

# REXPACO: An algorithm for high contrast reconstruction of the circumstellar environment by angular differential imaging

Olivier Flasseur<sup>1</sup>, Samuel Thé<sup>1</sup>, Loïc Denis<sup>2</sup>, Éric Thiébaud<sup>1</sup>, and Maud Langlois<sup>1</sup>

<sup>1</sup> Université de Lyon, Université Lyon1, ENS de Lyon, CNRS, Centre de Recherche Astrophysique de Lyon UMR 5574, 69230 Saint-Genis-Laval, France  
e-mail: [olivier.flasseur@univ-lyon1.fr](mailto:olivier.flasseur@univ-lyon1.fr)

<sup>2</sup> Université de Lyon, UJM-Saint-Etienne, CNRS, Institut d'Optique Graduate School, Laboratoire Hubert Curien UMR 5516, 42023 Saint-Etienne, France  
e-mail: [loic.denis@univ-st-etienne.fr](mailto:loic.denis@univ-st-etienne.fr)

Received 16 July 2020 / Accepted 11 May 2021

## ABSTRACT

*Context.* Direct imaging is a method of choice for probing the close environment of young stars. Even with the coupling of adaptive optics and coronagraphy, the direct detection of off-axis sources such as circumstellar disks and exoplanets remains challenging due to the required high contrast and small angular resolution. Angular differential imaging (ADI) is an observational technique that introduces an angular diversity to help disentangle the signal of off-axis sources from the residual signal of the star in a post-processing step.

*Aims.* While various detection algorithms have been proposed in the last decade to process ADI sequences and reach high contrast for the detection of point-like sources, very few methods are available to reconstruct meaningful images of extended features such as circumstellar disks. The purpose of this paper is to describe a new post-processing algorithm dedicated to the reconstruction of the spatial distribution of light (total intensity) received from off-axis sources, in particular from circumstellar disks.

*Methods.* Built on the recent PACO algorithm dedicated to the detection of point-like sources, the proposed method is based on the local learning of patch covariances capturing the spatial fluctuations of the stellar leakages. From this statistical modeling, we develop a regularized image reconstruction algorithm (REXPACO) following an inverse problems approach based on a forward image formation model of the off-axis sources in the ADI sequences.

*Results.* Injections of fake circumstellar disks in ADI sequences from the VLT/SPHERE-IRDIS instrument show that both the morphology and the photometry of the disks are better preserved by REXPACO compared to standard post-processing methods such as cADI. In particular, the modeling of the spatial covariances proves useful in reducing typical ADI artifacts and in better disentangling the signal of these sources from the residual stellar contamination. The application to stars hosting circumstellar disks with various morphologies confirms the ability of REXPACO to produce images of the light distribution with reduced artifacts. Finally, we show how REXPACO can be combined with PACO to disentangle the signal of circumstellar disks from the signal of candidate point-like sources.

*Conclusions.* REXPACO is a novel post-processing algorithm for reconstructing images of the circumstellar environment from high contrast ADI sequences. It produces numerically deblurred images and exploits the spatial covariances of the stellar leakages and of the noise to efficiently eliminate this nuisance term. The processing is fully unsupervised, all tuning parameters being directly estimated from the data themselves.

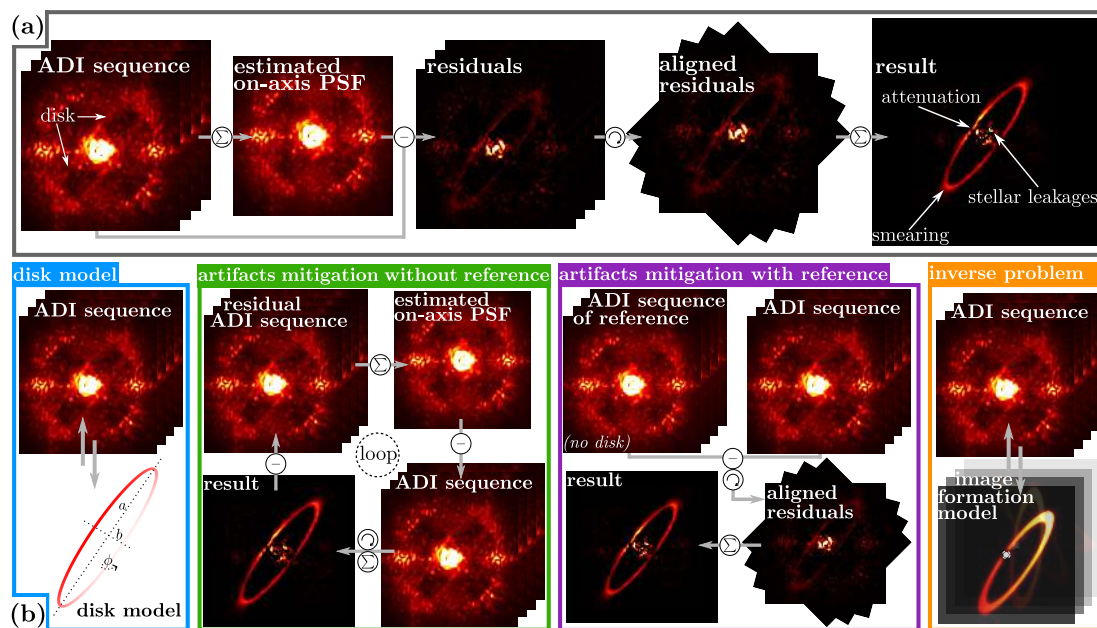
**Key words.** techniques: image processing – techniques: high angular resolution – methods: statistical – methods: data analysis

## 1. Introduction

Circumstellar disks are at the heart of the planet formation processes. Direct imaging in the near infrared is a unique method for addressing this question because planets can be detected together with the disk environment. Despite the very promising results from direct imaging, a small fraction of the expected circumstellar disks have been unveiled and characterized using high contrast direct imaging in the near infrared (Esposito et al. 2020; Garufi et al. 2020; Langlois et al. 2020). Recent protoplanetary and debris disk studies performed in intensity or in polarimetry have focused on morphological disk characteristics such as the presence of spirals (Benisty et al. 2015; Ren et al. 2018a; Muro-Arena et al. 2020), asymmetries, warps (Dawson et al. 2011; Kluska et al. 2020) and gaps (van Boekel et al. 2017) that could be signposts for the presence of exoplanets. In this context, protoplanetary disks allow unique studies of the exoplanet-disk

interactions (Keppler et al. 2018; Haffert et al. 2019; Mesa et al. 2019). In order to understand the physical processes governing these disks, it is essential to reconstruct their surface-brightness or scattering phase function but also to disentangle, in a post-processing step, the disks and the exoplanets.

From an observational point of view, the direct detection of circumstellar disks requires: (i) high angular resolution to resolve the close environment of stars (the typical size of disks is smaller than 1 arcsec) and (ii) high contrast to detect the faint light scattered by disks (in the infrared, an effective contrast better than  $10^5$  relative to the host star is typically required). Among the available ground-based observing facilities, the Spectro-Polarimetry High-contrast Exoplanet Research instrument (SPHERE; Beuzit et al. 2019) operating at the Very Large Telescope (VLT) offers such capabilities. It produces angular differential imaging (ADI; Marois et al. 2006) sequences using the pupil-tracking mode of the telescope as a means to enhance the



**Fig. 1.** Schematic representation of the main steps followed by the state-of-the-art post-processing methods for ADI sequences with circumstellar disks. (a) Principle of the classical approach based on the estimation and subtraction of an on-axis PSF. (b) Four main categories of algorithms designed to reduce reconstruction artifacts seen in (a): using a physics-based model of the disk, mitigating artifacts directly from the ADI sequence of interest or using an additional sequence of reference, or formalizing the reconstruction task as an inverse problem. REXPACO falls into the last category: methods based on an inverse problems approach.

achievable contrast. The presence of strong residual stellar leakages from the coronagraph, in the form of so-called speckles, is the current main limitation of the reachable contrast. A post-processing step is required to combine the images of the ADI sequence and cancel out most of the nuisance component<sup>1</sup> while preserving the signal from the off-axis sources<sup>2</sup>.

There is a large variety of methods dedicated to the post-processing of ADI sequences (see for example Pueyo 2018 for a review). Even though these methods are primarily designed for the detection of unresolved point-like sources, they are also widely used to process observations of disks (Milli et al. 2012) and share a common framework with most of the existing post-processing methods designed for disks. The common principle is to estimate a reference image<sup>3</sup> of the nuisance component (see Fig. 1a for a schematic illustration). This reference image is subtracted to each image of the sequence. The resulting residual images are aligned along a common direction to spatially superimpose the signals from the off-axis sources, and are added to each others. More sophisticated methods follow an inverse problems framework. In particular, the recent PACO algorithm (Flasseur et al. 2018a,b,c) jointly learns the average nuisance component and its fluctuations by estimating spatial covariances. Since the nuisance component is highly nonstationary, the underlying statistical model is estimated at a local scale of small patches.

<sup>1</sup> We refer to the term “nuisance component” to designate the contribution of light sources other than the object of interest, as well as different sources of noise.

<sup>2</sup> Since the signals from circumstellar disks and point-like sources come from directions other than the optical axis, we use the term “off-axis sources” to refer indifferently to these two types of objects.

<sup>3</sup> Since the nuisance component is mainly due to light coming from the optical axis, we refer to the term “on-axis point-spread function” (in short, “on-axis PSF”) to designate the estimated reference image of the nuisance component.

Most of these standard ADI post-processing algorithms are not suited to detect and reconstruct the light distribution of circumstellar disks. Their major drawback is the presence, in the output images, of strong artifacts that take the form of partial replicas, the removal of extended smooth components, smearing, and nonuniform attenuations due to the so-called self-subtraction phenomenon. As a direct consequence, both the morphology and the photometry of the disks are strongly corrupted. Very few specific solutions have been developed to tackle these limitations in the ADI processing of circumstellar disks. The existing approaches can be classified into four categories (see Fig. 1b for a schematic diagram of their principles).

A first family of methods introduce a forward-backward loop embedding a physics-based model of the disk (Milli et al. 2012; Esposito et al. 2013; Mazoyer et al. 2020). In a nutshell, the parameters constraining the disk morphology and flux are optimized by minimizing the residuals produced by the selected post-processing algorithm from the ADI sequence free from the current estimation of the disk contribution. This method is however limited as it requires a physics-based model of the disk that is sufficiently simple to be numerically optimized and yet sufficiently flexible to accurately model the actual disk.

A second category of methods attempt to avoid self-subtraction by reducing the disk contribution in the estimation of the reference on-axis PSF. For instance, Pairet et al. (2018) have proposed a modified version of the PCA-based post-processing algorithm that iteratively removes the estimated disk component to improve the estimation of the on-axis PSF. To limit self-subtraction, Ren et al. (2020) have proposed a data-imputation strategy: The areas of the field of view impacted by the disk are considered as missing-data based on prior knowledge of the disk location and replaced by regions free from any disk contribution.

So-called reference differential imaging (RDI; Gerard & Marois 2016; Wahhaj et al. 2020) is a third class of approaches that require an additional ADI sequence of a reference star that hosts no known off-axis sources to estimate the stellar

**Table 1.** Summary of the main notations.

Not.	Range	Definition
▷ Constants and related indexes		
$K$	$\mathbb{N}$	Number of pixels in a patch
$M$	$\mathbb{N}$	Number of pixels in a reconstructed image
$N$	$\mathbb{N}$	Number of pixels in a temporal frame
$T$	$\mathbb{N}$	Number of temporal frames
$n$	$[[1, N]]$	Pixel index
$t$	$[[1, T]]$	Temporal index
▷ Data quantities		
$\mathbf{r}$	$\mathbb{R}^{NT}$	ADI sequence (observed data)
$\mathbf{f}$	$\mathbb{R}^{NT}$	Nuisance component
$\mathbf{x}$	$\mathbb{R}_+^M$	Light distribution of off-axis sources
▷ Operators		
$\mathbf{A}$	$\mathbb{R}^{NT \times M}$	Off-axis PSF model: $\mathbf{A} = \mathbf{V}\mathbf{H}\mathbf{T}\mathbf{Q}$
$\mathbf{Q}$	$\mathbb{R}^{MT \times M}$	Field rotation
$\mathbf{T}$	$\mathbb{R}^{MT \times MT}$	Coronagraph attenuation
$\mathbf{H}$	$\mathbb{R}^{MT \times MT}$	Convolution by off-axis PSF
$\mathbf{V}$	$\mathbb{R}^{NT \times MT}$	Field of view cropping
$\mathbf{P}$	$\mathbb{R}^{K \times N}$	Patch extractor
▷ Estimated quantities		
$\widehat{\mathbf{m}}$	$\mathbb{R}^N$	Temporal mean of $\mathbf{f}$
$\widehat{\mathbf{C}}$	$\mathbb{R}^{K \times K}$	Empirical spatial covariance of $\mathbf{f}$
$\widetilde{\mathbf{C}}$	$\mathbb{R}^{K \times K}$	Shrunk spatial covariance of $\mathbf{f}$
$\widehat{\mathbf{\Omega}}$	$\mathbb{R}^{NT}$	Statistics of the nuisance component $\mathbf{f}$
$\widehat{\mathbf{x}}$	$\mathbb{R}_+^M$	Reconstructed distribution of light

contamination. The quality of the post-processing highly depends both on the selection of the reference star, and on the similarity of the observing conditions (Ruane et al. 2019). Ren et al. (2018b) also use a reference ADI sequence to build a non-negative decomposition (via nonnegative matrix factorization) of the nuisance component.

Finally, a fourth class of methods is emerging and states the reconstruction of the object of interest as an inverse problem. It consists in jointly estimating the nuisance term, the disk and, possibly, point-like sources given the data and low-complexity priors adapted to each kind of contribution (Pairet et al. 2019, 2021). The method we propose in this paper belongs to this family.

Based on the previous analysis of the astrophysical needs and of the state-of-the-art post-processing methods for circumstellar disk detection and reconstruction, the desirable properties of post-processing algorithms may be listed as follows: (i) detection of disks at high contrasts, (ii) reconstruction of a physically plausible image of the flux distribution with limited processing artifacts, and (iii) ability to disentangle the spatially extended contribution of disks from the spatially localized contribution of point-like sources.

In this paper, we attempt to address these three points by proposing a new post-processing method to extract the light distribution of extended features like circumstellar disks from ADI sequences. To get rid of the nuisance term, we formalize the reconstruction as an inverse problem where all the contributions appear explicitly and where the measured instrumental PSF is taken into account. As a result, the estimation of the object of interest is nearly free from contamination by the stellar leakages and from blurring by the off-axis PSF. The method proposed in this paper, named REXPACO (for Reconstruction of

EXtended features by PATCH COvariances), introduces dedicated constraints for each component to help disentangle the object of interest from the nuisance term. For the object of interest, these constraints take the form of specific regularizations favoring the smoothness of extended structures (e.g., disk-like component) or the sparsity of point-like sources. For the nuisance term, we consider the same statistical model as with PACO’s algorithm (Flasseur et al. 2018a,b,c), initially designed for exoplanet detection, but which has also proven very effective for the recovery of extended patterns in holographic microscopy (Flasseur et al. 2019). The power of PACO statistical model resides in its ability to capture the local behavior of the stellar leakages and the noise through the spatial covariances. Another advantage is that this model is directly learned from the ADI sequence and requires no other specific calibration data. Similarly, REXPACO is an unsupervised algorithm: All tuning parameters are automatically estimated from the data. The framework implemented by REXPACO is rather flexible. As an example, REXPACO can be coupled with PACO to better estimate off-axis point-like sources.

The paper is organized as follows. In Sect. 2, we formalize REXPACO algorithm. In Sect. 3, we illustrate the results obtained by REXPACO, both with semisynthetic data and with real observations of known circumstellar disks from the InfraRed Dual Imaging Spectrograph (IRDIS; Dohlen et al. 2008a,b) of the VLT/SPHERE instrument. Section 4 presents results from the combination of REXPACO with the PACO algorithm to unmix the spatially extended contribution of disk from the point-like contribution of exoplanets. Section 5 summarizes our contributions.

## 2. REXPACO: Reconstruction of EXtended features by PATCH COvariances modeling

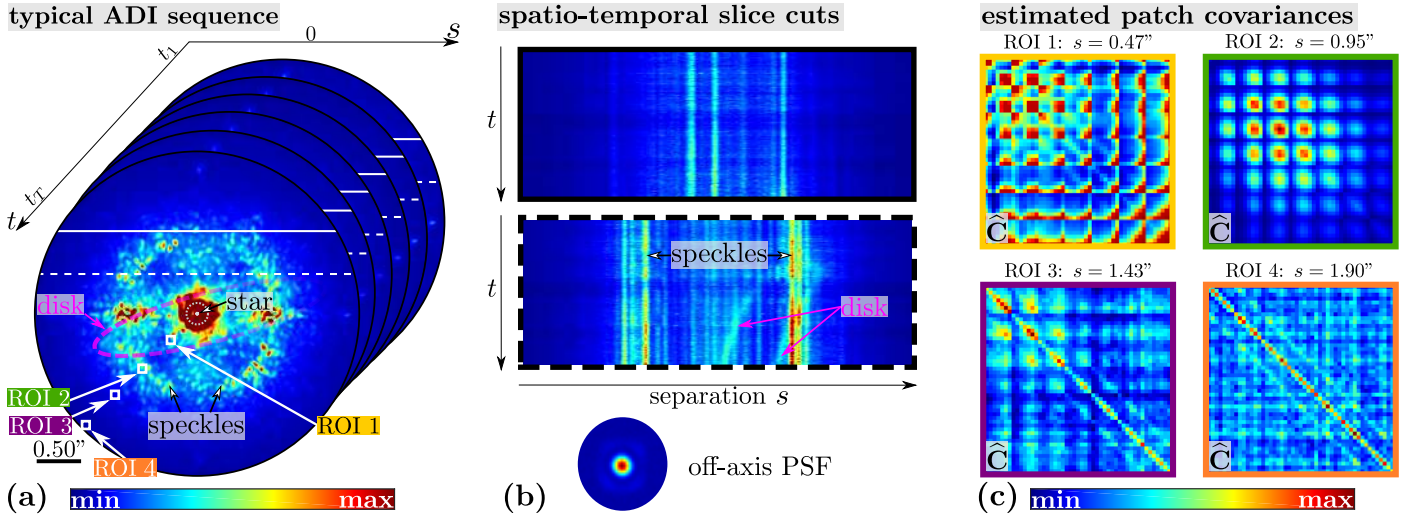
This section describes REXPACO, the method we propose to recover the spatial distribution  $\mathbf{x}$  of the light flux of off-axis sources from an ADI sequence  $\mathbf{r} \in \mathbb{R}^{NT}$  formed by  $T$  images of  $N$  pixels each. Table 1 summarizes the main notations used throughout the paper.

