RNAAS RESEARCH NOTES OF THE AAS

SpectralUnmix: A Torch-based Regularized Non-negative Matrix Factorization

To cite this article: Rafael S. de Souza *et al* 2026 *Res. Notes AAS* **10** 56

Draft version March 12, 2026
Typeset using LaTeX **RNAAS** style in AASTeX631

SpectralUnmix: A Torch-Based Regularized Non-negative Matrix Factorization

Rafael S. de Souza ,[1,2,3] Paula Coelho ,[4] Niranjana P ,[2] Ana L. Chies-Santos ,[2] and Rogério Riffel [2]

[1] *Centre for Astrophysics Research, University of Hertfordshire, College Lane, Hatfield, AL10 9AB, UK*
[2] *Instituto de Física, Universidade Federal do Rio Grande do Sul, Porto Alegre, RS 90040-060, Brazil*
[3] *Department of Physics & Astronomy, University of North Carolina at Chapel Hill, NC 27599-3255, USA*
[4] *Instituto de Astronomia, Geofísica e Ciências Atmosféricas, USP, Rua do Matão 1226, 05508-090, São Paulo, Brazil*

ABSTRACT

We present `SpectralUnmix`, an `R` package for regularized non-negative matrix factorization (NMF), implemented in `torch` with optional GPU acceleration. The package estimates low-rank non-negative representations through proximal-gradient updates and allows smoothness regularization along the spectral axis. As a compact demonstration, we apply the method to a subset of stellar spectra and compare the recovered NMF components with principal-component directions and representative stellar spectra. The package is released under the MIT license at this repository ; a copy has been deposited to Zenodo (de Souza 2026).

INTRODUCTION

Many spectral datasets can be represented as matrices in which rows correspond to samples and columns correspond to wavelength channels. A common objective in exploratory analysis is to describe such data through a small number of latent components and their corresponding weights. Principal component analysis (PCA; Jolliffe & Cadima 2016) has long been used in astronomy for this purpose (e.g. Ronen et al. 1999; Steiner et al. 2009; Ishida & de Souza 2013; Joseph et al. 2014; de Souza et al. 2014, 2022), providing an orthogonal basis that captures the dominant variance in spectral datasets. Non-negative matrix factorization (NMF) offers an alternative decomposition in which both the components and their weights are constrained to be non-negative (Lee & Seung 1999). Because astronomical spectra represent non-negative flux measurements, imposing non-negativity often yields components that are more directly interpretable in physical terms (e.g. Blanton & Roweis 2007; Hurley et al. 2013; Melchior et al. 2018). Such representations arise naturally in applications ranging from stellar libraries to spatially resolved spectroscopy. Our `R` package provides a general-purpose implementation of regularized NMF with `torch` as the computational backend. It is designed for exploratory decomposition problems where smooth, non-negative spectral components are desirable.

METHOD

Let $\mathbf{X}$ be a non-negative data matrix $\mathbf{X} \in \mathbb{R}_+^{N\times P}$, where N denotes the number of samples and P the number of spectral channels. We seek a low-rank non-negative representation $\mathbf{X} \approx \mathbf{AS}$, with $\mathbf{A} \in \mathbb{R}_+^{N\times K}$, and $\mathbf{S} \in \mathbb{R}_+^{K\times P}$. The rows of $\mathbf{S}$ define K latent spectra, while the columns of $\mathbf{A}$ contain the corresponding non-negative weights. We estimate $\mathbf{A}$ and $\mathbf{S}$ by solving

$$\min_{\mathbf{A},\mathbf{S}\geq 0} \frac{1}{2} \left\|\mathbf{X} - \mathbf{AS}\right\|_F^2 + \lambda \left\|\mathbf{SD}^\top\right\|_F^2 , \tag{1}$$

where $\mathbf{D} \in \mathbb{R}^{(P-1)\times P}$ is the first-order finite-difference operator along the spectral axis,

$$\mathbf{D} = \begin{bmatrix} -1 & 1 & 0 & \dots & 0 \\ 0 & -1 & 1 & \dots & 0 \\ \vdots & & & \ddots & \\ 0 & 0 & 0 & -1 & 1 \end{bmatrix}, \tag{2}$$

r.s.desouza@herts.ac.uk

so that the regularization penalizes large bin-to-bin variations and encourages smooth spectral components. In this case

$$\|\mathbf{S}\mathbf{D}^{\top}\|_F^2 = \sum_{k=1}^{K}\sum_{j=1}^{P-1}\left(S_{k,j+1} - S_{k,j}\right)^2 . \tag{3}$$

