A novel framework for the three-dimensional NLTE inverse problem

Jiří Štěpán[1], Tanausú del Pino Alemán[2,3], and Javier Trujillo Bueno[2,3,4]

[1] Astronomical Institute of the Academy of Sciences, Ondřejov, Czech Republic. e-mail: `jiri.stepan@asu.cas.cz`
[2] Instituto de Astrofísica de Canarias, E-38205 La Laguna, Tenerife, Spain
[3] Departamento de Astrofísica, Universidad de La Laguna, E-38206 La Laguna, Tenerife, Spain
[4] Consejo Superior de Investigaciones Científicas, Spain

ABSTRACT

Inversion of spectropolarimetric observations of the solar upper atmosphere is one of the most challenging goals in solar physics. If we account for all relevant ingredients of the spectral line formation process such as three-dimensional (3D) radiative transfer out of local thermodynamic equilibrium (NLTE), the task becomes extremely computationally expensive. Instead of generalizing 1D methods to 3D, we develop a new approach to the inverse problem. In our meshfree method we do not consider the requirement of 3D NLTE consistency as an obstacle, but as a natural regularization with respect to the traditional pixel-by-pixel methods. This leads to more robust and less ambiguous solutions. We solve the 3D NLTE inverse problem as an unconstrained global minimization problem avoiding repetitive evaluations of the Λ operator. Apart from 3D NLTE consistency, the method allows to easily include additional conditions of physical consistency such as zero divergence of the magnetic field. Stochastic ingredients make the method less prone to ending up in local minima of the loss function. Our method is capable of solving the inverse problem by orders of magnitude faster than it would be possible using grid-based methods. The method can provide accurate and physically consistent results if sufficient computing time is available, but also approximate solutions in case of very complex plasma structures or limited computing time.

Key words. Methods: numerical – Polarization – Radiative transfer – Sun: atmosphere

1. Introduction

Remote sensing of magnetic fields and other physical quantities of the outer solar atmosphere is a notoriously difficult problem. The new generation of solar telescopes such as DKIST (currently in the commissioning phase; Rimmele et al. 2020) and EST (currently in the preparatory phase; Jurčák et al. 2019) will provide spectropolarimetric observations of unprecedented quality, and there is an urgent need of plasma diagnostic tools to correctly interpret them. Yet, our modeling techniques lag behind the capabilities of such new observational facilities.

While the equations governing the intensity and polarization of spectral lines in solar prominences, filaments, and the chromosphere, are mostly well known, the forward modeling remains difficult mainly for numerical reasons: Accounting for all the relevant processes requires the solution of the problem of the generation and transfer of spectral line polarization in three-dimensional (3D) geometry. Since many spectral lines of interest form out of local thermodynamic equilibrium (NLTE), a complicated non-linear and non-local problem needs to be solved. This is the case for a number of spectral lines of high diagnostic potential whose optical thickness can approach or exceed unity, such as Hα at 6563 Å or the He I triplet at 10830 Å. Under NLTE conditions, one can expect a significant impact of the radiative transfer within the medium and, as we will show below, full 3D NLTE radiative transfer becomes inevitable already for small optical thicknesses of the order of one.

The ultimate goal of solar spectropolarimetry is to reliably infer the plasma properties from the observed data. This so-called inverse problem is even more difficult than the forward modeling one. In principle, one should explore the space of all possible model parameters and to eliminate the models that do not agree with the observations. This is not possible in practice because such parameter space is too large and a single 3D NLTE model evaluation already takes a substantial amount of computing (CPU) time. We therefore need to approach the 3D NLTE inverse problem (3DNIP) in a different way.

A common approach to solve the inverse problem in the photosphere, chromosphere, prominences, and filaments, is to use the so-called pixel-by-pixel approach in which every column of matter behind an observed pixel is treated as independent from any other. In the case of the photosphere and chromosphere, a number of techniques have been developed over the years (e.g., Ruiz Cobo & del Toro Iniesta 1992; Socas-Navarro et al. 2000, 2015; Asensio Ramos & Díaz Baso 2019; de la Cruz Rodríguez et al. 2019). For recent reviews on inversion methods, see del Toro Iniesta & Ruiz Cobo (2016) and de la Cruz Rodríguez & van Noort (2017). A number of inversion tools have also been developed for the case of prominences and filaments (e.g., Casini et al. 2003; Asensio Ramos et al. 2008; Lagg et al. 2009). Some algorithms go beyond the pure pixel-by-pixel approach by taking into account spatial correlation of the data (e.g., van Noort 2012; Asensio Ramos & de la Cruz Rodríguez 2015; Asensio Ramos & Díaz Baso 2019) but these solutions do not take into account NLTE radiative coupling among different regions of the plasma.

In the pixel-by-pixel approach to prominences, it is difficult to go beyond models in which the physical quantities are constant along the line of sight (LOS) (i.e., the so-called constant-property slab approximation) because there is simply no justification for considering and constructing more refined mod-

els. This approximation is indeed very rough: If we assume the prominence properties to be constant along the LOS, why should we not consider the properties to be constant along a perpendicular direction, i.e., in the plane of sky? But the spectra are indeed changing in the plane-of-the-sky and in the pixel-by-pixel inversions we are interpreting it as due to changing physical quantities. This approach is clearly inconsistent. An additional problem of the oversimplified models is the existence of a number of ambiguous solutions. Spectral lines that are typically optically thin, such as He I D_3 at 5877 Å, do not suffer so much from the effects of NLTE radiative transfer (López Ariste 2015), even though being subordinate lines they can be affected by such effects due to their coupling with optically thick transitions. However, due to their negligible optical thickness, they do not provide sufficient information about the variation of the plasma parameters along the LOS. The pixel-by-pixel approach can be successfully used but only under certain physical conditions. The observed plasma structure needs to be optically thin in all the directions and it must be sufficiently homogeneous along the LOS.

To the best of our knowledge, the 3DNIP has not yet been seriously studied and it has not been even clarified if it is practically possible. The existing inversion methods solve the problem by neglecting the effects of NLTE radiative transfer between different parts of the medium in the direction perpendicular to the LOS, but this leads to neglecting a crucial ingredient of spectral line formation. In this paper, we introduce a new framework for solving the 3DNIP problem taking into account 3D NLTE radiative transfer effects. We reformulate the problem in a mesh-free manner and we consider it to be an unconstrained global minimization problem. The method can provide both exact and approximate results depending on the available CPU time. Apart from the fact that it can be relatively easily implemented, we show that it can lead to orders-of-magnitude acceleration with respect to the grid-based methods. In addition, it is less prone to end up in local minima.

In contrast to other methods, we do not consider the requirement of NLTE coupling to be an obstacle but rather to be a very strong natural regularization. The fact that all parts of the domain are coupled by a nontrivial set of NLTE equations leads to a much more robust method than if the individual LOS were considered separately. While our method is very efficient, our focus here is more on the physical consistency rather than on computational speed.

This paper is the first one of a series. It mainly focuses on the inversion problem of solar prominence and filament data, but a similar approach can be applied, after some modifications, to the case of the chromosphere. In the following papers, we will address a number of details of the method and its practical applications.

The structure of the paper is the following. In Sect. 2 and Sect. 3, we briefly review the key ingredients of the numerical NLTE spectral synthesis and inverse problems, respectively, and we discuss the specifics of the 3D solution. We also recall the concept of sparse approximation of physical quantities and we discuss the benefits of such representation. In Sect. 4, we recall the standard minimization procedure of the inversion algorithms and we develop a new meshfree approach to the 3D inverse problem. In order to make this possible, we reformulate the inversion problem as a global optimization in which deviation from the NLTE consistency is used as a natural regularization together with other penalizations due to different physical inconsistencies. In the resulting unconstrained minimization algorithm, the atomic state is treated as independent of the magnetohydrodynamical (MHD) state of the atmosphere. We complete
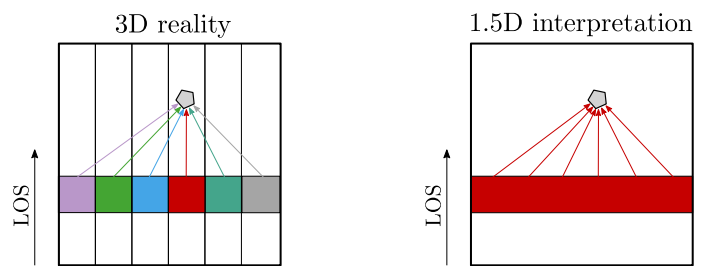
Fig. 1. Spatial coupling between lines of sight in a 3D (left) and in a 1.5D (right) model atmosphere. In 3D, the radiation field at a given point results from radiation transfer from different regions of the domain, both in the vertical and in the horizontal directions. In 1.5D, the horizontal variation is neglected. The 1.5D model can be sufficiently accurate in some cases but, in general, it often leads to serious errors in calculation of the line scattering polarization.

the algorithm in Sect. 5 by using a stochastic approach to the loss function minimization in order to reduce the risk of convergence to a local minimum and to make the solution more accurate and less demanding in terms of computer memory. We demonstrate the usability of the algorithm in Sect. 6, where we apply the method to invert the thermal and magnetic structure of an academic 3D NLTE problem and showing that the total inversion time is by two orders of magnitude smaller than it would be if using the grid-based techniques. We summarize our results in Sect. 7, where we also provide some future prospects. Two appendices provide technical details on the implementation of the method and on the numerical example discussed in Sect. 6.

2. A brief overview of the forward NLTE modeling

In this paper we develop a framework for the numerical inference of the three-dimensional distribution of physical quantities in the solar atmosphere from spectropolarimetric data. We assume that our spectropolarimetric observation consists of a data cube with $N_{pix} = N^2$ spatial dimensions and N_λ wavelengths for each of the four Stokes parameters $(I, Q, U, V) = (I_0, I_1, I_2, I_3)$, i.e., the specific intensity (I), the two components of the linear polarization (Q and U), and the circular polarization component (V). We also assume that the data are affected by Gaussian noise, but we do not consider any other instrumental degradation in this paper.

Although the main objective of this paper is to develop and describe a meshfree inference method, it is illustrative to compare its properties with a Cartesian grid-based method. To our knowledge, no such grid-based method has been developed before, apart from the 2.5D attempt by Štěpán et al. (2019; talk at the Solar Polarization Workshop 9[1]). However, the numerical properties of such a method can be easily estimated. In this section we consider, without loss of generality, that the dimensions of the grid in the grid-based examples is N^3, that is, the mesh has the same spatial resolution as the data. However, take into account that, in practice, it may be necessary to use a finer mesh depending on the particular spectral lines and the spatial extension of their formation region (e.g., Auer 2003).

Given the spatial distribution of physical quantities defining the thermal and magnetic properties of the medium (such as temperature, density, magnetic field, etc.), the forward NLTE solution gives the atomic level populations and quantum coherence that are consistent with the radiation field at every point, $\boldsymbol{r}$, of the

---

[1] https://www.mps.mpg.de/spw9

medium. Hereafter, we use the term MHD-like quantities to describe this set of quantities, $\{\psi(\boldsymbol{r}; k)\}_{k=1}^{N_\psi}$, where N_ψ is the number of such physical quantities. For example, if we consider the kinetic temperature of the plasma to be our first ($k = 1$) MHD-like quantity, we have $\psi(\boldsymbol{r}; 1) = T(\boldsymbol{r})$.

In the standard unpolarized NLTE forward problem, the specific intensity of the radiation at any given point, propagation direction, and wavelength can be formally written as[2]

$$I(\boldsymbol{r}, \boldsymbol{\Omega}, \lambda) = \Lambda(\boldsymbol{r}, \boldsymbol{\Omega}, \lambda)[n(\boldsymbol{r}')] , \qquad (1)$$

where Λ is an operator which depends on the MHD-like quantities and $n(\boldsymbol{r}')$ stands for the atomic level populations within the medium. The simplicity of Eq. (1) is only formal, as it represents a non-local coupling of the radiation field and the state of the matter. This equation is often referred to as the formal solution of the radiative transfer equation (RTE, Hubený & Mihalas 2014), or simply as the formal solution. It is worth noticing that, although this form of the forward problem in terms of the Λ operator is useful in theoretical derivations, the operator is usually not explicitly constructed; it represents the process of solving the RTE.

The mean radiation field in the co-moving reference frame at every point in the medium, i.e., the specific intensity integrated over the unit sphere and over the line absorption profile of spectral lines of interest is

$$\bar{J}(\boldsymbol{r}) = \oint \frac{d\boldsymbol{\Omega}}{4\pi} \int d\lambda \, I(\boldsymbol{r}, \boldsymbol{\Omega}, \lambda)\varphi(\boldsymbol{r}, \lambda) . \qquad (2)$$

where $\varphi(\boldsymbol{r}, \lambda)$ is the line's absorption profile. Likewise, we can define an averaged $\bar{\Lambda}$ operator that takes the above integration into account,

$$\bar{J}(\boldsymbol{r}) = \bar{\Lambda}(\boldsymbol{r})[n(\boldsymbol{r}')] . \qquad (3)$$

The atomic populations, the local mean radiation field, and the local MHD-like quantities are related via the equations of statistical equilibrium (ESE, Hubený & Mihalas 2014),

$$n(\boldsymbol{r}) = \mathcal{E}[\bar{J}(\boldsymbol{r})] . \qquad (4)$$

Together with Eq. (3), these equations make a system of coupled NLTE equations for which we must find the self-consistent distribution of atomic populations.

The general NLTE problem can only be solved numerically with iterative methods. The simplest of such methods is the so called Λ-iteration, which can be summarized in a compact form as

$$\bar{J}_{k+1}(\boldsymbol{r}) = \bar{\Lambda}(\boldsymbol{r})[n_k(\boldsymbol{r}')] , \qquad (5)$$
$$n_{k+1}(\boldsymbol{r}) = \mathcal{E}[\bar{J}_{k+1}(\boldsymbol{r})] , \qquad (6)$$

where the radiation field in iteration $k + 1$ is calculated, at every spatial point, using the atomic populations from the previous iteration k (Eq. 5), and the atomic populations are in turn updated using this new radiation field (Eq. 6). Solving Eq. (5) involves the solution of the RTE at all the domain points, propagation directions, and wavelengths, and it is usually the most computationally expensive part of the NLTE problem solution (see the next section for a discussion of the numerical aspects of this step). The ESE operator $\mathcal{E}$ in Eq. (6), which depends on the

MHD-like quantities, $\{\psi(\boldsymbol{r}; k)\}_{k=1}^{N_\psi}$, acts locally, and the computational cost of its application is rather negligible in comparison with the cost of Eq. (5) for all the relevant solar physics applications.

