Part II : Radiative transfer and Spectropolarimetry

Model fitting and spectropolarimetric inversions

Ivan Milić (CU/LASP/NSO)

March 5, 2019

Quick Summary

- So far we have studied the line formation mechanisms.
- Polarized emerged intensity is determined by the model.
- If we know the atmospheric model and the atomic model, and the relevant physics (e.g. LTE vs NLTE, etc.), we can calculate the emergent spectrum.
- How to do the opposite? **That is, how to infer a model atmosphere from the observed polarized spectrum?**

Let's remember the model parameters

- Temperature: Saha-Boltzman, line broadening, collisions.
- Gas pressure: Total number of particles, collisions.
- Line of sight velocity: Spectral line shifts, asymmetry.
- Microturbulent velocity: Line broadening.
- Magnetic field: line splitting (broadening), line polarization.

So what is our model?

Remember that the model atmosphere is basically a big table (matrix) with the values of all the relevant physical parameters.

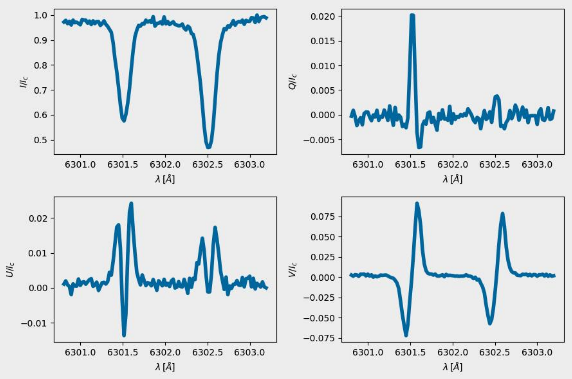
So: $\boldsymbol{T} = (T_0, T_1 .... T_{ND})$, $\boldsymbol{v}_{\mathbf{los}} = (v_0, v_1 .... v_{ND})$, etc...

Let's write it concisely as $\boldsymbol{M} = (T_0, T_1 ... T_{ND}, v_0, v_1 .... v_{ND}, B_0, B_1 ...)$

Length of $\boldsymbol{p}$ is $ND \times NP = NM$ (M for model).

In practice $NP = 7$, $ND = \mathcal{O}(10^2)$.

And what is our data?



I,Q,U,V of the selected piece of the spectrum. For spectrograph: few hundreds of wavelengths, for filtergraph, few tens. So, again $N = \mathcal{O}(10^2 - 10^3)$.

Generative model

To perform any sort of meaningful inference, we need to choose a *generative model*. That includes:

- Set of model parameters that describes the problem we are studying (e.g. if we only have spectroscopic observations probably we are not estimating $\boldsymbol{B}$).
- Physics that connects the model parameters to the observables. (e.g. we might choose LTE approximation + RTE, or we can go to NLTE, or simplify to Milne-Eddington approximation or even fixed slab model).
- Noise model. This is often overlooked and could lead to severe mistakes. We usually assume Gaussian distribution.

A few words about the noise

- We assume Gaussian noise, what does that mean?
- It means that if the true (uknown) value is I_0, we will measure $I = I_0 + \epsilon$, where:

$$p(\epsilon) = \frac{1}{\sqrt{\pi}\sigma} e^{-\epsilon^2/\sigma^2}.$$

A very simple generative model - pendulum

If we make a number of (reasonable) assumptions, we get the following equation describing the behaviour of the pendulum:

$$T = 2\pi\sqrt{\frac{l}{g}}.$$

Where T is the period of the pendulum, l is the length, and g the constant of gravitational acceleration. We get:

$$T^2 = 4\pi^2\frac{l}{g} + \epsilon_{T^2}.$$

Or simplified:

$$y = kx + m + \epsilon(\sigma).$$

Our goal

We want to determine g, that is k. We measured T^2 and l, that is x and y, and we assume there are no errors in x and that we know errors in y.

What now, what do we want?

We want **the most probable** values of k and m given the measured values of x and y.