### 2.1. Inverse problems formulation

REXPACO follows an inverse problems formulation: It computes an estimate  $\widehat{\mathbf{x}}$  from the sequence of observations  $\mathbf{r}$  by fitting a model of the data under given constraints. Since all observations  $\mathbf{r}$  are collected during a single night, the distribution of interest  $\mathbf{x}$  can be considered time-invariant and described by a vector  $\mathbf{x} \in \mathbb{R}_+^M$ , where  $M$  is the number of pixels of the reconstructed field of view and where  $\mathbb{R}_+$  is the set of nonnegative real numbers to account that the intensity is necessarily nonnegative everywhere. The ADI data sequence  $\mathbf{r}$  results from the contribution of two components as expressed by:

$$\mathbf{r} = \mathbf{A}\mathbf{x} + \mathbf{f}, \quad (1)$$

where  $\mathbf{A}$  is the linear operator ( $\mathbb{R}^M \rightarrow \mathbb{R}^{NT}$ ) modeling the effects of the off-axis instrumental point spread function (off-axis PSF), that is to say the non-coronagraphic image of the star, and  $\mathbf{f} \in \mathbb{R}^{NT}$  is a nuisance component accounting for the residual stellar light in the presence of aberrations (that is, the speckles) and for the contribution of noise. Due to the temporal variation of aberrations and to the noise, the nuisance component  $\mathbf{f}$  is the result of some random process. The quality of the statistical modeling of this process is crucial for the inversion of Eq. (1), and it represents a key element of REXPACO. Figure 2a displays an



**Fig. 2.** Typical ADI sequence from the VLT/SPHERE-IRDIS instrument: (a) Examples of temporal frames; (b) two spatio-temporal slices cut along the solid and dashed lines of (a) emphasizing the spatial variations of the structure of the signal; (c) estimated covariance matrices for four regions of interest (ROIs) at different angular separations  $s$  with the host star. Data set: HR 4796A, see Sect. 3.3.1 for observing conditions.

example of an ADI sequence from the star HR 4796A observed by VLT/SPHERE-IRDIS. HR 4796A is surrounded by a bright inclined debris disk (see Sect. 3.3.1 for a more complete description of its astrophysical properties). We selected this data set to illustrate the rationale of the different ingredients of REXPACO. Part a of Fig. 2 shows that the bright disk of HR 4796A can only be marginally detected by a close visual inspection of a given frame of the sequence. This is due to the very large contrast between the star and its environment: The contribution of the component of interest  $\mathbf{A}\mathbf{x}$  is typically fainter than the nuisance component  $\mathbf{f}$  by one to four orders of magnitude. This high contrast sets the gain that must be achieved by the post-processing method in order to recover the component of interest  $\mathbf{x}$  from the data. Part b of Fig. 2 shows two spatio-temporal slices extracted at two different locations (along the solid line: far from the host star, along the dashed line: closer to the host star) and illustrates the nonstationarity of the nuisance component. Near the star (i.e., at angular separations smaller than about 1.5 arcsec), the nuisance component is dominated by speckles. At larger angular separations, the nuisance component mainly results from a combination of photon noise, thermal background flux and detector readout. Part c of Fig. 2 shows the empirical spatial covariances of the data computed by temporal averaging through the ADI sequence in different regions of interest (ROIs). This figure not only reveals that the covariances of the fluctuations are spatially nonstationary but also that the covariances are non-negligible in a small neighborhood. In a nutshell, the nuisance component  $\mathbf{f}$  has strong spatial correlations and its distribution is spatially nonstationary and fluctuates over time, making the extraction of the object  $\mathbf{x}$  all the more difficult. We introduce in Sect. 2.3 a model of the probability density function  $p(\mathbf{f})$  of the nuisance component  $\mathbf{f}$  that is suitable to describe the statistical behavior of these fluctuations.

We formalize the reconstruction of the object  $\mathbf{x}$  from the data  $\mathbf{r}$  as an inverse problem: We seek the light distribution  $\hat{\mathbf{x}}$  that would produce observations, given the instrumental model  $\mathbf{A}$ , statistically close to the actual measurements. Image reconstruction methods based on the resolution of inverse problems are generally expressed in the form of an optimization problem. In a Bayesian framework, the solution of this optimization problem is interpreted as the maximum a posteriori estimator (see for

example Thiébaud 2006). Introducing an objective function  $\mathcal{L}$  to account for the instrumental effects, the statistical model of the measurements and regularization terms that promote more satisfying solutions (e.g., sparser and smoother solutions) and better reject noise, the problem writes:

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x} \geq 0} \mathcal{L}(\mathbf{r}, \mathbf{x}, \mathbf{A}, \mathbf{\Omega}, \boldsymbol{\mu}), \quad (2)$$

where:

$$\mathcal{L}(\mathbf{r}, \mathbf{x}, \mathbf{A}, \mathbf{\Omega}, \boldsymbol{\mu}) = \mathcal{D}(\mathbf{r}, \mathbf{A}\mathbf{x}, \mathbf{\Omega}) + \mathcal{R}(\mathbf{x}, \boldsymbol{\mu}), \quad (3)$$

and  $\arg \min_{\mathbf{x} \geq 0} \mathcal{L}(\mathbf{r}, \mathbf{x}, \mathbf{A}, \mathbf{\Omega}, \boldsymbol{\mu})$  denotes the set of nonnegative values of  $\mathbf{x}$  minimizing  $\mathcal{L}(\mathbf{r}, \mathbf{x}, \mathbf{A}, \mathbf{\Omega}, \boldsymbol{\mu})$ . The first term,  $\mathcal{D}(\mathbf{r}, \mathbf{A}\mathbf{x}, \mathbf{\Omega})$ , is the data-fidelity term that penalizes the discrepancy between the recorded data  $\mathbf{r}$  and the modeled contribution  $\mathbf{A}\mathbf{x}$  of the object of interest  $\mathbf{x}$  in the data. The data fidelity term also depends on parameters  $\mathbf{\Omega}$  that define the statistical distribution of the nuisance component  $\mathbf{f}$  (e.g., mean, variances or covariances, see Sect. 2.3). The second term,  $\mathcal{R}(\mathbf{x}, \boldsymbol{\mu})$ , is a regularization term enforcing some prior knowledge about the object properties to favor physically more plausible reconstructions and prevent noise amplification. The hyper-parameters  $\boldsymbol{\mu}$  tune the behavior of the regularization, in particular, its relative importance compared to the data fidelity term. In the following, we detail the ingredients of this general framework. The forward model  $\mathbf{A}$  of the off-axis PSF is developed in Sect. 2.2. The statistical model of the nuisance component is stated in Sect. 2.3 while the estimation of the parameters  $\mathbf{\Omega}$  of this model is described in Sect. 2.4. Different possible regularizations are considered in Sect. 2.5. Finally, a strategy for the automatic tuning of the regularization hyper-parameters is presented in 3.2.

## 2.2. Forward model: off-axis point spread function

The parameters  $\mathbf{x} \in \mathbb{R}_+^M$  represent the distribution of light due to the off-axis sources (e.g., circumstellar disks, exoplanets, brown dwarfs, background stars). During the acquisition of an ADI sequence, the field of view as seen by the detector rotates around the optical axis. After suitable instrument calibration, this apparent rotation is completely defined by the center of rotation and

the parallactic angles recorded with the data<sup>4</sup>. The contribution of  $\mathbf{x}$  in the  $t$ -th data frame  $\mathbf{r}_t$  is given by the linear model  $\mathbf{A}_t \mathbf{x}$  where  $\mathbf{A}_t : \mathbb{R}^M \rightarrow \mathbb{R}^N$  implements the off-axis point spread function for the considered frame: it models the apparent rotation of the field of view, the nonuniform attenuation by the coronagraph and the instrumental blurring. In other words,  $[\mathbf{A} \mathbf{x}]_t \equiv \mathbf{A}_t \mathbf{x}$  for all frames  $t \in \llbracket 1, T \rrbracket$  where  $\mathbf{A} : \mathbb{R}^M \rightarrow \mathbb{R}^{N \times T}$  is the linear operator introduced in Sect. 2.1 and such that  $\mathbf{A} \mathbf{x}$  models the contribution of  $\mathbf{x}$  to all data frames.

Considering the size of the image reconstruction problem, it is essential that our model  $\mathbf{A}$  of the off-axis PSFs not only be an accurate approximation of the actual instrumental effects but also that  $\mathbf{A}$  and its adjoint be fast to apply in an iterative reconstruction algorithm (see Appendix A). To obtain a numerically efficient model, we decompose each operator  $\mathbf{A}_t$  into different factors:  $\mathbf{Q}_t$  to model the temporally and spatially variant effects due to the rotation of the field of view,  $\mathbf{\Gamma}$  to account for the attenuation close to the optical axis due to the coronagraph, and  $\mathbf{H}$  to implement the instrumental blurring (considered in first approximation as spatially and temporally invariant). Operators  $\mathbf{Q}_t$  can be implemented by interpolation operations while operator  $\mathbf{H}$  performs bi-dimensional convolutions that can be computed efficiently using fast Fourier transforms (FFTs). Modeling the instrumental blurring as being shift-invariant (a convolution) is accurate in most of the field of view except close to the optical axis. The coronagraph induces a reduction of the light transmission at short angular separations (i.e., under and near the coronagraphic mask) and a deformation of the PSF. We model the variable light transmission by a diagonal operator  $\mathbf{\Gamma} = \text{diag}(\gamma)$  with  $\gamma$  a 2D transmission mask. The mask  $\gamma$  has a smooth profile with all entries in the range  $[0, 1]$ ; from 0 at the center of the coronagraph (i.e., no flux from off-axis sources transmitted) to 1 farther away, and being equal to 0.5 at the inner working angle (IWA) of the coronagraph.

We neglect the deformation of the PSF in our model. This approximation has a limited effect since the area affected by the PSF deformations (close to or under the coronagraphic mask) is dominated by the nuisance component. An alternate way to handle the measurements in the area of the coronagraph would consist of assuming that they suffer an infinite variance (so that they be completely discarded in the subsequent estimations). Both approaches lead to comparable reconstructions (i.e., differences in the reconstructed flux are several orders of magnitude lower than the disk magnitude) in all experimental cases we have considered. This model of the effect of the coronagraph on off-axis point sources as a simple attenuation varying with the angular position is consistent (for angular separations larger than 0.05 arcsec) with the calibrations of the coronagraphs considered in Beuzit et al. (2019) for the SPHERE instrument. For the results presented in Sect. 3, the experimental calibration of this transmission has been performed in the H-Johnson’s spectral band (i.e., at  $\approx 1.6 \mu\text{m}$ ) and extrapolated to the other wavelengths by using the experimental measurement and dedicated simulations, as described in Beuzit et al. (2019).

To summarize, our forward model of the contribution of the object of interest  $\mathbf{x}$  in the  $t$ th frame writes<sup>5</sup>:

$$\mathbf{A}_t \mathbf{x} = \mathbf{V} \mathbf{H} \mathbf{\Gamma} \mathbf{Q}_t \mathbf{x}, \quad (4)$$

where  $\mathbf{V}$  is a simple truncation operator to extract the  $N$  pixels corresponding to the actual data field of view from the (larger)  $M$ -pixels output of blurring operator  $\mathbf{H}$ . Due to the apparent rotation of the field of view during the ADI sequence, previously unseen areas fall within the corners of the sensor and, by combining all measurements, a region larger than  $N$  pixels can be reconstructed (up to 57% more pixels).

The factorization of the forward operator  $\mathbf{A}_t$  can straightforwardly be generalized to the case of a time-varying blur  $\mathbf{H}_t$  or transmission  $\mathbf{\Gamma}_t$  to account for an evolution of the observing conditions during the ADI sequence or a possible partial decentering of the coronagraph. Figure 3 illustrates the different steps to convert the object of interest  $\mathbf{x}$  into the model of its contribution into the ADI sequence recorded by the high contrast imaging instrument. In this figure, we adopt the same convention as for  $\mathbf{A}$  that, when the frame index  $t$  of an operator is omitted, the result of applying the operator for all frames is considered.

### 2.3. Statistical model of the nuisance component

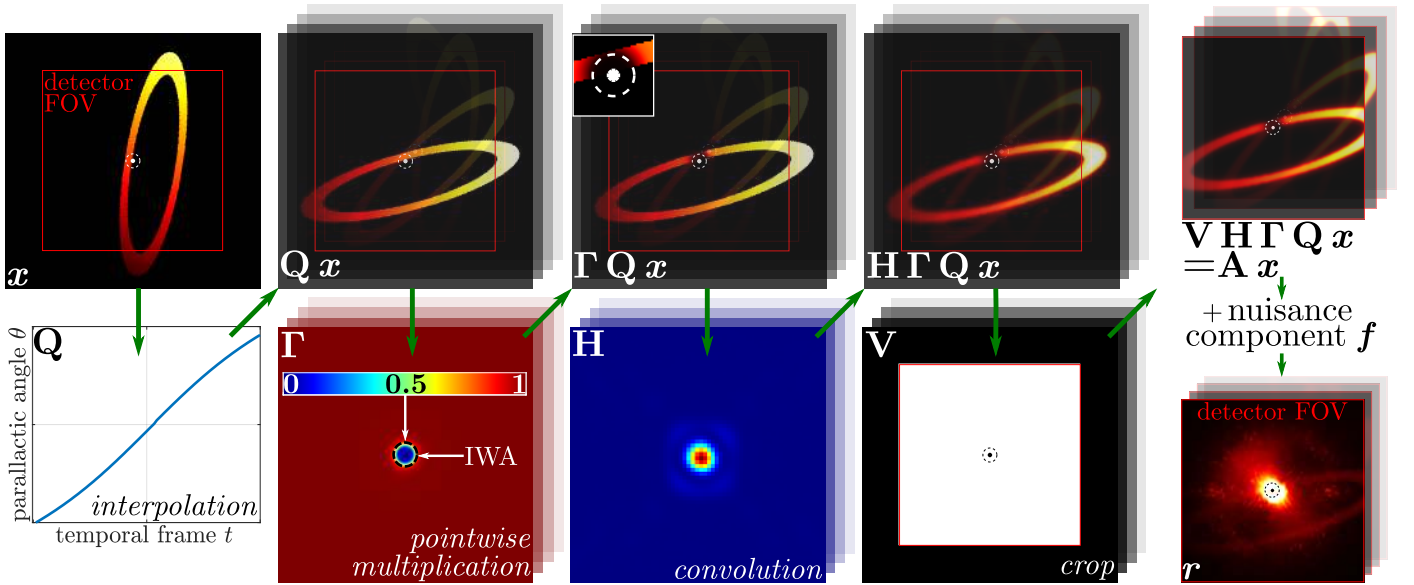
The nuisance component  $\mathbf{f}$  in ADI sequences generally presents strong and nonstationary spatial correlations (speckles) that fluctuate over time, especially near the host star where the stellar leakages dominate, see Fig. 2a. Also, the presence of several bad pixels (hot, dead or randomly fluctuating) is a common issue in high contrast imaging where the images are recorded by infrared detectors. Despite a prereduction stage identifying and correcting for defective pixels, some of them displaying large fluctuations only on a few frames remain after this processing (see for example Pavlov et al. 2008; Delorme et al. 2017 for the SPHERE instrument). We thus aim to account for the different causes of fluctuations of the nuisance component through a statistical model.

We recover an image of the objects of interest  $\mathbf{x}$  based on a co-log-likelihood term  $\mathcal{D}(\mathbf{r}, \mathbf{A} \mathbf{x}, \mathbf{\Omega})$  that captures the statistical distribution of the residuals between  $\mathbf{r}$  and the model  $\mathbf{A} \mathbf{x}$ . Like in our previous works on the PACO algorithm (Flasseur et al. 2018a,b,c, 2020a,b), we model the covariances locally at the scale of small patches of size  $K$  (a few tens of pixels in practice). This approximation is justified by the nonstationarity of the spatial correlations and can be seen as a trade-off between considering full-size spatial covariance matrices (of size  $N \times N$ , and hence too large to be estimated and stored in practice) and completely neglecting the spatial correlations. Our choice amounts to model the nuisance component  $\mathbf{f}_t$  of the  $t$ th frame as the realization of independent Gaussian processes for each  $K$ -pixel patch, without overlap. Given that model, the probability density function of the nuisance component is given by:

$$\forall t, \quad p(\mathbf{f}_t) \propto \prod_{n \in \mathbb{P}} \det^{-\frac{1}{2}}(\mathbf{C}_n) \times \exp\left[-\frac{1}{2} (\mathbf{P}_n(\mathbf{f}_t - \mathbf{m}))^\top \mathbf{C}_n^{-1} (\mathbf{P}_n(\mathbf{f}_t - \mathbf{m}))\right], \quad (5)$$

<sup>5</sup> The forward model could be written differently so that the operations involved in  $\mathbf{Q}_t$ ,  $\mathbf{\Gamma}$  and  $\mathbf{H}$  would be applied in a different order, subject to slight adaptations. For example, if the field rotation and translation modeled by operator  $\mathbf{Q}_t$  are applied after the convolution  $\mathbf{H}$ , the non-isotropic off-axis PSF has to be counter-rotated by the parallactic angles (i.e., the blurring is not the same for all  $t$ ) to counter-balance the effect of the permutation of  $\mathbf{Q}_t$  and  $\mathbf{H}$ . The same remark applies for the attenuation  $\mathbf{\Gamma}$  if the mask  $\gamma$  is not circular symmetrical.

<sup>4</sup> In practice for VLT/SPHERE-IRDIS, not all parallactic angles are measured and some of them are interpolated from others. Parallactic angles are also corrected for the estimated overheads of storage in a calibration and preprocessing step (Pavlov et al. 2008). Estimation errors of the rotation center and parallactic angles are not included in our statistical modeling.



**Fig. 3.** Schematic illustration of the forward model  $\mathbf{A} \mathbf{x} = \mathbf{V} \mathbf{H} \mathbf{\Gamma} \mathbf{Q} \mathbf{x}$  of REXPACO. First, the object  $\mathbf{x}$  is rotated by the parallactic angles by operator  $\mathbf{Q}$ . Then, each resulting temporal frame is attenuated by the (time independent) transmission of the coronagraph with  $\mathbf{\Gamma}$ , and convolved by the off-axis PSF by application of  $\mathbf{H}$ . Finally, the field of view seen by the sensor is extracted by  $\mathbf{V}$  to obtain the object contribution  $\mathbf{A} \mathbf{x}$ . The recorded ADI sequence is modeled by the sum of  $\mathbf{A} \mathbf{x}$  and of the nuisance component  $\mathbf{f}$ , as described by Eq. (1).

with  $\mathbf{P}_n : \mathbb{R}^N \rightarrow \mathbb{R}^K$  the operator extracting patch  $n$  from a given temporal frame for any index  $n$  in  $\mathbb{P}$  defining the partition of the data frame into nonoverlapping  $K$ -pixel patches, and  $\mathbb{P}$  is the list of the indices of the pixels at the center of the nonoverlapping patches. The vector  $\mathbf{m} \in \mathbb{R}^N$  denotes the temporal mean  $\mathbb{E}_t(\mathbf{f}_t)$  while  $\mathbf{C}_n \in \mathbb{R}^{K \times K}$  denotes the temporal covariance matrix within the patch  $n$ .

