

Optimality of the Maximum Likelihood estimator in Astrometry

Sebastian Espinosa¹, Jorge F. Silva¹, Rene A. Mendez², Rodrigo Lobos³, and Marcos Orchard¹

¹ Information and Decision Systems Group, Department of Electrical Engineering, Facultad de Ciencias Físicas y Matemáticas, Universidad de Chile, Beauchef 850, Santiago, Chile

e-mail: sebastian.espinosa, josilva, morchard @ing.uchile.cl

² Departamento de Astronomía, Facultad de Ciencias Físicas y Matemáticas, Universidad de Chile, Casilla 36-D, Santiago, Chile

e-mail: rmendez@uchile.cl

³ Department of Electrical Engineering, University of Southern California, 90007, California, USA

e-mail: rlobos@usc.edu

May 11, 2018

ABSTRACT

Context. Astrometry relies on the precise measurement of positions and motions of celestial objects. Driven by the ever-increasing accuracy of astrometric measurements, it is important to critically assess the maximum precision that could be achieved with these observations.

Aims. The problem of astrometry is revisited from the perspective of analyzing the attainability of well-known performance limits (the Cramér-Rao bound) for the estimation of the relative position of light-emitting (usually point-like) sources on a CCD-like detector using commonly adopted estimators such as the weighted least squares and the maximum likelihood.

Methods. Novel technical results are presented to determine the performance of an estimator that corresponds to the solution of an optimization problem in the context of astrometry. Using these results we are able to place stringent bounds on the bias and the variance of the estimators in close form as a function of the data. We confirm these results through comparisons to numerical simulations under a broad range of realistic observing conditions.

Results. The maximum likelihood and the weighted least square estimators are analyzed. We confirm the sub-optimality of the weighted least squares scheme from medium to high signal-to-noise found in an earlier study for the (unweighted) least squares method. We find that the maximum likelihood estimator achieves optimal performance limits across a wide range of relevant observational conditions. Furthermore, from our results, we provide concrete insights for adopting an adaptive weighted least square estimator that can be regarded as a computationally efficient alternative to the optimal maximum likelihood solution.

Conclusions. We provide, for the first time, close-form analytical expressions that bound the bias and the variance of the weighted least square and maximum likelihood implicit estimators for astrometry using a Poisson-driven detector. These expressions can be used to formally assess the precision attainable by these estimators in comparison with the minimum variance bound.

Key words. Astrometry, parameter estimation, information limits, Cramér-Rao bound, maximum likelihood, weighted least squares.

1. Introduction

Astrometry, which deals with the accurate and precise measurement of positions and motions of celestial objects, is the oldest branch of observational astronomy, dating back at least to Hipparchus of Nicaea in 190 BC. Since, from its very beginnings, this branch of astronomy has required measurements over time to fulfill its goals, it could be considered the precursor of the nowadays fashionable “time-domain astronomy”¹, preceding it by at least 20 centuries. In recent years, astrometry has experienced a “coming of age” motivated by the rapid increase in positional precision allowed by the use of all-digital techniques and space observatories (see, e.g., Reffert (2009), Fig. 1a in Høg (2017) for an overview spanning more than 2000 years of astrometry, Benedict et al. (2016) for a summary of the contributions from the HST (fine guide sensors), and, of course, the exquisite prospects from the Prusti et al. (2016), with applications ranging from fundamental astrophysics (Van Altena 2013; Cacciari et al. 2016), to cosmology (Lattanzi 2012)).

A number of techniques have been proposed to estimate the location and flux of celestial sources as recorded on digital detectors (CCD). In this context, estimators based on the use of a least-squares (LS) error principle have been widely adopted (King 1983; Stetson 1987; Alard & Lupton 1998). The use of this type of decision rule has been traditionally justified through heuristic reasons. First, LS methods are conceptually straightforward to formulate based on the observation model of these problems. Second, they offer computationally efficient implementations and have shown reasonable performance (Lee & Van Altena 1983; Stone 1989; Vakili & Hogg 2016). Finally, the LS approach was the classical method used when the observations were obtained with analog devices (Van Altena & Auer 1975; Auer & Van Altena 1978), which are well characterized by a Gaussian noise model for the observations. In the Gaussian case the LS is equivalent to the maximum likelihood (ML) solution (Chun-Lin (1993), Gray & Davisson

¹ As defined, e.g., by Wikipedia: https://en.wikipedia.org/wiki/Time_domain_astronomy

(2004), Cover & Thomas (2012)), and, consequently, the LS method was taken from the analogous to the digital observational (Poisson noise model) setting naturally.

Considering astrometry as an inference problem (of, usually, point sources), the astrometric community has been interested for a long time in understanding the fundamental performance limits (or information bounds) of this task (Lindgren (2010) and references therein). It is well understood by the community that the characterization of this precision limit offers the possibility of understanding the complexity of the task and how it depends on key attributes of the problem, like the quality of the observational site, the performance of the instrument (CCD), and the details of the experimental conditions (Mendez et al. 2013, 2014). On the other hand, it provides meaningful benchmarks to define the optimality of practical estimators in the process of comparing their performance with the bounds (Lobos et al. 2015).

Concerning the characterization and analysis of fundamental performance bounds, we can mention some works on the use of the parametric Cramér-Rao (CR) bound by Lindgren (1978); Jakobsen et al. (1992); Zaccheo et al. (1995); Adorf (1996); and Bastian (2004). The CR bound is a minimum variance bound (MVB) for the family of unbiased estimators (Rao 1945; Cramér 1946). In astrometry and joint photometry and astrometry, Mendez et al. (2013, 2014) have recently studied the structure of this bound, and have analyzed its dependency with respect to important observational parameters under realistic (ground-based) astronomical observing conditions. In this context, closed-form expressions for the Cramér-Rao bound were derived in a number of important settings (high pixel resolution and low and high signal-to-noise (S/N) regimes), and their trends were explored across different CCD pixel resolutions and the position of the object in the CCD array. As an interesting outcome of those studies, the analysis of the CR bound has allowed us to predict the optimal pixel resolution of the array, as well as providing a formal justification to some heuristic techniques commonly used to improve performance in astrometry, like *dithering* for undersampled images (Mendez et al. 2013, Sect. 3.3). Recently, an application of the CR bound to moving sources has been done by Bouquillon et al. (2017), indicating excellent agreement between our theoretical predictions, simulations, and actual ground-based observations of the Gaia satellite, in the context of the GBOT program (Altmann et al. (2014)). The use of the CR bound on other applications is also of interest, e.g., in assessing the performance of star trackers to guide satellites with demanding pointing constraints (Zhang et al. 2016), or to meaningfully compare positional differences from different catalogues (for an example involving the SDSS and Gaia see Lemon et al. (2017)). Finally, a formulation of the (non-parametric) Bayesian CR bound in astrometry, using the so-called “Van-Trees inequality” (Van Trees 2004), has been presented by our group in Echeverria et al. (2016): This approach is particularly well suited for objects at the edge of detectability, and where some prior information is available, and has been proposed for the analysis of Gaia data for faint sources, or for those with a poor observational history (Michalik et al. 2015; Michalik & Lindgren 2016).

From the perspective of astrometric estimators, Lobos et al. (2015) have studied in detail the performance of the widely adopted LS estimator. In particular extending the result in So et al. (2013), Lobos and collaborators derived lower and upper bounds for the mean square error (MSE) of the LS estimator. Using these bounds, the optimality of the LS estimator was analyzed, demonstrating that for high S/N there is a considerable gap between the CR bound and the performance of the LS estimator (indicating a lack of optimality of this estimator). This work showed that for the very low S/N observational regime (weak astronomical sources), the LS estimator is near optimal, as its performance closely follows the CR bound. The limitations of the LS method in the medium to high S/N regime proved in that work opens up the question of studying alternative estimators that could achieve the CR bound on these regimes, which is the main focus of this paper, as outlined below.

1.1. Contribution and organization

In this work we study the ML estimator in astrometry, motivated by its well-known optimality properties in a classical parametric estimation setting with independent and identically distributed measurements (i.i.d.) (Kendall et al. 1999). We know that in the i.i.d. case this estimator is efficient with respect to the CR limit (Kay 1993), but it is important to emphasize that the observational setting of astrometry deviates from the classical i.i.d. case and, consequently, the analysis of its optimality is still an open problem. In particular, we face the technical challenge of evaluating its performance, a problem that, to the best of our knowledge, has not been addressed by the astrometric community. Concerning the independent but not identically distributed case Bradley & Gart (1962) and Hoadley (1971) give conditions under which ML estimators are consistent² and asymptotically normal³. Those conditions, however, are technically difficult to prove in this context.

The main challenge here is the fact that, as in the case of the LS estimator (Lobos et al. 2015), the ML estimator is the solution of an optimization problem with a nonconvex cost function of the data. This implies that it is not possible to directly compute the performance of the method. To address this technical issue, we extend the approach proposed by Fessler (1996) to approximate the variance and the mean of an implicit estimator solution of a generic optimization problem of the data through the use of a Taylor approximation around the mean measurement (see Theorem 1 below). Our extension considers high order approximations of the function that allows us not only to estimate the performance of the ML estimator through an explicit nominal value, but also it provides a confidence interval around it. With this result we revisit the more general weighted least square (WLS) and ML methods providing specific upper and lower bounds for both methods (see Theorems 2 and 3). The main findings from our analysis of the bounds are two fold: first we show that the WLS exhibits a sub-optimality similar to that of the LS method for medium to high S/N regimes discovered by Lobos et al. (2015) and, second, that the ML estimator achieves the CR limit for medium to high S/N and, consequently, it is optimal on those regimes. This last result is remarkable because, in conjunction with the result presented in Lobos et al. (2015), we are able to identify estimators that achieve the fundamental performance limits of astrometry in all the S/N regimes for the problem.

² this is, as the sample size increases, the sampling distribution of the estimator becomes increasingly concentrated at the true parameter value

³ more precisely, whose distribution around the true parameter approaches a normal distribution as the sample size grows

The paper is organized as follows: Section 2 introduces the background, preliminaries and notation of the problem. Section 3 presents the main methodological contribution of this work. Section 4 presents the application to astrometry considering the WLS and the ML schemes. Finally, numerical analysis of these performance bounds are presented in Sect. 5 and the final remarks and conclusion are given in Sect. 6.

2. Preliminaries and background

We begin by introducing the problem of astrometry. For simplicity, we focus on the 1-D scenario of a linear array detector, as it captures the key conceptual elements of the problem⁴.

2.1. Relative astrometry as a parameter estimation problem

The main problem at hand is the inference of the relative position (in the array) of a point source. This source is modeled by two scalar quantities, the position of object $x_c \in \mathbb{R}$ in the array⁵, and its intensity (or brightness, or flux) that we denote by $\tilde{F} \in \mathbb{R}^+$. These two parameters induce a probability distribution $\mu_{x_c, \tilde{F}}$ over an observation space that we denote by \mathbb{X} . Formally, given a point source represented by the pair (x_c, \tilde{F}) , it creates a nominal intensity profile in a photon integrating device (PID), typically a CCD, which can be expressed by

$$\tilde{F}_{x_c, \tilde{F}}(x) = \tilde{F} \cdot \phi(x - x_c, \sigma), \quad (1)$$

where $\phi(x - x_c, \sigma)$ denotes the one dimensional normalized point spread function (PSF) and where σ is a generic parameter that determines the width (or spread) of the light distribution on the detector (typically a function of wavelength and the quality of the observing site, see Sect. 5) (Mendez et al. 2013, 2014).

The profile in Eq. (1) is not measured directly, but it is observed through three sources of perturbations. First, an additive background noise which captures the photon emissions of the open (diffuse) sky, and the noise of the instrument itself (the read-out noise and dark-current Janesick (2001); Howell (2006); Janesick (2007); McLean (2008)), modeled by \tilde{B}_i in Eq. (2). Second, an intrinsic uncertainty between the aggregated intensity (the nominal object brightness plus the background) and actual measurements, which is modeled by independent random variables that follow a Poisson probability law. Finally, we need to account for the spatial quantization process associated with the pixel-resolution of the PID as specified in Eqs. (2) and (3). Modeling these effects, we have a countable collection of independent and non-identically distributed random variables (observations or counts) $\{I_i : i \in \mathbb{Z}\}$, where $I_i \sim \text{Poisson}(\lambda_i(x_c, \tilde{F}))$, driven by the expected intensity at each pixel element i , given by

$$\lambda_i(x_c, \tilde{F}) \equiv \mathbb{E}\{I_i\} = \underbrace{\tilde{F} \cdot g_i(x_c)}_{\equiv \tilde{F}_i(x_c, \tilde{F})} + \tilde{B}_i, \quad \forall i \in \mathbb{Z} \quad (2)$$

and

$$g_i(x_c) \equiv \int_{x_i - \Delta x/2}^{x_i + \Delta x/2} \phi(x - x_c, \sigma) dx, \quad \forall i \in \mathbb{Z}, \quad (3)$$

where $\mathbb{E}\{\}$ is the expectation value of the argument and $\{x_i : i \in \mathbb{Z}\}$ denotes the standard uniform quantization of the real line-array with resolution $\Delta x > 0$, i.e., $x_{i+1} - x_i = \Delta x$ for all $i \in \mathbb{Z}$. In practice, the PID has a finite collection of measurement elements (or pixels) I_1, \dots, I_n , then a basic assumption here is that we have a good coverage of the object of interest, in the sense that for a given position x_c

$$\sum_{i=1}^n g_i(x_c) \approx \sum_{i \in \mathbb{Z}} g_i(x_c) = \int_{-\infty}^{\infty} \phi(x - x_c, \sigma) dx = 1. \quad (4)$$

At the end, the likelihood (probability) of the joint observations $I^n = (I_1, \dots, I_n)$ (with values in \mathbb{N}^n) given the source parameters (x_c, \tilde{F}) is given by

$$L(I^n; x_c, \tilde{F}) = f_{\lambda_1(x_c, \tilde{F})}(I_1) \cdot f_{\lambda_2(x_c, \tilde{F})}(I_2) \cdots f_{\lambda_n(x_c, \tilde{F})}(I_n), \quad \forall I^n \in \mathbb{N}^n, \quad (5)$$

where $f_\lambda(x) = \frac{e^{-\lambda} \lambda^x}{x!}$ denotes the probability mass function (PMF) of the Poisson law (Gray & Davisson 2004).

Finally, if \tilde{F} is assumed to be known⁶, the astrometric estimation is the task of defining a decision rule $\tau_n() : \mathbb{N}^n \rightarrow \Theta$, with $\Theta = \mathbb{R}$ being the parameter space, where given an observation I^n the estimated position is given by $\hat{x}_c(I^n) = \tau_n(I^n)$.

⁴ The analysis can be extended to the 2-D case as presented in Mendez et al. (2013).

⁵ This captures the angular position in the sky and it is measured in seconds of arc (arcsec thereafter), through the ‘‘plate-scale’’, which is an optical design feature of the instrument plus telescope configuration.

⁶ The joint estimation of photometry and astrometry is the task of estimating both (x_c, \tilde{F}) from the observations, see Mendez et al. (2014).

2.2. The Cramér-Rao bound

In astrometry the Cramér-Rao bound has been used to bound the variance (estimation error) of any unbiased estimator (Mendez et al. 2013, 2014). In general, let I^n be a collection of independent observations that follow a parametric PMF $f_{\bar{\theta}}$ defined on \mathbb{N} . The parameters to be estimated from I^n will be denoted in general by the vector $\bar{\theta} = (\theta_1, \theta_2, \dots, \theta_m) \in \Theta = \mathbb{R}^m$. Let $\tau_n(I^n) : \mathbb{N}^n \rightarrow \Theta$ be an unbiased estimator⁷ of $\bar{\theta}$, and $L(I^n; \bar{\theta}) = f_{\bar{\theta}}(I_1) \cdot f_{\bar{\theta}}(I_2) \cdots f_{\bar{\theta}}(I_n)$ be the likelihood of the observation $I^n \in \mathbb{N}^n$ given $\bar{\theta} \in \Theta$. Then, the Cramér-Rao bound (Rao 1945; Cramér 1946) establishes that if

$$\mathbb{E}_{I^n \sim f_{\bar{\theta}}} \left\{ \frac{\partial \ln L(I^n; \bar{\theta})}{\partial \theta_i} \right\} = 0, \quad \forall i \in \{1, \dots, m\}, \quad (6)$$

then, the variance (denoted by Var), satisfies that

$$Var(\tau_n(I^n)_i) \geq [I_{\bar{\theta}}(n)^{-1}]_{i,i}, \quad (7)$$

where $I_{\bar{\theta}}(n)$ is the *Fisher information* matrix given by

$$[I_{\bar{\theta}}(n)]_{i,j} = \mathbb{E}_{I^n \sim f_{\bar{\theta}}} \left\{ \frac{\partial \ln L(I^n; \bar{\theta})}{\partial \theta_i} \cdot \frac{\partial \ln L(I^n; \bar{\theta})}{\partial \theta_j} \right\} \quad \forall i, j \in \{1, \dots, m\}. \quad (8)$$

In particular, for the scalar case ($m = 1$), we have that for all $\theta \in \Theta$

$$\min_{\tau_n(\cdot) \in \mathcal{T}^n} Var(\tau_n(I^n)) \geq I_{\theta}(n)^{-1} = \mathbb{E}_{I^n \sim f_{\theta}} \left\{ \left[\left(\frac{d \ln L(I^n; \theta)}{d\theta} \right)^2 \right] \right\}^{-1}, \quad (9)$$

where \mathcal{T}^n is the collection of all unbiased estimators and $I^n \sim f_{\theta}$. For astrometry, Mendez et al. (2013, 2014) have characterized and analyzed the Cramér-Rao bound, leading to

Proposition 1. (Mendez et al. (2014, Sect. 2.4)) *If $\tilde{F} \in \mathbb{R}^+$ is fixed and known, and we want to estimate x_c from $I^n \sim f_{(x_c, \tilde{F})} = L(I^n; x_c, \tilde{F})$ in Eq. (5), then the Fisher information is given by*

$$I_{x_c}(n) = \sum_{i=1}^n \frac{(\tilde{F} \frac{dg_i(x_c)}{dx_c})^2}{\tilde{F} g_i(x_c) + \tilde{B}_i}, \quad (10)$$

which from Eq. (9) induces a MVB for the astrometric estimation problem, and where $\sigma_{CR}^2(n) \equiv I_{x_c}(n)^{-1}$ denotes the (astrometric) CR bound.