Let's focus on one measurement:

$$y_i = kx_i + m + \epsilon(\sigma)$$

For given values of k and m, the probability of getting value y_i is:

$$\frac{1}{\sqrt{\pi}\sigma} e^{-(y_i - kx_i - m)^2/\sigma_i^2}$$

or:

$$p_i \propto e^{-(y_i - f(x_i))^2/\sigma_i^2}$$

Now, the probability of getting the whole set of the observations is:

$$p = \prod_i p_i \propto \exp\left(-\sum_i (y_i - f(x_i))^2/\sigma_i^2\right)$$

Now, you would be tempted to maximize this probability...

But why?

What you would be maximizing here would be the $p(D|M)$. Probability of the (observed) data, given the model (to be inferred).

We are not interested in that. We want to find the most probable model for the data we have observed. That is:

$$p(M|D)_{\mathrm{max}}.$$

How to write that if we only know $p(D|M)$? (In our language $p(\boldsymbol{y}|k, m)$, or even $p(\boldsymbol{I_\lambda}|\boldsymbol{M})$)

We need to know more!

Bayesian probability (sorry I can't not tell you about this)

$$P(M \wedge D) = P(M|D) \times P(D) = P(D|M) \times P(M).$$

Do you all agree?

$$P(M|D) = \frac{P(D|M)P(M)}{P(D)}$$

Dictionary:

- $P(D|M)$ - likelihood
- $P(M)$ - prior probability
- $P(D)$ - evidence
- $P(M|D)$ - posterior probability. What we need!

What are we optimizing?

Finally, if we assume all the parameter values are *a priori* equally probable, then maximizing posterior $P(M|D)$ is the same as maximizing the likelihood $P(D|M)$.

$P(D|M) \propto \exp\left(-\sum_i (y_i - f(x_i))^2/\sigma_i^2\right)$ and is maximum for...

$$\chi^2(M) = \sum_i \frac{(y_i - f(x_i, M))^2}{\sigma_i^2} = \min$$

So, we minimize the $\chi^2(\boldsymbol{M})$ function. It is a $NM-$dimensional function where NM is the total number of parameters. This is an optimization problem.

Problem set-up

On one side, we have a model described by the tabulated values of parametes of relevant physical values:
$\boldsymbol{M} = (T_0, T_1 ... T_{ND}, v_0, v_1 .... v_{ND}, B_0, B_1 ...)$

On the other side, we have observed values of the stokes parameters:
$\boldsymbol{I}_\lambda = (I_0, I_1 ... Q_0, Q_1 ...)$

We make various assumptions using physical arguments and common sense to formulate the forward problem:

$$\boldsymbol{I}^{\mathrm{synth}} = f(\boldsymbol{p}).$$

Physical arguments and the common sense?

Obviously, line formation is a complicated problem. The model is actually solution of a differential equation (RTE). There are a number decisions to be made:

- LTE vs NLTE?
- What atomic model, what atomic constants, what models for various processes (collisions etc). If NLTE, how many atomic levels, etc?
- If NLTE: PRD or CRD? (more about this on Thursday) What effects to include in the polarization?
- Can we simplyfy the atmosphere somehow?

For example

Atmosphere model: temperature, density, velocity, magnetic field

↓

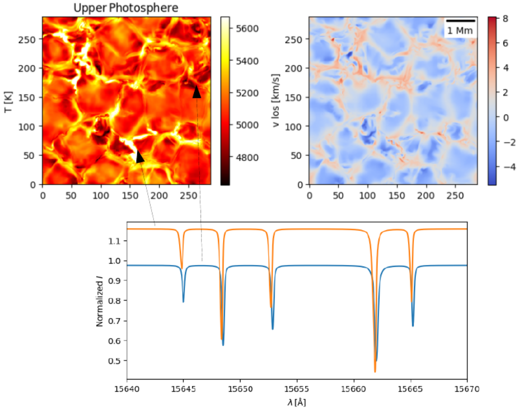
Use Saha-Boltzmann to calculate all relevant level populations

↓

Calculate opacity/emissivity everywhere in the atmosphere

↓

Solve Radiative Transfer Equation → Get the spectrum



MURAM quiet Sun simulation, courtesy of T. Riethmüller

Simplified solutions

Weak field approximation. Does not care how Stokes I is formed, only about the relationship between I and V:

$$V = -4.67 \times 10^{-13} \frac{dI}{d\lambda} \lambda_0^2 g_L B$$

Assumes:

- Magnetic splitting weaker than Doppler splitting $4.67 \times 10^{-13} \lambda_0^2 g_L B < \Delta\lambda_D$
- V formation influenced exclusively by η_I and η_V.
- Magnetic field constant with height.

