination MS coefficients.

The MS formula is given by:

$$MS(f,x) = \sum_i q_i f(x + i\Delta x) - f(x).$$

Applying the first derivative, we get:

$$
\begin{aligned}
\frac{dMS}{dx}(f,x) &= \left(\sum_i q_i f(x+i\Delta x) - f(x)\right)' \\
&= \left(\sum_i q_i f(x+i\Delta x)\right)' - f'(x) \quad \text{Applying lemma 11 to first term} \\
&= \sum_i q_i f'(x+i\Delta x) - f'(x) \\
&= MS\left(f', x\right).
\end{aligned}
$$

Now applying the second derivative to MS we obtain:

$$\begin{aligned}
\frac{d^2MS}{dx^2}(f,x) &= \frac{dMS}{dx}\left(f',x\right) \\
&= \left(\sum_i q_i f'(x+i\Delta x) - f'(x)\right)' \\
&= \left(\sum_i q_i f'(x+i\Delta x)\right)' - f''(x) \\
&= \sum_i q_i f''(x+i\Delta x) - f''(x) \\
&= MS\left(f'',x\right).
\end{aligned}$$

Note that the q_i coefficients remain the same for the each derivative.

Now, suppose that

$$\frac{d^nMS}{dx^n}(f,x) = MS\left(\frac{d^nf}{dx^n},x\right),$$

we will prove that

$$\frac{d^{n+1}MS}{dx^{n+1}}(f,x) = MS\left(\frac{d^{n+1}f}{dx^{n+1}},x\right).$$

Starting with:

$$\begin{aligned}
\frac{d^{n+1}MS}{dx^{n+1}}(f,x) &= \frac{dMS}{dx}\left(\frac{d^nf}{dx^n},x\right) \\
&= \left(\sum_i q_i \frac{d^nf}{dx^n}(x+i\Delta x) - \frac{d^nf}{dx^n}(x)\right)' \\
&= \left(\sum_i q_i \frac{d^nf}{dx^n}(x+i\Delta x)\right)' - \frac{d^{n+1}f}{dx^{n+1}}(x) \\
&= \sum_i q_i \frac{d^{n+1}f}{dx^{n+1}}(x+i\Delta x) - \frac{d^{n+1}f}{dx^{n+1}}(x) \\
&= MS\left(\frac{d^{n+1}f}{d^{n+1}x},x\right).
\end{aligned}$$

This completes the proof by mathematical induction. □

Theorem 2. *Let two point sources be described by Gaussian distributions satisfying the Sparrow condition. Then, the MS at the midpoint between the two sources is a local maximum if the neighborhood radius satisfies $h < 1.915\sigma$, where σ is the standard deviation of the Gaussian.*

Proof. We begin by proving that the first derivative of MS vanishes at the midpoint ($x = 0$), independently of σ. Let $G_a(x) = g_{-a}(x) + g_a(x)$, where $g_a(x) = e^{-\frac{(x-a)^2}{2\sigma^2}}$.

The MS of G_a is:

$$MS(G_a,x) = \sum_i q_i G_a(x+i\Delta x) - G_a(x).$$

Expanding:

$$MS(G_a,x) = \sum_i q_i \left(g_{-a}(x+i\Delta x) + g_a(x+i\Delta x)\right) - \left(g_{-a}(x) + g_a(x)\right).$$

By Lemmas 1, 2, and 10, and the symmetry of the Gaussian, we conclude:

$$\left.\frac{d}{dx}MS(G_a,x)\right|_{x=0} = 0.$$

In particular, for $a = \sigma$,

$$\frac{d}{dx} MS(G_\sigma, x)\Big|_{x=0} = 0.$$

Now, computing the second derivative applying Lemma 12:

$$\frac{d^2}{dx^2} MS(G_a, x) = MS(G''_a, x).$$

And using the Sparrow condition $G''_\sigma(0) = 0$, we find:

$$\begin{aligned}
\frac{d^2}{dx^2} MS(G_\sigma, x)\Big|_{x=0} &= 2\sum_{i>0} q_i G''_\sigma(i\Delta x \\
&= 2\sum_{i>0} q_i \left[\left(\frac{(-\sigma + i\Delta x)^2}{\sigma^4} - \frac{1}{\sigma^2}\right) e^{-\frac{1}{2}\frac{(-\sigma+i\Delta x)^2}{\sigma^2}} +) \left(\frac{(\sigma + i\Delta x)^2}{\sigma^4} - \frac{1}{\sigma^2}\right) e^{-\frac{1}{2}\frac{(\sigma+i\Delta x)^2}{\sigma^2}}\right] \\
&= \frac{2}{\sigma^2}\sum_{i>0} q_i \left[\left(\left(-1 + \frac{i\Delta x}{\sigma}\right)^2 - 1\right) e^{-\frac{1}{2}\left(-1+\frac{i\Delta x}{\sigma}\right)^2} + \left(\left(1 + \frac{i\Delta x}{\sigma}\right)^2 - 1\right) e^{-\frac{1}{2}\left(1+\frac{i\Delta x}{\sigma}\right)^2}\right].
\end{aligned}$$

Let $b_i = \frac{i\Delta x}{\sigma}$ (since $i\Delta x \le h$, the largest b_i is h/σ) and making some factors gives:

$$\frac{d^2}{dx^2} MS(G_\sigma, x)\Big|_{x=0} = \frac{2}{\sigma^2}\sum_{i>0} q_i b_i e^{-\frac{1}{2}(1+b_i^2)} \left((-2+b_i)e^{b_i} + (2+b_i)e^{-b_i}\right). \tag{14}$$

Our goal is to find out when expression 14 is negative. Since q_i, b_i, $e^{-\frac{1}{2}(1+b_i^2)}$ are all greater than zero, this reduces to analyze:

$$(-2+b_i)e^{b_i} + (2+b_i)e^{-b_i} < 0. \tag{15}$$

Thus, this reduces to

$$\begin{aligned}
-2\sinh(b_i) + b_i\cosh(b_i) &< 0 && (16)\\
2\tanh(b) - b &> 0. && (17)
\end{aligned}$$

The function $u(b) = 2\tanh(b) - b$ has a unique positive root at $b \approx 1.915$ (find it numerically, see Appendix 4A). So, with the property 10, we conclude to the following condition:

$$b_i = \frac{i\Delta x}{\sigma} < 1.915 \Rightarrow h < 1.915\sigma.$$

Hence, when $h < 1.915\sigma$, the second derivative is negative, and MS has a local maximum at the midpoint. □

In the previous theorem, if h satisfies $0 < h < 1.915\sigma$, then every neighborhood radius up to that value will also satisfy the condition. To ensure that the MS algorithm can resolve two Gaussian point sources beyond the Sparrow limit, the kernel radius should be less than approximately 1.915σ; in practice, it is convenient and effective to round this value to 2σ. This value is conveniently given by the discrete pixel grid in digital imaging.

