Draft version June 1, 2022
Typeset using LaTeX/show your work! modern style in AASTeX631

Mapping stellar surfaces III: An Efficient, Scalable, and Open-Source Doppler Imaging Model

Rodrigo Luger [1], Megan Bedell [1], Daniel Foreman-Mackey [1], Ian J. M. Crossfield [2], Lily L. Zhao [1] and David W. Hogg [1]

[1] *Center for Computational Astrophysics, Flatiron Institute, New York, NY*
[2] *Department of Physics and Astronomy, Kansas University, Lawrence, KS*

ABSTRACT

The study of stellar surfaces can reveal information about the chemical composition, interior structure, and magnetic properties of stars. It is also critical to the detection and characterization of extrasolar planets, in particular those targeted in extreme precision radial velocity (EPRV) searches, which must contend with stellar variability that is often orders of magnitude stronger than the planetary signal. One of the most successful methods to map the surfaces of stars is Doppler imaging, in which the presence of inhomogeneities is inferred from subtle line shape changes in high resolution stellar spectra. In this paper, we present a novel, efficient, and closed-form solution to the problem of Doppler imaging of stellar surfaces. Our model explicitly allows for incomplete knowledge of the local (rest frame) stellar spectrum, allowing one to learn differences from spectral templates while simultaneously mapping the stellar surface. It therefore works on blended lines, regions of the spectrum where line formation mechanisms are not well understood, or stars whose spots have intrinsically different spectra from the rest of the photosphere. We implement the model within the open source `starry` stellar modeling framework, making it fast, differentiable, and easy to use in both optimization and posterior inference settings. As a proof-of-concept, we use our model to infer the surface map of the brown dwarf WISE 1049-5319B, finding close agreement with the solution of Crossfield et al. (2014). We also discuss Doppler imaging in the context of EPRV studies and describe an interpretable spectral-temporal Gaussian process for stellar spectral variability that we expect will be important for EPRV exoplanet searches. This paper was prepared using the show your work! scientific article workflow, which exposes the data, source code, and package dependencies for all of the paper's figures, allowing readers to reproduce the results in this paper exactly and with minimal effort.

rluger@flatironinstitute.org

1. INTRODUCTION

High resolution spectra encode a vast amount of information about the structure, dynamics, and chemical composition of stars. Spectra of fast-rotating stars, in particular, also encode the two-dimensional brightness map of the photosphere via subtle line shape changes as inhomogeneities rotate from the blueshifted side to the redshifted side of the star. The process of inferring the map of a stellar "surface" given high resolution spectra taken at different rotational phases is known as *Doppler imaging*, a technique that dates back to at least Deutsch (1958). Most modern Doppler imaging studies, however, are based on the methodology developed in Khokhlova (1976), Goncharskii et al. (1977), and Goncharskii et al. (1982), and updated in Vogt & Penrod (1983) and Vogt et al. (1987). These studies formalised the technique of forward-modeling the line shape changes due to arbitrary brightness variations rotating in and out of view. The latter study, in particular, presented a now common technique for solving the inverse problem—inferring the map of the star given a spectroscopic dataset—under certain regularizing assumptions.

Since those pioneering studies several decades ago, Doppler maps have now been obtained for a variety of targets, ranging from giant and subgiant stars (Donati 1999; Strassmeier 1999; Roettenbacher et al. 2017) to massive main sequence and pre-main sequence stars (Collier Cameron & Unruh 1994; Hatzes 1995) to cool dwarfs (Rosén et al. 2015; Zaire et al. 2021). A Doppler map has even been obtained for a brown dwarf in the solar neighborhood (Crossfield et al. 2014). Future observations with extremely large telescopes may also allow one to obtain Doppler maps of directly imaged extrasolar planets (Snellen et al. 2014; Crossfield 2014).

As more and more objects are imaged in this way, it becomes tempting to interpret Doppler images in the broader context of stellar physics and evolution. For instance, one may be interested in understanding trends in the population-level magnetic properties of giant stars, or in how those properties change as a function of spectral type or stellar age. However, the vast majority of Doppler imaging studies produce *point estimates*: that is, they usually report the maximum likelihood solution for the stellar surface map, with no formal assessment of its uncertainty. In all but the simplest cases, maximum likelihood estimators are biased (e.g., Lehmann & Casella 1998), meaning any attempt to understand these estimates in bulk is prone to significant bias (see, e.g., Hogg et al. 2010). This problem is compounded by four factors.

First, Doppler imaging is usually an ill-posed problem, meaning there is no unique solution for the surface map given a spectroscopic dataset. This issue is typically well understood (e.g., Piskunov et al. 1990; Rice 2002), and most studies circumvent this issue by employing restrictive priors to ensure the uniqueness of the solution. These include maximum entropy regularization (Vogt et al. 1987) and Tikhonov regularization (e.g., Piskunov et al. 1990), which typically enforce sparsity and smoothness in the solution, respectively. However, while useful for obtaining a maximum likelihood solution, both choices can lead to significant bias in the inferred map, as they can

both suppress features and introduce spurious artifacts. In fact, a recent comparison of Doppler imaging, photometric light curve inversion, and interferometry revealed qualitatively diffferent maps of the the same star (Roettenbacher et al. 2017), due to the various degeneracies afflicting each method. **Second**, the Doppler imaging technique is generally more sensitive to longitudinal than latitudinal variations, often leading to spurious, latititudinally elongated features in the maximum likelihood solution (e.g., Unruh & Collier Cameron 1995; Crossfield 2014). **Third**, the Doppler imaging technique requires the availability of accurate spectral templates for the local rest frame spectrum of the star. Several studies have shown that errors in line templates or the presence of unresolved line blends in the observed spectrum can introduce signficant artifacts in the map solution (e.g., Rice et al. 1989; Unruh & Collier Cameron 1995). Moreover, changes in the local spectrum over the surface of the star (such as different line depths or the presence of different absorption lines within spots) can similarly bias the inferred map if these are not properly modeled. While some past studies have solved for variations in the local line profiles (Khokhlova 1976; Goncharskii et al. 1977, 1982), these usually make the assumption that the intensity variation is small across the surface of the star; that is, they do not simultaneously solve for the Doppler map of the surface brightness variations. **Fourth**, and finally, Doppler imaging results reported in the literature use a variety of custom data analysis codes, most of which are not open source (Collier Cameron 1997; Hackman et al. 2001; Carroll et al. 2012; Rosén et al. 2015; Asensio Ramos et al. 2021, to name a few); many studies also utilize data that is often not included in the resulting publication or made easily available online. This makes it difficult to reproduce results and compare Doppler images of different stars on the same footing. As argued above, detailed choices in a Doppler imaging analysis will have a significant effect on the reported point-estimate result, so this inability to compare uniformly analyzed observations is a significant hindrance to interpretation.

In this paper, the third in a series dedicated to mapping stellar surfaces (Luger et al. 2021b,c), we develop a novel algorithm for Doppler imaging that addresses each of these issues. The various pitfalls of Doppler imaging outlined above necessitate an algorithm that is efficient enough to be used in a Bayesian posterior inference context, where one can formally address the degeneracies at play and quantify their effect on the uncertainties of the inferred stellar maps. They also necessitate an algorithm that is flexible, capable of learning deviations from the template or capturing spatial variations in the rest-frame spectrum. And most of all, they necessitate a user-friendly, well-documented, and open source implementation, which is currently lacking the literature. Our algorithm addresses these various issues by building on our previous work (Luger et al. 2019), where we derived analytic solutions to the light curve mapping problem by expanding the stellar surface map in the spherical harmonic basis. While spherical harmonics are routinely used in Doppler imaging problems (Deutsch 1958, 1970; Falk & Wehlau 1974; Donati et al. 2006; Kochukhov et al. 2014), we present

a new closed-form solution to the forward model for Doppler imaging when working in this basis, making our model both precise and computationally efficient. We also expose a fundamental linearity of the model in terms of both the surface map and the rest frame spectrum, allowing us to learn both using a fast bilinear solver. Finally, we implement our algorithm within the open source `starry` code package with accompanying documentation, examples, and tutorials.[1] All of the figures in this paper are also open source; clickable icons next to each figure link to the exact version of the `Python` script used to generate them. To ensure the full reproducibility of this paper, we prepared it using the **show your work!** workflow[2], which exposes the datasets, source code, and all package dependencies needed to exactly reproduce the results.

The paper is organized as follows: in §2 and §3 we introduce the math behind the Doppler imaging problem and derive a closed-form expression for the model. In §4, §5, and §6 we demonstrate how to re-express the model as a linear operation on the input spectrum and stellar map and derive closed-form expressions for the two conditioned on the data and priors. In §7 we apply our techniques to a mock problem, and in §8 we apply it to a real spectral dataset of the brown dwarf WISE 1049-5319B. We further discuss our results in §9, and in §10 we focus on an immediate extension of our work: the development of an interpretable spectral-temporal Gaussian process for stellar spectral variability. Finally, we discuss the implementation of our Doppler imaging algorithm within the `starry` framework in §11, and present a summary of the paper in §12.

2. THE PROBLEM

Let $I(\lambda, \mathbf{x}, t)$ be the specific intensity observed at log wavelength $\lambda \equiv \ln \frac{\lambda}{\lambda_r}$ and at sky-projected Cartesian position $\mathbf{x} = (x, y, z)$ on the surface of the star at time t, where λ_r is an arbitrary reference wavelength, such as the central wavelength in the observed spectrum. We may express this intensity as

$$I(\lambda, \mathbf{x}, t) = I\Big(\lambda_0 + D(\mathbf{x}), \mathbf{x}, t\Big) \tag{1}$$

where λ_0 is the log wavelength in the rest frame and D is the Doppler shift in log wavelength space:

$$D(\mathbf{x}) = \frac{1}{2} \ln \left(\frac{1 + \beta(\mathbf{x})}{1 - \beta(\mathbf{x})} \right) \tag{2}$$

where $\beta = v(\mathbf{x})/c$ is the ratio of the radial velocity at a point on the surface of the star to the speed of light. In keeping with the literature, we take positive values of v (and D) to mean redshifts.

A common approach to computing the Doppler-shifted spectrum is to evaluate the spectrum at the rest frame wavelength λ_0 and interpolate back to the grid in λ. This

[1] https://github.com/rodluger/starry
[2] https://github.com/rodluger/showyourwork

is practical when computing the spectrum at a single *point* on the surface, but not ideal when one is interested in the *integral* over the visible surface of the star $\mathcal{S}$, which is typically all we can observe:

$$F(\lambda, t) \equiv \iint_{\mathcal{S}(\mathbf{x})} I(\lambda, \mathbf{x}, t) \mathrm{d}\mathcal{S}(\mathbf{x}) \ , \tag{3}$$

The difficulty in solving Equation (3) stems from the fact that $I(\lambda, \mathbf{x}, t)$ is difficult to write down in closed form, given the nonlinearity of the Doppler shift. The standard approach to solving this integral is therefore to discretize the surface of the star with a fine grid, evaluate the Doppler-shifted spectrum in each cell, and sum over the spatial axes to approximate the integral. Depending on the resolution of the grid, this may be either numerically inaccurate or computationally inefficient.

3. THE SOLUTION

3.1. *Doppler Deconvolution*

The observed spectrum is an intricate function of spatial, spectral, temporal, and velocity terms. The goal in this section is to deconvolve each of these terms to make solving the integral in Equation (3) easier. The first thing we will do is to express Equation (1) as a convolution:

$$I(\lambda, \mathbf{x}, t) = I(\lambda_0, \mathbf{x}, t) * \delta\big(\lambda_0 + D(\mathbf{x})\big) \tag{4}$$

where δ is the Dirac delta function and $*$ denotes the linear convolution operator, defined for two arbitrary functions g and h as the integral

$$(g * h)(t) \equiv \int_{-\infty}^{\infty} g(\tau) h(t - \tau) d\tau \tag{5}$$

for some independent coordinate t and dummy parameter τ. The convolution of $I(\lambda_0)$ with a delta function has the effect of shifting the spectrum by an amount D, returning a function of the shifted (observed) log wavelength, $\lambda = \lambda_0 + D$.

Next, we expand the spatial dependence of the specific intensity at the rest frame wavelength in terms of spherical harmonics on the unit disk:

$$I(\lambda_0, \mathbf{x}, t) = \sum_{l=0}^{l_{\max}} \sum_{m=-l}^{l} a_{lm}(\lambda_0, t) Y_{lm}(\mathbf{x}) \ , \tag{6}$$

where $Y_{lm}(\mathbf{x})$ is a spherical harmonic on the projected disk and $a_{lm}(\lambda_0, t)$ is the corresponding spherical harmonic coefficient at log wavelength in the rest frame λ_0 and time t. For convenience, we may write this equation in vector form:

$$I(\lambda_0, \mathbf{x}, t) = \boldsymbol{\psi}^\top(\mathbf{x})\, \mathbf{a}(\lambda_0, t) \ , \tag{7}$$

where

$$\mathbf{a}(\lambda_0, t) \equiv \Big(a_{0,0}(\lambda_0, t)\ a_{1,-1}(\lambda_0, t)\ a_{1,0}(\lambda_0, t)\ a_{1,1}(\lambda_0, t)\ ...\Big)^\top \tag{8}$$

is the vector of $N = (l_{\max} + 1)^2$ spherical harmonic coefficients and

$$\boldsymbol{\psi}^\top(\mathbf{x}) \equiv \Big(Y_{0,0}(\mathbf{x})\ Y_{1,-1}(\mathbf{x})\ Y_{1,0}(\mathbf{x})\ Y_{1,1}(\mathbf{x})\ ...\Big) \tag{9}$$

is the corresponding vector of N spherical harmonics. We may further decompose our expression by linearizing the time dependence of the spherical harmonic coefficients:

$$\mathbf{a}(\lambda_0, t) = \mathbf{R}(t)\,\mathbf{a}(\lambda_0)\ , \tag{10}$$

where $\mathbf{a}(\lambda_0) = \mathbf{a}(\lambda_0, t = t_0)$ for some reference time t_0. For rigid body rotation, the result is exact and $\mathbf{R}(t)$ is the $(N \times N)$ Wigner rotation matrix for real spherical harmonics (e.g. Collado et al. 1989), which is implicitly a function of the inclination, obliquity, and rotation period of the star.

The equation for the specific intensity in the rest frame now reads

$$I(\lambda_0, \mathbf{x}, t) = \boldsymbol{\psi}^\top(\mathbf{x})\,\mathbf{R}(t)\,\mathbf{a}(\lambda_0)\ , \tag{11}$$

where it is clear that we have fully separated the spatial, temporal, and spectral terms. Inserting this into Equation (4) and integrating over the visible portion of the star, we arrive at the expression for the observed spectrum:

$$\begin{aligned} F(\lambda, t) &= \iint\limits_{\mathcal{S}(\mathbf{x})} \boldsymbol{\psi}^\top(\mathbf{x})\,\mathbf{R}(t)\,\mathbf{a}(\lambda_0) * \delta\big(\lambda_0 + D(\mathbf{x})\big)\mathrm{d}\mathcal{S}(\mathbf{x}) \\ &= \iint\limits_{\mathcal{S}(\mathbf{x})} \boldsymbol{\psi}^\top(\mathbf{x})\delta\big(\lambda_0 + D(\mathbf{x})\big)\mathrm{d}\mathcal{S}(\mathbf{x})\,\mathbf{R}(t)\ *\ \mathbf{a}(\lambda_0)\ , \end{aligned} \tag{12}$$

where we made use of the commutativity of the convolution operator and the fact that the integral is taken only over the spatial dimensions. Note, importantly, that the convolution operation above is implicitly a vector operation—i.e., the convolution is taken for each spherical harmonic term individually and then summed over all terms.

Finally, we can simplify Equation (12) by noting that the delta function in the integrand allows us to reduce the double integral to a line integral:

$$\begin{aligned} \iint\limits_{\mathcal{S}(\mathbf{x})} \boldsymbol{\psi}^\top(\mathbf{x})\delta\big(\lambda_0 + D(\mathbf{x})\big)\mathrm{d}\mathcal{S}(\mathbf{x}) \ &= \int\limits_{\mathcal{C}(\lambda,\mathbf{x})} \boldsymbol{\psi}^\top(\mathbf{x})\mathrm{d}\mathcal{C}(\lambda, \mathbf{x}) \\ &\equiv \boldsymbol{\kappa}^\top(\lambda)\ . \end{aligned} \tag{13}$$

where the path $\mathcal{C}(\lambda, \mathbf{x})$ corresponds to the set of all points on the visible disk where $\lambda_0 + D(\mathbf{x}) = 0$. We thus have

$$F(\lambda, t) = \boldsymbol{\kappa}^\top(\lambda)\,\mathbf{R}(t) * \mathbf{a}(\lambda_0)\ . \tag{14}$$

Equation (14) is the deconvolution of the observed spectrum into velocity terms, temporal terms, and spectral/spatial terms, respectively. In the next section, we will discuss how to efficiently solve the line integral in Equation (13).

3.2. *Computing the kernel* $\boldsymbol{\kappa}^\top$

In the case that the star's rotation is rigid (i.e., the effect of differential rotation on the spectrum is small) and other effects such as convective blueshift may be ignored, the radial velocity at any point on the surface is simply proportional to the distance to the axis of rotation. Without loss of generality, if we assume the axis of rotation lies along the $y - z$ plane, we have

$$\beta(\mathbf{x}) = \frac{\omega\, x}{c} \sin i \tag{15}$$

and

$$D(\mathbf{x}) = \frac{1}{2} \ln \left(\frac{1 + \frac{\omega\, x}{c} \sin i}{1 - \frac{\omega\, x}{c} \sin i} \right) \ , \tag{16}$$

where ω is the stellar angular velocity and i is the stellar inclination with respect to $\hat{y}$, where $i = 0^\circ$ corresponds to a pole-on orientation.

The region of integration $\mathcal{S}(\mathbf{x})$ in Equation (13) is simply the unit disk, so $\mathcal{C}(\lambda, \mathbf{x})$ is the set of all points that satisfy both $\lambda_0 + D(\mathbf{x}) = 0$ and $x^2 + y^2 \leq 1$. Solving Equation (16) for x, we find that the curves $\mathcal{C}(\lambda, \mathbf{x})$ are simply vertical lines on the unit disk given by

$$x = \left(\frac{c}{\omega \sin i} \right) \left(\frac{\mathrm{e}^{-2\lambda_0} - 1}{\mathrm{e}^{-2\lambda_0} + 1} \right) \ . \tag{17}$$

The integral in Equation (13) is therefore just the line integral of the spherical harmonic basis over y:

$$\boldsymbol{\kappa}^\top(\lambda) = \int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} \boldsymbol{\psi}^\top(x, y) \mathrm{d}y \ . \tag{18}$$

In Appendix A we derive an analytic solution to this integral and show how to recursively solve it for all terms in the spherical harmonic basis. Figure 1 shows the $\boldsymbol{\kappa}^\top$ basis computed from the formulae above and Equation (18) up to spherical harmonic degree $l_{\mathrm{max}} = 10$.

This completes our derivation of the analytic expression for the observed spectrum (Equation 14).

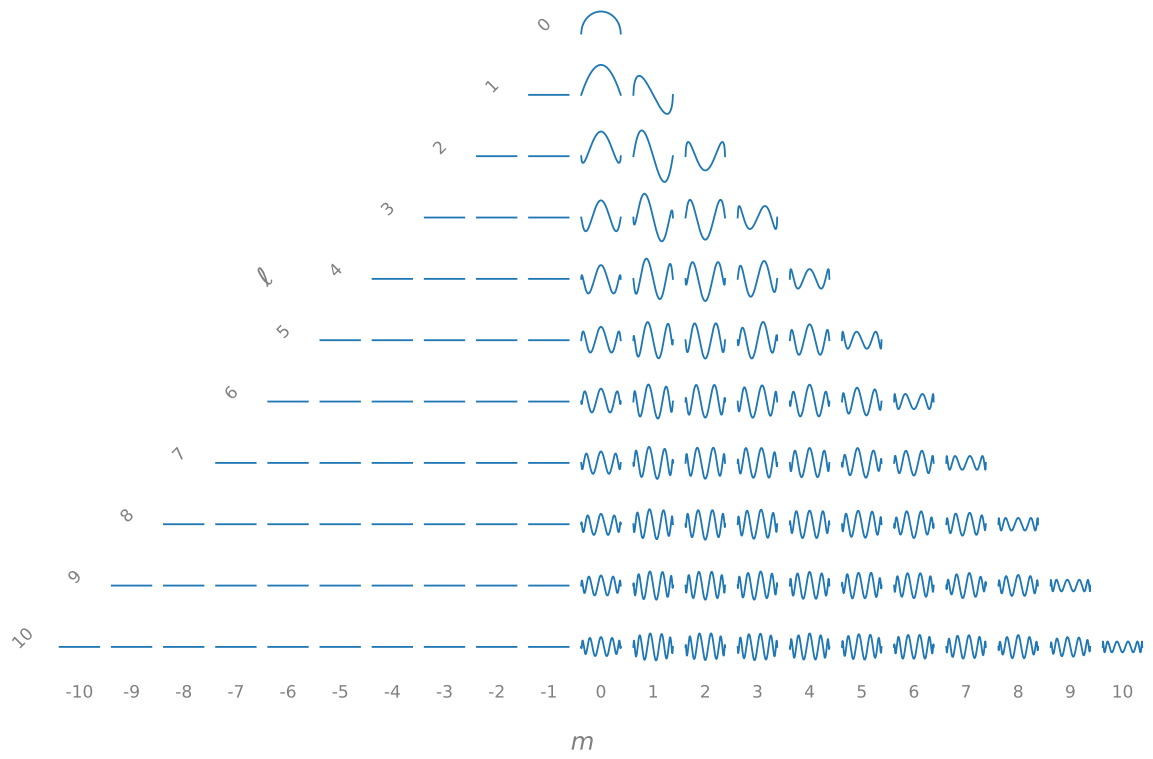

Figure 1. The Doppler $\boldsymbol{\kappa}^\top$ basis for a rigidly rotating star computed from Equation (18) and Equation (A17) up to spherical harmonic degree $l_{\rm max} = 10$. Rows correspond to the degree l and columns correspond to the order m. These functions encode the contribution of each spherical harmonic to the rotational broadening of features in the stellar spectrum. Because the rotational axis is chosen to be aligned with $\hat{y}$ (along the plane of the sky), none of the $m < 0$ harmonics contribute to the observed spectrum.

3.3. *Example*

Figure 2 shows an example application of the formulae derived in this section. We generated a synthetic stellar surface with a large, circular spot at a latitude of 30° and computed its spherical harmonic expansion up to $l_{\rm max} = 20$ in the form of the vector of coefficients $\mathbf{y}$. The stellar rest frame spectrum $I(\lambda_0)$ is taken to be a single Gaussian absorption line that is spatially constant everywhere save for a multiplicative factor equal to the local spot intensity, which is described by the spherical harmonic expansion. In the Figure, we rotated the star about an axis perpendicular to the line of sight (left panel) and computed the observed spectrum using Equation (14) (center panel). The right panel shows the difference between the observed spectrum and the spectrum when the spot is not in view. The effect of the spot on the absorption line is evident: a "bump" that travels from the blueshifted side of to the redshifted side of the line in sync with the change in longitude of the spot. This correspondence between line shape and spot location is the cornerstone of the Doppler imaging technique (compare to, e.g., Figures 1 and 4 in Vogt & Penrod 1983).