There are numerical techniques to accelerate the solution of the NLTE problem (Jacobi, Gauss-Seidel/SOR, multigrid; see, e.g., Rybicki & Hummer 1991; Steiner 1991; Trujillo Bueno & Fabiani Bendicho 1995; Fabiani Bendicho et al. 1997, and references therein) that can significantly reduce the number of necessary iterations. Recently, a very promising approach to the NLTE problem acceleration has been proposed by Janett et al. (2021) and Benedusi et al. (2021). However, the evaluation of the Λ operator is always a key ingredient of the forward problem solution. For brevity, we use the term Λ iteration to describe an iteration of any such numerical technique involving the evaluation of the Λ operator, regardless of a particular acceleration scheme.

In this work we solve the more general problem of the generation and transfer of polarized radiation. We consider the stationary NLTE problem in complete frequency redistribution (CRD) discussed in the monograph of Landi Degl'Innocenti & Landolfi (2004) (the more general partial frequency distribution (PRD) problem is beyond the scope of this paper). The specific intensity is then replaced by the vector of Stokes parameters, the mean radiation field by the radiation field tensors $\bar{J}_Q^K$, and the populations of the atomic levels by the atomic density matrix ρ_Q^K. For the sake of simplicity, we only consider bound-bound transitions, but it is straightforward to generalize the method to include the bound-free and free-free transitions as well. We take into account scattering polarization, the Hanle and Zeeman effects, the macroscopic velocity gradients, and the full 3D radiative transfer, i.e., all the relevant physical ingredients leading to the breaking of the symmetry of the scattering processes which critically affect the emergent polarization (see Štěpán et al. 2015; Štěpán 2015; del Pino Alemán et al. 2018; Jaume Bestard et al. 2021a).

Analogously to the MHD-like variables, we define the set of N_ξ atomic-like quantities $\{\xi(\boldsymbol{r}; k)\}_{k=1}^{N_\xi}$. The variables in this set are essentially the solution of the the NLTE forward problem. There is no unique way of defining these quantities, but there are two most useful representations. In the first, one specifies the components of the atomic density matrix, in which case the $\{\xi(\boldsymbol{r}; k)\}_{k=1}^{N_\xi}$ quantities correspond to the real and imaginary parts of the density matrix elements $\rho_Q^K(\boldsymbol{r}; \alpha J)$ in a multi-level atom (see Sect. 7.2 of Landi Degl'Innocenti & Landolfi 2004, for details). In the second representation, one specifies instead the components of the averaged radiation field tensors $\bar{J}_Q^K(\boldsymbol{r}; \ell)$ for each of the spectral lines, ℓ. Because the only condition the atomic-like quantities must fulfill is to fully describe the atomic state, the two given representations are equivalent, since the radiation field tensors, $\bar{J}_Q^K(\boldsymbol{r}; \ell)$, fully determine the atomic density matrix via the ESE. In Sect. 4 it will become clear why it is useful to introduce the $\{\xi(\boldsymbol{r}; k)\}_{k=1}^{N_\xi}$ variables and why some representations can be more advantageous than others depending on the specific inverse problem.

3. Grid-based approach to the NLTE inversion

3.1. Scaling of grid-based numerical forward solutions

The multi-dimensional NLTE problem of the generation and transfer of polarized radiation is typically solved on a discrete grid using the short characteristics method for the RTE (Kunasz & Auer 1988). For the sake of simplicity, we only consider

---

[2] For clarity, we omit the explicit mentioning of the term on the right hand side of the equation standing for the external illumination of the domain, as its inclusion does not contribute to the explanation.

Cartesian meshes in this work. As mentioned in Sect. 2, in solar physics problems the evaluation of the Λ operator is typically the most computationally intensive part of the NLTE forward solution, while the evaluation time of the ESE is negligible in comparison.

In 1D problems, the computing time per Λ iteration is proportional to the number of grid points, N, rays in the angular quadrature, N_Ω, and wavelengths, N_λ. The number of accelerated Λ iterations of the Jacobi method (one of the most commonly used techniques) needed for convergence down to the fixed error tolerance is proportional to N (Hackbusch 1985). Consequently, the total computing time for the NLTE forward solution is of the order of $O(N^2 N_\Omega N_\lambda)$.

In 3D geometry, the computing time per Λ iteration is also proportional to the number of grid points, N^3. However, regarding the number of iterations, the convergence rate is determined by the distance in the units of grid steps between the most distant points. This distance is of the order of $\sqrt[3]{N^3} = N$, hence the number of accelerated Λ iterations in 3D is usually similar to that of the 1D case. Therefore, the solution time of the 3D NLTE problem is typically of the order of $O(N^4 N_\Omega N_\lambda)$.

In order to provide a computing time example, we note that in the 3D NLTE code PORTA (Štěpán & Trujillo Bueno 2013) the CPU time needed for evaluating the Λ operator per grid point, propagation direction, wavelength, and Stokes parameter can be, typically, of the order of 10^{-6} seconds in a common CPU. Considering a small 3D model with $N = 100$ (total 100^3 grid points), angular quadrature with 100 propagation directions, and the four Stokes parameters of a single spectral line sampled in 50 wavelengths, a single Λ iteration would take about 5.5 hours. With the typically expected 100 Jacobi iterations, the total computation time of the NLTE forward solution is about 550 hours. The computing time in more realistic problems of the solar chromosphere, with typical meshes of about 500^3 grid points, becomes correspondingly larger (by two orders of magnitude) and using high-performance computing (HPC) becomes a requirement (see Štěpán et al. 2015; Štěpán & Trujillo Bueno 2016).

In order to bypass these high requirements in computational resources, it is a common practice to split the 3D problem of the N^3 grid into N^2 independent 1D problems with N grid points (the so-called 1.5D approximation). When scattering polarization, the Hanle effect, and the macroscopic velocity fields do not play a significant role in the polarization of the emergent spectral line radiation, such 1.5D solution is faster because the ensuing axial symmetry of each 1D model of every pixel/column allows to neglect the azimuthal dependence of the radiation field. Such approach would still be approximate, as the radiative interaction between such models is neglected, but it can still provide sufficiently accurate approximations in some applications. However, when the cylindrical symmetry is broken, both 1.5D and 3D solutions have similar computing demands and, moreover, the 1.5D approach can lead to significant errors, especially when accounting for scattering polarization (see Fig. 1).

3.2. Parameterization and discrete representation of variables

In the inverse problem one rarely tries to directly find the $\{\psi(\mathbf{r}; k)\}_{k=1}^{N_\psi}$ values in all the grid points because, even in 1D, the number of unknowns would be too large and, moreover, such approach leads to underdetermined problems. Instead, one can parameterize the spatial distribution of variables using a smaller number M_ψ of parameters $\{\psi_i(k)\}_{i=1}^{M_\psi}$ from which the value of

each individual variable can be reconstructed at any point within the medium:

$$\psi(\mathbf{r}; k) = f(\mathbf{r}; \psi_1(k), \cdots, \psi_{M_\psi}(k)). \tag{7}$$

The form of the function $f(\cdot)$ depends on the particular chosen representation of the variables. Examples of this approach are the node-based interpolation in 1D inversions (Ruiz Cobo & del Toro Iniesta 1992) and the wavelet expansion in 2D geometry proposed by Asensio Ramos & de la Cruz Rodríguez (2015).

From now on, we assume that each of the MHD-like quantities is represented by a set of M_ψ parameters. For simplicity of notation, here we choose to parameterize every MHD-like quantity using the same number M_ψ of coefficients. Therefore, the whole model, ψ, can be defined as an $N_\psi M_\psi$-elements vector of these sets of parameters,

$$\begin{aligned}\psi &= (\psi_1, \cdots, \psi_{N_\psi M_\psi}) \\ &= (\psi_1(1), \cdots, \psi_{M_\psi}(1), \psi_1(2), \cdots, \psi_{M_\psi}(N_\psi)).\end{aligned} \tag{8}$$

However, in practice, it can be advantageous to use different numbers of parameters for different quantities. We postpone the discussion of the particular parameterization of Eq. (7) to Sect. 4.1. For convenience, we assume that the elements of ψ are dimensionless parameters.

It is important to emphasize that, under certain conditions, we can aim at fulfilling the condition $M_\psi \ll N^3$, which leads to a significant reduction of the number of unknowns and makes the underdetermined problem a well defined one. Moreover, it can lead to a very significant reduction of the total CPU time needed for the inversion. We will discuss and apply this condition in the following sections.

3.3. Inverse problem definition and solution

Algorithm 1 Inversion as the minimization of χ^2 via successive solutions of self-consistent NLTE forward problems. Here we show the gradient descent method with step h. $\psi(i)$ stands for the i-th iteration of the ψ vector in Eq. (8) and ψ_j stands for its j-th element. The symbol $\bar{J}$ is a condensed notation for the averaged radiation field tensors calculated in the grid points and $\bar{\Lambda}_\psi$ corresponds to the $\bar{\Lambda}$ operator, constructed using the ψ vector.

1: **Initialization**: Randomly initialize the model's MHD-like quantities $\psi(0)$.
2: Given $\psi(0)$, solve the NLTE forward problem for $\bar{J}(0)$ via Λ iteration.
3: $i \leftarrow 0$
4: **repeat** {**Descent along the negative χ^2 gradient.**}
5: $\quad i \leftarrow i + 1$
6: $\quad$ **for** $j = 1$ to $N_\psi M_\psi$ **do** {**Loop over the model MHD-like variables.**}
7: $\qquad$ Calculate j-th element gradient $\nabla_j \chi^2 = \partial \chi^2(D; \psi(i-1))/\partial \psi_j$. {$\Lambda$ **iteration till NLTE consistency.**}
8: $\quad$ **end for**
9: $\quad \psi(i) \leftarrow \psi(i-1) - h\nabla\chi^2$. {**New estimate of the model parameters.**}
10: $\quad$ Given $\psi(i)$, solve the NLTE forward problem for $\bar{J}(i)$ via Λ iteration.
11: **until** $\chi^2(D; \psi(i)) \approx 1$.

Assume that our data, D, consist of a square matrix of $N_{\text{pix}} = N^2$ pixels and N_λ wavelengths for each of the four Stokes

parameters and that they are contaminated with Gaussian noise. The inverted 3D model is parameterized by the ψ vector defined in Eq. (8). The measure of the goodness of the fit of the model to the data is the familiar χ^2 function

$$\chi^2(D; \psi) = \frac{1}{N_{\text{pix}} N_\lambda} \sum_{i=1}^{N_{\text{pix}}} \sum_{j=1}^{N_\lambda} \sum_{k=0}^{3} w_k \frac{(I_{ijk}^{\text{M}}(\psi) - I_{ijk}^{\text{O}})^2}{\sigma_{ijk}^2}, \quad (9)$$

where $I_{ijk}^{\text{M}}(\psi)$ is the k-th Stokes parameter at pixel i and wavelength index j calculated from the model's vector ψ. Similarly, I_{ijk}^{O} is the corresponding observed signal, for the same k-th Stokes parameter at the same i-th pixel and j-th wavelength, with the noise variance σ_{ijk}^2. The w_k quantities are weights for the different Stokes parameters which must fulfill the normalization condition $\sum_{k=0}^{3} w_k = 1$ (see Appendix A.2). The goal of the inversion process is to find an estimate $\hat{\psi}$ of the model parameters leading to the best fit to the observed data,

$$\hat{\psi} = \arg \min_{\psi} \chi^2(D; \psi). \quad (10)$$

Given the normalization of the signals to their noise variance, it follows that the optimal fit fulfills $\chi^2 = 1$. The restriction of the number of model parameters to M_ψ per quantity is an example of regularization of the solution, i.e., additional constraints that cannot be derived from the observed data themselves. Examples of such regularizations in the general inversion theory are the requirement on the number of non-zero parameters (i.e., sparsity regularization) or on smoothness of the solution. Instead of Eq. (10), one then solves a problem of the form

$$\hat{\psi} = \arg \min_{\psi} \left[\chi^2(D; \psi) + g(\psi) \right], \quad (11)$$

with $g(\psi)$ being the regularization condition. Regularization can help make an ill-defined problem well defined and it can be understood as an Occam's razor condition.

It is important to emphasize that, in contrast to pixel-by-pixel inversion, in the minimization of the χ^2 in Eq. (9) all the pixels are being inverted together because, in general, every component of the ψ vector affects the model parameters in the whole computational domain (depending on the particular form of the right-hand side of Eq. 7). This leads to a more robust family of methods where correlations among the MHD-like quantities in different spatial points can be naturally taken into account, such as in Asensio Ramos & de la Cruz Rodríguez (2015). Moreover, it also allows us to take full 3D radiative transfer into account, as well as other physical constrains, such as the condition of zero divergence of the magnetic field, which are practically impossible to implement in the pixel-by-pixel approaches (see below).

Algorithm 1 shows a pseudocode for the traditional inversion process based on the minimization of the χ^2 function via the gradient descent method. Even though more efficient algorithms than the gradient descent one can be used to reduce the number of required iterations, the general structure remains the same, involving the calculation of the χ^2 function gradient with respect to each model parameter (loop in lines 6–8). To calculate each gradient component $\nabla \chi_j^2$ (line 7) using a rather simple first-order method, one needs to modify the ψ_j variable by a small amount δ, resulting in the model vector $\psi' = (\psi_1, \psi_2, \cdots, \psi_{j-1}, \psi_j + \delta, \psi_{j+1}, \cdots, \psi_{N_\psi M_\psi})$, calculate a new self-consistent solution for ψ', and then calculate the new χ^2 value. Then

$$\frac{\partial \chi^2(\psi)}{\partial \psi_j} \approx \frac{\chi^2(\psi') - \chi^2(\psi)}{\delta}. \quad (12)$$

Evaluating Eq. (12) requires a NLTE forward solution, $O(N_\Omega N_\lambda N^3)$, and evaluating $\chi^2(\psi')$, $O(N_\lambda N^3)$. Therefore, for $N_\psi M_\psi$ components the calculation of the gradient (loop 6–8) scales as $O(N_\psi M_\psi (N_\Omega N_\lambda N^3 + N_\lambda N^3)) = O(N_\psi M_\psi N_\Omega N_\lambda N^3)$, which is completely dominated by the NLTE forward solution. Assuming that the number of inversion iterations is N_{it} (loop 4–11) and assuming that evaluating Eq. (12) requires a single Λ iteration, the total number of Λ iterations required for the inverse solution according to Algorithm 1 is equal to $N_{\text{it}} N_\psi M_\psi$, implying a total scaling of $O(N_{\text{it}} N_\psi M_\psi N_\Omega N_\lambda N^3)$.

