Characterization of stellar companions from high-contrast long-slit spectroscopy data. The EXtraction Of SPEctrum of COmpanion (Exospeco) algorithm

S. Thé, É. Thiébaut, L. Denis, T. Wanner, R. Thiébaut, M. Langlois, F. Soulez

A&A, 678, A77 (2023)
https://doi.org/10.1051/0004-6361/202245565

Astronomy & Astrophysics

Characterization of stellar companions from high-contrast long-slit spectroscopy data

The EXtraction Of SPEctrum of COmpanion (EXOSPECO) algorithm

Samuel Thé[1], Éric Thiébaut[1], Loïc Denis[2], Thibault Wanner[1], Rémi Thiébaut[1], Maud Langlois[1], and Ferréol Soulez[1]

[1] Université de Lyon, Université Lyon1, ENS de Lyon, CNRS, Centre de Recherche Astrophysique de Lyon UMR 5574, 69230 Saint-Genis-Laval, France
e-mail: samuel.the@univ-lyon1.fr, eric.thiebaut@univ-lyon1.fr
[2] Université Jean-Monnet-Saint-Étienne, CNRS, Institut d'Optique Graduate School, Laboratoire Hubert Curien UMR 5516, 42023 Saint-Étienne, France
e-mail: loic.denis@univ-st-etienne.fr

Received 28 November 2022 / Accepted 19 July 2023

ABSTRACT

Aims. High-contrast long-slit spectrographs can be used to characterize exoplanets. The resulting spectroscopic data are, however, corrupted by stellar leakages that largely dominate other signals and make the process of extracting the companion spectrum very challenging. This paper presents a complete method to calibrate the spectrograph and extract the signal of interest.
Methods. The proposed method is based on a flexible direct model of the high-contrast long-slit spectroscopic data. This model explicitly accounts for the instrumental response and for the contributions of both the star and the companion. The contributions of these two components and the calibration parameters are jointly estimated by solving a regularized inverse problem. As this problem has no closed-form solution, we propose an alternating minimization strategy to effectively find the solution.
Results. We tested our method on empirical long-slit spectroscopic data and by injecting synthetic companion signals in these data. The proposed initialization and the alternating strategy effectively avoid the self-subtraction bias, even for companions observed very close to the coronagraphic mask. Careful modeling and calibration of the angular and spectral dispersion laws of the instrument clearly reduce the contamination by the stellar leakages. In practice, the outputs of the method are mostly driven by a single hyper-parameter that tunes the level of regularization of the companion's spectral energy distribution (SED).

Key words. infrared: planetary systems – methods: data analysis – techniques: imaging spectroscopy – instrumentation: spectrographs – instrumentation: adaptive optics

1. Introduction

High-contrast extreme adaptive optics (AO) systems such as SPHERE (Spectro-Polarimetry High-contrast Exoplanet REsearch; Beuzit et al. 2019), GPI (Gemini Planet Imager; Macintosh et al. 2006, 2014), or SCExAO (Jovanovic et al. 2015) have been developed to directly observe the close environment of stars in the visible and the near-infrared. The study of exoplanets and their formation is among the main scientific objectives of these instruments. One of the advantages of high-contrast extreme AO systems is that they can provide direct access to the light from the exoplanet, which is crucial when performing spectral characterizations. However, substantial contamination by the light from the host star occurs: in the visible and the near-infrared, in spite of the real-time correction by the AO system and of the masking of the host star by a coronagraph, the residual stellar light diffracted by the instrument is much brighter than that received from most exoplanets of interest. For this reason, dedicated post-processing methods have been developed to track evidence of the presence of exoplanets in data corrupted by strong stellar leakages. The number of published detection algorithms, LOCI (Lafreniere et al. 2007), TLOCI (Marois et al. 2013), KLIP (Soummer et al. 2012), MOODS (Smith et al. 2009), ANDROMEDA (Mugnier et al. 2009), PEX (Devaney & Thiébaut 2017), and PACO (Flasseur et al. 2018, 2020a,b), to name a few, reflects the scientific interest but also the intrinsic difficulty of trustfully detecting an exoplanet from sequences of high-contrast images. The most successful of these methods are those that account for the statistics of the stellar leakages (notably their correlations), no matter if they consist of sequences of images (Smith et al. 2009; Flasseur et al. 2018, 2020b), sequences of multi-spectral images from Integral field spectrographs (IFS; Flasseur et al. 2020a), or even multi-epoch sequences of images (Dallant et al. 2022).

After its detection, the direct characterization of an exoplanet is possible with high-contrast extreme AO systems equipped with a spectrograph. Both SPHERE and GPI are equipped with low-resolution IFS. In addition, SPHERE/IRDIS is equipped with a medium (MRS) resolution long-slit spectrograph (LSS) in J, H, and K bands, with the latter being also available at low (LRS) resolution[1] (Dohlen et al. 2008). With SPHERE/IRDIS/LSS, the spectrum of a detected companion can then be measured by aligning the slit of the spectrograph with the host star and the companion while the host star is occulted by an opaque mask combined with the slit. In an LSS image, the stellar leakages take the form of speckles spectrally dispersed along oblique lines generally brighter than the companion spectrum (see Fig. 1 for an example). In order to get rid of these

[1] $\lambda/\Delta\lambda = 35$ for the LRS mode or $\lambda/\Delta\lambda = 400$ for the MRS mode.

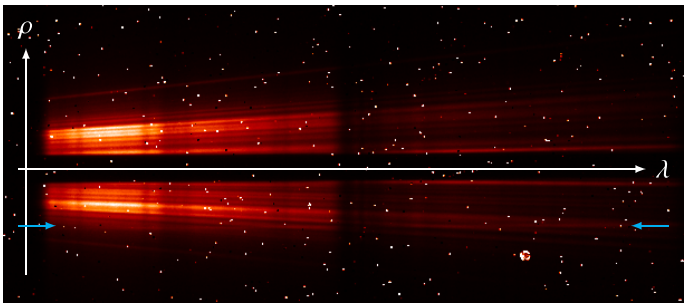

Fig. 1. Long-slit medium resolution spectroscopy data of HR 3549 taken by IRDIS, with horizontally the spectral axis λ and vertically the angular separation axis, ρ. The blue arrows indicate the position of the companion.

stellar leakages, Vigan et al. (2008, 2012) have developed a spectral deconvolution (SD) method following the work of Sparks & Ford (2002). The SD method consists of a filtering of the LSS image after a geometrical transform to align the speckles along a given direction. In practice, the SD method is quite sensitive to the alignment of the instrument, requires fixing defective pixels, and suffers from a self-subtraction bias. The latter is due to an overestimation of the stellar leakages caused by the presence of the companion. To improve on the SD method and reduce the self-subtraction bias, Mesa et al. (2016) adapted the strategy implemented in TLOCI (Marois et al. 2013) to the LSS data. In spite of these improvements, existing extraction methods suffer from a number of faults, most of them stemming from the requirement to geometrically transform the LSS data to align the dispersed speckles. In particular, they provide, at best, a least-squares estimation of the stellar leakages which is sub-optimal as the noise is not independent or identically distributed (i.i.d.) in the geometrically transformed images (Thiébaut et al. 2016). To overcome the drawbacks of existing methods, we propose formulating the extraction of the spectrum of a companion as an inverse problem. The inverse problem corresponds to the joint estimation of the contributions of the star and of its companion from the LSS data. Not only does this approach require no transformation of the LSS data (thus avoiding the introduction of correlations) but it also yields statistically optimal estimators. To cope with both the possible instrumental misalignment and the lack of a closed-form solution for the inverse problem, we implemented an alternating optimization strategy with optional self-calibration stages to solve the problem.

The outline of the paper is as follows. In Sect. 2, we present a model of the distribution of the light on the detector of a LSS instrument. This model is used to illustrate how (following a geometrical transform) stellar leakages can be partially removed by a truncated singular value decomposition (TSVD) before extracting the companion signal. Such an approach is representative of the optimal performance that can be reached by standard methods. We present in Sect. 3 our approach to jointly estimate the stellar leakages and the contribution of the companion without transforming the data. For the processing of LSS data, knowledge the spectro-angular coordinates of each detector pixel is requisite and we describe in Sect. 4 a numerical method to estimate the spatio-spectral dispersion laws from given calibration data. In Sect. 5, we validate the proposed method on both real data from SPHERE/IRDIS/MRS and on injections of synthetic companions in real data. We show the importance of the calibration (described in Appendix A) and compare our method with more standard approaches. Finally, we conclude in Sect. 6 and present some possible modifications of the proposed methods.

2. State-of-the-art processing

In spite of the coronagraphic mask in high-contrast data, the stellar leakages, taking the form of quasi-static dispersed speckles, largely dominate the signal of interest, namely, the spectrum of the companion. These speckles, whose contribution cannot be precisely determined by using other stars (i.e., by reference differential imaging; Xie et al. 2022) or by rotating the slit to hide the planet (Vigan 2016) are a major source of nuisance for extracting the companion spectrum. This section introduces a modeling of the LSS data that is used in Sect. 3 to design our spectrum extraction method. This model is also useful to explain how previous approaches perform the suppression of the stellar leakages (Vigan et al. 2008; Mesa et al. 2016).

2.1. Image formation

Figure 1 shows a single exposure captured by the LSS of SPHERE/IRDIS. The vector $\boldsymbol{d} \in \mathbb{R}^N$, with N the number of pixels, can be modeled by:

$$d_n = m(\rho_n, \lambda_n) + \varepsilon_n, \tag{1}$$

with $m(\rho, \lambda)$ the distribution of light in the detector plane at an angular coordinate, ρ, along the slit and wavelength, λ and ρ_n, and λ_n as the angular and spectral coordinates at the nth pixel, and ε_n the contribution of the noise. Our notations are summarized in Table 1. The light distribution in the detector plane is the sum of the contributions by the star and by the companion:

$$m(\rho, \lambda) = f_\star(\lambda)\, h_\star(\rho, \lambda) + f_\oplus(\lambda)\, h_\oplus(\rho, \lambda), \tag{2}$$

with $f_\star$ and $f_\oplus$ the spectral energy distributions (SEDs) of the star and of the companion as seen by the detector[2], $h_\star$, and $h_\oplus$ the point spread functions (PSFs) for a source at the respective angular positions of the star and of the companion, namely, the so-called on-axis and off-axis PSFs. The on-axis PSF, $h_\star(\rho, \lambda)$ explains the oblique bright lines due to stellar leakages in Fig. 1, while the stellar SED $f_\star(\lambda)$ explains the variations of intensity along these lines. As can be seen in Fig. 1, the companion signal, that is, $f_\oplus(\lambda)\, h_\oplus(\rho, \lambda)$, is barely distinguishable in the LSS data and it is mandatory to get rid of the stellar leakages $f_\star(\lambda)\, h_\star(\rho, \lambda)$.

2.2. Low-rank approximation of the stellar leakages

Devaney & Thiébaut (2017) have shown that, except in the vicinity of the coronagraphic mask, the chromatic PSF can be written in the form of a series expansion. Applying their model to the star and taking into account that our data have one angular dimension instead of two yields:

$$h_\star(\rho, \lambda) = \sum\nolimits_{k\geq 1} \gamma(\lambda)^k\, h_{\star,k}(\gamma(\lambda)\,(\rho - \rho_\star)), \tag{3}$$

with $\gamma(\lambda) = \lambda^{\mathrm{ref}}/\lambda$ a chromatic magnification factor relative to some arbitrary reference wavelength λ^{ref}, $\rho_\star$ the angular position of the star along the slit to account for a possible pointing error of the instrument, and $\{h_{\star,k}\}_{k=1,\ldots}$ a family of spatial PSF modes at the reference wavelength.

Since the stellar leakages dominate the signal in the LSS image, $\boldsymbol{d}$, Eq. (3) suggests applying specific image warping so as to form a 2D image, $\boldsymbol{d}^{\mathrm{warp}}$, whose first dimension may vary along

[2] These SEDs are affected by the chromatic transmission of the atmosphere and of the instrument.

Table 1. Notations.

Notation	Description
Subscript $\star$	Parameters of the stellar model
Subscript $\oplus$	Parameters of the companion model
$\boldsymbol{d} \in \mathbb{R}^N$	Science data
$\boldsymbol{m} \in \mathbb{R}^N$	Sampled model of $\boldsymbol{d}$
$\boldsymbol{w} \in \mathbb{R}^N$	Diagonal of the precision matrix of $\boldsymbol{d}$
$\boldsymbol{\lambda} \in \mathbb{R}^N$	Pixel-wise wavelengths
$\boldsymbol{\rho} \in \mathbb{R}^N$	Pixel-wise angular positions
$f_\star$	Continuous star SED
$\boldsymbol{x} \in \mathbb{R}^{N_x}$	Sampled star SED $f_\star$
$\boldsymbol{\lambda}_\star^{\mathrm{grd}} \in \mathbb{R}^{N_x}$	Sampling wavelengths for $\boldsymbol{x}$
$\mathbf{F}_\star \in \mathbb{R}^{N \times N_x}$	Interpolation operator: $\boldsymbol{x}$ to pixel-wise $f_\star$
$h_\star$	Continuous on-axis PSF
$\boldsymbol{y} \in \mathbb{R}^{N_y}$	Sampled on-axis PSF $h_\star$ at λ^{ref}
$\boldsymbol{\rho}_\star^{\mathrm{grd}} \in \mathbb{R}^{N_y}$	Sampling angles for $\boldsymbol{y}$
$\rho_\star$	Angular position of the star
$\boldsymbol{\nu}_\star \in \Omega_\star$	Calibration parameters of $h_\star$
$\Omega_\star$	Feasible set $h_\star$ parameters
$\mathbf{H}_\star \in \mathbb{R}^{N \times N_y}$	Interpolation operator: $\boldsymbol{y}$ to pixel-wise $h_\star$ at λ^{ref}
$f_\oplus$	Continuous companion SED
$\boldsymbol{z} \in \mathbb{R}^{N_z}$	Sampled companion SED $f_\oplus$
$\boldsymbol{\lambda}_\oplus^{\mathrm{grd}} \in \mathbb{R}^{N_x}$	Sampling wavelengths for $\boldsymbol{z}$
$\mathbf{F}_\oplus \in \mathbb{R}^{N \times N_z}$	Interpolation operator: $\boldsymbol{z}$ to pixel-wise $f_\oplus$
$h_\oplus$	Continuous off-axis PSF
$\boldsymbol{h}_\oplus \in \mathbb{R}^N$	Sampled off-axis PSF $h_\oplus$ at λ^{ref}
$\boldsymbol{\nu}_\oplus \in \Omega_\oplus$	Calibration parameters of $h_\oplus$
$\Omega_\oplus$	Feasible set of $h_\oplus$ parameters
$\rho_\oplus \in \mathbb{R}$	Angular position of the companion
$\boldsymbol{\mu} = (\mu_x, \mu_y, \mu_z)$	Hyper-parameters
$\boldsymbol{\gamma} \in \mathbb{R}^N$	Pixel-wise chromatic scaling factors
$\boldsymbol{\nu} = (\boldsymbol{\nu}_\star, \boldsymbol{\nu}_\oplus) \in \Omega$	Calibration parameters
$\Omega = \Omega_\star \times \Omega_\oplus$	Feasible set of calibration parameters
Λ	Spectral dispersion law
$\boldsymbol{a}$	Parameters of the spectral dispersion law
ϱ	Angular dispersion law
$\boldsymbol{s}$	Parameters of the angular dispersion law
$\Delta\rho$	Width of the coronagraphic mask

Notes. Lowercase letters are for continuous functions and scalars (e.g., $f_\star$), boldface lowercase letters for vectors (e.g., $\boldsymbol{x}$), and boldface uppercase letters for linear mappings, a.k.a. matrices (e.g., $\boldsymbol{F}_\star$). Vectors with a hat (e.g., $\hat{\boldsymbol{x}}$) are estimators. The main unknowns of the problem are $\boldsymbol{x}$, the sampled star SED, $\boldsymbol{y}$, the sampled on-axis PSF, and $\boldsymbol{z}$, the sampled companion SED.

the wavelength, while its second dimension varies along the coordinate $s = \gamma(\lambda)\,(\rho - \rho_\star)$. For more details, we refer to Fig. 2. According to Eqs. (1) and (3), the warped image is modeled and then approximated by:

$$d_{i,j}^{\mathrm{warp}} = m\Big(\rho_\star + s_j^{\mathrm{warp}}/\gamma\Big(\lambda_i^{\mathrm{warp}}\Big), \lambda_i^{\mathrm{warp}}\Big) + \varepsilon_{i,j}^{\mathrm{warp}}, \tag{4}$$

$$\approx \sum_{k \geq 1} \gamma\Big(\lambda_i^{\mathrm{warp}}\Big)^k f_\star\Big(\lambda_i^{\mathrm{warp}}\Big) h_{\star,k}\Big(s_j^{\mathrm{warp}}\Big), \tag{5}$$

where $\boldsymbol{\varepsilon}^{\mathrm{warp}}$ in Eq. (4) denotes the contribution of the noise in the warped image while the $\approx$ symbol in Eq. (5) is to account for the contributions of the potential companion and for the noise which have been neglected. In other words, the stellar leakages appear to have a simple separable decomposition in the warped image.

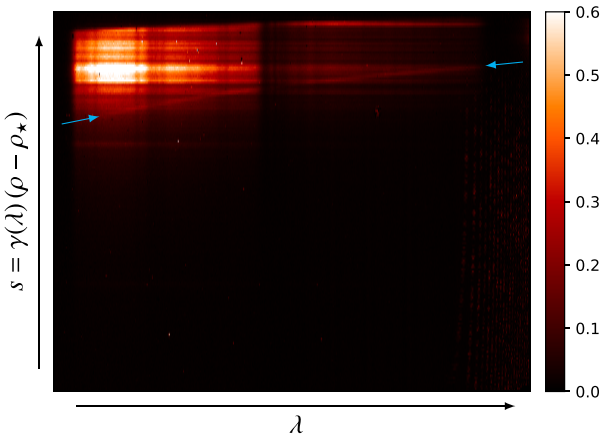

Fig. 2. Warped HR 3549 image. This figure shows the bottom half of the data shown in Fig. 1, corresponding to the side where lies the companion, warped so as to align the dispersed speckles of the stellar leakages. The warping is defined by the coordinates ρ and λ of the pixels given by the "complex" calibration model of the spectral and angular dispersion laws described in Appendix A. The companion signal can be seen as a faint curved track indicated by the blue arrows.

