Optimal transport-based dictionary learning and its application to Euclid-like Point Spread Function representation

Morgan Schmitz, Matthieu Heitz, Nicolas Bonneel, Fred Ngolè, David Coeurjolly, Marco Cuturi, Gabriel Peyré, Jean-Luc Starck

Optimal transport-based dictionary learning and its application to Euclid-like Point Spread Function representation

Morgan A. Schmitz[a], Matthieu Heitz[b], Nicolas Bonneel[b], Fred Ngolè[c], David Coeurjolly[b], Marco Cuturi[d], Gabriel Peyré[e], and Jean-Luc Starck[a]

[a]Astrophysics Department, IRFU, CEA, Université Paris-Saclay, F-91191 Gif-sur-Yvette, France
Université Paris-Diderot, AIM, Sorbonne Paris Cité, CEA, CNRS, F-91191 Gif-sur-Yvette, France
[b]Université de Lyon, CNRS/LIRIS, Lyon, France
[c]LIST, Data Analysis Tools Laboratory, CEA Saclay, France
[d]Centre de Recherche en Economie et Statistique, Paris, France
[e]DMA, ENS Ulm, Paris, France

ABSTRACT

Optimal Transport theory enables the definition of a distance across the set of measures on any given space. This Wasserstein distance naturally accounts for geometric warping between measures (including, but not exclusive to, images). We introduce a new, Optimal Transport-based representation learning method in close analogy with the usual Dictionary Learning problem. This approach typically relies on a matrix dot-product between the learned dictionary and the codes making up the new representation. The relationship between atoms and data is thus ultimately linear. By reconstructing our data as Wasserstein barycenters of learned atoms instead, our approach yields a representation making full use of the Wasserstein distance's attractive properties and allowing for non-linear relationships between the dictionary atoms and the datapoints.
We apply our method to a dataset of Euclid-like simulated PSFs (Point Spread Function). ESA's Euclid mission will cover a large area of the sky in order to accurately measure the shape of billions of galaxies. PSF estimation and correction is one of the main sources of systematic errors on those galaxy shape measurements. PSF variations across the field of view and with the incoming light's wavelength can be highly non-linear, while still retaining strong geometrical information, making the use of Optimal Transport distances an attractive prospect. We show that our representation does indeed succeed at capturing the PSF's variations.

Keywords: Optimal Transport, Dictionary Learning, Point Spread Function

1. INTRODUCTION

Feature learning, also known as representation learning and often associated with dimensionality reduction, is the branch of machine learning methods that deal with the preprocessing step of creating new features using the ones in existing data. The resulting features are more convenient and/or appropriate for whatever tasks must then be performed. As a typical example, one can think of images: in its rawest form, the data from an image is

Further author information: (Send correspondence to M.A.S.)
M.A.S.: E-mail: morgan.schmitz@cea.fr, Telephone: +33 (0)1 69 08 83 77
M.H.: E-mail: matthieu.heitz@univ-lyon1.fr
N.B.: E-mail: nicolas.bonneel@liris.cnrs.fr
F.N.: E-mail: fred-maurice.ngole-mboula@cea.fr
D.C.: E-mail: david.coeurjolly@liris.cnrs.fr
M.C.: E-mail: marco.cuturi@ensae.fr
G.P.: E-mail: gabriel.peyre@ens.fr
J-L.S.: E-mail: jean-luc.starck@cea.fr

a vector of pixel intensities. This can of course quickly lead to extremely high-dimensional data, not to mention most of the information is likely to be contained in a small subset of all those pixels. If one wanted to perform, say, a classification task on a set of images, using pixel intensities as the features would be cumbersome and likely lead to poor results - using a preliminary feature learning step, however, can give the data a more appropriate form and dramatically improve the accuracy of the classification to follow.
Dictionary learning is a subset of those methods wherein, instead of reconstructing data using a predetermined dictionary (*e.g.*, based on wavelets[1]), the dictionary itself is learned from the data. This is typically achieved by minimizing a loss function consisting of a similarity term ensuring the outcome successfully reconstructs the training data, and (optionally) a set of constraint terms that enforce the learned representation exhibits some desired properties (for instance, sparsity[2]). In its most common form, the dictionary is a matrix composed of a certain number of atoms, each of the same dimensionality as a datapoint. Data X is then reconstructed by performing a matrix dot product between this dictionary D and a set of codes Λ that are learned simultaneously: $X \approx D\Lambda$.
In this paper, we introduce a new approach where we replace this linear formulation by one based on Optimal Transport (see section 2). We then illustrate the use of our method on a toy example in subsection 3.1 and on the problem of modelling a space-based telescope's Point Spread Function (hereafter PSF) in subsection 3.2.