Now, we investigate the behavior of $\frac{d^2MS}{dx^2}(0)$ for $h > 1.915\sigma$. The derivation starts with the equation 14:

$$\frac{d^2MS}{dx^2}(0) = \frac{2}{\sigma^2}\sum_{i>0} q_i b_i e^{-\frac{1}{2}\left(1+b_i^2\right)} \left[-(2-b_i)\,e^{b_i} + (2+b_i)\,e^{-b_i}\right]$$

and is subsequently simplified by factoring and combining terms, ultimately focusing on the key expression:

$$\sum_{i>0}\left[\frac{b_i\left((2-b_i)\,e^{b_i} - (2+b_i)\,e^{-b_i}\right)}{e^{\frac{b_i^2}{2}}}\right].$$

This simplification is justified by the positivity and monotonicity of q_i, allowing the analysis to focus on the behavior of the summand itself.

Figure 7 provides a comprehensive analysis of how the cumulative second derivative G'' varies as a function of the parameter h, which determines the radius of the MS neighborhood and directly influences the function's shape and behavior.

Figures from 7b.i to 7b.iv provide a more detailed analysis by displaying G' as a function of the distance r for different values of h. Each plot corresponds to a different fixed h value ($h = 2,4,6,8$), and multiple curves represent different numbers of points ($N = 2,3,4,5,8,16$). The x-axis denotes the distance r, while the y-axis represents the cumulative function G'. In panel 7b.i ($h = 2$), the blue curve (representing $N = 2$) exhibits the steepest descent, reaching a deep minimum before stabilizing. As the number of points increases, the function flattens out, reducing the depth of the minimum. Panel 7b.ii ($h = 4$) follows a similar trend, but the minimum of G' shifts to a larger r value, and the overall variation decreases in magnitude. The curves with higher N remain less pronounced. The panel 7 b.iii ($h = 6$) and panel 7b.iv ($h = 8$) shows a continued trend in which the function flattens further and the minimum becomes less deep, indicating a reduction in variation as h increases. The shaded regions in each panel highlight areas where G' is negative. The presence and depth of these negative regions depend both on h and N. The vertical dashed lines mark reference points related to the minimum or characteristic distances where significant changes in G' occur. This figure provides insight into how h and N affect the behavior of G', showing that as h increases, the function becomes smoother and the sharp variations diminish.

Figure 8 presents a series of heatmaps that illustrate the sign of the second derivative of the Mean Shift function, denoted as $sg(MS''(0))$, varying the number of points in the neighborhood B_h. Each subplot corresponds to a different number of points uniformly distributed in B_h, ranging from 2 to 129, as indicated by the labels above each heatmap. The x-axis in each plot represents the parameter σ, while the y-axis represents h, both of which influence the behavior of $MS''(0)$. The color scheme differentiates regions where $MS''(0) > 0$ (yellow), and $MS''(0) < 0$ (blue). The yellow region generally appears above a diagonal boundary, indicating a transition in the function's behavior as h and σ vary. As the number of points increases, the yellow region, where $MS''(0) > 0$, gradually shrinks, suggesting that the conditions required to maintain a positive second derivative become more restrictive for larger datasets. The results provide insight into how the MS function behaves under different discretizations and smoothing conditions, particularly as the dataset size increases.

Figure panels 9a-c collectively explore how different parameters in the $MSSR$ algorithm affect the dip value —a proxy for 2D emitter separability— and thus the resolution enhancement. Panel 9a presents a series of heatmaps where the vertical axis corresponds to the spatial neighborhood radius h, and the horizontal axis to the Gaussian kernel width σ. Each column corresponds to a fixed image magnification amp and the rows vary by $MSSR$ $order$. The color encodes the dip value between two emitters located at the Sparrow limit. The black diagonal line, $h = 1.915\sigma$, indicates the theoretical threshold below which $MSSR$ is mathematically proven to resolve emitters beyond the Sparrow limit (as shown in Theorem 2). The dip values decrease noticeably when $MSSR$ order increases, especially for moderate amplitudes (e.g., amp = 3, 5), and when h/σ is large. For example, at order 3 and higher amplitudes, most of the heatmap shifts toward lower dip values (blue), indicating increased resolution. The dip value is directly related to the second derivative of the Mean Shift function: a more negative second derivative at the midpoint indicates a deeper dip, which corresponds to better separation of the emitters. Therefore, minimizing the dip through tuning of $MSSR$ parameters effectively enhances spatial resolution.

Panel 9b provides selected $MSSR$ results demonstrating the effects of varying three key parameters—amplitude, $MSSR$ order, and the h/σ ratio—on dip. The top row (varying amplitude, fixed order and ratio) shows that even as emitter brightness increases, the dip only slightly changes, indicating that resolution is relatively stable against amplitude in high SNR conditions. The middle row shows that increasing the MSSR order significantly reduces dip (from 0.784 at order 0 to just 0.001 at order 3), highlighting the strong impact of algorithmic depth. The bottom row shows results at fixed order and amplitude, with increasing h/σ ratios. As this ratio increases, the dip reduces, confirming the theoretical result that larger neighborhoods enhance resolution. Panel 9c summarizes these trends conceptually, illustrating that dip and resolution are inversely related: as dip decreases with optimized parameters, resolution improves. This visual encapsulation reinforces the idea that proper tuning of $MSSR$ parameters—especially the h/σ ratio and order—is crucial for super-resolving emitters at or below the Sparrow limit.

4 Appendix

Appendix A

To solve inequality 17 given by $2\ \tanh(b) - b > 0$, we need to know what values of b satisfy the condition. First, we will show that there exists a positive root of b. We then apply a numerical approach using Newton's method to determine the threshold value x_h.