We can now easily estimate the wall-clock time of a 3D inversion. Assuming, conservatively, that the number of inversion iterations is $N_{\text{it}} = 100$, considering that the number of MHD-like quantities would typically be of the order of $N_\psi = 10$, that the number of coefficients per quantity, M_ψ, is of the order of 10^2 (for very simple models) to 10^4 (more refined models; see below for further details), and that the CPU time for each Λ iteration is about 5.5 hours for a model with $N^3 = 100^3$ (Sect. 3.1), the total CPU time is $N_{\text{it}} N_\psi M_\psi \cdot 5.5$ hours, i.e., between 0.5 and 55 millions of CPU hours. Although enormous, such computing resources are available in today's HPC facilities. However, our experience shows that the approach sketched in Algorithm 1 often tends to end up in a local minimum of the χ^2 function and the calculation needs to be restarted. This happens already in the 2.5D problems we have tested and it can be expected that in full 3D geometry this problem is only going to get worse. In addition, due to the possible non-uniqueness of the solution, it is nevertheless worthwhile to repeat the calculations in order to discover alternative compatible solutions. Taking into account that spectropolarimetric observations with 100^2 pixels are very coarse by today standards, the CPU time necessary for grid-based 3D inversions quickly becomes unacceptable in practice.

4. Meshfree approach to the 3D NLTE inversion

In the multi-dimensional NLTE forward problem in solar-physics, the input model atmosphere is often the result of an MHD simulation, and the large-scale simulations of the outer solar atmosphere that are used in the NLTE synthesis are often computed using rectilinear Cartesian grids (e.g., Gudiksen et al. 2011). Such regular discretization is advantageous in the NLTE forward problem, but the existence of a grid is a complication in the sense that the radiation transfer needs to be performed in the particular topological order of the grid nodes and, consequently, the parallelization of the NLTE solution becomes a non-trivial task.

In the inverse problem we are not constrained by any a priori model to work with and, therefore, we are not obliged to adopt any particular spatial discretization. In this section, we introduce the basic ingredients and algorithms for a meshfree method for solving the 3DNIP. As shown below in this section, abandoning the space discretization entails a number of advantages, even though it requires to reformulate the standard inversion Algorithm 1. One of the main properties of the new method is that it allows to avoid the costly and repetitive application of the $\bar{\Lambda}$ operator (Sect. 2), at the expense of not automatically guaranteeing the full NLTE consistency of the problem at every step of the inversion.

4.1. Expansion of variables into a basis

In Sect. 3.2 we parameterized the spatial variation of each of the N_ψ MHD-like quantities, $\{\psi(\boldsymbol{r}; k)\}_{k=1}^{N_\psi}$, with M_ψ coefficients,

$\{\psi_i(k)\}_{i=1}^{M_\psi}$, which, together with the function $f(\cdot)$ in Eq. (7), allow us to compute the physical quantities at any point, r, within the medium. From now on, we particularize the general function $f(\cdot)$ to the linear function

$$\psi(\boldsymbol{r}; k) = \sum_{i=1}^{M_\psi} \psi_i(k)\phi_i(\boldsymbol{r}), \tag{13}$$

where $\{\phi_i(\boldsymbol{r})\}_{i=1}^{M_\psi}$ is a certain basis of orthonormal functions in the 3D computational domain, i.e.,

$$(\phi_i, \phi_j) = \int \mathrm{d}^3 r \, \phi_i(\boldsymbol{r})\phi_j(\boldsymbol{r}) = \delta_{ij}, \tag{14}$$

where $(\cdot, \cdot)$ stands for the inner product of the functions and the integration is done over the computational domain.

The choice of the basis functions is, in general, arbitrary, but different bases can have different degrees of suitability for approximating the distribution of quantities in different models and coordinate systems. A good approximation should fulfill $M_\psi \ll N^3$, so that the number of free parameters per quantity is much smaller than the number of mesh points in an equivalent 3D grid, and should allow the approximation of the spatial variation of the quantities on both large and small scales. Such an approximation is often possible because the physical quantities are at least piece-wise continuous. However, due to the nature of the NLTE inverse problem, it is not possible to know the most suitable basis a priori and, at the same time, it is usually not possible either to find an optimal sparse subset of the basis functions as in Asensio Ramos & de la Cruz Rodríguez (2015) due to the high CPU demands of such a process. Nevertheless, it is possible to improve the choice of the basis by considering heuristic methods and by using the fact that different quantities are often spatially (anti)correlated.

In this paper, we restrict ourselves to a fixed basis consisting of a set of typically smooth functions up to a certain order (such as in Appendix A.1, where we provide an example in the 3D Cartesian coordinates). In this sense, it is important to understand that the term "smooth model" that we occasionally use is often adequate but generally overly restrictive; our constraint is actually on the number of basis functions, M_ψ, rather than on their smoothness in any particular mathematical sense.

4.2. Sampling the radiation field in pilot points

What makes the 3DNIP problem more difficult than other inverse problems is that, even if we find the ψ correctly reproducing the observed data D, we still need to guarantee that the radiation and the atomic state are consistent at every point within the medium (see Sect. 2). Similarly to the expansion of the MHD-like variables in Sect. 4.1, we can do the same with the atomic-like quantities:

$$\xi(\boldsymbol{r}; k) = \sum_{i=1}^{M_\xi} \xi_i(k)\phi_i(\boldsymbol{r}), \tag{15}$$

where M_ξ is the number of basis functions. Analogously to Eq. (8), we can represent the information on these variables with a $N_\xi M_\xi$ elements vector,

$$\begin{aligned} \xi &= (\xi_1, \cdots, \xi_{N_\xi M_\xi}) \\ &= (\xi_1(1), \cdots, \xi_{M_\xi}(1), \xi_1(2), \cdots, \xi_{M_\xi}(N_\xi)). \end{aligned} \tag{16}$$

It is important to emphasize that, because we assume that in the family of models of interest it is also $N_\xi \ll N^3$, we can
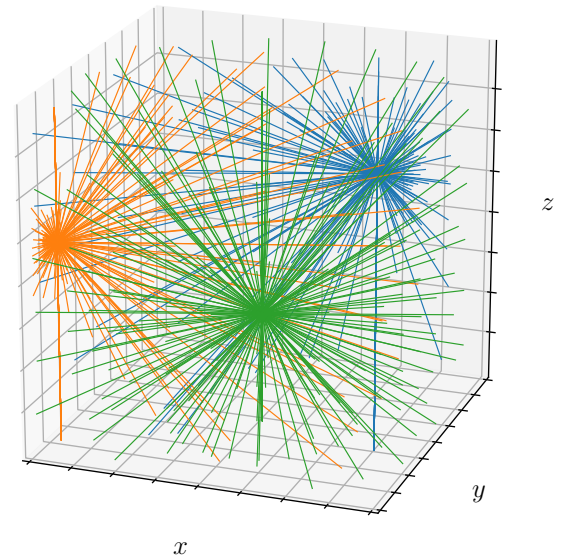
Fig. 2. Example of three pilot points and their associated long characteristics in a computational domain.

fully determine the ξ vector from a relatively small number of samples in the 3D domain. To show this, let $y(\boldsymbol{r})$ be an arbitrary real function that can be expanded in the $\{\phi_i(\boldsymbol{r})\}_{i=1}^M$ basis,

$$y(\boldsymbol{r}) = \sum_{i=1}^M q_i\phi_i(\boldsymbol{r}), \tag{17}$$

where the $q_i = (y, \phi_i)$ are the expansion coefficients, and let $\{\boldsymbol{r}_j\}_{j=1}^M$ be a set of points randomly distributed in the model domain. At each of these points Eq. (17) becomes $y(\boldsymbol{r}_j) = \sum_{i=1}^M q_i\phi_i(\boldsymbol{r}_j)$ or, equivalently, $\boldsymbol{y} = \boldsymbol{\Phi}\boldsymbol{q}$, where $\boldsymbol{y}$ and $\boldsymbol{q}$ are column vectors and $\boldsymbol{\Phi}$ is a M by M matrix with elements $\Phi_{ji} = \phi_i(\boldsymbol{r}_j)$. In our approach, we restrict ourselves to the sets of random points for which the matrix $\boldsymbol{\Phi}$ can be inverted which are, from a practical point of view, the vast majority of cases. Due to isomorphism of $\boldsymbol{q}$ and $\boldsymbol{y}$, any random sample $\boldsymbol{y}$ uniquely determines the vector $\boldsymbol{q} = \boldsymbol{\Phi}^{-1}\boldsymbol{y}$.

Consequently, if we apply the same reasoning to the atomic-like variable $\xi(\boldsymbol{r}; k)$ sampled in $N_\Lambda = M_\xi$ random points[3], these samples are sufficient to determine the $\xi(\boldsymbol{r}; k)$ vector in the whole 3D domain. Because $N_\Lambda \ll N^3$, such condensed description could lead to a significant optimization of the numerical solution in comparison to the full Λ iteration. From now on, we refer to the random points used for sampling the atomic-like quantities as pilot points.[4]

Regardless of the particular representation of the atomic-like quantities, their sampling in the pilot points needs to be done via the calculation of the mean radiation field tensors $\bar{J}_Q^K(\boldsymbol{r}_j; \ell)$

---

[3] We introduce here the new variable N_Λ to represent the number of sampling points and we make it equal to the number of basis functions, although it will be shown in Sect. 5 that it can be advantageous to give different values to these variables.

[4] We note that the same term is used in the geophysical inverse problem research in a slightly different context (Doherty et al. 2010).

(see Sect. 2). From these, the density matrix elements are easily obtained by solving the ESE.

Given an estimate of the ψ and ξ vectors, we can get the mean radiation field at the $\{r_j\}_{j=1}^{N_\Lambda}$ points by solving the RTE using the long characteristics (LC) method (Hubený & Mihalas 2014; de Vicente et al. 2021) with a suitable angular quadrature with N_Ω propagation directions (see Fig. 2). Conservatively assuming, for simplicity, that the number of discrete steps along the LC is N, then the time required for the calculation of the mean radiation field tensors is $O(N_\Lambda N_\Omega N_\lambda N)$, while the full Λ iteration is of $O(N_\Omega N_\lambda N^3)$. To calculate the χ^2 function, we need to solve the RTE in N^2 pixels in order to get the emerging Stokes parameters in the whole field of view. This can also be done with LC parallel to the LOS, once per pixel, and this formal solution is then of the order of $O(N^2 N_\lambda)$.

The first step towards the meshfree method discussed in this subsection is still conceptually similar to the traditional approach based on the Λ-iteration approaches: Instead of determining the radiation field and atomic state (represented formally by y in Eq. 17) in all the points of a fine-spaced rectilinear grid, one can only consider it in a smaller set of pilot points. Then the inverse of the Φ matrix provides us with the model parameters q. These can be used to reconstruct the atomic state at all points in the domain. This information can subsequently be used in the radiative transfer calculations along the long characteristics going through the whole domain. In the following section we show that this naive generalization does not bring many benefits over the traditional Λ-iteration approach and a more radical idea is needed.

4.3. 3DNIP as an unconstrained global minimization problem

The solution of the 3DNIP in the meshfree representation of quantities could be done with minor modifications to Algorithm 1. The difference would be in the application of the $\bar{\Lambda}$ operator at lines 2, 7, and 10 of the algorithm. This operator would be evaluated just in $N_\Lambda = M_\xi$ pilot points instead of in the whole 3D grid. However, a closer inspection of this approach reveals that it is seriously deficient.

At line 7 in Algorithm 1, we modify ψ_j by a small amount δ, we solve the ESE in the pilot points, calculate the new radiative transfer coefficients along the LC, and then we solve the RTE in order to get the new values of the $\bar{J}_Q^K(r_j)$ tensors. If δ is small, only a few such iterations are necessary. Finding $\bar{J}_Q^K(r_j)$ in the pilot points is of the order of $O(N_\Lambda N_\Omega N_\lambda N) = O(M_\xi N_\Omega N_\lambda N)$. The problem is that once the $\bar{J}_Q^K(r_j)$ in the pilot points have been found (or, equivalently, the new values of $\xi(r_j, k)$, see Sect. 4.2), one needs to find the expansion coefficients $\xi_i(k)$ of Eq. (15) necessary for evaluating the modified χ^2 function. This operation involves the Φ^{-1} matrix multiplication (see the previous section). With N_ξ atomic-like quantities, this operation is of the order of $O(N_\xi M_\xi^2)$. Moreover, evaluating the radiative transfer coefficients for the χ^2 calculation is of the order $O(N_\xi M_\xi N^3)$ because the RTE coefficients are linear combinations of M_ξ basis functions of N_ξ quantities evaluated in N^3 points. Finally, the calculation of the χ^2 is then of the order of $O(N^3 N_\lambda)$.

Consequently, the loop 6–8 would be $O(N_\psi M_\psi (M_\xi N_\Omega N_\lambda N + N_\xi M_\xi^2 + N_\xi M_\xi N^3 + N^3 N_\lambda)) = O(N_\psi M_\psi (M_\xi N_\Omega N_\lambda N + N_\xi M_\xi N^3 + N^3 N_\lambda))$, taking into account that $M_\xi < N^3$. In any problem of reasonable complexity, the dominant term is $N_\xi M_\xi N^3$, hence the computing demands of the method can be written as $O(N_\psi M_\psi N_\xi M_\xi N^3)$.
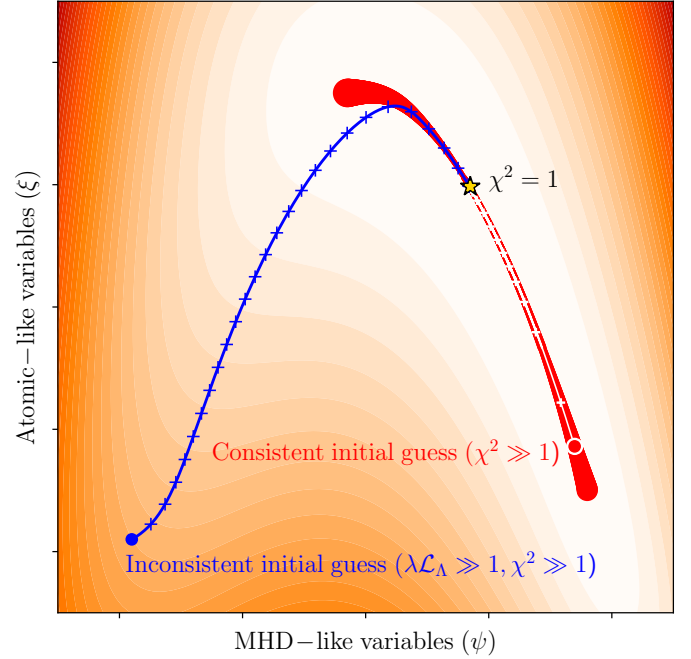

Fig. 3. Schematic visualization of the inversion process in the space spanned by the MHD-like and atomic-like variables. The background color indicates deviation from the NLTE consistency, $\mathcal{L}_\Lambda$, for every combination of the MHD- and atomic-like variables, with the brightest color representing the minimum value. The red line shows the common approach to the NLTE inversion as a sequence of physically consistent NLTE models of decreasing χ^2 value (constrained minimization). The blue line shows an unconstrained minimization with allowed inconsistent solutions during the inversion process.