Figure 3 shows one of the spectra in Figure 2, this time computed using both our formulae (blue line) and the traditional technique of discretizing the stellar surface,

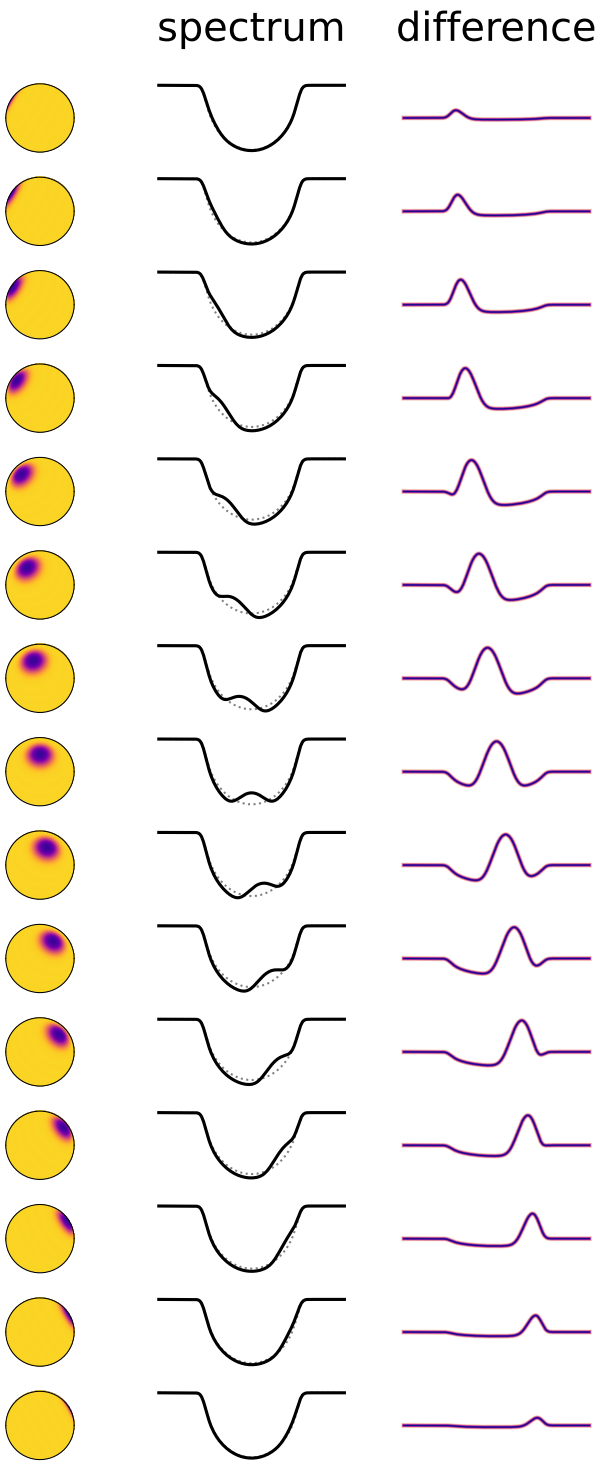

Figure 2. Time evolution of a Gaussian absorption line on a rigidly rotating, spotted star computed from the analytic formulae in §3.2. The stellar surface is modeled as a spherical harmonic expansion up to $l_{\rm max} = 20$, and the line shape is assumed to be the same everywhere; the spot simply downweights the local intensity at all wavelengths. As the spot rotates into view (left panel), the line shape changes slightly (center panel). The difference between the line when the spot is in view (solid) and when it is on the backside of the star (dotted) are shown in the right panel, where a Gaussian-like feature can be seen tracking the spot as it rotates from the blueshifted hemisphere to the redshifted hemisphere of the star.

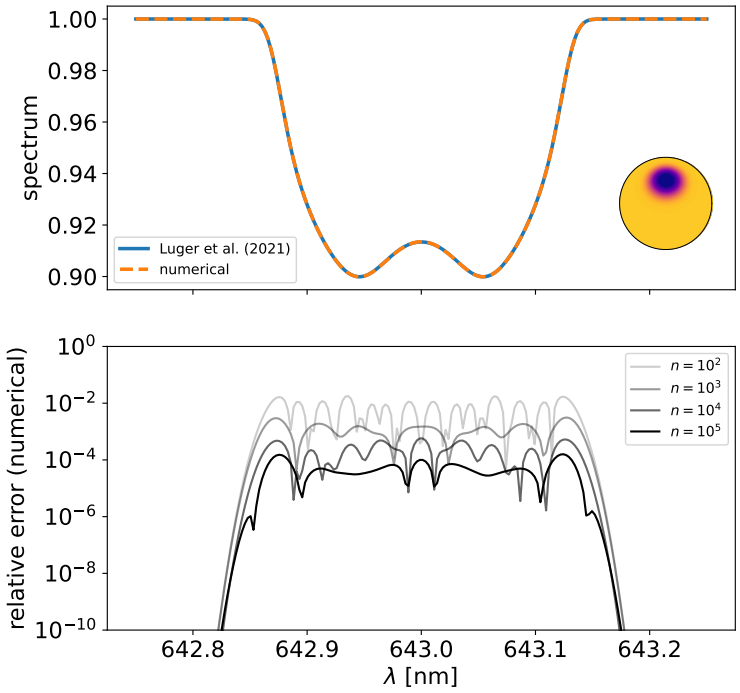

Figure 3. The observed spectrum when the spot is in view computed with our algorithm (blue) and numerically (orange). The problem setup is the same as that in Figure 2. The bottom panel shows the absolute value of the difference between the two spectra for different number of grid points in the numerical solution. As the number of grid points increases, the numerical solution approaches our solution. Note that here we assume the stellar surface can be expressed exactly as an $l_{\rm max} = 20$ spherical harmonic decomposition. When our algorithm is applied to real data, there will be errors due to the truncation of the expansion at finite spherical harmonic degree.

Doppler-shifting the spectrum in each grid cell according to the local radial velocity, interpolating back to a uniform wavelength grid, and summing over the spatial dimension (dotted orange line). We show the residuals in the lower panel, where it is clear that as the number of grid cells increases from 10^2 (light grey) to 10^5 (dark grey), the numerical solution approaches the solution obtained using our approach.

4. THE FORWARD PROBLEM

Although the convolution operator (Equation 5) is defined for continuous functions, the problem of Doppler imaging deals with discrete measurements of a spectrum in bins of log wavelength. Provided our wavelength grid is fine enough, we may therefore approximate Equation (5) with a discrete convolution. For discrete arrays $\mathbf{g}$ and $\mathbf{h}$, we have

$$\mathbf{g} * \mathbf{h} = \mathbf{T}(\mathbf{g})\,\mathbf{h} \tag{19}$$

where $\mathbf{T}$ is a Toeplitz matrix, a matrix whose diagonals are constant from top left to bottom right. In this case, the diagonals are constructed from the values of $\mathbf{g}$. If $\mathbf{g}$ has length L, element g_n is placed everywhere along the $k^{\rm th}$ diagonal of $\mathbf{T}$ for

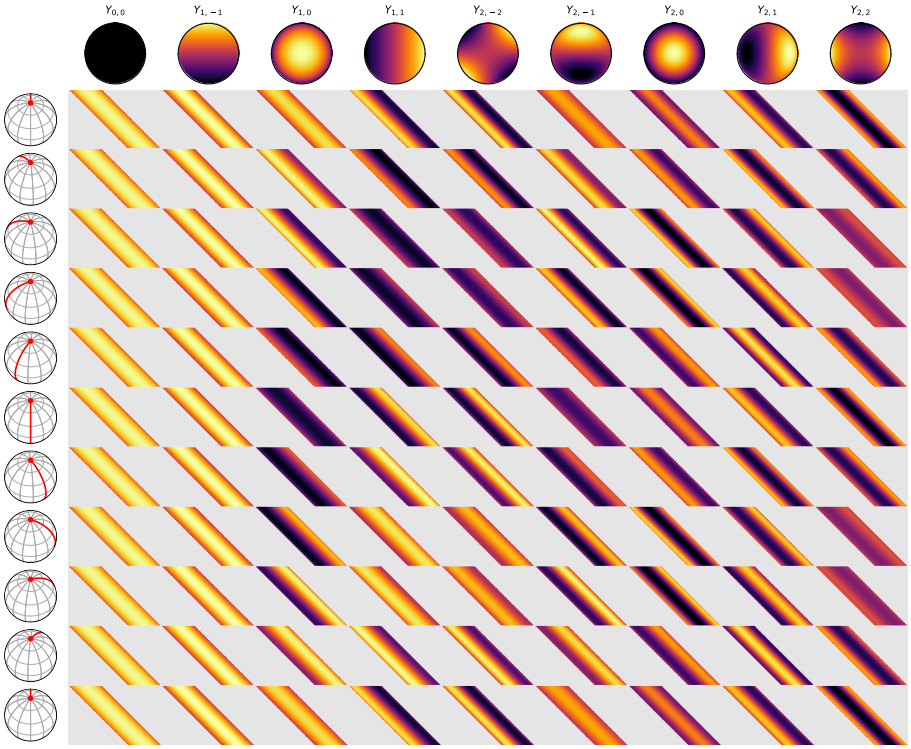

Figure 4. The Doppler design matrix $\mathbf{D}$ (Equation 22), constructed from a grid of Toeplitz matrices, each of shape $K \times K'$, where K is the number of wavelength bins in the broadened spectrum and $K' = K+W-1$ is the number of bins in the latent (template) spectrum, where W is the width of the convolution kernel. The N columns of $\mathbf{D}$ correspond to the Toeplitz matrices for each of the N spherical harmonics (shown at the top); the M rows correspond to the Toeplitz matrices rotated to each of the M stellar phases observed (indicated graphically to the left of $\mathbf{D}$). The color scale ranges from black (negative) to bright yellow (positive); light grey regions correspond to entries in the matrix that are identically zero.

$k = -L/2 + n$; all other entries of $\mathbf{T}$ are set to zero. Note that the matrix $\mathbf{T}$ need not be square; in fact, for our purposes, it is a $(K \times K')$ matrix, where K is the size of the output wavelength grid and $K' = K + W - 1$ is the size of the input wavelength grid (in the rest frame), where W is the width of the convolution kernel.

Given this formulation, we may re-write Equation (14) as a purely linear operation on $\mathbf{a}(\lambda_0)$:

$$\mathbf{f}_m = \mathbf{D}_m \, \mathbf{a} \ , \tag{20}$$

where $\mathbf{D}_m$ is a $(K \times NK')$ matrix constructed by horizontally concatenating the Toeplitz matrices for each of the N components of the convolution kernel $\boldsymbol{\kappa}^\top(\lambda)$ transformed by the rotation matrix $\mathbf{R}(t)$ evaluated at time $t = t_m$:

$$\mathbf{D}_m = \begin{pmatrix} \mathbf{T}(\boldsymbol{\kappa}_0^\top) & \mathbf{T}(\boldsymbol{\kappa}_1^\top) & \cdots & \mathbf{T}(\boldsymbol{\kappa}_N^\top) \end{pmatrix} (\mathbf{R}(t_m) \otimes \mathbf{I}_{K'}) \ . \tag{21}$$

where $\mathbf{I}_{K'}$ is the $(K' \times K')$ identity matrix and $\otimes$ denotes the Kronecker product. The $(K \times 1)$ vector $\mathbf{f}_m$ is the model for the flux observed at each wavelength at time $t = t_m$, and the $(NK' \times 1)$ vector $\mathbf{a}$ is the vector representation of the spatially-dependent

spectrum of the star. The latter is constructed by flattening the spherical harmonic expansion of the star such that the first K' terms in $\mathbf{a}$ correspond to the the values of $a_{0,0}$ at each wavelength λ_0, followed by the values of $a_{1,-1}$ at each wavelength, and so forth.

In general, our data will consist of observations made at several epochs, corresponding to different rotational phases of the star. If we concatenate all M spectra $\mathbf{f}_m$ into the $(MK \times 1)$ vector $\mathbf{f}$, we may write

$$\mathbf{f} = \mathbf{D}\,\mathbf{a} \ , \tag{22}$$

where $\mathbf{D}$ is the full $(MK \times NK')$ Doppler design matrix, constructed by vertically concatenating the individual matrices $\mathbf{D}_m$:

$$\mathbf{D} = \begin{pmatrix} \mathbf{D}_0 \\ \mathbf{D}_1 \\ \cdots \\ \mathbf{D}_m \end{pmatrix} \ . \tag{23}$$

Equation (22) represents the full linearization of the problem, where we have expressed the quantity we can observe, $\mathbf{f}$, as a linear operation on the quantity of interest, the spectral/spatial decomposition of the stellar surface, $\mathbf{a}$. Figure 4 shows an example of $\mathbf{D}$. Note, importantly, that this expression does not account for limb darkening or for the fact that spectra are often normalized to the continuum level; we account for these two effects in §6.1 and §6.2 below.

5. THE INVERSE PROBLEM

In the previous section we showed how to express the data vector $\mathbf{f}$, a series of spectra obtained at different epochs, as a purely linear operation on the map vector $\mathbf{a}$, the spherical harmonic decomposition of the specific intensity on the surface of the star. The advantage of this linearity is that, given suitable priors, it allows one to compute both the maximum likelihood solution for $\mathbf{a}$ and its uncertainty *analytically*. In particular, if one places a (multidimensional) Gaussian prior on $\mathbf{a}$ with mean $\boldsymbol{\mu}_\mathbf{a}$ and covariance $\boldsymbol{\Lambda}_\mathbf{a}$, the maximum *a posteriori* (MAP) solution for $\mathbf{a}$ is given by the solution to the L2 regularization problem (see, e.g., Luger et al. 2016):

$$\hat{\mathbf{a}} = \boldsymbol{\Sigma}_{\hat{\mathbf{a}}} \left(\mathbf{D}^\top \boldsymbol{\Sigma}_\mathbf{f}^{-1} \mathbf{f} + \boldsymbol{\Lambda}_\mathbf{a}^{-1} \boldsymbol{\mu}_\mathbf{a} \right) \ , \tag{24}$$

with posterior covariance given by

$$\boldsymbol{\Sigma}_{\hat{\mathbf{a}}} = \left(\mathbf{D}^\top \boldsymbol{\Sigma}_\mathbf{f}^{-1} \mathbf{D} + \boldsymbol{\Lambda}_\mathbf{a}^{-1} \right)^{-1} \ , \tag{25}$$

where $\boldsymbol{\Sigma}_\mathbf{f}$ is the data covariance matrix. These equations require the inversion of a few fairly large matrices, but as we will see later, it allows one to obtain the map of the star ($\hat{\mathbf{a}}$) and its uncertainty ($\boldsymbol{\Sigma}_{\hat{\mathbf{a}}}$) in under a second for a typical dataset.

There is, however, a catch: for most practical purposes, it is difficult to find a good Gaussian prior for $\mathbf{a}$. Usually we will have some prior information on what the spectrum of the star is, and perhaps some prior information on the distribution of surface features such as starspots. The problem is that even if these priors are Gaussian, the corresponding prior on $\mathbf{a}$ is *not*. To understand this, consider the case where the stellar rest frame spectrum is spatially constant, save for an amplitude that is spatially variable. We can then describe the rest frame spectrum by the $(K' \times 1)$ vector $\mathbf{s}$ and the amplitude by the $(N \times 1)$ vector $\mathbf{y}$ representing the spherical harmonic expansion of the intensity profile of the star. The vector $\mathbf{a}$ is then given by the (flattened) outer product of the two vectors:

$$\begin{aligned} \mathbf{a} &= \mathrm{vec}\,(\mathbf{A}) \\ &= \mathrm{vec}\left(\mathbf{s}\,\mathbf{y}^\top\right) \ , \end{aligned} \tag{26}$$

where vec denotes the vectorization operation, which in this case transforms the $(K' \times N)$ matrix $\mathbf{A}$ into the $(NK' \times 1)$ column vector $\mathbf{a}$ by vertically stacking all of its columns. In other words, the vector $\mathbf{a}$ consists of the set of K' intensities in each wavelength bin repeated N times, each time multiplied by the n^{th} spherical harmonic coefficient in the expansion of the surface intensity. Returning to our point about priors, note that all of the entries of $\mathbf{a}$ are the product of two independent random variables. Since the product of two random variables is generally non-Gaussian, even when the variables themselves are Gaussian,[3] so too is the prior on $\mathbf{a}$. This point makes Equation (24) tricky to use in practice, since a multivariate Gaussian is not generally the most suitable prior for $\mathbf{a}$.

That said, a Gaussian prior *is* typically suitable for both the spectrum $\mathbf{s}$ and the intensity $\mathbf{y}$ (just not their product). Spectral templates can provide us with a physically-motivated mean for the prior on $\mathbf{s}$, while the variance can be used to capture our degree of confidence in the model. Similarly, a multivariate Gaussian prior on the spherical harmonic coefficients $\mathbf{y}$ is flexible enough to capture prior information on the location and properties of starspots (Luger et al. 2021c). Given Gaussian priors on these two quantities, it is straightforward to show that because Equation (22) is linear in $\mathbf{a}$, it is also linear in both $\mathbf{s}$ and $\mathbf{y}$, so MAP solutions can be found analytically for $\mathbf{s}$ (conditioned on $\mathbf{y}$) and $\mathbf{y}$ (conditioned on $\mathbf{s}$). A solution to the full problem of determining both $\mathbf{s}$ and $\mathbf{y}$ can then be obtained by iteratively solving these two problems. We investigate this procedure below.

5.1. *Solution for the map*

Given a single rest-frame spectrum $\mathbf{s}$, the model for the flux vector may be written as

$$\mathbf{f} = \mathbf{D}\,\mathbf{S}\,\mathbf{y} \ , \tag{27}$$

[3] see, e.g., http://mathworld.wolfram.com/NormalProductDistribution.html

where

$$\mathbf{S} = \begin{pmatrix} \mathbf{s} & & & & & \\ & \mathbf{s} & & & & \\ & & \mathbf{s} & & & \\ & & & \mathbf{s} & & \\ & & & & \ddots & \end{pmatrix} \tag{28}$$

is a $(NK' \times N)$ block matrix constructed by repeating the column vector $\mathbf{s}$ N times in blocks along the main diagonal. Given prior mean $\boldsymbol{\mu}_\mathbf{y}$ and covariance $\boldsymbol{\Lambda}_\mathbf{y}$ on $\mathbf{y}$, the MAP solution is given by

$$\hat{\mathbf{y}} = \boldsymbol{\Sigma}_{\hat{\mathbf{y}}} \left(\mathbf{S}^\top \mathbf{D}^\top {\boldsymbol{\Sigma}_\mathbf{f}}^{-1} \mathbf{f} + {\boldsymbol{\Lambda}_\mathbf{y}}^{-1} \boldsymbol{\mu}_\mathbf{y} \right) \ , \tag{29}$$

with covariance

$$\boldsymbol{\Sigma}_{\hat{\mathbf{y}}} = \left(\mathbf{S}^\top \mathbf{D}^\top {\boldsymbol{\Sigma}_\mathbf{f}}^{-1} \mathbf{D} \mathbf{S} + {\boldsymbol{\Lambda}_\mathbf{y}}^{-1} \right)^{-1} \ . \tag{30}$$

If we choose a prior with diagonal covariance that is constant for each spherical harmonic order m,

$$\boldsymbol{\Lambda}_\mathbf{y} = \mathrm{diag} \left(\sigma_0^2 \quad \sigma_1^2 \quad \sigma_1^2 \quad \sigma_1^2 \quad \sigma_2^2 \quad \sigma_2^2 \quad \sigma_2^2 \quad \sigma_2^2 \quad \sigma_2^2 \quad \cdots \right) \ , \tag{31}$$

we are in fact imposing an isotropic prior with an angular power spectrum given by $C_l = \sigma_l^2$ (e.g., Baldi & Marinucci 2006). This is especially useful if the typical angular scale of features on the surface of the star is known or if a Gaussian prior can be placed on it. Alternatively, we may compute $\boldsymbol{\Lambda}_\mathbf{y}$ from the Gaussian process (GP) formalism in Luger et al. (2021c) to incorporate explicit prior information on the typical latitudes, sizes, and contrasts of spots. We discuss this point further in §9.2.

5.2. *Solution for the spectrum*

Conversely, given a spatial representation of the surface $\mathbf{y}$, the model for the flux vector may be written in terms of the spectrum $\mathbf{s}$ as

$$\mathbf{f} = \mathbf{D}\,\mathbf{Y}\,\mathbf{s} \ , \tag{32}$$

where

$$\mathbf{Y} = \begin{pmatrix} y_{0,0}\,\mathbf{I}_{K'} \\ y_{1,-1}\,\mathbf{I}_{K'} \\ y_{1,0}\,\mathbf{I}_{K'} \\ y_{1,1}\,\mathbf{I}_{K'} \\ \cdots \end{pmatrix} \tag{33}$$

is a $(NK' \times K')$ matrix constructed by vertically stacking N $(K' \times K')$ identity matrices, each multiplied by the corresponding coefficient of $\mathbf{y}$. Given prior mean $\boldsymbol{\mu}_\mathbf{s}$ and covariance $\boldsymbol{\Lambda}_\mathbf{s}$ on $\mathbf{s}$, the MAP solution is given by

$$\widehat{\mathbf{s}} = \boldsymbol{\Sigma}_{\widehat{\mathbf{s}}} \left(\mathbf{Y}^\top \mathbf{D}^\top {\boldsymbol{\Sigma}_\mathbf{f}}^{-1} \mathbf{f} + {\boldsymbol{\Lambda}_\mathbf{s}}^{-1} \boldsymbol{\mu}_\mathbf{s} \right) \ , \tag{34}$$

with covariance

$$\boldsymbol{\Sigma}_{\widehat{\mathbf{s}}} = \left(\mathbf{Y}^\top \mathbf{D}^\top {\boldsymbol{\Sigma}_\mathbf{f}}^{-1} \mathbf{D} \mathbf{Y} + {\boldsymbol{\Lambda}_\mathbf{s}}^{-1} \right)^{-1} \ . \tag{35}$$

5.3. *The full solution*

In the previous two sections we showed how, if the stellar rest-frame spectrum is spatially constant and known *a priori* (from, say, a template spectrum), the posterior distribution for the surface map $\mathbf{y}$ is *analytic* (Equations 29 and 30). Conversely, if the surface map is known, the posterior for the spectrum is also analytic (Equations 34 and 35). Both sets of solutions require Gaussian priors on the quantities of interest, but these are easily interpretable. A Gaussian prior on the spherical harmonic coefficients of the surface map corresponds to the angular power spectrum of the surface features, whereas a Gaussian prior on the spectrum can be used to correctly propagate uncertainties in the stellar spectrum from observational uncertainties on the stellar metallicity, temperature, etc.