2. OPTIMAL TRANSPORT

2.1 Wasserstein distances and entropy

Optimal Transport theory[3] enables one to define distances (the most common being the Wasserstein distance, W) on the set of measures over a given space, which, informally, can be thought of as the minimal cost of moving a heap of sand μ toward a hole in the ground ν, knowing the effort needed to move quantities of sand to any other location (see Figure 1).

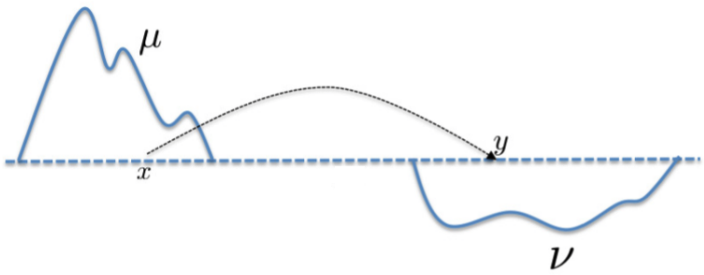

Figure 1. Graphical representation of the mass transportation problem. The minimal effort cost to transport one measure into the other defines an Optimal Transport distance between μ and ν.

Formally, let $\mathcal{P}(\Omega)$ be the set of measures on some space Ω. Let $\mu, \nu \in \mathcal{P}(\Omega)$, and $c : \Omega \times \Omega \mapsto \mathbb{R}$ a cost function. The Monge-Kantorovich problem is:

$$\inf \left\{ \int_{\Omega \times \Omega} c(x_1, x_2) d\pi(x_1, x_2),\ \pi \in \Pi(\mu, \nu) \right\}, \tag{1}$$

where $\Pi(\mu, \nu)$ is the set of transport plans between μ and ν, that is, the set of all measures on $\Omega \times \Omega$ whose marginals are equal to μ and ν. The so-called Wasserstein distance between μ and ν is defined as the value reached in (1) when c is a metric on Ω.
In the discrete case, that is, when Ω has cardinal N for some finite N, measures are histograms, *i.e.*:

$$\mathcal{P}(\Omega) = \Sigma_N := \left\{ u \in \mathbb{R}^N, \sum_{i=1}^{N} u_i = 1 \right\}.$$

The cost function c can then be expressed as a matrix $C \in \mathbb{R}^{N \times N}$ whose entries c_{ij} are the cost of transportation between the i^{th} and the j^{th} element of grid Ω. The Wasserstein distance is then defined simply as:

$$W(\mu, \nu) := \min_{T \in \Pi(\mu,\nu)} \langle T, C \rangle, \tag{2}$$

where the set of admissible transport plans are now:

$$\Pi(\mu, \nu) := \left\{ T \in \mathbb{R}_+^{N \times N}, T\mathbb{1}_N = \mu, T^\top \mathbb{1}_N = \nu \right\}.$$

While W is always defined, its computation was only possible in practice for very low values of N. However, it was recently proposed[4] to add an entropy penalty term to its formulation (2), yielding the approximate Wasserstein distances defined as:

$$W_\gamma(\mu, \nu) := \min_{T \in \Pi(\mu,\nu)} \langle T, C \rangle + \gamma H(T), \tag{3}$$

where $H(T) = \sum_{i,j} T_{ij} \log(T_{ij} - 1)$ and $\gamma > 0$ is a user-selected parameter. The distance can then be computed using an iterative and efficient scheme based on the celebrated Sinkhorn algorithm.[5]

2.2 Wasserstein barycenter and dictionary learning

We now consider a set of histograms $D = d_1, \ldots, d_S$. For any given weights $\lambda = \lambda_1, \ldots, \lambda_S$ such that $\sum_i \lambda_i = 1$, their Euclidean barycenter is defined as:

$$P_e(D, \lambda) := \operatorname*{argmin}_{u \in \mathbb{R}^N} \sum_{s=1}^{S} \lambda_s \|u - d_s\|_2^2 = \sum_{s=1}^{S} \lambda_s d_s.$$

Since (approximate) Wasserstein distances are made readily computable by the addition of an entropy term as in (3), Wasserstein barycenters are defined by analogy with their Euclidean counterpart as:[6]

$$P\left(D, \lambda\right) := \operatorname*{argmin}_{u \in \Sigma_N} \sum_{s=1}^{S} \lambda_s W_\gamma(d_s, u). \tag{4}$$