The singular value decomposition (SVD) of the warped image[3] is expressed as:

$$\boldsymbol{d}^{\mathrm{warp}} = \sum_{k=1}^{\min(N_1,N_2)} \boldsymbol{u}_k\, \sigma_k\, \boldsymbol{v}_k^\top, \tag{6}$$

where N_1 and N_2 are the dimensions of the warped image, $\boldsymbol{u}_k \in \mathbb{R}^{N_1}$ is the kth left singular vector of the decomposition, $\sigma_k \geq 0$ is the kth singular value, and $\boldsymbol{v}_k \in \mathbb{R}^{N_2}$ is the kth right singular vector. Comparing Eq. (5) and Eq. (6), we see that the SVD of $\boldsymbol{d}^{\mathrm{warp}}$ readily provides a decomposition similar to the contribution of the stellar leakages with, for each index, k, the left singular vector, $\boldsymbol{u}_k$, sampling $\gamma(\lambda)^k\, f_\star(\lambda)$ as a function of λ and the right singular vector, $\boldsymbol{v}_k$, sampling $h_{\star,k}(s)$ as a function of s (both up to a normalization factor that depends on k). The truncated singular value decomposition (TSVD) of the warped image is obtained by limiting the sum in the right-hand side of Eq. (6) to the $k_{\max} \leq \min(N_1, N_2)$ first terms; also, according to the Eckart-Young-Mirsky (Eckart & Young 1936; Mirsky 1960) theorem, it is the best possible approximation of $\boldsymbol{d}^{\mathrm{warp}}$ of rank $k_{\max}$ in the least-squares approach. Hence, it may be assumed that, for a suitable choice of $k_{\max}$, the TSVD of $\boldsymbol{d}^{\mathrm{warp}}$ provides a good approximation of the stellar leakages without being too much affected by the companion signal (if the companion is not too bright) and by the noise. A residual image that mostly depends on the companion can then be formed by subtracting the un-warped TSVD of the warped image $\boldsymbol{d}^{\mathrm{warp}}$ from the LSS image $\boldsymbol{d}$:

$$\boldsymbol{r}_\oplus = \boldsymbol{d} - \mathcal{U}\left(\sum_{k=1}^{k_{\max}} \boldsymbol{u}_k\, \sigma_k\, \boldsymbol{v}_k^\top\right), \tag{7}$$

where $\mathcal{U}$ denotes the un-warping operation[4]. As illustrated by Fig. 3, the signal of interest (i.e., the companion SED) is then

[3] Considered as a simple matrix.
[4] e.g., a simple separable 2D interpolation.

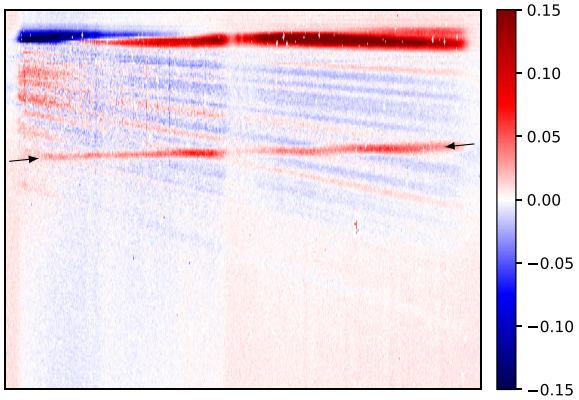

Fig. 3. Bottom half of the residual image $\boldsymbol{r}_\oplus$ for the HR 3549 data with stellar leakages estimated by the TSVD method as defined in Eq. (7) and with the warped image shown in Fig. 2. Compared to the original data shown in Fig. 1, the companion signal appears more distinctly (indicated by the two arrows).

easier to extract from the residual image, $\boldsymbol{r}_\oplus$. This can be done with standard aperture photometry tools.

As pointed out by Devaney & Thiébaut (2017), there are a number of issues in using the TSVD to get rid of the stellar leakages in multi-wavelengths high-contrast data. First, to produce a rectangular warped image (that can be interpreted as a matrix to perform the SVD), quite substantial regions of the original data, $\boldsymbol{d}$, have to be discarded (i.e., those near the coronagraphic mask and the edges of the formed image). This limits the range of admissible angular positions for the companion and gets rid of data that might be valuable to improve the estimation of the stellar leakages. Second, the presence of a companion in the original data, $\boldsymbol{d}$, yields a positive bias in the approximation of the stellar leakages by the TSVD. This results in a negative bias in the residual image and, hence, in the estimated companion SED. This artifact is known in the literature as "self-subtraction". Third, the least-squares fit performed by the TSVD of $\boldsymbol{d}^{\mathrm{warp}}$ is sub-optimal regarding the distribution of the noise in the warped image. Indeed, a least squares approach is only optimal for independent identically distributed (i.i.d.) noise, which is certainly not the case for $\varepsilon_{i,j}^{\mathrm{warp}}$: at least, the shot noise in the image, $\boldsymbol{d}$ has a non-uniform distribution and a side effect of the transform of $\boldsymbol{d}$ to yield the warped image, $\boldsymbol{d}^{\mathrm{warp}}$, is the introduction of correlations. Moreover, defective pixels, which are quite numerous for the kind of detectors used by NIR instruments such as LSS, must be corrected, usually by averaging their neighbors' values, before warping the image. This correction can only introduce additional correlations.

In spite of these drawbacks, the proposed processing methods (Vigan et al. 2008; Mesa et al. 2016) are similar to the TSVD approximation of warped LSS images (optimal linear combination of a set of images). Some refinements have been proposed to limit the self-subtraction bias (Mesa et al. 2016) but the other issues have been largely left unaddressed. In the rest of this paper, we propose a new "inverse problems" approach to solve all the aforementioned limitations.

3. Inverse problems approach

To avoid the issues resulting from warping the LSS image, we propose solving an inverse problem that consists of jointly estimating the parameters of the direct model of the data given in Eqs. (1) and (2) without transforming the data themselves. For an optimal information extraction, we model the likelihood of the data to consider the uneven quality of the data and, therefore, to account for defective pixels or missing data in a consistent way. Besides, our approach relies on a precise calibration of the spectro-spatial instrumental dispersion as seen by the detector. The proposed method includes auto-calibration stages to refine the calibration parameters and thus accounts for a possible misalignment of the science exposures.

3.1. Assumed continuous model

To simplify the on-axis PSF model in Eq. (3), we kept only the first and most significant of these modes. Thus we assume that:

$$h_\star(\rho,\lambda) = \gamma(\lambda)\, h_\star(\gamma(\lambda)\,(\rho-\rho_\star)), \tag{8}$$

with $h_\star(\rho) = h_{\star,1}(\rho)$ as the first spatial mode of the on-axis PSF. As shown in Sect. 5, this simple model of the stellar leakages already gives excellent results. Likewise, the chromatic off-axis PSF $h_\oplus$ can also be written as:

$$h_\oplus(\rho,\lambda) \approx \gamma(\lambda)\, h_\oplus(\gamma(\lambda)\,(\rho-\rho_\oplus)), \tag{9}$$

with $h_\oplus(\rho) = h_\oplus(\rho,\lambda^{\mathrm{ref}})$ the off-axis PSF at the reference wavelength λ^{ref} and $\rho_\oplus$ the angular position of the companion along the slit. We note that the $\gamma(\lambda)$ factor ensures that the on-axis and off-axis PSFs be normalized at all wavelengths provided the PSF at the reference wavelength be also normalized, namely, $\int h(\rho,\lambda)\,\mathrm{d}\rho = 1\ (\forall\lambda)$. These approximations for the on-axis and off-axis PSFs yield the following simplified model, presented in Fig. 4, that we assume for the rest of the paper:

$$\begin{aligned} m(\rho,\lambda) = \gamma(\lambda)\,[&f_\star(\lambda)\, h_\star(\gamma(\lambda)\,(\rho-\rho_\star)) \\ &+ f_\oplus(\lambda)\, h_\oplus(\gamma(\lambda)\,(\rho-\rho_\oplus))]. \end{aligned} \tag{10}$$

3.2. Discretized distribution

In order to fit the data, the model $m(\rho,\lambda)$ in Eq. (10) has to be estimated at each angular and spectral coordinates (ρ_n,λ_n) of the N pixels of the detector. These pixel coordinates can be identified by fitting angular and spectral dispersion laws to calibration data, as explained in Appendix A. Because of these angular and spectral dispersion laws, continuous functions are needed to model the SEDs of the star and the companion ($f_\star$ and $f_\oplus$) and their respective PSFs ($h_\star$ and $h_\oplus$) on the sensor pixel grid. Next, we explain how we parameterize these functions.

Our models of the star SED $f_\star(\lambda)$, of the on-axis PSF $h_\star(\rho)$, and of the companion SED $f_\oplus(\lambda)$ are given by the following linear interpolations:

$$f_\star(\lambda) = \sum_{j=1}^{N_x} \varphi_\star\left(\lambda-\lambda_{\star,j}^{\mathrm{grd}}\right) \underbrace{f_\star\left(\lambda_{\star,j}^{\mathrm{grd}}\right)}_{x_j}, \tag{11a}$$

$$h_\star(\rho) = \sum_{j=1}^{N_y} \psi_\star\left(\rho-\rho_{\star,j}^{\mathrm{grd}}\right) \underbrace{h_\star\left(\rho_{\star,j}^{\mathrm{grd}}\right)}_{y_j}, \tag{11b}$$

$$f_\oplus(\lambda) = \sum_{j=1}^{N_z} \varphi_\oplus\left(\lambda-\lambda_{\oplus,j}^{\mathrm{grd}}\right) \underbrace{f_\oplus\left(\lambda_{\oplus,j}^{\mathrm{grd}}\right)}_{z_j}, \tag{11c}$$

with $\varphi_\star\colon \mathbb{R}\to\mathbb{R}$, $\psi_\star\colon \mathbb{R}\to\mathbb{R}$, and $\varphi_\oplus\colon \mathbb{R}\to\mathbb{R}$ chosen interpolation functions, and where $\boldsymbol{\lambda}_\star^{\mathrm{grd}} \in \mathbb{R}^{N_x}$ is an evenly spaced

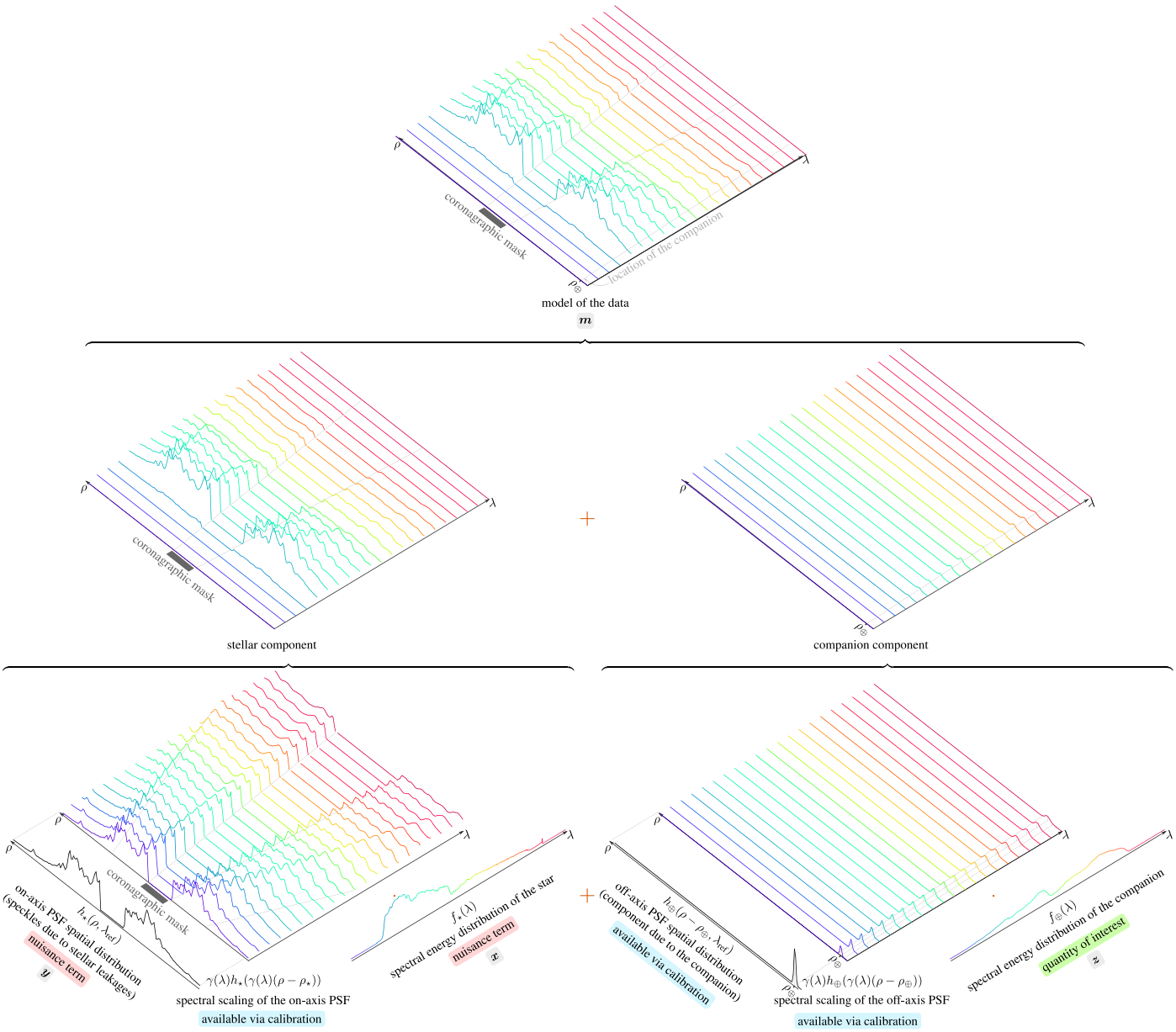

Fig. 4. Illustration of the direct model for high-contrast long-slit spectroscopy given in Eq. (10). The data are modeled as the sum of two components: a stellar component and a companion component. Extracting the SED of the companion also requires the estimation of the on-axis PSF and the SED of the host star. The label "available via calibration" denotes components that may be self-calibrated by EXOSPECO directly from the science data (see Sect. 3.5 for details).

grid of wavelengths to sample the star SED $f_\star$, $\boldsymbol{\rho}_\star^{\mathrm{grd}} \in \mathbb{R}^{N_y}$ is an evenly spaced grid of angles to sample the on-axis PSF $h_\star$, and $\boldsymbol{\lambda}_\oplus^{\mathrm{grd}} \in \mathbb{R}^{N_z}$ is an evenly spaced grid of wavelengths to sample the companion SED $f_\oplus$. At the coordinates (ρ_n, λ_n) of any pixel $n \in [\![1, N]\!]$ of the detector, our linear interpolation yields:

$$f_{\star,n} = f_\star(\lambda_n) = \sum_{j=1}^{N_x} \underbrace{\varphi_\star\left(\lambda_n - \lambda_{\star,j}^{\mathrm{grd}}\right)}_{F_{\star,n,j}} x_j = (\mathbf{F}_\star\, \boldsymbol{x})_n, \tag{12a}$$

$$h_{\star,n} = h_\star(\gamma_n\,(\rho_n - \rho_\star)) = \sum_{j=1}^{N_y} \underbrace{\psi_\star\left(\gamma_n\left(\rho_n - \rho_\star - \rho_{\star,j}^{\mathrm{grd}}\right)\right)}_{H_{\star,n,j}} y_j$$
$$= (\mathbf{H}_\star\, \boldsymbol{y})_n, \tag{12b}$$

$$f_{\oplus,n} = f_\oplus(\lambda_n) = \sum_{i=1}^{N_z} \underbrace{\varphi_\oplus\left(\lambda_n - \lambda_{\oplus,j}^{\mathrm{grd}}\right)}_{F_{\oplus,n,i}} z_i = (\mathbf{F}_\oplus\, \boldsymbol{z})_n, \tag{12c}$$

with $\gamma_n = \gamma(\lambda_n)$, and the matrices $\mathbf{F}_\star \in \mathbb{R}^{N\times N_x}$, $\mathbf{H}_\star \in \mathbb{R}^{N\times N_y}$, and $\mathbf{F}_\oplus \in \mathbb{R}^{N\times N_z}$ defined in Eqs. (12a)–(12c) represent interpolation operators[5]. These operators are applied to the vectors[6] $\boldsymbol{x} \in \mathbb{R}^{N_x}$, $\boldsymbol{y} \in \mathbb{R}^{N_y}$, and $\boldsymbol{z} \in \mathbb{R}^{N_z}$ defined in Eqs. (11a)–(11c). They form the unknown parameters of our models of the star SED $f_\star$, of the on-axis PSF $h_\star$, and of the companion SED $f_\oplus$.

[5] In practice, the interpolation operators are very sparse and only their non-zero entries need to be stored or computed on the fly.

[6] We use boldface lowercase letters to denote the vectors, that is, quantities that depend on a single index, while boldface uppercase letters denote linear operators, that is, quantities that depend on two indices.