Importantly, the framework developed here allows one to analytically compute not only the "optimal" solution to the Doppler imaging problem, but also its uncertainty. The problem of re-constructing surface images of stars from spectra and/or light curves is fundamentally ill-posed because of intrinsic degeneracies and a large null space (e.g., Cowan et al. 2017; Luger et al. 2019, 2021b). Coupled to uncertainties in properties like the stellar inclination, rotation period, and limb darkening coefficients, as well as the finite signal-to-noise ratio of the data, this makes the process of searching for the "optimal" solution flawed at best, and meaningless at worst. Access to the full covariance matrix of the solution is therefore essential to assessing our confidence in the various features of the inferred surface maps.

However, as we mentioned above, the Doppler imaging problem can be solved analytically for either the surface map or the spectrum, but not typically both. If neither are known exactly *a priori*, which is quite often the case, we can still find the MAP solution quickly by guessing at an initial spectrum $\mathbf{s}$ and iterating between optimizing for $\mathbf{y}$ and optimizing for $\mathbf{s}$. Note, importantly, that this bilinear problem is not generally convex: the likelihood space often has many local minima, which can in principle make it difficult to find the true solution. Nevertheless, given a good starting point for $\mathbf{s}$ and sufficient regularization on $\mathbf{y}$, we find that it typically converges to the correct solution; see §7.

Once the MAP solution has been found for both $\mathbf{s}$ and $\mathbf{y}$, the covariance of one conditioned on the other can be computed from the formulae above. However, this is necessarily a lower bound on the true covariance of the solution, since it neglects any

covariance *between* $\mathbf{s}$ and $\mathbf{y}$. To *correctly* account for this, one must resort to Monte Carlo methods or other sampling techniques. We discuss this further in §9.2.

6. FURTHER CONSIDERATIONS

Before we demonstrate our formalism in practice, we consider three modifications to our model that are often required when modeling real data. Below we discuss what to do when the relative brightness of the star across different observations is unknown (§6.1), how to account for the effects of limb darkening (§6.2), and how to model stars whose local spectra differ within and outside of starspots (§6.3).

6.1. *Unknown continuum normalization*

Consider the star depicted in Figure 2. When the spot is in view, absorption lines become distorted due to the suppression of either blueshifted or redshifted light. However, what is not shown in the figure is the fact that overall, the star also gets *dimmer*: the continuum level drops in the presence of the spot. Unfortunately, spectrographs are not usually designed to measure this, since the change in the star's overall brightness from one observation to the next is often dwarfed by (unknown) changes in the spectrograph stability, cloud cover, etc. For this reason, the spectra are usually normalized to have a fixed baseline continuum, and Equation (22) is no longer a good model for the data.

One option is to collect simultaneous photometry on the target in a band similar to the wavelength range observed, and use this light curve to (un-)normalize the spectral timeseries $\mathbf{f}$. However, when this is not an option, our model for $\mathbf{f}$ (Equation 22) must be amended:

$$\begin{aligned} \tilde{\mathbf{f}} &= \frac{\mathbf{f}}{\mathbf{B}\,\mathbf{a}} \\ &= \frac{\mathbf{D}\,\mathbf{a}}{\mathbf{B}\,\mathbf{a}} \,, \end{aligned} \tag{36}$$

where $\tilde{\mathbf{f}}$ is the normalized spectral timeseries, $\mathbf{B}$ is the baseline design matrix, and the vector division is performed elementwise. If c is an index in the observed spectrum corresponding to the continuum (at all epochs), the elements of $\mathbf{B}$ are computed from

$$B_{ij} = D_{i'j} \tag{37}$$

where

$$i' = c + \left\lfloor \frac{i}{K} \right\rfloor K \,. \tag{38}$$

In other words, the matrix $\mathbf{B}$ is constructed by repeating the c^{th} row of each epoch block $\mathbf{D}_m$ of $\mathbf{D}$ K times. Equation (36) is therefore equivalent to computing the M unnormalized spectra using Equation (22) and dividing each one by the value at index c.

Given this new equation, we must change how we approach the inverse problem (§5) if our data vector is the normalized spectral timeseries $\tilde{\mathbf{f}}$. When solving for the spectrum given the spherical harmonic vector (§5.2), we need only make the replacements $\mathbf{f} \rightarrow \tilde{\mathbf{f}} \circ (\mathbf{B}\,\mathbf{a})$ and $\mathbf{\Sigma_f} \rightarrow \mathbf{\Sigma}_{\tilde{\mathbf{f}}} \circ (\mathbf{B}\,\mathbf{a})^2$ in Equations (34) and (35), where $\circ$ denotes the elementwise product. However, when solving for the spherical harmonic vector given the spectrum (5.1), the problem is no longer linear in $\mathbf{y}$, as the vector of interest appears in both the numerator and the denominator:

$$\tilde{\mathbf{f}} = \frac{\mathbf{D}\,\mathbf{S}\,\mathbf{y}}{\mathbf{B}\,\mathbf{S}\,\mathbf{y}} \ . \tag{39}$$

Unfortunately, this equation does not admit a closed-form solution for $\mathbf{y}$. One can, however, still solve for the maximum *a posteriori* value of $\mathbf{y}$ in an efficient manner by iteratively solving for the baseline and then for the spherical harmonic coefficients. We describe a procedure for this in §7 below.

6.2. *Limb darkening*

Thus far we have avoided mention of limb darkening, whose effect on the observed spectrum can often be significant. A proper treatment of limb darkening requires solution of radiative transfer equations and should account for the temperature and composition gradients in the stellar atmosphere. However, in the spirit of data-driven inference, we can approximate the effect of limb darkening by weighting the spatial component of our map by the limb darkening profile of the star, which can either be imposed or learned from the data. For a polynomial limb darkening law of the form

$$\frac{I(\mu)}{I(\mu=1)} = 1 - \sum_{n=1}^{n_{\max}} u_n (1-\mu)^n \ , \tag{40}$$

where $\mu = z = \sqrt{1-x^2-y^2}$ is the radial coordinate on the projected disk and u_n is a limb darkening coefficient, the effect of limb darkening on the stellar map can be expressed exactly as a linear operation on the spherical harmonic coefficient vector (Luger et al. 2019; Agol et al. 2020). This can be understood by noting that all terms in Equation (40) are strictly polynomials in x, y, and z, all of which can be expressed exactly as sums of spherical harmonics (Luger et al. 2019). When weighting the surface intensity by the limb darkening profile, the resulting intensity is simply a product of spherical harmonics, which is itself a sum of spherical harmonics. Given a limb darkening law of degree $n_{\max}$, we can always construct a matrix $\mathbf{L}$ that transforms a spherical harmonic vector $\mathbf{y}$ of degree $l_{\max}$ to a limb-darkened spherical harmonic vector $\mathbf{y}'$ of degree $l_{\max} + n_{\max}$. As an example, consider a map of degree $l_{\max} = 1$ and

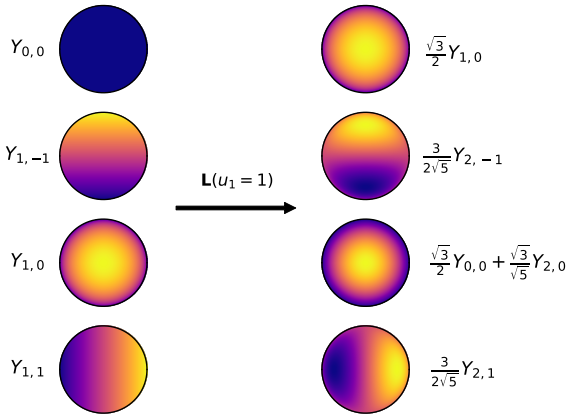

Figure 5. Illustration of the limb darkening transformation given by Equation (41) for a linear limb darkening law with $u_1 = 1$. Each row shows a spherical harmonic before (left) and after (right) transformation by $\mathbf{L}$.

the linear limb darkening law ($n_{\max} = 1$). The transformation matrix from $\mathbf{y}$ to $\mathbf{y}'$ is

$$\mathbf{L} = \frac{1}{1 - \frac{u_1}{3}} \begin{pmatrix} 1 - u_1 & 0 & \frac{u_1}{\sqrt{3}} & 0 \\ 0 & 1 - u_1 & 0 & 0 \\ \frac{u_1}{\sqrt{3}} & 0 & 1 - u_1 & 0 \\ 0 & 0 & 0 & 1 - u_1 \\ 0 & 0 & 0 & 0 \\ 0 & \frac{u_1}{\sqrt{5}} & 0 & 0 \\ 0 & 0 & \frac{2u_1}{\sqrt{15}} & 0 \\ 0 & 0 & 0 & \frac{u_1}{\sqrt{5}} \\ 0 & 0 & 0 & 0 \end{pmatrix} . \tag{41}$$

The columns of $\mathbf{L}$ are constructed from the coefficient vectors of each transformed spherical harmonic, which are in turn computed by multiplying each spherical harmonic by the spherical harmonic decomposition of the particular limb darkening law. Figure 5 shows this transformation with $u_1 = 1$ applied to each of the first four spherical harmonics.

Now, to incorporate limb darkening into our model, we must modify how we compute the Doppler design matrix $\mathbf{D}$ (Equation 23 and Figure 4). Recall that this matrix is constructed by vertically stacking the submatrices $\mathbf{D}_m$ given by Equation (21), which we now compute as

$$\mathbf{D}_m = \begin{pmatrix} \mathbf{T}(\boldsymbol{\kappa}_0^\top) & \mathbf{T}(\boldsymbol{\kappa}_1^\top) & \cdots & \mathbf{T}(\boldsymbol{\kappa}_{N'}^\top) \end{pmatrix} (\mathbf{L}\,\mathbf{R}(t_m) \otimes \mathbf{I}_{K'}) \,, \tag{42}$$

where we have replaced the rotation matrix $\mathbf{R}(t_m)$ with the linear transformation $\mathbf{L}\,\mathbf{R}(t_m)$, which first rotates the input spherical harmonic coefficient vector to the correct phase then limb darkens it. Note also that the terms in the convolution kernel

$\boldsymbol{\kappa}^\top$ are now indexed from 0 to $N' = (l_{\max} + n_{\max} + 1)^2$, where $n_{\max}$ is the degree of the limb darkening operator.

Note that it is also possible to model wavelength-dependent limb darkening in this fashion. The procedure is similar, except that the limb darkening coefficients u_n are now functions of λ, and each element of the matrix $\mathbf{L}$ is now a vector of length K'. The submatrices $\mathbf{D}_m$ may be written as

$$\mathbf{D}_m = \begin{pmatrix} \mathbf{T}(\boldsymbol{\kappa}_0^\top) & \mathbf{T}(\boldsymbol{\kappa}_1^\top) & \cdots & \mathbf{T}(\boldsymbol{\kappa}_{N'}^\top) \end{pmatrix} \begin{pmatrix} \mathrm{diag}(\mathbf{L}_{0,0}) & \mathrm{diag}(\mathbf{L}_{0,1}) & \cdots & \mathrm{diag}(\mathbf{L}_{0,N}) \\ \mathrm{diag}(\mathbf{L}_{1,0}) & \mathrm{diag}(\mathbf{L}_{1,1}) & \cdots & \mathrm{diag}(\mathbf{L}_{1,N}) \\ \cdots & \cdots & \ddots & \cdots \\ \mathrm{diag}(\mathbf{L}_{N',0}) & \mathrm{diag}(\mathbf{L}_{N',1}) & \cdots & \mathrm{diag}(\mathbf{L}_{N',N}) \end{pmatrix} (\mathbf{R}(t_m) \otimes \mathbf{I}_{K'}) \quad , \tag{43}$$

where $\mathrm{diag}(\mathbf{v})$ denotes the $(K' \times K')$ diagonal matrix constructed by placing the vector $\mathbf{v}$ along the main diagonal.

6.3. *Multiple spectral components*

In our discussion so far we have focused on the particular case where the spectral/spatial decomposition of the stellar surface can be expressed as the outer product of two vectors (Equation 26). As we mentioned previously, this corresponds to the case where the spectrum is constant across the surface of the star except for a spatially-variable amplitude: i.e., spots have the same spectrum as the rest of the star, but emit less flux overall. However, most of the formalism developed here also applies to the more general case where the spectrum at a particular point on the surface of the star is a linear combination of several different spectral components. In this case, the map vector $\mathbf{a}$ may be written as the sum over J spectra $\mathbf{s}_j$, each of which is dotted into a different vector of spherical harmonic coefficients $\mathbf{y}_j^\top$:

$$\mathbf{a} = \sum_{j=0}^{J-1} \mathrm{vec}\left(\mathbf{s}_j \, \mathbf{y}_j^\top\right) \tag{44}$$

$$= \mathrm{vec}\left(\mathbb{S} \, \mathbb{Y}^\top\right) \quad , \tag{45}$$

where $\mathbb{S}$ is the $(K' \times J)$ matrix whose columns are the J spectral components and $\mathbb{Y}^\top$ is the $(J \times N)$ matrix whose rows are the corresponding J spherical harmonic coefficient vectors.

The model remains linear in both the spherical harmonic coefficients and the spectral components, and the procedures outlined in §5 may still be followed to solve the inverse problem, provided we unroll these matrices into vectors, letting $\mathbf{s} = \mathrm{vec}(\mathbb{S})$ and $\mathbf{y} = \mathrm{vec}(\mathbb{Y})$. We present an example of inference on a two-component stellar spectrum in §7.4 below.

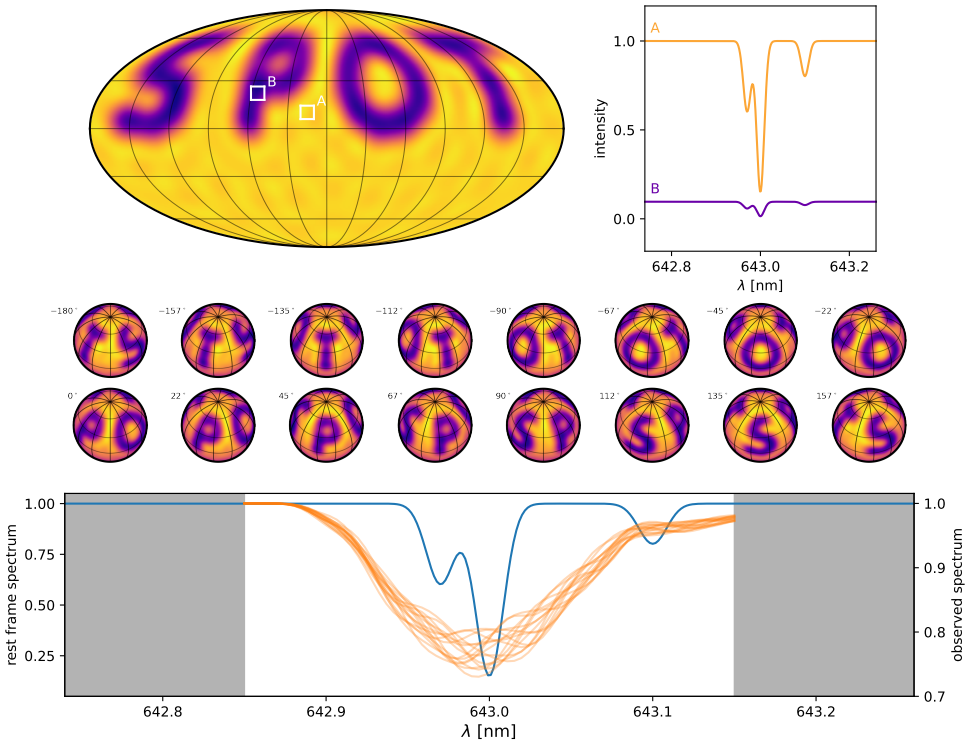

Figure 6. The basic setup for the `SPOT` problem. We compute the $l_{\rm max} = 15$ spherical harmonic expansion of a map where the word "SPOT" has been placed across the northern hemisphere (top left). Each point on the surface is given the same spectrum, composed of a sum of synthetic Gaussian absorption lines (blue curve in the bottom panel). The specific intensity at each point on the surface is equal to this spectrum weighted by the intensity of the map, which is about ten percent within the spot (point labeled `B`) and close to unity outside the spot (point labeled `A`). To generate the synthetic dataset (orange curves in the bottom panel), the star is inclined at 40° with respect to the line of sight and the disk-integrated spectrum is computed at 16 equally-spaced phases between −180° and 180°, assuming an equatorial velocity $v = 60$ km/s (middle panel). The grey regions indicate wavelength regions outside the observation window, but whose features can still contribute to the observed spectrum due to rotational broadening.

7. APPLICATION TO A SPOTTED STAR

In this section we apply the formalism developed in §5 and §6 to a synthetic dataset of a rapidly rotating spotted star. This section is motivated by the examples in Vogt et al. (1987), who generated rotationally-broadened spectra of a rapidly-rotating star with the letters "VOGT" painted across the northern hemisphere. In that paper, the authors generated synthetic line profiles of the Fe I λ6430 absorption line with equivalent width ∼10 km/s and from these computed rotationally broadened spectra for a star at an inclination $i = 40°$ rotating at projected velocity $v \sin i = 40$ km/s. The spectra were computed at 16 equally spaced rotational phases and downsampled to 32 wavelength bins per epoch, for a total of 512 data points.

We repeat the same experiment here using our new formalism, except we instead generate a stellar surface with the word "SPOT" written across the northern hemisphere (top left panel of Figure 6). We do this by computing the $l_{\rm max} = 15$

spherical harmonic decomposition of an image of the word, generating a set of $N = (l_{\rm max} + 1)^2 = 256$ spherical harmonic coefficients $\mathbf{y}$. We assume an equatorial velocity $v = 60$ km/s and an inclination $i = 40°$, such that $v \sin i = 38.6$ km/s (compare to $i = 40°$ and $v \sin i = 40$ km/s in Vogt et al. 1987).

Our input spectrum $\mathbf{s}$ was generated at extremely high resolution ($R \equiv \frac{\lambda}{\Delta\lambda} = 750,000$) centered on the Fe I λ6430 absorption line (blue curve in the bottom panel of Figure 6). The spectrum consists of a (purely synthetic) Gaussian absorption line of fractional depth 50% centered at $\lambda = 643$ nm with FWHM of 0.02 nm, or, in velocity space, $\Delta v = \text{FWHM} \times c/\lambda \approx 10$ km/s, equal to the value used in Vogt et al. (1987). Unlike Vogt et al. (1987), we add two additional small absorption lines, one of which is blended with the central absorption line. In total, the rest frame spectrum $\mathbf{s}$ consists of $K = 351$ wavelength bins in the range 642.85 nm $\leq \lambda \leq$ 643.15 nm plus an additional $W - 1 = 252$ bins of padding (126 on either side) to eliminate edge effects due to the convolution, for a total of $K' = 603$ wavelength bins. Finally, we linearly interpolate this spectrum to a wavelength grid uniform in log space to obtain the vector $\mathbf{s}$ evaluated at each λ.

Given $\mathbf{y}$ and $\mathbf{s}$, we use Equations (14) and (26) to compute the data vector $\mathbf{f}$, which we downsample to a grid of $K_{\rm obs} = 70$ wavelength bins in the range 642.85 nm $\leq \lambda \leq$ 643.15 nm, corresponding to a resolution of $R = 150,000$. We compute $\mathbf{f}$ at $M = 16$ equally spaced rotational phases; these phases are shown in orthographic projection (as an observer would see the star) in the center panel of Figure 6. The synthetic spectra for each phase are shown as the orange curves in the bottom panel of the figure (note the different scaling, as all spectral lines become shallower due to the broadening). We assume a wavelength-independent, quadratic limb darkening law with $u_1 = 0.5$ and $u_2 = 0.25$. Finally, we divide the flux by the baseline at each epoch (Equation 36) and add random Gaussian noise with covariance matrix $\mathbf{\Sigma_f} = \sigma_{\mathbf{f}}^2 \mathbf{I}$, with $\sigma_{\mathbf{f}} = 2 \times 10^{-4}$ to simulate a very high signal-to-noise ratio (SNR) spectrum. The typical fractional line depth in the observed spectrum is 0.2, so our choice of $\sigma_{\mathbf{f}}$ corresponds to SNR $\sim 10^3$ in each of the 70 wavelength bins of $\mathbf{f}$.

In the following sections we discuss various approaches to solving the Doppler imaging problem for this dataset. Unless otherwise noted, in all experiments we adopt a Gaussian prior for the spherical harmonic coefficients $\mathbf{y}$ with mean $\boldsymbol{\mu}_{\mathbf{y}} = (1 \quad \mathbf{0})$ and covariance $\mathbf{\Lambda_y} = 10^{-4}\mathbf{I}$, corresponding to a flat power spectrum prior. When solving for the rest frame spectrum, we also place a Gaussian prior on $\mathbf{s}$ with unit mean and covariance $\mathbf{\Lambda_s} = 10^{-3}\mathbf{I}$. These choices are arbitrary and primarily based on trial-and-error given knowledge of the true solutions. We discuss strategies for choosing priors for real datasets in §9.2 below.

7.1. *Known baseline and spectrum*

In our first experiment, we attempt to infer the stellar map assuming we have perfect knowledge of the rest frame spectrum $\mathbf{s}$ and the baseline, as well as the stellar

inclination, rotation period, and limb darkening coefficients. This is analogous to the case where a good physical model exists for the stellar spectrum and simultaneous photometry is available to constrain the baseline variations. As we showed in §5.1, the posterior distribution for the map coefficients $\mathbf{y}$ is analytic in this case, with mean given by Equation (29) and covariance given by Equation (30). Note that we must first scale the data vector $\tilde{\mathbf{f}}$ by the baseline (which in this case is known) to obtain $\mathbf{f}$.

Figure 7 shows the results. The left hand panel shows the inferred map alongside the data (black points) and best fit model (orange curves) at each of the 16 epochs. The panels on the right show the true map (top), the inferred map (center) and the posterior pointwise uncertainty (bottom) on a Mollweide grid. The uncertainty on the vector of pixels $\mathbf{p}$ where we evaluate the map is equal to the square root of the diagonal of the posterior covariance matrix for the pixel values, computed from

$$\boldsymbol{\Sigma}_{\mathbf{p}} = \mathbf{P}\boldsymbol{\Sigma}_{\hat{\mathbf{y}}}\mathbf{P}^{\top} \tag{46}$$

where $\mathbf{P}$ is the change-of-basis matrix from spherical harmonic coefficients to pixels on the grid.

The inferred map very closely matches the true map, aside from a few small artifacts in the southern hemisphere. Note, importantly, that regions south of the equator are almost never in view to the observer (below $-40°$, they are *never* in view); our inferred map is therefore informed primarily by our prior in southern latitudes. This is well captured by the uncertainty (bottom panel), which increases south of the equator, reaching a maximum at $-30°$, below which the uncertainty is constant at about 13%, a value associated with our particular choice of prior.