2.3. Achievability and performance of the LS estimator

Concerning the achievability of the CR bound with a practical estimator, Lobos et al. (2015, Proposition 2) have demonstrated that this bound cannot be attained, meaning that for any unbiased estimator $\tau_n(\cdot)$ we have that

$$Var(\tau_n(I^n)) > \sigma_{CR}^2, \quad (11)$$

where I^n follows the Poisson PMF $f_{(x_c, \tilde{F})}$ in Eq. (5).

This finding should be interpreted with caution, considering its pure theoretical meaning. This is because Eq. (11) does not exclude the possibility that the CR bound could be approximated arbitrarily close by a practical estimation scheme. Motivated by this refined conjecture, Lobos et al. (2015) proposed to study the performance of the widely adopted LS estimator⁸ with the goal of deriving operational upper and lower performance bounds of its performance that could be used to determine how far could this scheme depart from the CR limit. Then, from this result, it was possible to evaluate the goodness of the LS estimator for concrete observational regimes. For bounding the performance of the LS estimator, the challenge was that $\tau_{LS}(I^n)$ is an implicit function of the data (where no close-form expression is available) and, consequently, Lobos et al. (2015) derived a result to bound the estimation error and the variance of $\tau_{LS}(I^n)$. We can briefly summarize the main result presented in Lobos et al. (2015, Theorem 1) saying that under certain mild sufficient conditions (that were shown to be realistic for astrometry), there is a constant $\delta > 0$ (that depends on the observational regime, in particular the S/N) and a nominal variance σ_{LS}^2 , which is determined in closed-form in the result, from which it is possible to bound $Var(\tau_{LS}(I^n))$ by the simple expression

$$Var(\tau_{LS}(I^n)) \in \left(\frac{\sigma_{LS}^2(n)}{(1+\delta)^2}, \frac{\sigma_{LS}^2(n)}{(1-\delta)^2} \right), \quad (12)$$

⁷ In the sense that, for all $\bar{\theta} \in \Theta$, $\mathbb{E}_{I^n \sim f_{\bar{\theta}}} \{ \tau_n(I^n) \} = \bar{\theta}$.

⁸ This is the solution of $\tau_{LS}(I^n) = \arg \min_{\alpha \in \mathbb{R}} \sum_{i=1}^n (I_i - \lambda_i(\alpha))^2$, with $\lambda_i(\alpha) = \tilde{F} g_i(\alpha) + \tilde{B}_i$, α being a generic variable representing the astrometric position, $g_i(\cdot)$ is given by Eq. (3), and where $\arg \min$ represents the argument that minimizes the expression. More details are presented in Lobos et al. (2015).

where

$$\sigma_{LS}^2(n) = \frac{\sum_{i=1}^n (\tilde{F} g_i(x_c) + \tilde{B}_i) \cdot (g'_i(x_c))^2}{(\tilde{F} \sum_{i=1}^n (g'_i(x_c))^2)}. \quad (13)$$

Note that when δ is small, $\sigma_{LS}^2(n)$ tightly determines the performance of the LS estimator, and its comparison with σ_{CR}^2 can be used to evaluate the goodness of the LS estimator for astrometry. Based on a careful comparison, it was shown in Lobos et al. (2015, Sect. 4) that, in general, $\sigma_{LS}^2(n)$ is close to $\sigma_{CR}^2(n)$ for the small S/N regime of the problem. However for moderate to high S/N regimes, the gap between $\sigma_{LS}^2(n)$ and $\sigma_{CR}^2(n)$ becomes quite significant⁹.

These unfavorable findings for the LS method have motivated us to study alternatives schemes that could potentially approach better the Cramér-Rao bound for the rich observational context of medium to high S/N regimes. This will be the focus of the following sections, where in particular we explore the performance of the ML and WLS estimators, thus extending and generalizing the analysis done for the LS estimator by our group presented in Lobos et al. (2015).

3. Bounding the performance of an implicit estimator

Before we go to the case of the WLS and the ML estimators, we present a general result that bounds the performance of *any* estimator that is the solution of a generic optimization problem. Let us consider a vector of observations $I^n = (I_1, \dots, I_n) \in \mathbb{R}^n$ and a general so-called cost function $J(\alpha, I^n)$. Then the estimation of x_c from the data is the solution of the following optimization problem:

$$\tau_J(I^n) \equiv \arg \min_{\alpha \in \mathbb{R}} J(\alpha, I^n), \quad (14)$$

where α represents the position of the object in the context of astrometry. As in our previous work (Lobos et al. 2015), the challenge here is that this estimator is implicit because no closed-form expression of the data which solves Eq. (14) is assumed. In particular, this implies that both the variance and the estimation error of $\tau_J(I^n)$ can not be determined directly. To address this technical issue, we extend the approach proposed by Fessler (1996) to approximate the variance and the mean of an implicit estimator solution of a problem described by Eq. (14) through the use of a Taylor approximation around the mean measurement, i.e., $\bar{I}^n = \mathbb{E}_{I^n \sim f(x_c, F)}(I^n)$.

More precisely, we assume that $J(\alpha, I^n)$ has a unique global minimum at $\tau_J(I^n)$, and that it has a regular behavior, so its partial derivatives are zero, i.e.,

$$0 = \left. \frac{\partial}{\partial \alpha} J(\alpha, I^n) \right|_{\alpha=\tau_J(I^n)} \equiv \frac{\partial}{\partial \alpha} J(\tau_J(I^n), I^n). \quad (15)$$

Then we can obtain $\tau_J(I^n)$ by a first order Taylor expansion around the mean \bar{I}^n by

$$\tau_J(I^n) = \tau_J(\bar{I}^n) + \sum_{i=1}^n \frac{\partial}{\partial I_i} \tau_J(\bar{I}^n) (I_i - \bar{I}_i) + \underbrace{\frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \frac{\partial^2}{\partial I_i \partial I_j} \tau_J(\bar{I}^n - t(I^n - \bar{I}^n)) (I_i - \bar{I}_i) (I_j - \bar{I}_j)}_{\equiv e(\bar{I}, I - \bar{I})}. \quad (16)$$

with $t \in [0, 1]$ is fixed but unknown¹⁰. For simplicity, Eq. (16) can be written in matrix form as

$$\tau_J(I^n) = \tau_J(\bar{I}^n) + \nabla \tau_J(\bar{I}^n) \cdot (I^n - \bar{I}^n) + e(\bar{I}^n, I^n - \bar{I}^n), \quad (17)$$

where $\nabla = [\frac{\partial}{\partial I_1} \dots \frac{\partial}{\partial I_n}]$ denotes the row gradient operator and $e(\bar{I}^n, I - \bar{I}^n)$ is the residual error of the Taylor expansion. From Eq. (17) we can readily obtain the following expression for its variance

$$\text{Var}\{\tau_J(I^n)\} = \underbrace{\nabla \tau_J(\bar{I}^n) \text{Cov}\{I^n\} \nabla \tau_J(\bar{I}^n)^T}_{\equiv \sigma_J^2(n)} + \underbrace{\text{Var}\{e(\bar{I}^n, I^n - \bar{I}^n)\} + 2 \text{Cov}\{\nabla \tau_J(\bar{I}^n) (I^n - \bar{I}^n), e(\bar{I}^n, I^n - \bar{I}^n)\}}_{\equiv \gamma_J(n)}. \quad (18)$$

In Eq. (18) we recognize two terms: $\sigma_J^2(n)$ that captures the linear behaviour of $\tau_J(\cdot)$ around \bar{I}^n and $\gamma_J(n)$ which reflects the deviation from this linear trend. It should be noted that the above expression does not depend on $\tau(I^n)$ itself, but on its partial derivatives evaluated at the mean vector of observations. Then in the adoption of this approach to estimate $\text{Var}\{\tau_J(I^n)\}$, a key task is to determine $\nabla \tau_J(\bar{I}^n)$.

Remark 1. It is meaningful to note that Fessler (1996) only considered the linear term in his approximate analysis, obviating the residual term $\gamma_J(n)$ in Eq. (18). This first order reduction is not realistic for our problem because the solution of a problem like the one posed by Eq. (14) in astrometry has important non-linear components that need to be considered in the analysis of Eq. (18).

⁹ In particular, for the very high S/N regime and assuming $\Delta x/\sigma \ll 1$, Lobos et al. (2015, Proposition 3) shows that this gap reaches the condition $\frac{\sigma_{LS}^2(n)}{\sigma_{CR}^2(n)} \approx \frac{8}{3\sqrt{3}}$.

¹⁰ It follows that $\lim_{I \rightarrow \bar{I}} \frac{e(\bar{I}, I - \bar{I})}{\|I - \bar{I}\|_2} = 0$.

In an effort to analyze both the linear and non-linear aspects of a general intrinsic estimator solution to Eq. (14), the following result offers sufficient conditions to determine $\sigma_J^2(n)$ in closed-form, and to bound the magnitude of the residual term $\gamma_J(n)$ in Eq. (18).

Theorem 1. *Let us consider a fixed and unknown parameter $x_c \in \mathbb{R}$, the observations $I^n = (I_1, \dots, I_n)^T$ where $I_i \sim f_{x_c}$, and $\tau_J(I^n)$ the estimator solution of Eq. (14). If we satisfy the following two rather general conditions:*

- a) *the cost function $J(\alpha, I^n)$ is twice differentiable with respect to I^n and x_c , and the gradient of $\tau_J(\cdot)$ evaluated in the mean data \bar{I}^n offers the following decomposition*

$$\nabla \tau_J(\bar{I}^n) \cdot (I^n - \bar{I}^n) = a \sum_{i=1}^N b_i (I_i - \bar{I}_i) \quad (19)$$

with a and $\{b_i : i \in \{1, \dots, N\}\}$ constants, and,

- b) *the estimator evaluated in the mean data equals the true parameter $x_c \in \mathbb{R}$, this is,*

$$\tau_J(\bar{I}^n) = x_c, \quad (20)$$

then we can define two new quantities $\epsilon_J(n)$ and $\beta_J(n)$ (both > 0) and $\sigma_J^2(n)$ in Eq. (18) with analytical expressions (details presented in Appendix A) such that

$$\underbrace{|\mathbb{E}_{I^n \sim f_{x_c}} \{\tau_J(I^n)\} - x_c|}_{\text{bias}} \leq \epsilon_J(n) \quad (21)$$

and

$$\text{Var}_{I^n \sim f_{x_c}} \{\tau_J(I^n)\} \in (\sigma_J^2(n) - \beta_J(n), \sigma_J^2(n) + \beta_J(n)). \quad (22)$$

The proof of this result and the expression for $(\epsilon_J(n), \sigma_J^2(n), \beta_J(n))$ in Eqs. (21) and (22) are presented in detail in Appendix A.

Revisiting the equality in Eq. (18), Theorem 1 provides general sufficient conditions to bound the residual term $\gamma_J(n)$ and by doing that, a way of bounding the variance of $\tau_J(I^n)$ which is the solution of Eq. (14). In particular, it is worth noting that if the ratio $\frac{\beta_J(n)}{\sigma_J^2(n)} \ll 1$, then Eq. (22) offers a tight bound for $\text{Var}_{I^n \sim f_{x_c}} \{\tau_J(I^n)\}$. In this last context, $\sigma_J^2(n)$ (called the nominal value of the result) provides a very good approximation for $\text{Var}_{I^n \sim f_{x_c}} \{\tau_J(I^n)\}$.

On the application of this result to the WLS and ML estimators, we will see that the main assumption in Eq. (19) is satisfied in both cases (see Eqs. (B.8) and (C.6) in Appendix B and C, respectively), and from that $\sigma_J^2(n)$ is playing an important role to approximate the performance of ML and WLS in a wide range of observational regimes. In addition, the analysis of the bias in Eq. (21) shows that these estimators are unbiased for any practical purpose and, consequently, contrasting their performance (estimation error $\sim \sqrt{\text{variance}}$) with the CR bound is a meaningful way to evaluate optimality.

4. Application to astrometry

In this section we apply Theorem 1 to bound the variances of the ML and WLS estimators in the context of astrometry. Following the model presented in Sect. 2.1, $I^n = (I_1, \dots, I_n)^T$ denotes the measurements acquired by each pixel of the array, and where each of them follows a Poisson distribution given by

$$I_i \sim \text{Poisson}(\lambda_i(x_c)), \quad i = 1, \dots, n, \quad (23)$$

as expressed by Eqs. (2) and (3).

4.1. Bounding the variance of the WLS estimator

The WLS estimator, denoted by $\tau_{WLS}(I^n)$ in Eq. (25), is implicitly defined through a cost function given by

$$J_{WLS}(\alpha, I^n) = \sum_{i=1}^n w_i (I_i - \lambda_i(\alpha))^2, \quad (24)$$

where $(w_1, \dots, w_n)^T \in \mathbb{R}_+^n$ is a weight vector, and α is a general source position parameter. Specifically we have that

$$\tau_{WLS}(I^n) = \arg \min_{\alpha \in \mathbb{R}} J_{WLS}(\alpha, I^n). \quad (25)$$

Applying Theorem 1 we obtain the following result:

Theorem 2. *Let us consider the WLS estimator solution of Eq. (25), then we have that*

$$\underbrace{|\mathbb{E}_{I^n \sim f_{x_c}} \{\tau_{WLS}(I^n)\} - x_c|}_{\text{bias}} \leq \epsilon_{WLS}(n) \quad (26)$$

and

$$\text{Var}_{I^n \sim f_{x_c}} \{\tau_{WLS}(I^n)\} \in (\sigma_{WLS}^2(n) - \beta_{WLS}(n), \sigma_{WLS}^2(n) + \beta_{WLS}(n)), \quad (27)$$

where $\sigma_{WLS}^2(n)$ is given by

$$\sigma_{WLS}^2(n) = \frac{\sum_{i=1}^n w_i^2 \lambda_i(x_c) \left(\frac{\partial \lambda_i(\alpha)}{\partial \alpha} \right)^2 \Big|_{\alpha=x_c}}{\left(\sum_{i=1}^n w_i \left(\frac{\partial \lambda_i(\alpha)}{\partial \alpha} \right)^2 \Big|_{\alpha=x_c} \right)^2} \quad (28)$$

and $\beta_{WLS}(n)$ and $\epsilon_{WLS}(n)$ are well defined analytical expression of the problem (presented in Appendix B).

The proof of this result and the expressions for $\epsilon_{WLS}(n)$ and $\beta_{WLS}(n)$ in Eqs. (26) and (27), respectively, are elaborated in Appendix B. This result offers concrete expressions to bound the bias as well as the variance of the WLS estimator. For the bias bound in Eq. (26), it will be shown that $\epsilon_{WLS}(n)$ is very small (order of magnitudes smaller than x_c) for all the observational regimes explored in this work and, consequently, the WLS can be considered an unbiased estimator in astrometry, as it would be expected. Concerning the bounds for the variance in Eq. (27), we will show that for high and moderate S/N regimes the ratio $\beta_{WLS}(n)/\sigma_{WLS}^2(n) \ll 1$ and consequently in this context $\sigma_{WLS}^2(n)$ is a precise estimator of $\text{Var}_{I^n \sim f_{x_c}} \{\tau_{WLS}(I^n)\}$. For the very small S/N the results offers an admissible interval $\sigma_{WLS}^2(n) \pm \beta_{WLS}(n)$ around the nominal value $\sigma_{WLS}^2(n)$. Therefore in any context $\sigma_{WLS}^2(n)$ shows to be a meaningful approximation for the performance of the WLS.

Remark 2. If we focus on the analysis of the closed form expression $\sigma_{WLS}^2(n)$ as an approximation of $\text{Var}_{I^n \sim f_{x_c}} \{\tau_{WLS}(I^n)\}$ and we compare it with the CR bound $\sigma_{CR}^2(n)$ in Eqs. (10) and (11), we note that they are very similar in their structure. In particular, it follows that $\sigma_{WLS}^2(n) = \sigma_{CR}^2(n)$, if and only if, the weights of the WLS estimator are selected in the following way

$$w_i = K \cdot \frac{1}{\lambda_i(x_c)}, \quad \forall i \in \{1, \dots, n\}, \quad (29)$$

where K is an arbitrary constant ($K > 0$). In other words, the only way in which the performance of the WLS approximates the CR limit is if we select the weights as in Eq. (29). However, this selection uses the information of the true position x_c , which is unfeasible as it contradicts the very essence of the inference task (indeed, x_c is unknown, and we are trying to estimate it from the data). Another interpretation is that no matter how we choose the weights of the WLS estimator, it is not possible that the WLS is close to the CR bound for every position x_c , telling us that the WLS is intrinsically not optimal from the perspective of being close to the CR limit in all the possible astrometric scenarios. In particular, this impossibility result is very strong in the high S/N regimes where $\sigma_{WLS}^2(n) \approx \text{Var}_{I^n \sim f_{x_c}} \{\tau_{WLS}(I^n)\}$. This implication is consistent with the analysis presented by Lobos et al. (2015, Fig. 4), where it was shown that the variance of the LS estimator is significantly higher than then CR bound in the high S/N regime. This justifies the study of the ML estimator.

