Denoising Stellar Polarization Signals

Yannis Konidakis
Department of Computer Science
University of Crete
Heraklion, Crete, Greece
ikonidakis@csd.uoc.gr

Aikaterini Papadaki
Department of Computer Science
University of Crete
Heraklion, Crete, Greece
apapadaki@ics.forth.gr

Abstract

In the context of this project, the position of magnetized interstellar dust clouds in a specific Line Of Sight (LOS) is investigated given information of the starlight polarization. In particular, there will be examined a Signal Denoising problem. The aim is to denoise stellar polarization signals and there after to detect the change point into the signal, where a change point corresponds to the position of a magnetized interstellar dust cloud. The dataset consists of mock starlight polarization samples derived from astrophysical simulations and the denoising analysis pipeline is based on Coupled Dictionary Learning algorithm [1]. The corresponding code to the assignment can be found in *https://github.com/ya-joko/Denoising-Stellar-Polarization-Signals*.

1 Motivation

Our Galaxy is permeated by a large-scale magnetic field, which plays an integral role in a wide range of astrophysical processes. A detailed mapping of the Galactic magnetic field would revolutionize our understanding of astrophysical phenomena. However, our diagnostics of magnetic fields are few and hard to obtain. One major diagnostic of the magnetic field properties is *starlight polarization*.

The observed starlight polarization originates in interstellar absorption. In particular, starlight starts out unpolarized, and picks up polarization through partial absorption by a number of magnetized interstellar dust clouds [1], [2]. Measurements in small sky patches are put together forming a beam and it is assumed that in this beam all the stars fell the same interstellar medium, i.e. the same number of dust clouds. In other words, in a specific Line Of Sight (LOS) of observations there is a number of magnetized interstellar dust clouds which determine the starlight polarization values of the group of stars observed. In a different area of the sky, i.e. in a different LOS, that number of dust clouds differs.

In reverse, *observations at a specific LOS of the starlight polarization of a group of stars can reveal the number of magnetized interstellar dust clouds in that LOS*. The aim is to decompose an 1D representation of starlight polarization observations of a group of stars in a area of the sky and obtain a 3D visualization of the interstellar medium in that LOS.

2 Problem Definition

In the context of this project, the interstellar dust clouds separation problem will be investigated. The position of a magnetized interstellar dust cloud in a specific LOS will be examined, given information of the starlight polarization. The aim is to denoise polarization signals assisting the detection of the change point of the signal.

2.1 Polarization Signals Simulations

The constructed dataset consists of mock samples of starlight polarization data derived by astrophysical simulations following the procedure described below.

Each sample is a different LOS. There are 2*n features, the stokes parameters $[q_1, q_2, ..., q_n, u_1, u_2, ..., u_n]$ of all the n stars enclosed in the different groups of stars S_k at that LOS, as illustrated in **Fig. 1**.

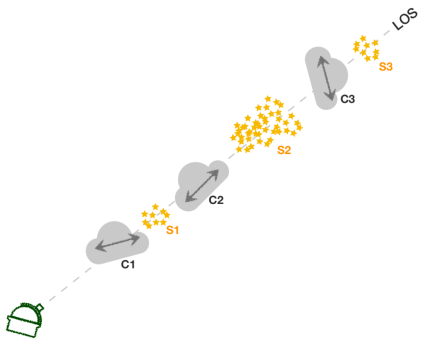


Figure 1: Line Of Sight, where C_k is a magnetized interstellar dust cloud and S_k a group of stars. [Pelgrims, V.]

The Stokes parameters Q and U are related to the The intensity of starlight in different directions. The starlight linear polarization P is characterized by the Stokes parameters Q, U and is given by the following relation **Eq. 1**

$$P = \sqrt{Q^2 + U^2} \tag{1}$$

For a given total intensity I, the degree of polarization is equal to the ratio **Eq. 2**

$$p = \frac{P}{I} = \frac{\sqrt{Q^2 + U^2}}{I} \tag{2}$$

Starlight starts picks up polarization through partial absorption by a number of magnetized interstellar dust clouds N_C. The Stokes Parameters of the linear starlight polarization through partial absorption by a number of magnetized interstellar dust clouds N_C are given by the following formulas **Eq. 3** and **Eq. 4**

$$q_V^{[S_k]} = \sum_{i<=k}^{N_c} q_V^{C_i} \tag{3}$$

$$u_V^{[S_k]} = \sum_{i<=k}^{N_c} u_V^{C_i} \tag{4}$$

For one homogeneous cloud of constant extinction E_{B-V} permeated by a uniform magnetic field **B** with inclination γ_B on the plane of the sky and with position angle ψ_B **Fig. 2**, the vector of reduced Stokes parameters of the linear polarization s that would be measured in absence of noise for a star that is background to the cloud (as seen from an observer), is given by

$$s_C^0 = \begin{pmatrix} q_{V,C}^0 \\ u_{V,C}^0 \end{pmatrix} = p_C^{max} cos^2 \gamma_B \begin{pmatrix} cos[2\psi_B] \\ sin[2\psi_B] \end{pmatrix}, \tag{5}$$

where P_{max} = 13% E_{B-V}. In general,E_{B-V} depends on the dust grain opacity- which depends on the physical properties of the grains and on the observation wavelength-, the density of the dust grains and the path length through the cloud.