7.2. *Known spectrum, unknown baseline*

We next investigate the case where the spectrum is known exactly, but the baseline is not. This corresponds to the case discussed in §6.1, in which each spectrum is normalized to its continuum, and the true difference in continuum levels across spectral epochs is not precisely known. This is likely to be the most common application of the methodology presented in this work.

Our approach is to iteratively solve Equation (39) by re-expressing it as

$$\tilde{\mathbf{f}} = \frac{\mathbf{D}\,\mathbf{S}\,\mathbf{y}}{\mathbf{b}} \ . \tag{47}$$

where the baseline $\mathbf{b}$ is given by

$$\mathbf{b} = \mathbf{B}\,\mathbf{S}\,\mathbf{y} \ . \tag{48}$$

We adopt a simple iterative scheme where we solve the L2 problem for $\mathbf{y}$ with $\mathbf{b}$ fixed, then update the value of $\mathbf{b}$ and repeat. Because of the high SNR of our data, we find that it is necessary to employ a simple tempering procedure, in which we artificially inflate the variance of the data by a factor T, which we slowly decrease

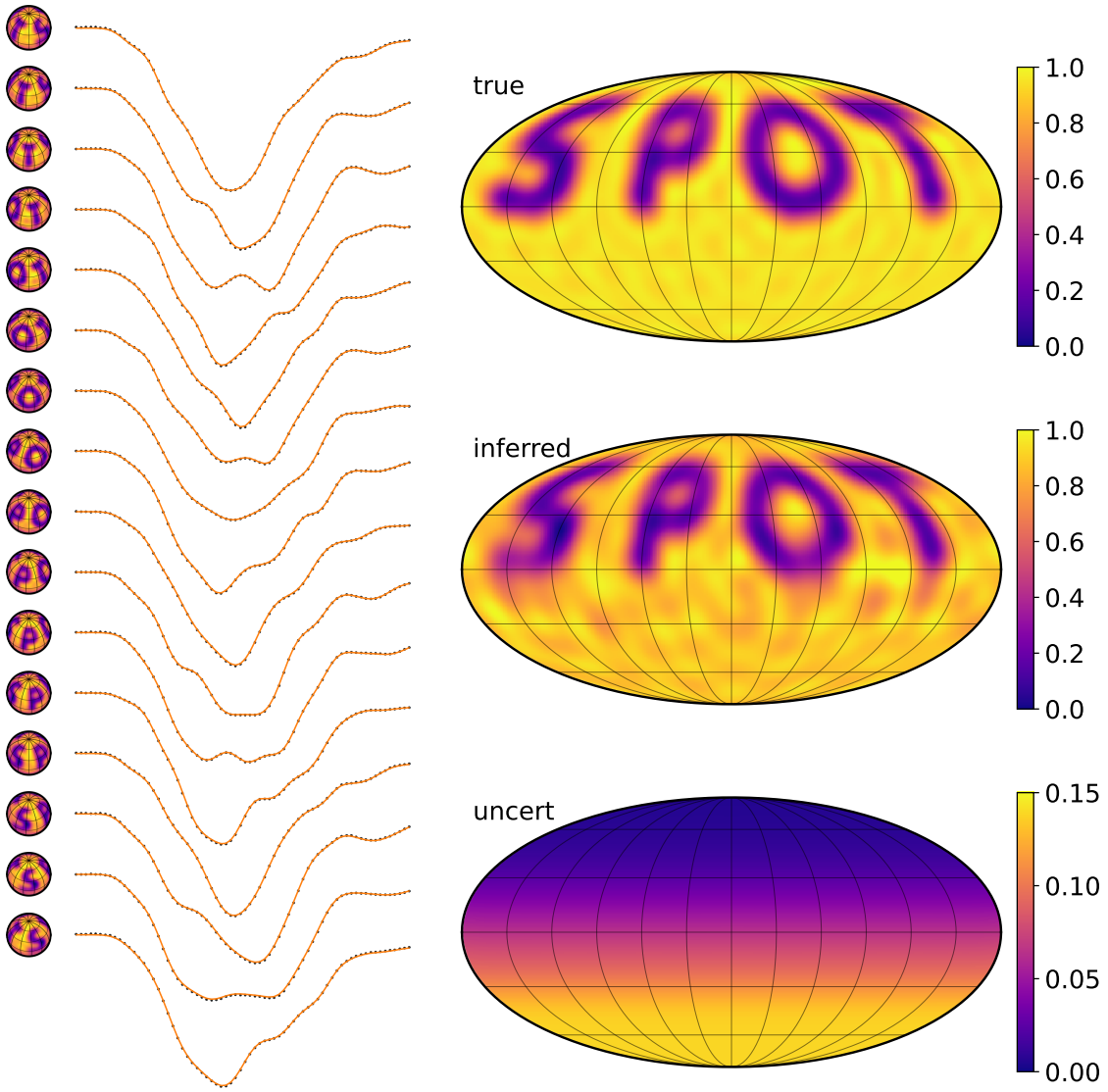

Figure 7. The solution to the `SPOT` problem at high SNR when both the rest frame spectrum and the baseline are known. The panel on the left shows the inferred map at each of the 16 epochs, alongside the observed spectrum (black points) and the best-fit model (orange lines). The Mollweide projection of the original map is shown at the top right; the inferred map is displayed below it. The bottom panel shows the posterior uncertainty on the inferred intensity, ranging from close to zero in the northern hemisphere to $\sigma_{\rm prior} \approx 13\%$ below a latitude of $-30°$. The latter is the uncertainty associated with our particular choice of prior, as the southernmost latitudes of the star are never visible to the observer. The optimization step in this experiment ran in under 1 second on a two-core virtual machine.

with each step until $T = 1$. This has the effect of smoothing out the likelihood surface, effectively preventing the solver from getting stuck in local minima. For this particular experiment, we set $\ln T_0 = 2$ and $\Delta \log T = -0.04$, for a total of 50 steps of the bilinear solver. We also add a small variance term $\sigma_b^2 = 10^{-2}$ to all entries of the data covariance $\mathbf{\Sigma_f}$; this has the effect of marginalizing over a small additive model term, which we find improves the convergence of the algorithm.

Our results are shown in Figure 8. As in the case where the continuum level was known exactly (Figure 7), we recover the stellar surface map with high fidelity in the northern hemisphere. However, there is now a strong dipolar artifact; the background intensity level is higher in the northern hemisphere than in the southern hemisphere. This is due to a degeneracy involving the $Y_{1,-1}$ spherical harmonic ( ), which, because a large portion of the southern hemisphere is never in view, cannot be constrained by the data.

Interestingly, while the map uncertainty still increases toward the south pole (bottom panel), its magnitude is smaller than that in Figure 7. This is counter-intuitive, since the continuum-normalized spectrum should contain *less* information than the unnormalized one. Indeed, the uncertainties shown in the figure are conditioned on our estimate of the baseline, so they are actually a strict *lower bound* on the map uncertainty. In order to constrain the true uncertainty on the solution to this bilinear problem, we must resort to approximate sampling methods such as Markov Chain Monte Carlo (MCMC). We discuss this further in §9.2.

7.3. *Unknown spectrum, unknown baseline*

Suppose we are handed a series of spectra in a wavelength regime where we do not have a good physical model for the line generation mechanisms, or a wavelength regime with many line blends for which we do not have reliable templates. Or perhaps the spectra are of a star whose rotational broadening does not dominate over other broadening mechanisms, in which case our template might not be accurate enough to allow us to tease out the small Doppler signals. In these cases, we may want to relax the assumption that we know the rest frame spectrum $\mathbf{s}$ exactly, and instead try to *learn* it from the data.

Given the same data as before, we now try to learn both the stellar map and the stellar spectrum, conditioned only on the wide Gaussian priors discussed above. We must follow the iterative procedure presented in §5.3, adjusting for the fact that we do not know the true baseline, either (§6.1).

Because the problem is generally not convex, we need a good starting guess for either the spectrum or the map coefficients. We note that we can very crudely approximate the rest frame spectrum by deconvolving the average observed spectrum with the first term of the convolution kernel $\boldsymbol{\kappa}^\top$ ( ); we do this by solving the linear problem in Equation (19) with aggressive regularization. In practice, this is a fairly poor approximation, but we find that it suffices as an initial guess. We then

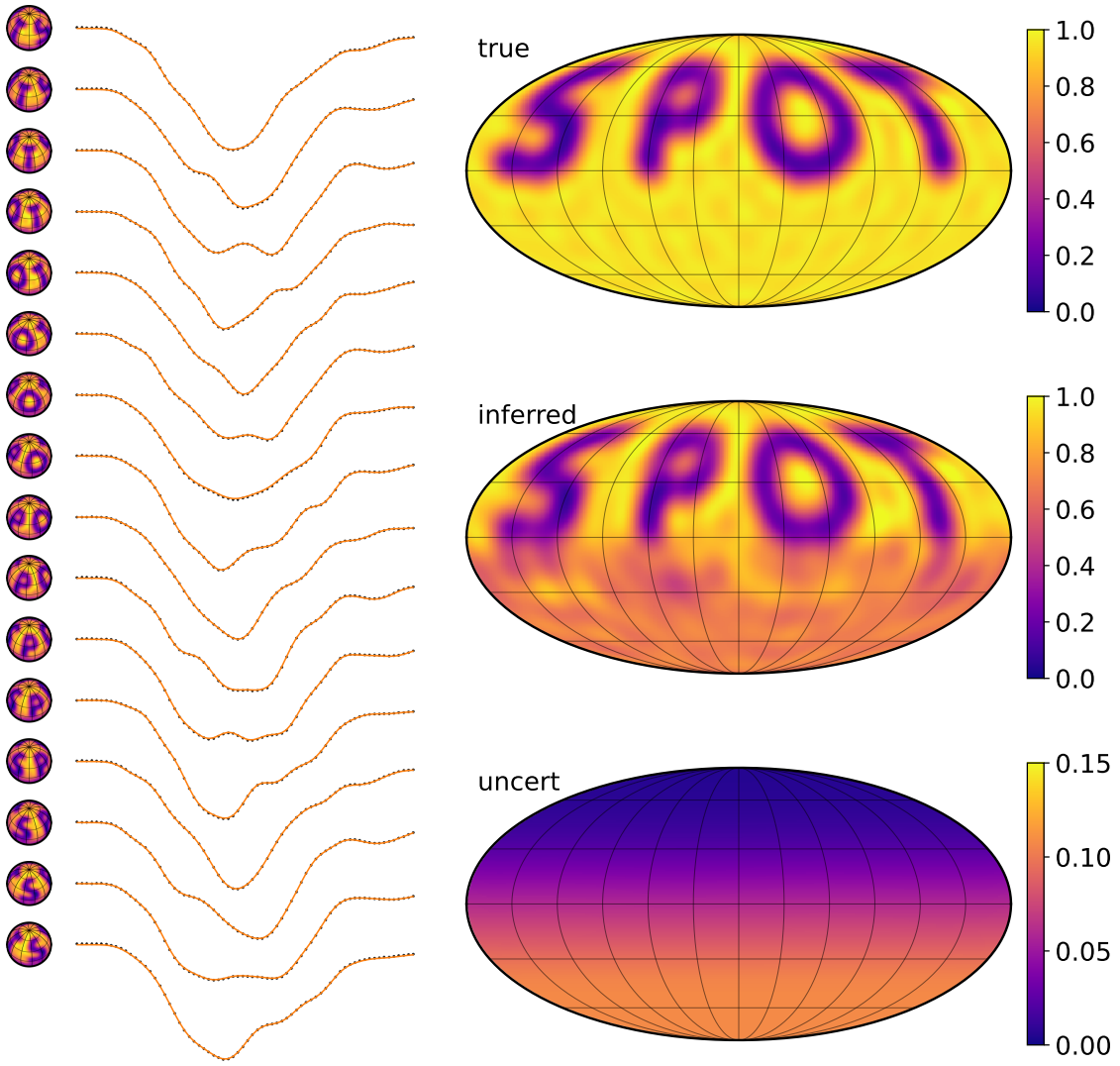

Figure 8. Similar to Figure 7, but assuming the input spectra are continuum-normalized. We are able to infer the surface map just as well as in the case where the true continuum level is known (Figure 7). As before, regions south of the equator are almost never in view, so that portion of our inferred map is dominated by the prior. Note that the uncertainty shown (lower panel) is a strict lower bound on the true uncertainty, as it is conditioned on our point estimate of the baseline level. The optimization step in this experiment ran in about 1 second.

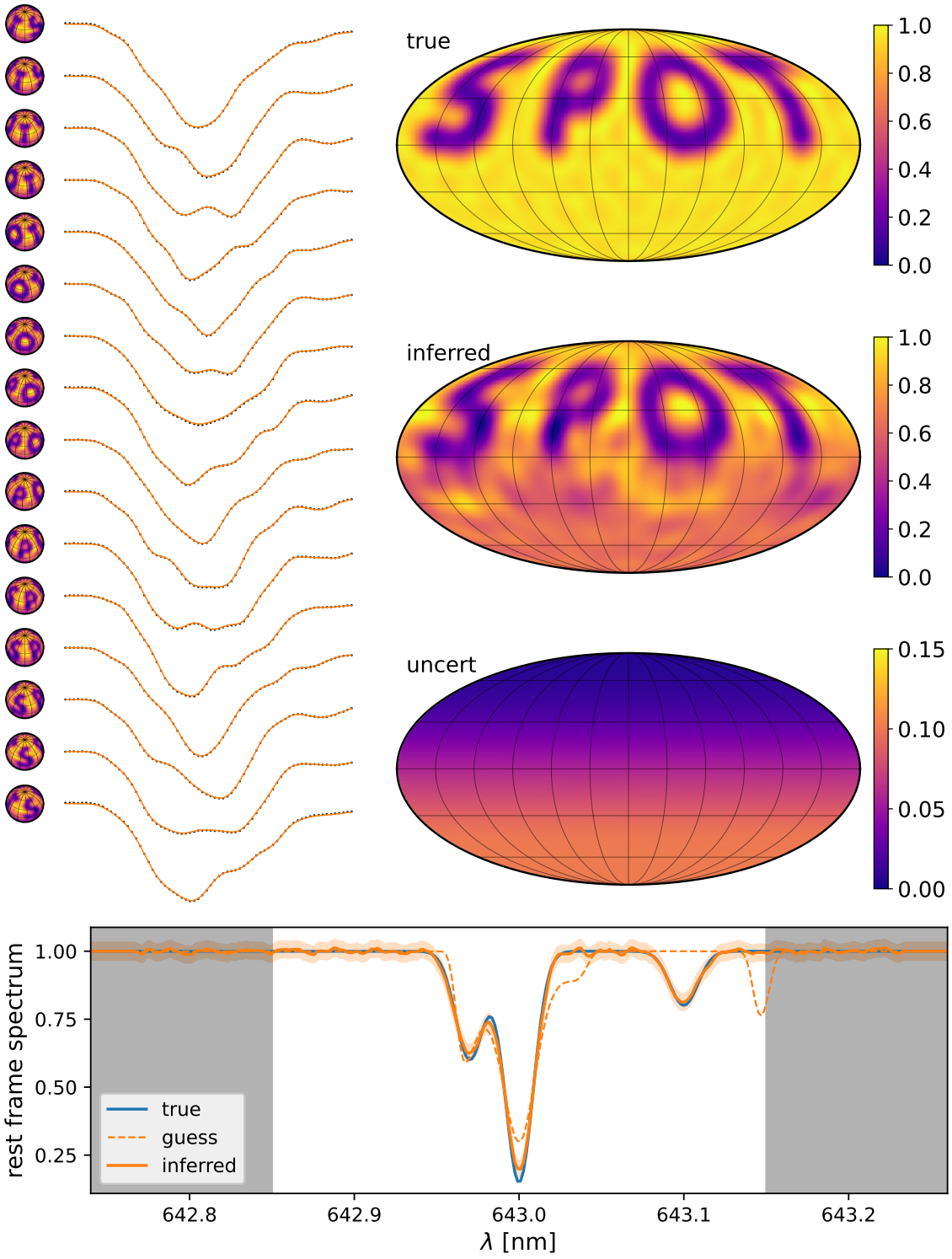

Figure 9. Similar to Figure 7, but assuming we have no information about either the spectrum, the baseline, or the map coefficients. The bottommost panel now shows the true rest frame spectrum (blue line), the initial guess based on a deconvolution (dashed orange line), and the inferred spectrum (orange line, with shaded 1σ uncertainty). The optimization step in this experiment ran in about 30 seconds.

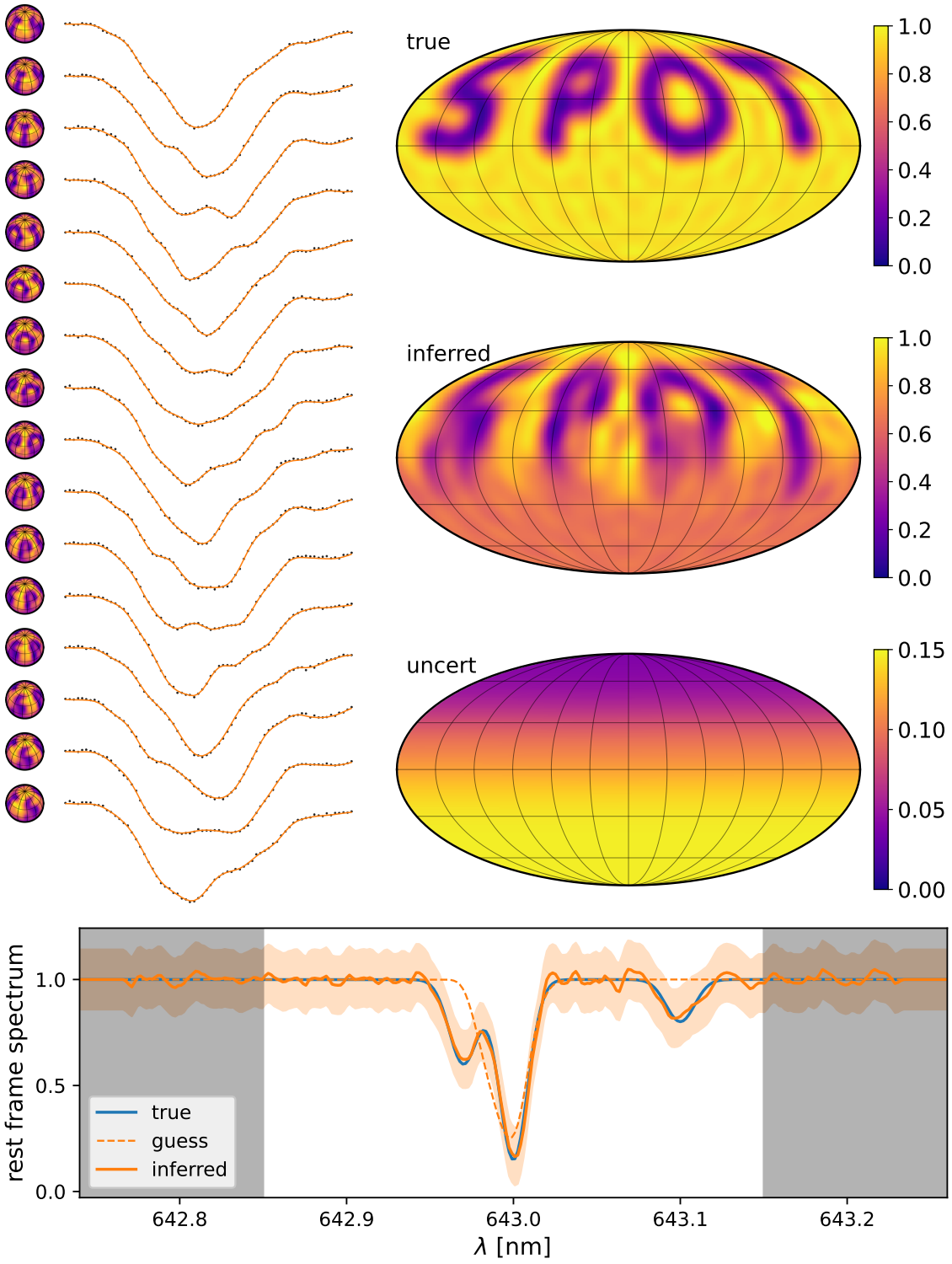

Figure 10. Same as Figure 9, but for 10× lower SNR. We still recover the `SPOT` map and the input spectrum, albeit with higher uncertainty, as expected. Once again, the optimization step in this experiment ran in about 30 seconds.

iterate between solving for $\mathbf{y}$ (§5.1) and solving for $\mathbf{s}$ (§5.2), followed by a step in which we update our estimate of the baseline. We adopt the same tempering scheme discussed above.

Figure 9 shows the results; the top portion of the figure is the same as in Figure 8. A panel has been added at the bottom showing the true rest frame spectrum (blue), the initial guess based on the deconvolution (dashed orange), and the inferred spectrum with 1-σ uncertainty shading (orange). We recover the `SPOT` feature as before, albeit with more noise, particularly south of the equator. Once again, our map uncertainties are somewhat underestimated, as expected, since they are computed conditioned on the optimum value of the baseline and the spectrum. On the other hand, we recover the spectrum exceedingly well, with low uncertainty and virtually no bias anywhere.

To test our technique on more realistic data, we repeat the experiment on data with 10 times higher uncertainty, setting $\sigma_{\mathbf{f}} = 2 \times 10^{-3}$. After some tuning, we find it necessary to increase our prior variances slightly to get good fits to the data: we set $\mathbf{\Lambda_y} = 2 \times 10^{-4}\mathbf{I}$ and $\mathbf{\Lambda_s} = 2 \times 10^{-2}\mathbf{I}$. Figure 10 shows the results. We again correctly infer the `SPOT` feature and the rest frame spectrum, although the surface map is considerably degraded, particularly at high spatial frequencies.

7.4. *Spatially-variable spectrum*

In our final experiment, we relax the assumption that the stellar rest frame spectrum is spatially constant, allowing for different spectral features inside and outside of the `SPOT`. This is the case for, say, chemically peculiar Ap stars, whose abundance of certain elements is significantly different within spotted regions on their surfaces. It is also the case in general for spots that are significantly cooler than the rest of the photosphere, where the relative abundance of different ionization states or molecular features may be different than in the hotter, unspotted regions.