4.2. Bounding the variance of the ML estimator

The ML estimator, denoted by $\tau_{ML}(I^n)$ in Eq. (31), is implicitly defined through a cost function

$$J(\alpha, I^n) = \sum_{i=1}^n I_i \ln(\lambda_i(\alpha)) - \lambda_i(\alpha), \quad (30)$$

where α is a general source position parameter. Specifically, given an observation I^n we have that

$$\begin{aligned} \tau_{ML}(I^n) &= \arg \max_{\alpha \in \mathbb{R}} J(\alpha, I^n), \\ &= \arg \min_{\alpha \in \mathbb{R}} \sum_{i=1}^n -I_i \ln(\lambda_i(\alpha)) + \lambda_i(\alpha). \end{aligned} \quad (31)$$

Applying Theorem 1 we obtain the following result:

Theorem 3. *Let us consider the ML estimator solution of Eq. (31), then we have that*

$$\underbrace{|\mathbb{E}_{I^n \sim f_{x_c}} \{\tau_{ML}(I^n)\} - x_c|}_{\text{bias}} \leq \epsilon_{ML}(n) \quad (32)$$

and

$$\text{Var}_{I^n \sim f_{x_c}} \{\tau_{ML}(I^n)\} \in (\sigma_{ML}^2(n) - \beta_{ML}(n), \sigma_{ML}^2(n) + \beta_{ML}(n)), \quad (33)$$

where

$$\sigma_{ML}^2(n) = \sigma_{CR}^2(n) = \left(\sum_{i=1}^n \frac{\left(\tilde{F} \frac{d g_i(x_c)}{d x_c} \right)^2}{\tilde{F} g_i(x_c) + \tilde{B}_i} \right)^{-1}, \quad (34)$$

and $\beta_{ML}(n)$ and $\epsilon_{ML}(n)$ are well defined analytical expression of the problem (presented in Appendix C).

The proof of this result and the expressions for $\epsilon_{ML}(n)$ and $\beta_{ML}(n)$ in Eqs. (32) and (33), respectively, are elaborated in Appendix C.

Remark 3. It is important to mention that the magnitude of $\epsilon_{ML}(n)$ is orders of magnitude smaller than x_c in all the observational regimes studied in this work (see this analysis in Sect. 5) and, consequently, for any practical purpose the ML is an unbiased estimator. This implies that the comparison with the CR bound is a meaningful indicator when evaluating the optimality of the ML estimator.

Remark 4. We observe that if the ratio $\beta_{ML}(n)/\sigma_{ML}^2(n)$ is significantly smaller than one, which is shown in Sect. 5 from medium to high S/N regimes, then $\text{Var}_{I^n \sim f_{x_c}}\{\tau_{ML}(I^n)\} \approx \sigma_{ML}^2(n)$. This is a very interesting result because we can approximate the performance of the ML estimator with $\sigma_{ML}^2(n)$. On this context, it is remarkable to have that the nominal value $\sigma_{ML}^2(n)$ is precisely the CR bound (see Eq. (34)), because this means that the ML estimator closely approximate this MVB in the interesting regime from moderate to very high S/N . Note that this medium-high S/N regime is precisely the context where the LS estimator shows significant deficiencies as presented in Lobos et al. (2015). Therefore, ML offers optimal performances in the regime where LS type of methods are not able to match the CR bound, which satisfactorily resolves the question posted by Lobos et al. (2015) on the study of schemes that could closely approach the CR bound in the high S/N regime.

5. Numerical analysis

In this section we evaluate numerically the performance bounds obtained in Sect. 4 for the WLS and ML estimators, and compare them with the astrometric CR bound in Proposition 1. The idea is to consider some realistic astrometric conditions to evaluate the expressions developed in Theorems 2 and 3 and their dependency on important observational conditions and regimes. As we shall see, key variables in this analysis are the tradeoff between the intensity of the object and the noise represented by the S/N value, and the pixel resolution of the CCD.

5.1. Experimental setting

We adopt some realistic design and observing variables to model the problem (Mendez et al. 2013, 2014). For the PSF, analytical and semi-empirical forms have been introduced, see for instance the ground-based model in King (1971) and the space-based models by King (1983) or Bendinelli et al. (1987). In this work we will adopt a Gaussian PSF, i.e., $\phi(x, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}}$ in Eq. (3), and where σ is the width of the PSF and is assumed to be known. This PSF has been found to be a good representation for typical astrometric-quality Ground-based data (Méndez et al. 2010). In terms of nomenclature, $FWHM \equiv 2\sqrt{2\ln 2}\sigma$ measured in arcsec, denotes the *Full-Width at Half-Maximum* parameter, which is an overall indicator of the image quality at the observing site (Chromey 2016).

The background profile, represented by $\{\tilde{B}_i, i = 1, \dots, n\}$ in Eq. (2), is a function of several variables, like the wavelength of the observations, the moon phase (which contributes significantly to the diffuse sky background), the quality of the observing site, and the specifications of the instrument itself. We will consider a uniform background across pixels underneath the PSF, i.e., $\tilde{B}_i = \tilde{B}$ for all i . To characterize the magnitude of \tilde{B} , it is important to first mention that the detector does not measure photon counts [e^-] directly, but a discrete variable in “*Analog to Digital Units* (ADUs)” of the instrument, which is a linear proportion of the photon counts (McLean (2008)). This linear proportion is characterized by the gain of the instrument G in units of [e^-/ADU]. G is just a scaling value, where we can define $F \equiv \tilde{F}/G$ and $B \equiv \tilde{B}/G$ as the intensity of the object and noise, respectively, in the specific ADUs of the instrument. Then, the background (in ADUs) depends on the pixel size Δx arcsec as follows

$$B = f_s \Delta x + \frac{D + RON^2}{G} [ADU], \quad (35)$$

where f_s is the (diffuse) sky background in $ADU \text{ arcsec}^{-1}$, while D and RON^2 , both measured in e^- model the dark-current and read-out-noise of the detector on each pixel, respectively. Note that the first component in Eq. (35) is attributed to the site, and its effect is proportional to the pixel size. On the other hand, the second component is attributed to errors of the PID (detector), and it is pixel-size independent. This distinction is central when analyzing the performance as a function of the pixel resolution of the array (see details in Mendez et al. (2013, Sect. 4)). More important is the fact that in typical ground-based astronomical observation, long exposure times are considered, which implies that the background is dominated by diffuse light coming from the sky, and not from the detector (Mendez et al. 2013, Sect. 4).

For the experimental conditions, we consider the scenario of a ground-based station located at a good site with clear atmospheric conditions and the specification of current science-grade CCDs, where $f_s = 1502.5 \text{ ADU arcsec}^{-1}$, $D = 0$, $RON = 5 \text{ e}^-$, $FWHM = 1 \text{ arcsec}$ (equivalent to $\sigma = 1/(2\sqrt{2\ln 2}) \text{ arcsec}$) and $G = 2 \text{ e}^- \text{ ADU}^{-1}$ (with these values we will have $B = 313 \text{ ADU}$ for $\Delta x = 0.2 \text{ arcsec}$ using Eq. (35)). In terms of scenarios of analysis, we explore different pixel resolutions for the CCD array $\Delta x \in [0.1, 0.65]$ measured in arcsec, and different signal strengths $\tilde{F} \in \{1080, 3224, 20004, 60160\}$, measured in e^- , which corresponds to $S/N \in \sim \{12, 32, 120, 230\}$. Note that increasing \tilde{F} implies increasing the S/N of the problem, which can be approximately measured by the ratio \tilde{F}/\tilde{B} . On a given detector plus telescope setting, these different S/N scenarios can be obtained by appropriately changing the exposure time (open shutter) that generates the image.

5.2. Bias analysis

Considering the upper bound terms $\epsilon_{WLS}(n)$ and $\epsilon_{ML}(n)$ for the bias error obtained from Theorems 2 and 3 for the WLS and ML, respectively, Fig. 1 presents the relative bias error for different S/N regimes and pixel resolutions. In the case of the ML estimator, the bounds for relative bias error are very small in all the explored S/N regimes and pixel resolutions meaning that for any practical purposes this estimator is unbiased as expected from theory (Gray & Davisson 2004). For the case of the WLS, we observe that from medium to high S/N the relative error bound obtained is very small but, meanwhile at low S/N , unbiasedness can not be fully guaranteed from the bound in Eq. (26). In general, our results show that both WLS and ML are unbiased estimators for astrometry in a wide range of relevant observational regimes (in particular from medium to high S/N) and, consequently, it is meaningful to analyze the optimality of these estimators in comparison with the CR bound in those regimes.

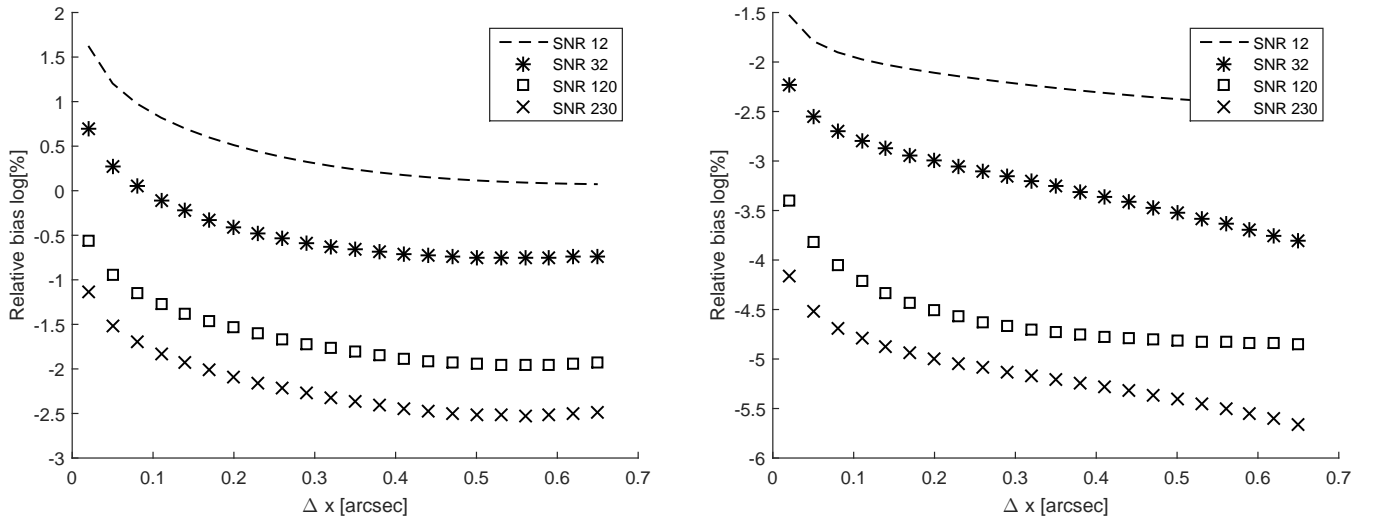


Fig. 1. Relative performance of the bias (as measured by $\log\left(100 \times \frac{\epsilon_f(n)}{x_c}\right)$) stipulated by Theorem 1 for the WLS estimator (left side, Eq. (26)) and the ML estimator (right side, Eq. (32)). Results are reported for different values of the source flux $\tilde{F} \in \{1080, 3224, 20004, 60160\}$, all in e^- (top to bottom symbols respectively), as a function of the detector pixel size. The 0% level corresponds to having achieved no bias.

In the following sections, we move to the analysis of the variance of the WLS and ML with particular focus on the medium to high S/N regimes across all pixel resolutions using the performance bounds derived in Eqs. (27) and (33), respectively.

5.3. Performance analysis of the WLS estimator

In this section, we evaluate numerically the expression derived in Theorem 2 to bound the variance of the WLS estimator in Eq. (27). For that we characterize the admissible regime predicted for the variance of the WLS estimator, i.e., the interval

$$\left(\sigma_{WLS}^2(n) - \beta_{WLS}(n), \sigma_{WLS}^2(n) + \beta_{WLS}(n)\right), \quad (36)$$

for $S/N \in \{12, 32, 120, 230\}$ and $\Delta x \in [0.01, 0.65]$ arcsec. In these bounds, we recognize its central value (or nominal value) $\sigma_{WLS}^2(n)$ in Eq. (28) and the length of the interval $2\beta_{WLS}(n)$ that is determined in closed form for its numerical evaluation in Eqs. (A.2) and (B.11). Note that $2\beta_{WLS}(n)$ can be considered an indicator of the precision of our result to approximate the variance of the WLS in astrometry.

5.3.1. Revisiting the uniform weight case

To begin the analysis, we consider the setting of uniform weights across pixels, i.e., the case of the LS estimator and, without loss of generality, we locate the object in the center of the field of view¹¹, which can be considered a reasonable scenario to represent the complexity of the astrometry task. At this point, it is important to remind the reader that from the analysis of $\sigma_{WLS}^2(n)$ in Remark 2, the nominal value $\sigma_{WLS}^2(n)$ is equivalent to the CR bound when the w_i are selected as a function of the true position in Eq. (29). In view of this observation, the selection of non-uniform weights can be interpreted as biasing the estimation to a particular area of the angular space, which goes in contradiction with the essence of the inference problem that estimates the position with no prior information, and only relies on the measured counts. From this interpretation, revisiting the LS estimator is an important first step in the analysis of the WLS framework.

¹¹ It is important to remind that the CR bound is a function of the value of the parameter to be estimated, in this case the position x_c , see the Fisher information in Eq. (10).

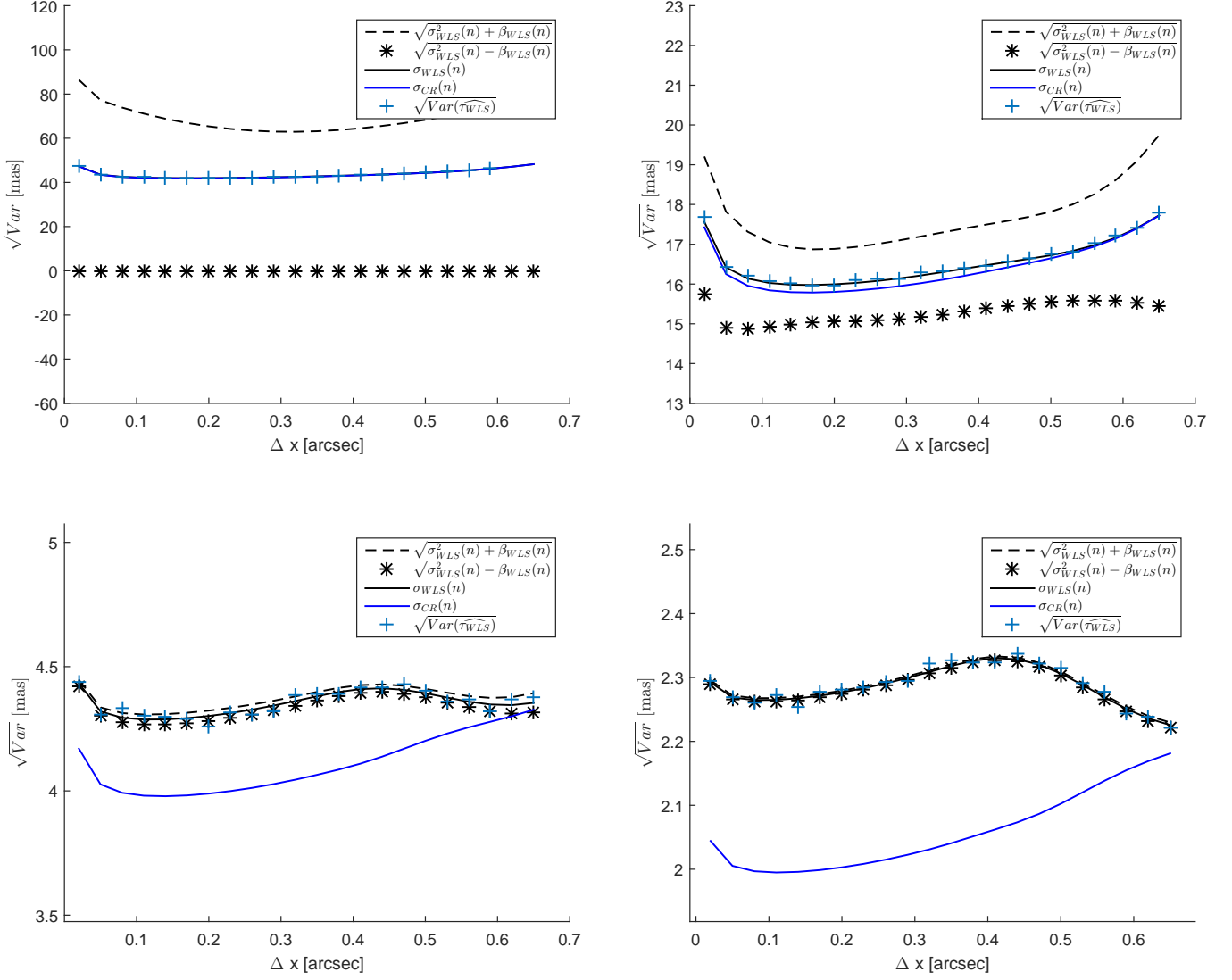


Fig. 2. Range of the square root of the variance performance (in miliarcsecond=mas) for the WLS method in astrometry using uniform weights (equivalent to the LS method) predicted by Theorems 1 and 2, Eq. (27). Results are reported for different representative values of \tilde{F} and across different pixel sizes (top-left to bottom-right): $\tilde{F} \in \{1080, 3224, 20004, 60160\} e^-$.