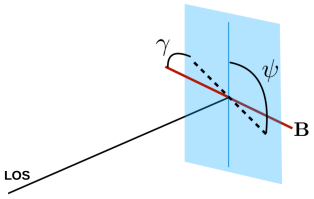


Figure 2: Homogeneous cloud permeated by a uniform magnetic field **B** with inclination γ_B on the plane of the sky and with position angle ψ_B.[Pelgrims, V.]

In the presence of one cloud, the corresponding Stokes parameters vector of a star i is given by the following formula

$$s_i = \begin{pmatrix} q_{V,i} \\ u_{V,i} \end{pmatrix} = s_{C,i} - n_i, \tag{6}$$

where n_i is the observational noise and $s_{C,i}$ is the Stokes parameters vector of the polarization omitted by the cloud C to a star i if the star is behind the cloud

$$s_{C,i} = \begin{cases} s^0_{C,i} + \delta s_{C,i} & , \ d_i > d_C \\ 0 & , \ d_i < d_C \end{cases}$$

where $s^0_{C,i}$ is given by **Eq. 3**, δs_C,i is the effect of the intrisic scatter in cloud C on star and d_i is the star distance. In this report $\delta_{C,i} = 0$, since there is assumed that there is not intrisic scattering effect. Using the Heaviside function H the above formula takes the following form

$$s_{C,i} = (s^0_{C,i} + \delta s_{C,i}) H(d_i - d_C).$$

For N_C homogeneous clouds the signal term accounts for the contribution from all intervening clouds

$$s_{C,i} = \sum_{k=1}^{N_C} (s^{0[k]}_{C,i} + \delta s^{[k]}_{C,i}) H(d_i - d_C). \tag{7}$$

2.2 Dataset Construction

The polarimetric dataset was constructed using different mock starlight polarization example samples. At each example a cloud with different properties is placed along the line of sight. The different dust cloud configurations are constructed from **Eq. 3** and **Eq. 4** for varying values for polarization and distance of the cloud parameters. The distance of each cloud d_C is such that 10% of the stars are foregrounds of the cloud. In this project, a cloud can be present in 39 different positions, starting from 280pc to 1610pc with step 35pc. The polarization P_{max} is sampled from an ensemble where the degree of polarization p is (0.3%, 0.4%) for the first dataset and (0.8%, 0.9%) for the second dataset. The polarization position angle ψ_B is sampled from the degrees (0.39, 1.17, 1.96, 2.74) . Each dust configuration is applied to the stars. The background stars get polarized by the cloud and foreground stars are kept unchanged.

In **Fig. 3** there is illustrated a generated mock starlight polarization example sample. On the left there is a signal at complete absence of noise. There in not included scatter in the signal induced by the cloud, i.e $n_i = 0$. There is neither observational noise added to the distance of the star through the uncertainty, i.e $s_\varpi = 0$, nor instrinsic scatter, i.e $\delta s_{C,i} = 0$. On the right, the scatter in the signal induced by the cloud is derived solely from observational uncertainties, with $n_i = 0.002$. Observational noise is also added to the positions of the stars with value $s_\varpi = 4.5 * 10^{-5}$, where ϖ is the parallaxes. Noise is considered un-correlated. Moreover, instrinsic scatter is included, with value $\delta s_{C,i} = 0.001$.

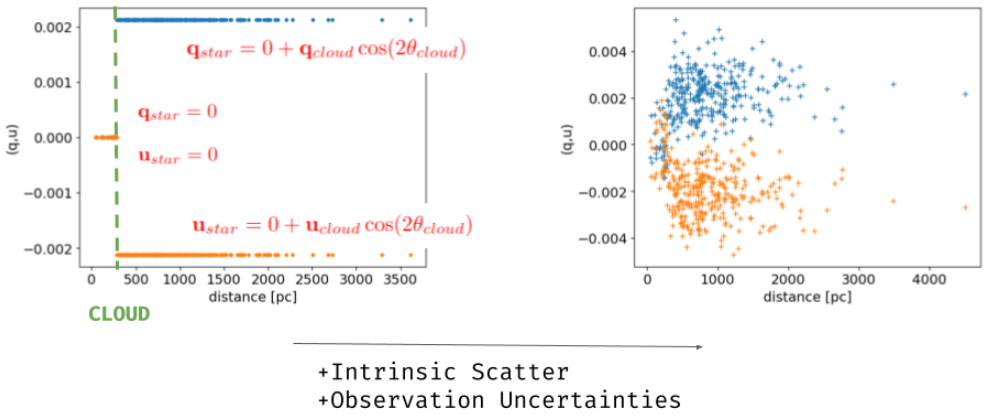


Figure 3: Clean Polarization Observations (left), Noisy Polarization Observations (right)

3 Problem Solution Methodology

3.1 Denoising Stellar Polarization Signals

In the domain of astrophysical signal processing, the inherent characteristics of stellar polarization signals exhibit a step function - like model, which encapsulates the sudden changes in polarization when light traverses through a dust cloud. The choice of a step function introduces a natural sparsity in the signal representation due to its localized and discrete nature, which aligns with the principles of sparse representation. SR aims to express each noisy signal $\mathbf{s}_n \in \mathbb{R}^M$ as a sparse linear combination of elements from a noisy dictionary matrix $\mathbf{D}_n \in \mathbb{R}^{M\times N}$ $\mathbf{s}_n = \mathbf{D}_n\mathbf{w}$, $\mathbf{w} \in \mathbb{R}^N$.
The best regularizer for sparsity promotion is known to be l_0-norm, however this leads to an intractable solution, thus l_1-norm is used. We are left with the following problem:

$$\mathbf{w}^* = \underset{\mathbf{w}}{\text{argmin}} ||\mathbf{s}_n - \mathbf{D}_n\mathbf{w}||_F^2 + \lambda||\mathbf{w}||_1$$

where λ is a regularization parameter. Afterwards, the denoised signal $\mathbf{s}_c \in \mathbb{R}^M$ is extracted as $\mathbf{s}_c = \mathbf{D}_c\mathbf{w}^*$, with $\mathbf{D}_c \in \mathbb{R}^{M\times N}$ the corresponding denoised dictionary.