The problem setup is the same as that discussed in §6.3, where our full map vector $\mathbf{a}$ is the outer product of a spectral matrix $\mathbb{S}$, whose J columns are the different spectral components, and a spherical harmonic coefficient matrix $\mathbb{Y}^\top$, whose J rows are the corresponding coefficient vectors. For simplicity, in this experiment we consider only $J = 2$ map components and further require that the surface map for the second component is the complement of the surface map for the first component. Specifically, the first component map ($\mathbf{y_0}$, the first row of $\mathbb{Y}^\top$) corresponds to the background photosphere, whose value is unity everywhere except within the `SPOT` feature, where the value is zero. The second component map ($\mathbf{y_1}$, the second row of $\mathbb{Y}^\top$) corresponds to the spot and is zero in the photosphere and unity within the spot. The two columns of $\mathbb{S}$ are $\mathbf{s_0}$ and $\mathbf{s_1}$, the unknown spectral components corresponding to the photosphere and the spot, repsectively. We set $\mathbf{s_0}$ to unity everywhere, with an absorption feature at 643 nm similar to the one in the previous experiments. The spot spectrum $\mathbf{s_1}$ corresponds to two absorption lines on either side of this feature, and the baseline intensity is set to half that of the photospheric background. The rest frame spectrum

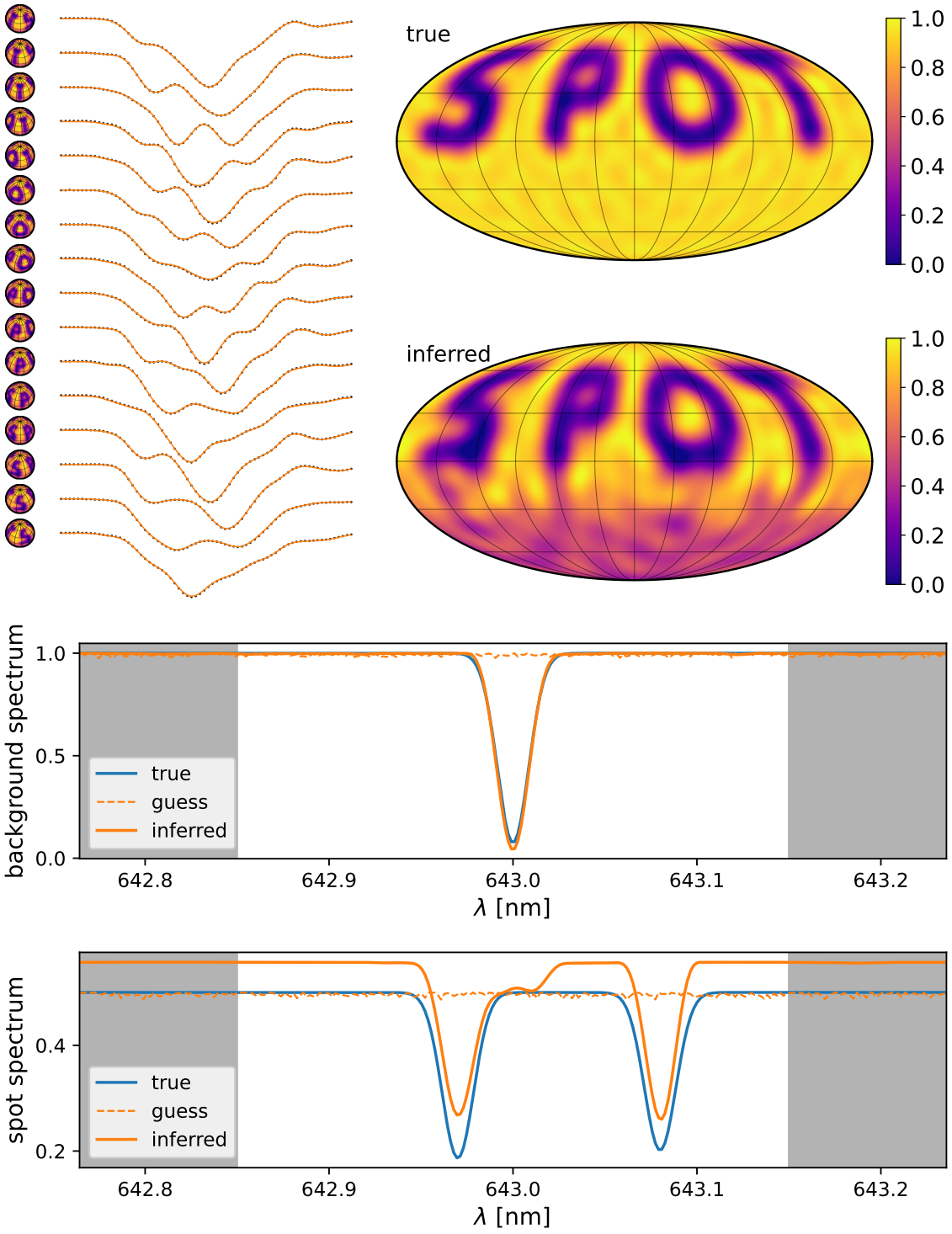

Figure 11. An example of a star whose spectrum is different within the spot and outside of the spot. We are able to infer both the surface map of the star (upper right) and the two component spectra (lower two panels), assuming virtually no prior knowledge of either. Starting from a random initial guess for the maps and the spectra, the optimization step in this experiment ran in about 20 minutes.

at any point on the surface of the star is then a linear combination of $\mathbf{s_0}$ and $\mathbf{s_1}$, weighted by the respective surface maps $\mathbf{y_0}$ and $\mathbf{y_1}$. We generate the synthetic dataset in the same way as in §7.1: 16 spectra, each with 70 wavelength bins observed with SNR $\sim 10^3$.

In our inference step, we again assume we have no knowledge of the map, the baseline, or either of the spectral components. While our bilinear method (§7.3) could in principle work here, we find that in practice it is very difficult to converge on the global optimum without either good priors or a good initial guess. We therefore optimize our model using the nonlinear gradient-descent `Adam` optimizer (Kingma & Ba 2014) within the `pymc3` probabilistic modeling framework (Salvatier et al. 2016). Since we are no longer restricted to using Gaussian priors on the variables (a requirement of our bilinear method), we place a unit-mean Laplacian (sparsity) prior on the two spectra with scale parameter $b = 10^{-3}$ and bounded to the range $[0, 1]$. We further place a Gaussian process (GP) prior on the spectra to enforce smoothness, adopting a squared exponential kernel with amplitude 10^{-1} and lengthscale 10^{-2} nm. We also solve for a single scale parameter $r \sim \text{Uniform}(0, 0.75)$, equal to the ratio of the continuum levels of the two spectra (whose true value is 0.5). Finally, in order to enforce that the two component surface maps sum to unity (as described above), we optimize for the actual *pixel* intensities of the photosphere map, on which we place a simple uniform prior on $[0, 1]$; the spot map is then simply the complement. The spherical harmonic coefficients are then computed via a fast, linear spherical harmonic transform (SHT) operation at each step, as in Bartolić et al. (2021). In order to satisfy the Nyquist criterion, our pixel grid consists of twice as many pixels as spherical harmonic coefficients.[4]

We initialize our two maps with uniform intensities and initialize our two spectra at unity, with small random negative deviations to break the symmetry. We then take 20,000 steps of the optimizer with a learning rate of 10^{-2}.

The results are shown in the bottom panel of Figure 11. Even though we assume no knowledge of the surface map—except its intensity is bounded in $[0, 1]$—we are able to recover the `SPOT` feature just as well as in the previous experiments. We also recover the two spectral components quite well. Our solution for the photospheric component, however, is noticeably better than that for the spot component, simply because effective SNR of the former is larger due to its greater surface area and higher baseline intensity. We also slightly overestimate the baseline level of the spot spectrum, but we correctly recover the locations, shapes, and depths of the two absorption lines. Note, importantly, that the nonlinear optimization step does not yield uncer-

[4] While it may seem strange to return to a discretized parametrization of the surface—after all the work we did to derive a closed-form solution in the spherical harmonic basis—our approach capitalizes on the strengths of both parametrizations: the ease of incorporating a nonnegativity prior in the pixel basis and the efficiency with which we can compute the model in the spherical harmonic basis. Thanks to the linear relationship between these two bases, the transformation between them accrues virtually no overhead.

tainties (or lower bounds, like the bilinear solver); for this, we must resort to MCMC sampling (see §9.2).

7.5. *Dependence on SNR*

Most of the examples in this paper considered very high SNR spectra. Figure 12 shows the effect of different SNR on the quality of the inferred map. As before, the SNR we report is the ratio of the typical line depth in the observed spectrum to the standard deviation of the added noise term, $\sigma_{\mathbf{f}}$. The panels in Figure 12 show the inferred map for SNR in the range 10—10,000, where SNR $\sim 1,000$ is the fiducial case considered in the examples so far. The setup is the same as that in §7.2, where the spectrum is known exactly but the map coefficients and the baseline are not.

As may be expected, the quality of the map decreases with lower SNR, but a reduction of the SNR by $\sim 20\times$ still allows for reasonable recovery of the `SPOT` feature. In general, as the SNR decreases further, the features begin to blur out, especially south of about $+30°$ latitude, where features are more limb-darkened at this inclination and in view for a smaller fraction of the time.

As the SNR increases, however, the quality of the `SPOT` feature does not improve appreciably, since a SNR of 500 is sufficient to recover a $l_{\max} = 15$ band-limited map. Interestingly, at constant spatial prior variance ($\mathbf{\Lambda_y} = 10^{-4}\mathbf{I}$), the map quality actually begins to degrade with higher SNR south of the equator due to overfitting. To circumvent this, we decrease the prior variance by a factor of 10–100 for the highest SNR cases, which penalizes excessive complexity in the solution.

7.6. *Dependence on inclination*

As in Vogt et al. (1987), we also investigate the dependence of the solution on the inclination of the star. Figure 13 shows the inferred map for the same problem setup as before, but this time varying the inclination between 10° (nearly face-on) and 90° (perfectly edge-on). We also add four small circular features at $-45°$ latitude to test our ability to infer features in the southern hemisphere (i.e., the side of the star facing away from the observer). We set the covariance of the spatial prior to $\mathbf{\Lambda_y} = 3 \times 10^{-5}\mathbf{I}$ and, as before, assume the inclination to be known exactly. We further keep the true equatorial velocity of the star constant at 60 km/s across all panels, so that the figure corresponds to the same star viewed at different orientations. Because of this, decreasing the inclination decreases $v \sin i$, reducing the width of the convolution kernel and therefore the spatial resolution of our inferred map. Increasing the inclination does not result in appreciable changes until an inclination of $\sim 80°$, above which it becomes difficult to differentiate between features in the northern and southern hemispheres, whose projected velocities are nearly equal. At $i = 90°$, there is a formal degeneracy between the two hemispheres, which can be seen from the mirror-image symmetry in the inferred map. Finally, as expected, the features in the southern hemisphere are not recovered below inclinations of about 60°, as they are never actually in view. At an inclination of 80° the spots are recovered fairly well,

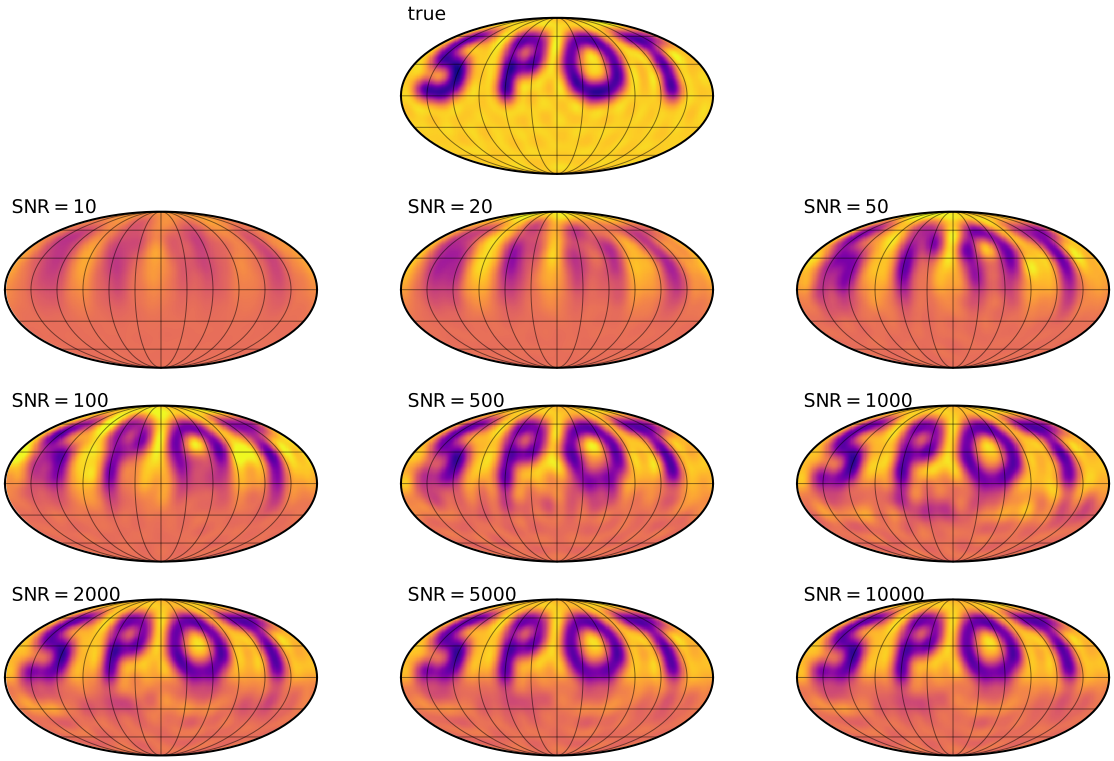

Figure 12. Recovered `SPOT` map as a function of SNR per resolution element. The observed spectra are continuum-normalized and the rest frame spectrum is assumed to be known exactly. The SNR=1,000 panel corresponds to the solution presented in Figure 8.

but never with as high fidelity as features in the observer-facing hemisphere. These results are broadly consistent with those of Vogt et al. (1987).

8. APPLICATION TO WISE 1049-5319B

Having shown that our model performs well on synthetic data, we now turn to a real dataset: the VLT/CRIRES observations of WISE 1049-5319B (Luhman 16B) presented in the Doppler imaging study of Crossfield et al. (2014). WISE 1049-5319B belongs to a binary brown dwarf system, which, at a distance of 2 pc, consists of the two closest known brown dwarfs to the Solar system. Despite its proximity, WISE 1049-5319B is intrinsically dim and, with a projected rotational velocity $v \sin i = 26.1$ km/s, is at the slow end of what can be imaged using the traditional Doppler imaging technique. Therefore, unlike those of classical Doppler imaging targets, the spectral lines of WISE 1049-5319B do not individually show large scale spot modulation (Figure 2) from one epoch to the next; nor are the spectral templates for brown dwarfs particularly well understood. Nonetheless, we chose to model WISE 1049-5319B specifically to show the power behind our new methodology, in particular its ability to learn both the surface map and the rest frame spectrum simultaneously from the entire dataset—not just from unblended spectral lines that have reliable templates.

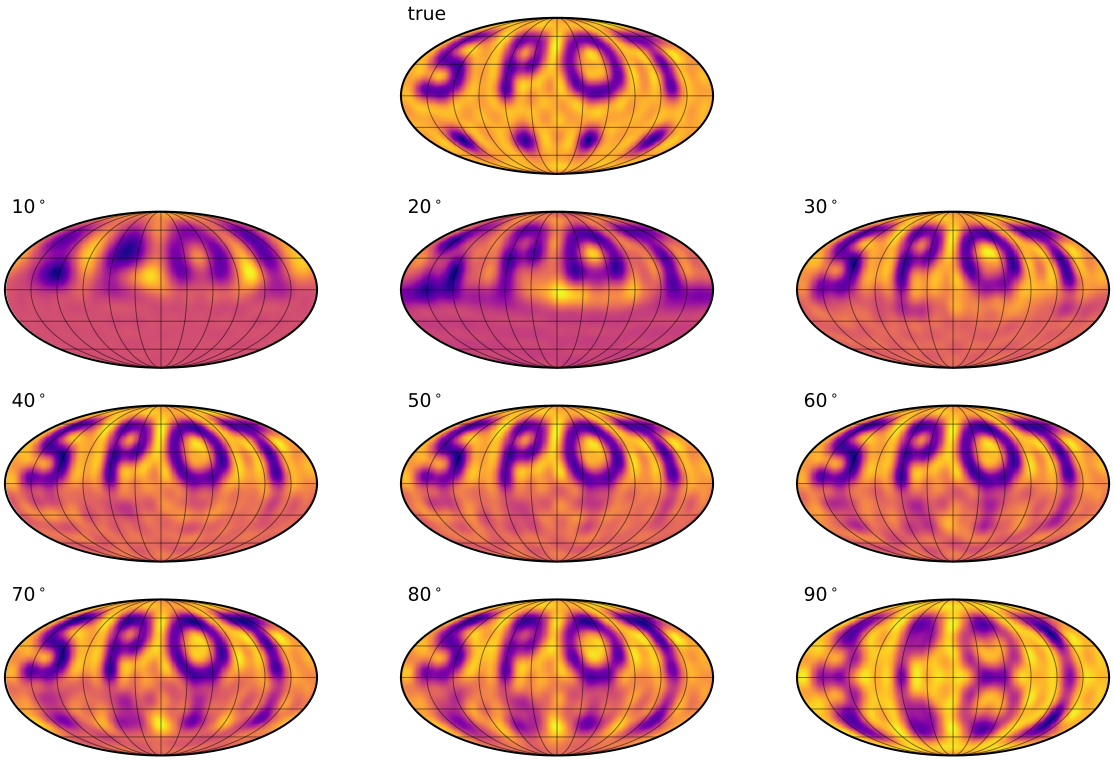

Figure 13. Recovered `SPOT` map as a function of stellar inclination. The observed spectra are continuum-normalized and the rest frame spectrum is assumed to be known exactly. In general, it is very difficult to infer features in the hemisphere facing away from the observer (in this case, the southern hemisphere). At low inclinations, those features are never in view, while at inclinations close to 90°, it becomes difficult to differentiate between features above and below the equator.

In the original Doppler imaging study of this target, Crossfield et al. (2014) took 56 spectra of each object (the angular resolution was high enough to obtain individual spectra of both brown dwarfs) over the four CRIRES detectors in the range $2.288 - 2.345\mu$m over a time span of 5 hours, with individual exposures of 300 s. The 5 hour baseline was enough to capture one full rotation of WISE 1049-5319B, whose period was constrained to be about 4.9 hours from its photometric variability (Gillon et al. 2013). These spectra were then downbinned to 14 epochs to boost the SNR, corrected for tellurics and spectrograph systematics, and continuum-normalized. Each of the 14 spectra consist of 1,024 intensity measurements in each of 4 detectors, for a total of $57,344$ data points. For more details on the data reduction, refer to Crossfield et al. (2014).

The black points in Figure 14 show the exact dataset used in the Doppler imaging analysis of Crossfield et al. (2014), except for the fact that we masked additional outliers at $> 4\sigma$ (identified by subtracting a version of the spectrum smoothed on a lengthscale of 20 pixels) as well as regions of the spectrum that had been visibly overcorrected for tellurics. This figure shows the same data as the lower panel of Extended Data Figure 2 in Crossfield et al. (2014).

In order to produce Doppler images of WISE 1049-5319B, Crossfield et al. (2014) used brown dwarf template spectra from the BT-Settl library (Allard et al. 2013), finding that the spectrum with $\log_{10} g = 5.0$ and $T_{\rm eff} = 1450$ K provided the best match to the dataset. This spectrum was then Doppler-shifted by the systemic radial velocity of the brown dwarf and broadened to match the spectrograph resolution. This final template spectrum is shown as the blue curves in Figure 15, where we again masked regions where data was discarded.

Crossfield et al. (2014) then employed the common technique of Least Squares Deconvolution (LSD; Donati et al. 1997) to transform each of the 14 spectra in Figure 14 into a single high-SNR line spanning 35 pixels, which was used as the input to their Doppler imaging code, based on that of Vogt et al. (1987). They discretized the stellar surface using 200 cells and solved for the optimal surface map with a maximum entropy prior as in Vogt et al. (1987), assuming a fiducial inclination $i = 70°$ roughly based on the photometric rotation period and the inferred value of $v \sin i$; given the low SNR of the dataset, they were unable to place meaningful constraints on i. We show their resulting surface brightness map in the right panel of Figure 16, seen at six different phases spanning a full rotation.

In our analysis, we use the same data (Figure 14) and the same spectral template (Figure 15), but instead following the methodology derived in this work. Our problem is similar to that in §7.3, in which we treat both the spectrum and the surface map as unknowns given continuum-normalized spectra as input. However, this time we use the template spectrum as the prior mean on the true rest frame spectrum $\mathbf{s}$, setting the prior standard deviation to one percent (i.e., a covariance $\mathbf{\Lambda_s} = 10^{-4}\mathbf{I}$) to allow the model to learn any small deviations from the template. We must also change the way we model the continuum normalization (§6.1), given the high density of spectral lines and the consequent difficulty of finding a reliable continuum level to normalize against. We therefore treat the unknown continuum normalization factor for each of the 14 epochs as a free parameter of the model, $f \sim \text{Uniform}(0.3, 3.0)$; we also allow for a small, time-independent, additive offset $\Delta \sim \mathcal{N}(0, 0.1^2)$ for each of the four detectors. We assume the same fiducial inclination $i = 70°$, a rotational period $P = 4.9$ days, and allow the linear limb darkening coefficient and projected velocity to vary subject to $u_1 \sim \text{Uniform}(0, 1)$ and $v \sin i \sim \mathcal{N}(26.1, 0.2^2)$ km/s, respectively, as in Crossfield et al. (2014). Finally, we place a uniform prior on the pixel intensities $\mathbf{p} \sim \text{Uniform}(0, 1)$ as in §7.4, linearly transforming to an $l_{\rm max} = 10$ spherical harmonic representation of the map to compute the model at each step. We must therefore solve for $2 \times (l_{\rm max} + 1)^2 = 242$ free parameters for the pixelized surface map, $3,748$ free parameters for the rest frame spectrum across each of the four detectors (equal to the size of our template spectrum after removing outliers), 18 parameters describing the continuum normalization, and 2 stellar parameters (the linear limb darkening coefficient and the projected velocity). In total, our model has just over $4,000$ free parameters.