On the specifics, the boundaries of the interval in Eq. (36) and its nominal values are illustrated in Fig. 2 for the different observational regimes. In addition, Fig. 2 shows the CR bound as a reference to evaluate the optimality of the LS scheme across settings. We observe that for the low $S/N \sim 12$ regime, the nominal values precisely match the CR bound, however our result is not conclusive regarding the performance of the LS estimator. In the regime $S/N \in (30, 50)$ (top right panel on Fig. 2), we notice an important reduction on the range of admissible performance, and our result becomes more informative and meaningful. In this context, the nominal values is very close to the CR bound, and we could assert that the LS estimator offers sufficiently good performance in the sense that is very competitive with the MVB. Importantly, when we move to the regime of relatively high S/N and very high S/N (from 100 to 300), our results are very accurate to predict the performance of the LS method, and we find that the gap between the CR and the nominal value is very significant (the deviation from the MVB is 16% and 30% for S/N 120 and 230 at $\Delta x = 0.2$ arcsec, respectively). This last result confirms one of the main findings presented in Lobos et al. (2015), who showed that for medium to high S/N the LS estimator is suboptimal with respect to the MVB. In Fig. 2 we also show the square root of the empirical variance ($Var(\tau_{WLS})$) with respect to the empirically-determined mean position \hat{x}_c (using the WLS estimator), all as derived from the simulations, showing good consistency with our predictions.

5.3.2. Non-uniform weight case

The sub-optimality of the WLS scheme from moderate to high S/N seen in Fig. 2 can be extended for any arbitrary selection of a fixed set of weights, as it would be expected. Given that the space of weights selection is literally unlimited, we use the insight

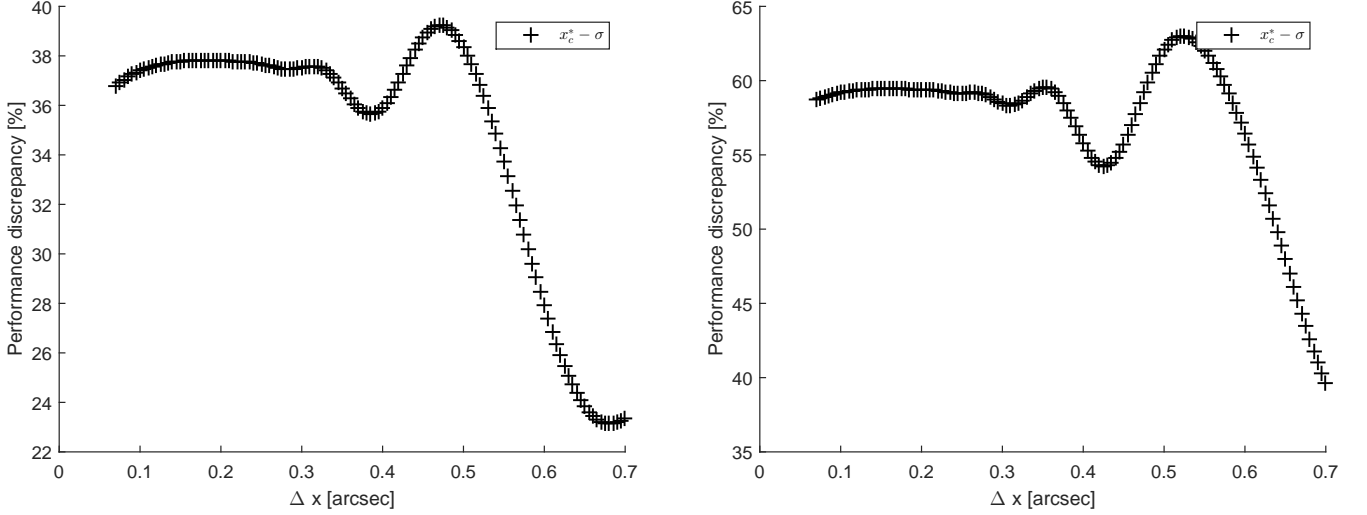


Fig. 3. Worse case discrepancies in Eq. (37) for the WLS estimator using the weights set indexed by the positions $\Theta = \{x_o^* - \sigma, x_o^* - 0.8 * \sigma, x_o^* - 0.6 * \sigma, x_o^* - 0.4 * \sigma, x_o^* - 0.2 * \sigma, x_o^*\}$. Results are reported for two S/N scenarios, namely $\bar{F} = 20004 e^-$ (Left) and $\bar{F} = 60160 e^-$ (Right), and across different pixel sizes.

obtained from Remark 2 that states that a selection of weights can be interpreted as an specific prior on the position of the object where the optimum, but unfortunately unknown, selection (achieving the CR bound) of weights in Eq. (29) is an explicit function of the unknown position of the object. Then, we consider a finite set of positions $\{x_{c,1}, \dots, x_{c,M}\}$ that uniformly partition the field of view, and their corresponding weights sets $\{w_i(x_{c,1}) : 1 = 1, \dots, n\}$, $\{w_i(x_{c,2}) : 1 = 1, \dots, n\} \dots$ and $\{w_i(x_{c,M}) : 1 = 1, \dots, n\}$ using Eq. (29) to cover a representative collections of weights for the problem of astrometry.

Then, for a specific selection of weights in our admissible collection (attributed to a prior belief of the position of the object in the field of view), we evaluate the worse discrepancy between $\sigma_{WLS}^2(n) - \beta_{WLS}(n)$ (which is the most favourable expression for the variance predicted from Theorem 2), and the CR bound in Proposition 1, across a collection of presumed object positions in the following range of positions

$$\Theta = \{x_o^* - \sigma, x_o^* - 0.8 * \sigma, x_o^* - 0.6 * \sigma, x_o^* - 0.4 * \sigma, x_o^* - 0.2 * \sigma, x_o^*\}$$

where x_o^* denotes the center of the array (which, as indicated at the beginning of Sect. 5.3.1, is equal to the true object position x_c) and $\sigma = FWHM/2 \sqrt{2 \ln 2}$ is the dispersion parameter of the PSF. The idea of using this worse case difference is justified from the fact that in this parameter estimation problem we do not know the position of the object, and consequently, the optimality of any WLS estimator should be evaluated in the worse case situation, as the scheme should be able to estimate the position of the object in any scenario (position). More precisely, for a given weight selection $\{w_i, i = 1, \dots, n\}$, we use the following worse case discrepancy

$$\sup_{x_c \in \Theta} \frac{(\sigma_{WLS}^2(n) - \beta_{WLS}(n)) - \sigma_{CR}^2(n)}{\sigma_{CR}^2(n)}. \quad (37)$$

For this analysis note that both $(\sigma_{WLS}^2(n) - \beta_{WLS}(n))$ and $\sigma_{CR}^2(n)$ are functions of the position x_c , Δx , and S/N .

Fig. 3 illustrates the worse case discrepancy in Eq. (37) for the medium and high S/N regimes where Theorem 2 provides an accurate and meaningful prediction of the performance of the WLS method, i.e., for $S/N \in \{120, 230\}$, and across $\Delta x \in [0.05, 0.7]$ arcsec. The discrepancy is quite significant, on the order of 37% and 60% in the range for $\Delta x \in [0.1, 0.3]$ arcsec for S/N 120 and 230, respectively.

To refine the worse case analysis presented in Fig. 3, and to evaluate in more detail the sensitivity of the discrepancy indicator given by Eq. (37), we evaluate the discrepancy of WLS using the weights associated to x_o^* (the center position of the array) with respect to the CR bound associated to the positions $\{x_o^* - \sigma, x_o^* - 0.8 * \sigma, x_o^* - 0.6 * \sigma, x_o^* - 0.4 * \sigma, x_o^* - 0.2 * \sigma, x_o^*\}$, to study how the discrepancy (measuring the non-optimality of the method) evolves when the adopted position moves far from the prior imposed by the WLS in the center of the array. Fig. 4 illustrates this behaviour, where it is possible to see that the discrepancy is very sensitive and proportional to the misassumption of the object position, where the worse case discrepancy in the maximum achievable location precision is on the order of 40% for pixel sizes in the range $[0.1, 0.6]$ arcsec for $S/N \sim 120$, and about 60% for pixel sizes in the range $[0.1, 0.6]$ arcsec for $S/N \sim 230$. These worse case scenario happens in both cases when the object is located the farthest from the prior assumption, as it would be expected.

The main conclusion derived from this CR bound analysis is that, independent of the weight selection adopted, as long as the weights are fixed, the WLS estimator is not able to achieve the CR bound in all observational regimes. More precisely, the discrepancy (measuring the non-optimality) in the less favourable case of an hypothetical and feasible position of the object is very significant, in the range of 40% – 60% for the important regime of high and very high S/N .

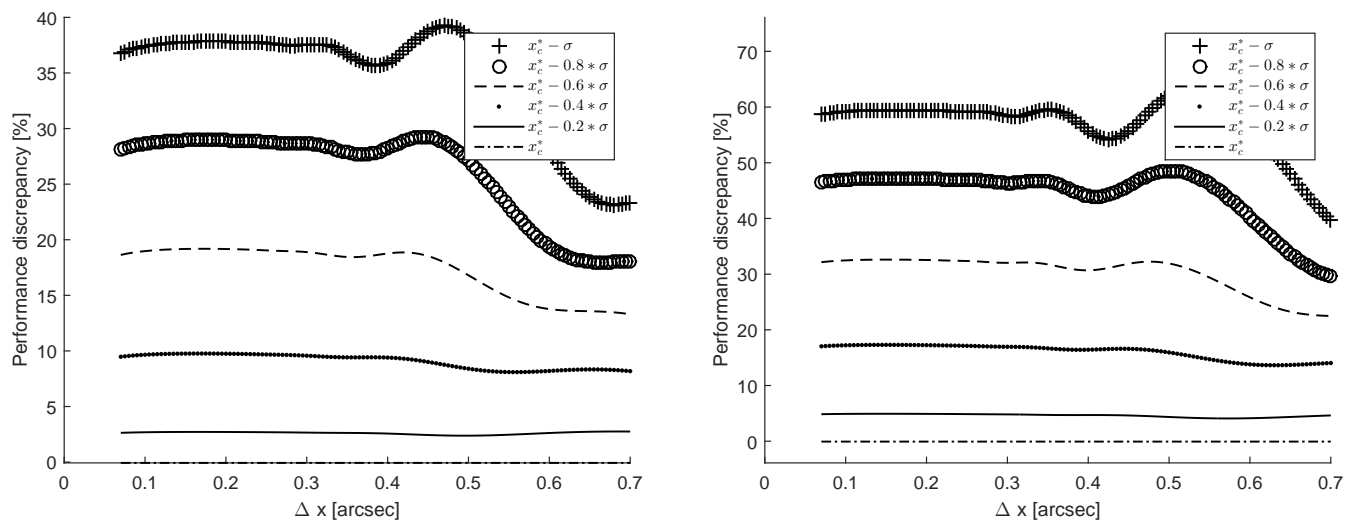


Fig. 4. Performance discrepancies (measuring the non-optimality) of the WLS estimator using the center position as a prior for the weight selection with respect to the CR bound obtained for the true object positions $\{x_o^* - \sigma, x_o^* - 0.8 * \sigma, x_o^* - 0.6 * \sigma, x_o^* - 0.4 * \sigma, x_o^* - 0.2 * \sigma, x_o^*\}$. Results are reported for two S/N scenarios, namely $\bar{F} = 20004 e^-$ (Left) and $\bar{F} = 60160 e^-$ (Right), and across different pixel sizes.

5.4. Performance analysis of the ML estimator

In this section we perform the same analysis done for the WLS in Sect. 5.3, but using the result in Theorem 3. In particular, we consider the admissible regime for the variance of the ML estimator given by

$$\left(\sigma_{ML}^2(n) - \beta_{ML}(n), \sigma_{ML}^2(n) + \beta_{ML}(n)\right)$$

in Eq. (33), where the nominal value in this case, $\sigma_{ML}^2(n)$ in Eq. (34), is precisely the CR bound, while the length of the interval $2\beta_{ML}(n)$ is given by Eqs. (A.2) and (C.9) in closed form.

Considering an object located in the center of the array, i.e., $x_c = x_o$, the performance curves are presented in Fig. 5 for $S/N \in \{12, 32, 120, 230\}$ and $\Delta x \in [0.1, 0.65]$ arcsec. First, we note that there is a significant difference in the predictions of our methodology for the ML estimator in comparison with what we predict in the WLS case. In fact, the results of our approach are very precise for the determination of the variance of the ML estimator in all the regimes, from small to high S/N , which is remarkable. More important it is the fact that, from these findings, we observe that the performance deviation from the MVB in the worse case (small S/N) is very small (see Table 1, first row), while for any practical purposes the variance of the ML estimator achieves the CR limit for all the other regimes, from medium to high S/N , which is a numerical corroboration of the optimality of the ML estimator in astrometry, as predicted theoretically by Theorems 1 and 3. In Fig. 5 we also show the square root of the empirical variance ($\text{Var}(\tau_{ML})$) with respect to the empirically-determined mean position \hat{x}_c (using the ML estimator), all as derived from the simulations, showing good consistency with our predictions.

Complementing this analysis, we conducted the same comparison considering different positions for the object within the array obtaining the same trends and conclusions. To summarize these results, Fig. 6 shows the value $100 \times \frac{\sqrt{\sigma_{ML}^2(n) + \beta_{ML}(n) - \sigma_{ML}(n)}}{\sigma_{ML}(n)}$, which is an indicator of the quality of the estimator (the smaller the better) for different scenarios of the position of the object $x_c \in \Theta = \{x_o^* - \sigma, x_o^* - 0.8 * \sigma, x_o^* - 0.6 * \sigma, x_o^* - 0.4 * \sigma, x_o^* - 0.2 * \sigma, x_o^*\}$. In particular, for all the evaluated positions, the relative discrepancy is bounded (relative to the CR bound) by values in the range of 0.025% and 0.012% for pixel resolution in the range $\Delta x \in [0.1, 0.2]$ arcsec for $S/N = 120$ and $S/N = 230$, respectively. Finally, Table 1 presents the relative discrepancy for all the range of S/N values considered in this study for the case $\Delta x = 0.2$ arcsec.

We conclude from this analysis that the ML estimator is nearly optimal for the medium, high and very high S/N regimes across pixel resolutions, achieving the MVB for astrometry. This is an interesting result, since it lends further support to the adoption of these types of estimators for very demanding astrometric applications, as has been done in the case of Gaia (Lindgren 2008). We note that Vakili & Hogg (2016) reach the same conclusion regarding the optimality of the ML method in comparison with the MVB, through simulations of 2D CCD images using a broad set of Moffat PSF stellar profiles. While their results are purely empirical, it is interesting that they test the ML using a different PSF from ours, and in a 2D scenario, and yet they reach the same conclusions. More recently, Gai et al. (2017) have tested (also empirically) the ML method in a 1D scenario (similar to ours), but in the context of a Gaia-like PSF. They find that the ML is unbiased (in agreement with our results, see Fig. 1), and, by comparing two implementations of the ML they conclude that they predict self-consistent and reliable results over a broad range of flux, background, and instrument response variations. It would still be quite interesting to compare the performance of those implementations against the CR MVB in order to further test our theoretical predictions.