3.2 Coupled Dictionary Learning

Coupled Dictionary Learning depends on creating a set of dictionaries that collectively represent the noisy $\mathbf{S}_n$ and clean $\mathbf{S}_c$ signal matrices. These dictionaries encode signals with sparse representations in relation to the trained dictionaries. The primary objective is to identify a paired set of dictionaries, $\mathbf{D}_n$ and $\mathbf{D}_c$, corresponding to the $\mathbf{S}_n$ and $\mathbf{S}_c$. In a formal sense, this pairing can be determined by solving a specific set of sparse decompositions.

$$\begin{aligned} &\underset{\mathbf{D}_c,\mathbf{D}_n,\mathbf{w}_c,\mathbf{w}_n}{\text{argmin}} \quad ||\mathbf{S}_c - \mathbf{D}_c\mathbf{W}_c||_F^2 + ||\mathbf{S}_n - \mathbf{D}_n\mathbf{W}_n||_F^2 + \lambda_c||\mathbf{W}_c||_1 + \lambda_n||\mathbf{W}_n||_1 \\ &\text{subject to} \quad \mathbf{W}_c = \mathbf{W}_n,\ ||\mathbf{D}_c(:,i)||_2 \le 1,\ ||\mathbf{D}_n(:,i)||_2 \le 1 \\ &\qquad\qquad \forall j, ||a_j||_0 \le L \end{aligned}$$

where $\mathbf{W}_n$ and $\mathbf{W}_c$ correspond to the sparse coefficient matrices for the noisy and clean signal matrices respectively, with λ_n and λ_c the sparsity regularization parameters.

The ADMM technique considers the distinct structure of individual variables, focusing on minimizing the augmented Lagrangian function associated with each variable:

$$\begin{aligned}\mathcal{L}(\mathbf{D}_c, \mathbf{D}_n, \mathbf{W}_c, \mathbf{W}_n, \mathbf{P}, \mathbf{Q}, Y_1, Y_2, Y_3) = \frac{1}{2}||\mathbf{D}_c\mathbf{W}_c - \mathbf{S}_c||_F^2 + \frac{1}{2}||\mathbf{D}_n\mathbf{W}_n - \mathbf{S}_n||_F^2 + \lambda_c||\mathbf{P}||_1 \\ + \lambda_n||\mathbf{Q}||_1 + \langle Y_1, \mathbf{P} - \mathbf{W}_c\rangle + \langle Y_2, \mathbf{Q} - \mathbf{W}_n\rangle + \langle Y_3, \mathbf{W}_c - \mathbf{W}_n\rangle \\ + \frac{c_1}{2}||\mathbf{P} - \mathbf{W}_c||_F^2 + \frac{c_2}{2}||\mathbf{Q} - \mathbf{W}_n||_F^2 + \frac{c_3}{2}||\mathbf{W}_c - \mathbf{W}_n||_F^2\end{aligned}$$

where Y_1, Y_2 and Y_3 represent the Lagrange multiplier matrices, $c_1 > 0$, $c_2 > 0$ and $c_3 > 0$ denote the step size parameters.

Adhering to the algorithmic approach of the ADMM scheme, we aim to identify the stationary point by iteratively solving for each variable while holding the others constant. The comprehensive steps of the algorithm for acquiring coupled dictionaries and the update formulas are outlined as follows:

- $\mathbf{W}_c = (\mathbf{D}_c^T\mathbf{D}_c + c_1\mathbf{I} + c_3\mathbf{I})^{-1} \cdot (\mathbf{D}_c^T\mathbf{S}_c + Y_1 - Y_3 + c_1\mathbf{P} + c_3\mathbf{W}_n)$
- $\mathbf{W}_n = (\mathbf{D}_n^T\mathbf{D}_n + c_2\mathbf{I} + c_3\mathbf{I})^{-1} \cdot (\mathbf{D}_n^T\mathbf{S}_n + Y_2 + Y_3 + c_2\mathbf{Q} + c_3\mathbf{W}_c)$
- $\mathbf{P} = S_{\lambda_c}(|\mathbf{W}_c - Y_1/c_1|)$
- $\mathbf{Q} = S_{\lambda_n}(|\mathbf{W}_n - Y_2/c_2|)$
- $\mathbf{D}_c^{k+1}(:, j) = \mathbf{D}_c^k(:, j) + \mathbf{S}_c\mathbf{W}_c(j, :)/(\mathbf{W}_c(j, :)\mathbf{W}_c^T(j, :) + \delta)$
- $\mathbf{D}_n^{k+1}(:, j) = \mathbf{D}_n^k(:, j) + \mathbf{S}_n\mathbf{W}_n(j, :)/(\mathbf{W}_n(j, :)\mathbf{W}_n^T(j, :) + \delta)$
- $Y_1^{k+1} = Y_1^k + c_1(\mathbf{P} - \mathbf{W}_c)$
- $Y_2^{k+1} = Y_2^k + c_2(\mathbf{Q} - \mathbf{W}_n)$
- $Y_3^{k+1} = Y_3^k + c_3(\mathbf{W}_c - \mathbf{W}_n)$

For analytical information on how the update equations are acquired see Appendix A (5).