Now we investigate the shape of $u(x)$. The maximum of the function $u(x) = 2\ \tanh(x) - x$ for $x > 0$ is given by

$$u'(x) = [2\ \tanh(x) - x]' = 1 - 2\ \tanh^2(x) = 0$$

$$\tanh(x) = \pm\frac{1}{\sqrt{(2)}}$$

So, $x_p = \text{atanh}\left(\frac{1}{\sqrt{(2)}}\right) \approx 0.8814$, because we are interested in non-negative values.

The second derivative is given by:

$$u''(x) = 4 * tanh(x) * (tanh(x)^2 - 1).$$

Since

$$x > 0 \Rightarrow 0 < tanh(x) < 1 \Rightarrow tanh(x)^2 - 1 < 0 \Rightarrow u(x)'' < 0.$$

This implies that $u(x)$ is always concave for $x > 0$ and the point x_p is the only local maximum for this infinite interval. The fact that

$$u(0) = 2\ \tanh(0)) - 0 = 0,$$

means that $u(x)$ is strictly increasing in $(0, x_p)$, moreover this function is positive for this interval. On the other hand, since x_p is a local maximum, then $u(x)$ is strictly decreasing for all $x > x_p$, in other words, is an injective function if $x \in [x_p, +\infty]$.

So,

$$\begin{aligned} u(x_p) &= 2\ \tanh(x_p)) - x_p \\ &= 2\frac{1}{\sqrt{(2)}} - \text{atanh}\left(\frac{1}{\sqrt{(2)}}\right) \\ &= \sqrt{(2)} - \text{atanh}\left(\frac{1}{\sqrt{(2)}}\right). \end{aligned}$$

Since $\text{atanh}\left(\frac{1}{\sqrt{(2)}}\right) < 1$ then $u(x_p) > 0$. On the other hand,

$$u(5) = 2\ \tanh(5)) - 5 \approx -3.0002 < 0$$

and $u(x)$ is a continuous function in $\mathbb{R}$. For Bolzano's theorem of continuous functions and the injectivity of u, there exists a unique point x_h such as satisfies that $x_h > x_p$ and $u(x_h) = 0$.

At this point, we apply Newton's method for finding roots. Note that by the selection of an initial approximation $x_0 \geq 3$, we guarantee the convergence of this procedure (see Figure 10). The solution is given by $x_h \approx 1.9150$ with an error on the order of 10^{-3} with four iterations. This value has high precision. We have demonstrated that $u(x) > 0$ if $x \in (0, x_h)$ and $u(x) < 0$ for $x > x_h$.

Appendix B

Figure 11 illustrates how inappropriate parameter choices in the *MSSR* algorithm can lead to artifacts or loss of resolution. In this analysis, the parameters evaluated were the spatial kernel radius (h, in pixels) and the image up-sampling factor (amp). For all cases, the *MSSR* order was fixed at 0.

Figure 11a shows an Argolight line pattern with the analyzed region highlighted. This segment contains two vertical lines separated approximately at the Sparrow limit (120 nm, see Main Article, section 2.2.3). In the raw image, the lines merge into a single broad maximum, characteristic of an under-resolved system. The inset provides a magnified view of the selected region, further illustrating how the two lines blur into a single unresolved feature. The theoretical expectation is that *MSSR* should restore two distinct lobes with a central dip, provided that parameters are chosen appropriately.

Figure 11b displays the case of $amp = 1$. For a small kernel radius ($h_s = 5$, blue trace), two well-defined peaks are recovered, consistent with the expected Sparrow-limit separation. This works because the peak-to-peak distance is about 3–4 pixels, and a kernel of radius 5 appropriately samples the local intensity distribution (covering ∼9–10 pixels span). In contrast, larger kernels ($h_s = 25$ and $h_s = 100$) oversmooth the intensity distribution. As a consequence, the central dip is lost and the two emitters collapse into a single blurred maximum, leading again to under-resolution.

Figure 11c shows the case of $amp = 5$, where the system becomes more sensitive to parameter tuning. With a very small kernel ($h_s = 5$), the algorithm amplifies high-frequency noise, producing spurious oscillations and false side peaks. This artificial splitting represents over-resolution and is a typical *MSSR* artifact for small kernel radii. With a very large kernel ($h_s = 100$), the opposite occurs: the profile is over-smoothed, the central dip vanishes, and the two sources are once again indistinguishable. Interestingly, an intermediate kernel ($h_s = 25$) provides the most faithful recovery. Here the total emitter separation is ∼50 pixels, and the kernel radius corresponds to roughly half of that span, allowing *MSSR* to recover a clear dip with two symmetric peaks.

Overall, this figure demonstrates that kernel size and magnification interact nonlinearly in *MSSR*. Small kernels at high magnification risk generating false features, while excessively large kernels suppress real spatial detail. Only a carefully tuned intermediate kernel radius recovers the correct Sparrow-limit configuration without artifacts. This emphasizes that *MSSR* can either improve or distort resolution depending on parameter selection, and that rigorous calibration is essential to avoid over-interpretation.

Appendix C

Table 1. Parameters used for Comparative Performance of Super-Resolution Methods at the Sparrow Limit.

Method	Parameters
Lucy-Richardson Deconvolution (*LRD*)	$PSF = 20$, $iter = 10$
Deblurring by Pixel Reassignment (*DPR*)	$PSF = 20$, $gain = 3$, $background = 30$
MeanShift SuperResolution (*MSSR*)	$PSF = 1$, $amp = 1$, $order = 0$
Superposition of point Sources (*SUPPOSe*)	$sources = 2$; $population = 1000$; $generations = 100$; $mutation_{rate} = .1$

References

1. Hecht, J. *Optics: Light for a New Age* (Jeff Hecht, 1987).
2. Chris, T., Karamchandani, S. & Biradar, T. Review of mean shift algorithm and its improvements. *Int. J. Comput. Appl.* **5**, 25–30 (2015).
3. Carreira-Perpinán, M. A. A review of mean-shift algorithms for clustering. *arXiv preprint arXiv:1503.00687* (2015).