For  $\mathbf{m}$  and  $\mathbf{C}_n$ , we use the same estimators that proved successful with PACO for exoplanet detection: the temporal mean  $\mathbf{m}$  is estimated by the sample mean  $\widehat{\mathbf{m}}$  defined by:

$$\widehat{\mathbf{m}} = \frac{1}{T} \sum_{t=1}^T (\mathbf{r}_t - \mathbf{A}_t \mathbf{x}), \quad (6)$$

and the covariance matrices  $\mathbf{C}_n$  are estimated using a shrinkage estimator (Ledoit & Wolf 2004; Chen et al. 2010):

$$\widehat{\mathbf{C}}_n = (1 - \widetilde{\rho}_n) \widetilde{\mathbf{C}}_n + \widetilde{\rho}_n \widetilde{\mathbf{F}}_n, \quad (7)$$

that combines the high-variance and low-bias sample covariance estimator  $\widetilde{\mathbf{C}}_n$  with the low-variance and high-bias estimator  $\widetilde{\mathbf{F}}_n$  that neglects all covariances. In other words,  $\widetilde{\mathbf{F}}_n$  is a diagonal matrix containing the pixel sample variances that are the diagonal entries of the sample covariance of the considered patch:

$$\widetilde{\mathbf{C}}_n = \frac{1}{T} \sum_{t=1}^T \widehat{\mathbf{u}}_{n,t} \widehat{\mathbf{u}}_{n,t}^\top, \quad (8)$$

with:

$$\widehat{\mathbf{u}}_{n,t} = \mathbf{P}_n (\mathbf{r}_t - \widehat{\mathbf{m}} - \mathbf{A}_t \mathbf{x}), \quad (9)$$

the residuals in patch  $n$  of frame  $t$ . With this specific choice for  $\widetilde{\mathbf{F}}_n$ , Eq. (7) rewrites:

$$\widehat{\mathbf{C}}_n = \widetilde{\mathbf{W}}_n \odot \widetilde{\mathbf{C}}_n, \quad (10)$$

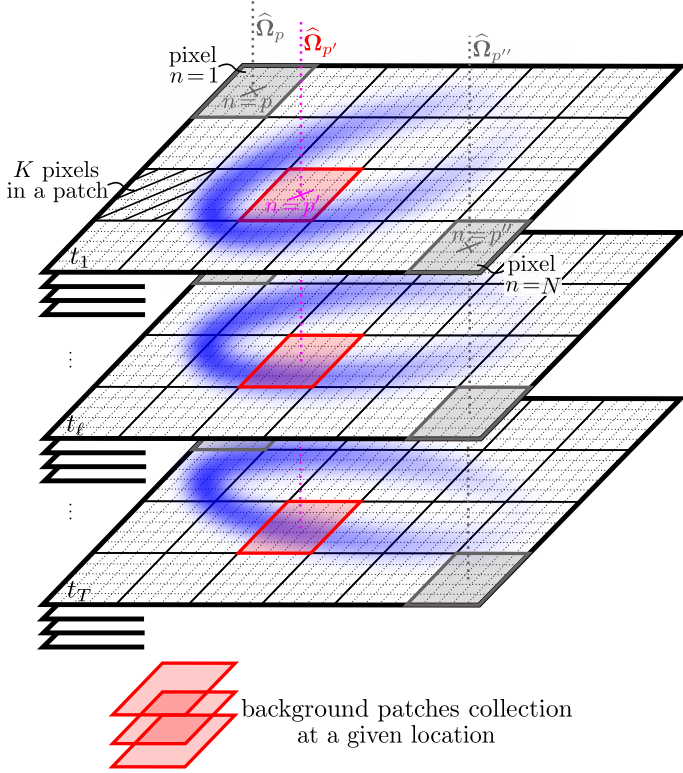
where  $\odot$  stands for the Hadamard product (i.e., entrywise multiplication), and  $\widetilde{\mathbf{W}}_n \in \mathbb{R}^{K \times K}$  is defined for  $\{k, k'\} \in \llbracket 1, K \rrbracket^2$  by:

$$[\widetilde{\mathbf{W}}_n]_{k,k'} = \begin{cases} 1 & \text{if } k = k', \\ 1 - \widetilde{\rho}_n & \text{if } k \neq k'. \end{cases} \quad (11)$$

The shrinkage parameter  $\widetilde{\rho}_n$  balances each estimator to reach a bias-variance trade-off. The extension of the results of Chen et al. (2010) to the diagonal covariance matrix  $\widetilde{\mathbf{F}}_n$  leads to the following data-driven expression for the shrinkage parameter  $\widetilde{\rho}_n$  (see Eq. (12) of Flasseur et al. 2018a):

$$\widetilde{\rho}_n = \frac{\text{tr}(\widetilde{\mathbf{C}}_n^2) + \text{tr}^2(\widetilde{\mathbf{C}}_n) - 2 \sum_{k=1}^K [\widetilde{\mathbf{C}}_n]_{k,k}^2}{(T+1) (\text{tr}(\widetilde{\mathbf{C}}_n^2) - \sum_{k=1}^K [\widetilde{\mathbf{C}}_n]_{k,k}^2)}. \quad (12)$$

It can be noted that most above defined quantities –  $\widehat{\mathbf{m}}$  in Eq. (6),  $\widetilde{\mathbf{C}}_n$  in Eq. (7),  $\widetilde{\mathbf{C}}_n$  in Eq. (8),  $\widehat{\mathbf{u}}_{n,t}$  in Eq. (9),  $\widetilde{\mathbf{W}}_n$  in Eq. (11) and  $\widetilde{\rho}_n$  in Eq. (12) – depend implicitly on the sought parameters  $\mathbf{x}$  although, and for the sake of simplicity, this is not always explicitly indicated. In Algorithms 1 and 2 presented in Sect. 2.4, these quantities are labeled with the iteration number as the parameters  $\mathbf{x}$  change at each iteration. Figure 4 illustrates the extraction of a collection of patches from an ADI sequence for locally learning the statistics of the nuisance component, in particular for the estimation of  $\widetilde{\mathbf{C}}_n$ . The number  $K$  of pixels in each patch, and hence the dimension of the local covariance matrices should be large enough to encompass the core of the off-axis PSF and yet not too large compared to the number  $T$  of samples available for the estimation. The optimal patch size is estimated using the same criterion (validated empirically) as in PACO and which yields an optimal patch size corresponding roughly to twice the off-axis PSF full width at half maximum (FWHM). This rule typically leads to  $50 \lesssim K \lesssim 100$ . Figure 2c gives examples of estimated patch covariance matrices  $\widehat{\mathbf{C}}_n$  for different angular separations. It shows that the structure of the correlations strongly depends on the distance to the star and that neglecting these correlations would be a crude approximation, especially near the star where the nuisance component is highly correlated and fluctuating. Figure 2c also emphasizes that the covariances are well preserved by the regularization (i.e.,  $\widetilde{\rho}_n$  is small enough



**Fig. 4.** Extraction of patch collections for local learning of the statistics of the nuisance component from ADI sequences. The contribution of off-axis sources is shown in blue.

for  $\tilde{\mathbf{C}}_n$  to represent the largest contribution). We have shown in Flasseur et al. (2018a, 2020b) that such a local model of the covariances approximates well the empirical distribution of the nuisance component. In Flasseur et al. (2020b), a small discrepancy was observed near the star due to the larger fluctuations<sup>6</sup> of the nuisance component in this area. The impact of a statistical mis-modeling in this area is further discussed in Sect. 3.3.2.

Given the patch-based statistics  $\hat{\mathbf{\Omega}} = \{\hat{\mathbf{m}}, \{\hat{\mathbf{C}}_n\}_{n \in \mathbb{P}}\}$  locally accounting for the fluctuations of the nuisance component, the co-log-likelihood rewrites:

$$\mathcal{D}(\mathbf{r}, \mathbf{A} \mathbf{x}, \hat{\mathbf{\Omega}}) = \frac{1}{2} \sum_{n \in \mathbb{P}} \left( \sum_{t=1}^T \|\hat{\mathbf{u}}_{n,t}\|_{\hat{\mathbf{C}}_n}^2 + T \log \det \hat{\mathbf{C}}_n \right), \quad (13)$$

where the summation over  $n$  is performed on the list  $\mathbb{P}$  of nonoverlapping  $K$ -pixel patches while the residuals  $\hat{\mathbf{u}}_{n,t}$  in the patch  $n$  for the frame  $t$  are defined in Eq. (9).

#### 2.4. Unbiased estimation of the mean and spatial covariances

The estimators  $\hat{\mathbf{m}}$ , in Eq. (6), and  $\hat{\mathbf{C}}_n$ , in Eq. (7), that appear in Eq. (13) both depend on the object flux  $\mathbf{x}$  which is unknown. To

<sup>6</sup> Large fluctuations in the nuisance component can be due to a slight decentering of the coronagraph or to a sudden degradation of the observing conditions. Finer statistical models of the nuisance component could be considered to account for these fluctuations. In particular, multivariate Gaussian scale mixture models (GSM; Wainwright & Simoncelli 2000) have been shown to be very effective in the context of exoplanet detection (Flasseur et al. 2020a,b). Replacing the multivariate Gaussian model consider in this paper by a GSM model requires specific developments that are left for future work.

---

#### Algorithm 1: REXPACO reconstruction (Alternating estimation approach).

---

**Input:** ADI sequence  $\mathbf{r}$ .

**Input:** Forward operator  $\mathbf{A}$ .

**Input:** Regularization parameters  $\boldsymbol{\mu}$ .

**Input:** Relative precision  $\eta \in (0, 1)$ .

**Output:** Unbiased estimate  $\hat{\mathbf{x}}$  of the light distribution of off-axis sources.

```

 $\hat{\mathbf{x}}^{[0]} \leftarrow \mathbf{0}_M$ 
 $i \leftarrow 0$ 
do
     $i \leftarrow i + 1$ 
    ▶ Step 1. Learn statistics of nuisance term.
     $\hat{\mathbf{m}}^{[i]} \leftarrow \frac{1}{T} \sum_{t=1}^T (\mathbf{r}_t - \mathbf{A}_t \hat{\mathbf{x}}^{[i-1]})$  ◀ Eq. (6)
    for  $n \in \mathbb{P}$  do
        for  $t \in \llbracket 1, T \rrbracket$  do
             $\mathbf{u}_{n,t}^{[i]} \leftarrow \mathbf{P}_n (\mathbf{r}_t - \hat{\mathbf{m}}^{[i]} - \mathbf{A}_t \hat{\mathbf{x}}^{[i-1]})$ 
             $\tilde{\mathbf{C}}_n^{[i]} \leftarrow \frac{1}{T} \sum_{t=1}^T \mathbf{u}_{n,t}^{[i]} \mathbf{u}_{n,t}^{[i] \top}$  ◀ Eq. (8)
             $\tilde{\rho}_n^{[i]} \leftarrow \frac{\text{tr}(\tilde{\mathbf{C}}_n^{[i]2}) + \text{tr}^2(\tilde{\mathbf{C}}_n^{[i]}) - 2 \sum_{k=1}^K [\tilde{\mathbf{C}}_n^{[i]}]_{k,k}^2}{(T+1)(\text{tr}(\tilde{\mathbf{C}}_n^{[i]2}) - \sum_{k=1}^K [\tilde{\mathbf{C}}_n^{[i]}]_{k,k}^2)}$  ◀ Eq. (12)
            for  $(k, k') \in \llbracket 1, K \rrbracket^2$  do
                 $[\tilde{\mathbf{W}}_n^{[i]}]_{k,k'} \leftarrow \begin{cases} 1 & \text{if } k = k' \\ 1 - \tilde{\rho}_n & \text{if } k \neq k' \end{cases}$  ◀ Eq. (11)
             $\hat{\mathbf{C}}_n^{[i]} \leftarrow \tilde{\mathbf{W}}_n^{[i]} \circ \tilde{\mathbf{C}}_n^{[i]}$  ◀ Eq. (10)
             $\hat{\mathbf{\Omega}}^{[i]} \leftarrow \{\hat{\mathbf{m}}^{[i]}, \{\hat{\mathbf{C}}_n^{[i]}\}_{n \in \mathbb{P}}\}$ 
        ▶ Step 2. Reconstruct an image of the objects.
         $\hat{\mathbf{x}}^{[i]} \leftarrow \arg \min_{\mathbf{x} \geq 0} \mathcal{D}(\mathbf{r}, \mathbf{A} \mathbf{x}, \hat{\mathbf{\Omega}}^{[i]}) + \mathcal{R}(\mathbf{x}, \boldsymbol{\mu})$  ◀ Eq. (2)
    while  $\|\hat{\mathbf{x}}^{[i]} - \hat{\mathbf{x}}^{[i-1]}\| > \eta \|\hat{\mathbf{x}}^{[i]}\|$ 

```

---

handle this problem, the simplest solution would be to neglect the object contribution as it is typically small compared to the nuisance component  $\mathbf{f}$ . Under this assumption, Eqs. (6) and (8) rewrite:

$$\hat{\mathbf{m}}^{[1]} = \frac{1}{T} \sum_{t=1}^T \mathbf{r}_t, \quad (14a)$$

$$\tilde{\mathbf{C}}_n^{[1]} = \frac{1}{T} \sum_{t=1}^T (\mathbf{P}_n (\mathbf{r}_t - \hat{\mathbf{m}}^{[1]})) (\mathbf{P}_n (\mathbf{r}_t - \hat{\mathbf{m}}^{[1]}))^{\top}, \quad (14b)$$

and correspond respectively to the sample mean and the sample covariances learned locally from the image patches.

We show in Appendix B that neglecting the light flux of off-axis objects when characterizing the nuisance component leads to biased estimations, a problem common to many ADI post-processing methods and known as the self-subtraction problem (Milli et al. 2012; Pairet et al. 2019). To recover a better estimation of the object flux, it is then necessary to develop an unbiased estimation of the statistics  $\mathbf{\Omega}$ . To do so, two strategies are possible: (i) an alternation between the reconstruction of  $\hat{\mathbf{x}}$  and the update of the statistics  $\hat{\mathbf{\Omega}}$  of the nuisance component  $\mathbf{f}$  from the residuals  $\mathbf{r} - \mathbf{A} \hat{\mathbf{x}}$ , until convergence, as done in Algorithm 1, or (ii) a joint estimation approach implemented by Algorithm 2. In this paper, we used  $\eta = 10^{-8}$  in the stopping condition of Algorithm 1 for all the reconstructions. Approach (ii) implemented by

---

**Algorithm 2:** REXPACO reconstruction  
(Joint estimation approach).

---

**Input:** ADI sequence  $\mathbf{r}$ .

**Input:** Forward operator  $\mathbf{A}$ .

**Input:** Regularization parameters  $\boldsymbol{\mu}$ .

**Output:** Unbiased estimate  $\widehat{\mathbf{x}}$  of the light distribution of off-axis sources.

► **Step 1.** Compute shrinkage matrices.

$$\{\widetilde{\mathbf{W}}_n\}_{n \in \mathbb{P}} \\ \widehat{\mathbf{m}}^{[1]} = \frac{1}{T} \sum_{t=1}^T \mathbf{r}_t \quad \leftarrow \text{Eq. (14a)}$$

**for**  $n \in \mathbb{P}$  **do**

$$\left[ \begin{array}{l} \widetilde{\mathbf{C}}_n^{[1]} \leftarrow \frac{1}{T} \sum_{t=1}^T \mathbf{P}_n (\mathbf{r}_t - \widehat{\mathbf{m}}^{[1]}) (\mathbf{r}_t - \widehat{\mathbf{m}}^{[1]})^\top \mathbf{P}_n^\top \quad \leftarrow \text{Eq. (14b)} \\ \widehat{\rho}_n^{[1]} \leftarrow \frac{\text{tr}(\widetilde{\mathbf{C}}_n^{[1]2}) + \text{tr}^2(\widetilde{\mathbf{C}}_n^{[1]}) - 2 \sum_{k=1}^K [\widetilde{\mathbf{C}}_n^{[1]}]_{k,k}^2}{(T+1)(\text{tr}(\widetilde{\mathbf{C}}_n^{[1]2}) - \sum_{k=1}^K [\widetilde{\mathbf{C}}_n^{[1]}]_{k,k}^2)} \quad \leftarrow \text{Eq. (12)} \\ \text{for } (k, k') \in \llbracket 1, K \rrbracket^2 \text{ do} \\ \left[ \begin{array}{l} [\widetilde{\mathbf{W}}_n^{[1]}]_{k,k'} \leftarrow \begin{cases} 1 & \text{if } k = k' \\ 1 - \widehat{\rho}_n & \text{if } k \neq k' \end{cases} \quad \leftarrow \text{Eq. (11)} \end{array} \right. \end{array} \right.$$

► **Step 2.** Joint minimization with fixed  $\{\widetilde{\mathbf{W}}_n\}_{n \in \mathbb{P}}$ .

$$\widehat{\mathbf{x}} \leftarrow \arg \min_{\mathbf{x} \geq \mathbf{0}} \mathcal{D}_{\text{joint}}(\mathbf{r}, \mathbf{x}) + \mathcal{R}(\mathbf{x}, \boldsymbol{\mu}) \quad \leftarrow \text{Eqs. (2) and (15)}$$


---

Algorithm 2 converges faster and is especially beneficial when reconstructing structures that are close to circular symmetry such as in Figs. 7 and 8. In order to be equivalent to Algorithm 1, Algorithm 2 requires replacing the data-fidelity term  $\mathcal{D}$  defined in Eq. (13) by (the proof is detailed in Appendix C):

$$\mathcal{D}_{\text{joint}}(\mathbf{r}, \mathbf{x}) = \frac{T}{2} \sum_{n \in \mathbb{P}} \log \det \widehat{\mathbf{C}}_n \\ + \frac{1}{2} \sum_{n \in \mathbb{P}} \text{tr}(\widehat{\mathbf{C}}_n^{-1} (\widetilde{\mathbf{W}}_n \odot \sum_{t=1}^T \widehat{\mathbf{u}}_{n,t} \widehat{\mathbf{u}}_{n,t}^\top)), \quad (15)$$

with  $\widehat{\mathbf{u}}_{n,t}$  the residuals in the patch  $n$  of frame  $t$  defined in Eq. (9). This modified data-fidelity corresponds to the co-log-likelihood under a Gaussian assumption except that shrinkage matrices  $\widetilde{\mathbf{W}}_n$  are introduced<sup>7</sup> so that the shrunk covariances  $\widehat{\mathbf{C}}_n$  defined in Eq. (10) are minimizers of  $\mathcal{D}_{\text{joint}}$ . The mean  $\widehat{\mathbf{m}}$  and shrunk covariances  $\widehat{\mathbf{C}}_n$  can thus be replaced by their closed-form expressions given in Eqs. (6) and (10) that depend on  $\mathbf{x}$  and the minimization be performed solely on  $\mathbf{x}$ . Beyond the faster convergence of Algorithm 2, the two algorithms differ slightly due to the shrinkage factors not being refined throughout the reconstruction in Algorithm 2. We found the impact of that discrepancy to be barely noticeable in the reconstruction results and hence would recommend the general usage of Algorithm 2 for its improved speed.

## 2.5. Regularization term

After discussing the data-fidelity term  $\mathcal{D}(\mathbf{r}, \mathbf{A}\mathbf{x}, \boldsymbol{\Omega})$  in the previous sections, we focus here on the regularization term  $\mathcal{R}(\mathbf{x}, \boldsymbol{\mu})$  and on the resolution of the inverse problem (2).

<sup>7</sup> Shrinkage matrices  $\{\widetilde{\mathbf{W}}_n\}_{n \in \mathbb{P}}$  are defined in Eq. (11) and we make the simplifying assumption that they are independent from  $\mathbf{x}$  so that they are estimated once for all given the data  $\mathbf{r}$ .

Regularization terms are introduced to enforce prior knowledge about the unknown object  $\mathbf{x}$  and to improve the conditioning of the inversion. Two classical regularizations are considered in the proposed method. The first one promotes smooth objects with sharp edges by favoring the sparsity of their spatial gradients. This is the goal of the so-called  $\ell_2 - \ell_1$  edge-preserving regularization (Charbonnier et al. 1997; Mugnier et al. 2004; Thiébaud 2006) widely used in image processing:

$$\mathcal{R}_{\text{smooth}}(\mathbf{x}, \epsilon) = \sum_{n=1}^N \sqrt{\|\Delta_n \mathbf{x}\|_2^2 + \epsilon^2}, \quad (16)$$

where  $\Delta_n$  approximates the spatial gradient at pixel  $n$  by finite differences and  $\epsilon > 0$  sets the edge-preservation behavior of the regularization ( $\epsilon = \sqrt{10^{-7}}$  in all the experiments of this paper): Local differences  $\Delta_n \mathbf{x}$  that are below  $\epsilon$  in magnitude are smoothed similarly as a quadratic regularization would, while larger differences are preserved (e.g., border of a disk). This classical regularization is sufficiently flexible to remain adapted to different disk morphologies, as illustrated by Figs. 5–8 where elliptical disks with sharp edges and spiral disks with smooth edges are considered. Since we aim to reconstruct a well-contrasted object on a dark background, we also consider a second regularization term promoting the sparsity of the object  $\mathbf{x}$  by penalizing its separable  $\ell_1$ -norm:

$$\mathcal{R}_{\ell_1}(\mathbf{x}) = \sum_{n=1}^N |x_n| = \sum_{x \geq 0} x_n. \quad (17)$$

This sparsity-promoting regularization term is also well adapted to restore point-like sources. We discuss in Sect. 4 how an alternating estimation strategy can be applied to further improve the separation of overlapping point-like sources and extended structures. Due to the positivity constraint in Eq. (2) that guarantees that reconstructed light fluxes are nonnegative (i.e.,  $\mathbf{x} \geq \mathbf{0}$ ), the  $\ell_1$ -norm in Eq. (17) boils down to  $\sum_n x_n$  which is a differentiable expression and smooth optimization techniques are applicable (see Appendix A). The two terms  $\mathcal{R}_{\text{smooth}}$  and  $\mathcal{R}_{\ell_1}$  are combined by:

$$\mathcal{R}(\mathbf{x}, \boldsymbol{\mu}) = \mu_{\text{smooth}} \mathcal{R}_{\text{smooth}}(\mathbf{x}, \epsilon) + \mu_{\ell_1} \mathcal{R}_{\ell_1}(\mathbf{x}), \quad (18)$$

where  $\boldsymbol{\mu} = \{\mu_{\text{smooth}}, \mu_{\ell_1}\}$  balances their relative weight with respect to the data-fidelity term  $\mathcal{D}$ . For a given set of parameters  $\boldsymbol{\mu}$ , we solve the constrained minimization problem (2) by running the VMLM-B algorithm (Thiébaud 2002). Some technical elements about this minimization technique can be found in Appendix A.