The interpolation functions ($\varphi_\star$, $\psi_\star$, and $\varphi_\oplus$) and the sampling lists ($\boldsymbol{\lambda}_\star^{\mathrm{grd}}$, $\boldsymbol{\rho}_\star^{\mathrm{grd}}$, and $\boldsymbol{\lambda}_\oplus^{\mathrm{grd}}$) may be chosen differently for each component of the model. If the spectral sampling lists and spectral interpolation functions are the same (as we chose for our experiments), then the two spectral interpolation operators $\mathbf{F}_\oplus$ and $\mathbf{F}_\star$ are the same. In our implementation of the method, we selected the Catmull & Rom (1974) cardinal cubic spline φ as the interpolation function: $\varphi_\star(\lambda) = \varphi(\lambda/\Delta\lambda_\star^{\mathrm{grd}})$, $\psi_\star(\rho) = \varphi(\rho/\Delta\rho_\star^{\mathrm{grd}})$, and $\varphi_\oplus(\lambda) = \varphi(\lambda/\Delta\lambda_\oplus^{\mathrm{grd}})$ with $\Delta\lambda_\star^{\mathrm{grd}}$, $\Delta\rho_\star^{\mathrm{grd}}$, and $\Delta\lambda_\oplus^{\mathrm{grd}}$ the sampling steps of $\boldsymbol{\lambda}_\star^{\mathrm{grd}}$, $\boldsymbol{\rho}_\star^{\mathrm{grd}}$, and $\boldsymbol{\lambda}_\oplus^{\mathrm{grd}}$.

For the off-axis PSF $h_\oplus(\rho)$ at the reference wavelength, we consider a simple parametric model. Since the principal lobe of the off-axis PSF represents most of the energy received from the companion, we assume a Gaussian approximation:

$$h_\oplus(\rho) = \frac{1}{\sqrt{2\pi}\sigma_\oplus} \exp\Big(-\frac{\rho^2}{2\sigma_\oplus^2}\Big). \tag{13}$$

Hence $\boldsymbol{h}_\oplus$, the sampled off-axis PSF at the reference wavelength for the companion, depends on the angular position of the companion $\rho_\oplus$ and on $\sigma_\oplus$ the standard deviation of the PSF at the reference wavelength. Other parametric models of the off-axis PSF could be considered with a simple adaptation of the algorithm proposed in Sect. 3.4.

Finally, we introduce the N-vectors $\boldsymbol{m} \in \mathbb{R}^N$, $\boldsymbol{\gamma} \in \mathbb{R}^N$ and $\boldsymbol{h}_\oplus \in \mathbb{R}^N$ defined by:

$$m_n = m(\rho_n, \lambda_n), \tag{14a}$$

$$\gamma_n = \gamma(\lambda_n) = \lambda^{\mathrm{ref}}/\lambda_n, \tag{14b}$$

$$h_{\oplus,n} = h_\oplus(\gamma_n\,(\rho_n - \rho_\oplus)), \tag{14c}$$

for $n \in [\![1, N]\!]$. The discretized model of the light distribution in Eq. (10) is then written as:

$$\boldsymbol{m}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}, \boldsymbol{\nu}) = \boldsymbol{\gamma} \odot (\mathbf{H}_\star(\boldsymbol{\nu})\,\boldsymbol{y}) \odot (\mathbf{F}_\star\,\boldsymbol{x}) + \boldsymbol{\gamma} \odot \boldsymbol{h}_\oplus(\boldsymbol{\nu}) \odot (\mathbf{F}_\oplus\,\boldsymbol{z}) \tag{15}$$

with $\odot$ the Hadamard product (entry-wise multiplication) and $\boldsymbol{\nu} = (\boldsymbol{\nu}_\star, \boldsymbol{\nu}_\oplus)$ the calibration parameters of the model, which are the other unknown parameters than $\boldsymbol{x}$, $\boldsymbol{y}$, or $\boldsymbol{z}$. With our Gaussian approximation of the off-axis PSF at the reference wavelength, the calibration parameters for the companion are $\boldsymbol{\nu}_\oplus = (\rho_\oplus, \sigma_\oplus)$. To account for a possible misalignment between the coronagraphic mask and the star, the calibration parameters for the star are just $\boldsymbol{\nu}_\star = (\rho_\star)$, with $\rho_\star$ the angular position of the star along the slit.

The signal-processing problem then amounts to estimating the companion's SED $\boldsymbol{z}$ as well as the other nuisance parameters of the model, $\boldsymbol{x}$, $\boldsymbol{y}$, and $\boldsymbol{\nu}$. A method to perform this task is proposed in the next section.

3.3. Objective function and regularization

After proper calibration of the detector, raw images are pre-processed to compensate for bias and gain non-uniformity and to identify defective pixels (i.e., pixels with a non-linear response). This pre-processing produces the long-slit spectroscopy data $\boldsymbol{d} \in \mathbb{R}^N$ considered here and modeled by $\boldsymbol{m}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}, \boldsymbol{\nu})$ in Eq. (15). Due to photon and detector noises as well as modeling inaccuracies, some discrepancies are expected between the data, $\boldsymbol{d}$, and our model $\boldsymbol{m}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}, \boldsymbol{\nu})$. Due to the observed flux level, there are enough photons detected per pixel for the data, $\boldsymbol{d}$, to approximately follow a Gaussian distribution of mean the model $\boldsymbol{m}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}, \boldsymbol{\nu})$ and of precision matrix[7], $\mathbf{W}$. Since we directly considered the data without any pixel interpolation (i.e., no image warping to align the dispersed speckles and no attempt to fix defective pixels), no correlations are introduced in the data and the pixels can be considered as mutually independent. The precision matrix is thus diagonal, $\mathbf{W} = \mathrm{diag}(\boldsymbol{w})$ where $\boldsymbol{w} \in \mathbb{R}^N$ collects the diagonal entries of $\mathbf{W}$ and is given by:

$$w_n = \begin{cases} 0 & \text{if } n\text{th pixel is invalid,} \\ 1/\mathrm{Var}(d_n) & \text{otherwise.} \end{cases} \tag{16}$$

where $\mathrm{Var}(d_n)$ can be estimated by different pre-processing methods (Mugnier et al. 2004; Berdeu et al. 2020). We consider pixels to be invalid when the model is shown to be incorrect; this includes defective pixels, pixels that are overly impacted by the coronagraphic mask, and pixels located outside of the field of view (see Fig. 6). We assume that the estimation of the variances and the identification of defective pixels are part of the pre-processing stage. The definition of the precision matrix in Eq. (16) amounts to assuming that the variance of invalid pixels is infinite. In other words, this expresses that the values of invalid pixels should not be considered at all. Given the large number of unknowns, the estimation of the stellar and companion components $\boldsymbol{x}$, $\boldsymbol{y}$, and $\boldsymbol{z}$ cannot be performed solely by fitting the data: regularity constraints are necessary to prevent noise amplification and cope with missing data (Titterington 1985). We consider regularized estimators obtained by minimizing the following criterion:

$$C(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}, \boldsymbol{\nu}, \boldsymbol{\mu}) = \|\boldsymbol{d} - \boldsymbol{m}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}, \boldsymbol{\nu})\|_{\mathbf{W}}^2 + \mathcal{R}_{x\,y\,z}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}, \boldsymbol{\mu}), \tag{17}$$

where the first term is a statistical distance between the model and the data (the co-log-likelihood) while $\mathcal{R}_{x\,y\,z}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}, \boldsymbol{\mu})$ is a regularization term parameterized by the vector $\boldsymbol{\mu}$ of so-called hyper-parameters. In the above equation, $\|\boldsymbol{u}\|_{\mathbf{W}}^2 = \boldsymbol{u}^\top \mathbf{W}\,\boldsymbol{u}$ denotes the squared Mahalanobis (1936) norm. Our estimators $\hat{\boldsymbol{x}}$, $\hat{\boldsymbol{y}}$, $\hat{\boldsymbol{z}}$, and $\hat{\boldsymbol{\nu}}$ of the parameters of interest are the ones that jointly minimize the criterion in Eq. (17):

$$(\hat{\boldsymbol{x}}(\boldsymbol{\mu}), \hat{\boldsymbol{y}}(\boldsymbol{\mu}), \hat{\boldsymbol{z}}(\boldsymbol{\mu}), \hat{\boldsymbol{\nu}}(\boldsymbol{\mu})) = \underset{\substack{\boldsymbol{x}\geq 0,\, \boldsymbol{y}\geq 0,\\ \boldsymbol{z}\geq 0,\, \boldsymbol{\nu}\in\Omega}}{\arg\min}\, C(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}, \boldsymbol{\nu}, \boldsymbol{\mu}). \tag{18}$$

These estimators depend on the hyper-parameters $\boldsymbol{\mu}$, as made explicit by the notation. As the parameters $\boldsymbol{x}$, $\boldsymbol{y}$, and $\boldsymbol{z}$ represent nonnegative quantities, their estimators are improved by enforcing nonnegativity as indicated by the inequality constraints in Eq. (18) such as $\boldsymbol{x} \geq 0$, which hold element-wise. The calibration parameters $\boldsymbol{\nu} = (\boldsymbol{\nu}_\star, \boldsymbol{\nu}_\oplus)$ are constrained to belong to a set $\Omega = \Omega_\star \times \Omega_\oplus$ where $\Omega_\star$ and $\Omega_\oplus$ are the respective feasible sets for the stellar and companion calibration parameters defined based on physical considerations.

The SEDs and the on-axis PSF at the reference wavelength being mutually independent, the regularization function can be decomposed as:

$$\mathcal{R}_{x\,y\,z}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}, \boldsymbol{\mu}) = \mu_x\,\mathcal{R}_x(\boldsymbol{x}) + \mu_y\,\mathcal{R}_y(\boldsymbol{y}) + \mu_z\,\mathcal{R}_z(\boldsymbol{z}). \tag{19}$$

The complete set of hyper-parameters is then:

$$\boldsymbol{\mu} = (\mu_x, \mu_y, \mu_z), \tag{20}$$

where $\mu_x > 0$, $\mu_y > 0$, and $\mu_z > 0$ tune the weights of the different regularization terms.

[7] The precision matrix is the inverse of the covariance matrix.

There are many regularizations that are suitable for our problem. Regularization terms should enforce some kind of continuity or smoothness of the sought uni-dimensional distributions. In the following (and for the sake of simplicity), we consider simple smoothness regularizations imposed by the quadratic penalty (Tikhonov & Arsenin 1977):

$$\mathcal{R}(\boldsymbol{u}) = \sum_{j=1}^{N_u-1} \underbrace{(u_{j+1} - u_j)}_{(\mathbf{D}\,\boldsymbol{u})_j}^2 = \|\mathbf{D}\,\boldsymbol{u}\|_2^2, \tag{21}$$

with N_u the size of $\boldsymbol{u} = \boldsymbol{x}$, $\boldsymbol{y}$, or $\boldsymbol{z}$, and $\mathbf{D} \in \mathbb{R}^{(N_u-1)\times N_u}$ a finite difference operator.

3.4. Alternating the minimization strategy

The joint minimization of the criterion defined in Eq. (17) requires us to cope with a highly non-linear function whose conditioning may be very bad and which depends on the scaling of the parameters. We propose to solve the problem by an alternated minimization strategy, that is, estimating each set of parameters given the others. Such a strategy consists in sequentially solving the following sub-problems:

$$\begin{aligned}
\hat{\boldsymbol{x}}(\boldsymbol{y}, \boldsymbol{r}_\star, \boldsymbol{\nu}_\star, \mu_x) &= \arg\min_{\boldsymbol{x}\geq 0} C(\boldsymbol{x}, \boldsymbol{y}, z, \boldsymbol{\nu}, \boldsymbol{\mu}) \\
&= \arg\min_{\boldsymbol{x}\geq 0} \left\{ \|\mathbf{A}_\star\,\boldsymbol{x} - \boldsymbol{r}_\star\|_{\mathbf{W}}^2 + \mu_x\,\mathcal{R}_x(\boldsymbol{x}) \right\}, && \text{(22a)} \\
\hat{\boldsymbol{y}}(\boldsymbol{x}, \boldsymbol{r}_\star, \boldsymbol{\nu}_\star, \mu_y) &= \arg\min_{\boldsymbol{y}\geq 0} C(\boldsymbol{x}, \boldsymbol{y}, z, \boldsymbol{\nu}, \boldsymbol{\mu}) \\
&= \arg\min_{\boldsymbol{y}\geq 0} \left\{ \|\mathbf{B}_\star\,\boldsymbol{y} - \boldsymbol{r}_\star\|_{\mathbf{W}}^2 + \mu_y\,\mathcal{R}_y(\boldsymbol{y}) \right\}, && \text{(22b)} \\
\hat{\boldsymbol{\nu}}_\star(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{r}_\star) &= \arg\min_{\boldsymbol{\nu}_\star\in\Omega_\star} C(\boldsymbol{x}, \boldsymbol{y}, z, \boldsymbol{\nu}, \boldsymbol{\mu}) \\
&= \arg\min_{\boldsymbol{\nu}_\star\in\Omega_\star} \|\boldsymbol{m}_\star(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{\nu}_\star) - \boldsymbol{r}_\star\|_{\mathbf{W}}^2, && \text{(22c)} \\
\hat{z}(\boldsymbol{r}_\oplus, \boldsymbol{\nu}_\oplus, \mu_z) &= \arg\min_{z\geq 0} C(\boldsymbol{x}, \boldsymbol{y}, z, \boldsymbol{\nu}, \boldsymbol{\mu}) \\
&= \arg\min_{z\geq 0} \left\{ \|\mathbf{A}_\oplus\,z - \boldsymbol{r}_\oplus\|_{\mathbf{W}}^2 + \mu_z\,\mathcal{R}_z(z) \right\}, && \text{(22d)} \\
\hat{\boldsymbol{\nu}}_\oplus(z, \boldsymbol{r}_\oplus) &= \arg\min_{\boldsymbol{\nu}_\oplus\in\Omega_\oplus} C(\boldsymbol{x}, \boldsymbol{y}, z, \boldsymbol{\nu}, \boldsymbol{\mu}) \\
&= \arg\min_{\boldsymbol{\nu}_\oplus\in\Omega_\oplus} \|\boldsymbol{m}_\oplus(z, \boldsymbol{\nu}_\oplus) - \boldsymbol{r}_\oplus\|_{\mathbf{W}}^2, && \text{(22e)}
\end{aligned}$$

with:

$$\begin{aligned}
\forall \boldsymbol{x},\ \mathbf{A}_\star\boldsymbol{x} &= \boldsymbol{\gamma} \odot (\mathbf{H}_\star(\boldsymbol{\nu})\,\boldsymbol{y}) \odot (\mathbf{F}_\star\boldsymbol{x}), && \text{(23a)} \\
\forall \boldsymbol{y},\ \mathbf{B}_\star\boldsymbol{y} &= \boldsymbol{\gamma} \odot (\mathbf{F}_\star\,\boldsymbol{x}) \odot (\mathbf{H}_\star(\boldsymbol{\nu})\boldsymbol{y}), && \text{(23b)} \\
\forall z,\ \mathbf{A}_\oplus z &= \boldsymbol{\gamma} \odot (\boldsymbol{h}_\oplus(\boldsymbol{\nu})) \odot (\mathbf{F}_\oplus z), && \text{(23c)} \\
\boldsymbol{r}_\star &= \boldsymbol{d} - \boldsymbol{m}_\oplus(z, \boldsymbol{\nu}_\oplus), && \text{(23d)} \\
\boldsymbol{r}_\oplus &= \boldsymbol{d} - \boldsymbol{m}_\star(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{\nu}_\star), && \text{(23e)} \\
\boldsymbol{m}_\star(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{\nu}_\star) &= \boldsymbol{\gamma} \odot (\mathbf{H}_\star(\boldsymbol{\nu})\,\boldsymbol{y}) \odot (\mathbf{F}_\star\,\boldsymbol{x}) = \mathbf{A}_\star\,\boldsymbol{x} = \mathbf{B}_\star\,\boldsymbol{y}, && \text{(23f)} \\
\boldsymbol{m}_\oplus(z, \boldsymbol{\nu}_\oplus) &= \boldsymbol{\gamma} \odot \boldsymbol{h}_\oplus(\boldsymbol{\nu}) \odot (\mathbf{F}_\oplus\,z) = \mathbf{A}_\oplus\,z. && \text{(23g)}
\end{aligned}$$

We enforced positivity constraints for the variables $\boldsymbol{x}$, $\boldsymbol{y}$, and z, while $\Omega_\star$ and $\Omega_\oplus$ (respectively) denote the feasible set of parameters $\boldsymbol{\nu}_\star$ and $\boldsymbol{\nu}_\oplus$. We note that $\boldsymbol{m}_\star(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{\nu}_\star)$ and $\boldsymbol{m}_\oplus(z, \boldsymbol{\nu}_\oplus)$ defined in Eqs. (23f) and (23g) are the respective contributions of the star and companion.

When the convex regularization defined in (21) is chosen and $\mathbf{A}_\star^\top\mathbf{W}\mathbf{A}_\star$, $\mathbf{B}_\star^\top\mathbf{W}\mathbf{B}_\star$, and $\mathbf{A}_\oplus^\top\mathbf{W}\mathbf{A}_\oplus$ are invertible, each of the problems (22a), (22b), and (22d) is strictly convex and thus has a unique solution that can be found by using existing algorithms[8]. This is another advantage of the alternated strategy. Since the original minimization problem (18) is not jointly convex with respect to all unknowns, only a local minimum is reached by the alternating minimization scheme, however.

Algorithm 1: FitStar – fit stellar parameters.