These barycenters can also be computed using an iterative scheme[7] based on the Sinkhorn algorithm. A kernel K that depends only on the cost function and the entropy parameter γ is iteratively scaled by two sets of vectors $a = a_1, \ldots, a_S, b = b_1, \ldots, b_S$ as follows:

$$K = \exp\left(-\frac{C}{\gamma}\right)$$

$$a_s^{(l)} = \frac{d_s}{K b_s^{(l-1)}} \tag{5}$$

$$P^{(l)}\left(D, \lambda\right) = \prod_{s=1}^{S} \left(K^\top a_s^{(l)}\right)^{\lambda_s} \tag{6}$$

$$b_s^{(l)} = \frac{P^{(l)}\left(D, \lambda\right)}{K^\top a_s^{(l)}}, \tag{7}$$

where $\forall s, b_s^{(0)} = 1_N$ and the exp, $\prod$ and $\div$ operators are applied element-wise.
Unlike its Euclidean counterpart which simply stacks the input histograms by weighted sums, because of its Optimal Transport-based formulation, the Wasserstein barycenter warps them together. Barycenters of two simple shapes are shown for several different weights as an illustration in Figure 2 for the Euclidean case and Figure 3 for the Wasserstein case.

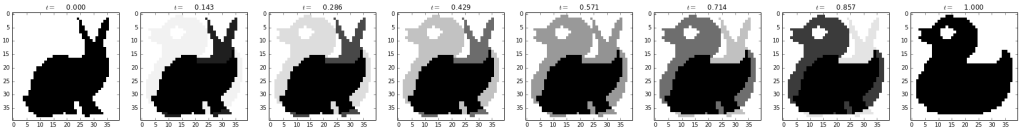

Figure 2. Evolution of the Euclidean barycenter of two images with varying weights.

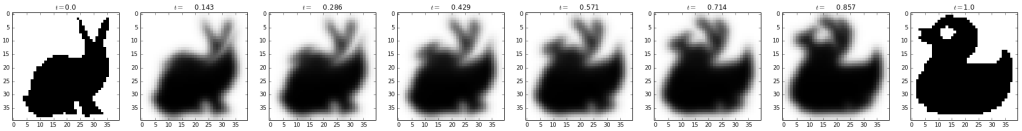

Figure 3. Evolution of the Wasserstein barycenter of two images with varying weights.

This property is the motivation for using the Wasserstein barycenter in a feature learning setting. Our method consists in replacing the usual, linear relation between atoms and codes in the dictionary learning problem ($X \approx D\Lambda$) by the approximate Wasserstein barycenter obtained after L iterations of the Sinkhorn algorithm (6): $X \approx \mathbf{P}^{(L)}(D, \Lambda)$, where $\Lambda \in \mathbb{R}^{S \times n}$ (with n the number of datapoints and S the user-specified number of atoms) and $\mathbf{P}^{(L)}(D, \Lambda) = \left(P^{(L)}(D, \lambda_1), \ldots, P^{(L)}(D, \lambda_n)\right)$.
The learning is performed by gradient descent, minimizing the following energy function:

$$\min_{D,\Lambda} \mathcal{E}_L(D, \Lambda) := \sum_{i=1}^{n} \mathcal{L}\left(P^{(L)}(D, \lambda_i), x_i\right), \tag{8}$$

Where each $x_i \in \mathbb{R}^N$ is a datapoint (for instance, an image with normalized pixel intensities), and $\mathcal{L}$ is an arbitrary loss function. By additivity, we can consider a single datapoint x without loss of generality (Λ is then made up of a single set of weights λ). Differentiating (8) yields:

$$\nabla_D \mathcal{E}_L(D, \Lambda) = \left[\partial_D P^{(L)}(D, \lambda)\right]^\top \nabla \mathcal{L}(P^{(L)}(D, \lambda), x) \tag{9}$$

$$\nabla_\lambda \mathcal{E}_L(D, \Lambda) = \left[\partial_\lambda P^{(L)}(D, \lambda)\right]^\top \nabla \mathcal{L}(P^{(L)}(D, \lambda), x). \tag{10}$$

The right-hand term is the gradient of the loss in its first argument and is usually readily computable, *e.g.*, if $\mathcal{L}(p, q) = \|p - q\|_2^2$ then $\nabla \mathcal{L}(p, q) = 2(p - q)$ (as will be the case for the applications in section 3). The left hand term is derived by automatic differentiation, that is, by application of the chain rule to the iterative updates (5)-(7). These derivations can either be done explicitly, 'by hand',[8] or, equivalently, by using an automatic differentiation tool such as the Theano library,[9] which computes the gradients after a number of operations equal to that of the forward loop used to compute the Wasserstein barycenter. This gives us a scheme to obtain the gradients in both dictionary and weights (9)-(10) which can then be used to find a local minimum in energy $\mathcal{E}_L$, yielding our representation. In the experiments that follow, we used an off-the-shelf L-BFGS solver.[10]