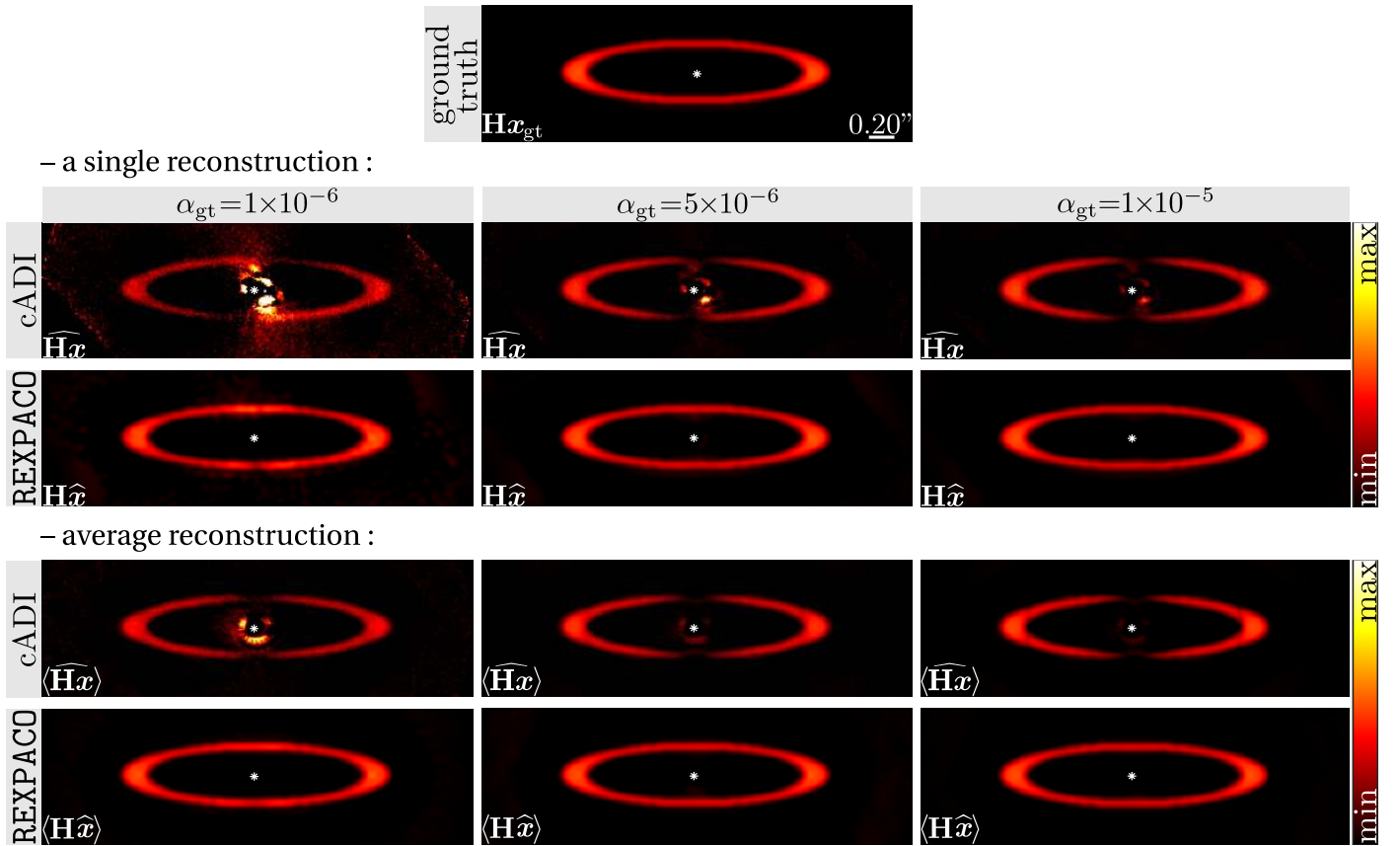
## 3. Results

### 3.1. Evaluation on simulated disks

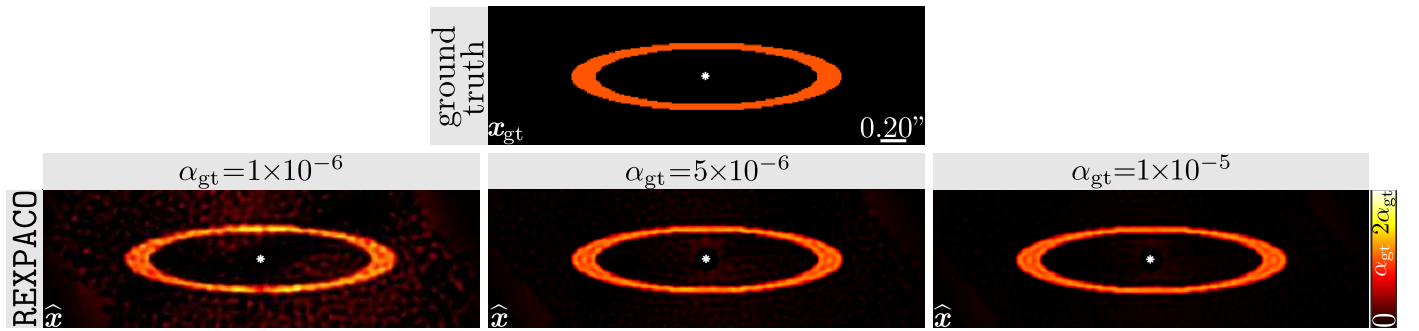
We first evaluate quantitatively, on numerical simulations, the ability of REXPACO to estimate a faithful image of the light flux distribution of off-axis sources. We consider two simulated disks with very different morphologies: (i) a spatially centered elliptical disk with an eccentricity about 0.95 and with sharp edges; (ii) a spiral disk with two arms whose center is shifted by ten pixels from the star center in one of two spatial directions. Contrary to the elliptical disk, the spiral disk has smooth edges. These simulated disks resemble the actual circumstellar disks presented in Sect. 3.3.2 so that our simulations can help to assess the quality of the reconstructions of real circumstellar disks.



(a) reconstructions at the resolution of the instrument :



(b) a deblurred reconstruction :

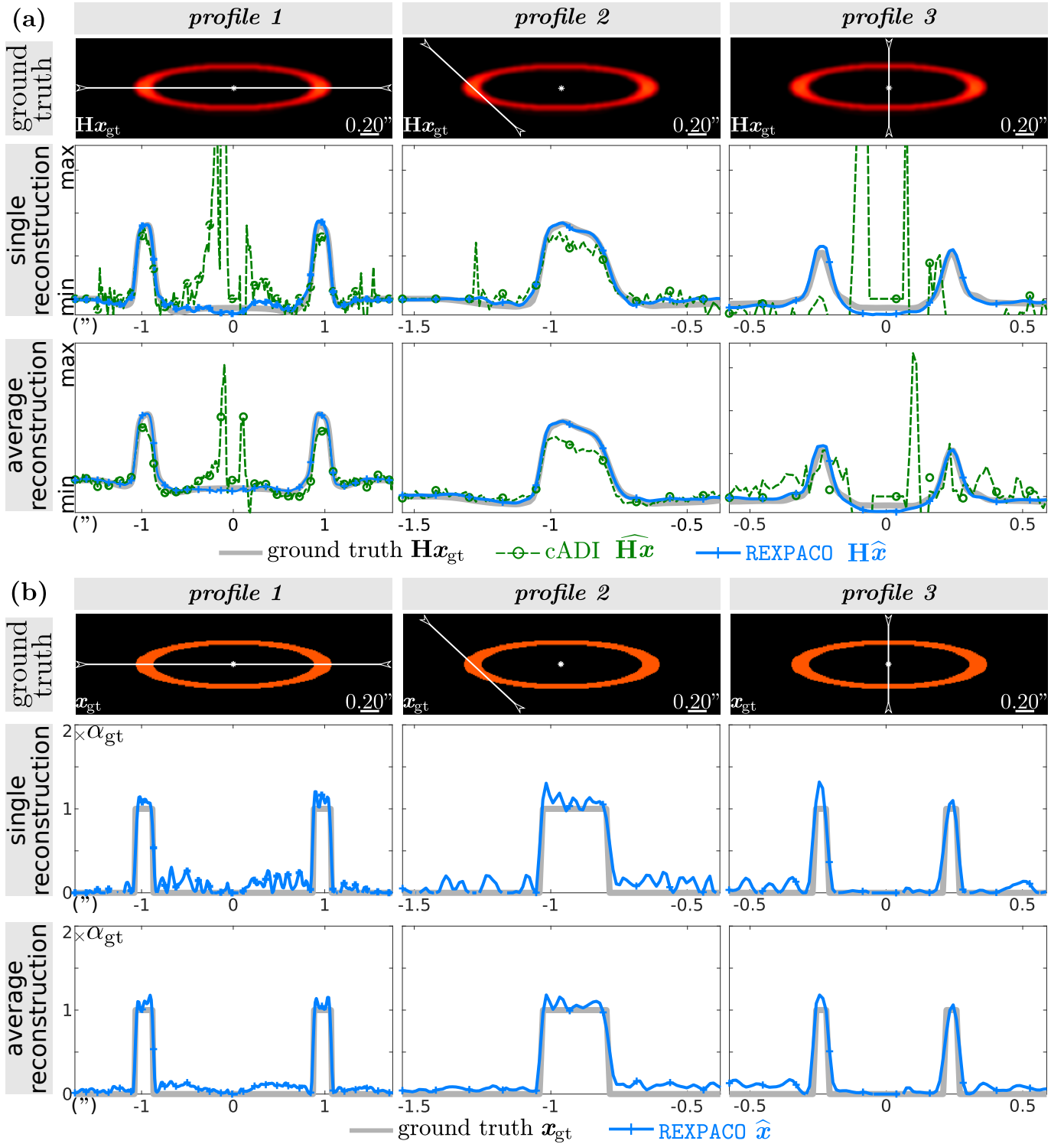


**Fig. 5.** Reconstructions of a simulated elliptical disk: (a) comparisons between cADI image combinations and reblurred REXPACO reconstructions; (b) resolution improvement achieved by deconvolution with REXPACO. The average reconstructions are computed over ten injections of the simulated disk with various orientations with respect to the background.

Each simulated disk is injected into a data set of HIP 80019 with no known off-axis source, at three different contrast levels:  $\alpha_{\text{gt}} \in \{1 \times 10^{-6}, 5 \times 10^{-6}, 1 \times 10^{-5}\}$ . Reconstructions of the elliptical disk are performed with Algorithm 1 while the spiral disk that requires many more iterations of Algorithm 1 to converge has been processed with Algorithm 2. A total of 60 reconstructions have been performed: for each disk and each level  $\alpha_{\text{gt}}$ , the simulated disk has been injected in one of ten different orientations with respect to the background in order to evaluate the mean and variance of the reconstructions.

Simulations of the elliptical disk are reported in Figs. 5 and 6. We compare REXPACO reconstructions to the cADI image combination technique. Since cADI does not perform a deconvolution, the comparison is performed at the resolution of the instrument,

that is to say the images produced by cADI are shown next to the reblurred reconstructions  $\widehat{\mathbf{H}\mathbf{x}}$  of REXPACO in Figs. 5a and 6a. Large errors can be noted in cADI images, in particular at small angular separations. The computation of the average reconstruction obtained for the ten different orientations of the disk with respect to the background indicates the presence of systematic errors with cADI: an under-estimation of the light-flux in the region of the disk that is closest to the star (an issue due to limited angular diversity). In contrast, REXPACO reconstructions are close to the ground truth, even for the lower level of contrast  $\alpha_{\text{gt}} = 10^{-6}$ . The deblurred reconstructions shown in Figs. 5b and 6b are in good agreement with the ground truth. Unsurprisingly, the reconstruction quality is higher when the disk is brighter: more spurious fluctuations are visible in the deblurred



**Fig. 6.** Line profiles extracted from the reconstructions at  $\alpha_{\text{gt}} = 10^{-6}$  shown in Fig. 5.

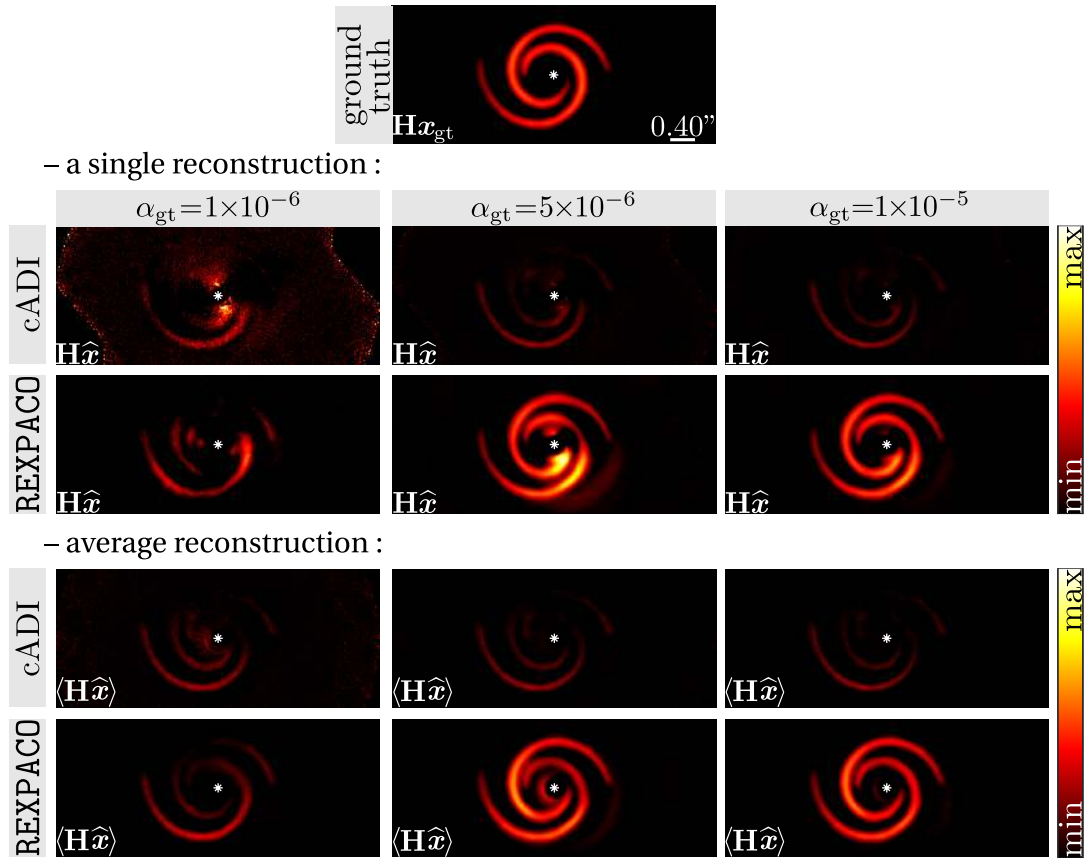
reconstruction at  $\alpha_{\text{gt}} = 10^{-6}$  than at  $\alpha_{\text{gt}} = 5 \times 10^{-6}$  or  $\alpha_{\text{gt}} = 10^{-5}$ . Although some discrepancies can be noted in the deblurred line profiles, the resolution is improved by the deconvolution process.

Simulations of the spiral disk are reported in Figs. 7 and 8. The comparison with cADI shows a clear improvement of the reconstructions with REXPACO: cADI fails to recover most of the disk. At the lowest contrast  $\alpha_{\text{gt}} = 10^{-6}$  and owing to the near circular-symmetry of the structures close to the star, the spiral disk is challenging to disentangle from the nuisance component. REXPACO is

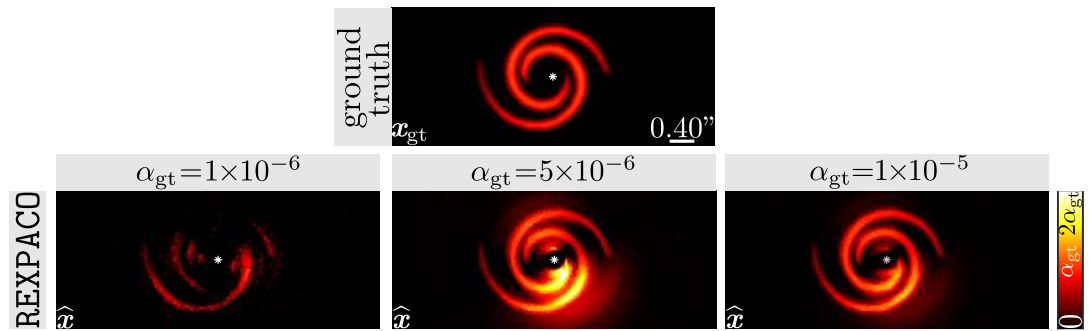
however able to recover some parts of the spiral arms although their flux is underestimated. At the highest contrast  $\alpha_{\text{gt}} = 10^{-5}$ , line profiles of Fig. 8 indicate that REXPACO reconstructions are strongly improved compared to cADI and that some errors remain in particular at the lowest angular separations.

The reconstructions are quantitatively evaluated by reporting the normalized root mean square error (N-RMSE, the lower the better) computed over different regions (the disk, the background, and the whole image):

(a) reconstructions at the resolution of the instrument :



(b) a deblurred reconstruction :



**Fig. 7.** Reconstructions of a simulated spiral disk: (a) comparisons between cADI image combinations and reblurred REXPACO reconstructions; (b) resolution improvement achieved by deconvolution with REXPACO. The average reconstructions are computed over ten injections of the simulated disk with various orientations with respect to the background.

$$N - \text{RMSE}(\mathbf{x}_{\text{gt}}, \hat{\mathbf{x}}) = \frac{\|\mathbf{x}_{\text{gt}} - \hat{\mathbf{x}}\|_2}{\|\mathbf{x}_{\text{gt}}\|_2}. \quad (19)$$

Table 2 reports N-RMSE values and indicates a clear improvement over cADI: the errors are typically divided by a factor of three to four (the error reduction is more modest for the more challenging reconstructions of the spiral disk at the lowest fluxes).

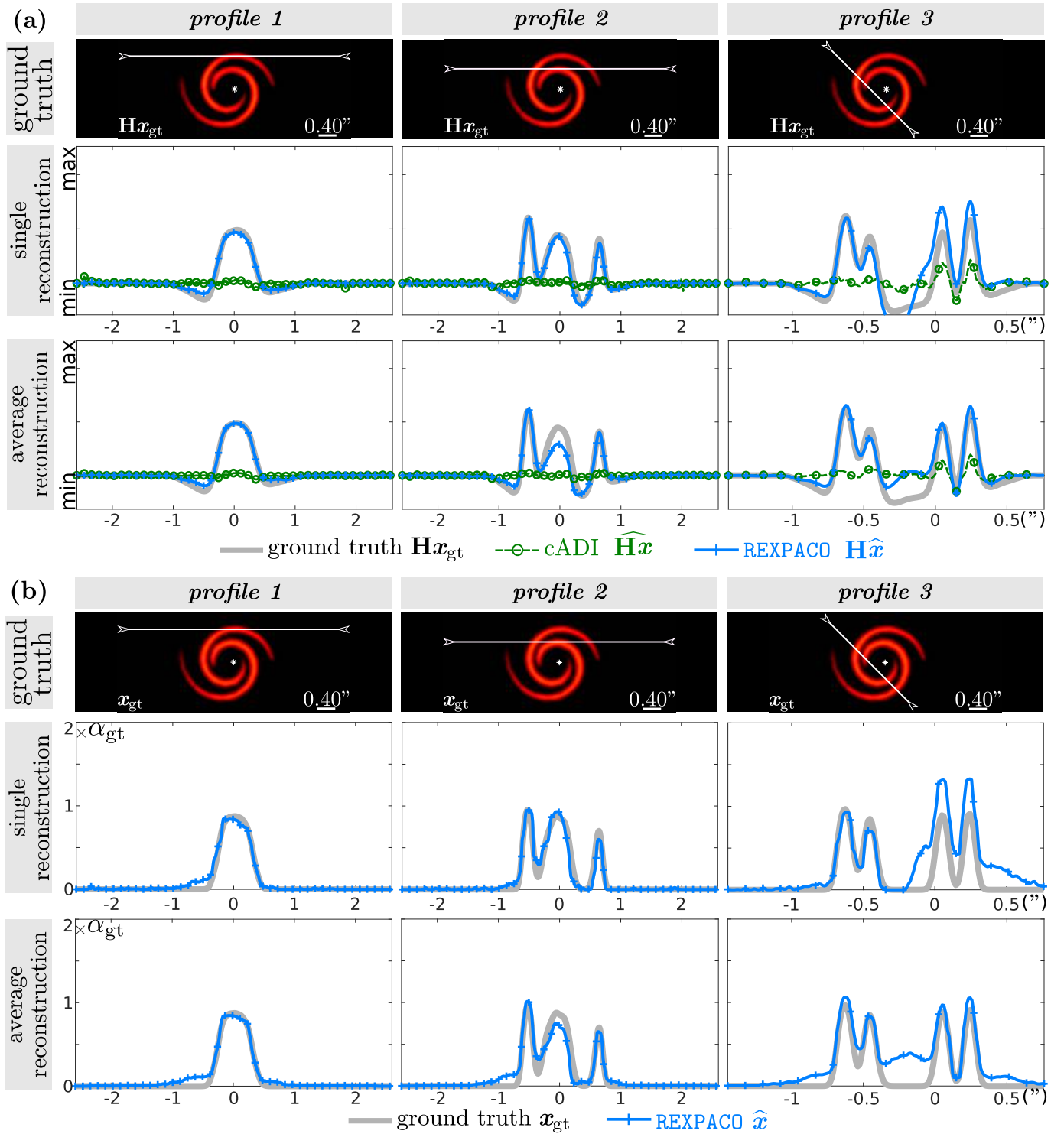
### 3.2. Automatic setting of the regularization parameters