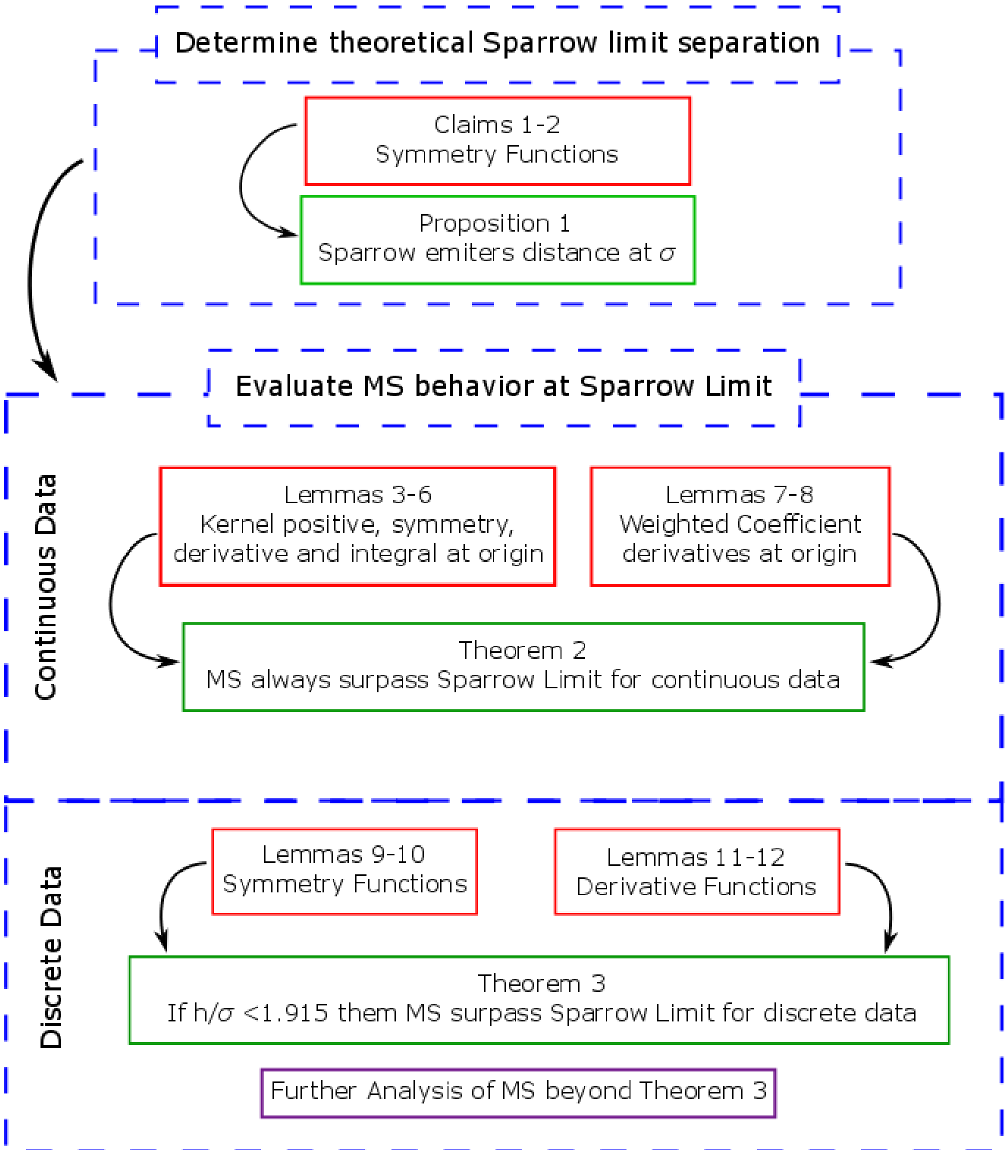

Figure 1. Logical flowchart of lemmas and theorems for *MS* resolution analysis at the Sparrow limit. Lemmas 1-2 establish symmetry properties of even functions and yield Theorem 1, which shows that two Gaussian emitters reach the Sparrow resolution at a separation of 2σ. Building on this, Lemmas 3–6 (symmetry and derivative conditions) underpin Theorem 2 for minimum separation in discrete data, while Lemmas 7–10 (kernel positivity, symmetry, derivative and integral behavior at the origin) together with Lemma 11 and 12 (zero derivatives of the coefficient at the origin) support Theorem 3 for the continuous-data case.