In Sect. 3.3 we estimated the CPU time demands per iteration of the grid-based method to be $O(N_\psi M_\psi N_\Omega N_\lambda N^3)$. Since in realistic models we can expect $N_\xi M_\xi > N_\Omega N_\lambda$, the meshfree approach seems to bring no advantage with respect to the grid-based methods. In order to make the mesh-free method really efficient, we need to avoid the approach based on the inversion of the Φ matrix and to develop a different technique.

Algorithm 1 is an example of a constrained minimization of $\chi^2(D; \psi, \xi)$ in which we impose on ξ the constrain that the problem is always NLTE consistent. At every step of the solution, the ξ variables are fully determined by ψ: $\xi = \xi(\psi)$, hence we can write $\chi^2(D; \psi, \xi) = \chi^2(D; \psi)$ with the implicit condition on the NLTE consistency. The inversion process can be understood as a sequence of solutions in the space spanned by ψ and ξ such that, at every step, the model is NLTE consistent. In Fig. 3, we represent such a process with the red curve, whose thickness is proportional to the value of χ^2. In the context of Algorithm 1, the horizontal axis stands for the independent variables while the vertical axis stands for the dependent ones. Algorithm 1 follows the consistency curve until a model with $\chi^2 \approx 1$ is found. In order to benefit from the meshfree formulation, we need a more significant change in the inversion algorithm.

4.3.1. NLTE consistency regularization

In order take advantage of the meshfree formulation of the inverse problem we need to substantially revise the traditional inversion algorithm. Let us define a vector of all model variables,

θ, as a union of the ψ and ξ vectors,

$$\theta = \{\theta_1, \cdots, \theta_{N_V}\} = \{\psi_1, \cdots, \psi_{N_\psi M_\psi}, \xi_1, \cdots, \xi_{N_\xi M_\xi}\}, \qquad (18)$$

where

$$N_V = N_\psi M_\psi + N_\xi M_\xi \qquad (19)$$

is the total number of variables in the model, and let us formulate the inverse problem as an unconstrained minimization in which we let all the θ variables acquire any value independently of each other. Instead of the inverse problem of Eq. (10), that is based on the minimization of the χ^2 function, we can base the inverse problem on the minimization of a more general loss function,

$$\mathcal{L}(D; \theta) = \chi^2(D; \theta) + \lambda \mathcal{L}_\Lambda(\theta), \qquad (20)$$

where $\chi^2(D; \theta)$ is the χ^2 function of Eq (9) but now calculated using the whole set of generally physically inconsistent parameters θ, λ is some positive constant, and $\mathcal{L}_\Lambda(\theta)$ is a non-negative differentiable function that stands for a penalty for deviation of θ from the NLTE consistency in the pilot points. In this formulation we allow the model parameters to take unphysical values during the inversion, but deviations from self-consistency are penalized. The goal of the minimization is to find the lowest value of $\mathcal{L}$, which is equivalent to finding the minimum value with $\chi^2 = 1$ and $\mathcal{L}_\Lambda = 0$, corresponding to a NLTE-consistent solution that simultaneously fits the observed data. In Fig. 3, such inversion process is schematically shown with the blue curve. Starting from a physically inconsistent model that does not fit the data, we aim to find the solution that is simultaneously consistent and fits the observations. Comparing Eqs. (20) and (11), we can interpret the $\lambda \mathcal{L}_\Lambda$ term as a regularization of the χ^2 minimization problem due to the requirement of NLTE consistency.[5]

As mentioned above, the loss function $\mathcal{L}_\Lambda(\theta)$ should be differentiable and its minimum value should be equal to zero for solutions that are NLTE-consistent at every pilot point, $\boldsymbol{r}_j$, i.e.,

$$\xi(\boldsymbol{r}_j; k) = \tilde{\xi}(\boldsymbol{r}_j, \theta; k), \qquad (21)$$

where $\xi(\boldsymbol{r}_j, k)$ is the atomic-like k-th variable calculated at the pilot point $\boldsymbol{r}_j$ from the current guess of the ξ vector and $\tilde{\xi}(\boldsymbol{r}_j, \theta; k)$ is the same atomic-like variable, at the same pilot point, resulting from the radiative transfer via LC and for the given model parameterization θ. A suitable choice for $\mathcal{L}_\Lambda$ is the norm

$$\mathcal{L}_\Lambda(\theta) = \frac{1}{N_\Lambda N_\xi} \sum_{j=1}^{N_\Lambda} \sum_{k=1}^{N_\xi} \left| \xi(\boldsymbol{r}_j; k) - \tilde{\xi}(\boldsymbol{r}_j, \theta; k) \right|^2 . \qquad (22)$$

The λ parameter in Eq. (20) needs to be estimated based on the particular problem and the required accuracy. If we define an acceptable error $\Delta \xi^2$, λ should fulfill $\lambda \Delta \xi^2 > 1$ in order to have the second term in Eq. (20) of at most the same order of magnitude as χ^2 in the final solution. Therefore, a rough estimate of λ can be made so that $\lambda > 1/\Delta \xi^2$. In practice, this choice needs to be done with care and with respect to each particular model and its parameterization. Moreover, the λ parameter does not have to be constant: the decrease of its value as the solution approaches convergence can reduce the oscillatory behavior of

---

[5] This analogy needs to be taken with care because regularizations are usually conditions imposed regardless of the physical model, but we can understand this particular regularization in the context of models neglecting the NLTE transfer effects — in this regard, the $\lambda \mathcal{L}_\Lambda$ term can be seen as a natural physical regularization imposing NLTE consistency.

the loss function and lead to better convergence, although we do not further discuss this in this work.

In addition to the NLTE consistency, the form of the loss function in Eq. (20) allows us to include additional penalties, such as deviations from local physical-laws consistency (in the sense that they are independently fulfilled at every point, in contrast with the radiatively coupled NLTE consistency). We include these penalties in an additional term $\gamma \mathcal{L}_L$ on the right-hand side of Eq. (20), i.e.,

$$\mathcal{L}(D; \theta) = \chi^2(D; \theta) + \lambda \mathcal{L}_\Lambda(\theta) + \gamma \mathcal{L}_L(\theta), \qquad (23)$$

where γ is another positive parameter of the problem (and similar considerations to those about the λ parameter above can be made about its value). Examples of what we mean by deviation from local physical consistency are the existence of negative atomic populations, the non-zero divergence of the magnetic field vector $\nabla \cdot \boldsymbol{B}$, or the deviation from magnetohydrostatic equilibrium of the plasma. In the stationary models that we consider in this paper, it is also straightforward to include a penalty for deviations from the equation of continuity $\nabla \cdot (\rho \boldsymbol{v}) = 0$ or other MHD equations.

The term $\mathcal{L}_L$ in the loss function can thus be understood as an additional natural regularization due physical consistency. It should be evaluated in a random set of N_L points that can typically be $N_L \gg N_\Lambda$ while still requiring a negligible amount of CPU time in comparison to $\mathcal{L}_\Lambda$ because there is no need to perform radiative transfer calculations in order to evaluate it.

Likewise, $\mathcal{L}_L \to 0$ as we are reaching a physically consistent solution. In Appendix A.3 we briefly discus the construction of the $\mathcal{L}_L$ penalty. In general, we want this function to be a convex differentiable function of θ with the minimum $\mathcal{L}_L = 0$ for the locally consistent solutions.

In summary, with the loss function given by Eq. (23), the 3DNIP problem can be formulated as

$$\hat{\theta} = \arg \min_\theta \mathcal{L}(D; \theta), \qquad (24)$$

that is, to find the estimate $\hat{\theta}$ for the given data D such that the $\mathcal{L}(D; \theta)$ loss function has its minimum value equal to 1. As we will see below, the condition $\mathcal{L}_\Lambda = \mathcal{L}_L = 0$ will be rarely satisfied exactly in practice, hence, assuming a suitable choice of λ and γ a typical solution fulfills $\chi^2 \approx 1$, $\mathcal{L}_\Lambda \ll 1$, and $\mathcal{L}_L \ll 1$.

4.3.2. The new inversion algorithm

Algorithm 2 3DNIP solution as an unconstrained $\mathcal{L}$ minimization.

1: **Initialization**: Randomly initialize the model parameters $\theta(0)$.
2: $i \leftarrow 0$
3: **repeat** {**Descent along the negative $\mathcal{L}$ gradient.**}
4:    $i \leftarrow i + 1$
5:    **for** $j = 1$ to N_V **do** {**Loop over the model parameters.**}
6:       Calculate the gradient element $\nabla_j \mathcal{L} = \partial \mathcal{L}(D; \theta(i - 1))/\partial \theta_j$.
7:    **end for**
8:    $\theta(i) \leftarrow \theta(i - 1) - h \nabla \mathcal{L}$. {**New estimate of the model parameters.**}
9: **until** $\mathcal{L} \approx 1$ (or, explicitly, $\chi^2(D; \theta(i)) \approx 1$ and $\mathcal{L}_\Lambda \ll 1$ and $\mathcal{L}_L \ll 1$).

We are now ready to devise a new inversion algorithm for the unconstrained NLTE inversion, Algorithm 2. Even though it

seems similar in structure to Algorithm 1, there are some very important differences. First, we randomly initialize all the variables θ, including the atomic-like variables ξ. Secondly, the main loop (lines 3–8) runs until a good fit to the data is found that, at the same time, is physically consistent. Finally, the gradient of the loss function in the inner loop (lines 5–7) is calculated with respect to all θ variables, that is, to both MHD-like, ψ, and atomic-like, ξ, variables.

In contrast to the step in line 7 in Algorithm 1, there is no need to evaluate the exact ξ variables from the radiation field tensors in pilot points because the ξ variables are explicitly known at every step. A small perturbation of any of the ψ_j or ξ_j coefficients by δ leads to a small modification of the RTE coefficients along the LC of the pilot points, whose calculation scales as $O(N_\Lambda N_\mathbf{\Omega} N_\lambda N) = O(M_\xi N_\mathbf{\Omega} N_\lambda N)$ (remember we imposed $N_\Lambda = M_\xi$ in Sect. 4.2). We emphasize that in this new approach the change of ψ variables is not followed by the calculation of the corresponding changes of ξ variables, because all the variables are now independent.

The calculation of the loss function gradient consists of three terms,

$$\frac{\partial \mathcal{L}}{\partial \theta_j} = \frac{\partial \chi^2}{\partial \theta_j} + \lambda \frac{\partial \mathcal{L}_\Lambda}{\partial \theta_j} + \gamma \frac{\partial \mathcal{L}_L}{\partial \theta_j} . \tag{25}$$

Regarding the scaling, the first term on the right-hand side is of the order of $O(N_\lambda N^3)$ (Sect. 3.3), the second one is of the order of $O(N_\Lambda N_\mathbf{\Omega} N_\lambda N) = O(M_\xi N_\mathbf{\Omega} N_\lambda N)$, and the last term is of the order of $O(N_L)$. Even though the inner loop in Algorithm 2 is over N_V variables, which is larger than $N_\psi M_\psi$ in Algorithm 1 because we need to explicitly consider the atomic-like variables, the order of magnitude of N_V and $N_\psi M_\psi$ is expected to be the same in practical applications if the number of spectral lines in the problem is relatively small. The comparison with the grid-based scaling of $O(N_\psi M_\psi N_\mathbf{\Omega} N_\lambda N^3)$ shows that, if the number of iterations of both inversion methods is similar, the meshfree algorithm can be faster because all three derivatives on the right-hand side of Eq. (25) are significantly faster than the NLTE-consistent derivative $\partial \chi^2(D; \psi)/\partial \psi_j$ of Algorithm 1.

In this work we have assumed that the set of basis functions (and, in particular, its dimension) of the MHD-like and atomic-like quantities (Eqs. (13) and (15), respectively) are known. In some particular cases they can be estimated empirically but, in general, neither M_ψ nor M_ξ are known a priori and need to be determined during the inversion process. M_ψ must be such that the model is capable to fit the observed data and N_L must be such that the computational domain is sufficiently finely sampled so that we can guarantee consistency of the physical quantities. This work is meant to be a discussion of the general framework of the inversion method rather than an exhaustive guideline for practical inversions and, consequently, we do not discuss here in detail the values of M_ψ and N_L. Instead, we simply assume that optimal values of M_ψ, $N_\Lambda = M_\xi$, and N_L are known and fixed.

However, what we want to stress here is one important aspect related to the NLTE-consistency criteria. In the grid-based methods, it is possible to reach NLTE consistency with a grid of any coarseness. Indeed, this solution is only approximate due to the finite discretization of the medium: it does provide the solution in the grid points, but the values of the atomic-like variables are only approximate and dependent on the discretization (Auer et al. 1994). Similarly, if we approximate the spatial distribution of the atomic-like variables by an expansion in M_ξ basis functions, we can find a consistent solution in $N_\Lambda = M_\xi$ pilot points but, due to the non-linearity of the NLTE problem, the consistency is not fully guaranteed in other points unless M_ξ is so high

(potentially infinite) that the solution is practically exact. In both grid-based and meshfree methods, we need to find the right compromise between accuracy and computation time.

In the grid-based methods one can estimate the error in the self-consistent solution by considering grids with different spatial resolutions (see Trujillo Bueno & Fabiani Bendicho 1995). Similarly, in the meshfree method we can increase the number of basis functions and of pilot points. When a sufficiently big $N_\Lambda = M_\xi$ is reached so that any increase in its value leads to a negligible change of $\mathcal{L}_\Lambda$ it is an indication of the achieved sufficient NLTE consistency. We will come back to this problem in Sect. 5.

4.4. Parallelization and memory requirements

The parallelization of multi-dimensional grid-based NLTE codes is difficult because the RTE needs to be solved following a particular order of grid points which depends on the radiation propagation direction. Different techniques based on domain decomposition and parallelization in the propagation directions and wavelengths have been developed in the past, e.g., the PORTA code (Štěpán & Trujillo Bueno 2013) implements domain decomposition only in the vertical direction, resulting in good scaling with the number of CPUs but relatively high memory constraints, while the MULTI3D code (Leenaarts & Carlsson 2009) implements a 3D domain decomposition, allowing for a larger distribution of sub-domains among CPUs but resulting in relatively worse scaling due to the need of iterating the boundary conditions of the domains.