Input: $\boldsymbol{r}_\star \in \mathbb{R}^N$, $\mathbf{W} \in \mathbb{R}^{N\times N}$, $\boldsymbol{x}^{[0]} \in \mathbb{R}^{N_x}$, $\boldsymbol{\nu}_\star^{[0]}$, $\mu_x > 0$, and $\alpha_0 > 0$.
Output: $\hat{\boldsymbol{x}}$, $\hat{\boldsymbol{y}}$, and $\hat{\boldsymbol{\nu}}_\star$ a local minimum of C in $\boldsymbol{x}$, $\boldsymbol{y}$, and $\boldsymbol{\nu}_\star$.

```
k = 0;
μ_y = 1;
while not converged do
    while true do
        ▶ Update on-axis PSF;
        y^[k+1] = ŷ(x^[k], r_⋆, ν_⋆^[k], α_k^{-2} μ_y);          ◁ Eq. (22b)
        α_{k+1/2} = α̂(x^[k], y^[k+1], μ_x, μ_y);                ◁ Eq. (C.4)
        if k ≥ 1 or α_{k+1/2} ≈ α_k then break;
        α_k = α_{k+1/2};
    ▶ Update star SED;
    x^[k+1] = x̂(y^[k+1], r_⋆, ν_⋆^[k], α_{k+1/2}^2 μ_x);         ◁ Eq. (22a)
    α_{k+1} = α̂(x^[k+1], y^[k+1], μ_x, μ_y);                    ◁ Eq. (C.4)
    ▶ Auto-calibration (optional);
    ν_⋆^[k+1] ← ν̂_⋆(x^[k+1], y^[k+1], r_⋆);                     ◁ Eq. (22c)
    k ← k + 1;
x̂ ← α_k x^[k];
ŷ ← y^[k]/α_k;
ν̂_⋆ ← ν_⋆^[k];
```

We solve for the two stellar components $\boldsymbol{x}$ and $\boldsymbol{y}$ following the alternated method proposed by Thé et al. (2020) to exploit the scaling indetermination of this problem (see Appendix C for details). This method is implemented by Algorithm 1 and takes as inputs the residuals $\boldsymbol{r}_\star = \boldsymbol{d} - \boldsymbol{m}_\oplus(z, \boldsymbol{\nu}_\oplus)$ (i.e., the data without the contribution of the companion), the precision matrix, $\mathbf{W}$, initial calibration parameters, $\boldsymbol{\nu}_\star^{[0]}$, the hyper-parameters, $\mu_x > 0$ (hyper-parameter μ_y is set to the arbitrary value 1 in Algorithm 1), initial estimates, $\boldsymbol{x}^{[0]}$, of the stellar SED, and initial estimate, $\alpha_0 > 0$, of the scaling parameter. We offer some remarks on Algorithm 1 below.

First, the initial stellar SED $\boldsymbol{x}^{[0]}$ must be such that $\mathcal{R}_x(\boldsymbol{x}^{[0]}) > 0$ to be able to apply formula (C.4) to compute the optimal scaling factor (i.e., a non-flat SED). The initial stellar SED can be provided by calibration data (see Appendix B); otherwise, it can be computed from the science data, $\boldsymbol{d}$, by the following weighted mean:

$$\forall j \in [\![1, N_x]\!] : \quad x_{\star,j}^{[0]} = \frac{\sum_{n\in\mathcal{X}_j} w_n\,d_n}{\sum_{n\in\mathcal{X}_j} w_n}, \tag{24}$$

with $w_n = W_{n,n}$ the nth diagonal term of the precision matrix and:

$$\mathcal{X}_j = \left\{ n \in [\![1, N]\!] \;\middle|\; \left|\lambda_{\star,j}^{\mathrm{grd}} - \lambda_n\right| = \min_{j'\in[\![1,N_x]\!]} \left|\lambda_{\star,j'}^{\mathrm{grd}} - \lambda_n\right| \right\} \tag{25}$$

[8] For example, in the unconstrained case and with quadratic regularizations, the solution of one of these sub-problem has a closed-form expression. Otherwise, each of these sub-problems can be solved by optimization algorithms such as quasi-Newton methods with bound constraints (e.g., Thiébaut 2002).

Algorithm 2: FitCompanion – fit companion parameters.

Input: residuals $\boldsymbol{r}_\oplus \in \mathbb{R}^N$, precision matrix $\mathbf{W}$, initial off-axis PSF parameters $\boldsymbol{\nu}_\oplus^{[0]}$, $\mu_z > 0$.
Output: $\hat{\boldsymbol{z}}$ and $\hat{\boldsymbol{\nu}}_\oplus$, a local minimum of C in $\boldsymbol{z}$ and $\boldsymbol{\nu}_\oplus$ given $\boldsymbol{m}_\star(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{\nu}_\star)$ the model of the stellar contribution.

$k = 0$;
while *not converged* **do**
  ▶ Update companion SED;
  $\boldsymbol{z}^{[k+1]} = \hat{\boldsymbol{z}}(\boldsymbol{r}_\oplus, \boldsymbol{\nu}_\oplus^{[k]}, \mu_z)$; ◁ Eq. (22d)
  ▶ Update off-axis PSF (optional);
  $\boldsymbol{\nu}_\oplus^{[k+1]} = \hat{\boldsymbol{\nu}}_\oplus(\boldsymbol{z}^{[k+1]}, \boldsymbol{r}_\oplus)$; ◁ Eq. (22e)
  $k \leftarrow k + 1$;
$\hat{\boldsymbol{z}} \leftarrow \boldsymbol{z}^{[k]}$;
$\hat{\boldsymbol{\nu}}_\oplus \leftarrow \boldsymbol{\nu}_\oplus^{[k]}$;

the set of pixels whose nearest wavelength in the model grid is the jth one. Since Algorithm 1 scales the final components $\boldsymbol{x}^{[k]}$ and $\boldsymbol{y}^{[k]}$ by the corresponding optimal scaling factor, $\alpha_0 = 1$ is a natural choice for the initial scaling factor in subsequent calls to Algorithm 1 (to refine the solution or after having improved the other parameters).

Second, the inner loop of Algorithm 1 avoids sensitivity to the initial scaling of the parameters (Thé et al. 2020). Then, the convergence criterion of Algorithm 1 is left unspecified. In our implementation, we chose to stop the algorithm when the relative change, in norm, between two consecutive iterates is smaller than a parameter ϵ ($= 10^{-3}$).

Although they represent very different physical quantities, the problem is quite symmetric in variables $\boldsymbol{x}$ and $\boldsymbol{y}$. Thus a variant of Algorithm 1 can be easily implemented to start with an initial estimate $\boldsymbol{y}^{[0]}$ of the stellar on-axis PSF at the reference wavelength instead of an initial estimate $\boldsymbol{x}^{[0]}$ of the stellar SED. For the very first run, this variant of Algorithm 1 is started with the weighted average of the on-axis PSF defined by:

$$\forall j \in [\![1, N_{\boldsymbol{y}}]\!] : \quad y_{\star,j}^{[0]} = \frac{\sum_{n \in \mathcal{Y}_j} w_n \, d_n}{\sum_{n \in \mathcal{Y}_j} w_n}, \tag{26}$$

with:

$$\mathcal{Y}_j = \left\{ n \in [\![1, N]\!] \;\middle|\; \left|\rho_{\star,j}^{\mathrm{grd}} - \rho_n\right| = \min_{j' \in [\![1, N_{\boldsymbol{y}}]\!]} \left|\rho_{\star,j'}^{\mathrm{grd}} - \rho_n\right| \right\}, \tag{27}$$

the set of pixels whose nearest angular position in the model grid is the jth one.

Finally, when the SED $\boldsymbol{z}$ of the companion is not yet known, it is sufficient to refer to Algorithm 1 with the weights of the pixels that are the most impacted by the companion set to zero (we denote the corresponding precision matrix as $\mathbf{W}_\star$) to estimate the components $\boldsymbol{x}$ and $\boldsymbol{y}$ of the stellar leakages without introducing a significant bias due to the contribution of the companion.

Algorithm 2 (FitCompanion) implements an alternated strategy to estimate the parameters $\boldsymbol{z}$ and $\boldsymbol{\nu}$ of the companion SED and its off-axis PSF at the reference wavelength. It takes as inputs the residuals $\boldsymbol{r}_\oplus = \boldsymbol{d} - \boldsymbol{m}_\star(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{\nu}_\star)$ (i.e., the data without the contribution of the star) and their respective weights $\mathbf{W}$, the hyper-parameter $\mu_z > 0$ and an initial estimate $\boldsymbol{\nu}_\oplus^{[0]} \in \Omega_\oplus$ of the parameters of the off-axis PSF at the reference wavelength.

Algorithm 3: ExoSpeco algorithm

Input: The data, $\boldsymbol{d}$, and its precision matrix, $\mathbf{W}$, the masked precision matrix, $\mathbf{W}_\star$, initial estimates, $\boldsymbol{x}^{[0]}$, $\boldsymbol{\nu}_\star^{[0]}$, and $\boldsymbol{\nu}_\oplus^{[0]}$, and hyper-parameters, $\mu_{\boldsymbol{x}} > 0$ and $\mu_z > 0$.
Output: $\hat{\boldsymbol{x}}$, $\hat{\boldsymbol{y}}$, $\hat{\boldsymbol{z}}$, $\hat{\boldsymbol{\nu}}_\star$, and $\hat{\boldsymbol{\nu}}_\oplus$ a local minimum of C.

$\mu_{\boldsymbol{y}} = 1$;
$\boldsymbol{z}^{[0]} = \mathbf{0}$;
$\hat{\alpha} = 1$;
$k = 0$;
while *not converged* **do**
  if $k = 0$ **then**
    ▶ Hide companion;
    $\mathbf{W}' = \mathbf{W}_\star$; ◁ Eq. (28)
  else
    ▶ Account for companion;
    $\mathbf{W}' = \mathbf{W}$;
  ▶ Update star leakage model;
  $\boldsymbol{r}_\star^{[k]} = \boldsymbol{d} - \boldsymbol{m}_\oplus(\boldsymbol{z}^{[k]}, \boldsymbol{\nu}_\oplus^{[k]})$; ◁ Eq. (23d)
  $(\boldsymbol{x}^{[k+1]}, \boldsymbol{y}^{[k+1]}, \boldsymbol{\nu}_\star^{[k+1]}) =$;
    FitStar$(\boldsymbol{r}_\star^{[k]}, \mathbf{W}', \boldsymbol{x}^{[k]}, \boldsymbol{\nu}_\star^{[k]}, \mu_{\boldsymbol{x}}, \hat{\alpha})$;
  ▶ Update companion model;
  $\boldsymbol{r}_\oplus^{[k+1]} = \boldsymbol{d} - \boldsymbol{m}_\star(\boldsymbol{x}^{[k+1]}, \boldsymbol{y}^{[k+1]}, \boldsymbol{\nu}_\star^{[k+1]})$; ◁ Eq. (23e)
  $(\boldsymbol{z}^{[k+1]}, \boldsymbol{\nu}_\oplus^{[k+1]}) =$ FitCompanion$(\boldsymbol{r}_\oplus^{[k+1]}, \mathbf{W}, \boldsymbol{\nu}_\oplus^{[k]}, \mu_z)$;
  $k \leftarrow k + 1$;
$(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}, \hat{\boldsymbol{z}}, \hat{\boldsymbol{\nu}}_\star, \hat{\boldsymbol{\nu}}_\oplus) \leftarrow (\boldsymbol{x}^{[k]}, \boldsymbol{y}^{[k]}, \boldsymbol{z}^{[k]}, \boldsymbol{\nu}_\star^{[k]}, \boldsymbol{\nu}_\oplus^{[k]})$;

These latter parameters can be given by the calibration described in Appendix B. Algorithm 2 also merits some remarks, which are given below.

First, the outputs of the algorithm only depend on the residual data $\boldsymbol{r}_\oplus$ defined in Eq. (23e) that need to be computed only once (on entry of the algorithm and not at each iterations).

Second, as in the case of Algorithm 1, various stopping criteria may be implemented to break the loop. Then, as with Algorithm 1, we can use the VMLM-B algorithm (Thiébaut 2002) to solve problem (22d) to estimate $\boldsymbol{z}$ under a non-negativity constraint.

In both algorithms, there are optional self-calibration steps performed by solving problem (22c) in Algorithm 1 (FitStar) and problem (22e) in Algorithm 2 (FitCompanion) to estimate the parameters of the on-axis and off-axis PSFs. These minimizations can be carried out by a derivative-free minimization algorithm. When there is a single calibration parameter, we use the Brent (2013) Fmin algorithm; if there are several parameters, we use one of Powell's derivative-free methods Newuoa or Bobyqa (Powell 2006, 2009) depending on the constraints defined by Ω.

3.5. The ExoSpeco algorithm

FitStar (Algorithm 1) and FitCompanion (Algorithm 2) are the building blocks of the ExoSpeco method given in Algorithm 3 for estimating all unknowns. A few additional remarks are given below.

First, for the first estimation of the stellar leakage parameters, it is beneficial to define a masked version, $\mathbf{W}_\star$, of the precision matrix of the data to avoid a significant bias of the first estimates due to the signal from the companion, which would slow down

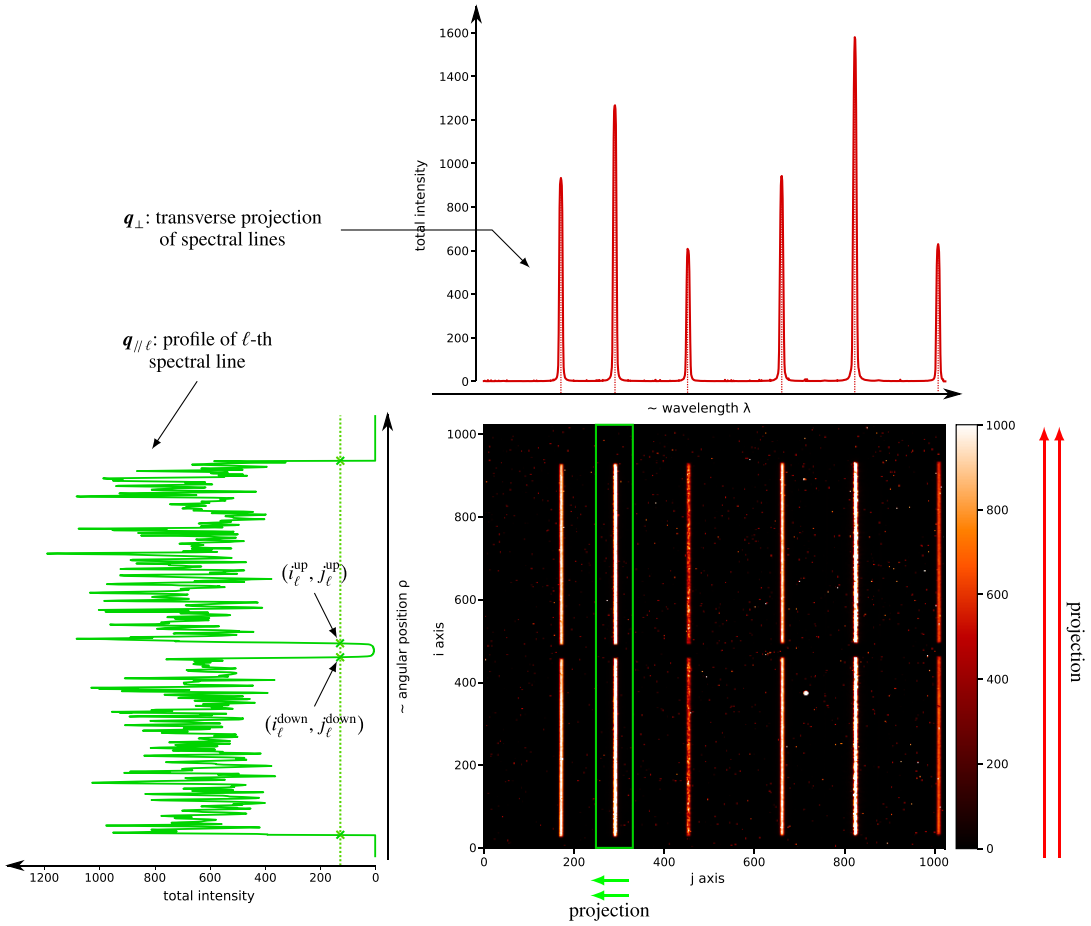

Fig. 5. Calibration data for the SPHERE/IRDIS instrument and for the observations of HR 3549 on 2015/12/28. Central panel: calibration image. Left and top panels: projections of the calibration data along the 2nd spectral line (in green) and across all spectral lines (in red).

the convergence of Algorithm 3. The masked precision matrix is simply given by:

$$\mathbf{W}_\star = \mathrm{diag}(\boldsymbol{w}_\star), \tag{28}$$

where the weights, $\boldsymbol{w}_\star$, are those of the precision matrix, $\mathbf{W}$, of the data except that they are set to zero for the pixels that are the most impacted by the companion:

$$\forall n \in [\![1, N]\!] : \quad w_{\star,n} = \begin{cases} 0 & \text{if } \gamma_n \, |\rho_n - \rho_\oplus| \leq \tau \\ w_n & \text{otherwise} \end{cases} \tag{29}$$

with $\tau > 0$ the angular half-width at the reference wavelength of the impacted region. In practice, τ is taken to be 2–3 times $\sigma_\oplus^{[0]}$ the initial angular standard deviation of the off-axis PSF at the reference wavelength.

Then, the model of the stellar leakage only depends on either μ_x or μ_y, the other being arbitrarily chosen. For this reason, Algorithm 3 takes as inputs only two hyper-parameters, μ_x and μ_z, the remaining hyper-parameter being set to $\mu_y = 1$.

After extracting the companion's spectrum by the ExoSPECO Algorithm 3, it is possible to express it as a contrast relative to the host star that can be multiplied by a reference spectrum of the star to get rid of the atmospheric absorption (see Appendix B).

The auto-calibration steps in FitSTAR (Algorithm 1) and FitCOMPANION (Algorithm 2) are optional and consist in the resolution of problems (22c) and (22e). As these problems are non-convex, activating the auto-calibration at the beginning of the method can lead to a local minimum. To avoid such a behavior, it is possible to start the self-calibration of $\boldsymbol{\nu}_\star$ and $\boldsymbol{\nu}_\oplus$ only after a few iterations of ExoSPECO (Algorithm 3).

Controlling the number of inner iterations to solve each subproblem could be done by changing the value of the stopping parameter ϵ (cf. remark 3 in Sect. 3.4): the smaller ϵ the more inner iterations are needed and conversely. But this is expected to also impact the number of outer iterations. Owing to the modest amount of time (2–3 min) taken by our implementation of ExoSPECO to solve the entire problem, we did not investigate whether the algorithm can be effectively accelerated by changing ϵ and keep the value $\epsilon = 10^{-3}$ suggested previously.

4. Calibration

The direct model in Eq. (15) assumes the physical coordinates (ρ_n, λ_n) of each pixel n of the detector are known. In this section, we offer a consistent approach to derive the spectro-angular dispersion laws of the instrument from calibration data.