3.3 Testing

For acquiring the denoised signal, we follow the SR framework as briefly described in 3.1. Given $\mathbf{D}_c \in \mathbb{R}^{M\times N}$ and $\mathbf{D}_n \in \mathbb{R}^{M\times N}$ from training step, where M is the size of each signal and N the size of dictionary

$$\mathbf{S}_n = \mathbf{D}_n\mathbf{W}_n.$$

In order to find $\mathbf{W}_n^* = \mathbf{W}_c^*$, we solve using Lasso the following minimization problem

$$\mathbf{W}_n^* = \underset{\mathbf{W}_n}{\text{argmin}}\ \frac{1}{2}||\mathbf{D}_n\mathbf{W}_n - \mathbf{S}_n||_F^2 + \lambda||\mathbf{W}_n||_1$$

where $||\cdot||_F$ denotes the Frobenius norm, and λ controls the sparsity of the solution.
The next step is to apply $\mathbf{W}_n^* = \mathbf{W}_c^*$ to calculate the reconstructed clean signal

$$\mathbf{S}_c^* = \mathbf{D}_c\mathbf{W}_c^*$$

In the context of this report, an Elastic Net regularization approach was also implemented. The motivation behind trying out Elastic Net on this problem lies in its ability to handle both L1 (Lasso) and L2 (Ridge) regularization terms, providing a balance between variable selection (sparsity) and parameter grouping.

$$\mathbf{W}_n^* = \text{argmin}\ \frac{1}{2}||\mathbf{D}_n\mathbf{W}_n - \mathbf{S}_n||_F^2 + \lambda(\alpha||\mathbf{W}_n||_1 + \frac{1-\alpha}{2}||\mathbf{W}_n||_2^2)$$

where λ denotes the regularization parameter controlling the overall strength of regularization, and $\alpha \in [0, 1]$ is the mixing parameter balancing l_1 and l_2 regularization.

3.4 Detecting Interstellar Magnetized Clouds

In order to locate the position of the interstellar dust cloud, we applied the following *"naive"* method. For reconstructed signal $\mathbf{s}_c \in \mathbb{R}^M$, let $\mathbf{ds}_c^*$ be a vector containing $(\mathbf{s}_c^*(i+1) - \mathbf{s}_c^*(i))$, for $i = 0 \dots M-1$. Then for the position of the magnetized cloud, x:

$$x = \text{argmax } \mathbf{ds}_c^*$$

Essentially, this considers x as the point of maximum slope. In order to calculate the actual distance in parsec:

$$x\ (pc) = x \cdot 10\ (pc).$$

4 Experimental Setup & Results

In this section the implementation of the above methodology will be displayed along with its corresponding results. There were examined 2 different datasets, with polarization steps (0.3, 0.4) and (0.8, 0.9). For each case, there were 623 different LOS used for training, resulting to 996 (q, u) training signals, and 125 different LOS for testing, resulting to 250 (q, u) testing signals. In **Fig. 4** an example of the signals used is displayed.

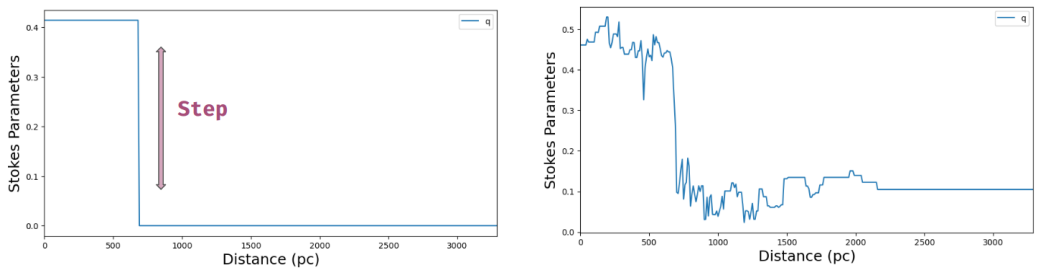
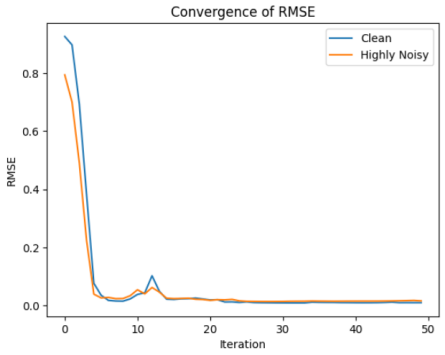


Figure 4: Clean Signal with polarization step 0.4 (left), Corresponding Noisy Signal (right)

4.1 Small Step Dataset

Below, there are the results derived from the analysis of the fist dataset with the smaller step. In **Fig. 5**, we present the learning curve of the training procedure which appears to decline, indicating that our model is being trained correctly. In particular, the RMSE between the original clean signals and the ones reconstructed by the model is approximately 0.012 and the RMSE between the original noisy signals and the ones reconstructed is approximately 0.014.