One of the advantages of our meshfree method is that parallelization of the inversion is straightforward: Each of the N_V variables in loop 5–7 in Algorithm 2 can be treated independently. Moreover, the RTE can also be solved independently for every LC associated with the pilot points or field of view pixels. This allows us to fully parallelize the solution with up to $N_V N_\Lambda N_\mathbf{\Omega}$ (or $N_V N^2$ in the case of the χ^2 evaluation) processes and the scaling with the number of processes will be practically linear. Given that these numbers exceed the tens of thousands in any practical application, the parallelization can be massive.

It follows, from its parameterization, that the amount of memory needed to store the model is proportional to N_V. For models that are computable in a reasonable time with current supercomputers, N_V would probably not exceed the order of a million significantly, what implies tens of megabytes to store the model. Consequently, the whole model parameterization can be stored in memory at every parallel process and no domain decomposition is ever needed.

However, from the computational point of view, it is advantageous to keep in the computer memory the LC of the pilot points and of the output pixels during the $\nabla \mathcal{L}$ calculations, as well as some information on the RTE coefficients. This guarantees that fast corrections of the RTE coefficients can be calculated in the inner loop of Algorithm 2 and the scaling discussed in Sect. 4.3 is guaranteed. The amount of these data can become too large with an increasing number of pilot points N_Λ or resolution of the observation. We will solve this drawback of the meshfree method in Sect. 5.

4.5. Suitability of the meshfree method

Let us study the suitability of the meshfree method by comparing the required computation time of both meshfree and grid-based methods for a broad set of models. For this analysis we assume
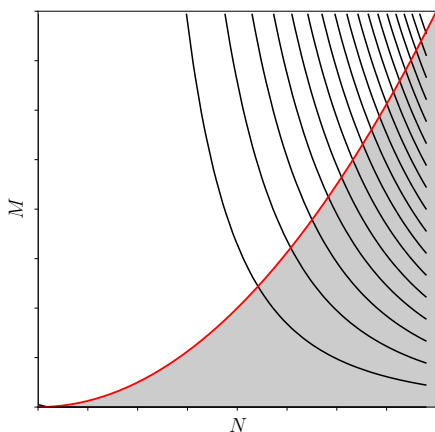
Fig. 4. Comparison between the efficiency of the meshfree and grid-based methods. The horizontal axis shows the grid resolution, while the vertical axis shows the number of coefficients per model parameter. The gray shaded area below the red curve $M = N^2$ shows the models more efficiently inverted using the meshfree method. The black curves connect models with identical required CPU time per iteration of a grid-based model. See the main text for further details.

that the number of iterations N_{it} needed by both methods is similar. It is difficult to analytically show that this is the general case, but our numerical experiments with 2.5D grid based and 3D meshfree models indicate that this is likely the case in many situations. Under this assumption, the comparison of the solution with both methods is equivalent to compare just one iteration.

A straightforward condition that the meshfree method must fulfill to be more efficient is that first and second terms on the right-hand side of Eq. (19) must be of comparable order, that is, the number of θ variables should not be too much larger than the number of ψ variables. Given that the number of spectral lines in the model is not too large, this will usually be the case.

By comparing the scaling of the calculation of each gradient component between the meshfree ($O(N_\lambda N^3 + M_\xi N_{\boldsymbol{\Omega}} N_\lambda N + N_{\text{L}})$, Sect. 4.3.2) and the grid-based ($O(N_{\boldsymbol{\Omega}} N_\lambda N^3)$, Sect. 3.3) methods, we can roughly estimate that the former will be faster if

$$\frac{1}{N_{\boldsymbol{\Omega}}} + \frac{M_\xi}{N^2} + \frac{N_{\text{L}}}{N_{\boldsymbol{\Omega}} N_\lambda N^3} \sim \frac{M_\xi}{N^2} < 1 \tag{26}$$

is satisfied, where we have taken into account that the first and third terms on the left-hand side of the inequality are clearly negligible.[6] In Fig. 4 we show this resulting condition with the shaded region below the red curve in a diagram of number of basis functions M per quantity versus the grid resolution. Assuming $M_\xi \approx M_\psi$, M represents any of the MHD- or atomic-like quantities. The gray area where the condition in Eq. (26) holds shows the models that are more efficiently calculated using the meshfree method.

We can thus conclude that the meshfree method is superior for models/observations with high resolution and somewhat limited spatial variability of the parameters (smaller M). Likewise, the grid-based methods can be more efficient in the case of low-resolution models/observations with abrupt spatial variability of the parameters.

The black curves in Fig. 4 correspond to curves with constant MN^3, proportional to the CPU time per iteration of the grid-based method. The area below each of them corresponds to

the models that can be inverted within a given maximum available CPU time. One important observation is that if one is given a limited CPU time (i.e., a particular black curve in the plot) and a particular model/observation resolution N, then it is always possible to find a meshfree solution for certain M below the red curve. Given the parabolic shape of the red curve, it follows that grid-based methods are increasingly less efficient compared to the meshfree method as the model/observation resolution increases.

5. Stochastic inversion

In the last section we have introduced several concepts leading to the formulation of a meshfree method to solve the 3DNIP. Even though the method seems promising, there are some problems that need to be solved to achieve a truly efficient algorithm, such as the fact that with $N_\Lambda = M_\xi$ fixed pilot points we cannot fully guarantee the NLTE consistency in the whole computational domain, or the memory allocation of a large number of LC. In addition, in this section we will discuss the problem of local minima of $\mathcal{L}$ and the artifacts introduced by the use of fixed angular quadratures in 3D.

5.1. Reshuffling of the pilot points and pixels: the stochastic algorithm

One of the issues of the meshfree method as introduced in Sect. 4, which we mentioned in Sect. 4.3.2, is the impossibility of ensuring the NLTE consistency in the whole domain by testing this consistency in just $M_\xi \ll N^3$ random pilot points. One possible solution to this problem could be, after the solution has converged, to change the set of pilot points and to check whether the solution is still NLTE consistent. While this approach would result in a slower method (we would be increasing the number of required iterations without changing the cost per iteration), it directs us toward a different approach: we can generate a new set of pilot points after every $n < N_{\text{it}}$ iterations, even before the convergence is reached for a given set of pilot points. In particular, we can use a small number of pilot points $N_\Lambda \ll M_\xi$, which are reshuffled after every n iterations (with a lower limit of $n = 1$). This allows us to eventually sample the whole 3D domain. Of course, changing the set of pilot points before reaching convergence leads to a much larger final number of iterations, but these are also much faster than in the standard algorithm, because of the condition $N_\Lambda \ll M_\xi$.

With this approach, estimating the $\mathcal{L}_\Lambda$ gradient becomes of the order of $O(N_\Lambda N_{\boldsymbol{\Omega}} N_\lambda N)$, much smaller than $O(M_\xi N_{\boldsymbol{\Omega}} N_\lambda N)$. The most time consuming part of the $\mathcal{L}$ gradient estimation would now be the χ^2 function gradient, still of the order of $O(N_\lambda N^3)$. We would have a very fast although inaccurate estimation of the $\mathcal{L}_\Lambda$ gradient, but an accurate and very slow estimation of the χ^2 gradient. We can thus balance the computing time of these quantities by applying a similar strategy to the calculation of the χ^2 gradient, namely, instead of evaluating χ^2 in N^2 pixels, we randomly generate $N_\chi < N^2$ pixels, to be reshuffled after every iteration, in which we calculate an approximation of this quantity. The balancing of the scaling of the $\mathcal{L}_\Lambda$ and χ^2 gradients can be done by taking $N_\Lambda N_{\boldsymbol{\Omega}} \approx N_\chi$.

Naturally, it is also possible to decrease N_{L} and to apply the same reshuffling strategy to the estimation of the $\mathcal{L}_{\text{L}}$ gradient.

This procedure can only provide an approximation to the gradient $\nabla \mathcal{L}$, but every iteration becomes significantly faster than before. This approach is used in the so-called stochastic gradient

---

[6] In the 3D radiative transfer we always have $N_{\boldsymbol{\Omega}} \sim 10^2$.

descent (SGD) class of methods and it provides several benefits over the traditional gradient descent (see below). The approximate loss function now takes the form

$$\mathcal{L}(D; \theta; N_\chi, N_\Lambda, N_L) =$$
$$\chi^2(D; \theta; N_\chi) + \lambda \mathcal{L}_\Lambda(\theta; N_\Lambda) + \gamma \mathcal{L}_L(\theta; N_L). \quad (27)$$

The expression for χ^2 is identical to that in Eq. (9), except for N_{pix} being replaced by N_χ. The general structure of Algorithm 2 remains valid, but the loss function at line 6 is replaced by that of Eq. (27). Moreover, the convergence criteria at line 9 must be modified in such a way that the convergence is guaranteed not only for a single sample of N_Λ and N_χ points, but for the whole domain. To this end, the stopping criteria should not only involve the current value of the loss function components and of their gradients, but also their values over time. We leave the discussion of this technical issue for later publications.

One last change is necessary at line 8 of Algorithm 2, where new values of the model parameters are calculated using the estimation of the gradient. A number of SGD algorithms have been developed in the recent years which lead to much better convergence rates than the naive approach outlined in Algorithm 2. In our calculations we have used the ADAM algorithm (Kingma & Ba 2014), which uses running averages of the gradient and of its second moment in order to estimate the new values of the problem variables. We have found that it provides good convergence results for the test presented in this work (see Sect. 6 for details).

In the stochastic approach outlined in this section we have replaced the small number of computationally intensive iterations from the method in Sect. 4 by a large number of very fast iterations. With a sufficient number of iterations, we can guarantee that the solution is consistent in the whole domain and not just in M_ξ fixed points. As we will show in the example below (Sect. 6), the number of stochastic iterations exceeds the thousands or tens of thousands, hence the computational domain can be effectively sampled with a small number of N_Λ, N_χ, and N_L points.

Note that, even though we are testing the NLTE consistency in just N_Λ pilot points during each iteration, we are also effectively probing the model along the LC themselves (cf. Fig. 2). Due to the non-local character of the radiative transfer problem, an error accumulated along the LC will produce an error at the pilot points, in contrast to the loss function of local quantities, $\mathcal{L}_L$.

Finally, the convergence analysis in the SGD is more difficult and thus we will discuss it in future publications. In this work we restrict ourselves to an empirical convergence test in Sect. 6 which shows that the method can converge very quickly. The numbers N_Λ, N_χ, and N_L determine how noisy the gradient and the convergence are, with higher values leading to a better convergence rate with a less noisy gradient at the cost of larger CPU time and memory requirements per iteration.

5.2. Convexity, local minima, and an analogy with deep learning

The NLTE problem is strongly non-linear. We are not aware of any existing rigorous mathematical analysis of the equations for multilevel systems, but it is very likely that the χ^2 and $\mathcal{L}$ functions are also non-convex; hence there are a number of local minima in which the inversion algorithm can end up. In fact, our numerical experiments with deterministic inversion algorithms show that it is not a rare occurrence.

The use of a SGD method helps to partially solve this problem because, unlike deterministic methods, the stochastic crawl-
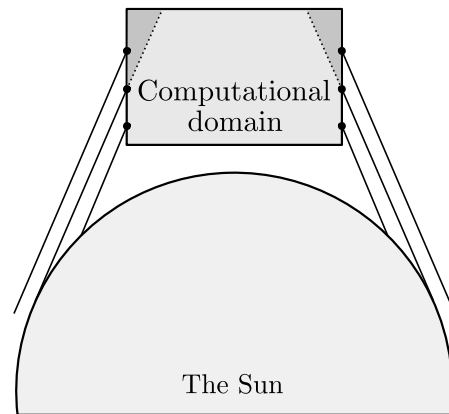

Fig. 5. Geometry of the illumination of the computational domain by the underlying solar surface using an angular quadrature. If the angular quadrature is identical at every point (see the inclined lines), artificial sharp jumps in the illumination at the boundary, which do not actually exist, will appear in the model. Such artifacts could make a full 3D inversion nearly impossible. The solution to this problem is to use a randomized quadrature at every pilot point.

ing through the parameter space in the SGD method is not condemned to remain in a local minimum. This is simply due to the fact that the exact shape of the loss function landscape (see also the background color in Fig. 3) is unknown and its estimation changes between iterations, hence the local minima may be passed through, in contrast to what happens in the standard gradient descent method with a more accurate and unchanging estimation of the loss function.

SGD methods have recently become important in the context of machine learning techniques. While our method is not based on these, an analogy can be made with the training of deep neural networks: As in our method, deep learning can be understood as a global optimization process with a substantial number of parameters. The network training proceeds by feeding a large example data set and evaluating the gradient of the loss function. In practice, it is unfeasible to use the whole database of examples in every iteration of the training. Instead, one feeds the network with a relatively small number of randomly chosen examples and the loss function gradient is only approximately calculated (the so-called mini-batch training). The use of a small number of random pilot points per iteration in our inversion method can be seen as analogous to the mini-batch training, while the calculation of the loss function gradient in all domain points would correspond to using the whole database of examples in every iteration of the training.

5.3. Angular quadratures artifacts

The accurate calculation of the $\bar{J}_Q^K$ tensors in the pilot points requires to solve the RTE along the LC in the directions of a sufficiently accurate angular quadrature. For a discussion of optimal quadratures for the transfer of polarized radiation see Štěpán et al. (2020) and Jaume Bestard et al. (2021b).

Under the physical conditions in the solar atmosphere, the number of propagation directions of a good quadrature is typically of the order of $N_\Omega \sim 10^2$. As with the truncation error in the grid-based methods, the angular discretization necessarily leads to some degree of numerical error, but if an appropriate quadrature is chosen, this error can be negligible. However, a problem arises with the interaction between the quadrature and

the boundary conditions of a 3D domain in the case of plasma structures embedded in the solar corona, i.e., in structures such as prominences and filaments. Given a fixed quadrature and the illumination from the underlying solar surface, there will be a sharp and artificial jump in the boundary conditions as demonstrated in Fig. 5.

An easy solution to this problem is possible in the framework of the stochastic approach: In the same way that the pilot points are randomized in every iteration, the orientation of the quadrature rays can also be randomized in each of these points. As a result, there are no preferable directions in the radiation transfer and artifacts similar to that in Fig. 5 cannot appear. For obvious reasons, this randomization of the angular quadrature orientation cannot be implemented in the grid-based methods relying on the short-characteristics method.

In our 2.5D grid-based experiment we have found that these artifacts can actually appear in practice and they significantly complicate the solution. As a byproduct of the stochastic approach, we have the means to solve another critical problem that would actually make the full 3D inversion practically impossible.