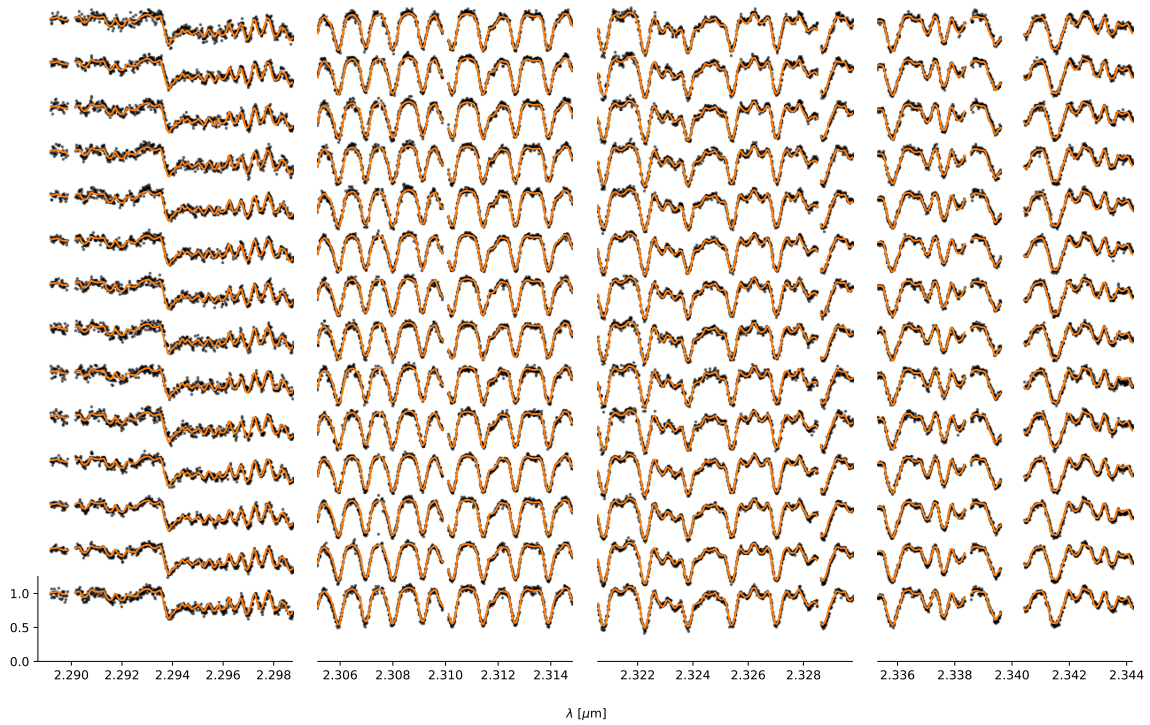

Figure 14. Calibrated and tellurics-corrected CRIRES spectra of WISE 1049-5319B from Crossfield et al. (2014) taken at 14 epochs spanning a full rotation of the brown dwarf (black points). Our maximum *a posteriori* solution is shown as the orange curves.

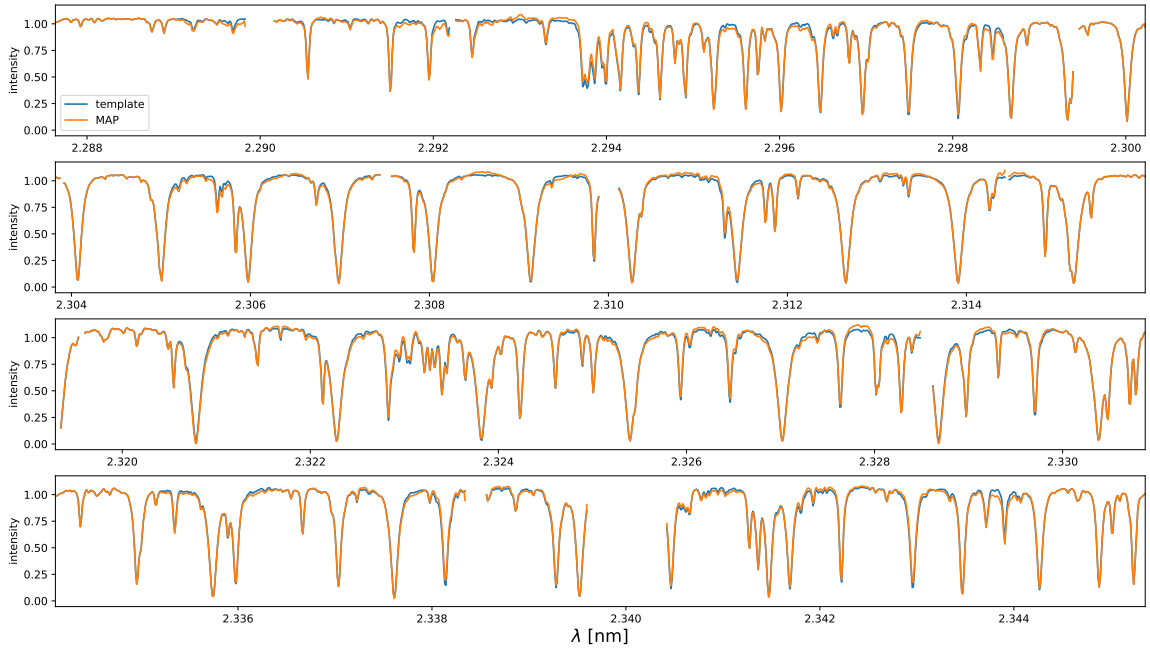

Figure 15. Unbroadened template spectra used to model the data in Crossfield et al. (2014) (blue curves). These were used as a prior on the spectra inferred in this work (orange curves). We infer a slightly different depth for some lines, as well as subtle changes in the continuum.

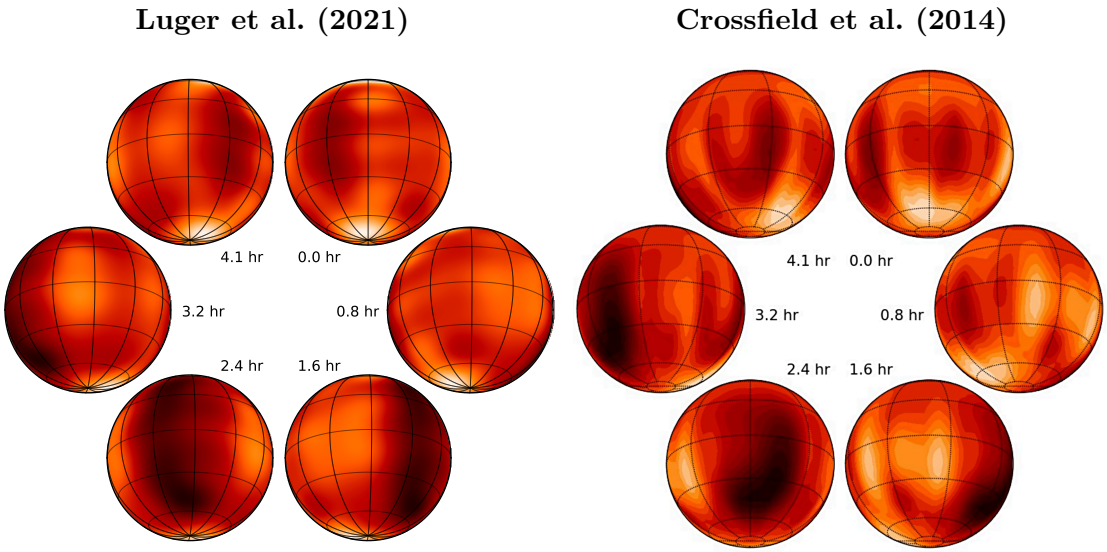

Figure 16. Maximum *a posteriori* surface maps of WISE 1049-5319B inferred from the spectra in Figure 14. *Left:* surface map inferred using the algorithm developed in this work. *Right*: surface map inferred in Crossfield et al. (2014), reproduced from Figure 2 in that paper. We infer the same two dominant features as in Crossfield (2014): a bright circumpolar spot, seen along the sub-observer meridian at $t = 0.0$ hr, and a dark, latitudinally elongated band prominent at $t = 2.4$ hr. While the exact latitude of the bright spot and the behaviour of the dark spot equatorward of $-30°$ latitude differ betwee the two solutions, both are point estimates obtained under different choices of prior and from significantly different methods. We take their qualitative agreement to be a successful validation of our method.

As in §7.4, our choice of parametrization and priors precludes the use of the fast bilinear approach outlined in §5. We therefore follow the same approach as in §7.4, optimizing our log-likelihood function using `Adam` with a learning rate of 10^{-3}. Since we have a very good guess for the template spectrum, this optimization is much faster than in §7.4, converging within $1,000$ steps. Because of our efficient gradient-based implementation (§11), the entire optimization takes only about 3 minutes of runtime on a 2018 Macbook laptop.

Our results are shown in Figures 14, 15, and 16. In Figure 14, our maximum *a posteriori* (MAP) model for each of the 14 spectra is shown as the orange line; it agrees with the data to within the measurement uncertainties. In Figure 15 we show our MAP *template* (rest frame) spectrum as the orange curve. Since we used the Crossfield et al. (2014) template as a prior (blue curve), our solution is very similar to the input spectrum. Nevertheless, we infer slightly different line depths for several of the lines at the level of a few percent; we also infer slightly different continuum levels in certain portions of the spectrum. Note, importantly, since we only performed an optimization (rather than full posterior inference), we cannot quantify the actual significance of these differences.

The right panel of Figure 16 shows our principal result: the inferred solution for the surface brightness map of WISE 1049-5319B; compare to the right panel, showing the solution inferred by Crossfield et al. (2014). Broadly speaking, the two surface

maps agree rather well: we infer the same circumpolar bright spot, of about the same size and at roughly the same phase, although the exact latitude is different. We also infer a similar dark feature on the "back" side of the object (visible near the sub-observer point at $t = 2.4$ hours). Bright equatorial regions also broadly match between both solutions. There are some important differences, such as the behavior of the $t = 2.4$ hour dark region beyond 60° of the visible pole. In general, we find that different choices for our priors—in particular the regularization on the pixels and the template spectrum, the prior on the limb darkening coefficient, and the value of the inclination of the star—result in maps that differ somewhat. We do, however, agree with Crossfield et al. (2014) in that the polar bright spot is the most robust of the inferred features, followed by the bottom portion of the dark spot seen at $t = 2.4$ hours. Once again, since this is simply an *optimization*, we are unable to place reliable uncertainties on our solutions. These require a sampling scheme like MCMC (§9.2).

9. DISCUSSION

9.1. *WISE 1049-5319B*

In §8 we validated our novel approach by reproducing the results of Crossfield et al. (2014). Despite the fact that we obtained a qualitatively similar surface map for WISE 1049-5319B (Figure 16), there are several aspects to our approach that could be improved in future work. First, we did not vary the inclination, but instead kept it fixed at the fiducial value $i = 70°$ used in Crossfield et al. (2014). The authors of that study argued that the SNR of the dataset was too low to constrain the inclination, and that their surface maps did not qualitatively change by varying the inclination between 60° and 90°, the range consistent with the joint knowledge of the rotation period and $v \sin i$. We find a similar result, but we note that the likelihood of the solution improves marginally when we allow the inclination to vary. A posterior inference scheme may thus be able to place stronger constraints on the value of i. Second, neither our study nor that of Crossfield et al. (2014) accounted for the finite exposure time of the observations; given that each of the 14 spectra were collected over 20 minutes, the object rotates through $\frac{20\,\mathrm{min}}{4.9\,\mathrm{hr}} \times 360° \approx 24°$ during each exposure. Nevertheless, this is comparable to the limiting resolution of our $l_{\mathrm{max}} = 10$ spherical harmonic expansion (equal to about $180°/l_{\mathrm{max}} = 18°$), so it is unlikely to substantially change our results. Third, as in Crossfield et al. (2014), we accounted for instrumental broadening in the template spectrum; in other words, the spectrum shown in Figure 15, the input into our Doppler imaging code, includes broadening due to the finite spectral resolution of CRIRES. A better approach is to compute the Doppler-broadened spectrum based on the true rest frame spectrum and *then* apply instrumental broadening (i.e., the order in which events take place in reality). Finally, and most importantly, our results (like those of Crossfield et al. (2014)) are just a point estimate. While the main features inferred in the map—the polar bright spot and the dark spot on the back side of the object—appear to be robust to changes in our priors, we are unable to quantify

our confidence in the map solution. Doing so with gradient-based posterior sampling methods is straightforward with the starry implementation of our algorithm, but we defer it to future work.

9.2. *Regularization, priors, and posteriors*

Inferring surface maps from unresolved astronomical observations is a hard problem. It is "hard" in the usual sense, requiring the development of complicated algorithms and the evaluation of computationally expensive models; but it is also "hard" in the sense that quite often the true answer is *unknowable* because the data simply do not contain sufficient information about it. This is especially the case in photometric studies, where one might seek to infer the distribution of spots on the surface of a star from its rotational light curve alone. Recently, Luger et al. (2021b) showed that this problem is extremely ill-posed: only a very small fraction of the information about the two-dimensional intensity distribution on the stellar surface is encoded in the light curve (see also Basri & Shah 2020). Even at infinite SNR, there exist an infinite number of solutions to the inverse problem, most of which look nothing alike. In a photometric study, the *only* way to obtain a unique solution is to supplement the dataset with external information, usually in the form of restrictive assumptions or explicit priors on the quantities of interest. Most of the information in the solution to a light curve inversion problem comes from these prior assumptions, so it is extremely important to get them right.

In the case of Doppler imaging datasets, the problem is much more well-posed because of the availability of an entire new dimension of information (wavelength). In the limit of infinite SNR, spectral resolution, and temporal sampling, and if the rest frame spectrum is known exactly, in many cases there *are* unique solutions to the visible portion of the stellar surface.[5] In this limit, the data is sufficiently informative and changing one's prior does not substantially change the inferred surface map. This is approximately the case in §7.1, in which we inferred the true surface map given a very high SNR dataset and perfect knowledge of the rest frame spectrum (see Figure 7). However, it is not the case in many of the other examples in that section, particularly when the rest frame spectrum is not known exactly and when the SNR of the dataset is finite. In those cases, good solutions to the surface map (ones in which the `SPOT` feature is clearly legible in the inferred map) can be obtained only under certain choices of our priors (or the strength of the regularization term). In the sections below we discuss how to choose these priors in a principled way.

9.2.1. *Gaussian priors*

In §7 we took advantage of the bilinear nature of the problem to compute the uncertainty (§7.1) and lower bounds on the uncertainty (§7.2 and §7.3) of the surface

[5] Recall that, except in the pathological case where $i = 90°$, there is always a region on the "back" side of the star that is never in view at any phase. No amount of statistical or algorithmic cleverness can ever tell us what that region looks like!

map solution. We found that the posterior uncertainty on the map ranges from about one percent at the north (visible) pole of the star to upwards of 10 percent at the south (hidden) pole (see Figures 7–9). It is important to keep in mind that these values are entirely dependent on our choice of prior (namely, a zero-mean Gaussian prior on the spherical harmonic coefficients with constant variance equal to 10^{-4}). In particular, the value of the posterior uncertainty close to the south pole of the star—which is never in view, and whose true intensity cannot be constrained from the data—is determined exclusively from the prior variance we chose above (via Equation 46). Its value is therefore as arbitrary as our choice of prior.[6] In §7 we chose a variance of 10^{-4} because we had access to the true maps and found, via trial and error, that this value yielded a good balance between underfitting and overfitting the synthetic data. In a realistic application, however, we could not do this, so we must naturally wonder: how should we go about choosing the prior variance in a principled way?

The Bayesian approach to this is to choose priors that truly encode our prior knowledge of the system. Absent prior observations of the system, these priors usually come from physics, such as our knowledge of the typical stellar photospheric temperatures or our understanding of the kinds of inhomogeneities we would expect on their surfaces (dark spots on stars, or patchy clouds on brown dwarfs). The most straightforward way to encode this information in our priors is to recognize that the prior variance on a spherical harmonic vector is closely related to a prior on its angular power spectrum (see, e.g., Baldi & Marinucci 2006; Wandelt 2012). Our choice for prior covariance in §7, namely $\mathbf{\Lambda_y} = 10^{-4}\mathbf{I}$, implies an isotropic, flat angular power spectrum prior on our surface maps with power equal to 10^{-4} in each spherical harmonic degree. Increasing this value will in general lead to solutions with more structure on all scales; decreasing it will favor more homogeneous maps. This may be an acceptable prior if truly nothing is known about the typical angular scale of features on the surface, but this is seldom the case. Both observations of the Sun and magnetohydrodynamic simulations of stellar magnetic fields give us prior constraints on the typical scales of starspots (e.g., Berdyugina 2005; Solanki et al. 2006), and climate models of brown dwarfs tell us about typical sizes of clouds and other atmospheric features (e.g., Tan & Showman 2019, 2020, 2021a,b). These can be encoded into our Gaussian prior by choosing a prior variance that scales with spherical harmonic degree, as in Equation (31), where the variance at each degree is parametrized as, e.g.,

$$\sigma_l^2(\alpha, \lambda, \tau) = \sigma_\lambda^2 \exp\left(-\frac{(l-\lambda)^2}{2\tau^2}\right) \quad , \tag{49}$$

where λ is the typical spherical harmonic degree of features, σ_λ^2 is the prior variance at that degree, and τ controls the uncertainty in λ. These hyperparameters can be either imposed or even *learned* from the data; see §9.2.4.

[6] While the exact *value* of the posterior variance scales with the prior variance, the *ratio* between the uncertainty in the north and the south does not. The gradient seen in the uncertainty maps of Figures 7–9 *is* therefore meaningful, as it directly encodes the spatial distribution of the information content of the data.

Sometimes, however, we may have even more information to encode in our priors, such as the typical *location* of the features. This might be the case for solar-like or lower mass stars, which have specific active latitudes at which spots form. An isotropic prior of the form discussed above cannot encode this, but a dense covariance matrix (with structure in the off-diagonal elements) *can*. Luger et al. (2021c) showed that it is possible to encode properties such as the distribution of spot latitudes, spot contrasts, and spot sizes into the prior covariance on the spherical harmonic coefficients; see Luger et al. (2021d) for an efficient implementation of this. Priors of this form can, once again, be imposed (if prior knowledge of the active latitudes or spot properties is available) or learned from the data by sampling over the corresponding hyperpriors. We discuss this in more detail in §9.2.4 below.

9.2.2. *Cross-validation*

The approach discussed above for choosing one's priors is distinctly Bayesian. Alternatively, one may take a data-driven approach: to learn the "optimal" value of the priors from the data itself. One common approach to this is *cross-validation*, in which one's prior is optimized to yield the best predictive power for the data. Typically, cross-validation consists of choosing a specific value for the prior, masking one ("leave-one-out-cross-validation") or several ("leave-p-out-cross-validation") data points, optimizing the parameters of the model given this smaller dataset, then using the optimized model to predict the values of the missing data (e.g., Stone 1974). The value of the prior is then chosen as the one with the highest predictive power, i.e., the smallest loss when fitting the missing data points (for an example of this, see §3.7 in Luger et al. 2018).

In this context, the prior is usually referred to as a "regularization" term, since this procedure is decidedly not Bayesian. Specifically, it regularizes the model to ensure a balance between underfitting (the model doesn't fit the data well) and overfitting (the model fits the data so well that it's also fitting the noise, leading to spurious results). In addition to helping one choose the value for (say) the prior covariance on the surface map, cross-validation can be used to optimize most of the numerical choices one makes when fitting a dataset, such as the degree $l_{\rm max}$ of the spherical harmonic expansion (or, in a discrete basis, the number of pixels in the surface grid) and the number of map components (§7.4).

9.2.3. *Other priors*

While we used Gaussian priors in most of our examples (§7), other kinds of priors might make more sense in different contexts. In particular, in §8 we enforced a uniform prior in the range $[0, 1]$ on the pixel intensities. This encodes our seemingly trivial but surprisingly restrictive (see Fienup 1982) belief that stellar intensities must be non-negative, which a Gaussian prior cannot do. Priors like this one break the analyticity of the MAP solution (§5), but they can be especially important in the limit that the data is not sufficiently constraining. Traditionally, many Doppler imaging studies

impose an additional *maximum entropy* constraint (Vogt et al. 1987), whose objective is to ensure a similar balance between underfitting and overfitting as discussed above. These can also easily be incorporated into our formalism, although we recommend using a method such as cross-validation to choose the regularization strength in this context.

9.2.4. *Posterior sampling*

In the limit of finite SNR, and when things like the true stellar spectrum, the stellar inclination, and the limb darkening profile are not well known (which is usually the case in Doppler imaging studies!), it is important to recognize the limitation of point estimates like those obtained in this work. The maps of WISE 1049-5319B presented in Figure 16, for instance, are just that: point estimates that carry no sense of the uncertainty associated with any of those features. Given the low SNR of the dataset (Figure 14), the unknown inclination of the star, and the uncertainty in parameters like the rotation period, $v \sin i$, and the limb darkening profile, the maps shown in Figure 16 are almost certainly *not* what WISE 1049-5319B actually looks like in reality. The fact that we obtained a solution that is qualitatively similar to that of Crossfield et al. (2014) is encouraging, but it is not sufficient. Doppler imaging studies are notorious for producing features that are artificially elongated in latitude, just like the dark spot seen in the figure (see, e.g., Unruh & Collier Cameron 1995). The Doppler imaging method is also less sensitive to azimuthally symmetric features, like the zonal bands that may be present on brown dwarfs as on Jupiter. Effects like these can bias the point estimates obtained in Doppler imaging studies, but they *can* be captured if one fully propagates uncertainties to the final result. This requires computing the posterior distribution of the model parameters (including the parametrization of the surface map), which is absolutely critical when one wishes to *interpret* a Doppler map.

In all but the simplest cases, maximum likelihood estimators are typically biased estimators of the true parameters (e.g., Lehmann & Casella 1998), meaning point estimates like those shown in Figure 16, as well as in many other Doppler imaging studies, are also biased. If the goal of a study is *only* to produce a map of the stellar surface, this may not be too much of an issue, especially because it is generally understood that Doppler maps are subject to artifacts like latitudinal elongation and suppression of azimuthally-symmetric features like bands, as mentioned above. Some of the features in Figure 16, for instance, are likely to be real, such as the bright polar spot and the dark spot on the back side of the object; but their exact properties are uncertain, and many of the other features in the maps may be spurious. However, generating a surface map is often *not* the ultimate end goal of Doppler imaging studies. Rather, we are usually interested in a more fundamental question: what does our inferred map tell us about the star's properties, or even about stellar physics in general? And therein lies the problem: as soon as we attempt to interpret the maps we've generated in terms of metadata such as the typical contrast of the star spots,

their typical latitudes, or the rate of differential rotation (based on comparing Doppler imaging maps taken at different epochs), we are subject to substantial bias due to the point estimate problem (for an example of this same problem in the context of inferring the exoplanet eccentricity distribution, see Hogg et al. 2010).

For this reason, it is crucial to treat Doppler imaging as a posterior inference problem, when at all possible; see Asensio Ramos et al. (2021) for recent work on a Doppler imaging model that harnesses approximate Bayesian techniques to perform posterior inference. In some restricted cases, we showed that is possible to obtain the full surface map posterior analytically with our method (§5), but only when all priors are Gaussian and the inclination, limb darkening profile, and template spectrum are known exactly. In all other cases, one must compute an approximation to the posterior distribution over all these quantities numerically, such as with MCMC. This is typically a computationally expensive problem, given the intricate degeneracies at play (which makes posterior sampling difficult), but also the sheer number of parameters, as we must sample over all the parameters representing our surface map (typically on the order of several hundred). When the input spectrum is unknown, the dimension of our problem can easily exceed a few thousand (§8).