Still, extremely fast and can be exploited by using lines with significantly different formation heights.

Simplified solutions

Milne-Eddington atmosphere. Assumes the source function increases linearly with depth. We describe formation of one line with 9 parameters only. Assumes:

- Magnetic field, line broadening, shift and damping are all constant with height.
- The source function linear in height.
- DOES NOT assume weak field.
- CAN model I,Q,U,V
- DOES NOT include any “real” physical parameters except magnetic field and los velocity (no temperature, density, etc.)

Analytical solution also super fast to calculate (and fit).

If we want more realism

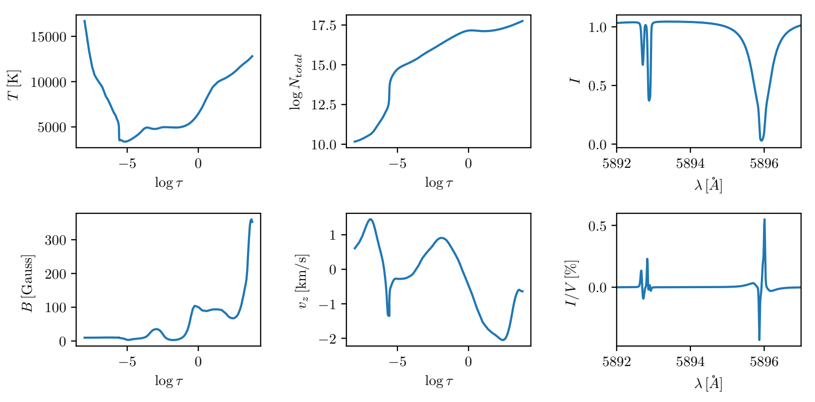


Figure 1: Temperature, particle density, magnetic field and velocity (left,middle) and the resulting spectra and the circular polarization (right)

How to fit?

Fitting is a minimization problem. If you tackle data, I **strongly** recommend you look into numerical recipes book by Press et al. (They are astrophysicists, btw :-))

- Grid (library) search
- Neldear-Mead downhill simplex
- Genetic algorithms
- Markov Chain Monte Carlo (MCMC) and similar (e.g. Nested Sampling) approaches (akka sampling approaches).
- Derivative based (gradient descent, Levenberg-Marquard etc.)

Levenberg-Marquard

This is a stabilized variant of Newton-Rhapson iteration. This method assumes that we are close to the final solution and that the following expression is valid (you can understand this intutively, but the real derivation is somewhat different):

$$\sum_j \frac{\partial f(x_i, \boldsymbol{p})}{\partial p_j} \delta p_j = (y_i - f(x_i, \boldsymbol{p}))$$

For us it would be:

$$\sum_j^{NM} \frac{\partial I_{s,i}(\boldsymbol{p})}{\partial p_j} = (I_{s,i}^{\mathrm{obs}} - I_{s,i}^{\mathrm{synth}}) = \delta I_{s,i}$$

We will call the partial derivative of the specific polarized intensity with respect to a physical quantity at given depth **response function.**

Make sure to read a different approach in the Jose Carlos book!

Levenberg-Marquard

$$\sum_{j}^{NM} \frac{\partial I_{s,i}(\boldsymbol{p})}{\partial p_j} = J_{j,i}\delta p_j = \delta I_i$$

(I absorbed s into i to be more concise, as we did few slides back).

Scheme:

- Assume a model atmosphere, i.e. vector $\boldsymbol{p}_0$.
- Calculate $\boldsymbol{I}(\boldsymbol{p})$ and evaluate $\hat{J}$ and $\delta\boldsymbol{I}$.
- Calculate correction to the parameter vector as by multiplying both sides with $\hat{J}^T$:
$$\hat{J}^T \hat{J}\delta\boldsymbol{p} = \hat{J}^T \delta\boldsymbol{I}$$
- Correct the model and re-start the loop. As soon as $\boldsymbol{I}^{\mathrm{calc}}$ is good enough, stop.

Levenberg-Marquard

What we described is basically Newthon-Rhapson iteration. This method converges very quickly once we are close to the final solution. But if we are not, it can behave very poorly.