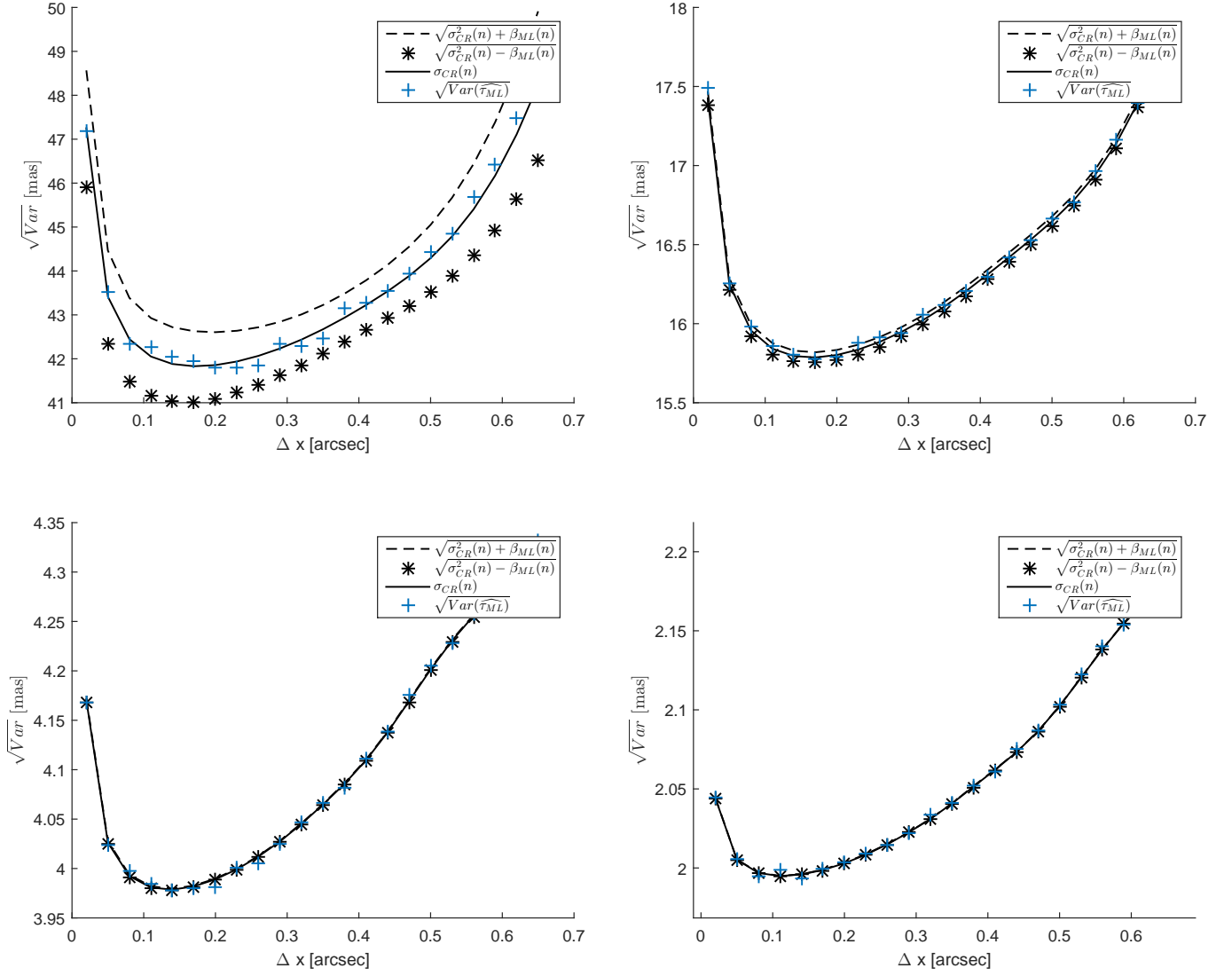


Fig. 5. Range of the square root of the variance performance (in miliarcsecond=mas) for the ML method in astrometry as predicted by Theorems 1 and 3, Eq. (33). Results are reported for different representative values of \tilde{F} and across different pixel sizes (top-left to bottom-right): $\tilde{F} \in \{1080, 3224, 20004, 60160\} e^-$.

Position x_c	$S/N=12$	$S/N=32$	$S/N=120$	$S/N=230$
x_o	3.8%	0.34%	0.032%	0.010%
$x_o - 0.2 * \sigma$	4.1%	0.27%	0.014%	0.009%
$x_o - 0.4 * \sigma$	4.3%	0.19%	0.022%	0.007%
$x_o - 0.6 * \sigma$	3.8%	0.29%	0.019%	0.009%
$x_o - 0.8 * \sigma$	3.9%	0.30%	0.022%	0.011%
$x_o - \sigma$	3.6%	0.40%	0.019%	0.008%

Table 1. Performance quality of the ML estimator relative to the Cramér-Rao bound expressed in terms of the indicator $100 \times \frac{\sqrt{\sigma_{ML}^2(n) + \beta_{ML}(n)} - \sigma_{ML}(n)}{\sigma_{ML}(n)}$ from the result in Theorem 3. The results are presented for different positions of the object $x_c \in \{x_o^* - \sigma, x_o^* - 0.8 * \sigma, x_o^* - 0.6 * \sigma, x_o^* - 0.4 * \sigma, x_o^* - 0.2 * \sigma, x_o^*\}$ (rows) and for different $S/N \in \{12, 32, 120, 230\}$ (columns).

5.5. Comments on an adaptive WLS estimator for astrometry

In Sect. 4.1 we presented results that offer a nominal prediction for the variance of the WLS method through Eq. (27) which turns out to be very accurate in the regime from medium to high S/N as shown in Sect. 5.3. Interestingly, Remark 2 tells us that this nominal values precisely match the CR limit for an optimal selection of weights given in Eq. (29), namely $w_i \sim 1/\lambda_i(x_c)$ for all $i = 1, \dots, n$ (compare Eqs. (28) and (34)). As this selection is unfeasible, because it requires the knowledge of x_c (see the expression

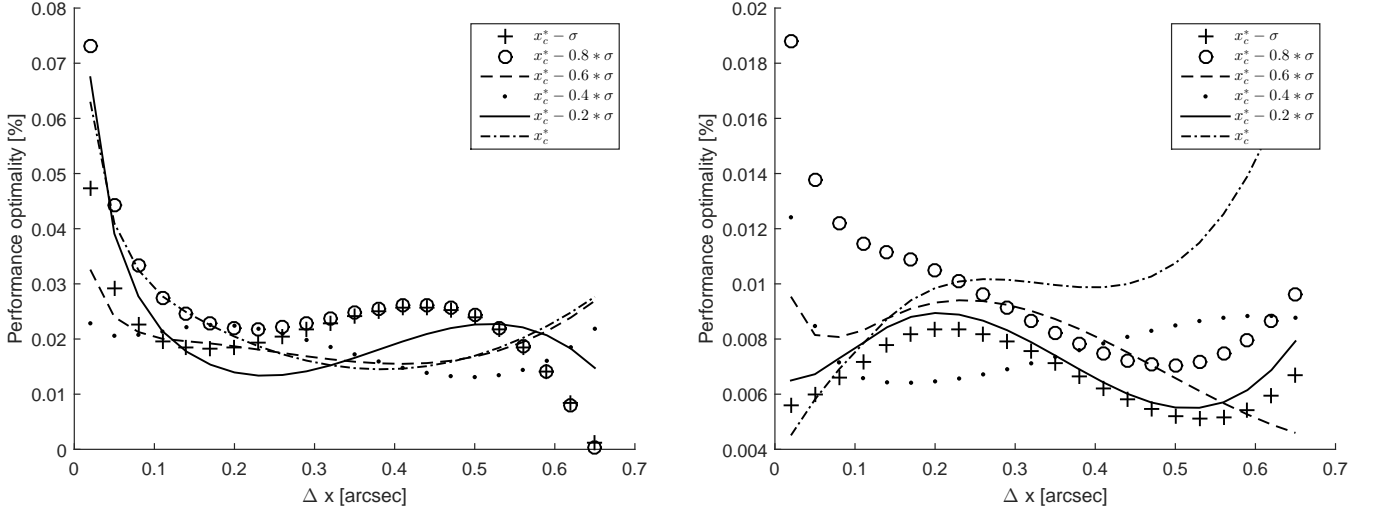


Fig. 6. Performance optimality of the ML estimator (computed as $100 \times \frac{\sqrt{\sigma_{ML}^2(n) + \beta_{ML}(n)} - \sigma_{ML}(n)}{\sigma_{ML}(n)}$) for different positions of the target object $x_c \in \Theta = \{x_o^* - \sigma, x_o^* - 0.8 * \sigma, x_o^* - 0.6 * \sigma, x_o^* - 0.4 * \sigma, x_o^* - 0.2 * \sigma, x_o^*\}$ in the array, as a function of pixel resolution. The left panel shows the case $\tilde{F} = 20004 e^-$, the right panel the case $\tilde{F} = 60160 e^-$.

in Eq. (2)), we can approximate this value by a noisy version of it, considering the fact that the expected value of the observations I_i that we measure at pixel i is $\lambda_i(x_c)$ using our model in Eq. (4). Therefore I_i can be interpreted as a noisy version of $\lambda_i(x_c)$ and

$$\hat{w}_i(I_i) = \frac{1}{I_i} \quad (38)$$

can be seen as a noisy version of the ideal weight $\frac{1}{\lambda_i(x_c)}$. Adopting this data-driven weighting approach, we would have an adaptive WLS method as the weights are not fixed but instead they are a function of the data $\{I_i : i = 1, \dots, n\}$. This selection of weights can be interpreted as an empirical version of the optimal weights that achieves the CR bound. Then, the problem reduces to solve

$$\tau_{AWLS}(I^n) = \arg \min_{\alpha \in \mathbb{R}} J_{AWLS}(\alpha, I^n) \quad (39)$$

where

$$J_{AWLS}(\alpha, I^n) = \sum_{i=1}^n \hat{w}_i(I_i) (I_i - \lambda_i(\alpha))^2. \quad (40)$$

Fig. 7 presents the performance of this scheme for the same regimes we have been exploring in this work, supporting the conjecture that this selection of weights resembles the optimal weight selection and in fact achieves MSE performances that are surprisingly close to the CR bound in all the observational regimes.

Our results in this sub-section show that some commonly adopted weighting schemes, using the analogous to Eq. (38), are a very good data-driven choice for methods that employ a WLS scheme, such is the case, e.g., of the well know PSF stellar fitting program (including astrometry) DAOPHOT, described in Stetson (1987, Eq. (10)).

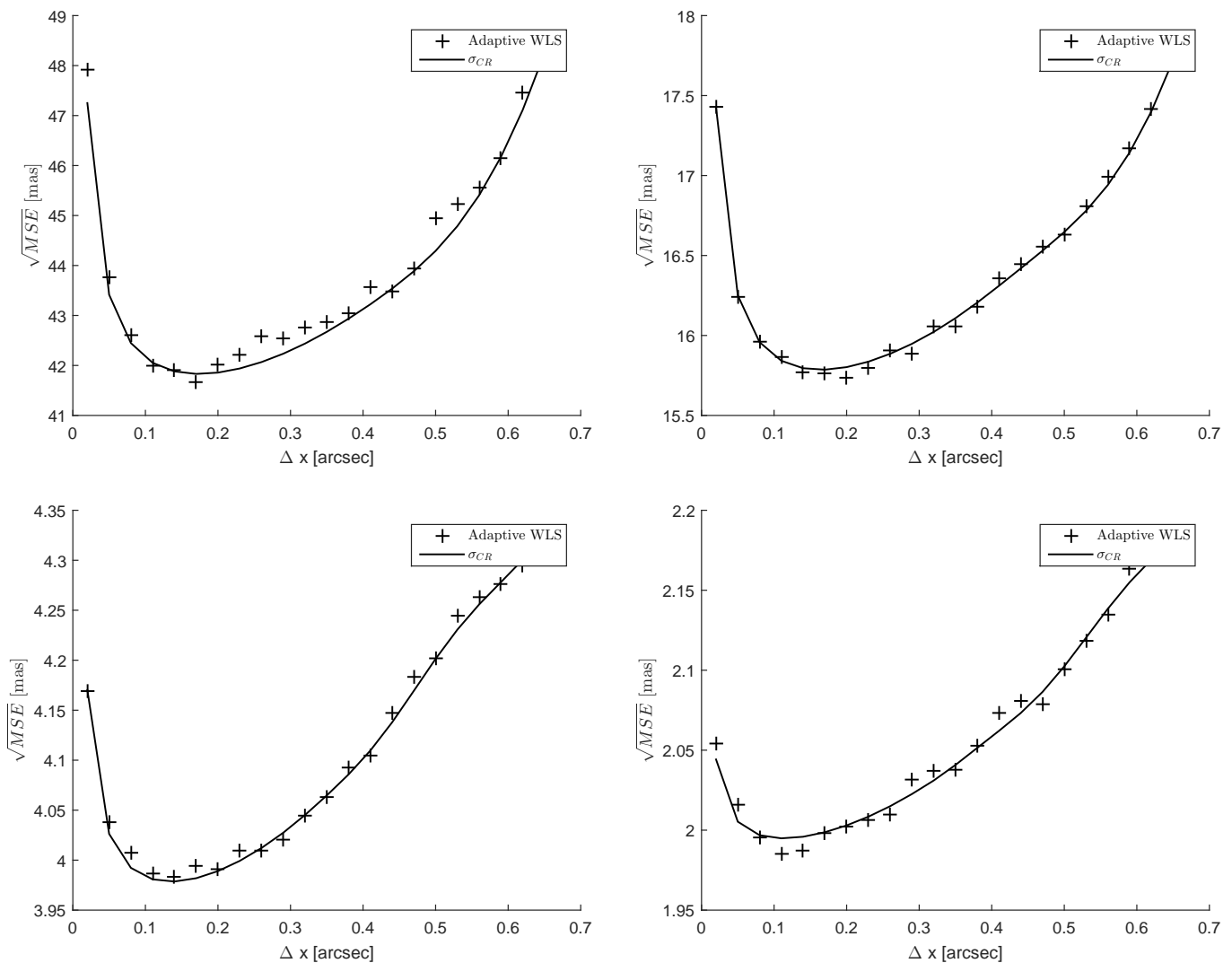


Fig. 7. Performance comparison between the \sqrt{MSE} of the adaptive WLS estimator and $\sigma_{CR}(n)$, both in mas. Results are reported for different \tilde{F} and across different pixel sizes: (top-left to bottom-right) $\tilde{F} \in \{1080, 3224, 20004, 60160\} e^-$.

5.6. Undersampled case

So far we have analyzed the performance of the ML and WLS estimators in regimes where the PSF (assumed to have a $FWHM = 1$ arcsec) is well sampled by the detector. In this section we analyze the undersampled case, where we have only a few pixels that capture most of the flux of the source. This scenario would be typical of nearly diffraction-limited images taken, e.g., from the space, or with an AO system from the ground.

Table 2 shows the performance of the bounds presented in Theorems 2 and 3 in the undersampled regime, where the PSF is concentrated in only a few pixels (order of 5 to 9 pixels), and where the flux is relatively large. For the space-based case, the background is reduced almost to the purely instrument noise (left numbers in the table), while in the ground-based case the background would be larger, but we would still have a very large Strehl ratio. From the table it can be seen that the performance is nearly optimal with the ML estimator, even in the severely undersampled regime (small FWHM values). On the other hand, the WLS deteriorates its performance very quickly as the image size decreases, reaching up to nearly 45% when the FWHM is on the order of the detector pixel size.

We emphasize that the above results are computed in the ideal situation of a uniform pixel response function (and a perfect flat-field correction), whereas in reality a non-uniform pixel response has a critical impact on the Cramér-Rao bound as demonstrated by Adorf (1996, Figure 1). In addition, in the case of stare-mode space-based images acquired under very small background, charge transfer inefficiency effects due to time-varying charge-traps in the detector would also blur (and systematically shift) the images, deteriorating the ultimate astrometric precision above the Cramér-Rao limit (for details, see, e.g., Bristow et al. (2006), especially their Figure 10).

FWHM	ML error [%]	WLS error [%]
0.2	0.03 / 0.02	25 / 45
0.3	0.03 / 0.02	5.5 / 8.0
0.4	0.01 / 0.03	1.9 / 2.8
0.5	0.01 / 0.04	1.4 / 2.3

Table 2. Performance quality of the ML and WLS estimators relative to the nominal bound expressed in terms of the indicator $100 \times \frac{\sqrt{\sigma_{ML}^2(n) + \beta_{ML}(n)} - \sigma_{ML}(n)}{\sigma_{ML}(n)}$ and $100 \times \frac{\sqrt{\sigma_{WLS}^2(n) + \beta_{WLS}(n)} - \sigma_{WLS}(n)}{\sigma_{WLS}(n)}$, respectively in the undersampled regime (space based and ground based are on the left and right, respectively). The results are presented for different values of $FWHM$ (measured in arcsec) with $\tilde{F} = 20004e^-$, $\tilde{B} = 25e^-$ (space based), $\tilde{B} = 626e^-$ (ground based) and $\Delta x = 0.2$ arcsec.

6. Summary, conclusions, and outlook

In this paper we study the performance of the WLS and ML estimators for relative astrometry on digital detectors subject to Poisson noise, in comparison with the best possible attainable precision given by the CR bound. Our study includes analytical results, and numerical simulations under realistic observational conditions to help us to corroborate our theoretical findings.

By generalizing the proposal by Fessler (1996) we are able to obtain, for the first time, close-form expressions for the variance and the mean of implicit estimators (as is the particular case of the WLS and ML schemes), which can be computed directly from the data (see Theorem 1, in particular Eqs. (21) and (22), and Appendix A). When specifying this result to astrometry with digital detectors, we are able to bound both the bias and the variance of the relative position of a celestial source on a CCD array as a function of all the relevant parameters of the problem (see Eqs. (26) and (27) or Eqs. (32) and (33) for the WLS and ML estimators respectively). We verified that the bias of the WLS and ML methods are negligible in all the observational regimes explored in this paper (see Fig. 1).

A careful analysis of our predictions confirms earlier results by Lobos et al. (2015) (for the LS method) in that the WLS method is, in general, sub-optimal (in comparison with the MVB given by the CR result), specially at high and very high S/N (see the two bottom panels on Fig. 2). However, a judicious data-driven selection of weights (called “adaptive” WLS method by us, Sect. (5.5)), improves the performance of the WLS substantially (see Fig. 7). This is an interesting result, given the widespread use and simple numerical implementation of the WLS method.

The ML method is found to have both a smaller bias than the WLS method (compare left and right panels of Fig. 1), although the bias on both methods is already quite small, and a tight correspondence to the MVB throughout the entire range of S/N regimes explored in this paper (Fig. 5). Therefore, the ML estimator for astrometry is consistently optimal, and should be the estimator of choice for high-precision applications.

This paper, along with Lobos et al. (2015), completes an in-depth study of the performance of commonly used estimators in astrometry using PIDs, and sets the stage for the development of codes that could efficiently implement astrometric ML estimators on 2D detectors, incorporating also the simultaneous measurement of fluxes, as explored by Gai et al. (2017).