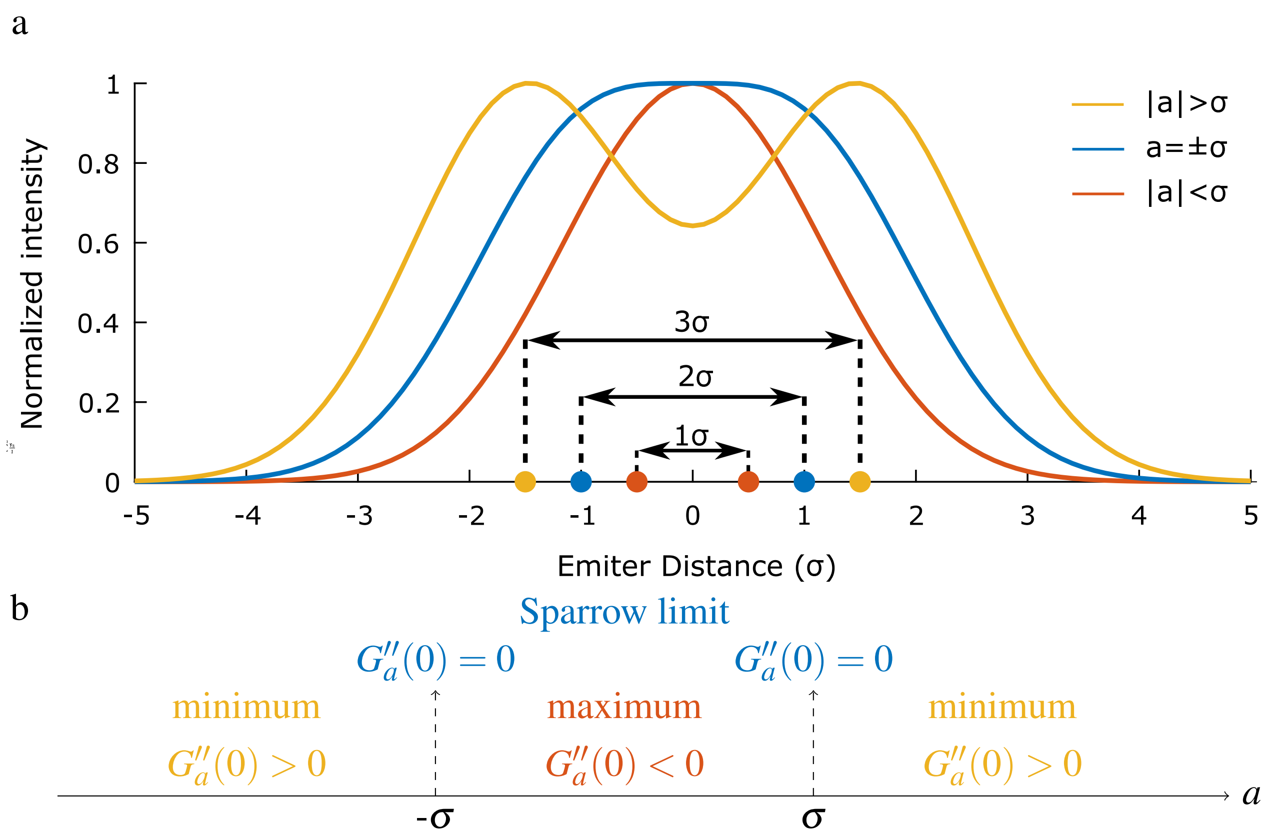

Figure 2. Influence of Gaussian emitters separation on intensity profiles and the second derivative at the origin. (a) Normalized intensity profiles resulting from the sum of two Gaussian emitters at various separations. Yellow: resolved ($|a| > \sigma$); blue: Sparrow limit ($|a| = \sigma$); red: unresolved ($|a| < \sigma$). (b) Classification of midpoint curvature based on $G''_a(0)$, showing conditions for minimum, maximum, and inflexion point behavior.

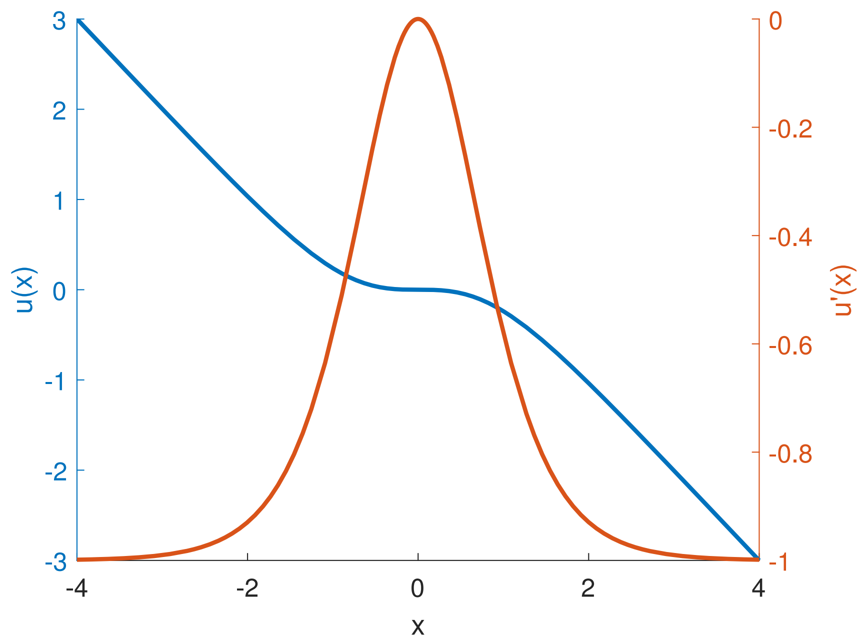

Figure 3. Plots of the function $u(x) = tanh(x) - x$ (blue curve, left axis) and its derivative $u'(x) = -tanh^2(x)$ (orange curve, right axis), shown over the interval $x \in [-4, 4]$.

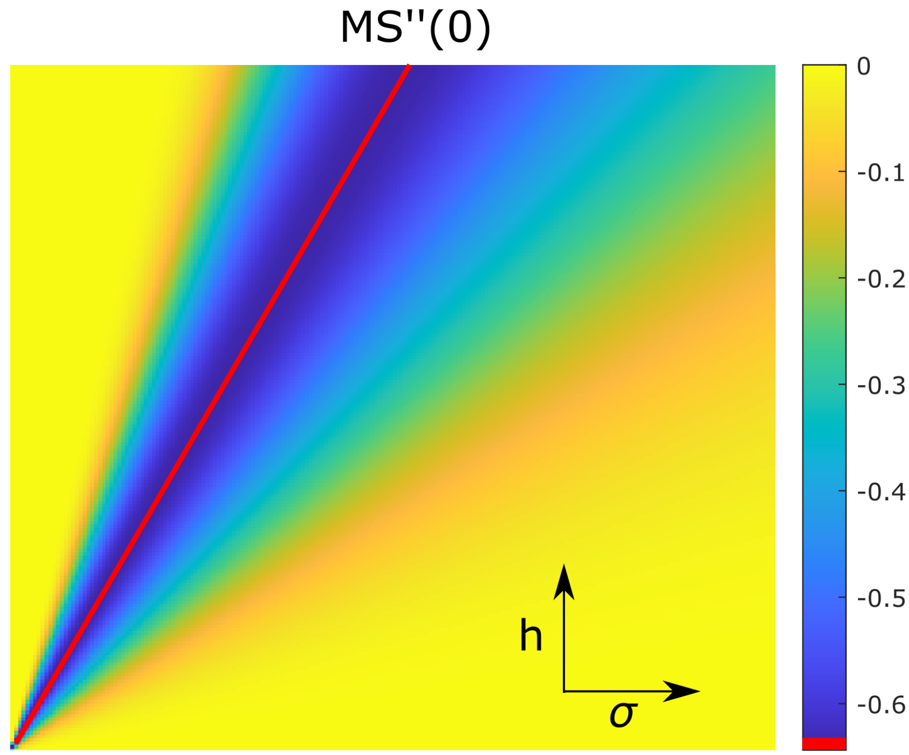

Figure 4. Value of the second derivative $MS''(0)$ for continuous data as a function of h and σ. Darker regions (blue) indicate stronger curvature and better resolution enhancement by MS. Bright regions (yellow) represent values near zero. The diagonal red line marks the empirical threshold derived in Theorem 2.