Levenberg and Marquard stabilized the method in the following way:

$$\hat{A} = \hat{J}^T \hat{J} + \lambda \mathrm{diag} \hat{J}^T \hat{J}.$$

If λ is small, we have Newton Rapson.

If λ is big diagonal dominates, and we basically revert to the gradient descent.

Gradient descent is slow, but behaves well even far away from the solution. By dynamically changing λ we can tune between the two extremes and achieve optimal convergence.

Levenberg-Marquard inversion

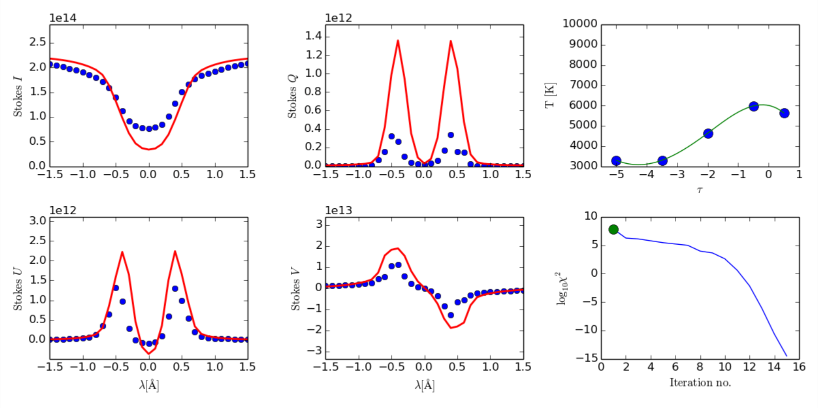


Figure 2: An example of a fitting procedure. The atmosphere is adjusted until the fit between observed (blue) and fitted (red) profiles is achieved.

Two questions remain

1. How do we calculate the derivatives $\frac{\partial I_i(\boldsymbol{p})}{\partial p_j}$? (Response functions).
2. Is it possible to solve for so many unknowns?

Calculating Response Functions

How do we calculate the derivatives $\frac{\partial I_i(\boldsymbol{p})}{\partial p_j}$?

We can try a numerical approach:

$$\frac{\partial I_i(\boldsymbol{p})}{\partial p_j} = \frac{I_i(\boldsymbol{p} + \Delta p_j) - I_i(\boldsymbol{p} - \Delta p_j)}{2\Delta p_j}$$

Here I used Δ instead of δ to emphasize that this now a "perturbation" we are inducing in order to calculate this derivative.

But you can also, painstakingly, step by step, "propagate" the derivatives through all the numerical solvers we are using here, and obtain the same results with much more coding but much less computational time.

Today there are also so called "automatic differentiators" that go through your code and take the derivative of the expressions employing chain rule. Magic stuff.

How do the response functions look like?

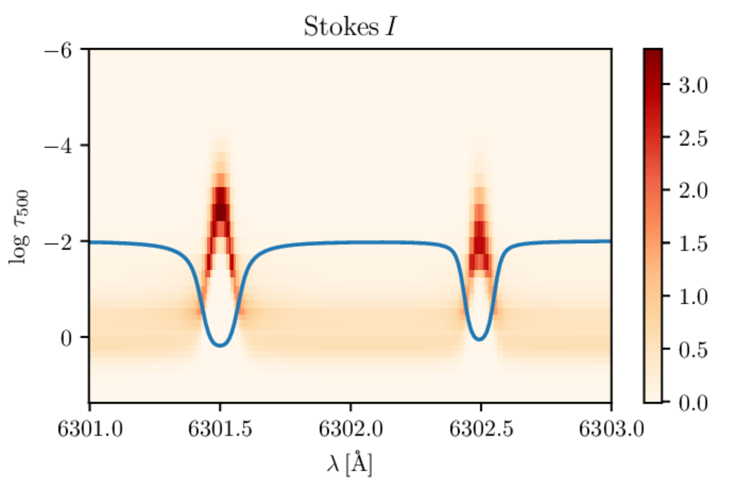


Figure 3: Response function of 6300 Å line pair to temperature.