7. Acknowledgments

SE acknowledges support from CONICYT-PCHA/MagisterNacional/2016 - 22160840. The authors want to thank the support of the Advanced Center for Electrical and Electronic Engineering, AC3E, Basal Project FB0008, PIA ACT1405, and the Chilean Centro de Excelencia en Astrofísica y Tecnologías Afines (CATA) BASAL PFB/06 from CONICYT. The authors also acknowledge support by CONICYT/FONDECYT grants # 1151213, 1170044, and 1170854. RAM also acknowledges support from the Project IC120009 Millennium Institute of Astrophysics (MAS) of the Iniciativa Científica Milenio del Ministerio de Economía, Fomento y Turismo de Chile. Finally, we thank an anonymous referee for his/her careful reading of our manuscript, and for pointing out the exploration of the relevant space-based undersampled regime that led to the discussion in Section 5.6.

Appendix A: Proof of Theorem 1

We begin presenting the expressions for $(\epsilon_j(n), \beta_j(n), \sigma_j^2(n))$ to complete the statement of the result.

$$\epsilon_j(n) = \max_{t \in [0,1]} \left| \mathbb{E}_{I^n \sim f_{sc}} \left\{ \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \frac{\partial^2}{\partial I_i \partial I_j} \tau_J(\bar{I}^n - t(I^n - \bar{I}^n))(I_i - \bar{I}_i)(I_j - \bar{I}_j) \right\} \right|, \quad (\text{A.1})$$

$$\beta_j(n) = \epsilon'_j(n) + 2\delta'_j(n), \quad (\text{A.2})$$

where

$$\epsilon'_j(n) = \max_{t \in [0,1]} \mathbb{E}_{I^n \sim f_{sc}} \left\{ \left(\frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \frac{\partial^2}{\partial I_i \partial I_j} \tau_J(\bar{I}^n - t(I^n - \bar{I}^n))(I_i - \bar{I}_i)(I_j - \bar{I}_j) \right)^2 \right\}, \quad (\text{A.3})$$

$$\delta'_j(n) = \max_{t \in [0,1]} \left| \mathbb{E}_{I^n \sim f_{sc}} \left\{ (\nabla \tau_J(\bar{I}^n) \cdot (I^n - \bar{I}^n)) \cdot \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \frac{\partial^2}{\partial I_i \partial I_j} \tau_J(\bar{I}^n - t(I^n - \bar{I}^n))(I_i - \bar{I}_i)(I_j - \bar{I}_j) \right\} \right|, \quad (\text{A.4})$$

and, finally,

$$\sigma_j^2(n) = \left[\frac{\partial^2 J(\tau_J(\bar{I}^n), \bar{I}^n)}{\partial \alpha^2} \right]^{-1} \left[\frac{\partial^2 J(\tau_J(\bar{I}^n), \bar{I}^n)}{\partial \alpha \partial I_i} \right] Cov\{I^n\} \left[\frac{\partial^2 J(\tau_J(\bar{I}^n), \bar{I}^n)}{\partial \alpha \partial I_i} \right]^T \left[\frac{\partial^2 J(\tau_J(\bar{I}^n), \bar{I}^n)}{\partial \alpha^2} \right]^{-1}. \quad (\text{A.5})$$

Proof of Theorem 1: Using the chain rule in the cost function $J(\alpha, I^n)$ and taking the partial derivative $\frac{\partial}{\partial I_i}$ of both sides in (15), we have that

$$0 = \frac{\partial^2}{\partial \alpha^2} J(\tau(I^n), I^n) \frac{\partial}{\partial I_i} \tau(I^n) + \frac{\partial^2}{\partial \alpha \partial I_i} J(\tau(I^n), I^n), \quad i = 1, \dots, n. \quad (\text{A.6})$$

Thus, we have n equations with one unknown, and it holds for any I^n . Defining the operators

$$\nabla^{20}(\cdot) = \frac{\partial^2}{\partial \alpha^2}, \quad \nabla^{11}(\cdot) = \frac{\partial^2}{\partial \alpha \partial I_i}, \quad (\text{A.7})$$

of dimensions 1×1 and $1 \times n$, respectively, we can express (A.6) in matrix form as

$$0 = \nabla^{20} J(\tau(I^n), I^n) \nabla \tau(I^n) + \nabla^{11} J(\tau(I^n), I^n). \quad (\text{A.8})$$

Assuming that the matrix $\nabla^{20} J(\tau(I^n), I^n)$ is non singular, we can calculate $\nabla \tau(I^n)$ from (A.8)

$$\nabla \tau(I^n) = -[\nabla^{20} J(\tau(I^n), I^n)]^{-1} \nabla^{11} J(\tau(I^n), I^n). \quad (\text{A.9})$$

Finally, using (A.9), evaluating at \bar{I}^n , and then replacing in (18), we have that

$$\begin{aligned} \sigma_j^2(n) &= -[\nabla^{20} J(\tau(\bar{I}^n), \bar{I}^n)]^{-1} \nabla^{11} J(\tau(\bar{I}^n), \bar{I}^n) Cov\{I^n\} (-[\nabla^{20} J(\tau(\bar{I}^n), \bar{I}^n)]^{-1} \nabla^{11} J(\tau(\bar{I}^n), \bar{I}^n))^T \\ &= [\nabla^{20} J(\tau(\bar{I}^n), \bar{I}^n)]^{-1} \nabla^{11} J(\tau(\bar{I}^n), \bar{I}^n) Cov\{I^n\} [\nabla^{11} J(\tau(\bar{I}^n), \bar{I}^n)]^T [\nabla^{20} J(\tau(\bar{I}^n), \bar{I}^n)]^{-1}. \\ &= \left[\frac{\partial^2 J(\tau(\bar{I}^n), \bar{I}^n)}{\partial \alpha^2} \right]^{-1} \left[\frac{\partial^2 J(\tau(\bar{I}^n), \bar{I}^n)}{\partial \alpha \partial I_i} \right] Cov\{I^n\} \left[\frac{\partial^2 J(\tau(\bar{I}^n), \bar{I}^n)}{\partial \alpha \partial I_i} \right]^T \left[\frac{\partial^2 J(\tau(\bar{I}^n), \bar{I}^n)}{\partial \alpha^2} \right]^{-1}. \end{aligned} \quad (\text{A.10})$$

Moving into the residual term $\gamma_j(n)$ in (18) captured by $\beta_j(n)$, we must consider the variance of the error function $Var\{e(\bar{I}^n, I^n - \bar{I}^n)\}$ and the covariance $Cov\{\nabla \tau_J(\bar{I}^n)(I^n - \bar{I}^n), e(\bar{I}^n, I^n - \bar{I}^n)\}$. For the first, we have that

$$\begin{aligned} Var\{e(\bar{I}^n, I^n - \bar{I}^n)\} &= Var \left\{ \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \frac{\partial^2 \tau}{\partial I_i \partial I_j} (\bar{I}^n + t(I^n - \bar{I}^n))(I_i - \bar{I}_i)(I_j - \bar{I}_j) \right\} \\ &\leq \mathbb{E} \left\{ \left(\frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \frac{\partial^2 \tau}{\partial I_i \partial I_j} (\bar{I}^n + t(I^n - \bar{I}^n))(I_i - \bar{I}_i)(I_j - \bar{I}_j) \right)^2 \right\} \\ &\leq \max_{t \in [0,1]} \mathbb{E} \left\{ \underbrace{\left(\frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \frac{\partial^2 \tau}{\partial I_i \partial I_j} (\bar{I}^n + t(I^n - \bar{I}^n))(I_i - \bar{I}_i)(I_j - \bar{I}_j) \right)^2}_{=\epsilon'_j(n)} \right\} \end{aligned} \quad (\text{A.11})$$

On the other hand, for the covariance, using the main assumption in (19), it is clear that

$$\mathbb{E}_{I^n \sim f_{x_c}} \left\{ \nabla \tau_{ML}(\bar{I}^n) \cdot (I^n - \bar{I}^n) \right\} = \mathbb{E}_{I^n \sim f_{x_c}} \left\{ a \sum_{i=1}^n b_i (I_i - \bar{I}_i) \right\} = 0. \quad (\text{A.12})$$

From this,

$$\begin{aligned} & |Cov\{\nabla \tau(\bar{I}^n)(I^n - \bar{I}^n), e(\bar{I}^n, I^n - \bar{I}^n)\}| \\ &= \left| \mathbb{E} \left\{ \nabla \tau(\bar{I}^n)(I^n - \bar{I}^n) \left(e(\bar{I}^n, I^n - \bar{I}^n) - \mathbb{E} \left(e(\bar{I}^n, I^n - \bar{I}^n) \right) \right) \right\} \right| \\ &= \left| \mathbb{E}_{I^n \sim f_{x_c}} \left\{ (\nabla \tau(\bar{I}^n) \cdot (I^n - \bar{I}^n)) \cdot \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \frac{\partial^2}{\partial I_i \partial I_j} \tau(\bar{I}^n - t(I^n - \bar{I}^n))(I_i - \bar{I}_i)(I_j - \bar{I}_j) \right\} \right| \\ &\leq \underbrace{\max_{t \in [0,1]} \left| \mathbb{E}_{I^n \sim f_{x_c}} \left\{ (\nabla \tau(\bar{I}^n) \cdot (I^n - \bar{I}^n)) \cdot \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \frac{\partial^2}{\partial I_i \partial I_j} \tau(\bar{I}^n - t(I^n - \bar{I}^n))(I_i - \bar{I}_i)(I_j - \bar{I}_j) \right\} \right|}_{=\delta'_J(n)} \end{aligned} \quad (\text{A.13})$$

Finally, replacing (A.11) and (A.13) in the definition of $\gamma_J(n)$, we have that:

$$\begin{aligned} |\gamma_J(n)| &\leq Var\{e(\bar{I}^n, I^n - \bar{I}^n)\} + 2|Cov\{\nabla \tau(\bar{I}^n)(I^n - \bar{I}^n), e(\bar{I}^n, I^n - \bar{I}^n)\}| \\ &\leq \epsilon'_J(n) + 2\delta'_J(n) = \beta_J(n). \end{aligned} \quad (\text{A.14})$$

For the bias expression of the result in (21), using the hypothesis in (20), we can take expectation at both sides of (17) to obtain that

$$\begin{aligned} |\mathbb{E}_{I^n \sim f_{x_c}} \{\tau(I^n)\} - x_c| &= \left| \mathbb{E}_{I^n \sim f_{x_c}} \left\{ a \sum_{i=1}^N b_i (I_i - \bar{I}_i) + e(\bar{I}^n, I^n - \bar{I}^n) \right\} \right| \\ &= \left| \mathbb{E}_{I^n \sim f_{x_c}} \left\{ e(\bar{I}^n, I^n - \bar{I}^n) \right\} \right| \\ &= \left| \mathbb{E}_{I^n \sim f_{x_c}} \left\{ \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \frac{\partial^2}{\partial I_i \partial I_j} \tau(\bar{I}^n - t(I^n - \bar{I}^n))(I_i - \bar{I}_i)(I_j - \bar{I}_j) \right\} \right| \\ &\leq \underbrace{\max_{t \in [0,1]} \left| \mathbb{E}_{I^n \sim f_{x_c}} \left\{ \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \frac{\partial^2}{\partial I_i \partial I_j} \tau(\bar{I}^n - t(I^n - \bar{I}^n))(I_i - \bar{I}_i)(I_j - \bar{I}_j) \right\} \right|}_{=\epsilon_J(n)}. \end{aligned} \quad (\text{A.15})$$

Appendix B: Proof of Theorem 2

Proof: The proof and, in particular, the derivation of $\sigma_{WLS}^2(n)$, $\beta_{WLS}(n)$ and $\epsilon_{WLS}(n)$ simply reduces to an straightforward application of Theorem 1. For that we need to first validate the assumptions of Theorem 1. If we begin with Eq. (A.9)

$$\nabla \tau(I^n) = -[\nabla^{20} J(\tau_J(I^n), I^n)]^{-1} \nabla^{11} J(\tau_J(I^n), I^n), \quad (\text{B.1})$$

and then we calculate the gradient terms in the RHS of (B.1) for our WLS context, it follows that

$$\begin{aligned} \nabla^{20} J_{WLS}(\alpha, I^n) &= \frac{\partial^2}{\partial \alpha^2} J_{WLS}(\alpha, I^n) \\ &= 2 \sum_{i=1}^n w_i \left(\left(\frac{\partial \lambda_i(\alpha)}{\partial \alpha} \right)^2 + (\lambda_i(\alpha) - I_i) \frac{\partial^2 \lambda_i(\alpha)}{\partial \alpha^2} \right), \end{aligned} \quad (\text{B.2})$$

$$\nabla^{11} J_{WLS}(\alpha, I^n) = \left(\frac{\partial^2}{\partial \alpha \partial I_1} J_{WLS}(\alpha, I^n), \dots, \frac{\partial^2}{\partial \alpha \partial I_n} J_{WLS}(\alpha, I^n) \right)^T \quad (\text{B.3})$$

$$= -2 \left(w_1 \frac{\partial \lambda_1(\alpha)}{\partial \alpha}, \dots, w_n \frac{\partial \lambda_n(\alpha)}{\partial \alpha} \right)^T. \quad (\text{B.4})$$

Following (B.1), we need to evaluate Eqs. (B.2) and (B.4) at $\alpha = \tau_{WLS}(\bar{I}^n)$. For that, we have the following¹²

$$\tau_{WLS}(\bar{I}^n) = \underset{\alpha \in \mathbb{R}}{\operatorname{argmin}} \sum_{i=1}^n w_i (\lambda_i(x_c) - \lambda_i(\alpha))^2. \quad (\text{B.5})$$

¹² considering that $\bar{I}_i = E\{I_i\} = \lambda_i(x_c)$.

Then we will use the following result:

Proposition 2. *Under the assumption of a Gaussian PSF, $\tau_{WLS}(\bar{I}^n) = x_c$.*

Notice that this proposition is the second assumption used in Theorem 1. Using this proposition, we obtain that

$$\begin{aligned} \nabla^{20} J_{WLS}(\tau_{WLS}(\bar{I}^n), \bar{I}^n) &= 2 \sum_{i=1}^n w_i \left(\left(\frac{\partial \lambda_i(\alpha)}{\partial \alpha} \right)^2 \Big|_{\alpha=x_c} + (\lambda_i(x_c) - \lambda_i(x_c)) \frac{\partial^2 \lambda_i(\alpha)}{\partial \alpha^2} \Big|_{\alpha=x_c} \right), \\ &= 2 \sum_{i=1}^n w_i \left(\frac{\partial \lambda_i(\alpha)}{\partial \alpha} \right)^2 \Big|_{\alpha=x_c}, \end{aligned} \quad (\text{B.6})$$

$$\begin{aligned} \nabla^{11} J_{WLS}(\tau_{WLS}(\bar{I}^n), \bar{I}^n) &= -2 \left(w_1 \frac{\partial \lambda_1(\alpha)}{\partial \alpha}, \dots, w_n \frac{\partial \lambda_n(\alpha)}{\partial \alpha} \right)^T \Big|_{\alpha=\tau_{WLS}(\bar{I}^n)}, \\ &= -2 \left(w_1 \frac{\partial \lambda_1(\alpha)}{\partial \alpha}, \dots, w_n \frac{\partial \lambda_n(\alpha)}{\partial \alpha} \right)^T \Big|_{\alpha=x_c}. \end{aligned} \quad (\text{B.7})$$

Finally, applying (B.6) and (B.7) in (A.9) we have that

$$\begin{aligned} \nabla \tau_{WLS}(\bar{I}^n) \cdot (I^n - \bar{I}^n) &= -[\nabla^{20} J(\tau_{WLS}(\bar{I}^n), \bar{I}^n)]^{-1} [\nabla^{11} J(\tau_{WLS}(\bar{I}^n), \bar{I}^n)] (I^n - \bar{I}^n) \\ &= \underbrace{\left[\sum_{i=1}^n w_i \left(\frac{\partial \lambda_i(\alpha)}{\partial \alpha} \right)^2 \Big|_{\alpha=x_c} \right]^{-1}}_a \cdot \underbrace{\sum_{j=1}^n w_j \frac{\partial \lambda_j(\alpha)}{\partial \alpha} \Big|_{\alpha=x_c}}_{b_j} (I_j - \mathbb{E}(I_j)), \end{aligned} \quad (\text{B.8})$$

which offers the decomposition needed for the application of Theorem 1 (Eq. (19)). For the value of $\sigma_{WLS}^2(n)$ in (A.5), since the observations are independent and follow a Poisson distribution, we have that

$$\text{Cov}\{I_i, I_j\} = \begin{cases} \text{Var}\{I_i\} = \lambda_i(x_c), & \text{if } i = j, \\ 0 & \sim. \end{cases} \quad (\text{B.9})$$

Then if we replace (B.6), (B.7) and (B.9) in (A.5), we have that

$$\sigma_{WLS}^2(n) = \frac{\sum_{i=1}^n w_i^2 \lambda_i(x_c) \left(\frac{\partial \lambda_i(\alpha)}{\partial \alpha} \right)^2 \Big|_{\alpha=x_c}}{\left(\sum_{i=1}^n w_i \left(\frac{\partial \lambda_i(\alpha)}{\partial \alpha} \right)^2 \Big|_{\alpha=x_c} \right)^2}. \quad (\text{B.10})$$

On the other hand, the expression for $\beta_{WLS}(n)$ and $\epsilon_{WLS}(n)$ can be determined from the evaluation of (A.2) and (A.1), respectively. Looking at them, the problem reduces to determine the key term $\frac{\partial^2 \tau_{WLS}}{\partial I_i \partial I_j}(\bar{I}^n + t(I^n - \bar{I}^n))$. For that, if we use (Fessler 1996, Eq. (17)) we can obtain the following identity¹³

$$\begin{aligned} \frac{\partial^2 \tau_{WLS}}{\partial I_i \partial I_j}(\bar{I}^n + t(I^n - \bar{I}^n)) &= \frac{-1}{\left[\sum_{i=1}^n \frac{\partial^2 \lambda_i(\alpha)}{\partial \alpha^2} \cdot (\lambda_i(\alpha) - (\bar{I}_i + t(I_i - \bar{I}_i))) 2w_i + 2w_i \left(\frac{\partial \lambda_i(\alpha)}{\partial \alpha} \right)^2 \right]^2} \\ &\quad \left[\sum_{i=1}^n \frac{\partial^3 \lambda_i(\alpha)}{\partial \alpha^3} \cdot (\lambda_i(\alpha) - (\bar{I}_i + t(I_i - \bar{I}_i))) 2w_i + 6w_i \frac{\partial^2 \lambda_i(\alpha)}{\partial \alpha^2} \frac{\partial \lambda_i(\alpha)}{\partial \alpha} \right] \\ &\quad \left[\frac{(2w_j \frac{\partial \lambda_j(\alpha)}{\partial \alpha})}{\left[\sum_{i=1}^n \frac{\partial^2 \lambda_i(\alpha)}{\partial \alpha^2} \cdot (\lambda_i(\alpha) - (\bar{I}_i + t(I_i - \bar{I}_i))) 2w_i - 2w_i \left(\frac{\partial \lambda_i(\alpha)}{\partial \alpha} \right)^2 \right]} - \left(2w_j \frac{\partial^2 \lambda_j(\alpha)}{\partial \alpha^2} \right) \right] \\ &\quad \left(2w_i \frac{\partial \lambda_i(\alpha)}{\partial \alpha} \right) - \left(2w_i \frac{\partial^2 \lambda_i(\alpha)}{\partial \alpha^2} \right) \cdot \left(2w_j \frac{\partial \lambda_j(\alpha)}{\partial \alpha} \right) \Big|_{\alpha=\tau_{WLS}(\bar{I}^n + t(I^n - \bar{I}^n))}, \end{aligned} \quad (\text{B.11})$$

which concludes the result.