5.4. Parallelization and memory requirements

The parallelization of the stochastic method is as straightforward as in the deterministic method described in Sect. 4. Given that in every iteration we need to solve the radiative transfer along $N_V(N_\Omega N_\Lambda + N_\chi)$ independent long characteristics, and since in practice we always have N_V, $N_{pix} > 10^2$ and $N_\Omega \approx 10^2$, and N_Λ is such that, typically, $N_\Omega N_\Lambda \approx N_\chi \gtrsim 10^2$, it is clear that the method can be massively parallelized and that the scaling with the number of CPUs should be practically linear.

One of the most significant benefits of the stochastic method over the deterministic one is its random access memory requirement. As discussed in Sect. 4.4, it is necessary to store the data of the LC during the $\nabla\mathcal{L}$ estimation, and increasing M_ξ and N could dramatically increase the memory requirements to the point of being a limiting factor of the method. However, in the stochastic method $N_\Lambda \ll M_\xi$ and $N_\chi \ll N^2$ and thus the issue is nonexistent. Moreover, we can always choose N_Λ and N_χ such that the method fits any memory constraint.

6. Example application

In this section we apply the previously developed stochastic meshfree method to invert the physical properties of a solar prominence-like 3D structure. For the sake of simplicity in the demonstration of the method, and without loss of generality, we use a dimensionless academic model. The chosen model is intentionally very simple and its purpose is to test the inversion algorithm.

6.1. Model definition and synthetic observations

The computational domain is a 3D cube with dimensions $[-1, 1]^3$ suspended above the solar surface and observed along an LOS perpendicular to one of its faces (see Fig. 6). In this example we assume that the scattering geometry is fully known a priori during the inversion process and that we can expect the observed plasma to be completely confined within the cubic box. Needless to say that none of these assumptions would be generally satisfied in an actual observation. The boundary conditions for the illumination are chosen such that they resemble the irra-
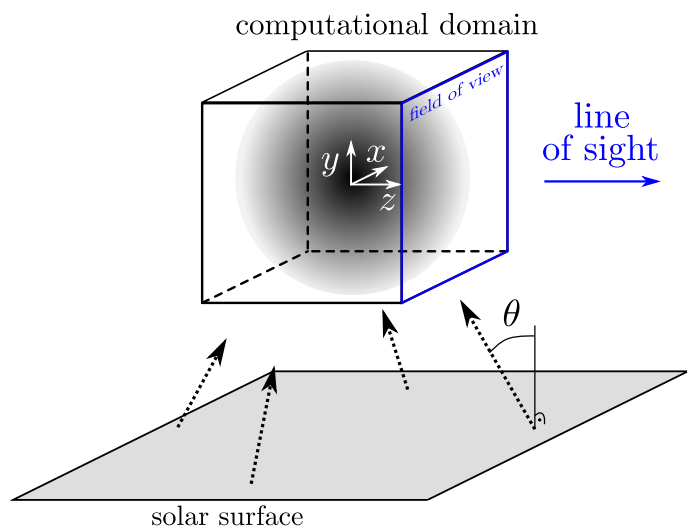
Fig. 6. Geometry of the model and the observation. The computational domain (black cube) is illuminated by the underlying Sun (gray surface). The limb darkening of the incident continuum radiation depends on the θ heliocentric angle (see the text for details). The LOS, which is in the direction of the z axis, and the local solar vertical direction (y axis) form a 90° angle. In this academic example, the solar surface is approximated by an infinite plane.

diation from the underlying solar photosphere. This unpolarized spectrally flat radiation is limb-darkened according to the rule

$$I(\mu) = \begin{cases} 1 - \frac{1}{2}(1-\mu^2) & \text{for } \mu > 0 \\ 0 & \text{for } \mu \leq 0 \end{cases}, \tag{28}$$

where $\mu = \cos\theta$, with θ the heliocentric angle (i.e., the angle between the propagation direction and the normal to the solar surface).

We consider an academic problem of a normal Zeeman triplet susceptible to the Hanle and Zeeman effects. Further details on the spectral line are given in Appendix B.1. For the sake of simplicity, we consider a spherically symmetric plasma distribution with the line opacity decreasing and the line Doppler width increasing with the distance from the domain center. This model qualitatively represents cold prominence plasma embedded in a hot surrounding corona. Details of the particular parameterization can be found in Appendix B.2, together with the simple, albeit non-trivial, configuration of the magnetic field vector shown in the left panel of Fig. 7. For additional details on the model, see Appendix B.

Assuming that our observation consists of the four Stokes profiles, sampled in $N_\lambda = 47$ wavelengths, for a $N_{pix} = 64 \times 64$ field of view, we have synthesized the observation with the PORTA code in an atmospheric model with a spatial grid of $N^3 = 64^3$ points. In order to mimic more realistic observations, we have added random Gaussian noise with standard deviation $\sigma = 4 \cdot 10^{-4}$ in units of the disk center intensity. Given the maximum observed intensity, the noise-to-signal ratio is always larger than 10^{-3} in the observations. The synthetic observations are shown in the top panel of Fig. 8.

In Fig. 9 we show the same synthetic data, calculated in the same model atmosphere but without any magnetic field. While the circular polarization (Stokes V) is obviously zero and the intensity is practically unaffected, the linear polarization components (Stokes Q and U) are significantly different from the magnetized case due to the missing Hanle effect. Even though the
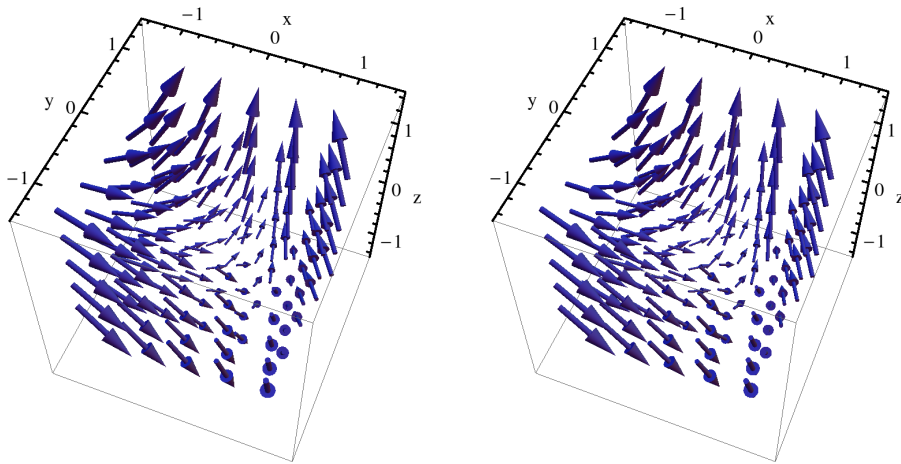
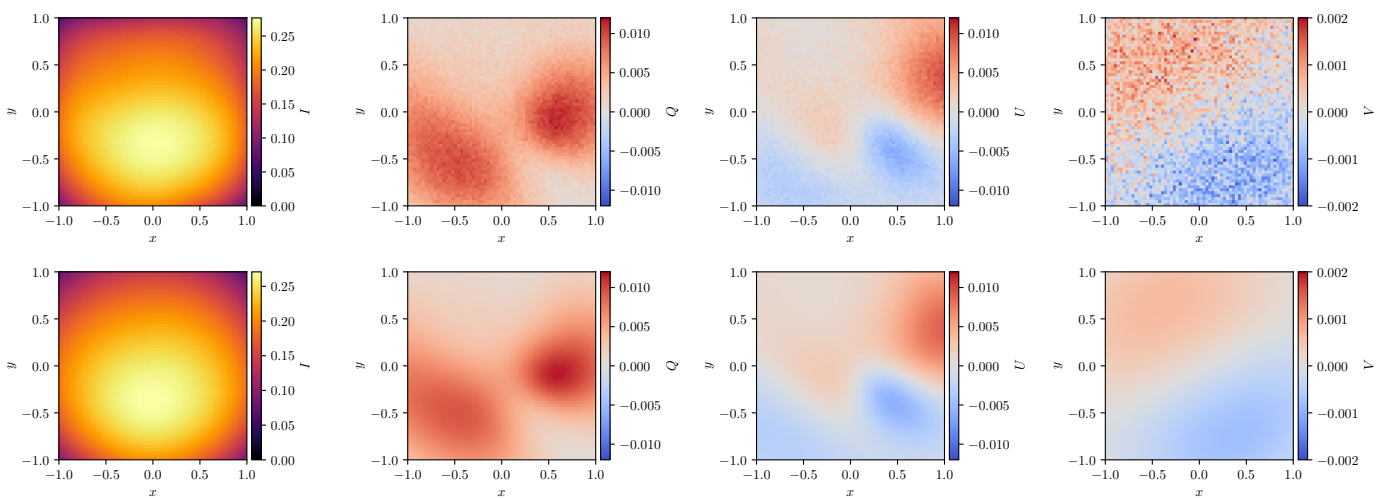**Fig. 7.** Magnetic field vector in the synthetic (left) and inverted (right) models.

Fig. 8. Synthetic observation (top row) and emergent radiation from the inverted model (bottom row). From left to right: I, Q, U in the line center, and V at around the wavelength $\lambda \approx 1.4$. The Stokes parameters are in the disk-center intensity units, $I(\mu = 1)$. The positive Q direction is parallel to the solar limb (parallel to the x axis, cf. Fig. 6). The synthetic signal is contaminated with Gaussian noise with $\sigma = 4 \cdot 10^{-4}$, in the disk-center intensity units.
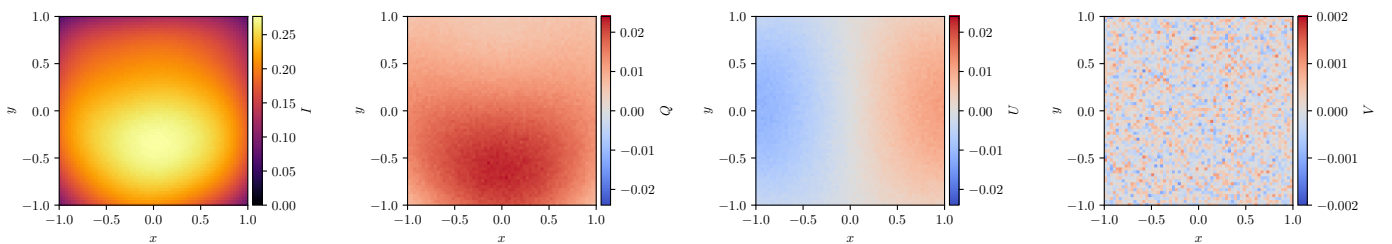
Fig. 9. Same as the top row in Fig. 8, but after switching off the magnetic field. This figure demonstrates that even though the maximum optical thickness of the model is only around $\tau \approx 1$, the symmetry breaking effects due to 3D radiative transfer play a significant role and the Stokes U signal is far from zero. Neglecting 3D radiative transfer could lead to serious errors in the interpretation of the observations.

optical thickness of the plasma structure is of the order of 1 (see Eq. B.10), the symmetry breaking due to the 3D radiative transfer within the medium is sufficiently strong to produce non-zero U signals of the same order as the Stokes Q ones. The inversion of this data using the pixel-by-pixel approach would lead to erroneous conclusions about the presence of magnetic fields.

6.2. Basis functions and the inversion setup

The MHD-like ψ quantities to invert are χ_{L}, $\Delta\lambda_{\mathrm{D}}$, Γ_x, Γ_y, and Γ_z, i.e., the line opacity, the line Doppler width, and the Cartesian magnetic field components defined in Appendix B.1. While in this particular model all the quantities are dimensionless by def-
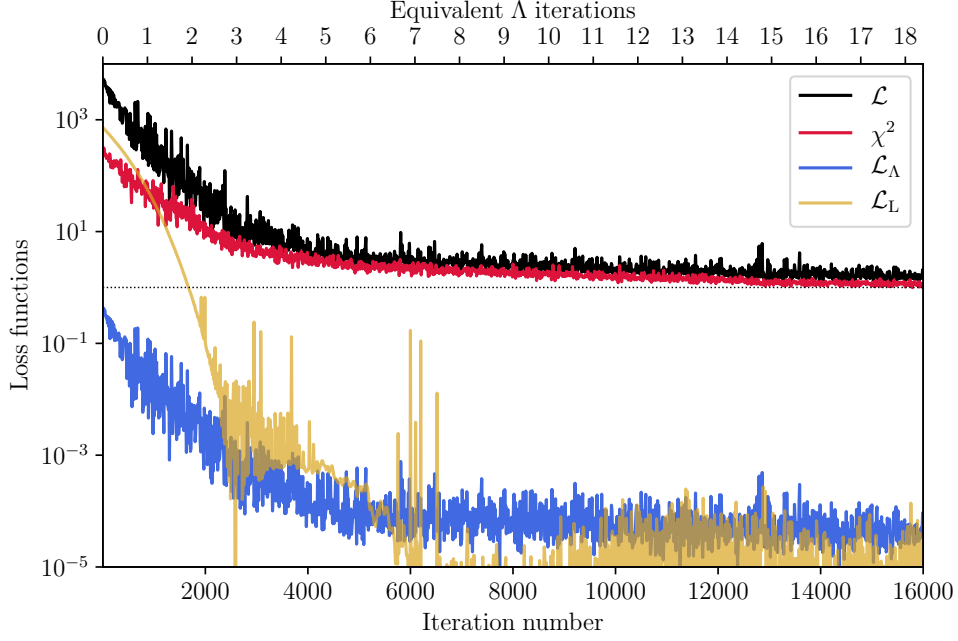
Fig. 10. Convergence of the stochastic meshfree method. The individual curves correspond to the respective terms in Eq. (27) shown in the legend. The bottom horizontal axis shows the number of stochastic iterations while the top horizontal axis shows the CPU time in units of full Λ-iterations in a $N^3 = 64^3$ grid-based model.
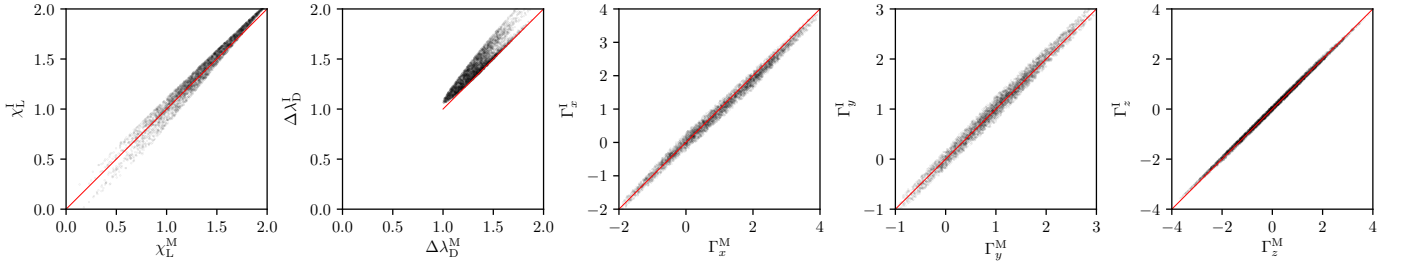

Fig. 11. Scatter plots of the true model (horizontal axes) and inverted (vertical axes) model quantities in 5 000 randomly distributed points within the domain. From left to right: line opacity, line Doppler width, and x-, y-, and z-component of the magnetic field vector.
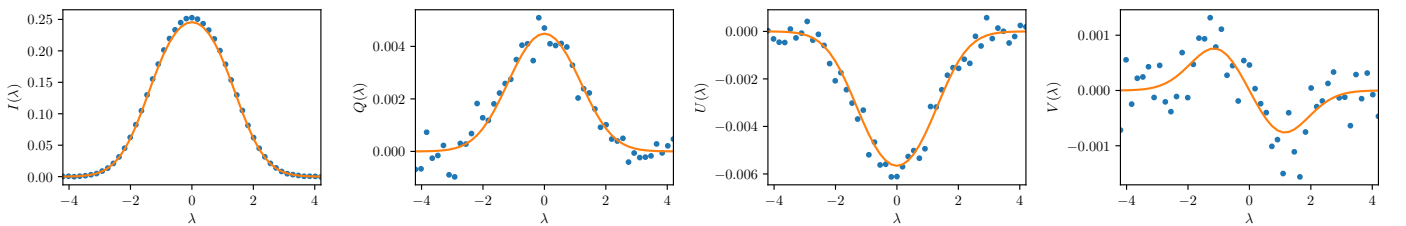

Fig. 12. Example of the inversion fit (orange curve) to the observed (blue dots) Stokes profiles at the spatial point $(x, y) = (0.75, -0.25)$

inition, in realistic applications a suitable normalization should be applied. The model atmosphere is static, hence we are not considering any macroscopic velocity field. Both elastic and inelastic collisions can also be (and are) neglected. In this simple model, the line opacity is equivalent to the density, while the Doppler width is equivalent to the kinetic temperature.