3. APPLICATIONS

3.1 Toy example

We first illustrate our method on a toy example. The dataset consists of a set of discretized Gaussian distributions on an evenly spaced grid of size 11. Each Gaussian is translated slightly on the grid. We then perform dimensionality reduction, fixing the number of atoms (or components) to $S = 2$, *via* Principal Components Analysis (PCA), Non-Negative Matrix Factorization (NMF), that is, dictionary learning with a constraint of positivity for both the dictionary and the codes, and our method. Despite the simplicity of the transformation, linear methods fail to describe it with only two components, while our method does reconstruct discretized Gaussians. Some datapoints, the atoms learned by all three methods, and the reconstructions are shown in Figure 4.

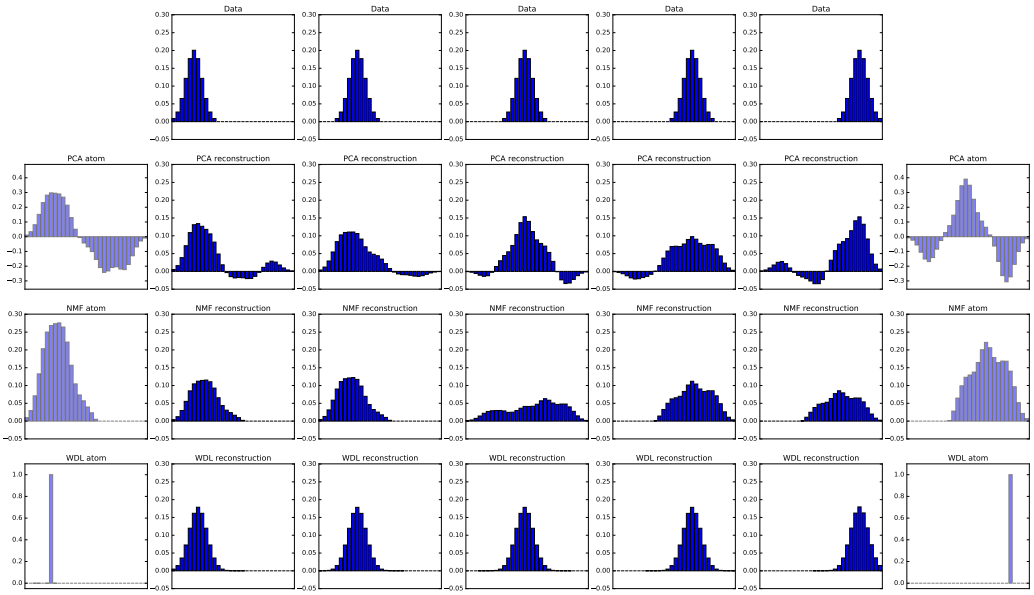

Figure 4. Learning two atoms from a set of translated discretized Gaussian distributions. Training data is shown in the first row. Each column underneath shows their reconstruction for all three methods: PCA, NMF and our method (WDL) for the second to fourth row, respectively. The two learned atoms learned by each method are shown at the ends of each row, in light purple.

Unlike PCA, both NMF and our approach are obtained by minimizing a non-convex function, and gradient descent can thus converge to a local minimum. We thus relaunched both several times (the results shown in Figure 4 are for the run that yielded the lowest value of the energy to minimize).
Not only does our method yield much better reconstructions, each of them are still histograms, unlike those of both PCA and NMF (which take negative values and do not sum to 1, respectively), as are the atoms we learn. Wasserstein barycenters of a set of a Gaussian distributions are known to also be Gaussian, so it might sound surprising that the atoms we learn are Diracs rather than the two extreme Gaussians. This is because the entropy term added to the definition of the Wasserstein distance in (3) causes the optimal transport plan to get blurred, which in turn induces a blurring of the Wasserstein barycenters. This can be mitigated by choosing lower values of γ, with $W_\gamma \to W$ as $\gamma \to 0$. However, picking too small a value for γ can lead to computational problems as values within the scaling vectors a and b (see (5), (7)) can then tend to infinity. However, as illustrated with this toy example, the blurring induced by the entropy term is taken into account by our method, which learns

sharper atoms - in this case, the atoms follow a Dirac and not a Gaussian distribution, but we have observed the same phenomenon in practice with more complicated types of data (*e.g.*, natural images).

3.2 Chromatic PSF representation

Modern observational cosmology aims at deriving tight constraints on the cosmological parameters contained within the standard cosmological model. One of the aims of this approach is to understand the nature of Dark Energy, and whether or not the λCDM model holds. Under λCDM, Dark Energy is a cosmological constant which is responsible for the observed acceleration of the expansion of the universe.

