SCARLET: Source separation in multi-band images by Constrained Matrix Factorization

Peter Melchior[a], Fred Moolekamp[a], Maximilian Jerdee[a], Robert Armstrong[a], Ai-Lei Sun[b], James Bosch[a], Robert Lupton[a]

[a]Department of Astrophysical Sciences, Princeton University, Princeton, NJ 08544, USA
[b]Department of Physics & Astronomy, Johns Hopkins University, Bloomberg Center, 3400 N. Charles St., Baltimore, MD 21218, USA

Abstract

We present the source separation framework SCARLET for multi-band images, which is based on a generalization of the Non-negative Matrix Factorization to alternative and several simultaneous constraints. Our approach describes the observed scene as a mixture of components with compact spatial support and uniform spectra over their support. We present the algorithm to perform the matrix factorization and introduce constraints that are useful for optical images of stars and distinct stellar populations in galaxies, in particular symmetry and monotonicity with respect to the source peak position. We also derive the treatment of correlated noise and convolutions with band-dependent point spread functions, rendering our approach applicable to coadded images observed under variable seeing conditions. SCARLET thus yields a PSF-matched photometry measurement with an optimally chosen weight function given by the mean morphology in all available bands. We demonstrate the performance of SCARLET for deblending crowded extragalactic scenes and on an AGN jet – host galaxy separation problem in deep 5-band imaging from the Hyper Suprime-Cam Stategic Survey Program. Using simulations with prominent crowding we show that SCARLET yields superior results to the HSC-SDSS deblender for the recovery of total fluxes, colors, and morphologies. Due to its non-parametric nature, a conceptual limitation of SCARLET is its sensitivity to undetected sources or multiple stellar population within detected sources, but an iterative strategy that adds components at the location of significant residuals appears promising. The code is implemented in Python with C++ extensions and is available at https://github.com/fred3m/scarlet.

Keywords: methods: data analysis, techniques: image processing, galaxies: structure, Non-negative matrix factorization

1. Introduction

Modern astronomical wide-field surveys cover large areas of the sky at ever increasing depths, revealing more objects and low-surface-brightness features of extended objects that were previously too faint to detect. These gains drive investigations into galactic, extragalactic and cosmological phenomena at an unprecedented level of detail and statistical power. On the other hand, because of the enhanced sensitivity, a larger fraction of the observed area is associated with detectable objects, thereby increasing the chance that multiple objects overlap. This so-called "blending" constitutes a major concern for the analysis of existing and upcoming surveys, especially those that observe from the ground.

The majority of methods for measuring the properties of celestial objects assume that every object can be considered isolated. If that assumption holds, well-defined and accurate measurements of the flux, position, shape, and morphology can routinely be made with methods that are either based on moments of the light distribution within some aperture or on parametric fits to the images.

However, the notion of isolated objects is becoming increasingly obsolete. With a limiting magnitude of $i \approx 24$, DES[1]

(Dark Energy Survey Collaboration, 2016) finds that 30% of galaxies suitable for weak-lensing measurements are affected by blending (Samuroff et al., 2018). For HSC[2] (Aihara et al., 2018a), whose Wide survey has a limiting magnitude of $i \approx 26$, Bosch et al. (2018) find that 58% of measured objects are in blended groups, a dramatic increase despite a substantially better average seeing than DES. LSST[3] (Ivezic et al., 2008) expects to reach $i \approx 27$ after 10 years of operations , and it is estimated that 63% of observed galaxies will have Sérsic model photometry that is altered by more than 2% due to the presence neighbors (Sanchez et al., in prep.). Even more problematically, in a comparison study of HST and Subaru imaging of a galaxy cluster field, where the Subaru data had similar seeing and depth so as to serve as a proxy for LSST, Dawson et al. (2016) found that 14% of observed galaxies are blended but not recognized as such in the ground-based images. Slightly larger numbers for unrecognized blends are found for HSC in a study that inserted fake objects into real survey images to infer how many of them could be recovered (Murata et al., in prep.).