Because we adopted a two-level atom with unpolarized lower state, for this example, the upper-level density matrix $\rho_Q^K(u)$ and the mean radiation field tensors $\bar{J}_Q^K$ are completely equivalent in describing the atomic state. In this example we choose the $\bar{J}_Q^K$ components as our atomic-like ξ variables: $\bar{J}_0^0$, $\bar{J}_0^2$, $\mathfrak{R}\bar{J}_1^2$, $\mathfrak{I}\bar{J}_1^2$, $\mathfrak{R}\bar{J}_2^2$, and $\mathfrak{I}\bar{J}_2^2$, where $\mathfrak{R}$ and $\mathfrak{I}$ stand for the real and imaginary parts, respectively. The components with rank $K = 1$ remain identically zero because of the adopted model. We normalize the atomic like variables with the disk-center intensity and, in addition, we scale the components with $K = 2$ with a factor 20 because we can expect the anisotropy of

the radiation field to be of the order of few percent of the mean intensity.

For the function basis to approximate the spatial distribution of the problem quantities we choose the Chebyshev polynomials of the first kind, using the tetrahedron subset (see the central panel of Fig. A.1). The maximum orders of the basis functions for the MHD-like quantities are $p(\chi_L) = p(\Delta\lambda_D) = 2$ and $p(\Gamma_x) = p(\Gamma_y) = p(\Gamma_z) = 1$, which, according to Eq. (A.2) entails $M_\psi(\chi_L) = M_\psi(\Delta\lambda_D) = 10$ and $M_\psi(\Gamma_x) = M_\psi(\Gamma_y) = M_\psi(\Gamma_z) = 4$ with the total number of MHD-like variables equal to 32. In this example we can afford to determine the basis from the a priori knowledge of the parameterization of the model (cf. Eqs. B.4 to B.8) but, in general, one would have to use adequate techniques to determine them (see Sect. 4.3.2).

We have empirically found via numerical experimentation that the order of the basis of the atomic-like variables sufficient to reach NLTE consistency in our example is $p(\bar{J}_Q^K) = 3$ which, using Eq. (A.2), gives us $M_\xi = 20$ and the total number of atomic-like variables is $N_\xi M_\xi = 5 \cdot 20 = 120$.

Given the relative amplitudes of the Stokes signals in the observations (see top panels in Fig. 8), we set the weights $w_I = 1$, $w_Q = w_U = 20$, and $w_V = 200$ in the χ^2 function in Eq. (9) (these are the values of w_k before their normalization to 1). The weight of the NLTE regularization factor is set to $\lambda = 10^4$ with the aim to reduce the residual NLTE error below 10^{-4}. We set $\gamma = 1$ for the $\mathcal{L}_L$ term, a value sufficient to end up with a locally consistent solution in this example, as we show below.

The model is initialized with random values of the variables and we use $N_\Lambda = 3$ pilot points, $N_\chi = 10$ probing pixels, and $N_L = 10$. The angular quadrature in the pilot points is the 88-point $L = 11$ quadrature of Štěpán et al. (2020). Alternatively, one might prefer even more accurate quadratures by Jaume Bestard et al. (2021b). In this example, we do not consider the randomization of the quadrature orientation described in Sect. 5.3 because the boundary conditions of Eq. (28) are such that the problem does not suffer from disk-edge artifacts.

We have implemented the meshfree stochastic method in a parallel C code and solved the inversion problem using the ADAM algorithm with parameters $\alpha = 0.001$, $\beta_1 = 0.9$, $\beta_2 = 0.999$, and $\epsilon = 10^{-8}$ (see Kingma & Ba 2014, for details).

6.3. Results

In Fig. 10, we show the convergence of the total loss function $\mathcal{L}$, as well as of the penalty functions $\mathcal{L}_\Lambda$ and $\mathcal{L}_L$, and of the χ^2 function. The χ^2 eventually converges almost to 1 and $\mathcal{L}_\Lambda$ and $\mathcal{L}_L$ decrease to the values corresponding to small deviations from the full consistency. It is worth to say that we have not used any quantitative stopping criteria nor did we extensively experimented with the hyperparameters affecting the convergence. The results in Fig. 10 are only to demonstrate the general behavior of the convergence in the numerical experiments we have performed and the problem requires more thorough analysis.

In terms of CPU time, every iteration shown in Fig. 10 requires about 5 seconds using a common contemporary CPU. The total computing time of this inversion with 16 000 iterations is of about 22 hours. In other words, this is equivalent to about 20 seconds per pixel. It is interesting to compare this time with an estimation of the inversion time of the very same model using the standard χ^2 minimization of a grid-based method. In the PORTA code, that could be used as a Λ-iteration engine, the CPU time per mesh point, per wavelength, and per quadrature propagation direction is of about $t = 10^{-6}$ seconds using the same CPU. One

Λ iteration with 4 Stokes parameters therefore requires about $4N^3 N_\lambda N_\Omega t \approx 4\,300$ seconds. Assuming, being very optimistic, that we would only need one Λ iteration to calculate a derivative of χ^2 with respect to each of the ψ variables and if we neglect the computing time required to solve the NLTE problem at the end of every iteration and the time needed to synthesize the emergent radiation in every step of the χ^2 gradient calculation, we would need about $32 \cdot 4\,300 \approx 38$ hours per iteration because the number of the MHD-like variables is 32. Since a realistic number of such iterations is at least $N_{it} = 100$, the inversion time using such method would require about 3 800 hours, i.e., about a factor 170 more than our solution.[7] We note that this significant difference is not just due to the fact that our model is quite smooth, because in both approaches the CPU time scales with the level of smoothness quantified by M_ψ.

The scatter plots in Fig. 11 show the correlation of the synthetic and the inverted model quantities at the same spatial points. The diagonal red lines in every panel show the span of the correct values and, in the case of a perfect inversion, all the black points should lay on them. Even though the match is not perfect due to the limited accuracy of the basis, the presence of noise, and due to the slightly premature stopping of the iterative process, the agreement seems to be quite satisfying for all five quantities in the whole 3D domain. In the right panel of Fig. 7 we visualize the inferred magnetic field vector, which is indistinguishable at plot level from the one in the original model (left panel of the same figure).

The bottom row in Fig. 8 shows the emergent radiation corresponding to the inferred model, to be compared with the top row in the same figure, we can see that the agreement is remarkable. An example of the good quality of the fit to the observed Stokes profiles at a particular pixel in the field of view is shown in Fig. 12.

7. Discussion and conclusions

3DNIP is generally considered to be one of the most challenging unsolved problems in the theory of radiation transfer. In this paper we have presented a first attempt to solve it. Our approach is not to generalize, in a brute-force manner, the standard 1D methods into the 3D geometry because such an approach is doomed to failure for a number of reasons.

In inversion methods based on the pixel-by-pixel approach, the NLTE radiative coupling of different regions (in the direction perpendicular to the LOS) of the plasma is usually considered to be a complication. We have developed a meshfree approach that takes this coupling as an advantageous natural regularization which leads to more robust, more physically correct, and computationally much faster solutions.

The new method employs the idea that 3DNIP can be solved as an unconstrained minimization problem in which unphysical solutions are allowed during the iterative process. We show that the method has the potential to be much faster than methods based on 3D grids with the recurring evaluations of the Λ operator. This approach promises to provide fast solutions, especially in case of relatively smooth models, but it can also provide at least approximate solutions in complex models that would be completely unsolvable using grid-based techniques.

---

[7] However, in practice, the number of Λ operator evaluations would be larger than one per gradient component in the grid-based methods, hence the speedup of the solution in the meshfree method would be $\gtrsim 340$.

We can summarize the main advantages of the proposed framework as follows:

1. Consistent 3D NLTE solution with scattering polarization, Hanle, and Zeeman effects fully taken into account.
2. Additional conditions for physical consistency, such as $\nabla \cdot \boldsymbol{B} = 0$ or the equation of continuity, can be naturally incorporated as penalty terms to the loss function.
3. Since all the pixels are inverted together, the solution is more robust and less sensitive to noise in the data than the pixel-by-pixel methods.
4. Stochastic angular quadratures make it possible to avoid unphysical discontinuities in the boundary illumination in the case of the prominence/filament geometry.
5. Long characteristics lead to more accurate calculation of the radiation field in the pilot points than short characteristics in the grid-based models.
6. There is no numerical error due to the finite spacing of the 3D mesh.
7. The method based on the modern algorithms of stochastic gradient descent is less prone to end up in a local minimum of the loss function.
8. The method can be trivially parallelized for massive HPC facilities. On the other hand, even a single desktop computer can be used to infer at least some information about the global structuring of the plasma properties — something that would not be possible with grid-based methods.
9. Much smaller memory requirements than in the grid-based methods. There is no need for domain decomposition in order to reduce the memory demands.
10. Given the lack of a grid, implementation of the code is easier than in the grid-based methods.

We may list the following disadvantages of the method:

1. The method is less suitable for problems with a large number of spectral lines or atomic levels because the number of variables to be inferred could grow and slow down the calculation (see Eq. 19).
2. In case of abrupt changes of physical parameters, the parameterized model can have difficulties to accurately describe such discontinuities. This problem is inherent not only to the presented method, but to every method using an expansion of the physical parameters into a predefined basis.
3. Given the representation of the atomic-like quantities in terms of an expansion in a finite set of basis functions, there may be certain residual NLTE inconsistency in the solution. However, this inconsistency can be reduced to any desired precision level.

We intentionally do not discuss a number of important issues related to the 3D inversion problem. Among these issues are the problems of possible ambiguities, local minima, location of the observed structures along the LOS, the convergence criteria, various possibilities for adaptivity of the inversion algorithm, and many others. Last but not least, applications with realistic spectral lines need to be studied.

References

Abramowitz, M. & Stegun, I. A. 2014, Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables (Martino Fine Books)
Asensio Ramos, A. & de la Cruz Rodríguez, J. 2015, A&A, 577, A140
Asensio Ramos, A. & Díaz Baso, C. J. 2019, A&A, 626, A102
Asensio Ramos, A., Trujillo Bueno, J., & Landi Degl'Innocenti, E. 2008, ApJ, 683, 542
Auer, L. 2003, in Astronomical Society of the Pacific Conference Series, Vol. 288, Stellar Atmosphere Modeling, ed. I. Hubeny, D. Mihalas, & K. Werner, 405
Auer, L., Bendicho, P. F., & Trujillo Bueno, J. 1994, A&A, 292, 599
Benedusi, P., Janett, G., Belluzzi, L., & Krause, R. 2021, A&A, 655, A88
Casini, R., López Ariste, A., Tomczyk, S., & Lites, B. W. 2003, ApJ, 598, L67
de la Cruz Rodríguez, J., Leenaarts, J., Danilovic, S., & Uitenbroek, H. 2019, A&A, 623, A74
de la Cruz Rodríguez, J. & van Noort, M. 2017, Space Sci. Rev., 210, 109
de Vicente, A., del Pino Alemán, T., & Trujillo Bueno, J. 2021, ApJ, 912, 63
del Pino Alemán, T., Trujillo Bueno, J., Štěpán, J., & Shchukina, N. 2018, ApJ, 863, 164
del Toro Iniesta, J. C. & Ruiz Cobo, B. 2016, Living Reviews in Solar Physics, 13, 4
Doherty, J., Fienen, M., & Hunt, R. 2010, Approaches to highly parameterized inversion: Pilot-point theory, guidelines, and research directions: U.S. Geological Survey Scientific Investigations Report 2010-5168
Fabiani Bendicho, P., Trujillo Bueno, J., & Auer, L. 1997, A&A, 324, 161
Gudiksen, B. V., Carlsson, M., Hansteen, V. H., et al. 2011, A&A, 531, A154
Hackbusch, W. 1985, Multi-Grid Methods and Applications (Springer-Verlag, Berlin)
Hubený, I. & Mihalas, D. 2014, Theory of Stellar Atmospheres
Janett, G., Benedusi, P., Belluzzi, L., & Krause, R. 2021, A&A, 655, A87
Jaume Bestard, J., Trujillo Bueno, J., Štěpán, J., & del Pino Alemán, T. 2021a, ApJ, 909, 183
Jaume Bestard, J., Štěpán, J., & Trujillo Bueno, J. 2021b, A&A, 645, A101
Jurčák, J., Collados, M., Leenaarts, J., van Noort, M., & Schlichenmaier, R. 2019, Advances in Space Research, 63, 1389
Kingma, D. P. & Ba, J. 2014, arXiv e-prints, arXiv:1412.6980
Kunasz, P. & Auer, L. H. 1988, J. Quant. Spectr. Rad. Transf., 39, 67
Lagg, A., Ishikawa, R., Merenda, L., et al. 2009, in Astronomical Society of the Pacific Conference Series, Vol. 415, The Second Hinode Science Meeting: Beyond Discovery-Toward Understanding, ed. B. Lites, M. Cheung, T. Magara, J. Mariska, & K. Reeves, 327
Landi Degl'Innocenti, E. & Landolfi, M. 2004, Polarization in Spectral Lines, Vol. 307
Leenaarts, J. & Carlsson, M. 2009, in Astronomical Society of the Pacific Conference Series, Vol. 415, The Second Hinode Science Meeting: Beyond Discovery-Toward Understanding, ed. B. Lites, M. Cheung, T. Magara, J. Mariska, & K. Reeves, 87
López Ariste, A. 2015, Magnetometry of Prominences, ed. J.-C. Vial & O. Engvold, Vol. 415, 179
Rimmele, T. R., Warner, M., Keil, S. L., et al. 2020, Sol. Phys., 295, 172
Ruiz Cobo, B. & del Toro Iniesta, J. C. 1992, ApJ, 398, 375
Rybicki, G. B. & Hummer, D. G. 1991, A&A, 245, 171
Socas-Navarro, H., de la Cruz Rodríguez, J., Asensio Ramos, A., Trujillo Bueno, J., & Ruiz Cobo, B. 2015, A&A, 577, A7
Socas-Navarro, H., Trujillo Bueno, J., & Ruiz Cobo, B. 2000, ApJ, 530, 977
Steiner, O. 1991, A&A, 242, 290
Trujillo Bueno, J. & Fabiani Bendicho, P. 1995, ApJ, 455, 646
van Noort, M. 2012, A&A, 548, A5
Štěpán, J. 2015, in Polarimetry: From the Sun to Stars and Stellar Environments, ed. K. N. Nagendra, S. Bagnulo, R. Centeno, & M. Jesús Martínez González, Vol. 305, 360–367
Štěpán, J., Jaume Bestard, J., & Trujillo Bueno, J. 2020, A&A, 636, A24
Štěpán, J. & Trujillo Bueno, J. 2013, A&A, 557, A143
Štěpán, J. & Trujillo Bueno, J. 2016, ApJ, 826, L10
Štěpán, J., Trujillo Bueno, J., Leenaarts, J., & Carlsson, M. 2015, ApJ, 803, 65