Several cosmological probes exist, each of them deriving constraints on our cosmological models by exploiting the signal we receive from various different sources and our current understanding of the Universe. Under General Relativity, light's path follows the curvature of space-time induced by massive objects, thus generating an effect similar to that of lenses in optics. Weak Lensing is a particularly promising cosmological probe that uses these effects as signal by measuring the shapes of distant galaxies, and inferring the amount of mass the photons have been affected by as they traveled towards us. This enables us to probe the Large Scale Structure of the Universe, and gives us a wealth of information on cosmological parameters, including those related to Dark Matter (as we cannot observe it directly, but it does cause gravitational lensing effects).

Images produced by telescopes are distorted for several different reasons, including the instrument's optics. The PSF is the kernel of these convolutions, and varies both chromatically (with the incoming light's wavelength) and spatially (across the instrument's field of view). Understanding and accounting for the PSF is naturally of paramount importance in Weak Lensing surveys, as they can otherwise affect the observed shape of galaxies, thus reducing (or even biasing) the Weak Lensing signal. Up until now, major Weak Lensing surveys have been conducted from ground-based telescopes,[11–13] in which case atmospheric effects largely dominate the contributions to the PSF.

Several upcoming surveys[14,15] are planned to image very large portions of the sky, thus allowing us to measure the shape of billions of galaxies. In particular, ESA's Euclid[16] is a planned space-based telescope that makes Weak Lensing analysis one of its main science goals. Due to the sheer amount of galaxies Euclid will observe, the main sources of error in the Weak Lensing signal will be systematic rather than statistical. PSF estimation is in turn one of the main sources of systematic error in such studies.

The use of optimal transport in the context of PSF modelling has already been studied[17,18] to deal with spatial variations. However, because Euclid will observe from space, chromatic variations of the PSF will also need to be precisely accounted for, as opposed to past ground surveys where these variations were very small in comparison to the atmospheric effects. At any given position in the field of view, chromatic variations happen based only on the incoming light's wavelength (see Figure 5), which makes the same sort of approach we used on our toy example in subsection 3.1 appealing. Namely, we use a set of simulated, Euclid-like PSFs at different wavelengths as our training data, and apply dimensionality reduction to learn only $S = 2$ atoms. Since the variations occur based on the incoming light wavelength, we can inject prior information in our method by initializing each PSF's weight as a function of wavelength, by choosing $\forall i, \lambda_i := [t, 1-t]$ where t is the projection of the data's wavelength in $[0, 1]$. Conversely, the atoms are initialized as uniform histograms (*i.e.*, each pixel's value is $1/N$). In Figure 6, we show the reconstructions we obtain for the PSFs of the dataset that were featured in Figure 5.

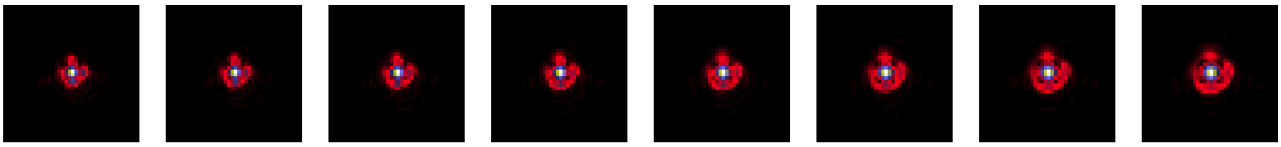

Figure 5. Chromatic variations of Euclid-like PSF between 550 and 900nm.

Despite being initialized without any prior information, the two atoms we learn appear very close to the two PSFs at extreme wavelength, as shown in Figure 7. Much like in the simplistic case of subsection 3.1, this shows that our method captures the fact that the training data is composed of intermediate state in a transformation between two extremes. By comparison, the two first principal components learned by PCA shown in Figure 8 have no relation to actual PSFs.

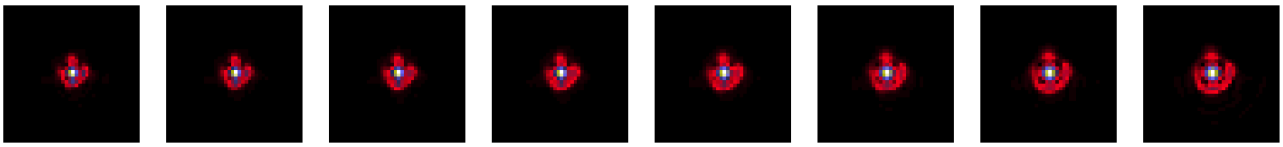

Figure 6. Reconstruction of Euclid-like PSF with our method for incoming wavelengths of 550 to 900nm.