The setting of the parameters  $\boldsymbol{\mu}$  is of uttermost importance since the results of the regularized reconstructions depend both on the characteristics of the data set (e.g., observing conditions, amplitude of the field rotation) and on the properties of the objects to

reconstruct (e.g., morphology, contrast). The parameters  $\boldsymbol{\mu}$  can be tuned manually by trial and error until the reconstruction is qualitatively acceptable, but this approach is not completely satisfactory since it is time-consuming and user-dependent. Several methods have been developed in the literature to automatically tune regularization parameters by minimizing a quantitative criterion (Craven & Wahba 1978; Wahba 1985; Stein 1981). In this work, we consider the Stein’s unbiased risk estimator approach (SURE; Stein 1981), which minimizes an estimate of the mean-square error (MSE) in the data space:

$$\text{MSE}(\boldsymbol{\mu}) = \|\mathbf{A}(\mathbf{x}_{\text{gt}} - \hat{\mathbf{x}}_{\boldsymbol{\mu}}(\mathbf{r}))\|_{\mathbf{C}^{-1}}^2, \quad (20)$$

where  $\mathbf{x}_{\text{gt}}$  is the (unknown) ground truth object, and  $\hat{\mathbf{x}}_{\boldsymbol{\mu}}(\mathbf{r})$  denotes the reconstructed object  $\hat{\mathbf{x}}$  obtained from the data  $\mathbf{r}$  when



**Fig. 8.** Line profiles extracted from the reconstructions at  $\alpha_{gt} = 10^{-5}$  shown in Fig. 7.

using regularization parameters  $\boldsymbol{\mu}$ . For any given set of parameters  $\boldsymbol{\mu}$ , the criterion  $\text{SURE}(\boldsymbol{\mu})$  provides an unbiased estimation of  $\text{MSE}(\boldsymbol{\mu})$  that does not require knowledge of the true object  $\boldsymbol{x}_{gt}$  (Stein 1981):

$$\text{SURE}(\boldsymbol{\mu}) = \|\boldsymbol{r} - \widehat{\boldsymbol{m}} - \mathbf{A} \widehat{\boldsymbol{x}}_{\boldsymbol{\mu}}(\boldsymbol{r})\|_{C^{-1}}^2 + 2 \text{tr}(\mathbf{A} \mathbf{J}_{\boldsymbol{\mu}}(\boldsymbol{r})) - NT, \quad (21)$$

where  $\mathbf{J}_{\boldsymbol{\mu}}(\boldsymbol{r}) = \partial \widehat{\boldsymbol{x}}_{\boldsymbol{\mu}}(\boldsymbol{r}) / \partial \boldsymbol{r}$  is the Jacobian matrix of the mapping  $\boldsymbol{r} \rightarrow \widehat{\boldsymbol{x}}_{\boldsymbol{\mu}}(\boldsymbol{r})$  with respect to the components of the data

$\boldsymbol{r}$ :  $[\mathbf{J}_{\boldsymbol{\mu}}(\boldsymbol{r})]_{a,b} = \partial [\widehat{\boldsymbol{x}}_{\boldsymbol{\mu}}(\boldsymbol{r})]_a / \partial [\boldsymbol{r}]_b$ . Since there is no closed-form expression for  $\widehat{\boldsymbol{x}}_{\boldsymbol{\mu}}(\boldsymbol{r})$  (it is obtained by an iterative process), it is complex to derive  $\text{tr}(\mathbf{A} \mathbf{J}_{\boldsymbol{\mu}}(\boldsymbol{r}))$ . An alternative proposed by Girard (1989) considers a Monte Carlo perturbation method to numerically approximate this term by finite-differences:

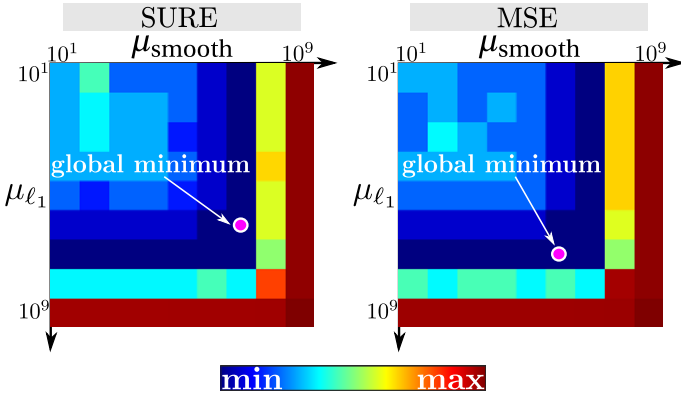
$$\text{tr}(\mathbf{A} \mathbf{J}_{\boldsymbol{\mu}}(\boldsymbol{r})) \approx \xi^{-1} \boldsymbol{b}^T \mathbf{A} [\widehat{\boldsymbol{x}}_{\boldsymbol{\mu}}(\boldsymbol{r} + \xi \boldsymbol{b}) - \widehat{\boldsymbol{x}}_{\boldsymbol{\mu}}(\boldsymbol{r})], \quad (22)$$

where  $\boldsymbol{b} \in \mathbb{R}^{NT}$  is an independent and identically distributed pseudo-random vector of unit variance and  $\xi$  is the amplitude

**Table 2.** Quantitative evaluation of the reconstructions on simulated elliptical and spiral disks: N-RMSE, defined in Eq. (19), for the reconstructions shown in Figs. 5–8.

Score	Algorithm	$\alpha_{\text{gt}} = 1 \times 10^{-6}$	$\alpha_{\text{gt}} = 5 \times 10^{-6}$	$\alpha_{\text{gt}} = 1 \times 10^{-5}$
Elliptical disk, see Figs. 5 and 6				
N – RMSE $(\mathcal{D}(\mathbf{x}_{\text{gt}}), \mathcal{D}(\widehat{\mathbf{x}}))$	REXPACO	0.20	0.18	0.17
N – RMSE $(\mathbf{x}_{\text{gt}}, \widehat{\mathbf{x}})$	REXPACO	0.51	0.32	0.25
N – RMSE $(\mathbf{H}\mathbf{x}_{\text{gt}}, \mathbf{H}\widehat{\mathbf{x}})$	cADI	0.65	0.45	0.43
N – RMSE $(\mathbf{H}\mathbf{x}_{\text{gt}}, \mathbf{H}\widehat{\mathbf{x}})$	REXPACO	<b>0.19</b>	<b>0.13</b>	<b>0.07</b>
Spiral disk, see Figs. 7 and 8				
N – RMSE $(\mathcal{D}(\mathbf{x}_{\text{gt}}), \mathcal{D}(\widehat{\mathbf{x}}))$	REXPACO	0.75	0.66	0.32
N – RMSE $(\mathbf{x}_{\text{gt}}, \widehat{\mathbf{x}})$	REXPACO	0.76	0.81	0.35
s N – RMSE $(\mathbf{H}\mathbf{x}_{\text{gt}}, \mathbf{H}\widehat{\mathbf{x}})$	cADI	0.88	0.87	0.89
N – RMSE $(\mathbf{H}\mathbf{x}_{\text{gt}}, \mathbf{H}\widehat{\mathbf{x}})$	REXPACO	<b>0.75</b>	<b>0.40</b>	<b>0.24</b>

**Notes.** The N-RMSE is also given on the restrictions  $\mathcal{D}(\mathbf{x}_{\text{gt}})$  and  $\mathcal{D}(\widehat{\mathbf{x}})$  to the area actually covered by the disks. The best scores are highlighted in bold fonts.



**Fig. 9.** Comparison between the SURE criterion accounting for the fluctuations of the nuisance component through a local learning of the covariances (*left*, see Eq. (23)) and the true MSE (*right*, see Eq. (20)). The pink circles represent the global minimum of each criterion. Data set: HIP 80019, see Sect. 3.3.1 for observing conditions.

of the perturbation. While the precise setting of  $\xi$  is not a crucial point of the method, it must be chosen both large enough to prevent errors due to numerical underflows in the computation of the difference  $\widehat{\mathbf{x}}_{\mu}(\mathbf{r} + \xi\mathbf{b}) - \widehat{\mathbf{x}}_{\mu}(\mathbf{r})$  and small enough so that the approximation in Eq. (22) remains valid (the effects of the perturbation being nonlinear in the model). In practice, when we evaluate Eq. (22), we set  $\xi = 0.1 \times \text{MAD}(\mathbf{r})$ , where the median absolute deviation  $\text{MAD}(\mathbf{r}) = \text{median}(|\mathbf{r} - \text{median}(\mathbf{r})|)$  is a robust estimator of the standard-deviation.

As in our previous derivation of the statistical model of the nuisance component, we approximate the full covariance matrix  $\mathbf{C}$  that appears in Eq. (21) as block-diagonal, with blocks corresponding to the partition of the image into nonoverlapping patches. The patch covariance model and the perturbation approximation lead to the following risk estimator:

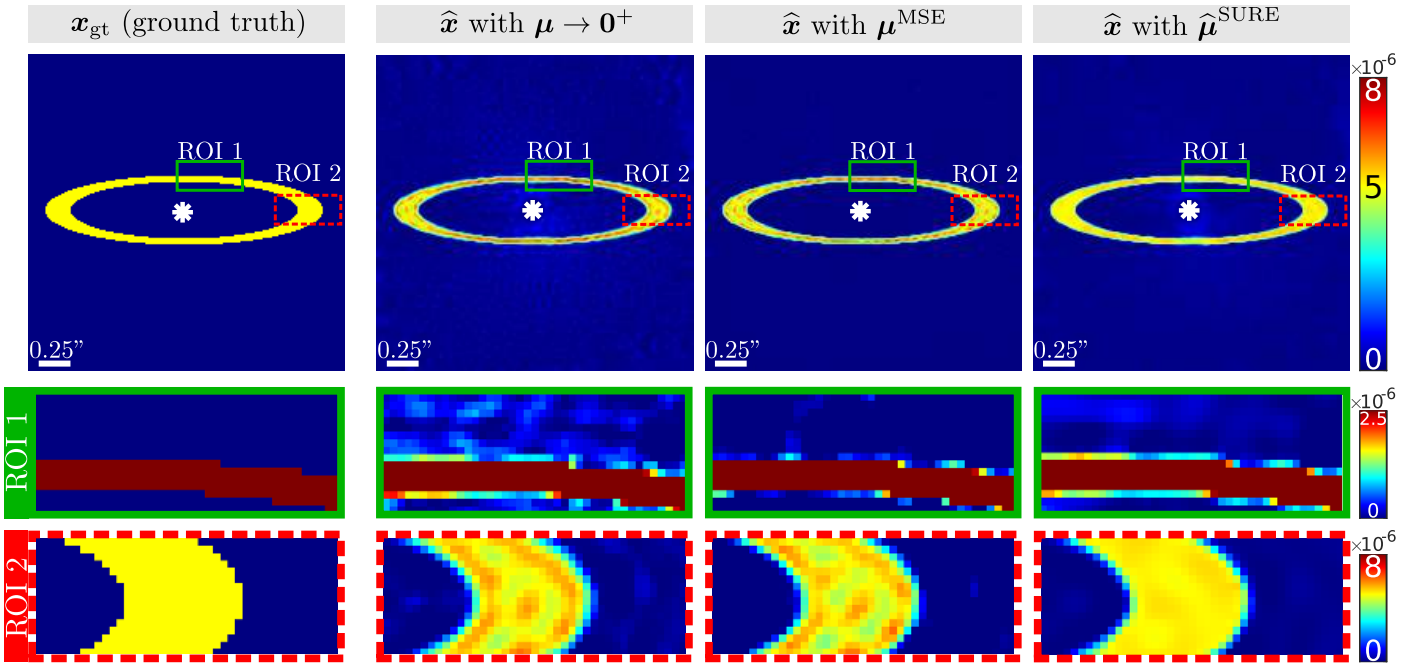
$$\text{SURE}(\boldsymbol{\mu}) \approx \sum_{n \in \mathbb{P}} \sum_{t=1}^T \|\mathbf{P}_n(\mathbf{r}_t - \widehat{\mathbf{m}} - \mathbf{A}\widehat{\mathbf{x}}_{\mu}(\mathbf{r}))\|_{\mathbf{C}_n}^2 + (2/\xi) \mathbf{b}^T \mathbf{A} [\widehat{\mathbf{x}}_{\mu}(\mathbf{r} + \xi\mathbf{b}) - \widehat{\mathbf{x}}_{\mu}(\mathbf{r})] - NT. \quad (23)$$

The optimal value  $\widehat{\boldsymbol{\mu}}^{\text{SURE}}$  of the regularization parameters is obtained by minimizing the quantity (23) with respect to  $\boldsymbol{\mu}$ .

We have validated on a numerical example the automatic tuning of the regularization parameters  $\boldsymbol{\mu}$ : The simulated elliptical disk is injected at the level of contrast  $\alpha_{\text{gt}} = 10^{-5}$  on the data set of HIP 80019. Our SURE risk estimator given in Eq. (23) is evaluated for different values of the parameters  $\boldsymbol{\mu}$ . Figure 9 compares this criterion with the true MSE: The two criteria reach a global minimum for similar hyper-parameter values. The SURE criterion is intrinsically non-convex and additional non-convexities arise due to the sensitivity of the evaluation of the term  $\text{tr}(\mathbf{A}\mathbf{J}_{\mu}(\mathbf{r}))$  by finite differences. Rather than proceeding by local minimization of  $\text{SURE}(\boldsymbol{\mu})$ , it is therefore safer to systematically evaluate the criterion on a 2D grid in order to identify the global minimum.

Figure 10 illustrates the impact of the regularization in three cases: an under-regularized setting (i.e.,  $\boldsymbol{\mu} \rightarrow \mathbf{0}^+$ ), the automatic setting  $\widehat{\boldsymbol{\mu}}^{\text{SURE}}$  based on the minimization of the SURE criterion (23), and the best-possible setting  $\boldsymbol{\mu}^{\text{MSE}}$  computed using the ground truth by minimizing the MSE criterion (20). The morphology of the reconstructed disk is improved by the regularization: It is smoother and sharper thanks to the edge-preserving term. The background of  $\widehat{\mathbf{x}}$  is also less noisy (closer to zero) due to the  $\ell_1$ -norm penalization. It can be noted that in this simulation, the reconstructed object  $\widehat{\mathbf{x}}$  with  $\widehat{\boldsymbol{\mu}}^{\text{SURE}}$  is flatter and sharper than the disk reconstructed with the optimal parameters because the optimal parameters lead to a better reconstruction outside of the disk (reconstructed values are very close to zero: the ground truth value) at the expense of a slightly noisier reconstruction of the disk. In Fig. 9, the location of the global minimum (denoted by a circle) corresponds to parameters  $\widehat{\mu}_{\text{smooth}}^{\text{SURE}} > \mu_{\text{smooth}}^{\text{MSE}}$  and  $\widehat{\mu}_{\ell_1}^{\text{SURE}} < \mu_{\ell_1}^{\text{MSE}}$ , which leads to a stronger smoothing of the object with  $\widehat{\boldsymbol{\mu}}^{\text{SURE}}$  and a better rejection of close-to-zero noise in the background with  $\widehat{\boldsymbol{\mu}}^{\text{MSE}}$ . This qualitative observation is also confirmed by Table 3 which reports RMSE<sup>8</sup> scores between the ground truth  $\mathbf{x}_{\text{gt}}$  and the reconstructed flux distribution  $\widehat{\mathbf{x}}$  ( $\text{RMSE} = \frac{1}{M} \|\mathbf{x}_{\text{gt}} - \widehat{\mathbf{x}}\|_2$ ). Table 3 shows that the quality of the reconstruction obtained with the regularization parameters  $\widehat{\boldsymbol{\mu}}^{\text{SURE}}$  estimated by minimizing the SURE criterion (22) is very

<sup>8</sup> Here, we use this metric instead of N-RMSE defined in Eq. (19) since we aim to compare the absolute error made both inside and outside the spatial support of the disk.



**Fig. 10.** Influence of the regularization on the REXPACO reconstructions.  $x_{\text{gt}}$  (1st column) is the ground truth elliptical disk injected at the level of flux  $\alpha_{\text{gt}} = 1 \times 10^{-5}$ . The flux distribution  $\hat{x}$  is reconstructed with a slight amount of regularization ( $\mu \rightarrow 0^+$ ; 2nd column), with  $\mu^{\text{MSE}}$  derived by minimizing the optimal MSE criterion (3rd column, see Eq. (20)) and its estimate  $\hat{\mu}^{\text{SURE}}$  through the SURE criterion (4th column, see Eq. (23)). *Bottom panel:* zooms on two regions of interest (ROIs) extracted from the reconstructions given in the *top panel*. Data set: HIP 80019, see Sect. 3.3.1 for observing conditions.

close to the best-possible reconstruction obtained by minimizing the MSE criterion (20).

### 3.3. Evaluation on data sets with circumstellar disks

In this section, we evaluate the ability of REXPACO to reconstruct light flux distributions of actual circumstellar disks from ADI sequences. The results are compared to two existing methods, cADI and PCA (see Sect. 1 for their respective principle), both designed for the detection of point-like sources but used extensively with different tuning parameters to process data with disks (see Milli et al. 2012; Pairet et al. 2019 for studies of the resulting post-processing artifacts). We also compare REXPACO to PACO. The rationale behind this last comparison is to assess the benefit of the reconstruction framework of REXPACO in addition to the common statistical modeling of the spatial covariances shared by the two methods. For cADI and the PCA, we used the SpeCal pipeline (Galicher et al. 2018) which is the post-processing standard of the SPHERE data center (Delorme et al. 2017). The PACO implementation is based on the algorithm described in Flasseur et al. (2018a,b,c). A MATLAB<sup>TM</sup> implementation of the REXPACO main routines is available online<sup>9</sup> as a purpose of illustration. It is based on GLOBALBIOIM<sup>10</sup>, an opensource software for image reconstruction through an inverse problem framework (Soubies et al. 2019). Before post-processing with the tested algorithms, the data sets are preprocessed and calibrated with the prerduction tools (Pavlov et al. 2008; Maire et al. 2016) of the SPHERE consortium available at the SPHERE Data Center (Delorme et al. 2017). In particular, background, flat-field, bad pixels, parallac-

**Table 3.** RMSE scores of REXPACO reconstructions of a simulated elliptical disk, for three different regularization levels.

Score ( $\times 10^{-5}$ )	$\mu \rightarrow 0^+$	$\mu^{\text{MSE}}$	$\hat{\mu}^{\text{SURE}}$
RMSE (disk)	4.37	4.41	<b>2.99</b>
RMSE (background)	5.19	<b>2.54</b>	3.96
RMSE (whole field)	6.79	<b>4.96</b>	5.09

**Notes.** The best scores are highlighted in bold fonts.

tic angles, wavelength, true-north and astrometric calibrations are performed during this step.

#### 3.3.1. Data sets description

For the comparisons, we have selected four data sets from the VLT/SPHERE-IRDIS instrument obtained by the observation of stars hosting known circumstellar disks of various morphological structures (e.g., face-on versus edge-on configuration, presence of inner and outer gaps, spirals, etc.). The rationale of such a selection is to test the versatility and the robustness of REXPACO in reconstructing the flux distribution of circumstellar disks with various morphologies. The considered data sets were recorded on the following targets. Table 4 summarizes the main observing conditions of the corresponding ADI sequences.