Other parameters, other Stokes components

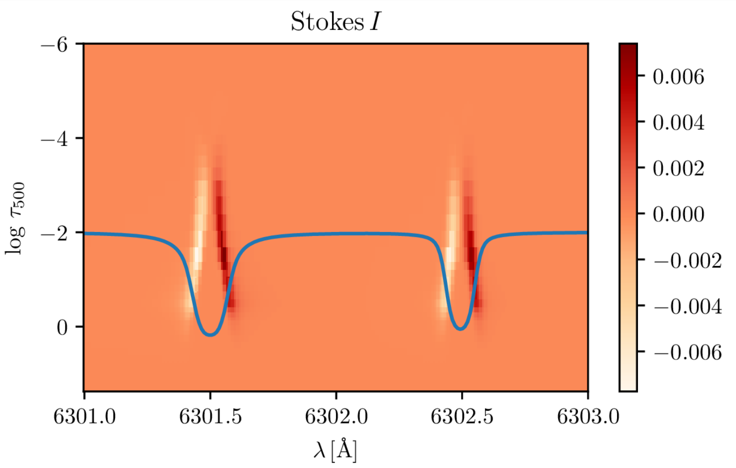


Figure 4: Response function of Stoke V 6300 Å line pair to LOS velocity.

What do RFs tell us?

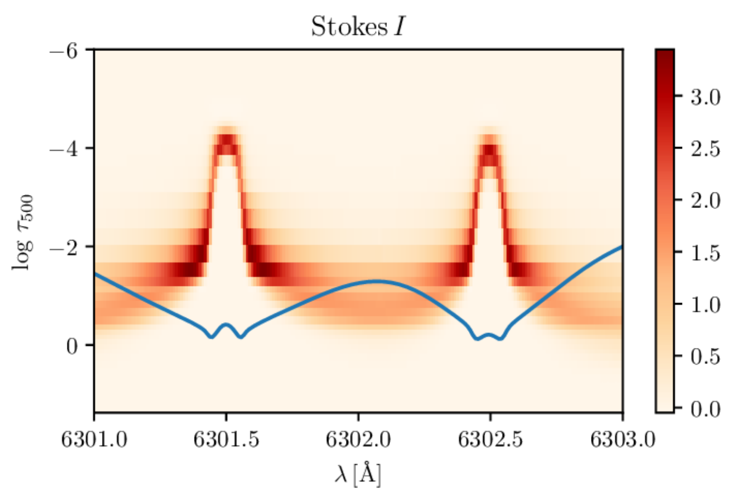


Figure 5: Response function of artificially amplified, by a factor of 100, 6300 Å line pair to temperature.