The first term in Equation 1 measures reconstruction fidelity, while the second penalizes channel-to-channel roughness in the recovered spectra. The standard NMF is recovered when $\lambda = 0$. The optimization is implemented in `torch` via alternating proximal-gradient updates (Xu & Yin 2013).

ANALYSIS

Here we apply `SpectralUnmix` to a subset of 80 representative stellar spectra from the spectral library of Coelho (2014), spanning four broad classes: hot stars, A/F stars, solar-like stars, and cool dwarfs, with 20 spectra per class. We fit a four-component NMF model and compare the recovered components with PCA obtained from the same data. Figure 1 summarizes this toy experiment. Panel A shows the class prototype spectra; panels B and C display the PCA and NMF eigenspectra, respectively, with colors indicating their one-to-one assignment to the prototype classes. For each component–class pair we compute the χ^2 distance between the corresponding eigenspectrum and the class prototype. Finally, panel D summarizes the class–component distances. In this example, the recovered NMF components trace physically recognizable spectral shapes while remaining smooth and non-negative by construction.

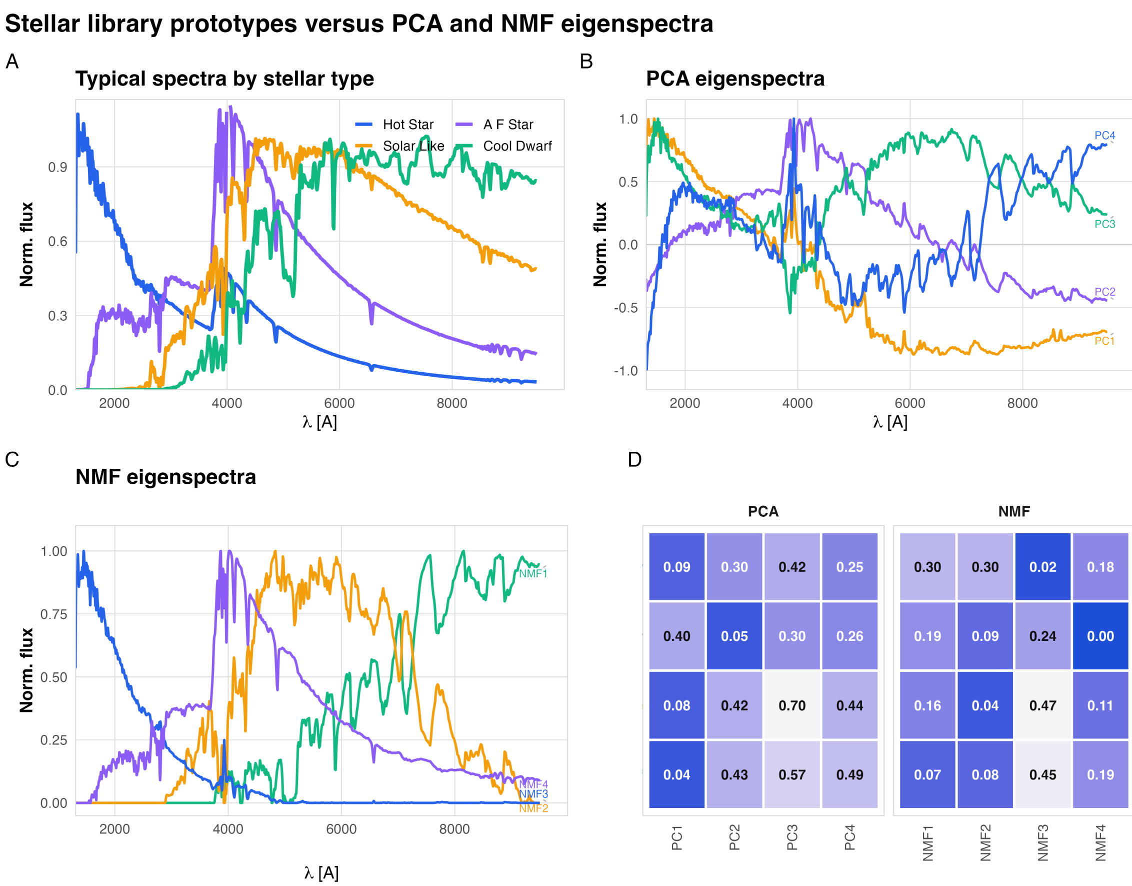

Figure 1. Panel A shows four representative normalized spectra out a sample of 80. Panels B and C compare PCA and NMF eigenspectra derived from the same data; colors indicate the stellar class assigned to each component through a one-to-one matching based on a simple χ^2-like distance between normalized spectral shapes. Panel D summarizes the class–component distances for PCA and NMF.

CONCLUSIONS