4.1. Calibration data

Calibration data takes the form of an image (e.g., as shown in Fig. 5) obtained by illuminating the spectrograph slit with N_λ laser sources[9]. This produces N_λ mono-chromatic lines on the

[9] $N_\lambda = 6$ at wavelengths 0.9877 µm, 1.1237 µm, 1.3094 µm, 1.5451 µm, 1.73 µm, and 2.015 µm for SPHERE/IRDIS.

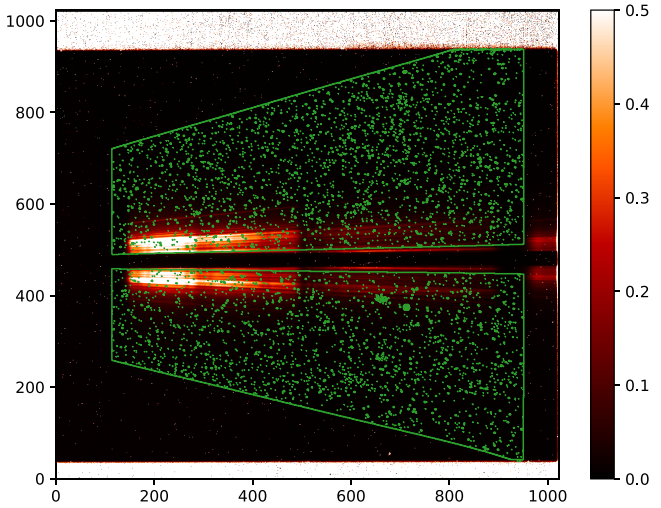

Fig. 6. Valid pixel mask for the HR 3549 data observed on 2015-12-28 in MRS mode.

detector, each being interrupted by the coronagraphic mask. The calibration image $\boldsymbol{d}_{\rm cal}$ is of size $I \times J$ and, for the calibration procedure, we denote it using $n \sim (i, j)$ the one to one mapping between the pixel number, n, and its indices $i \in [\![1, I]\!]$ and $j \in [\![1, J]\!]$ along the first and second dimensions of the detector.

The calibration image shall have been pre-processed to compensate for bias and non-uniform response of the detector. Furthermore, we assume a given mask of valid pixels:

$$w_{{\rm msk},i,j} = \begin{cases} 1 & \text{if the pixel } (i, j) \text{ is valid,} \\ 0 & \text{else.} \end{cases} \tag{30}$$

We consider a pixel as being invalid if its value cannot follow the assumed direct model given in Eq. (15). Invalid pixels include pixels outside the field of view, pixels under the coronagraphic mask or close to this mask, and defective pixels whose level does not linearly depend on the illumination. Figure 6 shows the mask of valid pixels for the HR 3549 data: the field of view and the coronagraphic mask are outlined by the two green trapezes, while the defective pixels are marked by green dots.

4.2. Dispersion laws

There are two dispersion laws to calibrate: $\Lambda(i, j)$ for the wavelength and $\varrho(i, j)$ for the separation angle along the slit. To determine the best approximation of these spectro-angular dispersion laws, we compared three models:

First, we have the "standard model" which assumes that the dispersion laws are uni-dimensional polynomials with spectral and angular directions aligned with the detector axes:

$$\Lambda_{\rm sta}(i, j) = \sum_{p=0}^{P_\lambda} a_p \, j^p, \tag{31a}$$

$$\varrho_{\rm sta}(i, j) = \sum_{p=0}^{P_\rho} s_p \, i^p, \tag{31b}$$

with P_λ and P_ρ the degrees of the polynomials and $\{a_p\}_{p\in[\![0,P_\lambda]\!]}$ and $\{s_p\}_{p\in[\![0,P_\rho]\!]}$ their coefficients. For $P_\rho = 1$ and $P_\lambda = 3$–5, the "standard model" reproduces what is done in the software by Vigan et al. (2008) usually used to process SPHERE/LSS data.

Next, the model of medium complexity, also assuming 1D polynomials for the dispersion laws but accounting for misalignment angles ϕ_λ and $\phi_\rho \approx \phi_\lambda + 90°$, respectively, between the spectral and angular directions and the detector axes:

$$\Lambda_{\rm med}(i, j) = \sum_{p=0}^{P_\lambda} a_p \, (i \, \sin\phi_\lambda + j \, \cos\phi_\lambda)^p, \tag{32a}$$

$$\varrho_{\rm med}(i, j) = \sum_{p=0}^{P_\rho} s_p \, (i \, \sin\phi_\rho + j \, \cos\phi_\rho)^p. \tag{32b}$$

We note that taking $\phi_\lambda = 0°$ and $\phi_\varrho = 90°$ yields the standard model.

Finally, a more complex model, which assumes 2D polynomials for the dispersion laws and, depending on the degree of these polynomials, can account for more complex image distortions than a simple rotation:

$$\Lambda_{\rm EXOSPECO}(i, j) = \sum_{p_1=0}^{P_\lambda} \sum_{p_2=0}^{P_\lambda - p_1} a_{p_1,p_2} \, i^{p_1} \, j^{p_2}, \tag{33a}$$

$$\varrho_{\rm EXOSPECO}(i, j) = \sum_{p_1=0}^{P_\rho} \sum_{p_2=0}^{P_\rho - p_1} s_{p_1,p_2} \, i^{p_1} \, j^{p_2}. \tag{33b}$$

To summarize, the considered dispersion laws are polynomials of respective degree, P_λ and P_ρ. Their calibration amounts to fitting their coefficients $\boldsymbol{a}$ and $\boldsymbol{s}$ given the calibration image, $\boldsymbol{d}_{\rm cal}$, as explained in the next sub-sections.

4.3. Calibration of the spectral dispersion law Λ

To calibrate the spectral dispersion law Λ, we extract from the calibration image $\boldsymbol{d}_{\rm cal}$ (see Fig. 5) the N_λ lists of pixel coordinates following the path of each spectral line on the detector and estimate the coefficients $\boldsymbol{a}$ by a least-squares fit:

$$\hat{\boldsymbol{a}} = \arg\min_{\boldsymbol{a}} \sum_{\ell=1}^{N_\lambda} \sum_{(i,j)\in C_\ell(\phi_\lambda)} (\lambda_\ell - \Lambda(i, j))^2, \tag{34}$$

where $C_\ell(\phi_\lambda)$ denotes the list of, possibly fractional, pixel coordinates (i, j) along the ℓth spectral line on the detector. Since $\Lambda(i, j)$ linearly depends on the coefficients $\boldsymbol{a}$, then the solution $\hat{\boldsymbol{a}}$ of the above problem has a closed-form expression (Lawson & Hanson 1974) that is easy to compute.

To extract the paths $C_\ell(\phi_\lambda)$ of the spectral lines, we first compute a transverse projection $\boldsymbol{q}_\perp(\phi_\lambda)$ of the calibration image $\boldsymbol{d}_{\rm cal}$ tuning the projection angle ϕ_λ, so as to maximize the peak values in the resulting projection. This transverse projection is plotted in red in the top panel of Fig. 5 and corresponds to $\phi_\lambda \approx 0°$ for the considered calibration data. Equations (A.1) formally define how we carefully compute the projection avoiding invalid pixels. We then use the procedure described in Appendix A.2 to locate the position of the N_λ most significant peaks in the transverse projection $\boldsymbol{q}_\perp(\phi_\lambda)$, which can be seen as a mean cross-section of the spectral lines. Finally, we use the method described in Appendix A.3 to extract the coordinates of the points defining the N_λ paths $C_\ell(\phi_\lambda)$. These coordinates are given by the centers of gravity (again accounting for invalid pixels thanks to the mask) of the calibration data in small sliding rectangular windows along each spectral lines (see Appendix A.3 for details).

4.4. Calibration of the angular dispersion law ϱ

To calibrate the angular dispersion law ϱ, we extract from the calibration image $\boldsymbol{d}_{\rm cal}$ (see Fig. 5), the positions of the edges of the coronagraphic mask for each of the N_λ spectral lines and

estimate the coefficients $\boldsymbol{s}$ of the polynomial and the width $\Delta\rho$ of the mask by a least-squares fit:

$$(\hat{\phi}_\rho, \hat{\Delta\rho}, \hat{\boldsymbol{s}}) = \underset{\phi_\rho, \Delta\rho, \boldsymbol{s}}{\arg\min} \left\{ \sum_{\ell=1}^{N_\lambda} \left(\varrho(i_\ell^{\text{down}}, j_\ell^{\text{down}}) + \Delta\rho/2 \right)^2 + \sum_{\ell=1}^{N_\lambda} \left(\varrho(i_\ell^{\text{up}}, j_\ell^{\text{up}}) - \Delta\rho/2 \right)^2 \right\} \tag{35}$$

where $(i_\ell^{\text{down}}, j_\ell^{\text{down}})$ and $(i_\ell^{\text{up}}, j_\ell^{\text{up}})$ denote the coordinates of the edges of the coronagraphic mask respectively on the downhill and uphill sides along the profile of the ℓth spectral line. To solve this problem, we exploit that the criterion is quadratic in the unknowns $\boldsymbol{s}$ and $\Delta\rho$, which thus have a closed-form solution (Lawson & Hanson 1974) that depends on ϕ_ρ. Replacing this closed-form solution in the criterion yields an uni-variate objective function that only depends on ϕ_ρ and which we minimize by Brent (2013) FMIN method starting at $\phi_\rho = \phi_\lambda + 90°$.

As explained in Appendix A.4, the coordinates of the edges of the coronagraphic mask for the ℓth spectral line are obtained from the longitudinal profile $\boldsymbol{q}_{/\!/\,\ell}(\phi_\lambda)$ of the line which is the weighted projection, in a direction perpendicular to the transversal projection $\boldsymbol{q}_\perp(\phi_\lambda)$, of the calibration data in a window encompassing the line. The longitudinal profile $\boldsymbol{q}_{/\!/\,\ell}(\phi_\lambda)$ for the second ($\ell = 2$) line is plotted in green in the left panel of Fig. 5 and the corresponding window is outlined in green in the central panel of Fig. 5.

4.5. Comparison of the calibration models

To compare the calibration models considered in Sect. 4.2, we applied EXOSPECO (Algorithm 3) on a scientific dataset of the star HR 3549 observed in MRS mode of SPHERE/IRDIS on 2015/12/28. Figure 7 shows the residuals $\boldsymbol{r} = \boldsymbol{d} - \boldsymbol{m}$, that is the difference between the data and their model, computed for the different spectro-angular dispersion laws. This figure shows that the root mean square (rms) of the residuals are significantly reduced when using more flexible calibration models than the standard one. The improvement brought by the medium complexity model compared to the standard model proves that accounting for a slight angular misalignment between the spatial and spectral directions and the detector axes is important. Compared to the medium model, the complex model is able to account other image distortions than a simple rotation and thus achieves a better suppression of the stellar leakages. These results motivate the choice of the complex dispersion model in EXOSPECO to reduce the rms level of the residual stellar leakages by a factor of ∼2 compared to the standard model and should therefore result in a better extraction of the companion contribution. We recall that the simple standard calibration model is similar to what is usually used for these data in other studies.

5. Validation and tuning of the method

To fully validate the method, we propose a study of the method on both real data and data where a synthetic companion was injected. This study allows us to both evaluate the modeling of the stellar leakages (and, thus, its subtraction in the residuals) and the extraction of the companion via a comparison with a ground truth spectrum. This study is described below.

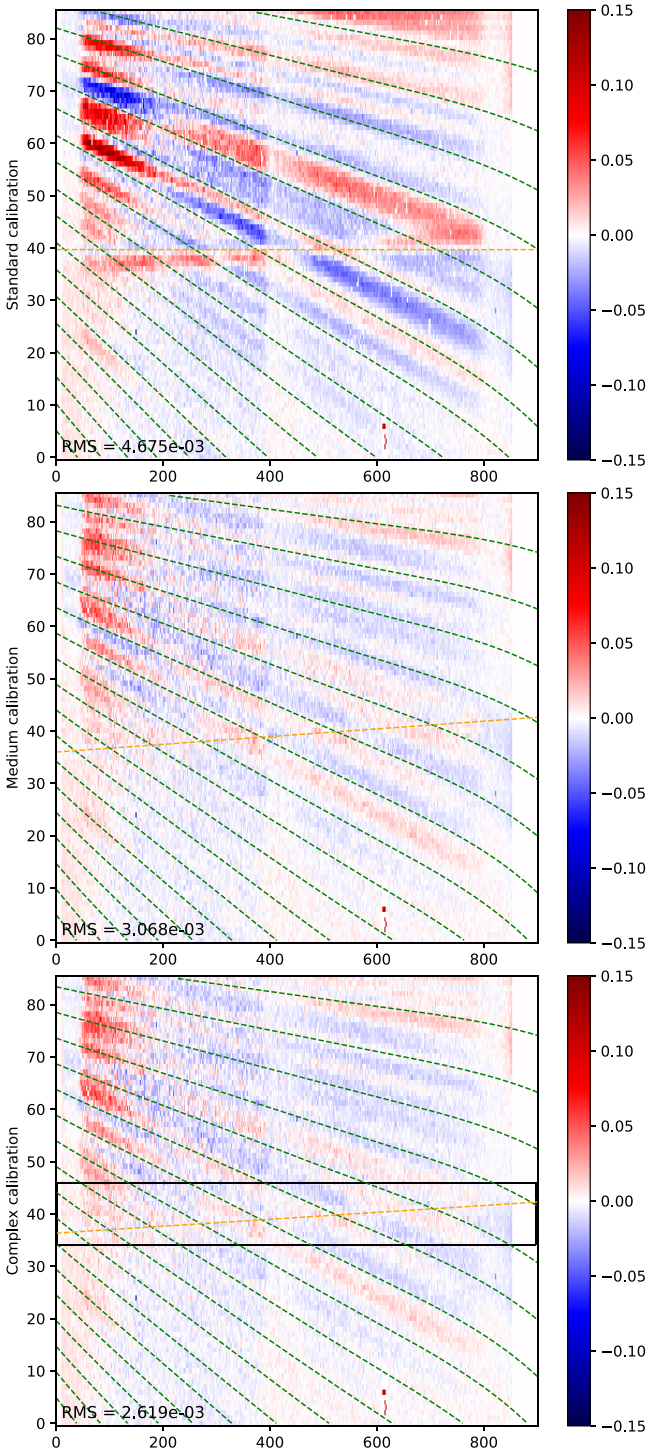

Fig. 7. Residuals between the HR 3549 scientific data and the model of the stellar leakages, assuming a standard (top figure), medium (center), and complex (bottom) calibration models. The rms values of the residuals are given for each model. The iso-levels of ρ/λ which are approximately followed by the dispersed stellar speckles are plotted as green dashed lines. The position of the companion at the different wavelengths (i.e., at $\rho = \rho_\oplus$) is plotted as an orange dashed line.

5.1. Reduction of the self-subtraction

One of the main features of EXOSPECO is that it jointly estimates the contributions of the star and of the companion whose parameters are iteratively refined until convergence. Figure 8 shows the residuals close to the companion (in the region outlined by the

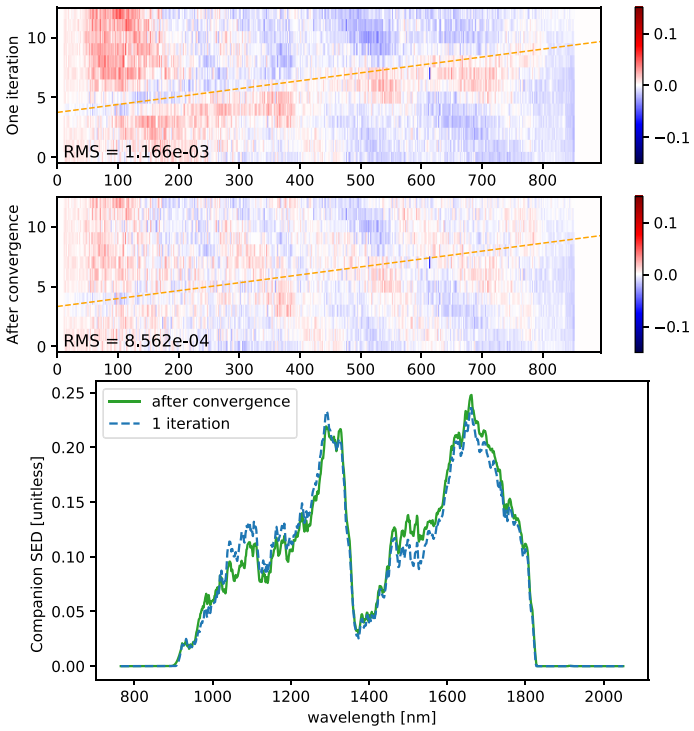

Fig. 8. Residuals between the HR 3549 scientific data and the model of the stellar leakages near the companion (defined by the black rectangle in the bottom of Fig. 7) and for the complex model of the spatial and spectral dispersion laws after one iteration (top) and after convergence (middle) of EXOSPECO. The rms of the residuals in this region are significantly reduced after convergence. The orange dashed line indicates the position of the companion at the different wavelengths. The SEDs of the companion (including atmospheric absorption) extracted from these residuals are plotted in the bottom-most panel.

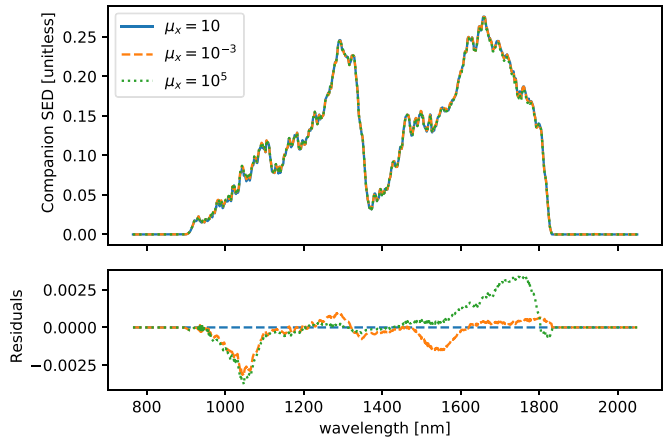

Fig. 9. Top: profiles of the companion SED z, for different levels of the stellar hyper-parameter μ_x and with $\mu_y = 1$ and $\mu_z = 10^5$. Bottom: differences between the profile for $\mu_x = 10$ and the profiles for $\mu_x = 10^{-3}$ (dashed orange) and $\mu_x = 10^5$ (dotted green).