HR 4796A (HD 109573A) is the primary member of a binary system of the TW Hydrae association with an age about 12 Myr (Bell et al. 2015) located at a distance about 72.8 pc (Van Leeuwen 2007). HR 4796A hosts a debris disk in face-on configuration first observed by the *Hubble* Space Telescope (Schneider et al. 1999). Since then, its morphology has been studied intensively with ground-based facilities (Milli et al. 2017, 2019). The relatively bright disk (contrast between  $10^{-5}$  and  $10^{-4}$ ) exhibits a

<sup>9</sup> [https://github.com/olivier-flasseur/rexpaco\\_demo.git](https://github.com/olivier-flasseur/rexpaco_demo.git)

<sup>10</sup> <https://biomedical-imaging-group.github.io/GlobalBioIm/>

**Table 4.** Observing conditions of ADI sequences from the VLT/SPHERE-IRDIS instrument considered in this paper.

Target	ESO ID	Obs. date	$T$	$\lambda$ ( $\mu\text{m}$ )	$\Delta_{\text{par}}$ ( $^\circ$ )	NDIT	DIT (s)	$\tau_0$ (ms)	Seeing ( $''$ )	Related paper
HR 4796A	095.C-0298	2015-02-03	110	1.59	48.5	8	32	11.8	0.67	Milli et al. (2017)
RY Lupi	097.C-0865	2016-04-16	80	1.59	71.0	4	64	61.4	0.36	Langlois et al. (2018)
SAO 206462	095.C-0298	2015-05-15	63	2.11	63.6	4	64	85.0	0.50	Maire et al. (2017)
PDS 70	100.C-0481	2018-02-24	90	2.11	95.7	3	96	71.5	0.46	Müller et al. (2018)
HIP 80019 <sup>(*)</sup>	095.C-0389	2015-04-21	116	2.11	42.8	11	8	23.2	0.80	–

**Notes.** Columns are: target name, ESO survey ID, observation date, number  $T$  of available temporal frames, spectral filter  $\lambda$ , total amount of field rotation  $\Delta_{\text{par}}$  of the field of view, number NDIT of sub-integration exposures, individual exposure time DIT, DIMM coherence time  $\tau_0$ , seeing value at the beginning of the observations and the first paper reporting analysis of the same data. All the observations are performed with the apodized Lyot coronagraph (APLC; Carillet et al. 2011) of the VLT/SPHERE instrument. <sup>(\*)</sup>The ADI sequence of HIP 80019 was made of 176 temporal frames but we have discarded the 60 first frames manually. This sequence contains no detectable off-axis sources based on the application of PACO and REXPACO algorithms. If off-axis sources are present at an undetectable level of contrast, they would introduce a negligible bias in comparison to the levels of contrast considered in our experiments. This ADI sequence is used in Sect. 2 to perform injections of fake disks.

thin ring structure with a semimajor axis about 77 au, an inclination by about  $76^\circ$ , and a high surface brightness which suggest the presence of exoplanets although no companions have been detected to date.

RY Lupi (HIP 78094) is a T Tauri star of the Scorpius-Centaurus association with an age about 10–20 Myr (Pecaut et al. 2012) located at a distance about 151 pc (Brown et al. 2016). It hosts a transition disk directly observed for the first time in scattered light with the SPHERE instrument (Langlois et al. 2018). The disk is close to an edge-on configuration with an inclination about  $70^\circ$ , and it exhibits asymmetries and warped features near the star due to its interaction with the star magnetosphere. Observations in ADI and PDI have shown that the structure of the disk could be interpreted as spiral arms possibly due to the presence of putative low-mass exoplanets orbiting exterior to the spiral arms (Langlois et al. 2018).

SAO 206462 (HD 135344B) is the secondary member of a binary system of the Upper Centaurus Lupus constellation with an age about 9 Myr (Müller et al. 2011) located at a distance about 157 pc (Brown et al. 2016). SAO 206462 hosts a transition disk near a face-on configuration with an inclination about  $11^\circ$ , first resolved both in thermal emission (Doucet et al. 2006) and in scattered light (Grady et al. 2009). The disk exhibits two spiral arms, asymmetric features and an inner cavity. Its structure is of high interest since it could be due to the presence of low-mass exoplanets inside the arms or the cavity as shown by recent analysis of high contrast and high resolution observations (Maire et al. 2017).

PDS 70 (V1032 Centauri) is a T Tauri star of the Scorpius-Centaurus association with an age about 5 Myr (Müller et al. 2018) located at a distance about 113 pc (Brown et al. 2016). It hosts a protoplanetary disk and two confirmed exoplanets (PDS 70b and PDS 70c) in formation inside the disk. The protoplanetary disk hosted by PDS 70 was first resolved with the Naos Conica instrument (NaCo; Lenzen et al. 2003) at the VLT (Riaud et al. 2006). It exhibits complex structures in the form of several arcs, outer and inner gaps, and potential spiral arms at the north side of the outer disk (Keppler et al. 2018; Mesa et al. 2019). The exoplanet PDS 70b has been detected by direct imaging with the VLT/SPHERE instrument (Keppler et al. 2018) in near infrared, and the exoplanet PDS 70c has been unveiled with the VLT/MUSE instrument in  $H_\alpha$  (Haffert et al. 2019). To date, this is the unique system embedding multiple nascent exoplanets. The analysis of the structure of this system is extremely valuable to understand exoplanet formation mechanisms and raises still-open questions like the presence of additional putative exoplanets (Mesa et al. 2019).

### 3.3.2. Reconstruction of circumstellar disks

We apply REXPACO to the four selected ADI sequences presented in Table 4. For each, the computation takes a few hours (MATLAB<sup>TM</sup> implementation, processor Intel i9-9880H at 2.3 GHz). Figure 11 displays the results obtained with PACO, PCA and cADI. For the PCA, we have tested several choices for the numbers of modes used in the decomposition and selected the best choice, corresponding to a small number of modes in order to limit the issue of self-subtraction (see Pairet et al. 2019 for a discussion about this effect). The images produced by PACO are obtained with a maximum likelihood estimator of the light flux derived for point-like sources. For REXPACO, the results take the form of reconstructed flux distributions  $\hat{\mathbf{x}}$ . We also give the blurred versions  $\mathbf{H}\hat{\mathbf{x}}$  to simplify the comparison with methods that do not perform a deconvolution. For the PCA and cADI, the results take the form of residual images obtained after subtraction of the estimated on-axis PSF. These residual images are normalized to the star flux that is encoded in the on-axis PSF provided with the data. Thus, the resulting images can also be interpreted as blurred flux distributions of the off-axis sources.

Figure 11 shows that the mean levels of flux estimated by PCA and cADI differ significantly from the flux estimated by REXPACO and PACO<sup>11</sup>: The PCA and cADI are sensitive to several causes of artifacts severely altering both the morphology and the photometry of the disks. In particular, cADI and the PCA suffer from positive and negative replicas and also from the self-subtraction phenomenon. cADI performs much better in the presence of extended features like disks due to its limited attenuation of the disk signal. However, cADI is also sensitive to the ADI artifacts creating morphological distortions like nonphysical discontinuities in the structure of the disks, as discussed in Sect. 1. This is especially the case for the thin and bright ring surrounding HR 4796A. As discussed in Sect. 2.4, the REXPACO reconstruction exhibits an almost continuous elliptical structure, a flux asymmetry on the west side of the ring as predicted by intensity scattering models (Milli et al. 2017), and the scattering structures predicted by radiative transfer models near the disk extremities (Lagrange et al. 2012). The morphological structure of the reconstructed disk is very close the one obtained from polarimetric observations that are almost exempt from artifacts at such separations (Milli et al. 2019). The faithfulness of the REXPACO reconstruction is also supported by the

<sup>11</sup> Fluxes estimated by REXPACO and PACO are approximately comparable. The residual flux difference is due to the fact that PACO is based on a point-like source assumption which can lead to bias in the estimated flux of extended features.

simulations presented in Sect. 2.4 (see Figs. 5 and 6) where we injected a simulated elliptical disk with comparable eccentricity and spatial extent than the HR 4796A disk. In this latter case, no significant artifacts were identified in the reconstruction even though the simulated disk was one order of magnitude fainter than the HR 4796A disk. The REXPACO reconstruction of RY Lupi also exhibits reduced ADI artifacts. As seen on Fig. 11, the REXPACO reconstructions show better the smooth and faint extended disk structures. In addition, the reconstructed flux distribution uncovers new details around the brightest part of the disk such as spirals as well as the bottom and top disk planes.

The comparison with PACO is also interesting. While this latter algorithm demonstrates significantly better performance than the PCA to detect point-like sources (Flasseur et al. 2018a), it performs worse than the PCA and cADI in the presence of extended features since it assumes that the pattern to detect is known and takes the form of a point-like source. Besides, parts of the disks are lost in the PACO reconstructions due to the subtraction of the mean component of the ADI sequence<sup>12</sup>. This effect is especially problematic for SAO 206462 since the nuisance component is hidden by the disk in almost all temporal frames due to the quasi-circular symmetry of the disk. The combination of the reconstruction framework of REXPACO with its iterative strategy reducing bias of the statistics of the nuisance component leads to a physically more exploitable image of the flux distribution. The quality of the reconstruction is supported by the simulations performed in Sect. 2.4 (see Figs. 7 and 8) where a numerical model of a spiral disk with two arms of comparable spatial extent is injected to the data. These numerical tests have shown that REXPACO is able to reconstruct a faithful image of the flux distribution of such a type of disk for levels of contrast  $\gtrsim 5 \times 10^{-6}$ , that is to say about five to ten times fainter than the mean reconstructed level of flux of the SAO 206462 disk. The bright asymmetry in the southwest side of the disk could be interpreted as a reconstruction artifact possibly due to a local mis-modeling of the nuisance component, but we did not experience such an effect in our simulations for the typical level of flux of SAO 206462. Besides, the REXPACO reconstruction exhibits detailed structures of the spiral arms with a better rejection of the nuisance component near the star than the ADI and RDI post-processings performed in Maire et al. (2017) on the same sequence of observations.

Finally, the case of PDS 70 is particularly interesting due to the complex structure of the disk and to the presence of two confirmed exoplanets at formation stage inside the disk. Concerning the disk structure, the outer ring reconstructed by REXPACO is free from the self-subtraction artifacts that were visible on the cADI, PCA and PACO reductions. The bright asymmetry of the disk in the west side of the outer disk and the possible spiral-shaped features pointed by Müller et al. (2018) in the north side of the disk are also detectable in the REXPACO reconstruction. Concerning the point-like sources, the exoplanet PDS 70b exhibits a bright flux at  $\lambda = 2.11 \mu\text{m}$  (K1 spectral band) as previously reported by Müller et al. (2018), Mesa et al. (2019). Re-detecting the exoplanet PDS 70c is more difficult since it is embedded inside the disk material as reported by Haffert et al. (2019), Mesa et al. (2019). We propose in Sect. 4 a

processing strategy to disentangle the contribution of point-like sources from the spatially extended contribution of the circumstellar disk. Besides, a putative point-like feature (PLF) has been recently identified by Mesa et al. (2019) based on the analysis of several near-infrared angular plus spectral differential imaging (ASDI) sequences by various post-processing methods. This PLF can be (marginally) detected from the REXPACO reconstruction while it is indistinguishable in the same ADI sequence processed by PCA (see Fig. 1 of Mesa et al. 2019), as it is also the case for our PCA reduction given in Fig. 11.

Figure 12 shows the importance of the modeling of spatial covariances in REXPACO: Regularized inversions with a statistical model of the nuisance component that neglects all covariances (but still accounts for the nonstationary variances) are given in Fig. 12a, to compare with REXPACO reconstruction shown in Fig. 12b. When spatial covariances are omitted, the reconstructions are worse. This is particularly visible on HR 4796A, in the background and close to the star, and on SAO 206462, close to the star.

Figure 13 completes this study by showing that the method for automatically setting the regularization hyper-parameters discussed in Sect. 3.2 leads to satisfying reconstructions: The HR 4796A disk is smoother, with sharper edges, and the background is less noisy when automatically selected hyper-parameters  $\hat{\mu}^{\text{SURE}}$  are used compared to the case of under-regularization  $\mu \rightarrow 0^+$ . Conversely, over-regularization ( $\mu = 20 \times \hat{\mu}^{\text{SURE}}$ ) leads to strong artifacts mimicking the self-subtraction effects, especially near the star, due to the large penalization of the  $\ell_1$ -norm of the reconstructed flux distribution. The extremities of the disk ring also exhibit staircase-artifacts. These effects are signs of an over-regularization by the edge-preserving term.

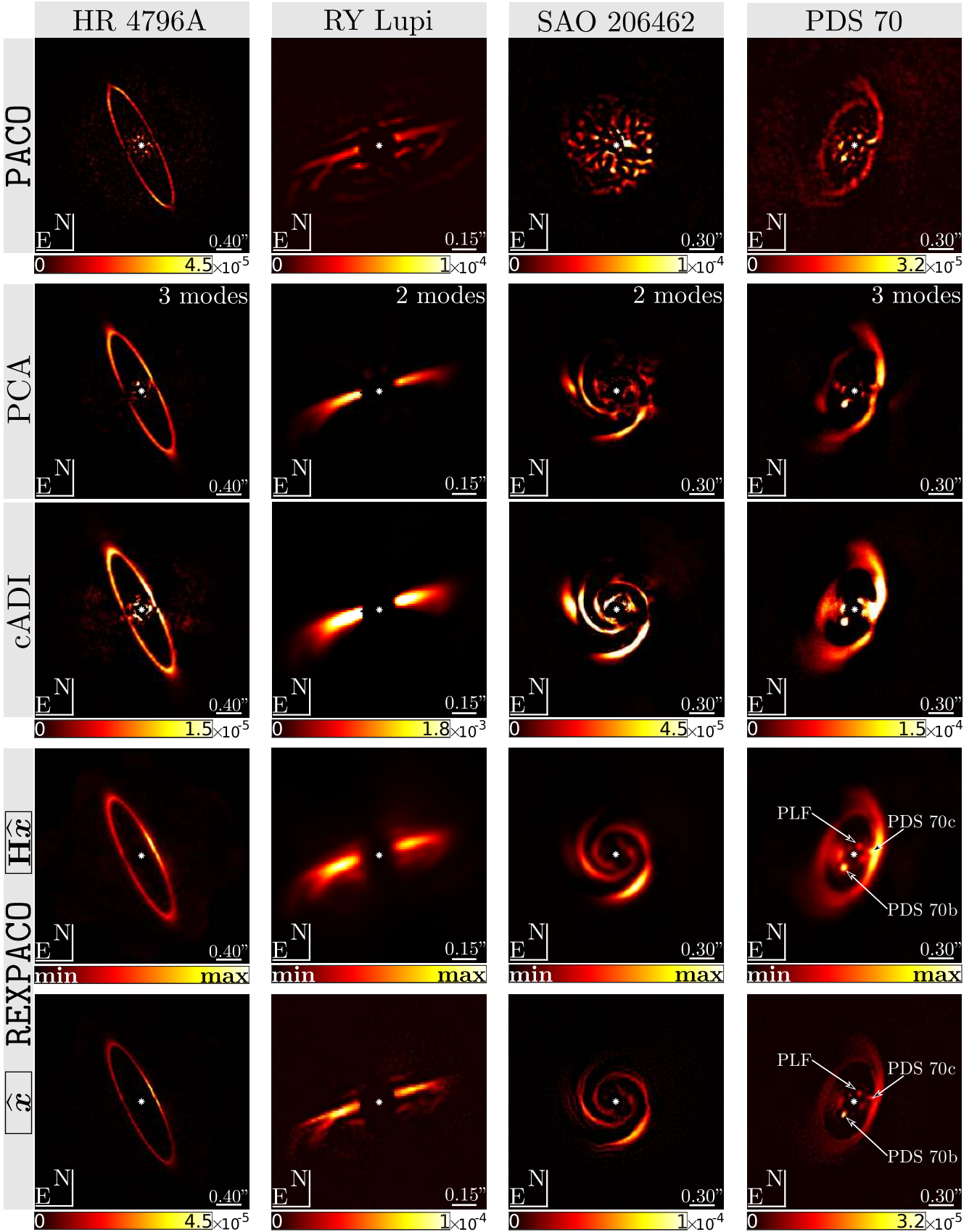
### 3.3.3. Discussion of the results

Based on numerical injections of fake disks (with different orientations and typical morphologies), we have shown that REXPACO is capable of reconstructing a faithful image of the flux distribution at contrasts up to  $10^6$  for elliptical disks with a large eccentricity (about 0.95). For disks in face-on configuration, our simulations have shown that the contrast achieved by REXPACO drops by a factor between two and ten due to the increased difficulty in disentangling the disk signal from the nuisance component in such a configuration. While these first quantitative results should be complemented and refined by more extensive simulations with diverse disks and observing conditions, they give first clues about the operating range of REXPACO. In addition, REXPACO better preserves both the morphology and the photometry of the disks than the standard cADI method used routinely to process ADI sequences in the presence of circumstellar disks.

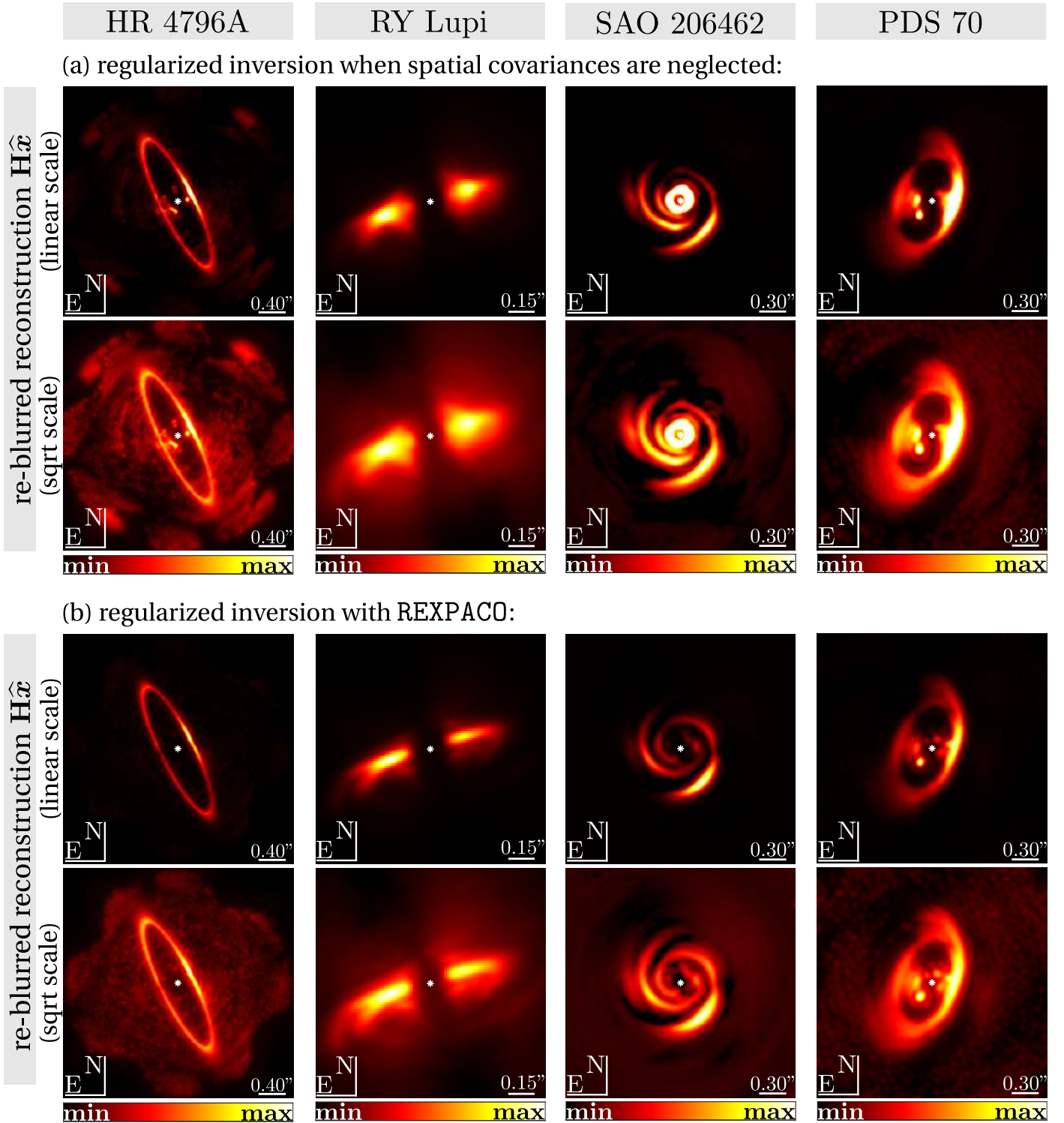
We have also applied REXPACO on four ADI sequences recorded by the VLT/SPHERE-IRDIS instrument on stars hosting known circumstellar disks with various morphologies. While there is no ground truth for these objects, we have shown that REXPACO copes with the classical ADI artifacts (e.g., nonphysical distortions and large residual stellar leakages). The images produced by REXPACO allow to recover structures in the circumstellar material that were previously identified only with methods complementing ADI such as reference or polarization differential imaging. We expect that the high contrast capability of REXPACO combined with its ability to reconstruct faithful images of the circumstellar material will be helpful in detecting new circumstellar disks and in refining the astrophysical interpretation of their flux distributions.