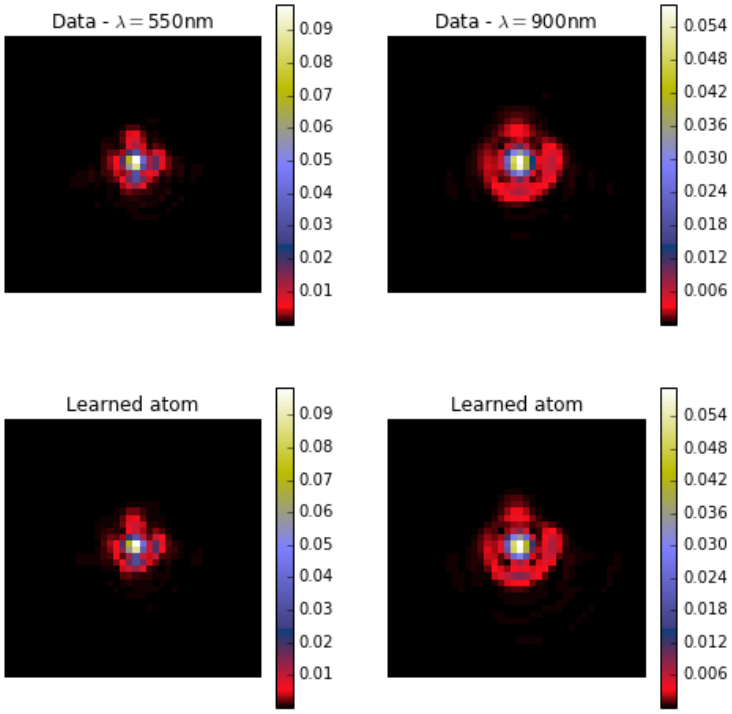

Figure 7. Extreme wavelength PSFs in the dataset and atoms learned by our method.

Similarly, the codes associated to these atoms are not monotonous with regards to the data's wavelength (see right-hand side of Figure 9). While the weights we learn differ from the linear relation we imposed at initialization, they remain very clearly linked to the wavelength, as shown in the leftmost figure of Figure 9.

In this case, unlike the example shown in subsection 3.1, NMF atoms also somewhat resemble the two PSFs at extreme wavelength, as shown in Figure 10. However, our method reaches lower values of the reconstruction error. In this case, we used $\mathcal{L} = 1/2\|.\|_2^2$, and our method reached a value of 1.71×10^{-3}, while NMF converged at 2.62×10^{-3}. Looking at the individual reconstruction errors, the gap between the two methods was particularly pronounced for datapoints lying in the middle of the spectrum, which indicates that the simple stacking of the two extreme wavelength PSFs (similar to what we observed with Euclidean barycenters in Figure 2) fails at capturing the actual warping undergone by the PSF as the wavelength varies.

In practice, we use the stars (as they should be close to points sources in the absence of PSF) present in Euclid's

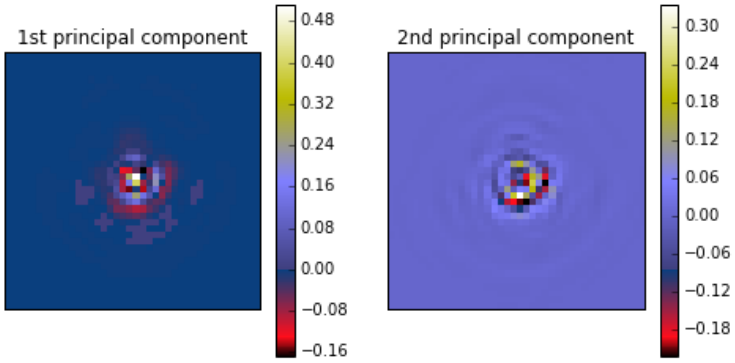

Figure 8. First two principal components from a PCA.