¹³ The derivation of this identity is presented in Appendix B.2.

Appendix B.1: Proof of Proposition 2

Proof: Using the function $h(\alpha) = \sum_{i=1}^n w_i (\lambda_i(x_c) - \lambda_i(\alpha))^2$, we need to show that the minimum is reached only at $\alpha = x_c$. From this, we have that $h(\alpha) \geq 0$ and it achieves its minimum at x_c . To prove uniqueness, let us assume that there is another position $x_c^* \neq x_c$ at which h is zero. Then

$$\begin{aligned} h(x_c^*) &= \sum_{i=1}^n w_i (\lambda_i(x_c) - \lambda_i(x_c^*))^2 = 0 \\ \Leftrightarrow \lambda_i(x_c) &= \lambda_i(x_c^*), \quad \forall i \in \{1, \dots, n\}. \end{aligned} \quad (\text{B.12})$$

The last identity is not possible, because if we use a Gaussian PSF there is at least one $i \in \{1, \dots, n\}$ such that $\lambda_i(x_c) \neq \lambda_i(x_c^*)$.

Appendix B.2: Proof of Eq. (B.11)

Proof: Recall (Fessler 1996, Eq. (17)) and considering $J_{WLS}(\alpha, I^n)$ as the cost function we have that

$$\begin{aligned} \frac{\partial^2 \tau_{WLS}}{\partial I_i \partial I_j}(\bar{I}^n + t(I^n - \bar{I}^n)) &= \left[-\frac{\partial^2 J_{WLS}(\alpha, \bar{I}^n + t(I^n - \bar{I}^n))}{\partial \alpha^2} \right]^{-1} \left(\left[\frac{\partial^3 J_{WLS}(\alpha, \bar{I}^n + t(I^n - \bar{I}^n))}{\partial \alpha^3} \cdot \frac{\partial \tau_{WLS}(\bar{I}^n + t(I^n - \bar{I}^n))}{\partial I_j} \right. \right. \\ &\quad \left. \left. + \frac{\partial^3 J_{WLS}(\alpha, \bar{I}^n + t(I^n - \bar{I}^n))}{\partial \alpha^2 \partial I_j} \right] \cdot \frac{\partial \tau_{WLS}(\bar{I}^n + t(I^n - \bar{I}^n))}{\partial I_i} \right. \\ &\quad \left. + \frac{\partial^3 J_{WLS}(\alpha, \bar{I}^n + t(I^n - \bar{I}^n))}{\partial \alpha^2 \partial I_i} \cdot \frac{\partial \tau_{WLS}(\bar{I}^n + t(I^n - \bar{I}^n))}{\partial I_j} + \frac{\partial^3 J_{WLS}(\alpha, \bar{I}^n + t(I^n - \bar{I}^n))}{\partial \alpha \partial I_i \partial I_j} \right) \Bigg|_{\alpha=\tau_{WLS}(\bar{I}^n + t(I^n - \bar{I}^n))}, \end{aligned} \quad (\text{B.13})$$

where from the definition of $J_{ML}(\alpha, I^n)$ we have that

$$\frac{\partial^2 J_{WLS}(\alpha, \bar{I}^n + t(I^n - \bar{I}^n))}{\partial \alpha^2} = \sum_{i=1}^n \frac{\partial^2 \lambda_i(\alpha)}{\partial \alpha^2} \cdot (\lambda_i(\alpha) - (\bar{I}_i + t(I_i - \bar{I}_i)))2w_i + 2w_i \left(\frac{\partial \lambda_i(\alpha)}{\partial \alpha} \right)^2, \quad (\text{B.14})$$

$$\frac{\partial^3 J_{WLS}(\alpha, \bar{I}^n + t(I^n - \bar{I}^n))}{\partial \alpha^3} = \sum_{i=1}^n \frac{\partial^3 \lambda_i(\alpha)}{\partial \alpha^3} \cdot (\lambda_i(\alpha) - (\bar{I}_i + t(I_i - \bar{I}_i)))2w_i + 6w_i \frac{\partial^2 \lambda_i(\alpha)}{\partial \alpha^2} \frac{\partial \lambda_i(\alpha)}{\partial \alpha}, \quad (\text{B.15})$$

$$\frac{\partial^3 J_{WLS}(\alpha, \bar{I}^n + t(I^n - \bar{I}^n))}{\partial \alpha^2 \partial I_i} = - \left(2w_i \frac{\partial^2 \lambda_i(\alpha)}{\partial \alpha^2} \right), \quad (\text{B.16})$$

$$\frac{\partial^3 J_{WLS}(\alpha, \bar{I}^n + t(I^n - \bar{I}^n))}{\partial \alpha \partial I_i \partial I_j} = 0. \quad (\text{B.17})$$

Concerning $\frac{\partial \tau_{ML}(\bar{I}^n + t(I^n - \bar{I}^n))}{\partial I_i}$ it is just the i -th component of the gradient in Eq. (B.1), then we use (B.2) and (B.4)

$$\begin{aligned} \frac{\partial \tau_{WLS}(\bar{I}^n + t(I^n - \bar{I}^n))}{\partial I_i} &= \frac{-\frac{\partial^2 J_{WLS}(\alpha, \bar{I}^n + t(I^n - \bar{I}^n))}{\partial \alpha \partial I_i}}{\frac{\partial^2 J_{WLS}(\alpha, \bar{I}^n + t(I^n - \bar{I}^n))}{\partial \alpha^2}} \\ &= \frac{\left(2w_i \frac{\partial \lambda_i(\alpha)}{\partial \alpha} \right)}{\sum_{i=1}^n \frac{\partial^2 \lambda_i(\alpha)}{\partial \alpha^2} \cdot (\lambda_i(\alpha) - (\bar{I}_i + t(I_i - \bar{I}_i)))2w_i + 2w_i \left(\frac{\partial \lambda_i(\alpha)}{\partial \alpha} \right)^2}. \end{aligned} \quad (\text{B.18})$$

Finally, replacing (B.14), (B.15), (B.16), (B.17) and (B.18) in (B.13), and evaluating in $\alpha = \tau_{WLS}(\bar{I}^n + t(I^n - \bar{I}^n))$ we obtain the desired result

$$\begin{aligned} \frac{\partial^2 \tau_{WLS}}{\partial I_i \partial I_j}(\bar{I}^n + t(I^n - \bar{I}^n)) &= \frac{-1}{\left[\sum_{i=1}^n \frac{\partial^2 \lambda_i(\alpha)}{\partial \alpha^2} \cdot (\lambda_i(\alpha) - (\bar{I}_i + t(I_i - \bar{I}_i)))2w_i + 2w_i \left(\frac{\partial \lambda_i(\alpha)}{\partial \alpha} \right)^2 \right]^2} \\ &\quad \left[\left[\sum_{i=1}^n \frac{\partial^3 \lambda_i(\alpha)}{\partial \alpha^3} \cdot (\lambda_i(\alpha) - (\bar{I}_i + t(I_i - \bar{I}_i)))2w_i + 6w_i \frac{\partial^2 \lambda_i(\alpha)}{\partial \alpha^2} \frac{\partial \lambda_i(\alpha)}{\partial \alpha} \right] \cdot \right. \\ &\quad \left. \frac{\left(2w_j \frac{\partial \lambda_j(\alpha)}{\partial \alpha} \right)}{\left[\sum_{i=1}^n \frac{\partial^2 \lambda_i(\alpha)}{\partial \alpha^2} \cdot (\lambda_i(\alpha) - (\bar{I}_i + t(I_i - \bar{I}_i)))2w_i - 2w_i \left(\frac{\partial \lambda_i(\alpha)}{\partial \alpha} \right)^2 \right]} \right] - \left(2w_j \frac{\partial^2 \lambda_j(\alpha)}{\partial \alpha^2} \right) \\ &\quad \left. \left(2w_i \frac{\partial \lambda_i(\alpha)}{\partial \alpha} \right) - \left(2w_i \frac{\partial^2 \lambda_i(\alpha)}{\partial \alpha^2} \right) \cdot \left(2w_j \frac{\partial \lambda_j(\alpha)}{\partial \alpha} \right) \right] \Bigg|_{\alpha=\tau_{WLS}(\bar{I}^n + t(I^n - \bar{I}^n))}. \end{aligned} \quad (\text{B.19})$$

Appendix C: Proof of Theorem 3

Proof: Again the proof and the derivation of $\sigma_{ML}^2(n)$, $\beta_{ML}(n)$ and $\epsilon_{ML}(n)$ reduce to apply Theorem 1. First, we need to validate the assumption of Theorem 1. Beginning with the equality in Eq. (B.1), it follows that

$$\begin{aligned}\nabla^{20} J_{ML}(\alpha, I^n) &= \frac{\partial^2}{\partial \alpha^2} J_{ML}(\alpha, I^n) \\ &= -\sum_{i=1}^n I_i \frac{1}{\lambda_i^2(\alpha)} \left(\frac{\partial \lambda_i(\alpha)}{\partial \alpha} \right)^2 + \sum_{i=1}^n \left(I_i \frac{1}{\lambda_i(\alpha)} - 1 \right) \frac{\partial^2 \lambda_i(\alpha)}{\partial \alpha^2},\end{aligned}\quad (C.1)$$

$$\begin{aligned}\nabla^{11} J_{ML}(\alpha, I^n) &= \left(\frac{\partial^2}{\partial \alpha \partial I_1} J_{ML}(\alpha, I^n), \dots, \frac{\partial^2}{\partial \alpha \partial I_n} J_{ML}(\alpha, I^n) \right)^T \\ &= \left(\frac{1}{\lambda_1(\alpha)} \frac{\partial \lambda_1(\alpha)}{\partial \alpha}, \dots, \frac{1}{\lambda_n(\alpha)} \frac{\partial \lambda_n(\alpha)}{\partial \alpha} \right)^T.\end{aligned}\quad (C.2)$$

For evaluating these two expressions at $\alpha = \tau_{ML}(\bar{I}^n)$ as required in (B.1), we use that¹⁴

$$\tau_{ML}(\bar{I}^n) = \operatorname{argmin}_{\alpha \in \mathbb{R}} \sum_{i=1}^n -\lambda_i(x_c) \ln(\lambda_i(\alpha)) + \lambda_i(\alpha).\quad (C.3)$$

Then we will use the following result

Proposition 3. *Under the assumption of a Gaussian PSF, $\tau_{ML}(\bar{I}^n) = x_c$.*

Notice again, that this proposition is the second assumption used in Theorem 1. From this proposition, it follows that

$$\begin{aligned}\nabla^{20} J(\tau_{ML}(\bar{I}^n), \bar{I}^n) &= -\sum_{i=1}^n \lambda_i(x_c) \frac{1}{\lambda_i^2(x_c)} \left(\frac{\partial \lambda_i(\alpha)}{\partial \alpha} \right)^2 \Big|_{\alpha=x_c} + \sum_{i=1}^n \left(\frac{\lambda_i(x_c)}{\lambda_i(x_c)} - 1 \right) \frac{\partial^2 \lambda_i(\alpha)}{\partial \alpha^2} \Big|_{\alpha=x_c}, \\ &= -\sum_{i=1}^n \frac{1}{\lambda_i(x_c)} \left(\frac{\partial \lambda_i(\alpha)}{\partial \alpha} \right)^2 \Big|_{\alpha=x_c},\end{aligned}\quad (C.4)$$

$$\begin{aligned}\nabla^{11} J(\tau_{ML}(\bar{I}^n), \bar{I}^n) &= \left(\frac{\partial^2}{\partial \alpha \partial I_1} J(\alpha, I^n), \dots, \frac{\partial^2}{\partial \alpha \partial I_n} J(\alpha, I^n) \right)^T \Big|_{\alpha=\tau_{ML}(\bar{I}^n)} \\ &= \left(\frac{1}{\lambda_1(\alpha)} \frac{\partial \lambda_1(\alpha)}{\partial \alpha}, \dots, \frac{1}{\lambda_n(\alpha)} \frac{\partial \lambda_n(\alpha)}{\partial \alpha} \right)^T \Big|_{\alpha=x_c}.\end{aligned}\quad (C.5)$$

Finally, we apply (C.4) and (C.5) in (A.9) to obtain that

$$\begin{aligned}\nabla \tau_{ML}(\bar{I}^n) \cdot (I^n - \bar{I}^n) &= -[\nabla^{20} J(\tau_{ML}(\bar{I}^n), \bar{I}^n)]^{-1} [\nabla^{11} J(\tau_{ML}(\bar{I}^n), \bar{I}^n)] (I^n - \bar{I}^n) \\ &= -\underbrace{\left[\sum_{i=1}^n \frac{1}{\lambda_i(x_c)} \left(\frac{\partial \lambda_i(\alpha)}{\partial \alpha} \right)^2 \Big|_{\alpha=x_c} \right]^{-1}}_a \cdot \underbrace{\sum_{j=1}^n \frac{1}{\lambda_j(x_c)} \frac{\partial \lambda_j(\alpha)}{\partial \alpha} \Big|_{\alpha=x_c}}_{b_j} (I_j - \mathbb{E}(I_j)),\end{aligned}\quad (C.6)$$

that shows that the sufficient condition in (19) of Theorem 1 is satisfied. For computing the value $\sigma_{ML}^2(n)$ in (A.5), we have that

$$\operatorname{Cov}\{I_i, I_j\} = \begin{cases} \operatorname{Var}\{I_i\} = \lambda_i(x_c), & \text{if } i = j, \\ 0 & \sim . \end{cases}\quad (C.7)$$

since the observations are independent and follow a Poisson distribution. Then, replacing (C.4), (C.5) and (C.7) in (A.5) we have that

$$\sigma_{ML}^2(n) = \frac{1}{\sum_{i=1}^n \frac{1}{\lambda_i(x_c)} \left(\frac{\partial \lambda_i(\alpha)}{\partial \alpha} \right)^2 \Big|_{\alpha=x_c}},\quad (C.8)$$

¹⁴ Considering that $\bar{I}_i = E\{I_i\} = \lambda_i(x_c)$.

which resolves the identity in (34). Finally $\beta_{ML}(n)$ and $\epsilon_{ML}(n)$ comes from evaluating (A.2) and (A.1) in this ML context. For that we only need to determine $\frac{\partial^2 \tau_{ML}}{\partial I_i \partial I_j}(\bar{I}^n + t(I^n - \bar{I}^n))$. Using (Fessler 1996, Eq. (17)), we can obtain the following identity¹⁵

$$\begin{aligned} \frac{\partial^2 \tau_{ML}}{\partial I_i \partial I_j}(\bar{I}^n + t(I^n - \bar{I}^n)) &= \frac{-1}{\left[\sum_{i=1}^n \frac{\partial^2 \lambda_i(\alpha)}{\partial \alpha^2} \cdot \frac{\bar{I}_i + t(I_i - \bar{I}_i)}{\lambda_i(\alpha)} - \frac{\bar{I}_i + t(I_i - \bar{I}_i)}{\lambda_i^2(\alpha)} \cdot \left(\frac{\partial \lambda_i(\alpha)}{\partial \alpha} \right)^2 \right]^2} \\ &\left[\left[\sum_{i=1}^n \frac{\partial^3 \lambda_i(\alpha)}{\partial \alpha^3} \cdot \frac{\bar{I}_i + t(I_i - \bar{I}_i)}{\lambda_i(\alpha)} - 3 \frac{\partial^2 \lambda_i(\alpha)}{\partial \alpha^2} \frac{\partial \lambda_i(\alpha)}{\partial \alpha} \frac{\bar{I}_i + t(I_i - \bar{I}_i)}{\lambda_i^2(\alpha)} + 2 \frac{\bar{I}_i + t(I_i - \bar{I}_i)}{\lambda_i^3(\alpha)} \left(\frac{\partial \lambda_i(\alpha)}{\partial \alpha} \right)^3 \right] \right. \\ &\quad \left. \left(-\frac{1}{\lambda_j(\alpha)} \frac{\partial \lambda_j(\alpha)}{\partial \alpha} \right) \right] + \frac{\partial^2 \lambda_j(\alpha)}{\partial \alpha^2} \frac{1}{\lambda_j(\alpha)} - \frac{1}{\lambda_j(\alpha)} \left(\frac{\partial \lambda_j(\alpha)}{\partial \alpha} \right)^2 \right] \\ &\left(-\frac{1}{\lambda_i(\alpha)} \frac{\partial \lambda_i(\alpha)}{\partial \alpha} \right) + \left(\frac{\partial^2 \lambda_i(\alpha)}{\partial \alpha^2} \frac{1}{\lambda_i(\alpha)} - \frac{1}{\lambda_i(\alpha)} \left(\frac{\partial \lambda_i(\alpha)}{\partial \alpha} \right)^2 \right) \cdot \left(-\frac{1}{\lambda_j(\alpha)} \frac{\partial \lambda_j(\alpha)}{\partial \alpha} \right) \Bigg|_{\alpha=\tau_{ML}(\bar{I}^n + t(I^n - \bar{I}^n))}, \end{aligned} \quad (C.9)$$

which concludes the result.