<sup>12</sup> PACO implements a dedicated strategy to cope with this issue for point-like sources: The candidate companions above a given threshold are characterized in terms of astrometry and photometry after their detection. This second step prevents a statistical bias in the estimated quantities.





**Fig. 11.** Images of the flux distribution reconstructed by REXPACO comparatively to the PACO, PCA and cADI algorithms. For REXPACO, both the deblurred reconstruction  $\hat{x}$  and its reblurred version  $H\hat{x}$  are given for comparison with the other considered methods that do not produce deblurred images of the flux distribution. Data sets: HR 4796A, RY Lupi, SAO 206462 and PDS 70, see Sect. 3.3.1 for observing conditions.

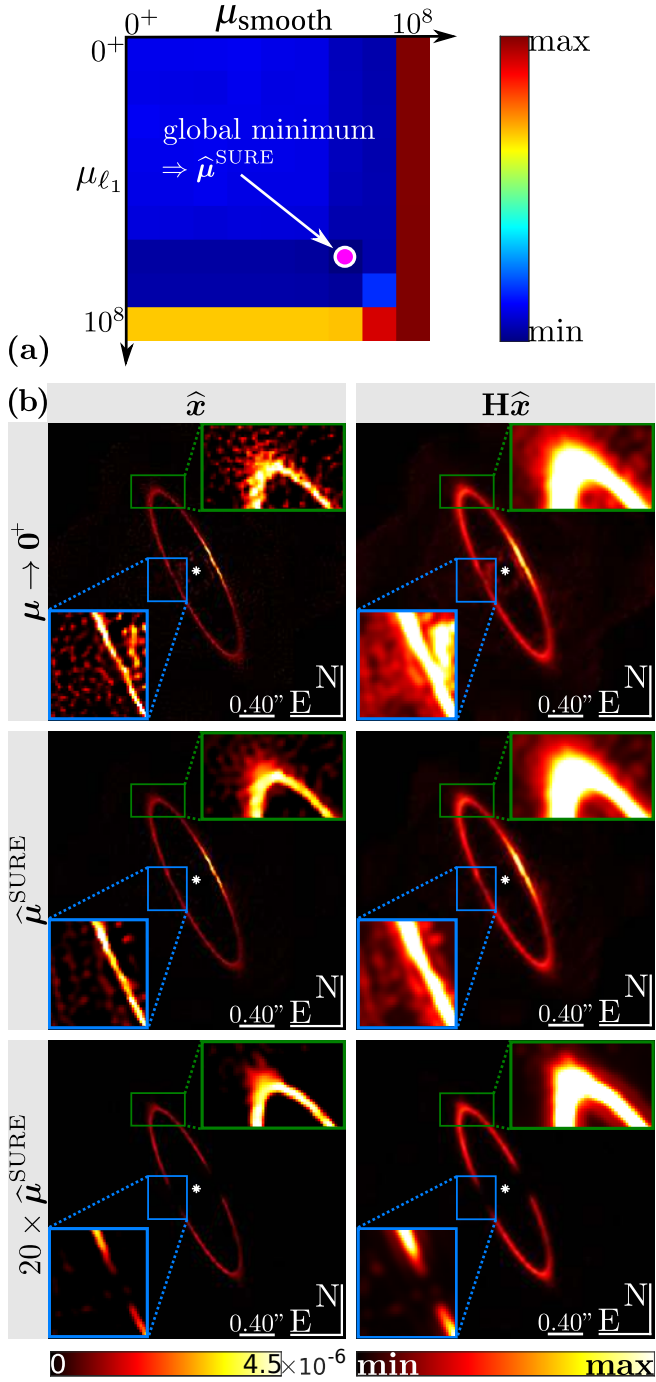


**Fig. 12.** Importance of the modeling of spatial covariances. Comparison between (a) regularized inversions with a model of the nuisance component that accounts for the nonstationary variances but neglect covariances, and (b) REXPACO inversions. Reconstructions are given first with a linear scale, then with a square-root scaling in order to better distinguish the lowest reconstructed values. Data sets: HR 4796A, RY Lupi, SAO 206462 and PDS 70, see Sect. 3.3.1 for observing conditions.

#### 4. Unmixing point-like sources and extended sources

For systems formed of both an extended disk and point-like sources such as exoplanets, the image reconstruction can be improved by combining REXPACO algorithm with the exoplanet

detection and characterization method PACO. Both algorithms share a common statistical modeling of the background and can thus be combined in a principled way. In our experiments, this approach proved superior to our attempts to reconstruct with REXPACO a composite object  $\mathbf{x} = \mathbf{x}_{\text{ext}} + \mathbf{x}_{\text{pnt}}$  where  $\mathbf{x}_{\text{ext}}$  accounts for extended (smooth) structures while  $\mathbf{x}_{\text{pnt}}$  accounts



**Fig. 13.** Effect of the setting of the regularization parameters on REXPACO reconstructions. (a) SURE criterion derived from Eq. (23); (b) comparative reconstructions (first row:  $\boldsymbol{\mu} \rightarrow \mathbf{0}$ , i.e., under-regularization; second row:  $\hat{\boldsymbol{\mu}}^{\text{SURE}} = \{10^6, 10^6\}$ , i.e., estimated optimal regularization; third row:  $20 \times \hat{\boldsymbol{\mu}}^{\text{SURE}} = \{2 \times 10^7, 2 \times 10^7\}$ , i.e., over-regularization). Two regions of interest are also displayed with a saturated colormap near the north side extremity of the disk and near the star. Data set: HR 4796A, see Sect. 3.3.1 for observing conditions.

for point-like sources and with a regularization designed to favor these specific structures in each component:

$$\mathcal{R}(\mathbf{x}, \boldsymbol{\mu}) = \mu_{\text{smooth}} \mathcal{R}_{\text{smooth}}(\mathbf{x}_{\text{ext}}, \epsilon) + \mu_{\ell_1} \mathcal{R}_{\ell_1}(\mathbf{x}_{\text{pnt}}).$$

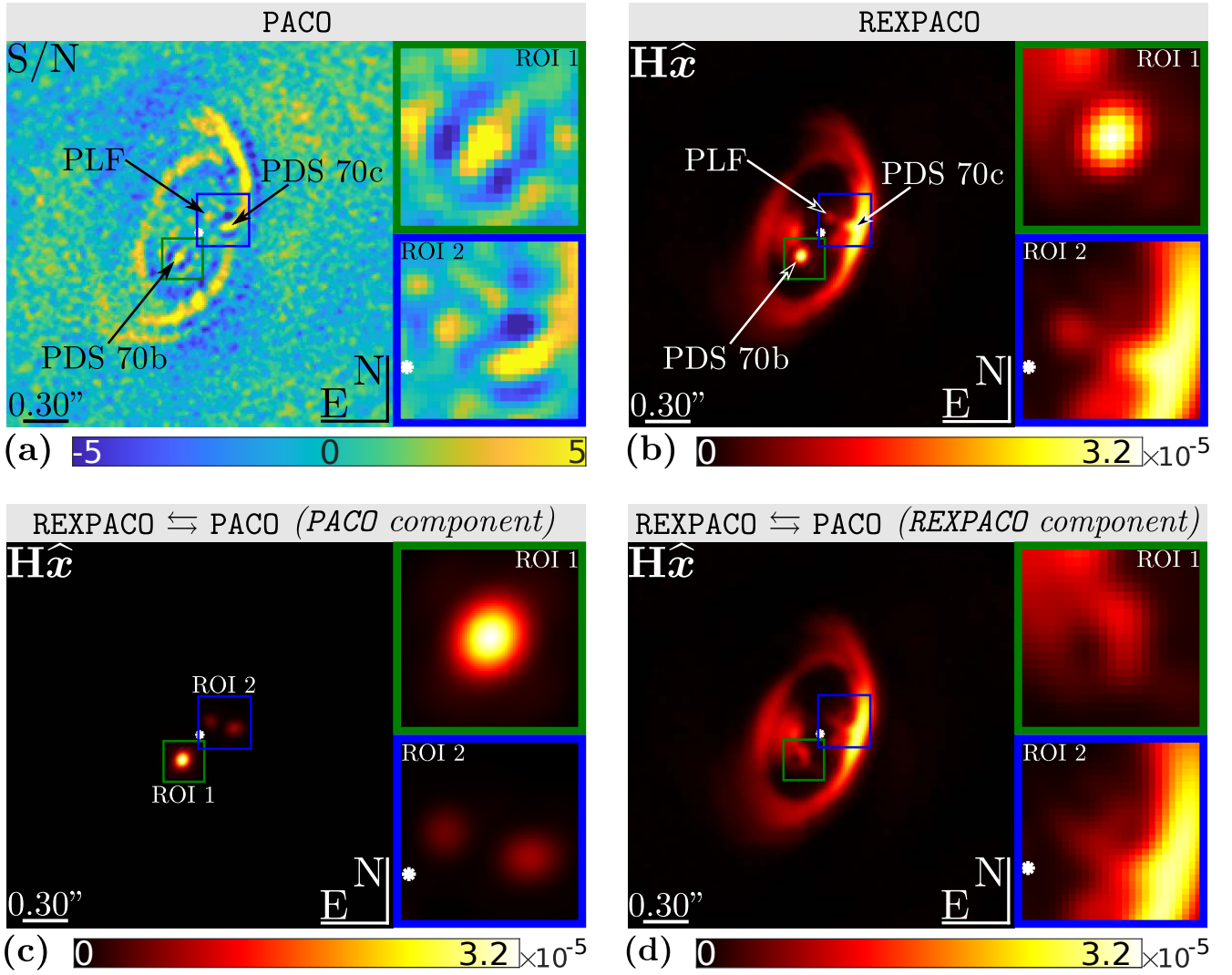
The two main arguments in favor of combining REXPACO and PACO are the ability of PACO to account for the sub-pixel location of point-like sources (while a sparse component  $\mathbf{x}_{\text{pnt}}$  is restricted

to the discretization of the pixel grid) and direct control on the number of point-like sources (while  $\mathbf{x}_{\text{pnt}}$  either displayed many more nonzero entries than the actual number of point sources, or missed the weaker point-like sources).

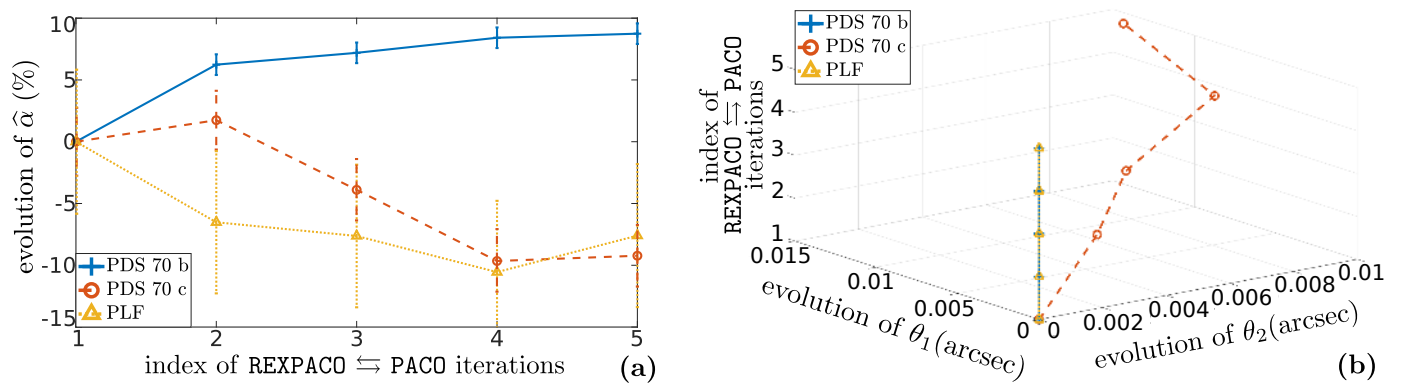
In a nutshell, the proposed combination of the two algorithms works as follows: (i) PACO and REXPACO are launched independently on a given ADI sequence; (ii) based on the S/N map produced by PACO and on the flux distribution produced by REXPACO, the user identifies manually putative point-like sources; (iii) REXPACO and PACO are then alternatively applied. During step (iii), the astrometry and photometry of the selected point-like sources are refined by PACO, based on the residual data obtained after subtraction of the disk contribution as currently reconstructed by REXPACO. Next, the flux distribution of the disk can be refined by REXPACO on updated residuals obtained after subtraction of the refined point-sources contribution estimated by PACO. Step (iii) is repeated until convergence of the disk and point-like sources components. In practice, we stop the alternating procedure of step (iii) when the relative evolution of both the disk and the point-like sources components is smaller than  $10^{-8}$  between two consecutive iterations.

This strategy can be particularly valuable for the processing of ADI sequences of stars that host point-like sources near a circumstellar disk. In this paper, we illustrate our combined approach on the PDS 70 system hosting two confirmed exoplanets (PDS 70b and PDS 70c) inside a protoplanetary disk (Keppler et al. 2018; Haffert et al. 2019), see Sect. 3.3.2 for a discussion about the REXPACO reconstruction. Based on the PACO S/N map and REXPACO flux distribution provided by step (i) of the combined method, we manually selected the rough location of the exoplanets PDS 70b and PDS 70c, see Figs. 14a–b. We also considered the point-like feature (PLF) first identified by Mesa et al. (2019)<sup>13</sup>, see Figs. 14a–b. Figures 14c–d show the contribution of the three selected point-like features (PDS 70b, PDS 70c and the PLF) estimated by PACO and the disk contribution reconstructed by REXPACO after convergence of the proposed alternating scheme. Figures 14c–d show the good separation of point-like sources and of the extended disk structures surrounding them, thus illustrating the efficiency of this approach to disentangle the sparse contribution of the point-like sources from the spatially extended contribution of the disk. It is worth noticing that near the location of PDS 70b, the spatially extended component retrieved by the REXPACO/PACO combination exhibits an extended faint circular structure that could be a hint of a circumplanetary disk as proposed by Christiaens et al. (2019) and Isella et al. (2019). It is not yet possible to conclude on the reality of this structure recovered by REXPACO/PACO, which could also be the result of reconstruction artifacts (e.g., due to a partial unmixing of the sparse and spatially extended components). This open question would require further analysis, for instance based on numerical simulations by means of injection of fake exoplanets and of a physics-based model of the PDS 70 disk to study the influence of the exoplanet-disk interactions on the reconstruction results. Such a study is left for future work. Concerning PDS 70c, the extended feature detected in the disk component near this exoplanet is in good agreement with the submillimeter continuum emission detection (Isella et al. 2019) which confirms the hypothesis that this exoplanet is still in its accretion phase.

<sup>13</sup> Based on estimated photometry, the study of Mesa et al. (2019) concludes that this PLF is most likely a part of the disk. In our present work, the goal is not to discuss the status of the PLF but rather to illustrate the principle of the proposed algorithm.



**Fig. 14.** Combining REXPACO with PACO on a data set of PDS 70 at  $\lambda = 2.110 \mu\text{m}$ : (a) S/N map from PACO; (b) flux distribution from REXPACO; (c) contribution of the three selected point-like sources estimated with PACO; (d) contribution of the disk reconstructed with REXPACO.



**Fig. 15.** Combining REXPACO with PACO on a data set of PDS 70 at  $\lambda = 2.110 \mu\text{m}$ : (a) evolution of the estimated photometry; (b) evolution of the estimated astrometry, for the three selected point-like sources.

Figure 15 shows how the photometry and astrometry of the three point sources evolve throughout the alternations between REXPACO and PACO. Accounting for the presence of the disk modifies the estimated photometry of the three selected point-like sources by about 10%. The astrometry of PDS 70b and of the PLF estimated with PACO are not significantly modified by

the proposed alternating scheme: These two point-like sources are not located inside a bright arc of the disk so that the disk material introduces only a small bias on PACO estimations. On the other hand, the estimated astrometry of the exoplanet PDS 70c located inside a very bright arm of the disk is modified by about 0.010 arcsec (i.e., about 1 pixel in the two spatial

directions, PACO astrometry is evaluated on a subpixel grid with four nodes per pixel). The astrophysical interpretation of these results would require a dedicated study on multiple epochs. If the astro-photometric estimations we derived are confirmed, they could be significant corrective factors of the spectral energy distribution and of the orbit of the exoplanet PDS 70c.

## 5. Conclusion

In this paper, we have introduced a new post-processing algorithm, named REXPACO, dedicated to the reconstruction of flux distributions from ADI data sets in the presence of extended features, like circumstellar disks. To the best of our knowledge, REXPACO is one of the first methods specifically designed for this task. It encompasses the statistical modeling of the fluctuations of the nuisance component (i.e., stellar leakages and additional sources of noise) at the scale of small patches of a few tens of pixels. The resulting statistics of the nuisance term are accounted for in a regularized reconstruction framework following an inverse problem approach based on a forward image formation model of the off-axis sources in the ADI sequences. We have shown that suitable spatial regularizations ( $\ell_1$  norm and  $\ell_2 - \ell_1$  edge-preserving) reduce residual reconstruction artifacts. Regularization parameters are estimated by minimization of SURE. Our numerical experiments have shown that this strategy leads to a quasi-optimal setting of these parameters. The images of the circumstellar environment produced by our algorithm are deblurred from the instrumental PSF which is accounted for in the forward model of the image formation. These different ingredients make our approach very versatile and capable of recovering both sharp edges and smooth transitions of the circumstellar material. The REXPACO algorithm is fully unsupervised: The patch size, the statistics of the nuisance component and the weights given to the regularizations are estimated in a data-driven fashion. Besides, we have illustrated the benefits of the combination of REXPACO with PACO, two algorithms dedicated to complementary tasks, to analyze ADI sequences in the presence of both circumstellar disks and candidate point-like sources. The combination of the two algorithms is able to disentangle the spatially extended contribution of circumstellar disks from the spatially localized contribution of point-like sources.

While the statistical model of REXPACO is able to capture most of the fluctuations of the nuisance component, slight stellar residuals are still present in the reconstructions. The fidelity of the underlying statistical model to the empirical distribution of the nuisance component is a key point to extract accurately the signal from off-axis sources. Beyond accurate modeling of the nuisance term, the angular motion through the ADI sequence is also a key point to unmix the light coming from off-axis objects such as a circumstellar disk and point-like sources from the stellar light that leaks from the coronagraph. For instance, objects with a circular symmetry cannot be correctly recovered since the temporal diversity brought by ADI is not sufficient in this case. Only additional information such as the spectral measurements obtained with angular plus spectral differential imaging (ASDI) in addition to the temporal diversity of ADI would make the reconstruction possible in those cases. The refinement and adaptation of REXPACO to the two above-mentioned points require specific refinements that will be developed in a future paper.

*Acknowledgements.* We thank A. Boccaletti (LESIA, Paris, France) who provided the transmission of the SPHERE coronagraphs. We also thank the anonymous referee for her/his careful reading of the manuscript as well as her/his insightful comments and suggestions. This work has made use of the SPHERE

Data Center, jointly operated by OSUG/IPAG (Grenoble, France), PYTHEAS/LAM/CESAM (Marseille, France), OCA/Lagrange (Nice, France), Observatoire de Paris/LESIA (Paris, France), and Observatoire de Lyon/CRAL (Lyon, France). This work has been supported by the Région Auvergne-Rhône-Alpes under the project DIAGHOLO, by the French National Programs (PNP and PNPS), and by the Action Spécifique Haute Résolution Angulaire (ASHRA) of CNRS/INSU co-funded by CNES.