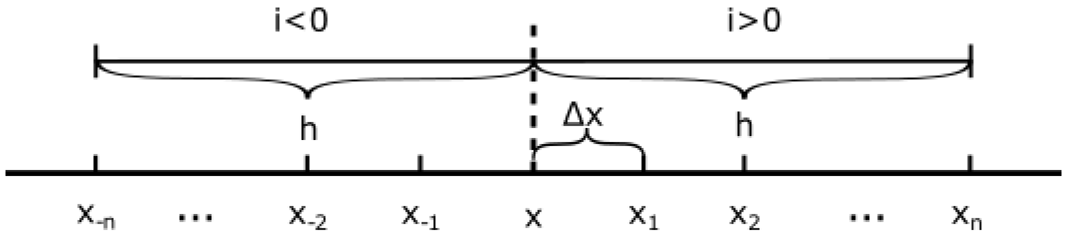

Figure 5. Schematic representation of the distribution of points x_i within the neighborhood B centered at x. The points are uniformly spaced by a distance Δx and symmetrically distributed with respect to $x = 0$. The total neighborhood radius is h, such that the furthest neighbor satisfies $|x_i| < h$.

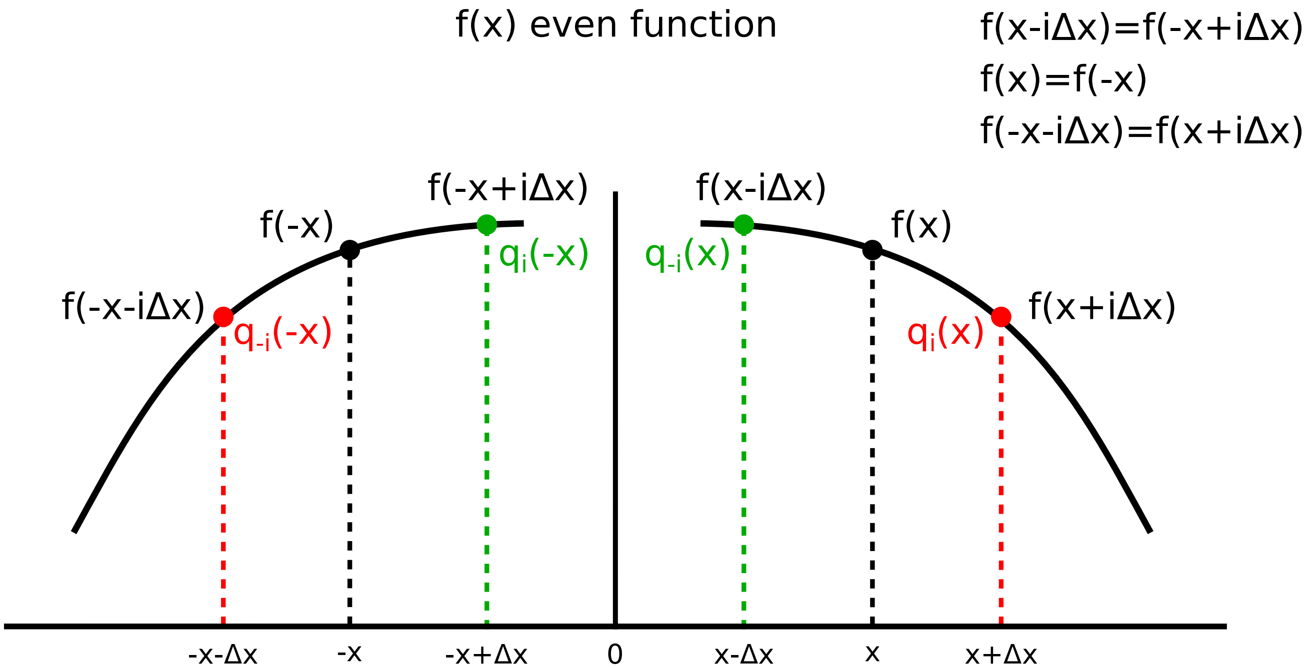

Figure 6. Symmetry of f(x) and associated weight coefficients illustrates even function properties used in Lemma 9. Highlighting the symmetry of its values and the corresponding coefficients $q_i(x)$ and $q_i(-x)$ at symmetric points. The diagram demonstrates that for $\Delta x > 0$, $f(x \pm i\Delta x)$ and $f(-x \pm i\Delta x)$ are mirror images, and the associated coefficients satisfy $q_i(x) = q_i(-x)$ and $q_{-i}(x) = q_i(-x)$.

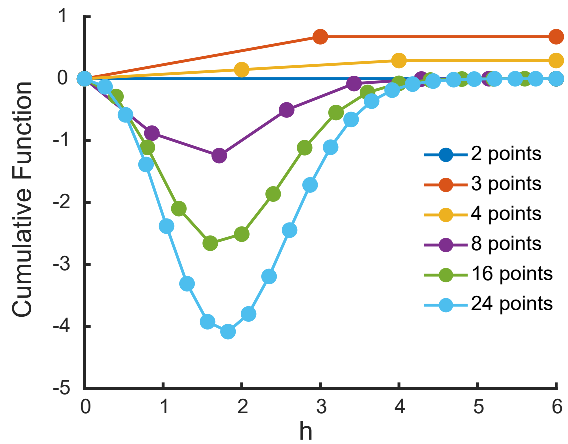

Figure 7. Cumulative derivative function G'; as a function of spatial distance for different kernel radii h (panels i-iv: $h = 2, 4, 6, 8$) and point counts $N = 2 - 16$ Shaded regions indicate intervals where $G'' < 0$, corresponding to local maxima in the cumulative response. As the kernel radius h increases, G'' flattens, reducing resolution.