Appendix A: Implementation details

A.1. Expansion in the basis functions

The choice of the orthonormal basis functions $\{\phi_i(\boldsymbol{r})\}_{i=1}^M$, with M standing for either M_ψ or M_ξ, is arbitrary in the sense that any such set with $M = N^3$ is capable of equivalently describe the spatial variation of the physical quantities in a grid-based model with N^3 points. However, we need to choose a subset of the basis functions with $M \ll N^3$ that will allow to accurately describe such spatial variation. Therefore, different sets of functions and coordinate systems will provide different results.

One of possible choices for a 3D basis is to construct it as a product of three one-dimensional basis functions,

$$\phi_i(\boldsymbol{r}) \rightarrow \phi_{klm}(\boldsymbol{r}) = \phi_k(x)\phi_l(y)\phi_m(z) , \qquad (A.1)$$

where the functions ϕ_n are taken from a suitable 1D basis. For instance, one can use $\phi_k(x) = T_k(x)$ with $T_k(x)$ being a Chebyshev polynomial of the first kind and order k (Abramowitz & Stegun 2014). Alternatively, one could use the basis of discrete cosine transform or any other set of orthogonal functions.

It is natural to restrict the orders k, l, and m to be smaller or equal to a certain integer p. This would typically imply a restriction to the $(p + 1)^3$ smoothest functions (see the left panel of Fig. A.1 for $p = 5$). Since the computing time is always a concern, we can even use a smaller subset of this basis. In our numerical experiments, we have tested the shapes of the bases shown in the middle and right panels of Fig. A.1. Apart from the $(p + 1)^3$ basis in the left panel, two alternative bases with the same order of functions but smaller M are the bases satisfying the conditions $k + l + m \leq p$ with

$$M = (p + 1)(p + 2)(p + 3)/6 \qquad (A.2)$$

(middle panel of Fig. A.1) and $k^2 + l^2 + m^2 \leq p^2$ with

$$M \approx \pi(p + 1)^3/6 \qquad (A.3)$$

(right panel in Fig. A.1). Using such subsets can lead to a significant CPU time saving while the approximation capabilities may only be slightly deteriorated. The choice of the subset is indeed problem dependent and we will discuss it in future publications. At this point, we only note that it is not necessary to use the same subset of basis functions for every axis. For instance, reducing the $\hat{z}$ basis to $\phi_0(z) = 1$ makes the problem effectively 2.5D with physical quantities being constant along the line of sight (given the geometry in Fig. 6).

Apart from the possibility to construct the 3D basis out of the one-dimensional functions in the Cartesian coordinates, it might be useful to use orthogonal functions in different coordinate systems, such as in spherical or cylindrical coordinates.

A.2. Normalization of the quantities

In the calculation of the χ^2 function, one usually uses different weights $\{w_k\}_{k=0}^3$ for different Stokes parameters (see Eq. 9). This is due to the fact that the Stokes parameters are usually of different orders of magnitude. If the weights of the Stokes parameters were all the same, the inversion algorithm would try to fit the intensity profile while the sensitivity to polarization components would be very small.

A similar situation can occur with the θ parameters of the problem variables. For instance, the atomic level population ρ_0^0 is usually much larger than the level alignment ρ_0^2. Since we want

the loss function of Eq. (22) to be sensitive to the errors of variables with different order of magnitude, it is desirable to choose the θ_i variables in such a way they have similar orders of magnitude. This can be achieved by using a proper normalization. The same applies to all the variables in the θ vector that should be suitably normalized in order to have similar orders of magnitude for their "typical" physical values.

A.3. Local consistency regularization

The third term on the right-hand side of Eqs. (23) and (27) stands for a penalty for deviation of the model parameters from physical consistencies other than the NLTE consistency. These include penalties for negative atomic level population, unphysical degree of atomic polarization, violation of the magnetic field zero divergence condition ($\nabla \cdot \boldsymbol{B} = 0$), violation of the equation of continuity, and any other condition depending on the particular problem. Evaluation of $\mathcal{L}_L$ does not involve radiative transfer calculations but rather checking the values and possibly derivatives of quantities in N_L points. This calculation is therefore very fast. The $\mathcal{L}_L$ term is important as an additional physical regularization of the problem that allows us to incorporate relatively complex physical conditions into a small piece of a numerical code.

The loss function of the local physical consistency, $\mathcal{L}_L$, can be written as an average over N_L random points in the following form:

$$\mathcal{L}_L = \frac{1}{N_L} \sum_{i=1}^{N_L} \sum_\alpha g_\alpha(\boldsymbol{r}_i; \theta) , \qquad (A.4)$$

where θ stands for the vector of model variables of Eq. (18) and the summation over α stands for different problem-dependent penalties g_α. These non-negative differentiable penalty functions should take zero value for the physically consistent solutions at any point $\boldsymbol{r}_i$.

This very general formulation can be easily understood considering particular examples. If we use x to represent a particular positive quantity that is constructed out of θ (such as atomic level population, plasma temperature, or any function of the θ variables), it is useful to define a penalty of the form

$$g_\alpha(x) \equiv g_+(x; t) = \begin{cases} 0 & \text{for } x > 0 \\ (x/t)^2 & \text{for } x \leq 0 \end{cases} , \qquad (A.5)$$

where the scaling parameter t is to be chosen depending on the variable and it controls the sensitivity of the loss function to this particular quantity – the smaller is t the stronger are the penalties for violation of the condition. In order to avoid convoluted notation, we do not explicitly indicate the point of g_+ evaluation, $\boldsymbol{r}_i$; the quantity x in g_α is always evaluated in all the N_L points and the resulting penalties are summed according to Eq. (A.4). If x is a quantity that must always be larger than some other quantity or a constant, y, we can use a penalty of the form

$$g_\alpha(x) \equiv g_>(x, y; t) = \begin{cases} 0 & \text{for } x > y \\ \left(\dfrac{x - y}{t}\right)^2 & \text{for } x \leq y \end{cases} . \qquad (A.6)$$

Analogously, if a quantity needs to be zero for physical reasons, the function

$$g_\alpha(x) \equiv g_0(x; t) = (x/t)^2 \qquad (A.7)$$

provides a way to penalize deviations from the zero value. An example of such quantity is the magnetic field divergence, $x = \nabla \cdot$
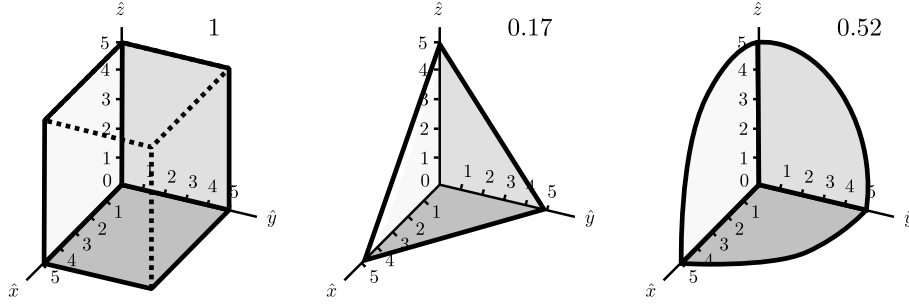
Fig. A.1. Schematic representation of three bases of the same order $p = 5$ for approximating the spatial variation of the problem variables in the computational domain. The numbers in the top right part of each diagram indicate the fraction of the number of the considered basis functions with respect to the full basis in the left diagram.

$\boldsymbol{B} = \partial_x B_x + \partial_y B_y + \partial_z B_z$. Additional local physical constrains that can be easily incorporated into $\mathcal{L}_L$ are, for instance, the equation of continuity in the stationary models, $x = \nabla \cdot (\rho \boldsymbol{v})$ with the g_0 penalty, the condition of magneto-hydrostatic equilibrium, etc.[8]

Appendix B: Example model details

This appendix describes some details of the academic model discussed in Sect. 6.

B.1. Model atom and the Hanle and Zeeman effects

Our example model atom is a normal Zeeman triplet with angular momenta $J_l = 0$ and $J_u = 1$ for the lower and upper levels, respectively. Given that the collisions in our model are considered negligible, the atomic density matrix of the upper level at a given spatial point, $\rho_Q^K(u)$, is fully determined by the local mean radiation field tensor, $\bar{J}_Q^K$, and the magnetic field $\boldsymbol{B}$. Instead of $\boldsymbol{B}$ in the units of G, we use the dimensionless vector

$$\boldsymbol{\Gamma} = 8.79 \cdot 10^6 g_u \frac{\boldsymbol{B}}{A_{ul}} , \tag{B.1}$$

where g_u is the upper-level's Landé factor and A_{ul} is the Einstein coefficient of spontaneous emission in s^{-1} (see Landi Degl'Innocenti & Landolfi 2004, for details). It follows that $\Gamma = |\boldsymbol{\Gamma}|$ is the factor commonly used to quantify the Hanle effect (see, e.g., the discussion after Eq. 5 of del Pino Alemán et al. 2018, where Γ_u is used instead of Γ). For $\Gamma = 1$, the spectral line is most sensitive to the Hanle effect, and we call the corresponding magnetic field the critical Hanle field.

The irreducible components of the line source function are given by Eq. (5) of del Pino Alemán et al. (2018) where $w_{J_u J_l}^{(2)} = w_{10}^{(2)} = 1$, $\epsilon = 0$, and $\delta = 0$. The matrix elements M_{ij} in that equation are defined in terms of the orientation of the magnetic field vector with respect with the local vertical and can be found in the Appendix therein. The source functions of the Stokes parameters then follow from Eq. (6) of the same paper.

The line absorption profiles are Gaussian with the Doppler width $\Delta\lambda_D$:

$$\varphi(\lambda) = \frac{1}{\Delta\lambda_D \sqrt{\pi}} \exp\left(-\frac{\lambda^2}{\Delta\lambda_D^2}\right) \tag{B.2}$$

---

[8] We use the same symbol t in Eqs. (A.5)–(A.7) in order to avoid cluttered notation but its value indeed depends on the context and the order of magnitude of x.

where the wavelength λ is defined as the distance to the line center for convenience.

In this example we also account for the longitudinal Zeeman effect assuming that the magnetic field is sufficiently weak so that the Zeeman splitting is much smaller than the line Doppler width and thus only the Stokes V signal is affected by the Zeeman effect. The radiative transfer equation for circular polarisation in the weak field limit is then (Landi Degl'Innocenti & Landolfi 2004)

$$\frac{dV(\lambda)}{ds} = -\chi_L \varphi(\lambda) V(\lambda)$$
$$-\Gamma \frac{\alpha}{\Delta\lambda_D} \cos\theta_\Gamma \frac{\lambda}{\Delta\lambda_D} \varphi(\lambda) \chi_L \left[I(\lambda) - S_L\right] , \tag{B.3}$$

where $\varphi(\lambda)$ is the local absorption profile, whose Gaussian form has been taking into account to derive this formal expression, $\Delta\lambda_D$ is the Doppler width, χ_L is the line opacity, S_L is the intensity line source function, and θ_Γ is the angle between the propagation direction and the magnetic field vector. The value of α is proportional to the effective Landé factor and we set it to $\alpha = 0.004$ for our academic case. If $\alpha \ll 1$, the weak field limit is satisfied because we typically have $\Gamma \sim 1$ and $\Delta\lambda_D \sim 1$.

B.2. Spatial variation of the physical quantities

The plasma in our synthetic example model is a cloud with a spherically symmetric distribution of line opacity and Doppler width. For the sake of demonstrating the method, we define a simple functional form of the relevant quantities:

$$\chi_L(\boldsymbol{r}) = 2(1 - r^2) , \tag{B.4}$$
$$\Delta\lambda_D(\boldsymbol{r}) = 1 + r^2 , \tag{B.5}$$
$$\Gamma_x = 1 - 2x - y , \tag{B.6}$$
$$\Gamma_y = 1 + x + y , \tag{B.7}$$
$$\Gamma_z = -x + 2y + z , \tag{B.8}$$

where $r = \sqrt{x^2 + y^2 + z^2}$ is the radial distance from the center of the $[-1, 1]^3$ model. The above magnetic field configuration is shown in Fig. 7.

It follows from Eqs. (B.4) and (B.5) that the maximum total optical thickness of the medium is in the domain center ($x = y = 0$) and it is equal to

$$\tau_{\max} = \int_{-1}^{1} dz\, \chi_L(z) \varphi(\lambda = 0, z) \tag{B.9}$$

$$= \frac{2}{\sqrt{\pi}} \int_{-1}^{1} dz\, \frac{1 - z^2}{1 + z^2} = \frac{2}{\sqrt{\pi}}(\pi - 2) \approx 1.3 . \tag{B.10}$$