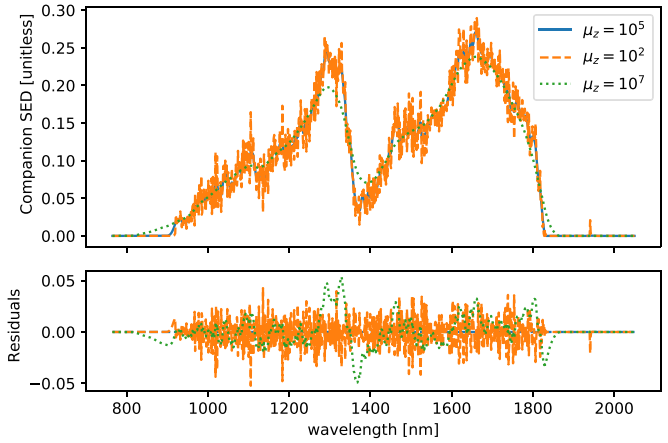

Fig. 10. Profiles of the companion SED z for different levels of the companion regularization ($\mu_z = 10^2$ in dashed orange, $\mu_z = 10^5$ in blue, and $\mu_z = 10^7$ in dotted green) in the top panel. Differences between the profile for $\mu_z = 10^5$ and the profiles for $\mu_z = 10^2$ (dashed orange) and $\mu_z = 10^7$ (dotted green) at the bottom. For all these results, the stellar hyper-parameters are $\mu_x = 10$ and $\mu_y = 1$.

black rectangle in the bottom of Fig. 7) with the complex model of the spatial and spectral dispersion laws in two cases: after the first outer iteration of EXOSPECO and after convergence of the algorithm. In the first outer iteration of EXOSPECO, the model of the stellar leakages is estimated by masking the region most impacted by the companion, as in Eq. (28), similarly to what is done by conventional methods. In all other outer iterations of EXOSPECO, the contribution of the other component is taken into account when fitting a given component (star or companion). As shown by the bottom panel of Fig. 8, there is a noticeable bias in the estimated companion's SED after the first outer iteration. This so-called "self-subtraction bias" is mostly avoided by adopting the proposed alternating strategy.

5.2. Tuning of the regularization parameters

As described in Appendix C, the fact that the model of the star leakages is bi-linear makes it possible to tune the regularization of this component by a single hyper-parameter, that is, the other hyper-parameter being held fixed. Thanks to this, the solution found by EXOSPECO only depends on two hyper-parameters: one for the star, say, μ_x (while $\mu_y = 1$ is imposed) and one for the companion, μ_z. In this section, we highlight the incidence on the companion SED extracted by EXOSPECO of these remaining hyper-parameters using the same scientific data set as in Sect. 4.5.

Figure 9 shows the SEDs of the companion estimated by EXOSPECO for different values of the star regularization hyper-parameters ($\mu_y = 1$ and $\mu_x = 10^{-3}$, 10, and 10^5). For such a broad range of values, the differences between the extracted companion SEDs are smaller than 1%. The stellar regularization hyper-parameters have thus a limited impact on the resulting companion SED. The tuning of μ_x can thus reasonably be done by visual inspection.

On the contrary, as Fig. 10 shows, the hyper-parameter μ_z has a strong impact on the resulting companion SED. This is expected as μ_z directly tunes the strength of the smoothness constraint for the companion SED z. This hyper-parameter has thus to be carefully chosen to find the best compromise between a solution that is too smooth (e.g., for $\mu_z = 10^7$ in Fig. 10) or too noisy (e.g., for $\mu_z = 10^2$ in Fig. 10). It is worth noticing that the correct value of μ_z strongly depends on the considered data, so $\mu_z = 10^5$, which seems to be a good choice for the HR 3549 data (see Fig. 10), should not be considered as a universal value.

Many methods have been proposed to automatically tune the hyper-parameter(s) of an inverse problem: a generalized cross-validation (CGV; Golubet al. 1979), Stein's unbiased risk estimate (SURE; Stein 1981), hierarchical Bayesian method (Molina 1994), or the L-curve (Hansen & O'Leary 1993) to

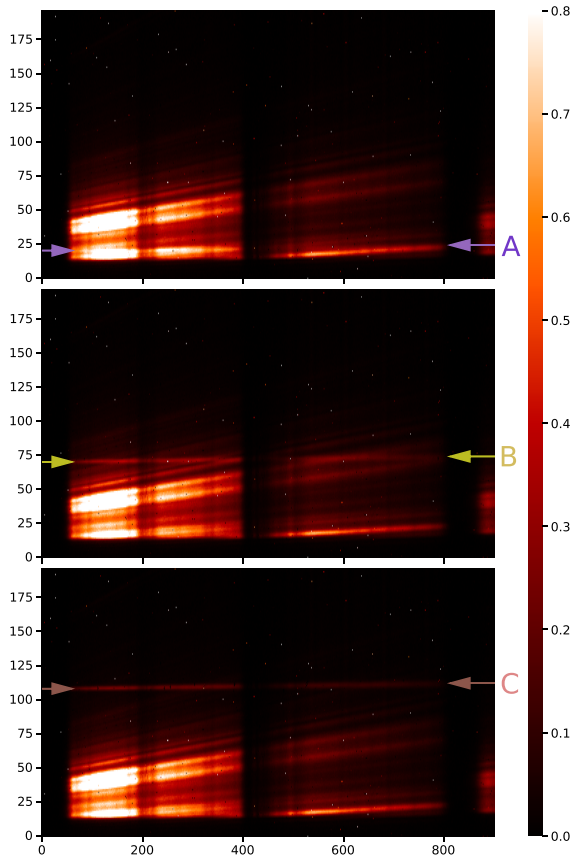

Fig. 11. Scientific data of HIP 65426 with a synthetic companion whose contrast is $\chi = 2 \times 10^{-4}$ relative to the star and injected at angular separations $\rho_\oplus - \rho_\star = 273$ mas (A), 890 mas (B), and 1353 mas (C) indicated by the arrows.

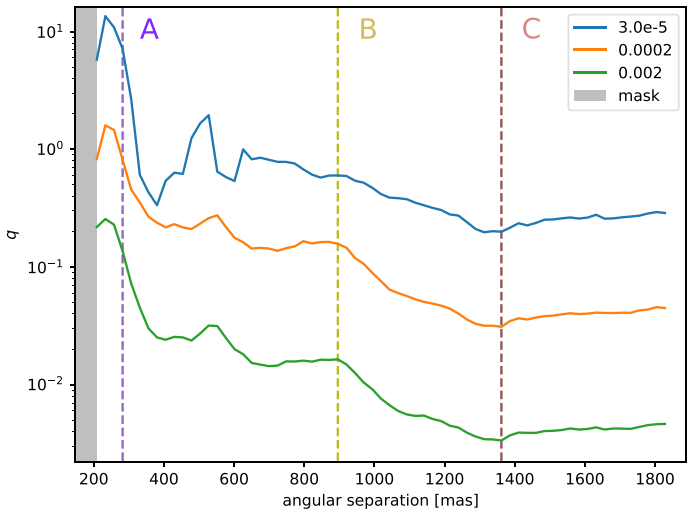

Fig. 12. Relative error q defined in Eq. (36) for synthetic companions injected in the scientific data of HIP 65426 (with the same spectra as the host star) as a function of the angular separation, $\rho_\oplus - \rho_\star$, and for contrasts of $\chi = 3 \times 10^{-5}$ (blue), 2×10^{-4} (orange), and 2×10^{-2} (green). The grayed area represents the region invalidated by the coronagraphic mask. The angular separations of the three cases presented in Fig. 11 are highlighted by the dashed lines labeled A, B, and C.

mention a few that could potentially be used with our extraction algorithm. Implementing and testing these methods for EXOSPECO is beyond the scope of this paper. However, since the companion SED found by EXOSPECO does not strongly depend on the tuning of the stellar regularization, our method is mostly driven by a single hyper-parameter, μ_z, the level of the regularization for the companion SED. This greatly reduces the complexity of tuning the EXOSPECO algorithm.

5.3. Extraction of simulated spectrum in real data

To validate the EXOSPECO method, we injected the contribution of a synthetic companion in existing SPHERE/IRDIS MRS data $\boldsymbol{d}$ of the star HIP 65426 observed on 2019-05-20. Although HIP 65426 star hosts a planet (Chauvin et al. 2017; Carter et al. 2023), the frame was selected for the de-rotation angles hiding the planet outside the slit. The off-axis PSF $h_\oplus$ of the synthetic companion follows the model in Eqs. (13) and (14c) with $\sigma_\oplus$ set to match the diffraction limit of the telescope at the reference wavelength λ^{ref} and with different angular positions $\rho_\oplus$ on the side of the coronagraphic mask where no companion was detected. The "ground truth" SED of the synthetic companion is $\boldsymbol{z}_{\text{gt}} = \chi \, \boldsymbol{x}_{\text{flux}}$, where $\chi > 0$ is the mean contrast of the companion relative to the star (without a coronagraph) and $\boldsymbol{x}_{\text{flux}}$ is the SED of the star HIP 65426 calibrated as explained in Appendix B. We used a constant contrast for all wavelengths (i.e., the SED of the star and of the companion are the same, up to the contrast χ). Figure 11 shows examples of generated data with a synthetic companion whose contrast with respect to the star is $\chi = 2 \times 10^{-4}$ and which is injected at different angular separations, $\rho_\oplus - -\rho_\star$.

To assess the quality of the extracted companion's SED $\hat{\boldsymbol{z}}$, we compute the following relative error:

$$q = \frac{\sum_{j=1}^{N_z} \left| z_{\text{gt},j} - \hat{z}_j \right|}{\sum_{j=1}^{N_z} \left| z_{\text{gt},j} \right|}. \tag{36}$$

In the following tests, the value of μ_z, the regularization level of the companion's SED, has been tuned so as to minimize the relative error q. Figure 12 plots the relative error q for synthetic companions injected at angular separations $\rho_\oplus - \rho_\star$ ranging from 200 mas to 1850 mas and with contrasts $\chi = 3 \times 10^{-5}$, 2×10^{-4}, and 2×10^{-3}. Clearly, the quality of the recovered SEDs degrades as the companion gets closer to the mask. This is expected because, when getting closer to the mask, not only are the stellar leakages brighter (thus causing more photon noise in the residuals) but the approximation by the assumed off-axis PSF model also worsens. For angular separations larger than ~600 mas and for all considered contrasts, the quality of the recovered SEDs improves as the separation increases until a plateau is reached at $\rho_\oplus - \rho_\star \sim 1400$ mas, where the dominant source of nuisance is the readout noise.

Figure 13 shows examples of recovered companion SEDs $\hat{\boldsymbol{z}}$ at angular separations $\rho_\oplus - \rho_\star = 273$ mas (A), 890 mas (B), and 1353 mas (C) for the same contrasts χ as in Fig. 12. Figure 13 confirms that the relative error q does reflect the ability of our method to reliably recover the companion SED. When $q \leq 0.1$ (the green curves for angular separations B and C and the orange curve for angular separation C), all the features of the SED are correctly recovered. For $q \sim 0.2$ (the green curve for case A, the orange curve for case B, and the blue curve for case C), the global shape of the SED is restored but with small spectral features smoothed out and some photometric biases. These cases prove that it is possible to extract a coarse but still exploitable SED for bright companions quite close to the mask, typically $\chi \geq 10^{-3}$ for $\rho_\oplus - \rho_\star \sim 250$ mas, from a single MRS exposure.

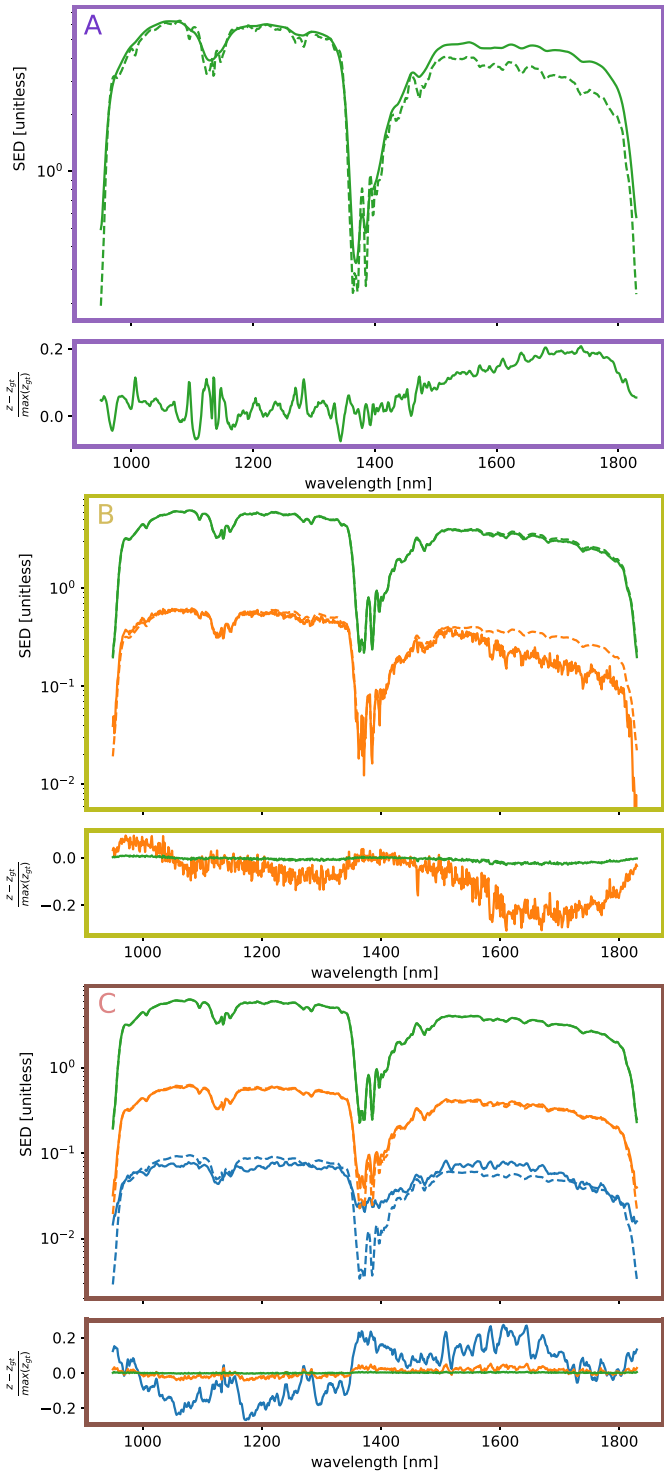

Fig. 13. Examples of recovered companion SEDs $\hat{z}$ in the same conditions as in Figs. 11 and 12 for synthetic companions injected at angular separations $\rho_\oplus - \rho_\star = 273$ mas (A), 890 mas (B), and 1353 mas (C) with contrasts of $\chi = 3 \times 10^{-5}$ (blue curves), 2×10^{-4} (orange curves), and 2×10^{-2} (green curves). The ground truth SED $z_{\rm gt}$ is plotted in dashed lines. The normalized residuals are plotted below each panel.

The angular separation must be larger for fainter companions; for example, $\rho_\oplus - \rho_\star \geq 1200$ mas for $\chi \sim 2 \times 10^{-5}$. The photometric biases in the most difficult cases (the green curve in case A, the orange one in case B, and the blue one in case C) clearly indicates that the removal of the modeled stellar contribution leaves non-negligible residuals compared to the companion.

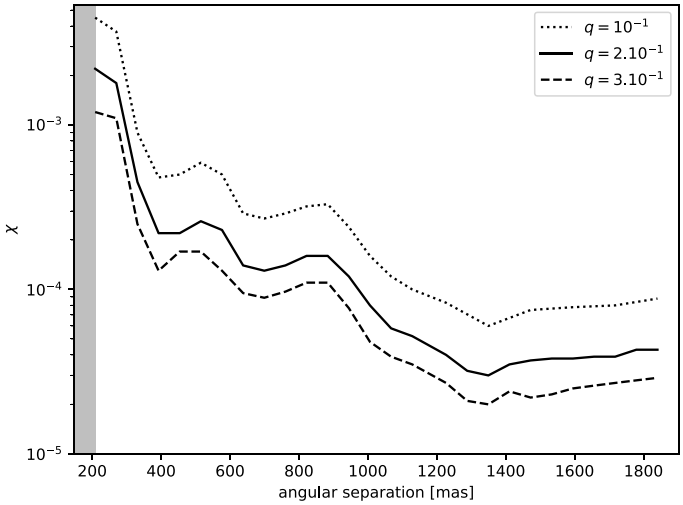

Fig. 14. Minimal contrast χ required to achieve a given relative error q as a function of the angular separation. The conditions are the same as in Figs. 11 and 12.

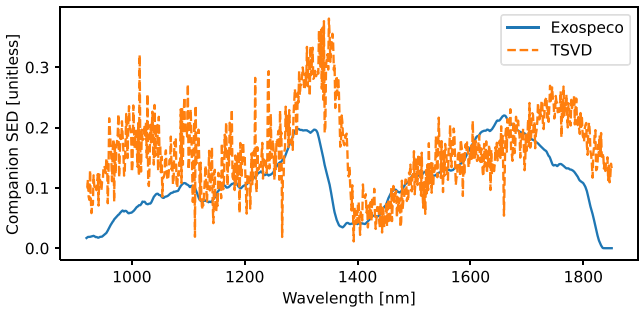

Fig. 15. Comparison of the SEDs extracted from the HR 3549 data by EXOSPECO and by a standard TSVD method. See text for details.

A possible improvement could be to use a more complex model of the on-axis PSF and consider more than one mode in the series expansion of Eq. (3).

To summarize the performances of the current version of EXOSPECO for a single data frame of the HIP 65426 observations, Fig. 14 plots the minimal contrast needed to achieve a given relative error q as a function of the angular separation. The figure shows that by tolerating a relative error as high as $q = 0.3$, a companion with a contrast up to $\chi \sim 2 \times 10^{-5}$ can be characterized. In our conclusions, we explain how to extend EXOSPECO to jointly process several data frames in order to increase the sensitivity of the algorithm.