RMSE of Trained Signals	
(Reconstructed Clean, Original Clean)	0.012
(Reconstructed Noisy, Original Noisy)	0.014

Figure 5: Training Learning Curve (left), Root Mean Squared Errors of Training Signals (right)

In **Fig. 6**, there are two examples of the results derived from the testing procedure. On the left we can see a clean test signal, on the center the corresponding noisy signal, and on the right the reconstructed clean signal generated by our model. Cumulatively for all the test signals, the RMSE is approximately equal to 0.121.

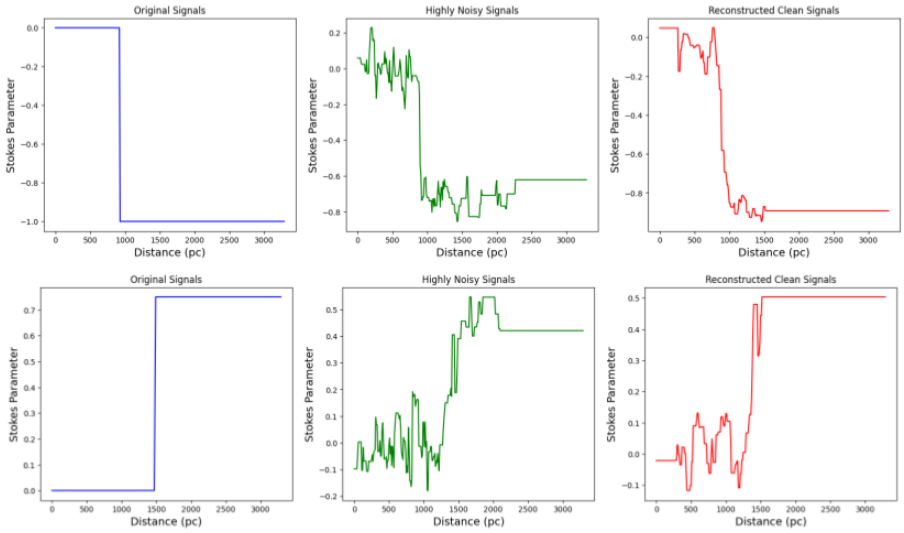
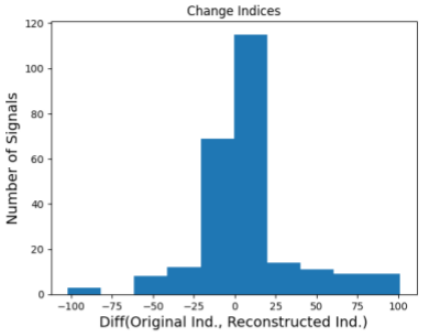


Figure 6: Examples of Reconstructed Test Signals

Furthermore, there were derived some statistics that examine the distribution of the change point position detected at each test signal. These results are presented in **Fig. 7**. The mean absolute difference between the true change index at a signal and the one detected by the model is approaching 15 indices. This means that the true change point can be detected at the denoised signals with distance uncertainty equal to 150 pc. Without denoising the signals, the mean absolute difference between the true change index and the one detected is approaching 24 indices, i.e 240 pc uncertainty. In detail, the percentage of the true detections is 25% and the percentage of the true detections with 50pc uncertainty is 74%. More statistical results are listed in **Fig. 7**.

Change Indices	
True / All	~ 25%
(True ∓ 50pc) / All	~ 74%
(True ∓ 150pc) / All	~ 82%
(True ∓ 250pc) / All	~ 85%

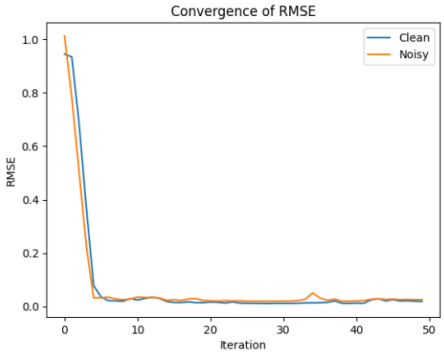


Mean |Difference| of Change Indices	
(Change Ind. Recon. Clean, Change Ind. Original Clean)	~ 15 indices ~ 150 pc
(Change Ind. Noisy, Change Ind. Original Clean)	~ 24 indices ~ 240pc

Figure 7: Change Detection Indices Distribution

4.2 Large Step Dataset

Below, we can see the results derived from the analysis of the second dataset with the larger step. In **Fig. 8**, we provide learning curve of the training procedure. Again, we can see that the curve declines, indicating that our model is being trained correctly. The RMSE between the original clean signals and the ones reconstructed by the model is approximately 0.018 and the RMSE between the original noisy signals and the ones reconstructed is approximately 0.025.



RMSE of Trained Signals	
(Reconstructed Clean, Original Clean)	0.018
(Reconstructed Noisy, Original Noisy)	0.025

Figure 8: Training Learning Curve (left), Root Mean Squared Errors of Training Signals (right)

In **Fig. 9**, there are two examples of the results derived from the testing procedure. On the left we can see is a clean test signal, on the center the corresponding noisy signal, and on the right the reconstructed clean signal. Cumulatively for all the test signals, the RMSE is approximately equal to 0.089.