Fortunately, our model is computationally efficient to evaluate and, given its analyticity, trivial to differentiate using an autodifferentiation scheme (see §11). It can therefore be used out-of-the-box within gradient-based sampling schemes such as `pymc3` (Salvatier et al. 2016), in particular the No-U-Turn Sampler (NUTS), a variant of Hamiltonian Monte Carlo that harnesses gradient information to greatly speed up the convergence of the MCMC chains (for a recent application of HMC in the context of spectroscopy, see Kawahara et al. 2021). Our model may also be useful in the context of variational inference (VI), and in particular automatic differentiation variational inference (ADVI; Kucukelbir et al. 2016), which casts posterior inference as an optimization problem; both VI and ADVI are also implemented in `pymc3`. The script used to infer the MAP solution for the map of WISE 1049-5319B (see Figure 16) uses `pymc3` to perform the optimization; extending it to perform posterior inference is straightforward. However, doing so is beyond the scope of the present paper, given that this task is still likely to be challenging given the very large number of parameters ($\sim$4,000). We will therefore revisit this system in the context of posterior estimation in a future paper.

One final word regarding posterior inference: given that our end goal is often to infer something about the physics of stellar atmospheres rather than exactly what they look like, posterior sampling can be employed to explicitly marginalize over all possible surface maps to directly constrain our posterior beliefs about the physics. An example of this is the model suggested in §9.2.1 above, where we express our prior covariance in terms of hyperparameters controlling the amplitude of the surface features and their typical lengthscale (Equation 49) or in terms of the typical size, contrast, and latitude of spots (Luger et al. 2021c,d). When the value of those hyperparameters is

not known *a priori*, we can include them as free variables in our model and perform MCMC to compute their marginal distribution (i.e., marginalized over everything else, including all possible surface maps that are consistent with the data). We can even imagine doing this in an ensemble setting under a fully hierarchical model, in which we attempt to constrain these properties from the analysis of many Doppler imaging datasets of different stars simultaneously. This procedure is likely computationally prohibitive, but the linearity of our model (§4) allows us to extend the Gaussian process formalism of Luger et al. (2021c) to Doppler imaging datasets, allowing us to marginalize over the surface maps consistent with the data *analytically*; see §10 below.

9.3. *Assumptions*

Both the derivation and application of our model make several assumptions, many of which we have already discussed. Nevertheless, we present a detailed list of these assumptions here for clarity:

- We assume there is a well-defined stellar "surface", an effective layer from which the star emits its light. This is certainly an oversimplification, as stars are fundamentally three-dimensional objects with complex vertical temperature gradients.
- We assume the emission is isotropic. We relax this assumption by allowing for limb darkening (§6.2), but this is treated as a suppression of flux toward the limb rather than by self-consistently solving the equations of radiative transfer.
- We assume the radial velocity field of the stellar surface is a simple dipole, controlled only by the rigid body rotation of the star. This ignores effects like convective blueshift and differential rotation. We discuss incorporating these into our work in §9.4 below.
- We assume the surface map is constant over time, and that all time variability comes from stellar rotation. We discuss modeling temporal surface variability in the context of a stochastic process in §10 below.
- We assume that the surface map may be well modeled by a spherical harmonic expansion truncated at some degree $l_{\rm max}$. The spherical harmonic basis is complete, meaning this assumption is rigorously correct in the limit $l_{\rm max} \to \infty$. In our experiments, we chose $l_{\rm max} = 10 - 20$, which allows us to model surface features on scales larger than about 18° and 9°, respectively.
- We assume the rest frame spectrum is defined on a grid with sufficiently high resolution that the convolution with the Doppler kernel may be approximated as a discrete linear convolution (Equation 19).

- In our base case (§3), we assume the star has a single rest frame spectrum, and that spatial changes are strictly changes in overall intensity. We relax this assumption in §6.3, where we assume that variations in the spectrum with location can be expressed as the outer product of a set of J spectral principal components $\mathbb{S}$ and a set of J spatial principal components $\mathbb{Y}$. If we choose $J = (l_{\rm max} + 1)^2$ (the number of spherical harmonic coefficients in the expansion of each component), this allows us to model *any* spectral-spatial configuration on the surface of the star, band-limited to degree $l_{\rm max}$. Truncation of this basis at smaller values of J limits our ability to model arbitrary spatial variability in the spectrum.

All of the assumptions listed above relate to our *forward* model (§4). Throughout the paper, we make several additional assumptions when solving the *inverse* problem, that is, solving for the map and/or spectrum given a dataset. None of these assumptions are necessary for the application of our model to other datasets, but we include them here for completeness:

- We assume the data are correctly calibrated and free of significant outliers, and that the noise model (e.g., Gaussian errors with known standard deviation) is known.
- We assume there are either no telluric lines, or that these have been removed in a pre-processing step. These may also be modeled simultaneously, but that is beyond the scope of the present work.
- We assume the SNR of the data is high enough that the maximum *a posteriori* solution to the surface map is meaningful. In general, however, we stress the importance of performing posterior inference to constrain the uncertainties and one's degree of confidence in the solution.
- In all of our examples, we assume the stellar rotation period and inclination are known exactly. In practice, this is rarely ever the case, and one should optimize for (or, ideall, sample over) these variables.

9.4. *Limitations and future work*

There are a few important limitations of the algorithm presented here and its implementation in `starry` that we will address in future papers in this series. First, we do not currently address the effects of differential rotation and convective blueshift. Both are straightforward to account for by modifying the path $\mathcal{C}(\lambda, \mathbf{x})$ of the line integral in Equation (13) to correspond to the isovelocity lines in the velocity field accounting for these two effects. Under the common assumption of linear differential rotation, in which the shear is proportional to the square of the sine of the latitude, the velocity field may be written exactly as a cubic polynomial in the Cartesian coordinates of the stellar surface (Short et al. 2018), therefore admitting an exact representation in

terms of an $l_{\max} = 3$ spherical harmonic expansion.[7] Similarly, the effect of convective blueshift on the velocity field can be modeled with a low degree spherical harmonic expansion consisting only of radial ($m = 0$) modes. In the Appendix, we showed that in the absence of these effects, the curves $\mathcal{C}(\lambda, \mathbf{x})$ are straight lines, so the line integral in Equation (13) is trivial to solve. In the presence of differential rotation and convective blueshift, these curves become polynomials, making the integration trickier. While we were unable to find simple closed-form expressions for the resulting line integrals, they may still exist; we will investigate this in future work. Even if a closed-form solution is unavailable, it is likely still possible to efficiently evaluate the integrals numerically.

Another effect we do not presently model is stellar oblateness. Many of the Doppler imaging targets are massive stars, which tend to be fast-rotating and therefore oblate. We recently expanded the starry framework to model stellar oblateness Dholakia et al. (2021), and will similarly adapt our Doppler imaging framework in an upcoming paper. In the absence of differential rotation and convective blueshift, oblateness can be accounted for by simply weighting the integration limits in Equation (A1) by the semi-minor axis of the elliptical boundary of the star in projection. Accounting for those effects may require numerical integration, as discussed above.

In a similar vein, our present model involves an integration over the entire stellar disk. This is the desired approach in Doppler imaging studies, but does not apply in the context of Doppler *tomography* (e.g., Johnson et al. 2017), in which the star is observed during a transit of an orbiting companion to infer, e.g., the degree of misalignment between the companion's orbit and the stellar spin axis. Our formalism can be adapted to Doppler tomography by changing the limits of integration in Equation (A1) to exclude the region occulted by the companion; we will also investigate this in upcoming work. We will also explore an extension of the method to the case of Zeeman-Doppler imaging (Marsh & Horne 1988), which makes use of spectropolarimetric observations to map the stellar magnetic field topology.

Regarding systematic effects, we do not explicitly account for instrumental broadening. The observed spectrum is a convolution of the Doppler-shifted, disk-integrated stellar spectrum (whose model our algorithm calculates) with the point spread function (PSF) of the spectograph. This convolution can be modeled as a dot product of a Toeplitz matrix with the Doppler design matrix $\mathbf{D}$, much in the same way as Equation (19). In a similar vein, our formalism assumes the data has been corrected for telluric contamination. This is typically done by removing a telluric model (based on a template spectrum) from the data in a pre-processing step (as in Crossfield et al. 2014), but in the data-driven approach of this paper, they could also be learned from the data, similar to the approach in Bedell et al. (2019). We will investigate this in future work.

[7] https://starry.readthedocs.io/en/latest/notebooks/RVDerivation

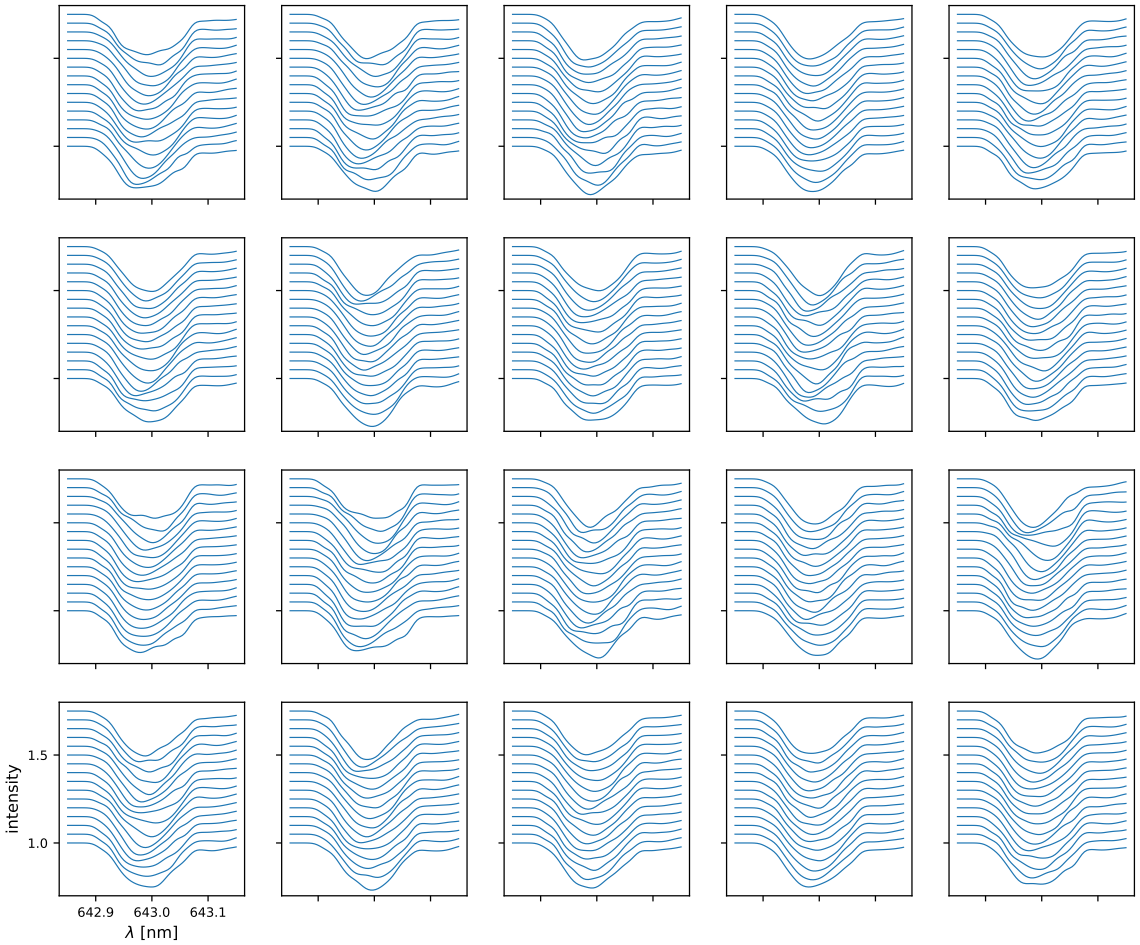

Figure 17. 20 samples from a spectral-temporal Gaussian process constructed from the algorithm presented here. Each panel shows a different sample, each of which consists of 16 epochs, offset vertically for visibility. The rest frame spectrum is the same as in Figure 6 and the star is inclined at $i = 40°$ with equatorial velocity $v = 50$ km/s. The star is assumed to have spots with radius $r = 15°$, 25 percent contrast, and active latitude $\phi = 30° \pm 5°$.

10. A SPECTRAL-TEMPORAL GAUSSIAN PROCESS

There is one immediate extension of the formalism developed in this work that merits a more in-depth discussion: the adaption of the Doppler imaging method into a Gaussian process (GP) for stellar spectroscopy. While the goal of Doppler imaging is often to infer the actual surface map of the star, sometimes the goal may be to infer higher-level statistical information, such as the active latitude of spots or their typical sizes. And sometimes, the goal of a spectroscopic analysis may not be to infer something about the star at all, such as in the case of radial velocity surveys of exoplanets. In both of these cases, one does not really care about what the *actual* map of the star is irrelevant; rather, the goal is to *marginalize* over all the possible stellar surfaces consistent with the data to obtain the marginal posterior of the quantities of interest, such as the active latitude or the properties of an orbiting exoplanet. In §9.2.4, we discussed how one might perform this marginalization via MCMC sampling.

However, thanks to the linearity of the model developed here, we can use the formalism developed in Luger et al. (2021c) to do this *analytically*.

Luger et al. (2021c) presented an interpretable GP for stellar photometric variability, based on the fact that it is possible to analytically compute the covariance matrix $\mathbf{\Lambda_y}$ describing the distribution of spherical harmonic coefficients characterizing a given stellar surface conditioned on certain bulk spot properties $\boldsymbol{\theta}$, such as their radius, latitude, and contrast distributions.[8] Since the model for the (photometric) flux is linear in the spherical harmonic coefficients (Luger et al. 2019), the covariance of the GP for stellar light curves is then simply

$$\mathbf{\Sigma}_{\text{photo}} = \mathbf{A}\,\mathbf{\Lambda_y}(\boldsymbol{\theta})\,\mathbf{A}^\top \ , \tag{50}$$

where $\mathbf{A}$ is the design matrix that transforms from spherical harmonic coefficients to flux, which implicitly depends on the stellar inclination, rotation period, etc.

Conditioned on a fixed spectrum, the Doppler imaging problem is also linear in $\mathbf{y}$ (Equation 27), so we can write the covariance of the GP for a spectroscopic timeseries dataset as

$$\mathbf{\Sigma}_{\text{Doppler}} = \mathbf{D}\,\mathbf{S}\,\mathbf{\Lambda_y}(\boldsymbol{\theta})\,\mathbf{S}^\top\,\mathbf{D}^\top \ , \tag{51}$$

where $\mathbf{S}$ is given by Equation (28) and $\mathbf{D}$ is the Doppler design matrix (Equation 23). The likelihood of the Gaussian with covariance given by Equation (51) is the probability of the observed spectra conditioned on spot hyperparameters $\boldsymbol{\theta}$, *marginalized* over all possible stellar surface maps. Since this marginalization is done analytically, one need only sample over the parameters of interest (the components of $\boldsymbol{\theta}$, as well as the properties of the orbiting planet, in the case of a radial velocity study) in an inference problem. Thus, this GP formalism can reduce the dimensionality of one's model from several hundred (the number of parameters used to model the surface map) to a handful (the hyperparameters of the GP, plus the parameters of the planet model, if applicable).

As an example, Figure 17 shows twenty samples from the GP described above. We condition the GP on the same spectrum used in §7 and on a stellar inclination $i = 40°$ and equatorial velocity $v = 50{,}000$ km/s, and choose hyperparameters $\boldsymbol{\theta}$ corresponding to spots with mean radius $r = 15°$, mean fractional contrast $c = 0.25$, and active latitude $\phi = 30° \pm 5°$. Each panel shows a full spectral-temporal dataset sampled from this GP, where individual spectra are vertically offset from each other for visibility.

The main application of our GP, of course, will not be for drawing samples, but for doing inference. In particular, we expect this GP to be useful for extreme precision radial velocity (EPRV) planet searches, which must contend with stellar variability signals that are often orders of magnitude stronger than the planetary signals of

[8] In Luger et al. (2021c), this covariance matrix is denoted $\mathbf{\Sigma_y}$.

interest. Our GP can be used as a principled, physically-motivated way to marginalize over the stellar signal to robustly detect the Keplerian signal of an orbiting planet. It is even straightforward to account for time variability in the stellar surface with this formalism. Luger et al. (2021c) discussed how to incorporate a stochastic time-variable component into the photometric variability GP, which can be directly incorporated into Equation (51) via the inclusion of a decoherence timescale τ in the set of GP hyperparameters $\boldsymbol{\theta}$.

The next paper in the *Mapping Stellar Surfaces* series will focus on further development of this GP, including accounting for uncertainty in the spectral template, accounting for the nonlinearity introduced by the continuum normalization, marginalizing over unknowns such as the inclination and limb darkening parameters, and rigorously testing the formalism on both synthetic and real datasets.

11. IMPLEMENTATION IN starry

We implement the Doppler imaging algorithm presented here in the starry package (Luger et al. 2019, 2021a), an open source `Python` library for stellar and planetary light curve analysis. All of the figures in this paper were generated via the Doppler imaging API in starry. The scripts used to generate them are available on GitHub[9] and can be directly accessed via the GitHub icons next to each figure caption. These are well-commented and can be used as a starting point for other Doppler imaging projects.

The algorithm is implemented using `theano/aesara` (Bergstra et al. 2010; Willard et al. 2021), a just-in-time compiling framework for `Python` with powerful autodifferentiation and optimization capabilities. As such, the algorithm is fully differentiable and thus easy to use within gradient-based optimization and/or inference schemes such as Hamiltonian Monte Carlo. The analysis presented in Figures 9 and 14—16 harnesses this capability within the `pymc3` modeling framework; refer to the corresponding scripts for details.

The implementation of the algorithm in starry is highly optimized. Even though we derived the model for the flux as a purely linear operation on the spectral/spatial expansion of the stellar surface (Equation 14), it is more efficient in practice to compute it as a convolution using the highly optimized `conv2d` operations in `theano/aesara`. Typical runtimes and scalings for the starry algorithm are shown in Figure 18, where we plot the evaluation of the full forward model for all spectra in a dataset as a function of several parameters. Equation (14) requires first the rotation of the $(l_{\rm max}+1)^2$ $\boldsymbol{\kappa}^\top$ kernels (each of length W) to the correct stellar rotational phase, an operation that scales approximately as $Wl_{\rm max}^3$, followed by $(l_{\rm max}+1)^2$ linear convolutions of the rotated $\boldsymbol{\kappa}^\top$ kernels with the spectral components (each of length K), which scales approximately as $WKl_{\rm max}^2$. Since typically $K \gg l_{\rm max}$, the latter computation dominates and the entire evaluation of the model for M spectral epochs takes $\mathcal{O}(MWKl_{\rm max}^2)$

[9] https://github.com/rodluger/paparazzi

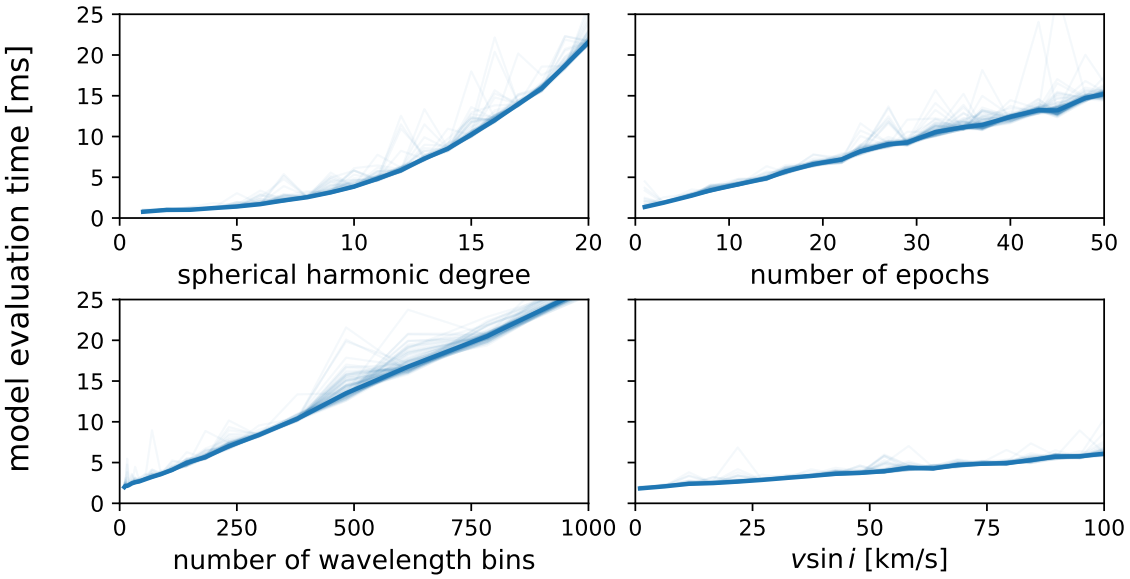

Figure 18. Model evaluation time in milliseconds as a function of various parameters, for spherical harmonic degree $l_{\max} = 10$, $M = 10$ spectral epochs, $K = 200$ wavelength bins, and $v \sin i = 50$ km/s. Thin lines correspond to individual trials; the thick lines show the median evaluation time. Timing tests were performed on a 2-core Linux virtual GitHub Actions machine.

operations. This is consistent with the relationships shown in Figure 18: the runtime is linearly proportional to the number of spectral epochs M, the number of wavelength bins K, and the width of the convolution kernel $W \propto v \sin i$; and it scales quadratically with the degree of the spherical harmonic expansion $l_{\max}$. In general, evalution of the model for a typical dataset takes on the order of 10 or a few tens of milliseconds.

12. CONCLUSIONS

We have derived a novel method to infer the surface brightness maps of stars from high resolution spectroscopic timeseries data. Our method is based on that of Vogt et al. (1987), but it explicitly exposes the linear dependence of the observed spectrum on both the representation of the surface map (expressed as a vector of spherical harmonic coefficients) and the template (rest frame) spectrum, making it fast, differentiable, and, in certain cases, trivially invertible. It also naturally extends to stellar surfaces with multiple spectral components, such as a star whose spectrum differs between spot and photosphere. Ours is fundamentally a data-driven approach to the problem of Doppler imaging, where one can not only impose knowledge of the template spectrum but also *learn* it from the data. This is particularly useful in the case where reliable spectral templates are not available, such as in certain wavelength regimes or for certain objects, like the brown dwarf considered in §8. Importantly, we derive a model for the *entire observed spectrum*. Most Doppler imaging studies rely on techniques like least squares deconvolution (LSD; Donati et al. 1997) to in-

fer a mean spectral line shape at each epoch, which is then treated as the input to the Doppler imaging step. These studies usually focus on a few distinct, unblended, and well-understood spectral lines, which are transformed into a single mean profile evaluated over a relatively small number of wavelength bins. Our approach, however, is to model the entire observed spectrum at full resolution, including line blends and wavelength regimes where the template spectrum may be unreliable. This is made computationally possible given our highly optimized `Python` implementation within the `starry` stellar modeling framework (Luger et al. 2019), which is openly available on GitHub with ample documentation. By using the entire dataset and learning the template spectrum in regions where it is unavailable, we are able to maximize the information content of the inputs. Our approach therefore has potential applications to slow rotators, low signal-to-noise spectra, and spectra of objects, such as brown dwarfs and directly imaged exoplanets, whose templates are not well known *a priori*.