5.4. Comparison with TSVD extraction

We compared EXOSPECO to a standard approach based on the TSVD method described in Sect. 2.2 to remove the stellar leakages. Figure 15 shows the companion SED extracted from the HR 3549 data by EXOSPECO and by the TSVD approach. For the latter method, the SED of the companion was extracted by local averaging in a 7 pixel height sliding window along the companion signal in the residual image given by Eq. (7) and shown in Fig. 3. In both cases, the same complex calibration model of the spectral and angular dispersion laws described in Appendix A has been used. In spite of this identical calibration, the two extracted SEDs are notably different. Thanks to the optimal extraction in the maximum likelihood sense and to

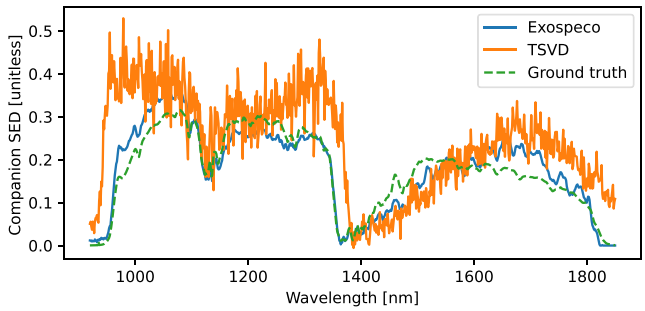

Fig. 16. Comparison of the SEDs extracted from semi-synthetic data by EXOSPECO and by a standard TSVD method. The angular separation and the contrast of the injected companion are respectively $\rho_\oplus - \rho_\star = 785$ mas and $\chi = 2 \times 10^{-4}$. The green dashed curve represents the ground truth injected spectrum which is that of HIP 65426 multiplied by χ.

the spectral regularization, the SED extracted by EXOSPECO is smoother and less noisy. At a coarser resolution, the two SEDs display quite different spectral features. However, without a known ground truth, the two SEDs cannot be ranked. For this reason, we also compared the results given by the two methods on a synthetic injection done as described in Sect. 5.3. Figure 16 clearly demonstrates that not only does EXOSPECO produce less noisy results, but that they also offer a better reflection of reality.

6. Conclusion

In this paper, we present a novel algorithm, EXOSPECO, to extract the spectrum of a companion from high-contrast long-slit spectroscopic data. The most challenging part of such a processing is to disentangle the signal of interest from the stellar leakages since they are much brighter. Compared to existing methods, our algorithm avoids any transform of the data, whether it is to align the speckles of the stellar leakages at all wavelengths or to fix defective pixels. EXOSPECO has also the advantage of jointly extracting the parameters describing the stellar leakages (the star spectrum and the on-axis PSF), the companion spectrum, the off-axis PSF, and, optionally, some calibration parameters. By using non-uniform statistical weights for the data pixels, our approach is optimal in the maximum likelihood sense, namely, it takes into account all available measurements and consistently treats defective pixels as missing data. The joint optimization problem having no closed-form solution, we proposed an alternating minimization strategy which has proven to be effective. In spite of the numerous parameters coming into play in the algorithm, the outputs of the method are, in practice, mostly driven by a single hyper-parameter that tunes the level of regularization of the companion SED.

Although it is not directly part of the spectrum extraction algorithm, we have shown that careful calibration of the instrument is critical to get rid of the contamination by the stellar leakages. For that purpose, we described a refined calibration method of the spectral and spatial dispersion laws from available calibration data. In particular, SPHERE/LSS data present a misalignment of the principal directions of dispersion with the detector axes as well as a geometrical shear. If they are not accounted for, we show that these distortions have a detrimental impact on the result of the processing, whether it is by EXOSPECO or by the current state-of-the-art method. A few remaining calibration parameters that may depend on the observing conditions, such as the off-axis PSF size and the precise locations of the star and of the companion along the slit, can be optionally adjusted by a self-calibration procedure built into EXOSPECO. Thanks to this calibration step, our method significantly reduces the "self-subtracting" bias by better disentangling the stellar leakages component from the companion component. A Julia (Bezanson et al. 2017) implementation of EXOSPECO is freely available[10], with an implementation of the calibration method described in Sect. 4 and Appendix A[11].

Based on tests carried on empirical long-slit spectroscopic data and on injections of a synthetic companion signal in these data, we demonstrated that the proposed approach effectively avoids the self-subtraction bias, even very close to the coronagraphic mask. We provided curves to predict the minimal contrast required to achieve a given quality of extraction of the companion SED. Reliable extraction of a companion SED can be achieved from a single data frame at contrasts as low as a few 10^{-5}. The proposed method could boost the characterization of known (faint) exoplanets at a spectral resolution substantially higher than currently possible with SPHERE IFS ($R \sim 35 - 50$) and for contrasts much better than achievable with IRDIS MRS using state of the art methods. By capturing more the stellar contamination efficiently, the method we propose does not require independent and thus imperfect calibration of the speckles by rotating the slit to hide the planet signal. This will typically gain at least 50% in telescope time while reaching (and even surpassing) the same contrast limit. This new method also paves the way to combining polarimetry and spectroscopic measurements with IRDIS LSS mode (R. Holstein, priv. comm.).

As it is based on an inverse problems framework, EXOSPECO is very flexible and can be adapted to various kinds of data (such as data sequences or data from other instruments). An example of such an extension of EXOSPECO is the joint processing of multiple frames that can be done as follows. Assuming T LSS exposures $\boldsymbol{d}_1$ to $\boldsymbol{d}_T$ of the same object are collected during a night, they can be combined into a single criterion that extends Eq. (17):

$$\begin{aligned} C(\boldsymbol{x}, \boldsymbol{y}_1, \ \boldsymbol{y}_2, \ldots, \boldsymbol{y}_T, z, \boldsymbol{\nu}, \boldsymbol{\mu}) \\ = \sum_{t=1}^{T} \|\boldsymbol{d}_t - \boldsymbol{m}(\boldsymbol{x}, \boldsymbol{y}_t, z, \boldsymbol{\nu})\|^2_{\mathbf{W}_t} + \mu_{\boldsymbol{x}} \mathcal{R}_{\boldsymbol{x}}(\boldsymbol{x}_t) \\ + \mu_{\boldsymbol{y}} \sum_{t=1}^{T} \mathcal{R}_{\boldsymbol{y}}(\boldsymbol{y}_t) + \mu_z \mathcal{R}_z(z), \end{aligned} \tag{37}$$

where the statistical independence of noise between frames is considered (a natural assumption). This criterion can be minimized in $\boldsymbol{x}$, $\boldsymbol{y}_1$, …, $\boldsymbol{y}_T$, z, and $\boldsymbol{\nu}$ following the same alternating method as described in Sect. 3.4, only with more steps in order to estimate the on-axis PSF at each of the T frames. Such a joint processing has the potential to improve the estimation of companion SEDs and push further the achievable contrast limit.

Finally, to better disentangle the stellar leakages from the companion spectrum, the model of the on-axis PSF could be improved by taking into account more spatial modes of the series expansion in Eq. (3). Indeed, as demonstrated in Devaney & Thiébaut (2017), accounting for more such modes significantly improves the modeling of the stellar leakages, especially near the coronagraphic mask. Such an improvement would not call into question the founding principles of EXOSPECO, but it would require an optimization strategy to be adopted.

[10] https://github.com/SJJThe/Exospeco
[11] https://github.com/SJJThe/ExospecoCalibration

Acknowledgements. This work was supported by the Action Spécifique Haute Résolution Angulaire (ASHRA) of CNRS/INSU co-funded by CNES. SPHERE is an instrument designed and built by a consortium consisting of IPAG (Grenoble, France), MPIA (Heidelberg, Germany), LAM (Marseille, France), LESIA (Paris, France), Laboratoire Lagrange (Nice, France), INAF – Osservatorio di Padova (Italy), Observatoire de Genève (Switzerland), ETH Zürich (Switzerland), NOVA (Netherlands), ONERA (France) and ASTRON (Netherlands) in collaboration with ESO. SPHERE was funded by ESO, with additional contributions from CNRS (France), MPIA (Germany), INAF (Italy), FINES (Switzerland) and NOVA (Netherlands). SPHERE also received funding from the European Commission Sixth and Seventh Framework Programmes as part of the Optical Infrared Coordination Network for Astronomy (OPTICON) under grant number RII3-Ct-2004-001566 for FP6 (2004–2008), grant number 226604 for FP7 (2009–2012) and grant number 312430 for FP7 (2013–2016).

References

Berdeu, A., Soulez, F., Denis, L., Langlois, M., & Thiébaut, É. 2020, A&A, 635, A90
Beuzit, J. L., Vigan, A., Mouillet, D., et al. 2019, A&A, 631, A155
Bezanson, J., Edelman, A., Karpinski, S., & Shah, V. B. 2017, SIAM Rev., 59, 65
Brent, R. 2013, Algorithms for Minimization Without Derivatives (Dover Publications)
Carter, A. L., Hinkley, S., Kammerer, J., et al. 2023, ApJ, 951, L20
Catmull, E., & Rom, R. 1974, in Computer Aided Geometric Design, eds. R. E. Barnhill, & R. F. Riesenfeld (Academic Press), 317
Chauvin, G., Desidera, S., Lagrange, A.-M., et al. 2017, A&A, 605, A9
Dallant, J., Langlois, M., Thiébaut, É. M., & Flasseur, O. 2022, Proc. SPIE, 12185, 1218537
Devaney, N., & Thiébaut, É. 2017, MNRAS, 472, 3734
Dohlen, K., Langlois, M., Saisse, M., et al. 2008, SPIE Conf. Ser., 7014, 70143L
Eckart, C., & Young, G. 1936, Psychometrika, 1, 211
Flasseur, O., Denis, L., Thiébaut, É., & Langlois, M. 2018, A&A, 618, A138
Flasseur, O., Denis, L., Thiébaut, É., & Langlois, M. 2020a, A&A, 637, A9
Flasseur, O., Denis, L., Thiébaut, É., & Langlois, M. 2020b, A&A, 634, A2
Golub, G. H., Heath, M., & Wahba, G. 1979, Technometrics, 21, 215
Hansen, P. C., & O'Leary, D. P. 1993, SIAM J. Sci. Comput., 14, 1487
Jovanovic, N., Martinache, F., Guyon, O., et al. 2015, PASP, 127, 890
Lafreniere, D., Marois, C., Doyon, R., Nadeau, D., & Artigau, E. 2007, ApJ, 660, 770
Lawson, C. L., & Hanson, R. J. 1974, Solving Least Squares Problems (Prentice-Hall)
Macintosh, B., Graham, J., Palmer, D., et al. 2006, Proc SPIE, 6272, 177
Macintosh, B., Graham, J. R., Ingraham, P., et al. 2014, PNAS, 111, 12661
Mahalanobis, P. C. 1936, in Proceedings of the National Institute of Sciences of India, II, 49–55
Marois, C., Correia, C., Véran, J.-P., & Currie, T. 2013, Proc. Int. Astron. Union, 8, 48
Mesa, D., Vigan, A., D'Orazi, V., et al. 2016, A&A, 593, A119
Mirsky, L. 1960, Q. J. Math., 11, 50
Molina, R. 1994, IEEE Trans. Pattern Anal. Mach. Intell., 16, 1122
Mugnier, L., Fusco, T., & Conan, J.-M. 2004, J. Opt. Soc. Am. A, Opt. Image Sci. Vis., 21, 1841
Mugnier, L. M., Cornia, A., Sauvage, J.-F., et al. 2009, J. Opt. Soc. Am. A, 26, 1326
Powell, M. J. D. 2006, in Nonconvex Optimization and Its Applications, eds. G. Di Pillo, & M. Roma (Springer Science), 255
Powell, M. 2009, Technical Report, Department of Applied Mathematics and Theoretical Physics, UK
Smith, I., Ferrari, A., & Carbillet, M. 2009, IEEE Trans. Signal Process., 57, 904
Soummer, R., Pueyo, L., & Larkin, J. 2012, ApJ, 755, L28
Sparks, W. B., & Ford, H. C. 2002, ApJ, 578, 543
Stein, C. M. 1981, Ann. Stat., 9, 1135
Thé, S., Thiébaut, É., Denis, L., & Soulez, F. 2020, in 28th European Signal Processing Conference, EUSIPCO 2020, Amsterdam, Netherlands, January 18–21, 2021 (IEEE), 2358
Thiébaut, É. 2002, Proc. SPIE, 4847, 174
Thiébaut, É., Devaney, N., Langlois, M., & Hanley, K. 2016, in Adaptive Optics Systems V, eds. E. Marchetti, L. M. Close, & J.-P. Véran (SPIE-Intl Soc Optical Eng)
Tikhonov, A. N., & Arsenin, V. I. 1977, Solutions of Ill-posed Problems (Washington: Winston; New York: distributed solely by Halsted Press)
Titterington, D. M. 1985, A&A, 144, 381
Vigan, A. 2016, Astrophysics Source Code Library [record ascl:1603.001]
Vigan, A., Langlois, M., Moutou, C., & Dohlen, K. 2008, A&A, 489, 1345
Vigan, A., Bonnefoy, M., Chauvin, G., Moutou, C., & Montagnier, G. 2012, A&A, 540, A131
Xie, C., Choquet, E., Vigan, A., et al. 2022, A&A, 666, A32

Appendix A: Calibration of the spectro-angular dispersion laws

This appendix provides some details about the methods used for the calibration of the spectro-angular laws described in Sect. 4 and some figures to support the results discussed in Sect. 4.5. The considered calibration data $\boldsymbol{d}_{\text{cal}}$ is the image in the central panel of Fig. 5.

A.1. Transverse projection

To locate the positions of the spectral lines, we compute a weighted *transverse* projection of the calibration image $\boldsymbol{d}_{\text{cal}}$:

$$q_{\perp k}(\phi_\lambda) = \frac{\sum_{i,j} \xi_{\perp\, i,j,n}(\phi_\lambda)\, d_{\text{cal},i,j}}{\sum_{i,j} \xi_{\perp\, i,j,n}(\phi_\lambda)}, \tag{A.1a}$$

with weights given by:

$$\xi_{\perp\, i,j,k}(\phi_\lambda) = w_{\text{msk},i,j}\, \varphi_{\text{proj}}(i\, \sin\phi_\lambda + j\, \cos\phi_\lambda - k) \tag{A.1b}$$

and for a projection angle, ϕ_λ, chosen to maximize the peak values of the resulting projection (plotted in red in the top panel of Fig. 5 for $\phi_\lambda \approx 0°$). In practice, we take $\varphi_{\text{proj}}(t) = \max(1 - |t|, 0)$, the linear B-spline, as the interpolating function for the projection. We note that thanks to the weighting by the mask of valid pixels $\boldsymbol{w}_{\text{msk}}$ defined in Eq. (30), invalid pixels have no incidence on the computed projection.

A.2. Detection of the spectral peaks

We use algorithm 4 with tolerance parameter $\delta_\perp = 10$ pixels to find the N_λ most significant peaks in the transverse projection $\boldsymbol{q}_\perp(\phi_\lambda) \in \mathbb{R}^{N_q}$ computed according to Eq. (A.1).

Algorithm 4: Find the most significant peaks

Input: N_λ, $\boldsymbol{q}_\perp(\phi_\lambda)$, and $\delta_\perp$.
Output: $\mathcal{P}(\phi_\lambda)$.
$\mathcal{P} \leftarrow \varnothing$; ◁ start with an empty list
$z = \boldsymbol{q}_\perp(\phi_\lambda)$; ◁ copy profile in workspace array
for $\ell = 1, \ldots, N_\lambda$ **do**
  $k_\ell = \arg\max_k z_k$; ◁ find position of largest value in z
  $\mathcal{P} \leftarrow \mathcal{P} \cup \{k_\ell\}$; ◁ update list of positions
  for $k \in [\![\max(1, k_\ell - \delta_\perp), \min(k_\ell + \delta_\perp, N_q)]\!]$ **do**
    $z_k \leftarrow -\infty$; ◁ invalidate nearby pixels
return $\mathcal{P}$;

A.3. Extraction of the paths of the spectral lines

Given the projection angle ϕ_λ and the list $\mathcal{P}(\phi_\lambda)$ of the N_λ most significant peaks in the transverse projection, $\boldsymbol{q}_\perp(\phi_\lambda)$, we build the ℓth spectral path C_ℓ as a list of points along the ℓth spectral line. The coordinates, $(i^{\text{path}}_{\ell,m}, j^{\text{path}}_{\ell,m})$, of the mth such point are given by computing the center of gravity of the calibration data in a small rectangular window, $\mathcal{W}_{\ell,m}(\phi_\lambda)$, sliding along the considered spectral lines:

$$(i^{\text{path}}_{\ell,m}, j^{\text{path}}_{\ell,m}) = \frac{\sum_{(i,j)\in\mathcal{W}_{\ell,m}(\phi_\lambda)} w_{\text{msk},i,j} \times (i,j)}{\sum_{(i,j)\in\mathcal{W}_{\ell,m}(\phi_\lambda)} w_{\text{msk},i,j}}, \tag{A.2}$$

computed for all non-empty[12] sliding window $\mathcal{W}_{\ell,m}$ of size ~ 1 pixel along the spectral line and $2\,\delta_\perp + 1$ pixels in the perpendicular direction:

$$\mathcal{W}_{\ell,m}(\phi_\lambda) = \left\{ \begin{array}{l} (i,j) \in [\![1,I]\!] \times [\![1,J]\!] \text{ such that} \\ \left|i\, \sin\phi_\lambda + j\, \cos\phi_\lambda - k_\ell\right| \leq \delta_\perp + \frac{1}{2} \\ \text{and } \left|i\, \cos\phi_\lambda - j\, \sin\phi_\lambda - m\right| \leq \frac{1}{2} \end{array} \right\}, \tag{A.3}$$

where k_ℓ is the ℓth index in the list $\mathcal{P}(\phi_\lambda)$ of the N_λ most significant peaks in the transverse projection $\boldsymbol{q}_\perp(\phi_\lambda)$. Again, we note that thanks to the weighting by the mask of valid pixels, invalid pixels have no incidence on the computed coordinates. In practice, we use the same value for the half-width of the sliding windows and for the minimal separation between peaks in the transverse projection, that is $\delta_\perp = 10$ pixels for the considered calibration data.