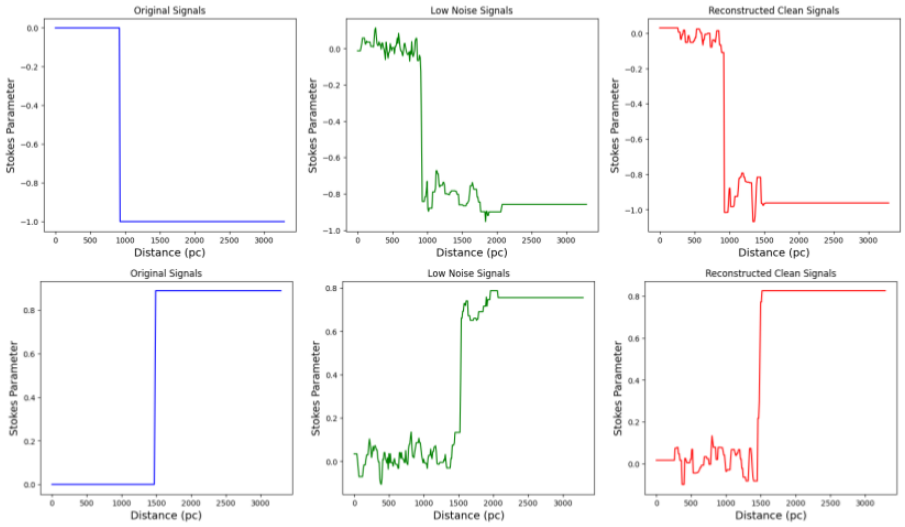
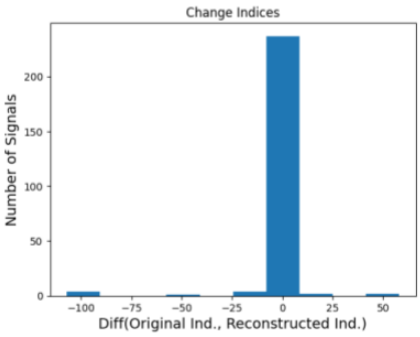


Figure 9: Examples of Reconstructed Test Signals

As before, we extracted statistical information about the distribution of the change point position detected at each test signal. These results are presented in **Fig. 10**. The mean absolute difference between the true change index at a signal and the one detected by the model is approximately 2 indices. This means that the true change point can be detected using the denoised signals with distance uncertainty equal to 20 pc. Without denoising the signals, the mean absolute difference between the true change index and the one detected is approximately 4 indices, i.e 40 pc uncertainty. In detail, the percentage of the true detections is 78% and the percentage of the true detections with 50pc uncertainty is 95%. More statistical results are listed in **Fig. 10**.

Change Indices	
True / All	~ 78%
(True ∓ 50pc) / All	~ 95%
(True ∓ 150pc) / All	~ 99%
(True ∓ 250pc) / All	~ 99%



Mean |Difference| of Change Indices	
(Change Ind. Recon. Clean, Change Ind. Original Clean)	~ 2 indices ~ 20 pc
(Change Ind. Noisy, Change Ind. Original Clean)	~ 4 indices ~ 40 pc

Figure 10: Change Detection Indices Distribution

5 Discussion & Conclusions

Our conclusions and proposed future steps include the presence of varying polarization magnitudes within the same dataset, as well as the need for separated training for q and u parameters. Additionally, in order to capture the shared information of the q and u parameters, we would suggest a similar application with 2-dimensional signals (consisting of both u and q signals) where coupled dictionary learning would applied with 3-dimensional dictionaries. Furthermore, the incorporation of more training data and a larger dictionary size, would further assist in achieving a more robust capture of underlying patterns. Finally, our future strategy would be to feed the denoised signals into a Neural Network designed for cloud detection, as this would improve the model's naive way of predicting the position of a cloud. These findings will help us improve our methods and advance astrophysical signal analysis.

References

[1] Pelgrims, V., Clark, S., Hensley, B., Panopoulou, G., Pavlidou, V., Tassis, K., Eriksen, H.K. & Wehus, I.K. (2021) Evidence for line-of-sight frequency decorrelation of polarized dust emission in Planck data. *Astronomy & Astrophysics*

[2] Pelgrims V., Macías-Pérez J.F. & Ruppin F. (2021) Galactic magnetic field reconstruction using the polarized diffuse Galactic emission: Formalism and application to Planck data. *Astronomy & Astrophysics*

[3] Konstantina Fotiadou, Grigorios Tsagkatakis, Bruno Moraes, Filipe B. Abdalla, Panagiotis Tsakalides (2017) Denoising Galaxy Spectra with Coupled Dictionary Learning. *EUSIPCO*

A Update Equations for ADMM

$$\begin{aligned}\mathcal{L}(\mathbf{D}_c, \mathbf{D}_n, \mathbf{W}_c, \mathbf{W}_n, \mathbf{P}, \mathbf{Q}, Y_1, Y_2, Y_3) = &\frac{1}{2}||\mathbf{D}_c\mathbf{W}_c - \mathbf{S}_c||_F^2 + \frac{1}{2}||\mathbf{D}_n\mathbf{W}_n - \mathbf{S}_n||_F^2 \\ &+ \lambda_c||\mathbf{P}||_1 + \lambda_n||\mathbf{Q}||_1 + \langle Y_1, \mathbf{P} - \mathbf{W}_c\rangle \\ &+ \langle Y_2, \mathbf{Q} - \mathbf{W}_n\rangle + \langle Y_3, \mathbf{W}_c - \mathbf{W}_n\rangle \\ &+ \frac{c_1}{2}||\mathbf{P} - \mathbf{W}_c||_F^2 + \frac{c_2}{2}||\mathbf{Q} - \mathbf{W}_n||_F^2 + \frac{c_3}{2}||\mathbf{W}_c - \mathbf{W}_n||_F^2\end{aligned}$$