`SpectralUnmix` provides an extensible implementation of regularized NMF for spectral data analysis using `torch` in `R`. The package supports GPU acceleration and optional smoothness constraints, making it suitable for exploratory decomposition of high-dimensional spectral datasets. Although demonstrated here on a simple stellar library example, the framework is general and can be readily applied to a wide range of astronomical problems, including stellar population studies, hyperspectral imaging, and integral-field spectroscopy.

RSS acknowledges support from the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq, Brazil), process nos. 446508/2024-1 and 315026/2025-1.

REFERENCES

Blanton, M. R., & Roweis, S. 2007, AJ, 133, 734, doi: 10.1086/510127

Coelho, P. R. T. 2014, MNRAS, 440, 1027, doi: 10.1093/mnras/stu365

de Souza, R. S. 2026, SpectralUnmix: A Torch-Based Regularized Non- negative Matrix Factorization, v0.1.0, Zenodo, doi: 10.5281/zenodo.18982563

de Souza, R. S., Maio, U., Biffi, V., & Ciardi, B. 2014, MNRAS, 440, 240, doi: 10.1093/mnras/stu274

de Souza, R. S., Quanfeng, X., Shen, S., Peng, C., & Mu, Z. 2022, Astronomy and Computing, 41, 100633, doi: 10.1016/j.ascom.2022.100633

Hurley, P. D., Oliver, S., Farrah, D., Lebouteiller, V., & Spoon, H. W. W. 2013, MNRAS, 437, 241, doi: 10.1093/mnras/stt1875

Ishida, E. E. O., & de Souza, R. S. 2013, MNRAS, 430, 509, doi: 10.1093/mnras/sts650

Jolliffe, I. T., & Cadima, J. 2016, Philosophical Transactions of the Royal Society of London Series A, 374, 20150202, doi: 10.1098/rsta.2015.0202

Joseph, R., Courbin, F., Metcalf, R. B., et al. 2014, A&A, 566, A63, doi: 10.1051/0004-6361/201423365

Lee, D. D., & Seung, H. S. 1999, Nature, 401, 788, doi: 10.1038/44565

Melchior, P., Moolekamp, F., Jerdee, M., et al. 2018, A&C, 24, 129, doi: 10.1016/j.ascom.2018.07.001

Ronen, S., Aragón-Salamanca, A., & Lahav, O. 1999, MNRAS, 303, 284, doi: 10.1046/j.1365-8711.1999.02222.x

Steiner, J. E., Menezes, R. B., Ricci, T. V., & Oliveira, A. S. 2009, MNRAS, 395, 64, doi: 10.1111/j.1365-2966.2009.14530.x

Xu, Y., & Yin, W. 2013, SIAM Journal on Imaging Sciences, 6, 1758, doi: https://doi.org/10.1137/120887795