## References

- Bell, C. P., Mamajek, E. E., & Naylor, T. 2015, *MNRAS*, 454, 593  
 Benisty, M., Juhasz, A., Boccaletti, A., et al. 2015, *A&A*, 578, L6  
 Beuzit, J.-L., Vigan, A., Mouillet, D., et al. 2019, *A&A*, 631, A155  
 Brown, A. G., Vallenari, A., Prusti, T., et al. 2016, *A&A*, 595, A2  
 Carbillet, M., Bendjoya, P., Abe, L., et al. 2011, *Exp. Astron.*, 30, 39  
 Charbonnier, P., Blanc-Féraud, L., Aubert, G., & Barlaud, M. 1997, *IEEE Trans. Image Proc.*, 6, 298  
 Chen, Y., Wiesel, A., Eldar, Y. C., & Hero, A. O. 2010, *IEEE Trans. Signal Proc.*, 58, 5016  
 Christiaens, V., Cantalloube, F., Casassus, S., et al. 2019, *ApJ*, 877, L33  
 Craven, P., & Wahba, G. 1978, *Numer. Math.*, 31, 377  
 Dawson, R. I., Murray-Clay, R. A., & Fabrycky, D. C. 2011, *ApJ*, 743, L17  
 Delorme, P., Meunier, N., Albert, D., et al. 2017, in *SF2A-2017: Proceedings of the Annual meeting of the French Society of Astronomy and Astrophysics*, eds. C. Reylé, P. Di Matteo, F. Herpin, et al., 347  
 Dohlen, K., Langlois, M., Saisse, M., et al. 2008a, in *SPIE Astronomical Telescopes + Instrumentation, International Society for Optics and Photonics*, 70143L  
 Dohlen, K., Saisse, M., Origine, A., et al. 2008b, *SPIE Astron. Telescopes Instrum.*, 7018  
 Doucet, C., Pantin, E., Lagage, P., & Dullemond, C. 2006, *A&A*, 460, 117  
 Esposito, T. M., Fitzgerald, M. P., Graham, J. R., & Kalas, P. 2013, *ApJ*, 780, 25  
 Esposito, T. M., Kalas, P., Fitzgerald, M. P., et al. 2020, *AJ*, 160, 24  
 Flasseur, O., Denis, L., Thiébaud, É., & Langlois, M. 2018a, *A&A*, 618, A138  
 Flasseur, O., Denis, L., Thiébaud, É., & Langlois, M. 2018b, in *IEEE International Conference on Image Processing*, 2735  
 Flasseur, O., Denis, L., Thiébaud, É. M., & Langlois, M. 2018c, in *SPIE Astronomical Telescopes + Instrumentation, Int. Soc. Opt. Photonics*, 10703, 107032R  
 Flasseur, O., Denis, L., Thiébaud, É., Olivier, T., & Fournier, C. 2019, in *European Signal Processing Conference*, 1  
 Flasseur, O., Denis, L., Thiébaud, É., & Langlois, M. 2020a, *A&A*, 637, A9  
 Flasseur, O., Denis, L., Thiébaud, É., & Langlois, M. 2020b, *A&A*, 634, A2  
 Galicher, R., Boccaletti, A., Mesa, D., et al. 2018, *A&A*, 615, A92  
 Garufi, A., Avenhaus, H., Pérez, S., et al. 2020, *A&A*, 633, A82  
 Gerard, B. L., & Marois, C. 2016, in *SPIE Astronomical Instrumentation + Telescopes, Int. Soc. Opt. Photonics*, 9909, 1544  
 Girard, D. A. 1989, *Numer. Math.*, 56, 1  
 Grady, C., Schneider, G., Sitko, M., et al. 2009, *ApJ*, 699, 1822  
 Haffert, S., Bohn, A., de Boer, J., et al. 2019, *Nat. Astron.*, 3, 749  
 Isella, A., Benisty, M., Teague, R., et al. 2019, *ApJ*, 879, L25  
 Keppler, M., Benisty, M., Müller, A., et al. 2018, *A&A*, 617, A44  
 Kluska, J., Berger, J. P., Malbet, F., et al. 2020, *A&A*, 636, A116  
 Lagrange, A.-M., Milli, J., Boccaletti, A., et al. 2012, *A&A*, 546, A38  
 Langlois, M., Pohl, A., Lagrange, A.-M., et al. 2018, *A&A*, 614, A88  
 Langlois, M., Gratton, R., Lagrange, A. M., et al. 2020, *A&A*, [arXiv:2103.03976]  
 Ledoit, O., & Wolf, M. 2004, *J. Multivariate Anal.*, 88, 365  
 Lenzen, R., Hartung, M., Brandner, W., et al. 2003, in *SPIE Astronomical Telescopes + Instrumentation, Int. Soc. Opt. Photonics*, 4841, 944  
 Maire, A. L., Langlois, M., Dohlen, K., et al. 2016, in *Ground-based and Airborne Instrumentation for Astronomy VI, Int. Soc. Opt. Photonics*, 9908, 990834  
 Maire, A.-L., Stolker, T., Messina, S., et al. 2017, *A&A*, 601, A134  
 Marois, C., Lafrenière, D., Doyon, R., Macintosh, B., & Nadeau, D. 2006, *ApJ*, 641, 556  
 Mazoyer, J., Arriaga, P., Hom, J., et al. 2020, in *Ground-based and Airborne Instrumentation for Astronomy VIII, Int. Soc. Opt. Photonics*, 11447, 1144759  
 Mesa, D., Keppler, M., Cantalloube, F., et al. 2019, *A&A*, 632, A25  
 Milli, J., Mouillet, D., Lagrange, A.-M., et al. 2012, *A&A*, 545, A111  
 Milli, J., Vigan, A., Mouillet, D., et al. 2017, *A&A*, 599, A108  
 Milli, J., Engler, N., Schmid, H., et al. 2019, *A&A*, 626, A54  
 Mugnier, L. M., Fusco, T., & Conan, J.-M. 2004, *J. Opt. Soc. Am. A*, 21, 1841  
 Müller, A., van den Ancker, M., Launhardt, R., et al. 2011, *A&A*, 530, A85  
 Müller, A., Keppler, M., Henning, T., et al. 2018, *A&A*, 617, L2  
 Muro-Arena, G., Benisty, M., Ginski, C., et al. 2020, *A&A*, 635, A121

- Pairat, B., Cantalloube, F., & Jacques, L. 2018, in [International Traveling Workshop on Interactions Between Sparse Models and Technology](#)
- Pairat, B., Jacques, L., & Cantalloube, F. 2019, in [Signal Processing with Adaptive Sparse Structured Representations](#), 1, 1
- Pairat, B., Cantalloube, F., & Jacques, L. 2021, [MNRAS](#), 503, 3724
- Pavlov, A., Möller-Nilsson, O., Feldt, M., et al. 2008, in [SPIE Astronomical Telescopes + Instrumentation](#), Int. Soc. Opt. Photonics, 7019, 701939
- Pecaut, M. J., Mamajek, E. E., & Bubar, E. J. 2012, [ApJ](#), 746, 154
- Pueyo, L. 2018, [Handbook of Exoplanets](#), 705
- Ren, B., Dong, R., Esposito, T. M., et al. 2018a, [ApJ](#), 857, L9
- Ren, B., Pueyo, L., Zhu, G. B., Debes, J., & Duchêne, G. 2018b, [ApJ](#), 852, 104
- Ren, B., Pueyo, L., Chen, C., et al. 2020, [ApJ](#), 892, 74
- Riaud, P., Mawet, D., Absil, O., et al. 2006, [A&A](#), 458, 317
- Ruane, G., Ngo, H., Mawet, D., et al. 2019, [AJ](#), 157, 118
- Schneider, G., Smith, B. A., Becklin, E., et al. 1999, [ApJ](#), 513, L127
- Soubies, E., Soulez, F., Mccann, M. T., et al. 2019, [Inverse Prob.](#), 35
- Stein, C. M. 1981, [Ann. Stat.](#), 1135
- Thiébaud, É. 2002, in [SPIE Astronomical Telescopes + Instrumentation](#), Int. Soc. Opt. Photonics, 4847, 174
- Thiébaud, É. 2006, [Optics in Astrophysics](#) (Springer), 397
- van Boekel, R., Henning, T., Menu, J., et al. 2017, [ApJ](#), 837, 132
- Van Leeuwen, F. 2007, [A&A](#), 474, 653
- Wahba, G. 1985, [Ann. Stat.](#), 13, 1378
- Wahhaj, Z., Milli, J., Romero, C., et al. 2020, [A&A](#), 648, A26
- Wainwright, M. J., & Simoncelli, E. P. 2000, [Advances in Neural Information Processing Systems](#), 855

## Appendix A: Constrained minimization of the cost function

The restored object  $\mathbf{x}$  is the minimum of the objective function  $\mathcal{C}(\mathbf{r}, \mathbf{x}, \mathbf{A}, \mathbf{\Omega}, \boldsymbol{\mu})$  defined in Eq. (2) on the nonnegative orthant (i.e.,  $\mathbf{x}$  is everywhere nonnegative). We numerically solve this constrained optimization problem with the VMLM-B algorithm (Thiébaud 2002) as implemented in OPTIMPACKLEGACY<sup>14</sup>. VMLM-B is a limited memory quasi-Newton method implementing optional bound constraints on the sought parameters. This method has proven to be very effective for solving large-scale (millions or even billions of variables) nonlinear problems with or without bound constraints that are typical of inverse problems for image restoration. The requirements for running VMLM-B are to provide the bounds (if any) and a numerical function to compute the objective function and its gradient for a given set of parameters  $\mathbf{x}$ .

The gradient of the objective function  $\mathcal{C}(\mathbf{r}, \mathbf{x}, \mathbf{A}, \mathbf{\Omega}, \boldsymbol{\mu})$  with respect to  $\mathbf{x}$  is given by:

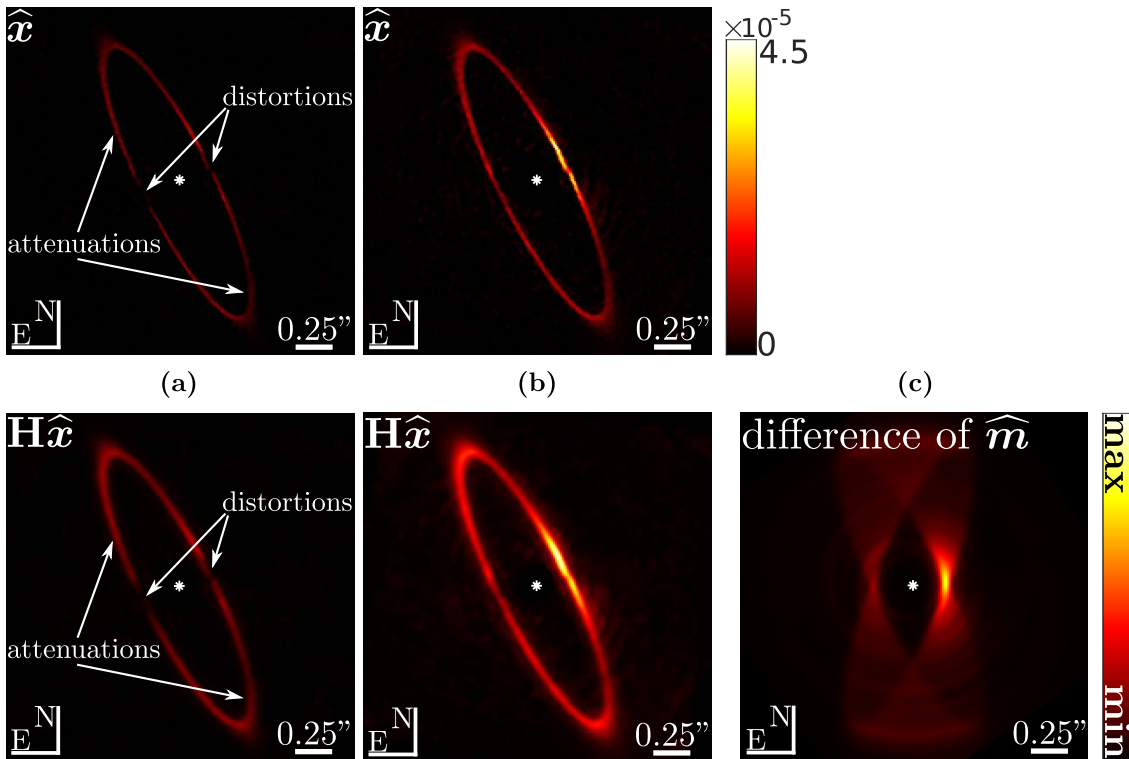
$$\begin{aligned} \nabla_{\mathbf{x}} \mathcal{C}(\mathbf{r}, \mathbf{x}, \mathbf{A}, \mathbf{\Omega}, \boldsymbol{\mu}) = & \underbrace{\sum_{n \in \mathbb{P}} \sum_{t=1}^T \mathbf{A}_t^\top \mathbf{P}_n^\top [\widehat{\mathbf{C}}_n (\mathbf{P}_n (\mathbf{A}_t \mathbf{x} - \mathbf{r}_t + \widehat{\mathbf{m}}))]}_{=\nabla_{\mathbf{x}} \mathcal{D}} \\ & + \underbrace{\mu_{\text{smooth}} \frac{\Delta_n^\top \Delta_n \mathbf{x}}{\sqrt{\|\Delta_n \mathbf{x}\|_2^2 + \epsilon^2}}}_{=\nabla_{\mathbf{x}} \mathcal{R}_{\text{smooth}}} + \underbrace{\mu_{\ell_1} \mathbf{1}_N}_{\nabla_{\mathbf{x}} \mathcal{R}_{\ell_1}}. \end{aligned} \quad (\text{A.1})$$

The computational burden for the gradient is similar to that of the objective function.

## Appendix B: Importance of removing off-axis sources when learning the statistical model of the nuisance component

Figure B.1a illustrates on the example of HR 4796A that estimating the statistics  $\mathbf{\Omega}$  of the nuisance component directly from the data with Eqs. (14) leads to self-subtraction artifacts: A portion of the light of off-axis sources gets removed from the measurements because it is erroneously attributed to the nuisance component (the mean  $\widehat{\mathbf{m}}^{[1]}$  defined in Eq. (14a)).

Figure B.1b shows that refining the statistics of the nuisance component to compensate for the contribution of the off-axis sources improves the reconstruction: Parts of the disk (like its bright west side) that were previously missed are covered. Figure B.1c displays the difference between the mean component  $\widehat{\mathbf{m}}$  estimated with Eq. (14a) and with Algorithm 1. This difference is largest near the star where a significant part of disk contribution is lost in the average of the nuisance component  $\widehat{\mathbf{m}}$  when using Eq. (14a). This explains the bias and artifacts observed in Fig. B.1a.



**Fig. B.1.** Illustration of the influence of the statistics  $\widehat{\mathbf{\Omega}} = \{\widehat{\mathbf{m}}, \{\widehat{\mathbf{C}}_n\}_{n \in \mathbb{P}}\}$  of the nuisance component on the REXPACO reconstructions. (a)  $\widehat{\mathbf{\Omega}}$  is computed with Eqs. (14) ignoring the contribution of the object; (b)  $\widehat{\mathbf{\Omega}}$  is computed with Algorithm 1; (c) difference of the mean component  $\widehat{\mathbf{m}}$  estimated with Eq. (14a) and Algorithm 1. For (a) and (b), both the deconvolved flux distribution  $\widehat{\mathbf{x}}$  and its convolved version  $\mathbf{H}\widehat{\mathbf{x}}$  are given. The difference of estimated means  $\widehat{\mathbf{m}}$  given in (c) is represented with the same colorbar than the convolved objects  $\mathbf{H}\widehat{\mathbf{x}}$  given in (a) and (b) to ease comparisons. Data set: HR 4796A, see Sect. 3.3.1 for observing conditions.

<sup>14</sup> <https://github.com/emmt/OptimPackLegacy>

### Appendix C: Hierarchical estimation of the statistics of the nuisance component and of the flux distribution

We first recall the expression of the shrinkage estimator  $\widehat{\boldsymbol{\Omega}} = \{\widehat{\mathbf{m}}, \{\widehat{\mathbf{C}}_n\}_{n \in \mathbb{P}}\}$  of the nuisance component given in Eqs. (6)–(10):

$$\begin{cases} \widehat{\mathbf{m}}(\mathbf{x}) = \frac{1}{T} \sum_{t=1}^T (\mathbf{r}_t - \mathbf{A}_t \mathbf{x}) \\ \widehat{\mathbf{C}}_n(\mathbf{x}) = \widetilde{\mathbf{W}}_n \odot \widetilde{\mathbf{C}}_n(\mathbf{x}), \forall n \in \mathbb{P} \end{cases} \quad (\text{C.1})$$

with the empirical covariance matrix defined by:

$$\widetilde{\mathbf{C}}_n(\mathbf{x}) = \frac{1}{T} \sum_{t=1}^T \widehat{\mathbf{u}}_{n,t}(\mathbf{x}) \widehat{\mathbf{u}}_{n,t}(\mathbf{x})^\top, \quad (\text{C.2})$$

and the residuals

$$\widehat{\mathbf{u}}_{n,t}(\mathbf{x}) = \mathbf{P}_n (\mathbf{r}_t - \widehat{\mathbf{m}}(\mathbf{x}) - \mathbf{A}_t \mathbf{x}). \quad (\text{C.3})$$

The shrinkage matrix  $\widetilde{\mathbf{W}}_n$  is defined by Eq. (11) and is considered fixed (i.e., independent from  $\mathbf{x}$ ). In the following we show that these estimators are minimizers of the modified cost function  $\mathcal{D}_{\text{joint}}$  defined in Eq. (15):

$$\left( \widehat{\mathbf{m}}(\mathbf{x}), \{\widehat{\mathbf{C}}_n(\mathbf{x})\}_{n \in \mathbb{P}} \right) = \arg \min_{\mathbf{m}, \{\mathbf{C}_n\}_{n \in \mathbb{P}}} \mathcal{D}_{\text{joint}}(\mathbf{r}, \mathbf{x}, \mathbf{m}, \{\mathbf{C}_n\}_{n \in \mathbb{P}}) \quad (\text{C.4})$$

with

$$\begin{aligned} \mathcal{D}_{\text{joint}}(\mathbf{r}, \mathbf{x}, \mathbf{m}, \{\mathbf{C}_n\}_{n \in \mathbb{P}}) &= \frac{T}{2} \sum_{n \in \mathbb{P}} \log \det \mathbf{C}_n \\ &+ \frac{1}{2} \sum_{n \in \mathbb{P}} \text{tr} \left( \mathbf{C}_n^{-1} \left( \widetilde{\mathbf{W}}_n \odot \sum_{t=1}^T \mathbf{u}_{n,t} \mathbf{u}_{n,t}^\top \right) \right), \end{aligned} \quad (\text{C.5})$$

and

$$\mathbf{u}_{n,t} = \mathbf{P}_n (\mathbf{r}_t - \mathbf{m} - \mathbf{A}_t \mathbf{x}). \quad (\text{C.6})$$

*Proof.* For a fixed  $\mathbf{x}$ , the differentiation of  $\mathcal{D}_{\text{joint}}$  with respect to  $\boldsymbol{\Omega}_n = \{\mathbf{m}, \mathbf{C}_n\}$  leads to:

$$\begin{aligned} d\mathcal{D}_{\text{joint}} \Big|_{\boldsymbol{\Omega}_n} &= \frac{T}{2} \text{tr} \left( \mathbf{C}_n^{-1} d\mathbf{C}_n \right) + \\ &\frac{1}{2} \text{tr} \left[ -2\mathbf{C}_n^{-1} \left( \widetilde{\mathbf{W}}_n \odot \sum_{t=1}^T \mathbf{u}_{n,t} (d\mathbf{m})^\top \right) \right. \\ &\left. - \mathbf{C}_n^{-1} d\mathbf{C}_n \mathbf{C}_n^{-1} \left( \widetilde{\mathbf{W}}_n \odot \sum_{t=1}^T \mathbf{u}_{n,t} \mathbf{u}_{n,t}^\top \right) \right]. \end{aligned} \quad (\text{C.7})$$

The necessary first-order optimality condition:

$$d\mathcal{D}_{\text{joint}} \Big|_{\boldsymbol{\Omega}_n} = 0$$

leads to:

$$\begin{cases} \mathbf{C}_n^{-1} \left( \widetilde{\mathbf{W}}_n \odot \sum_{t=1}^T \mathbf{u}_{n,t} \right) = \mathbf{0} \\ T \mathbf{I}_K - \mathbf{C}_n^{-1} \left( \widetilde{\mathbf{W}}_n \odot \sum_{t=1}^T \mathbf{u}_{n,t} \mathbf{u}_{n,t}^\top \right) = \mathbf{0}, \end{cases} \quad (\text{C.8})$$

where  $\mathbf{I}_K \in \mathbb{R}^{K \times K}$  stands for the identity matrix. Since  $\mathbf{C}_n$  is necessary non-singular, the system of Eq. (C.8) gives:

$$\begin{cases} \mathbf{m} = \frac{1}{T} \sum_{t=1}^T (\mathbf{r}_t - \mathbf{A}_t \mathbf{x}) \\ \mathbf{C}_n = \widetilde{\mathbf{W}}_n \odot \frac{1}{T} \sum_{t=1}^T \mathbf{u}_{n,t} \mathbf{u}_{n,t}^\top, \end{cases} \quad (\text{C.9})$$

which corresponds to the general form of the estimators given in Eq. (C.1).