Differentiating over $\mathbf{W}_c$:

$$\frac{\partial \mathcal{L}}{\partial \mathbf{W}_c} = -\mathbf{D}_c^T(\mathbf{S}_c - \mathbf{D}_c\mathbf{W}_c) - Y_1 - c_1(\mathbf{P} - \mathbf{W}_c) + Y_3 + c_3(\mathbf{W}_c - \mathbf{W}_n)$$

For $\frac{\partial \mathcal{L}}{\partial \mathbf{W}_c} = 0$,

$$\begin{aligned}&\frac{\partial \mathcal{L}}{\partial \mathbf{W}_c} = 0 \Leftrightarrow \\ -\mathbf{D}_c^T(\mathbf{S}_c - \mathbf{D}_c\mathbf{W}_c) - Y_1 - c_1(\mathbf{P} - \mathbf{W}_c) + Y_3 + c_3(\mathbf{W}_c - \mathbf{W}_n) = 0 &\Leftrightarrow \\ \mathbf{D}_c^T\mathbf{D}_c\mathbf{W}_c + c_1\mathbf{W}_c = \mathbf{D}_c^T\mathbf{S}_c + Y_1 + c_1\mathbf{P} - Y_3 + c_3\mathbf{W}_n &\Leftrightarrow \\ (\mathbf{D}_c^T\mathbf{D}_c + c_1\mathbf{I} + c_3\mathbf{I})\mathbf{W}_c = \mathbf{D}_c^T\mathbf{S}_c + Y_1 + c_1\mathbf{P} - Y_3 + c_3\mathbf{W}_n &\Leftrightarrow \\ \mathbf{W}_c = (\mathbf{D}_c^T\mathbf{D}_c + c_1\mathbf{I} + c_3\mathbf{I})^{-1}\cdot(\mathbf{D}_c^T\mathbf{S}_c + Y_1 - Y_3 + c_1\mathbf{P} + c_3\mathbf{W}_n)\end{aligned}$$

Differentiating over $\mathbf{W}_n$:

$$\frac{\partial \mathcal{L}}{\partial \mathbf{W}_n} = -\mathbf{D}_n^T(\mathbf{S}_n - \mathbf{D}_n\mathbf{W}_n) - Y_2 - c_2(\mathbf{Q} - \mathbf{W}_n) - Y_3 + c_3(\mathbf{W}_n - \mathbf{W}_c)$$

For $\frac{\partial \mathcal{L}}{\partial \mathbf{W}_n} = 0$

$$\begin{aligned}&\frac{\partial \mathcal{L}}{\partial \mathbf{W}_n} = 0 \Leftrightarrow \\ -\mathbf{D}_n^T(\mathbf{S}_n - \mathbf{D}_n\mathbf{W}_n) - Y_2 - c_2(\mathbf{Q} - \mathbf{W}_n) - Y_3 + c_3(\mathbf{W}_n - \mathbf{W}_c) = 0 &\Leftrightarrow \\ \mathbf{D}_n^T\mathbf{D}_n\mathbf{W}_n + c_2\mathbf{W}_n + c_3\mathbf{W}_n = \mathbf{D}_n^T\mathbf{S}_n + Y_2 + c_2\mathbf{Q} + Y_3 + c_3\mathbf{W}_c &\Leftrightarrow \\ (\mathbf{D}_n^T\mathbf{D}_n + c_2\mathbf{I} + c_3\mathbf{I})\mathbf{W}_n = \mathbf{D}_n^T\mathbf{S}_n + Y_2 + c_2\mathbf{Q} + Y_3 + c_3\mathbf{W}_c &\Leftrightarrow \\ \mathbf{W}_n = (\mathbf{D}_n^T\mathbf{D}_n + c_2\mathbf{I} + c_3\mathbf{I})^{-1}\cdot(\mathbf{D}_n^T\mathbf{S}_n + Y_2 + c_2\mathbf{Q} + Y_3 + c_3\mathbf{W}_c)\end{aligned}$$

Differentiating over $\mathbf{P}$:

$$\frac{\partial \mathcal{L}}{\partial \mathbf{P}} = \lambda_c \cdot \text{sign}(\mathbf{P}) + Y_1 + c_1(\mathbf{P} - \mathbf{W}_c)$$

For $\frac{\partial \mathcal{L}}{\partial \mathbf{P}} = 0$

$$\begin{aligned}&\frac{\partial \mathcal{L}}{\partial \mathbf{P}} = 0 \Leftrightarrow \\ \lambda_c \cdot \text{sign}(\mathbf{P}) + Y_1 + c_1(\mathbf{P} - \mathbf{W}_c) = 0 &\Leftrightarrow \\ c_1\mathbf{P} = -\lambda_c \cdot \text{sign}(\mathbf{P}) - Y_1 + c_1\mathbf{W}_c &\Leftrightarrow \\ \mathbf{P} = -\frac{1}{c_1}(\lambda_c \cdot \text{sign}(\mathbf{P}) - Y_1 + c_1\mathbf{W}_c)\end{aligned}$$