A.4. Detection of the edges of the spectral bands

Given the projection angle, ϕ_λ, and the list, $\mathcal{P}(\phi_\lambda)$, of the N_λ most significant peaks in the transverse projection $\boldsymbol{q}_\perp(\phi_\lambda)$, we compute the longitudinal profile of each spectral line as the following weighted projection:

$$q_{/\!/\,\ell,k}(\phi_\lambda) = \frac{\sum_{(i,j)\in\mathcal{D}_\ell(\phi_\lambda)} \xi_{/\!/\,i,j,k}(\phi_\lambda)\, d_{\text{cal},i,j}}{\sum_{(i,j)\in\mathcal{D}_\ell(\phi_\lambda)} \xi_{/\!/\,i,j,k}(\phi_\lambda)}, \tag{A.4a}$$

with weights given by:

$$\xi_{/\!/\,i,j,k}(\phi_\lambda) = w_{\text{msk},i,j}\, \varphi_{\text{proj}}(i\, \cos\phi_\lambda - j\, \sin\phi_\lambda - k), \tag{A.4b}$$

and where k is the index along the projection and $\mathcal{D}_\ell(\phi_\lambda)$ is a narrow rectangular window (in green in the central panel of Fig. 5) to isolate the pixels of the calibration image, $\boldsymbol{d}_{\text{cal}}$, impacted by the considered spectral line:

$$\mathcal{D}_\ell(\phi_\lambda) = \left\{ \begin{array}{l} (i,j) \in [\![1,I]\!] \times [\![1,J]\!] \text{ such that} \\ \left|i\, \sin\phi_\lambda + j\, \cos\phi_\lambda - k_\ell\right| \leq \delta_\perp + \frac{1}{2} \end{array} \right\}, \tag{A.5}$$

with, as before, $\delta_\perp \approx 10$ pixels the half-width of the region.

Detecting the edges of the central hole due to the coronagraphic mask in the resulting profile (plotted in green in the left panel of Fig. 5) can be done by a quite simple procedure. Each value at index k of the profile is compared with the next one up to a threshold value τ. If $q_{/\!/\,\ell,k} \leqslant \tau \leqslant q_{/\!/\,\ell,k+1}$ and $q_{/\!/\,\ell,k} < q_{/\!/\,\ell,k+1}$, then an ascending edge is detected. If $q_{/\!/\,\ell,k} \geqslant \tau \geqslant q_{/\!/\,\ell,k+1}$ and $q_{/\!/\,\ell,k} > q_{/\!/\,\ell,k+1}$, it is a descending edge. From the four edges detected in the ℓth spectral line (indicated by the crosses in the left panel of Fig. 5), the second and third ones correspond to the coronagraphic mask. Retrieving these edges in $C_\ell(\phi_\lambda)$ yields the coordinates $(i^{\text{down}}_\ell, j^{\text{down}}_\ell)$ and $(i^{\text{up}}_\ell, j^{\text{up}}_\ell)$ required in Sect. 4.4 for the calibration of the angular dispersion law.

A.5. Results on the calibration data

The results of the proposed calibration models are displayed in Fig. A.1, as blue lines for the spectral law and green lines for the spatial law. In practice, degrees $P_\lambda = 5$ and $P_\rho = 1$ (limited by the small number of points for ρ) were chosen. The three models described in Sect. 4.2 are tested, that is a standard model, one of a medium complexity and one more complex.

As can be seen on the different zooms shown in Fig. A.2, 2D polynomials are needed to explain local distortions. Choosing this model, we plot on Fig. A.3 some iso-wavelength (blue)

[12] in the sense that it contains at least one valid pixel

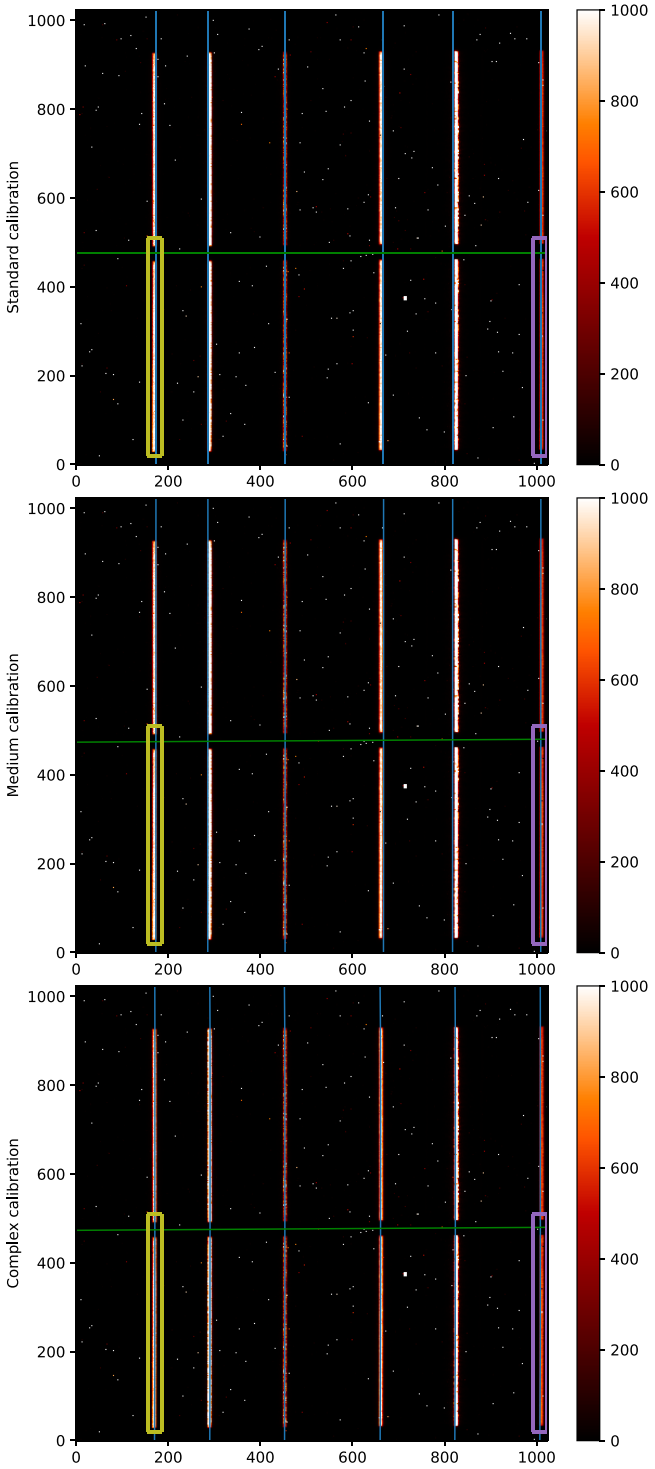

Fig. A.1. Iso-wavelength curves at the wavelengths of the calibration sources (blue lines) and iso-angular distance of the center of the coronagraphic mask of (green lines) presented on top of the calibration data, $\boldsymbol{d}_{\rm cal}$. The upper panel presents the results for the simple model, the central panel shows the results for the medium model, while the bottom panel shows the results of using the complex model.

and iso-angular distance (green) curves, on a zoom in of the HR3549 dataset. This figure highlights how well our proposed models for the dispersion laws are following the speckles, compared to the standard model. A strong shear effect due to the

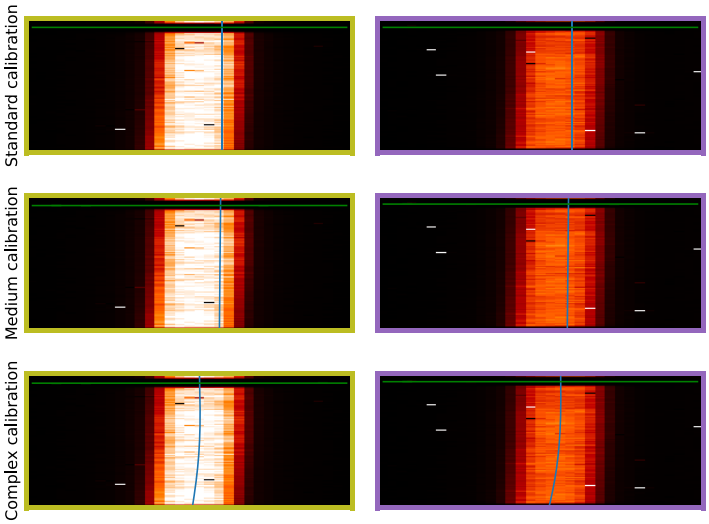

Fig. A.2. Magnified images of the two regions outlined by the yellow and purple rectangles for the three models described in Sect. 4.2. To best see the differences between models, the magnifications are different in the two dimensions.

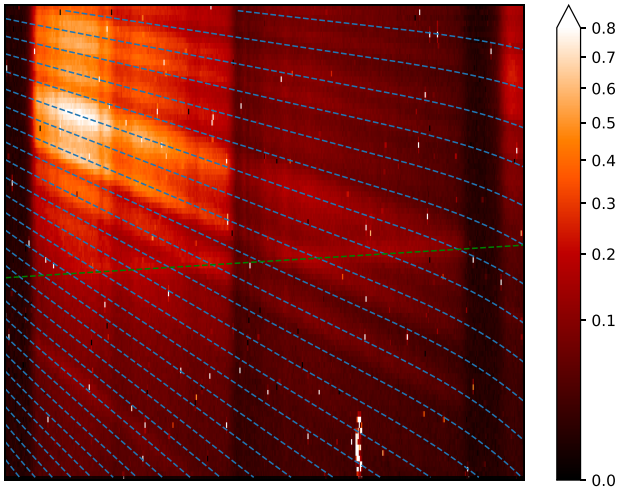

Fig. A.3. Iso-wavelength and iso-angular distance superposed to the HR 3549 data observed on 2015-12-28 with IRDIS in MRS mode.

dispersive elements is visible and taken into account by our model. For all these models, we took polynomials of degrees $P_\lambda = 5$ and $P_\rho = 1$ for the spectral and spatial dispersion laws.

Appendix B: Calibration of the contrast

In order to express the SED of the companion in terms of contrast with respect to the star, we use specific calibration data, $\boldsymbol{d}_{\rm flux}$, (shown in Fig. B.1, top panel) for which the star is placed in the spectrograph slit but shifted away from the coronagraphic mask and with a neutral density inserted in the optical path to avoid detector saturation. Applying FitCompanion (algorithm 2) to $\boldsymbol{d}_{\rm flux}$ and dividing the resulting SED by the spectral transmission of the neutral filter yields the SED of the star $\boldsymbol{x}_{\rm flux}$ (shown in Fig. B.1, bottom panel). The so derived parameters of the off-axis PSF can be later used in FitCompanion (algorithm 2) to extract the companion SED.

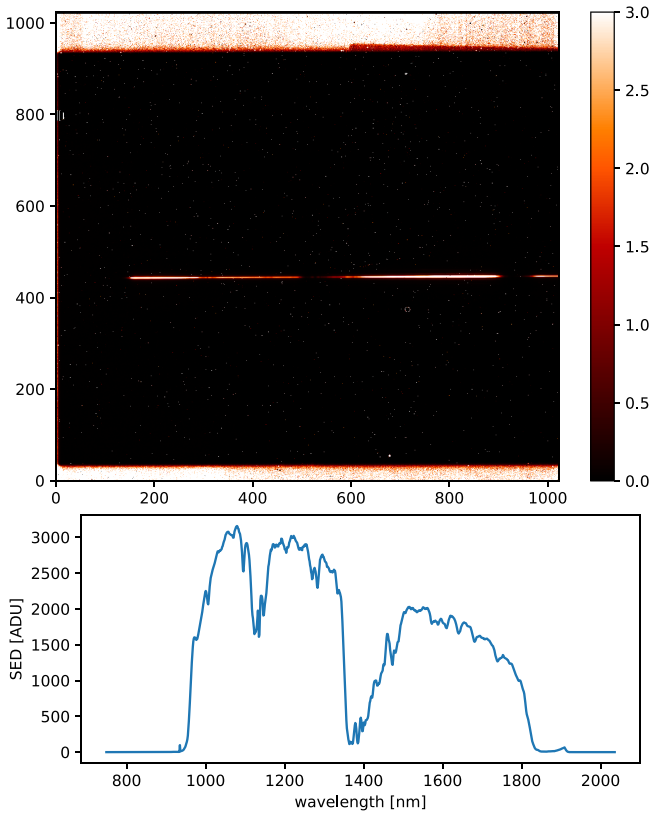

Fig. B.1. Calibration of the SED of the star HIP 65426. *Top:* Calibration image $\boldsymbol{d}_{\text{flux}}$ observed on 2019-05-20 with the MRS mode of SPHERE/IRDIS. *Bottom:* SED of the star extracted by FitStar (algorithm 2) and corrected from the density filter.

Appendix C: Exploiting the scaling indetermination

The estimated components $\hat{\boldsymbol{x}}$, $\hat{\boldsymbol{y}}$, $\hat{z}$, and $\hat{\boldsymbol{\nu}}$ of the direct model defined in Eq. (15) depend on hyper-parameters which include the regularization weights μ_x, μ_y, and μ_z. In this appendix, we show how to adapt the approach of Thé et al. (2020) to reduce the effective number of regularization parameters and also accelerate the minimization.

We first note that the regularizations considered for the stellar components are homogeneous functions of degree 2, namely, the following property holds:

$$\mathcal{R}_{\boldsymbol{u}}(\alpha\,\boldsymbol{u}) = \alpha^2\,\mathcal{R}_{\boldsymbol{u}}(\boldsymbol{u}), \tag{C.1}$$

whatever the component $\boldsymbol{u} = \boldsymbol{x}$ or $\boldsymbol{y}$ considered and $\alpha \geq 0$.

The contribution of the star, the first right-hand side term in Eq. (15), is a bilinear function of the parameters $\boldsymbol{x}$ and $\boldsymbol{y}$. As a consequence:

$$\boldsymbol{m}(\alpha\,\boldsymbol{x}, \boldsymbol{y}/\alpha, z, \boldsymbol{\nu}) = \boldsymbol{m}(\boldsymbol{x}, \boldsymbol{y}, z, \boldsymbol{\nu}), \tag{C.2}$$

holds for any scaling factor $\alpha > 0$.

Combining the properties in Eqs. (C.1) and (C.2) with the definition of the objective function in Eq. (17) and that of the regularization in Eq. (19), it can be seen that for any $\alpha > 0$:

$$C(\alpha\,\boldsymbol{x}, \boldsymbol{y}/\alpha, z, \boldsymbol{\nu}, \mu_x, \mu_y, \mu_z) = C(\boldsymbol{x}, \boldsymbol{y}, z, \boldsymbol{\nu}, \alpha^2\,\mu_x, \alpha^{-2}\,\mu_y, \mu_z). \tag{C.3}$$

In other words, scaling the unknowns $\boldsymbol{x}$ and $\boldsymbol{y}$ without changing the model is equivalent to scaling their regularization weights. Exploiting this, it is possible to compute an optimal scaling factor:

$$\begin{aligned}\hat{\alpha}(\boldsymbol{x}, \boldsymbol{y}, \mu_x, \mu_y) &= \arg\min_{\alpha>0} C(\alpha\,\boldsymbol{x}, \boldsymbol{y}/\alpha, z, \boldsymbol{\nu}, \mu_x, \mu_y, \mu_z) \\ &= \arg\min_{\alpha>0} C(\boldsymbol{x}, \boldsymbol{y}, z, \boldsymbol{\nu}, \alpha^2\,\mu_x, \alpha^{-2}\,\mu_y, \mu_z) \\ &= \arg\min_{\alpha>0} \left\{\alpha^2\,\mu_x\,\mathcal{R}_x(\boldsymbol{x}) + \alpha^{-2}\,\mu_y\,\mathcal{R}_y(\boldsymbol{y})\right\} \\ &= \left(\frac{\mu_y\,\mathcal{R}_y(\boldsymbol{y})}{\mu_x\,\mathcal{R}_x(\boldsymbol{x})}\right)^{\frac{1}{4}}. \end{aligned} \tag{C.4}$$

Plugging this expression in the definition of the criterion yields:

$$\begin{aligned} C^+(\boldsymbol{x}, \boldsymbol{y}, z, \boldsymbol{\nu}, \mu_x, \mu_y, \mu_z) &= \min_{\alpha>0} C(\alpha\,\boldsymbol{x}, \boldsymbol{y}/\alpha, z, \boldsymbol{\nu}, \mu_x, \mu_y, \mu_z) \\ &= \|\boldsymbol{d} - \boldsymbol{m}(\boldsymbol{x}, \boldsymbol{y}, z, \boldsymbol{\nu})\|^2_{\mathbf{W}} + \mu_{x,y}\,\mathcal{R}_{x,y}(\boldsymbol{x}, \boldsymbol{y}) + \mu_z\,\mathcal{R}_z(z) \end{aligned} \tag{C.5}$$

where:

$$\mathcal{R}_{x,y}(\boldsymbol{x}, \boldsymbol{y}) = \sqrt{\mathcal{R}_x(\boldsymbol{x})\,\mathcal{R}_y(\boldsymbol{y})}, \tag{C.6}$$

and:

$$\mu_{x,y} = 2\,\sqrt{\mu_x\,\mu_y}. \tag{C.7}$$

This shows that the regularizations of the stellar components $\boldsymbol{x}$ and $\boldsymbol{y}$ are entangled and that the effect of these regularizations on the shape of these components is effectively controlled by a single hyper-parameter; here, this is $\mu_{x,y}$. As shown by Eq. (C.4), the ratio μ_x/μ_y of the hyper-parameters controls the scaling, not the shape, of $\boldsymbol{x}$ and $\boldsymbol{y}$, one of the two can be fixed. This saves us the hassle of adjusting both parameters at the same time.