As a proof-of-concept, we qualitatively reproduced the Doppler map of the brown dwarf WISE 1049-5319B (Luhman 16B) inferred in Crossfield et al. (2014) from VLT/CRIRES observations; see Figure 16. We also tested our model's ability to accurately infer the surface maps and rest frame spectra from synthetic high SNR spectral timeseries observations, even in the limit of minimally informative priors on both quantities. Finally, we showed that we can infer the spectrum within spots when it is significantly different from that of the surrounding photosphere, again in the limit of high SNR and uninformative priors.

Given the linearity of our model, under certain choices of prior one can obtain the posterior distribution of surface maps analytically and in under a second for most datasets. However, given the usual uncertainty on stellar properties such as the inclination angle, the equatorial velocity, and the limb darkening profile, as well as the common need for restrictive priors on the surface map, most realistic applications of Doppler imaging require the use of computationally expensive numerical methods to infer the posterior distribution of the parameters. For this reason, Doppler imaging is traditionally performed in an optimization setting: most studies thus far yield only point estimates of the inferred surface maps, with no formal uncertainties. As we argued in §9.2.4, this can lead to significant biases when these maps are interpreted to infer higher-level properties of stars, such as the typical sizes and active latitudes of star spots, or the average spot coverage fraction. These biases are due to both the use of shrinkage priors (such as the commonly used maximum entropy constraint) and the fact that point estimates usually include spurious artifacts (like latititudinally elongated features) while suppressing certain surface modes (like azimuthally-symmetric features). To accurately infer spot properties from Doppler maps, one must treat the problem probabilistically, by explicitly marginalizing over unknown quantities to obtain a posterior *distribution* over the quantities of interest. The model presented here, as well as its implementation in `starry`, is easy to use in a posterior inference setting. In particular, our implementation harnesses the

autodifferentation capabilities of the `theano/aesara` (Bergstra et al. 2010; Willard et al. 2021) package within the `pymc3` (Salvatier et al. 2016) probabilistic modeling framework, allowing for efficient posterior inference using Hamiltonian Monte Carlo (HMC) techniques. We also present a sketch of how to incorporate our model into an interpretable, physically-motivated Gaussian process (GP) formalism for stellar spectroscopic timeseries datasets, a method that is highly relevant to extreme precision radial velocity (EPRV) searches for exoplanets. We will explore this application in more detail in the next paper in this series.

While Doppler imaging has classically been used to infer the surfaces of giant or luminous pre-main sequence stars, the study of Crossfield et al. (2014) used the method to infer the first Doppler map of a brown dwarf, which we reproduced in this work. Future studies could help us understand the intricate atmospheric dynamics of brown dwarfs (Tan & Showman 2021a,b), as well as the global properties of exoplanets, which may be amenable to Doppler imaging given the advent of a new generation of extremely large ground-based telescopes (Crossfield 2014). A likely first target will be the directly-imaged giant planet β Pictoris b, which is only an order of magnitude dimmer than WISE 1049-5319B; Snellen et al. (2014) showed that future observations with the planned Extremely Large Telecope (ELT) should be able to produce a global map of its surface. Given the expected low SNR of such observations and the considerable uncertainty in the rest-frame spectra of directly imaged exoplanets, an efficient, probabilistic Doppler imaging approach like the one developed in this work will be imperative to maximizing the scientific output of those studies.

This paper was prepared using the **show your work!** open source scientific article workflow[10]. The PDF was generated automatically from a GitHub repository[11] using a frozen `conda` environment on GitHub Actions CI[12], ensuring that all the results are exactly reproducible and easily extensible with minimal effort from the user. The algorithm derived in this paper is fully implemented in the `starry Python` package, which is installable via `pip` or directly from GitHub[13], with ample documentation, examples, and tutorials.[14]

We would like to thank Eric Agol, Christina Hedges, Rachael Roettenbacher, and the Astronomical Data Group at the Center for Computational Astrophysics (CCA) for useful comments, suggestions, and insights. RL and LLZ are supported by Flatiron Fellowships at the CCA, a division of the Simons Foundation.

[10] https://github.com/rodluger/showyourwork
[11] https://github.com/rodluger/paparazzi
[12] https://github.com/features/actions
[13] https://github.com/rodluger/starry
[14] https://starry.readthedocs.io

APPENDIX

A. SOLVING THE $\boldsymbol{\kappa}^\top$ INTEGRAL

In this section we will show how to compute the integral in Equation (18), yielding the terms of the convolution kernel $\boldsymbol{\kappa}^\top$. It is convenient to first change basis from spherical harmonics to polynomials on the sphere:

$$\begin{aligned}\boldsymbol{\kappa}^\top(\mathring{\lambda}) &= \int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} \boldsymbol{\phi}^\top\Big(x, y\Big)\mathrm{d}y\,\mathbf{A_1} \\ &\equiv \boldsymbol{\rho}^\top(\mathring{\lambda})\,\mathbf{A_1}\ ,\end{aligned} \tag{A1}$$

where $\mathbf{A_1}$ is the change of basis matrix from spherical harmonics to polynomials (Equation B11 in Luger et al. 2019) and

$$\boldsymbol{\phi}^\top(\mathbf{x}) \equiv \begin{pmatrix} 1 & x & z & y & x^2 & xz & xy & yz & y^2 & ... \end{pmatrix}^\top , \tag{A2}$$

is the polynomial basis (Equation 7 Luger et al. 2019). The n^{th} term of the polynomial basis is a polynomial in x, y, and $z = \sqrt{1 - x^2 - y^2}$:

$$\phi_n = x^i y^j z^k \tag{A3}$$

where i, j, k are integers given by

$$\begin{aligned} i &= \lfloor \Lambda - \Delta \rfloor \\ j &= \lfloor \Delta \rfloor \\ k &= \lceil \Delta \rceil - \lfloor \Delta \rfloor \end{aligned} \tag{A4}$$

with

$$\begin{aligned} \Lambda &= \lfloor \sqrt{n} \rfloor \\ \Delta &= \frac{n - \Lambda^2}{2}\ . \end{aligned} \tag{A5}$$

We may now explicitly write down the n^{th} term of the vector $\boldsymbol{\rho}^\top(\mathring{\lambda})$ in Equation (A1),

$$\rho_n(\mathring{\lambda}) = \rho_{i,\,j,\,k}(\mathring{\lambda}) = x^i \int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} y^j \left(1 - x^2 - y^2\right)^{\frac{k}{2}}\ \mathrm{d}y\ . \tag{A6}$$

When $i = k = 0$, the integral is easy to evaluate:

$$\rho_{i,\,j,\,0}(\aleph_0) = \begin{cases} 2\,(1-x^2)^{\frac{j+1}{2}} & j \text{ even} \\ 0 & j \text{ odd}\ . \end{cases} \tag{A7}$$

The general term may be computed from the recurrence relation

$$\rho_{0,\,j,\,0}(\aleph_0) = \frac{j-1}{j+1}\big(1-x^2\big)\rho_{0,\,j-2,\,0} \tag{A8}$$

provided

$$\begin{aligned} \rho_{0,\,0,\,0} &= 2\sqrt{1-x^2} \\ \rho_{0,\,1,\,0} &= 0\ . \end{aligned} \tag{A9}$$

When $i = 0$ and $k = 1$, we may substitute $y = \sin\psi\sqrt{1-x^2}$ in the integrand to obtain

$$\rho_{0,\,j,\,1}(\aleph_0) = (1-x^2)^{\frac{j+2}{2}} \int\limits_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \sin^j\psi\cos^2\psi\,\mathrm{d}\psi\ . \tag{A10}$$

We may solve this integral using the recurrence relation

$$\int \sin^j\psi\cos^m\psi\,\mathrm{d}\psi = -\frac{\sin^{j-1}\psi\cos^{m+1}\psi}{j+m} + \frac{j-1}{j+m}\int \sin^{j-2}\psi\cos^m\psi\mathrm{d}\psi\ , \tag{A11}$$

which, in our case, simplifies to

$$\int\limits_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \sin^j\psi\cos^2\psi\,\mathrm{d}\psi = \frac{j-1}{j+2}\int\limits_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \sin^{j-2}\psi\cos^2\psi\mathrm{d}\psi\ . \tag{A12}$$

We thus arrive at the recurrence relation

$$\rho_{0,\,j,\,1} = \frac{j-1}{j+2}\big(1-x^2\big)\rho_{0,\,j-2,\,1} \tag{A13}$$

with initial conditions

$$\begin{aligned} \rho_{0,\,0,\,1} &= \frac{\pi}{2}\big(1-x^2\big) \\ \rho_{0,\,1,\,1} &= 0\ . \end{aligned} \tag{A14}$$

Given these expressions, recursing upward in i is easy:

$$\rho_{i,\,j,\,k} = x\,\rho_{i-1,\,j,\,k}\ . \tag{A15}$$

This completes our derivation. To summarize, given the boundary conditions

$$\begin{aligned}\rho_{0,\,0,\,0} &= 2\sqrt{1-x^2} \\ \rho_{0,\,0,\,1} &= \frac{\pi}{2}\big(1-x^2\big) \\ \rho_{0,\,1,\,k} &= 0\ , \end{aligned} \tag{A16}$$

we may compute any term in $\boldsymbol{\rho}^\top$ via the expressions

$$\begin{aligned}\rho_{0,\,j,\,k} &= \frac{j-1}{j+1+k}\big(1-x^2\big)\rho_{0,\,j-2,\,k} \\ \rho_{i,\,j,\,k} &= x\,\rho_{i-1,\,j,\,k}\ . \end{aligned} \tag{A17}$$

The terms of the convolution kernel $\boldsymbol{\kappa}^\top$ are then obtained by dotting this vector into the change of basis matrix $\mathbf{A_1}$, as discussed above.

Table 1. Common notation used in this paper

Symbol	Description	Reference
	Variables	
a_{lm}	spectral spherical harmonic coefficient	Equation (6)
$\mathbf{a}$	vector of spectral spherical harmonic coefficients	Equation (8)
$\mathbf{A}$	matrix form of $\mathbf{a}$	Equation (26)
$\mathbf{B}$	baseline design matrix	§6.1
c	speed of light	Equation (15)
$\mathcal{C}$	curve of constant radial velocity	Equation (13)
D	logarithmic Doppler shift	Equation (2)
$\mathbf{D}$	Doppler design matrix	Equation (22)
$\mathbf{f}$	vector of observed flux values	Equation (22)
F	flux	Equation (3)
i	inclination	Equation (15)
I	specific intensity	Equation (1)
$\mathbf{I}$	identity matrix	
l	spherical harmonic degree	
$\mathbf{L}$	limb darkening operator	Equation (41)
m	spherical harmonic order	
$\mathbf{P}$	Change of basis matrix from Y_{lm} to pixels	Equation (46)
$\mathbf{R}$	Wigner rotation matrix	Equation (10)
$\mathbf{s}$	vector representation of the spectrum	Equation (26)
$\mathbf{S}$	block matrix representation of $\mathbf{s}$	Equation (28)
$\mathbb{S}$	matrix of J spectral components	Equation (44)
$\mathcal{S}$	projected visible area of the star	Equation (3)
t	time	§2
T	artificial temperature in bilinear solver	§7
T_0	initial artificial temperature setting	§7
$\Delta \log T$	artificial temperature step size	§7
u	limb darkening coefficient	§6.2
$\mathbf{x}$	Cartesian position on the sky	§2
y_{lm}	spherical harmonic coefficient	§5
$\mathbf{y}$	vector of spherical harmonic coefficients	Equation (26)
Y_{lm}	spherical harmonic	Equation (6)
$\mathbf{Y}$	block matrix representation of $\mathbf{y}$	Equation (33)
$\mathbb{Y}$	matrix of J spherical harmonic components	Equation (44)

Table 1. (continued from previous page)

Symbol	Definition	Reference
	Variables (Greek)	
β	relativistic parameter	Equation (2)
δ	Dirac delta function	Equation (4)
$\boldsymbol{\kappa}^\top$	convolution kernel basis vector	Equation (13)
λ	wavelength	§2
λ_r	reference wavelength	§2
$\mathfrak{X}$	$\ln \frac{\lambda}{\lambda_\mathrm{r}}$ (observer frame)	§2
$\mathfrak{X}_0$	$\ln \frac{\lambda}{\lambda_\mathrm{r}}$ (rest frame)	§2
$\boldsymbol{\Lambda}$	prior covariance matrix	
$\boldsymbol{\mu}$	prior mean vector	
$\boldsymbol{\psi}^\top$	spherical harmonic basis vector	Equation (9)
$\boldsymbol{\Sigma}$	covariance matrix	
ω	angular velocity	Equation (15)
	Dimensions	
J	# of spectral components	§6.3
K	# of wavelength bins in the broadened spectrum	§4
K'	# of wavelength bins in the latent spectrum	§4
K_obs	# of wavelength bins in the observed spectrum	§7
M	# of epochs observed	§4
N	# of spherical harmonic coefficients	§4
N'	# of spherical harmonic coefficients after limb darkening	§6.2
W	# of wavelength bins in the convolution kernel	§4
	Operators & Other symbols	
diag	diagonal matricization	Equation (43)
vec	vectorization	Equation (26)
$*$	convolution	Equation (5)
$\circ$	element-wise product	§6.1
$\otimes$	Kronecker product	Equation (21)
$\lceil\,\rceil$	ceiling	Equation (A4)
$\lfloor\,\rfloor$	floor	Equation (A4)
$\mathbf{0}$	the zero vector, $(0 \quad 0 \quad \cdots \quad 0)^\top$	
$\mathbf{1}$	the one vector, $(1 \quad 1 \quad \cdots \quad 1)^\top$	

REFERENCES

Agol, E., Luger, R., & Foreman-Mackey, D. 2020, AJ, 159, 123, doi: 10.3847/1538-3881/ab4fee

Allard, F., Homeier, D., Freytag, B., et al. 2013, Memorie della Societa Astronomica Italiana Supplementi, 24, 128. https://arxiv.org/abs/1302.6559

Asensio Ramos, A., Diaz Baso, C., & Kochukhov, O. 2021, arXiv e-prints, arXiv:2108.09266. https://arxiv.org/abs/2108.09266

Baldi, P., & Marinucci, D. 2006, arXiv Mathematics e-prints, math/0606709. https://arxiv.org/abs/math/0606709

Bartolić, F., Luger, R., Foreman-Mackey, D., Howell, R. R., & Rathbun, J. A. 2021, arXiv e-prints, arXiv:2103.03758. https://arxiv.org/abs/2103.03758

Basri, G., & Shah, R. 2020, ApJ, 901, 14, doi: 10.3847/1538-4357/abae5d

Bedell, M., Hogg, D. W., Foreman-Mackey, D., Montet, B. T., & Luger, R. 2019, AJ, 158, 164, doi: 10.3847/1538-3881/ab40a7

Berdyugina, S. V. 2005, Living Reviews in Solar Physics, 2, 8, doi: 10.12942/lrsp-2005-8

Bergstra, J., Breuleux, O., Bastien, F., et al. 2010, in Proceedings of the Python for Scientific Computing Conference (SciPy)

Carroll, T. A., Strassmeier, K. G., Rice, J. B., & Künstler, A. 2012, A&A, 548, A95, doi: 10.1051/0004-6361/201220215

Collado, J. R. A., Rico, J. F., López, R., Paniagua, M., & Ramírez, G. 1989, Computer Physics Communications, 52, 323 , doi: https://doi.org/10.1016/0010-4655(89)90107-0

Collier Cameron, A. 1997, MNRAS, 287, 556, doi: 10.1093/mnras/287.3.556

Collier Cameron, A., & Unruh, Y. C. 1994, MNRAS, 269, 814, doi: 10.1093/mnras/269.3.814

Cowan, N. B., Chayes, V., Bouffard, É., Meynig, M., & Haggard, H. M. 2017, Monthly Notices of the Royal Astronomical Society, 467, 747, doi: 10.1093/mnras/stx133

Crossfield, I. J. M. 2014, A&A, 566, A130, doi: 10.1051/0004-6361/201423750

Crossfield, I. J. M., Biller, B., Schlieder, J. E., et al. 2014, Nature, 505, 654, doi: 10.1038/nature12955

Deutsch, A. J. 1958, in IAU Symposium, Vol. 6, Electromagnetic Phenomena in Cosmical Physics, ed. B. Lehnert, 209

Deutsch, A. J. 1970, ApJ, 159, 985, doi: 10.1086/150376

Dholakia, S., Luger, R., & Dholakia, S. 2021, arXiv e-prints, arXiv:2109.03250. https://arxiv.org/abs/2109.03250

Donati, J. F. 1999, MNRAS, 302, 457, doi: 10.1046/j.1365-8711.1999.02096.x

Donati, J. F., Semel, M., Carter, B. D., Rees, D. E., & Collier Cameron, A. 1997, MNRAS, 291, 658, doi: 10.1093/mnras/291.4.658

Donati, J. F., Howarth, I. D., Jardine, M. M., et al. 2006, MNRAS, 370, 629, doi: 10.1111/j.1365-2966.2006.10558.x

Falk, A. E., & Wehlau, W. H. 1974, ApJ, 192, 409, doi: 10.1086/153072

Fienup, J. 1982, Applied optics, 21, 2758, doi: 10.1364/AO.21.002758

Gillon, M., Triaud, A. H. M. J., Jehin, E., et al. 2013, A&A, 555, L5, doi: 10.1051/0004-6361/201321620

Goncharskii, A. V., Stepanov, V. V., Khokhlova, V. L., & Yagola, A. G. 1982, Soviet Ast., 26, 690

Goncharskii, A. V., Stepanov, V. V., Kokhlova, V. L., & Yagola, A. G. 1977, Soviet Astronomy Letters, 3, 147

Hackman, T., Jetsu, L., & Tuominen, I. 2001, A&A, 374, 171, doi: 10.1051/0004-6361:20010682

Hatzes, A. P. 1995, ApJ, 451, 784, doi: 10.1086/176265

Hogg, D. W., Myers, A. D., & Bovy, J. 2010, ApJ, 725, 2166, doi: 10.1088/0004-637X/725/2/2166

Johnson, M. C., Cochran, W. D., Addison, B. C., Tinney, C. G., & Wright, D. J. 2017, AJ, 154, 137, doi: 10.3847/1538-3881/aa8462

Kawahara, H., Kawashima, Y., Masuda, K., & Crossfield, I. J. M. 2021, arXiv e-prints, arXiv:2105.14782. https://arxiv.org/abs/2105.14782
Khokhlova, V. L. 1976, Astronomische Nachrichten, 297, 203, doi: 10.1002/asna.19762970408
Kingma, D. P., & Ba, J. 2014, arXiv e-prints, arXiv:1412.6980. https://arxiv.org/abs/1412.6980
Kochukhov, O., Lüftinger, T., Neiner, C., Alecian, E., & MiMeS Collaboration. 2014, A&A, 565, A83, doi: 10.1051/0004-6361/201423472
Kucukelbir, A., Tran, D., Ranganath, R., Gelman, A., & Blei, D. M. 2016, arXiv e-prints, arXiv:1603.00788. https://arxiv.org/abs/1603.00788
Lehmann, E. L., & Casella, G. 1998, Theory of Point Estimation, 2nd edn. (New York, NY, USA: Springer-Verlag)
Luger, R., Agol, E., Bartolić, F., & Foreman-Mackey, D. 2021a, arXiv e-prints, arXiv:2103.06275. https://arxiv.org/abs/2103.06275
Luger, R., Agol, E., Foreman-Mackey, D., et al. 2019, AJ, 157, 64, doi: 10.3847/1538-3881/aae8e5
Luger, R., Agol, E., Kruse, E., et al. 2016, AJ, 152, 100, doi: 10.3847/0004-6256/152/4/100
Luger, R., Foreman-Mackey, D., Hedges, C., & Hogg, D. W. 2021b, AJ, 162, 123, doi: 10.3847/1538-3881/abfdb8
Luger, R., Foreman-Mackey, D., & Hedges, C. 2021c, AJ, 162, 124, doi: 10.3847/1538-3881/abfdb9
Luger, R., Foreman-Mackey, D., & Hedges, C. 2021d, The Journal of Open Source Software, 6, 3071, doi: 10.21105/joss.03071
Luger, R., Kruse, E., Foreman-Mackey, D., Agol, E., & Saunders, N. 2018, AJ, 156, 99, doi: 10.3847/1538-3881/aad230
Marsh, T. R., & Horne, K. 1988, MNRAS, 235, 269, doi: 10.1093/mnras/235.1.269
Piskunov, N. E., Tuominen, I., & Vilhu, O. 1990, A&A, 230, 363
Rice, J. B. 2002, Astronomische Nachrichten, 323, 220
Rice, J. B., Wehlau, W. H., & Khokhlova, V. L. 1989, Astronomy and Astrophysics, 208, 179
Roettenbacher, R. M., Monnier, J. D., Korhonen, H., et al. 2017, ApJ, 849, 120, doi: 10.3847/1538-4357/aa8ef7
Rosén, L., Kochukhov, O., & Wade, G. A. 2015, ApJ, 805, 169, doi: 10.1088/0004-637X/805/2/169
Salvatier, J., Wiecki, T. V., & Fonnesbeck, C. 2016, PeerJ Computer Science, 2, e55, doi: 10.7717/peerj-cs.55
Short, D. R., Orosz, J. A., Windmiller, G., & Welsh, W. F. 2018, AJ, 156, 297, doi: 10.3847/1538-3881/aae889
Snellen, I. A. G., Brandl, B. R., de Kok, R. J., et al. 2014, Nature, 509, 63, doi: 10.1038/nature13253
Solanki, S. K., Inhester, B., & Schüssler, M. 2006, Reports on Progress in Physics, 69, 563, doi: 10.1088/0034-4885/69/3/R02
Stone, M. 1974, Journal of the Royal Statistical Society: Series B (Methodological), 36, 111, doi: https://doi.org/10.1111/j.2517-6161.1974.tb00994.x
Strassmeier, K. G. 1999, A&A, 347, 225
Tan, X., & Showman, A. P. 2019, ApJ, 874, 111, doi: 10.3847/1538-4357/ab0c07
—. 2020, ApJ, 902, 27, doi: 10.3847/1538-4357/abb3d4
—. 2021a, MNRAS, 502, 678, doi: 10.1093/mnras/stab060
—. 2021b, MNRAS, 502, 2198, doi: 10.1093/mnras/stab097
Unruh, Y. C., & Collier Cameron, A. 1995, MNRAS, 273, 1, doi: 10.1093/mnras/273.1.1
Vogt, S. S., & Penrod, G. D. 1983, PASP, 95, 565, doi: 10.1086/131208
Vogt, S. S., Penrod, G. D., & Hatzes, A. P. 1987, ApJ, 321, 496, doi: 10.1086/165647
Wandelt, B. 2012, Gaussian Random Fields in Cosmostatistics, doi: 10.1007/978-1-4614-3508-2_5

Willard, B. T., Osthege, M., Ho, G., et al. 2021, pymc-devs/aesara:, rel-2.0.7, Zenodo, doi: 10.5281/zenodo.4695331
Zaire, B., Donati, J. F., & Klein, B. 2021, MNRAS, 504, 1969, doi: 10.1093/mnras/stab1019