Appendix C.1: Proof of Proposition 3

Proof: Let us consider the function $g_n : \mathbb{R}_+^n \rightarrow \mathbb{R}$ given by

$$g_n(y_1, \dots, y_n)_{\lambda_i^n} = \sum_{i=1}^n -\lambda_i \ln(y_i) + y_i. \quad (C.10)$$

We note that

$$\min_{y_i^n \in \mathbb{R}_+^n} g_n(y_1, \dots, y_n)_{\lambda_i^n} = \sum_{i=1}^n \min_{y_i \in \mathbb{R}_+} g_1(y_i)_{\lambda_i}, \quad (C.11)$$

where applying first order condition $y_i = \lambda_i, \forall i \in \{1, \dots, n\}$. Returning to our problem in (C.3) where $\lambda_i = \bar{I}_i = \lambda_i(x_c)$ and $y_i = \lambda_i(\alpha)$, it is clear, considering the Gaussian profile in PSF, that

$$\lambda_i(\alpha) = \lambda_i(x_c) \quad \forall i \in \{1, \dots, n\} \text{ if } \alpha = x_c, \quad (C.12)$$

which concludes the result.

Appendix C.2: Proof of Eq. (C.9)

Proof: Recall (Fessler 1996, Eq. (17)) and considering $J_{ML}(\alpha, I^n)$ as the cost function we have that:

$$\begin{aligned} \frac{\partial^2 \tau_{ML}}{\partial I_i \partial I_j}(\bar{I}^n + t(I^n - \bar{I}^n)) &= \left[-\frac{\partial^2 J_{ML}(\alpha, \bar{I}^n + t(I^n - \bar{I}^n))}{\partial \alpha^2} \right]^{-1} \left(\left[\frac{\partial^3 J_{ML}(\alpha, \bar{I}^n + t(I^n - \bar{I}^n))}{\partial \alpha^3} \cdot \frac{\partial \tau_{ML}(\bar{I}^n + t(I^n - \bar{I}^n))}{\partial I_j} \right. \right. \\ &\quad \left. \left. + \frac{\partial^3 J_{ML}(\alpha, \bar{I}^n + t(I^n - \bar{I}^n))}{\partial \alpha^2 \partial I_j} \right] \cdot \frac{\partial \tau_{ML}(\bar{I}^n + t(I^n - \bar{I}^n))}{\partial I_i} \right. \\ &\quad \left. \left. + \frac{\partial^3 J_{ML}(\alpha, \bar{I}^n + t(I^n - \bar{I}^n))}{\partial \alpha^2 \partial I_i} \cdot \frac{\partial \tau_{ML}(\bar{I}^n + t(I^n - \bar{I}^n))}{\partial I_j} + \frac{\partial^3 J_{ML}(\alpha, \bar{I}^n + t(I^n - \bar{I}^n))}{\partial \alpha \partial I_i \partial I_j} \right) \Bigg|_{\alpha=\tau_{ML}(\bar{I}^n + t(I^n - \bar{I}^n))}, \end{aligned} \quad (C.13)$$

where from the definition of $J_{ML}(\alpha, I^n)$ we have that

$$\frac{\partial^2 J_{ML}(\alpha, \bar{I}^n + t(I^n - \bar{I}^n))}{\partial \alpha^2} = \sum_{i=1}^n \frac{\partial^2 \lambda_i(\alpha)}{\partial \alpha^2} \cdot \frac{\bar{I}_i + t(I_i - \bar{I}_i)}{\lambda_i(\alpha)} - \frac{\bar{I}_i + t(I_i - \bar{I}_i)}{\lambda_i^2(\alpha)} \cdot \left(\frac{\partial \lambda_i(\alpha)}{\partial \alpha} \right)^2, \quad (C.14)$$

$$\frac{\partial^3 J_{ML}(\alpha, \bar{I}^n + t(I^n - \bar{I}^n))}{\partial \alpha^3} = \sum_{i=1}^n \frac{\partial^3 \lambda_i(\alpha)}{\partial \alpha^3} \cdot \frac{\bar{I}_i + t(I_i - \bar{I}_i)}{\lambda_i(\alpha)} - 3 \frac{\partial^2 \lambda_i(\alpha)}{\partial \alpha^2} \frac{\partial \lambda_i(\alpha)}{\partial \alpha} \frac{\bar{I}_i + t(I_i - \bar{I}_i)}{\lambda_i^2(\alpha)} + 2 \frac{\bar{I}_i + t(I_i - \bar{I}_i)}{\lambda_i^3(\alpha)} \left(\frac{\partial \lambda_i(\alpha)}{\partial \alpha} \right)^3, \quad (C.15)$$

$$\frac{\partial^3 J_{ML}(\alpha, \bar{I}^n + t(I^n - \bar{I}^n))}{\partial \alpha^2 \partial I_i} = \frac{\partial^2 \lambda_i(\alpha)}{\partial \alpha^2} \cdot \frac{1}{\lambda_i(\alpha)} - \frac{1}{\lambda_i^2(\alpha)} \cdot \left(\frac{\partial \lambda_i(\alpha)}{\partial \alpha} \right)^2, \quad (C.16)$$

¹⁵ The derivation of this result is presented in Appendix C.2.

$$\frac{\partial^3 J_{ML}(\alpha, \bar{I}^n + t(I^n - \bar{I}^n))}{\partial \alpha \partial I_i \partial I_j} = 0. \quad (\text{C.17})$$

Concerning $\frac{\partial \tau_{ML}(\bar{I}^n + t(I^n - \bar{I}^n))}{\partial I_i}$ it is just the i -th component of the gradient in Eq. (B.1), then we use (C.1) and (C.2)

$$\begin{aligned} \frac{\partial \tau_{ML}(\bar{I}^n + t(I^n - \bar{I}^n))}{\partial I_i} &= \frac{-\frac{\partial^2 J_{ML}(\alpha, \bar{I}^n + t(I^n - \bar{I}^n))}{\partial \alpha \partial I_i}}{\frac{\partial^2 J_{ML}(\alpha, \bar{I}^n + t(I^n - \bar{I}^n))}{\partial \alpha^2}} \\ &= \frac{\left(-\frac{1}{\lambda_i(\alpha)} \frac{\partial \lambda_i(\alpha)}{\partial \alpha}\right)}{\sum_{i=1}^n \frac{\partial^2 \lambda_i(\alpha)}{\partial \alpha^2} \cdot \frac{\bar{I}_i + t(I_i - \bar{I}_i)}{\lambda_i(\alpha)} - \frac{\bar{I}_i + t(I_i - \bar{I}_i)}{\lambda_i^2(\alpha)} \cdot \left(\frac{\partial \lambda_i(\alpha)}{\partial \alpha}\right)^2}. \end{aligned} \quad (\text{C.18})$$

Finally, replacing (C.14), (C.15), (C.16), (C.17) and (C.18) in (C.13), and evaluating in $\alpha = \tau_{ML}(\bar{I}^n + t(I^n - \bar{I}^n))$ we obtain the desired result

$$\begin{aligned} \frac{\partial^2 \tau_{ML}(\bar{I}^n + t(I^n - \bar{I}^n))}{\partial I_i \partial I_j} &= \frac{-1}{\left[\sum_{i=1}^n \frac{\partial^2 \lambda_i(\alpha)}{\partial \alpha^2} \cdot \frac{\bar{I}_i + t(I_i - \bar{I}_i)}{\lambda_i(\alpha)} - \frac{\bar{I}_i + t(I_i - \bar{I}_i)}{\lambda_i^2(\alpha)} \cdot \left(\frac{\partial \lambda_i(\alpha)}{\partial \alpha}\right)^2\right]^2} \\ &\left[\left[\sum_{i=1}^n \frac{\partial^3 \lambda_i(\alpha)}{\partial \alpha^3} \cdot \frac{\bar{I}_i + t(I_i - \bar{I}_i)}{\lambda_i(\alpha)} - 3 \frac{\partial^2 \lambda_i(\alpha)}{\partial \alpha^2} \frac{\partial \lambda_i(\alpha)}{\partial \alpha} \frac{\bar{I}_i + t(I_i - \bar{I}_i)}{\lambda_i^2(\alpha)} + 2 \frac{\bar{I}_i + t(I_i - \bar{I}_i)}{\lambda_i^3(\alpha)} \left(\frac{\partial \lambda_i(\alpha)}{\partial \alpha}\right)^3\right] \right. \\ &\left. \left(-\frac{1}{\lambda_j(\alpha)} \frac{\partial \lambda_j(\alpha)}{\partial \alpha}\right) + \frac{\partial^2 \lambda_j(\alpha)}{\partial \alpha^2} \frac{1}{\lambda_j(\alpha)} - \frac{1}{\lambda_j(\alpha)} \left(\frac{\partial \lambda_j(\alpha)}{\partial \alpha}\right)^2\right] \\ &\left(-\frac{1}{\lambda_i(\alpha)} \frac{\partial \lambda_i(\alpha)}{\partial \alpha}\right) + \left(\frac{\partial^2 \lambda_i(\alpha)}{\partial \alpha^2} \frac{1}{\lambda_i(\alpha)} - \frac{1}{\lambda_i(\alpha)} \left(\frac{\partial \lambda_i(\alpha)}{\partial \alpha}\right)^2\right) \cdot \left(-\frac{1}{\lambda_j(\alpha)} \frac{\partial \lambda_j(\alpha)}{\partial \alpha}\right) \Bigg|_{\alpha=\tau_{ML}(\bar{I}^n + t(I^n - \bar{I}^n))}. \end{aligned} \quad (\text{C.19})$$

References

- Adorf, H.-M. 1996, in *Astronomical Data Analysis Software and Systems V*, Vol. 101, 13
- Alard, C. & Lupton, R. H. 1998, *The Astrophysical Journal*, 503, 325
- Altmann, M., Bouquillon, S., & Taris, F. 2014, in *Proc. SPIE*, Vol. 9149, 91490P
- Auer, L. & Van Altena, W. 1978, *The Astronomical Journal*, 83, 531
- Bastian, U. 2004, GAIA technical note, 2004 BASNOCODE
- Bendinelli, O., Parmeggiani, G., Piccioni, A., & Zavatti, F. 1987, *The Astronomical Journal*, 94, 1095
- Benedict, G. F., McArthur, B. E., Nelan, E. P., & Harrison, T. E. 2016, *Publications of the Astronomical Society of the Pacific*, 129, 012001
- Bouquillon, S., Mendez, R., Altmann, M., et al. 2017, *Astronomy & Astrophysics*, 606, A27
- Bradley, R. A. & Gart, J. J. 1962, *Biometrika*, 49, 205
- Bristow, P., Kerber, F., & Rosa, M. 2006, in *The 2005 HST Calibration Workshop: Hubble After the Transition to Two-Gyro Mode*, 299
- Cacciari, C., Pancino, E., & Bellazzini, M. 2016, *Astronomische Nachrichten*, 337, 899
- Chromey, F. R. 2016, *To measure the sky: an introduction to observational astronomy* (Cambridge University Press)
- Chun-Lin, L. 1993, in *Symposium-International Astronomical Union*, Vol. 156, Cambridge University Press, 113–116
- Cover, T. M. & Thomas, J. A. 2012, *Elements of information theory* (John Wiley & Sons)
- Cramér, H. 1946, *Scandinavian Actuarial Journal*, 1946, 85
- Echeverria, A., Silva, J. F., Mendez, R. A., & Orchard, M. 2016, *Astronomy & Astrophysics*, 594, A111
- Fessler, J. A. 1996, *IEEE Transactions on Image Processing*, 5, 493
- Gai, M., Busonero, D., & Cancelliere, R. 2017, *Publications of the Astronomical Society of the Pacific*, 129, 054502
- Gray, R. M. & Davisson, L. D. 2004, *An introduction to statistical signal processing* (Cambridge University Press)
- Hoadley, B. 1971, *The Annals of mathematical statistics*, 1977
- Høg, E. 2017, arXiv preprint arXiv:1707.01020
- Howell, S. B. 2006, *Handbook of CCD astronomy*, Vol. 5 (Cambridge University Press)
- Jakobsen, P., Greenfield, P., & Jedrzejewski, R. 1992, *Astronomy and astrophysics*, 253, 329
- Janesick, J. R. 2001, *Scientific charge-coupled devices*, Vol. 83 (SPIE press)
- Janesick, J. R. 2007, *Photon transfer* (SPIE press San Jose)
- Kay, S. M. 1993, *Fundamentals of statistical signal processing. Vol 1, Estimation theory*
- Kendall, M., Stuart, A., Ord, J., & Arnold, S. 1999, *Arnold [etc.]*, London [etc.]
- King, I. R. 1971, *Publications of the Astronomical Society of the Pacific*, 83, 199
- King, I. R. 1983, *Publications of the Astronomical Society of the Pacific*, 95, 163
- Lattanzi, M. 2012, *Mem. SA It*, 83, 1033
- Lee, J.-F. & Van Altena, W. 1983, *The Astronomical Journal*, 88, 1683
- Lemon, C. A., Auger, M. W., McMahon, R. G., & Kposov, S. E. 2017, *Monthly Notices of the Royal Astronomical Society*, 472, 5023
- Lindgren, L. 1978, in *International Astronomical Union Colloquium*, Vol. 48, Cambridge University Press, 197–217
- Lindgren, L. 2008, *Gaia DPAC public document GAIA-C3-TN-LU-LL-078* (November 2008)
- Lindgren, L. 2010, *ISSI Scientific Reports Series*, 9, 279
- Lobos, R. A., Silva, J. F., Mendez, R. A., & Orchard, M. 2015, *Publications of the Astronomical Society of the Pacific*, 127, 1166
- McLean, I. S. 2008, *Electronic imaging in astronomy: detectors and instrumentation* (Springer Science & Business Media)
- Méndez, R. A., Costa, E., Pedreros, M. H., et al. 2010, *Publications of the Astronomical Society of the Pacific*, 122, 853

- Mendez, R. A., Silva, J. F., & Lobos, R. 2013, *Publications of the Astronomical Society of the Pacific*, 125, 580
- Mendez, R. A., Silva, J. F., Orostica, R., & Lobos, R. 2014, *Publications of the Astronomical Society of the Pacific*, 126, 798
- Michalik, D. & Lindegren, L. 2016, *Astronomy & Astrophysics*, 586, A26
- Michalik, D., Lindegren, L., Hobbs, D., & Butkevich, A. G. 2015, *Astronomy & Astrophysics*, 583, A68
- Prusti, T., De Bruijne, J., Brown, A. G., et al. 2016, *Astronomy & Astrophysics*, 595, A1
- Rao, C. R. 1945, *Bull. Calcutta Math. Soc.*, 37, 81
- Reffert, S. 2009, *New Astronomy Reviews*, 53, 329
- So, H. C., Chan, Y. T., Ho, K., & Chen, Y. 2013, *IEEE Signal Processing Magazine*, 30, 162
- Stetson, P. B. 1987, *Publications of the Astronomical Society of the Pacific*, 99, 191
- Stone, R. C. 1989, *The Astronomical Journal*, 97, 1227
- Vakili, M. & Hogg, D. W. 2016, arXiv preprint arXiv:1610.05873
- Van Altena, W. F. 2013, *Astrometry for astrophysics: methods, models, and applications* (Cambridge University Press)
- Van Altena, W. F. & Auer, L. 1975, in *Image Processing Techniques in Astronomy* (Springer), 411–418
- Van Trees, H. L. 2004, *Detection, estimation, and modulation theory, part I: detection, estimation, and linear modulation theory* (John Wiley & Sons)
- Zaccheo, T., Gonsalves, R., Ebstein, S., & Nisenson, P. 1995, *The Astrophysical Journal*, 439, L43
- Zhang, J., Hao, Y. C., Wang, L., & Long, Y. 2016, *Publications of the Astronomical Society of the Pacific*, 128, 035003