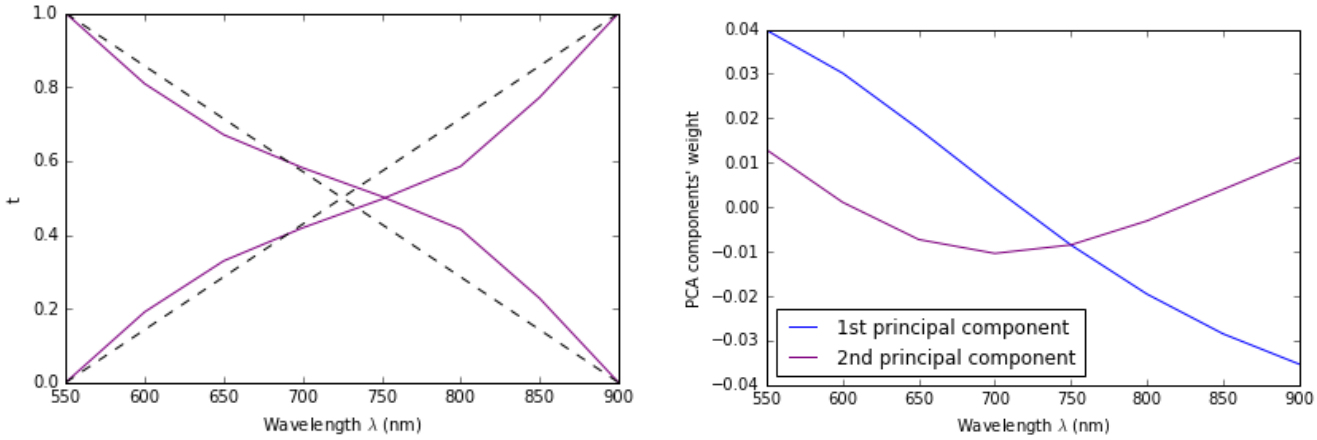

Figure 9. Left-hand side: weights reached by our method at convergence. Right-hand side: PCA-learned codes corresponding to the first two principal components.

field of view as measures of the PSF at this given position and for their precise Spectral Energy Distribution (SED). Ultimately, what matters for Weak Lensing studies is the estimation of the PSF at the position of the galaxies[18] and for this galaxy's SED. This means that from a set of observed PSFs with arbitrary (but known) star SEDs, we need to estimate the monochromatic PSF components so that we can then recombine them using the galaxies' estimated SED. This study shows that linear methods (NMF) do seem to somewhat capture the variation of the PSF - however, in a real-life setting where the dataset consists of a set of stacked polychromatic PSF measurements, one would need to add very strict constraints on the atoms to be learned to force the representation to capture the variation. Because our approach is based on Optimal Transport, this constraint would naturally appear if we were to cast the decomposition of polychromatic PSFs into monochromatic components as an Optimal Transport problem where we modelize the chromatic variations as the displacement interpolation of two extreme PSFs, that is, the set of their Wasserstein barycenters for all possible weights in $[t, 1-t], t \in [0, 1]$.

4. CONCLUSION

This paper introduces a new feature learning method akin to Dictionary Learning. In the latter, however, the relationship between learned atoms and data remains linear, while our method breaks free from this setting by using an Optimal Transport-based formulation. This is made possible by recent developments in numerical Optimal Transport based on the addition of an entropy term and the definition of the Wasserstein barycenter in

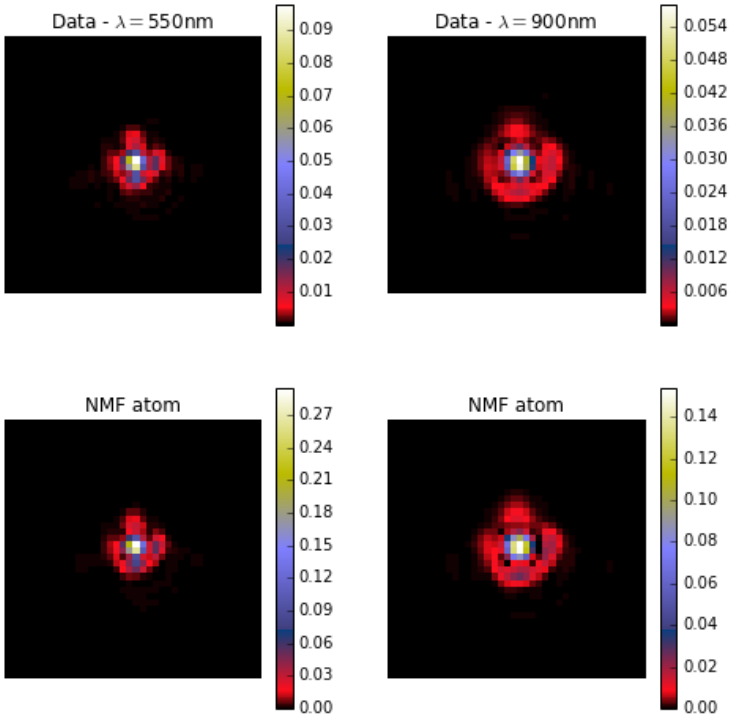

Figure 10. Extreme wavelength PSFs in the dataset and atoms learned by NMF.

analogy with the Euclidean barycenter.