If we denote S_{λ_c} the soft thresholding with parameter λ_c, the above can be written as

$$\mathbf{P} = S_{\lambda_c}(|\mathbf{W}_c - Y_1/c_1|)$$

Differentiating over $\mathbf{Q}$:

$$\frac{\partial \mathcal{L}}{\partial \mathbf{Q}} = \lambda_n \cdot \text{sign}(\mathbf{Q}) + Y_2 + c_2(\mathbf{Q} - \mathbf{W}_n)$$

For $\frac{\partial \mathcal{L}}{\partial \mathbf{Q}} = 0$

$$\begin{aligned}&\frac{\partial \mathcal{L}}{\partial \mathbf{Q}} = 0 \Leftrightarrow \\ \lambda_n \cdot \text{sign}(\mathbf{Q}) + Y_2 + c_2(\mathbf{Q} - \mathbf{W}_n) = 0 &\Leftrightarrow \\ c_2\mathbf{Q} = -\lambda_n \cdot \text{sign}(\mathbf{Q}) - Y_2 + c_2\mathbf{W}_n &\Leftrightarrow \\ \mathbf{Q} = -\frac{1}{c_2}(\lambda_n \cdot \text{sign}(\mathbf{Q}) - Y_2 + c_2\mathbf{W}_n)\end{aligned}$$

If we denote S_{λ_n} the soft thresholding with parameter λ_n, the above can be written as

$$\mathbf{Q} = S_{\lambda_n}\left(|\mathbf{W}_n - Y_2/c_2|\right)$$

Differentiating over $\mathbf{D}_c$:

$$\frac{\partial \mathcal{L}}{\partial \mathbf{D}_c} = (\mathbf{D}_c\mathbf{W}_c - \mathbf{S}_c)\mathbf{W}_c^T$$

For $\frac{\partial \mathcal{L}}{\partial \mathbf{D}_c} = 0$

$$\begin{aligned}
\frac{\partial \mathcal{L}}{\partial \mathbf{D}_c} = 0 \Leftrightarrow \\
(\mathbf{D}_c\mathbf{W}_c - \mathbf{S}_c)\mathbf{W}_c^T = 0 \Leftrightarrow \\
\mathbf{D}_c\mathbf{W}_c\mathbf{W}_c^T - \mathbf{S}_c\mathbf{W}_c^T = 0 \Leftrightarrow \\
\mathbf{D}_c = (\mathbf{S}_c\mathbf{W}_c^T)\cdot(\mathbf{W}_c\mathbf{W}_c^T)^{-1}
\end{aligned}$$

So, we update each column j of $\mathbf{D}_c$ as

$$\mathbf{D}_c^{(k+1)}(:,j) = \mathbf{D}_c^{(k)}(:,j) + (\mathbf{S}_c\mathbf{W}_c(j,:))\cdot(\mathbf{W}_c(j,:)\mathbf{W}_c(j,:)^T + \delta)^{-1}$$

Differentiating over $\mathbf{D}_n$:

$$\frac{\partial \mathcal{L}}{\partial \mathbf{D}_n} = (\mathbf{D}_n\mathbf{W}_n - \mathbf{S}_n)\mathbf{W}_n^T$$

For $\frac{\partial \mathcal{L}}{\partial \mathbf{D}_n} = 0$

$$\begin{aligned}
\frac{\partial \mathcal{L}}{\partial \mathbf{D}_n} = 0 \Leftrightarrow \\
(\mathbf{D}_n\mathbf{W}_n - \mathbf{S}_n)\mathbf{W}_n^T = 0 \Leftrightarrow \\
\mathbf{D}_n\mathbf{W}_n\mathbf{W}_n^T - \mathbf{S}_n\mathbf{W}_n^T = 0 \Leftrightarrow \\
\mathbf{D}_n = (\mathbf{S}_n\mathbf{W}_n^T)\cdot(\mathbf{W}_n\mathbf{W}_n^T)^{-1}
\end{aligned}$$

So, we update each column j of $\mathbf{D}_n$ as

$$\mathbf{D}_n^{(k+1)}(:,j) = \mathbf{D}_n^{(k)}(:,j) + (\mathbf{S}_n\mathbf{W}_n(j,:))\cdot(\mathbf{W}_n(j,:)\mathbf{W}_n(j,:)^T + \delta)^{-1}$$

Differentiating over Y_1:

$$\frac{\partial \mathcal{L}}{\partial Y_1} = \mathbf{P} - \mathbf{W}_c$$

So, we update Y_1 as

$$Y_1^{(k+1)} = Y_1^{(k)} + c_1(\mathbf{P} - \mathbf{W}_c)$$

Differentiating over Y_2:

$$\frac{\partial \mathcal{L}}{\partial Y_2} = \mathbf{Q} - \mathbf{W}_n$$

So, we update Y_2 as

$$Y_2^{(k+1)} = Y_2^{(k)} + c_2(\mathbf{Q} - \mathbf{W}_n)$$

Differentiating over Y_3:

$$\frac{\partial \mathcal{L}}{\partial Y_3} = \mathbf{W}_c - \mathbf{W}_n$$

So, we update Y_3 as

$$Y_3^{(k+1)} = Y_3^{(k)} + c_3(\mathbf{W}_c - \mathbf{W}_n)$$

.