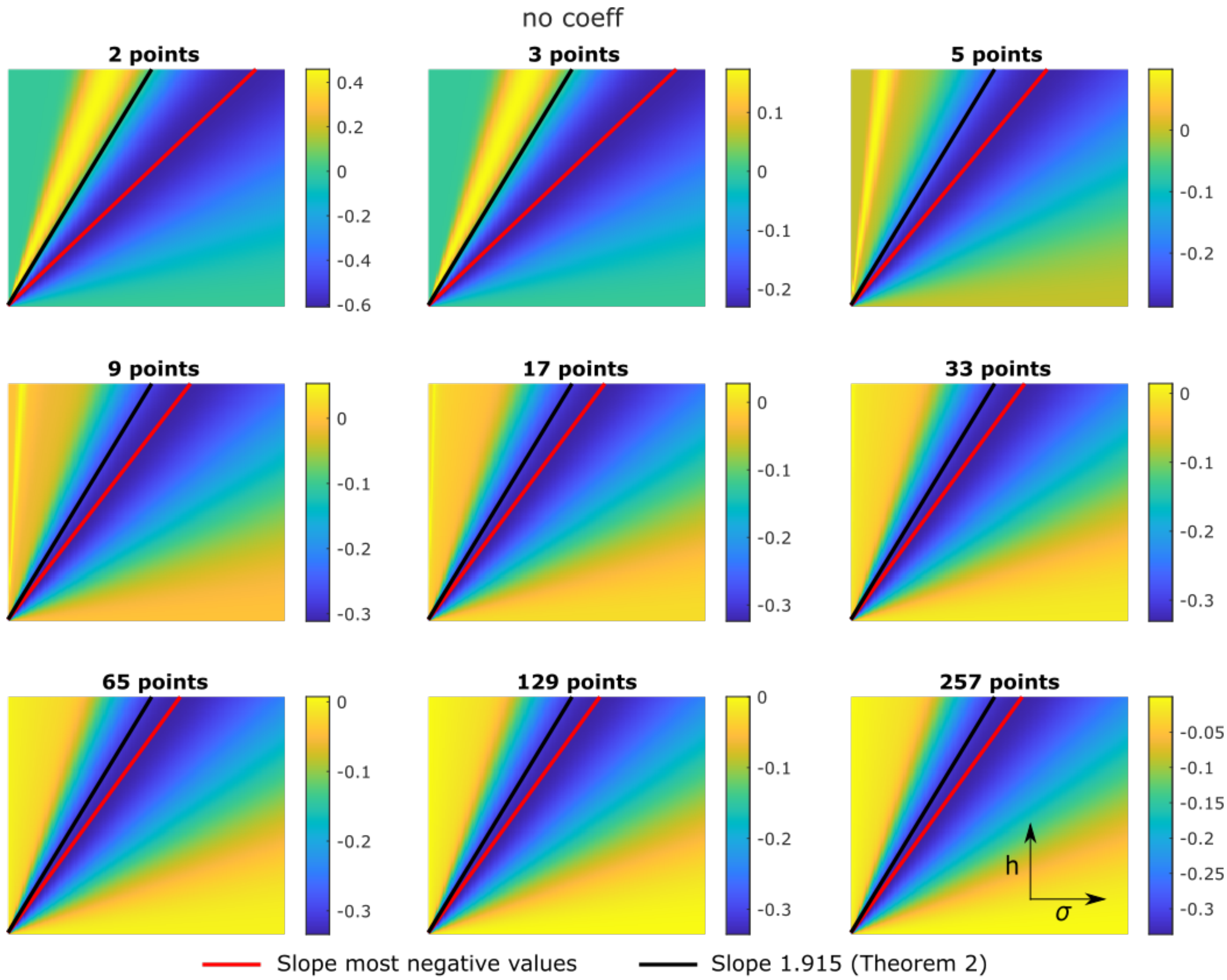

Figure 8. Heatmaps showing the sign of the second derivative of the MS function at $x = 0$, $sign(MS''(0))$, for varying neighborhood sizes and standard deviations. Yellow: $MS''(0) > 0$; cyan-emerald: sign transition region; blue: $MS''(0) < 0$. Increasing the number of discrete points shrinks the yellow region, suggesting convergence to the continuous case behavior.