Beyond offering a non-linear Dictionary Learning approach, we show that the Optimal Transport geometry can capture the variations that appear in real life settings - in this case, the chromatic variation of the Euclid visible light instrument.

For illustration purposes, the present work focused on displacement interpolation, that is, the case where we consider the Wasserstein barycenters of two measures. However, this setting is but a particular case of what our method can achieve, and the generalization to more than two atoms is immediate.

ACKNOWLEDGMENTS

M.A.S. is supported by the Centre National d'Etudes Spatiales (CNES). This work was supported by the European Community through the grant DEDALE (contract no. 665044) within the H2020 Framework Program.

REFERENCES

[1] Mallat, S., [*A wavelet tour of signal processing*], Academic press (1999).

[2] Mairal, J., Bach, F., Ponce, J., and Sapiro, G., "Online learning for matrix factorization and sparse coding," *Journal of Machine Learning Research* **11**(Jan), 19–60 (2010).

[3] Villani, C., [*Topics in optimal transportation*], no. 58, American Mathematical Soc. (2003).

[4] Cuturi, M., "Sinkhorn distances: Lightspeed computation of optimal transport," in [*Advances in Neural Information Processing Systems*], 2292–2300 (2013).

[5] Sinkhorn, R., "Diagonal equivalence to matrices with prescribed row and column sums," *The American Mathematical Monthly* **74**(4), 402–405 (1967).

[6] Agueh, M. and Carlier, G., "Barycenters in the wasserstein space," *SIAM Journal on Mathematical Analysis* **43**(2), 904–924 (2011).

[7] Benamou, J.-D., Carlier, G., Cuturi, M., Nenna, L., and Peyré, G., "Iterative bregman projections for regularized transportation problems," *SIAM Journal on Scientific Computing* **37**(2), A1111–A1138 (2015).

[8] Bonneel, N., Peyré, G., and Cuturi, M., "Wasserstein barycentric coordinates: Histogram regression using optimal transport," *ACM Transactions on Graphics (Proceedings of SIGGRAPH 2016)* **35**(4) (2016).

[9] Theano Development Team, "Theano: A Python framework for fast computation of mathematical expressions," *arXiv e-prints* **abs/1605.02688** (May 2016).

[10] Morales, J. L. and Nocedal, J., "Remark on "algorithm 778: L-bfgs-b: Fortran subroutines for large-scale bound constrained optimization"," *ACM Transactions on Mathematical Software (TOMS)* **38**(1), 7 (2011).

[11] Heymans, C., Van Waerbeke, L., Miller, L., Erben, T., Hildebrandt, H., Hoekstra, H., Kitching, T. D., Mellier, Y., Simon, P., Bonnett, C., et al., "Cfhtlens: the canada–france–hawaii telescope lensing survey," *Monthly Notices of the Royal Astronomical Society* **427**(1), 146–166 (2012).

[12] de Jong, J. T., Verdoes Kleijn, G. A., Kuijken, K. H., and Valentijn, E. A., "The kilo-degree survey," *Experimental Astronomy* , 1–20 (2013).

[13] Collaboration, D. E. S. et al., "The dark energy survey," *arXiv preprint astro-ph/0510346* (2005).

[14] Ivezic, Z., Tyson, J., Abel, B., Acosta, E., Allsman, R., AlSayyad, Y., Anderson, S., Andrew, J., Angel, R., Angeli, G., et al., "Lsst: from science drivers to reference design and anticipated data products," *arXiv preprint arXiv:0805.2366* (2008).

[15] Spergel, D., Gehrels, N., Breckinridge, J., Donahue, M., Dressler, A., Gaudi, B., Greene, T., Guyon, O., Hirata, C., Kalirai, J., et al., "Wide-field infrared survey telescope-astrophysics focused telescope assets wfirst-afta final report," *arXiv preprint arXiv:1305.5422* (2013).

[16] Laureijs, R., Amiaux, J., Arduini, S., Augueres, J.-L., Brinchmann, J., Cole, R., Cropper, M., Dabin, C., Duvet, L., Ealet, A., et al., "Euclid definition study report," *arXiv preprint arXiv:1110.3193* (2011).

[17] Irace, Z. and Batatia, H., "Motion-based interpolation to estimate spatially variant psf in positron emission tomography," in [*Signal Processing Conference (EUSIPCO), 2013 Proceedings of the 21st European*], 1–5, IEEE (2013).

[18] Ngolè-Mboula, F. and Starck, J.-L., "Psf field learning based on optimal transport distances," *arXiv preprint arXiv:1703.06066* (2017).