Response functions for NLTE lines

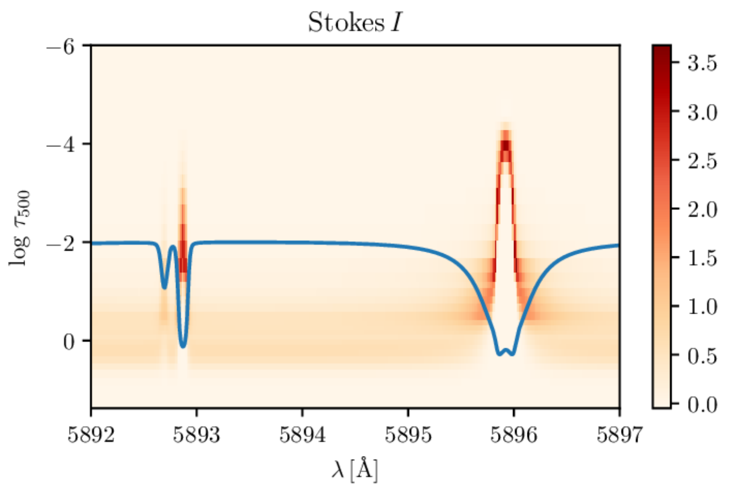


Figure 6: Response function of Sodium D1 line to temperature - LTE case

Response functions for NLTE lines

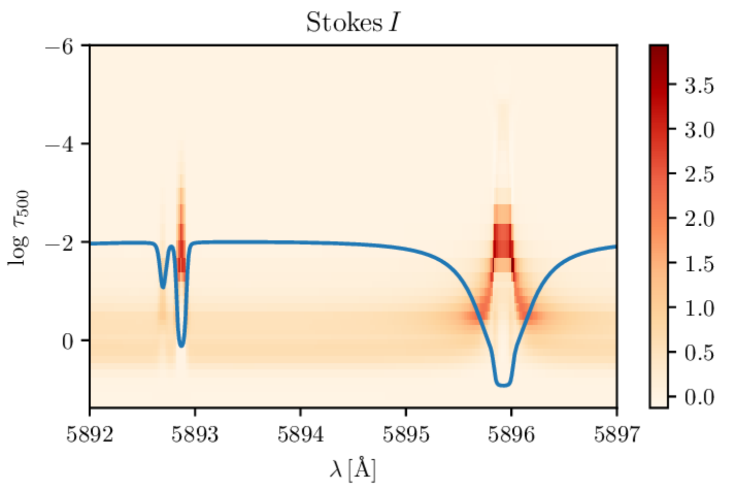


Figure 7: Response function of Sodium D1 line to temperature - NLTE case

So are we done? Can we do everything?

We now know what response function means. We can calculate them. We know how to minimize. Enough?

No.

So are we done? Can we do everything?

We now know what response function means. We can calculate them. We know how to minimize. Enough?

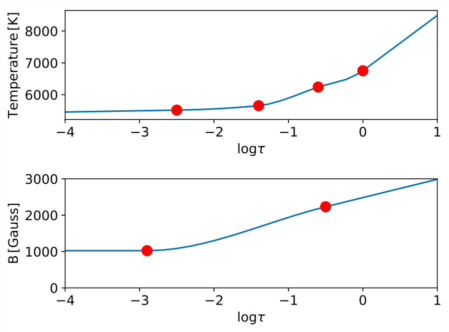
No.

The final problem is that inversion in general and spectropolarimetric inversion in particular is an ill-posed problem. Parameters are degenerate (cross-talk) + the problem is highly non-linear.

We try to simplify things somehow. To do that we use nodes

Instead of representing the full depth dependence of the atmosphere, we restrict it to few points. We map $\boldsymbol{p}$ to $\boldsymbol{\alpha}$, where dimension of $\boldsymbol{\alpha}$ is much smaller. Obviously, we need new response functions:

$$\frac{\partial I_i}{\partial \alpha_j} = \sum_{j'} \frac{\partial I_i}{\partial p_{j'}} \frac{\partial p_{j'}}{\partial \alpha_j}$$



To summarize

Our model: Fixed positions of the nodes. Model parameters are values of the quantities in the nodes. From nodes we can create the whole atmosphere and generate spectrum. That defines the forward problem.

Our data: Observed Stokes spectra + assumed noise model.

What do we do: We maximize $p(M|D)$ (M is model, D is data). That often reduces to maximizing $p(D|M)$ over parameter space, i.e. minimizing χ^2.

We take the values that minimize χ^2 values as the "estimated values", given the assumed model.

Inversion in one figure

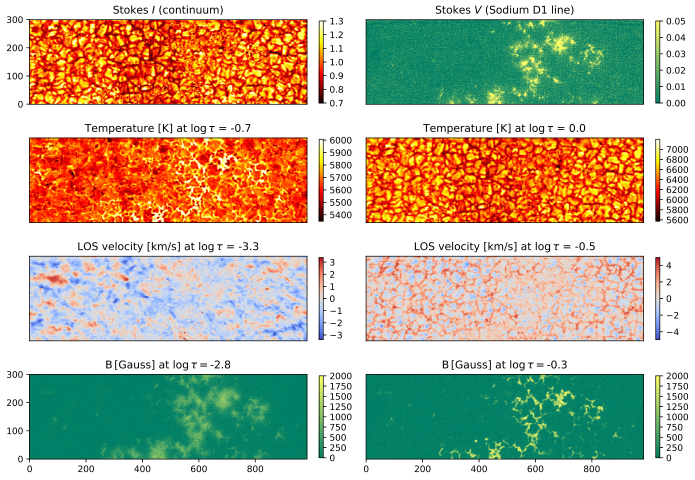


Figure 8: Spatial distributions of the observables (top two panels), and model parameters (bottom six panels)

Why are we doing such a complicated thing?

We are getting physically meaningful, hopefully self-consistent results.

For example, we get depth-dependent magnetic or velocity field, that best explains our observed spectra.

However, this is far from being finished and there is a lot of physics still missing.

Questions? Critics? Complaints? Suggestions?