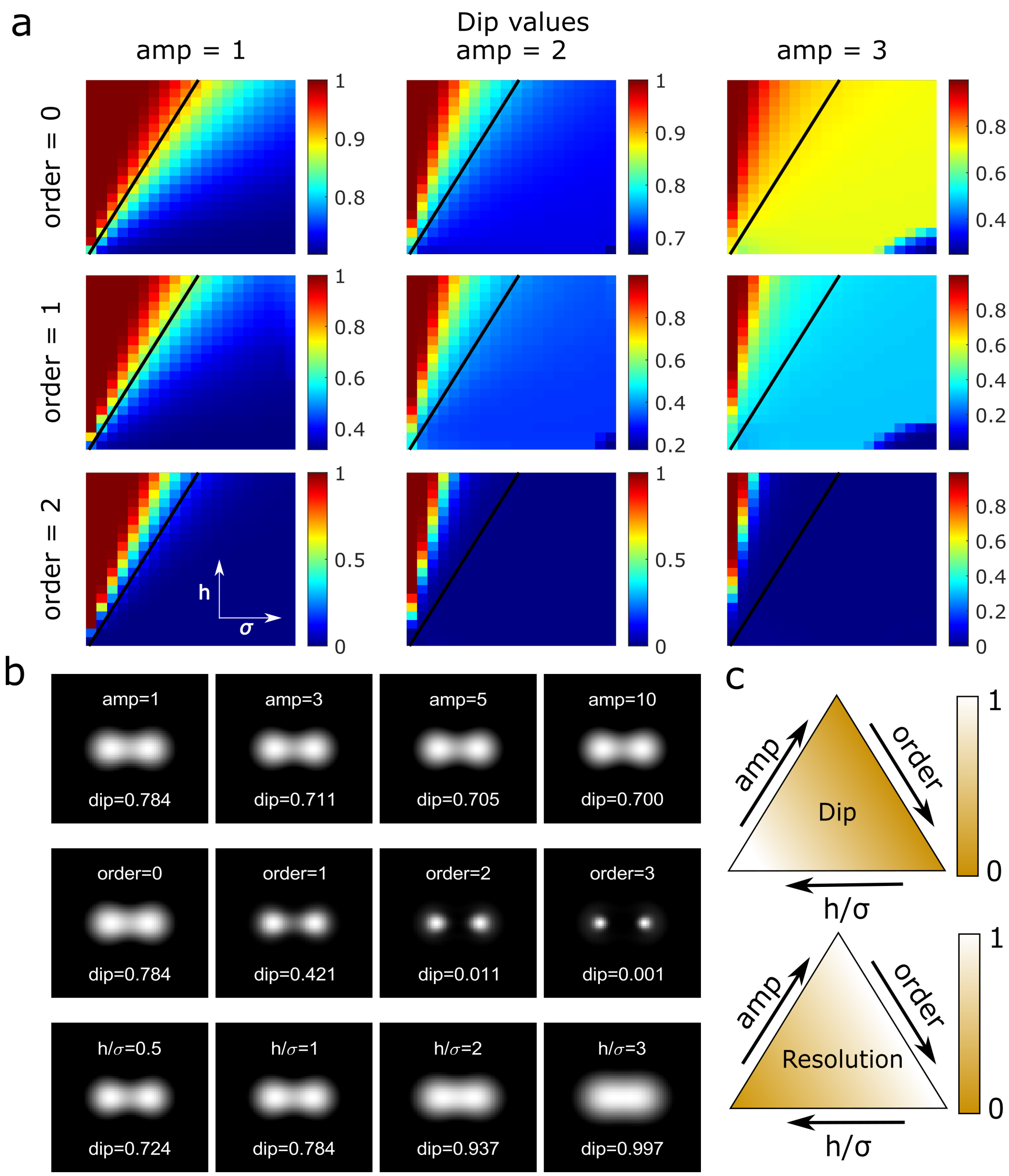

Figure 9. Influence of parameters on dip value and resolution in *MSSR* processing. (a) Heatmaps showing the dip values between two Gaussian emitters separated by the Sparrow limit, after applying *MSSR*. Columns vary by amplitude (*amp* = 1, 3, 5, 10); rows vary by *MSSR* iterations (*order* 0-3). The horizontal axis indicates the kernel standard deviation σ, and the vertical axis the neighborhood radius h. The black diagonal line marks the theoretical Sparrow threshold $h = 1.915\sigma$. (b) Representative *MSSR* outputs illustrating how dip changes with amplitude (top row), *MSSR* order (middle row), and the h/σ ratio (bottom row). Each image includes the computed dip value beneath it, indicating the minimum intensity between the two emitters. (c) Conceptual triangle diagrams summarizing the relationship between dip and resolution with parameters *amp*, *order*, and h/σ.

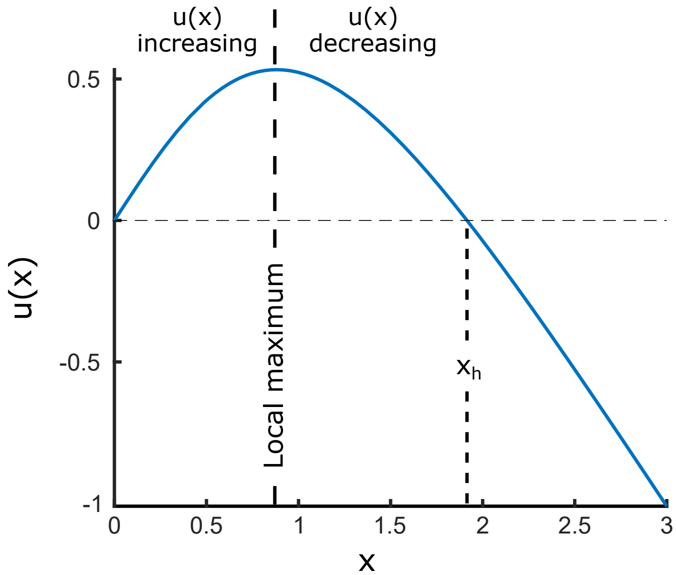

Figure 10. Behavior of function $u(x)$ and its roots.

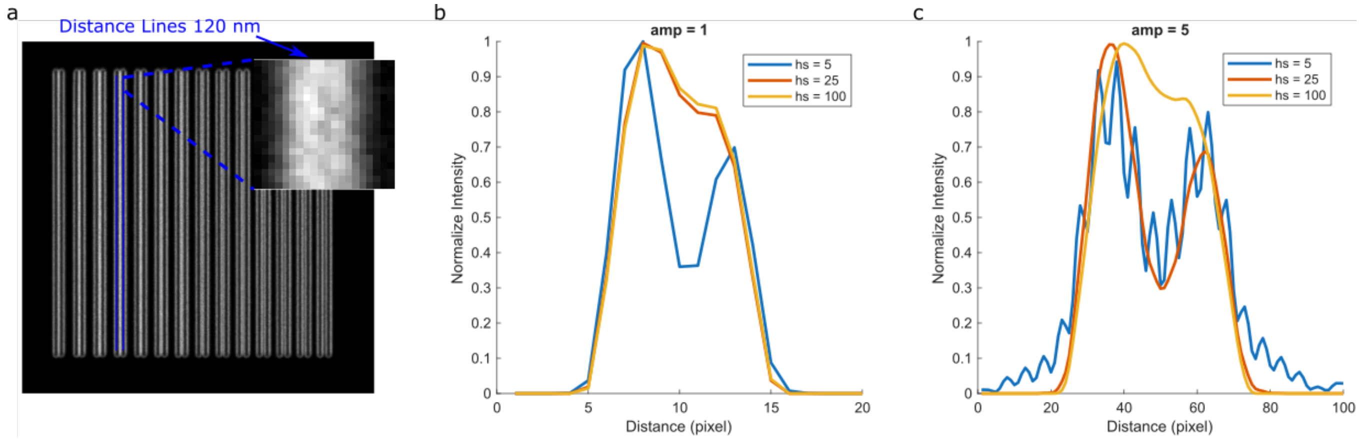

Figure 11. Analysis of *MSSR* parameter dependence on Argolight line pattern data. (a) Raw confocal image with the analyzed region marked, containing two vertical lines separated at the Sparrow limit. The inset shows a magnified view of this region, where the two lines appear merged into a single unresolved feature. (b) Intensity profiles obtained with $amp = 1$ for different kernel radii ($h_s = 5, 25, 100$). (c) Intensity profiles obtained with $amp = 5$ for the same set of kernel radii.