This lack of separability between objects necessitates the employment of techniques that analyze entire scenes with overlapping objects. For direct measurements, such as moments within apertures, there is no accurate way to correct for the excess light from overlapping objects because such a correction

---

[1]https://www.darkenergysurvey.org

[2]http://hsc.mtk.nao.ac.jp/ssp/
[3]http://lsst.org

depends e.g. on radial profiles of the objects involved, which cannot be determined well for blends. As a consequence, multiple objects need to be *modeled* either iteratively (by masking all but one) or simultaneously, often requiring sophisticated and fine-tuned schemes to prevent unstable or physically implausible solutions (e.g. Barden et al., 2012; Drlica-Wagner et al., 2018). Any such scheme is suitable to extract and separate the objects in celestial scenes, i.e. to "deblend" those scenes. Differences between the schemes include the propagation of errors, which conceptually favors simultaneous approaches, and whether the desired measurements are generated directly from deblender models or by passing them on to established measurement algorithms for isolated objects.

Traditional deblending approaches in astronomy are achromatic, i.e. they employ information from only a single image. SExtractor detects blending by thresholding an image at a range of intensity levels and searching for sets of pixels that are connected at a lower threshold but split into several connected regions at a higher one (Bertin & Arnouts, 1996). As a consequence of the splitting approach, the association of pixels to objects is unique and exclusive, i.e. in the internal representation of blended objects they do not overlap. This unrealistic notion has necessitated mitigation strategies or fine-tuning to prevent "over-deblending" of larger galaxies caused by smaller and fainter companions or interlopers (e.g. Rix et al., 2004).

The deblender in the SDSS *Photo* pipeline (Lupton, in prep.) does allow for overlap between nearby objects and, consequently, needs to estimate the portion of any pixel's flux that is due to each object. It uses a two-step approach, in which first a template is constructed for each object based on the requirement that pixel values symmetrically across the object's peak pixel be identical; they are generally not, so the minimum pixel value of those two pixels is adopted for both. Then, the original image values are projected onto those templates, associating each pixel's flux to different objects in proportion to the amplitude of the respective templates. Despite very few assumptions, the method mostly separates sources into physically plausible objects but struggles with situations where a central object is symmetrically surrounded by neighbors, for instance a blend with three peaks in a row (Lupton, in prep.).

More recently, Zhang et al. (2015) and Connor et al. (2017) proposed variants of inpainting techniques, where the relevant pixels of blended objects are replaced by an estimate of a local variable background, working inwards from an initially defined outline. The portion of the pixel flux above the background estimate is attributed to the respective object. These approaches implicitly account for blending by assuming that the background captures the flux contributions from neighboring objects. This is particularly useful when recovering small objects in multi-scale blending situations like the cores of galaxy clusters and removing them from the scene so that large objects can be measured separately.

While effective in many cases, all of the deblending schemes outlined above employ heuristic arguments for how to separate overlapping sources. They also perform the pixel–object association sequentially, one object at a time, thus losing the advantages of a simultaneous solution, for instance the ability to

explore the degeneracies that arise because the objects are not isolated. However, it is our opinion that the biggest limitation stems from the restriction to a single image, and therefore a single filter band, while most modern surveys observe the sky in several filters. A visual inspection of multi-band images clearly suggests that color can serve as a powerful discriminator between different objects, even with severe overlap (see Figure 1).

MuSCADeT (Joseph et al., 2016) addresses both limitations by building a joint model of multi-band image data. As their model is non-parametric, the number of degrees of freedom is large, which leads to many possible degeneracies in the solutions, so they demand that the spatial distribution of each source be sparse in the starlet (a form of isotropic wavelet) domain. The resulting solutions extract preferentially compact features down to the noise level, using a set of previously identified colors for each feature. For applications to wide-field multi-band data, we cannot generally assume to a priori identify the color of an object in the scene because there might not be a single pixel whose color is uncontaminated by other objects. We therefore seek the ability to update both spectral and morphological characteristics of the objects. While sharing noticeable similarities with MuSCADeT regarding the use of a non-parametric constrained morphological model, one can consider our approach an extension that also updates the source spectra as well as a generalization that allows an arbitrary number of constraints to be placed on each source.

The outline of the paper is as follows: We introduce our approach, dubbed SCARLET, in Section 2, demonstrate its performance on real data and simulations in Section 3, and conclude in Section 4.

2. Methodology

We base our deblender on the assumption that astronomical scenes are superpositions of multiple components, each with

1. a spatially compact support and

2. a constant spectrum over that support.

For stellar fields this is obviously true, but even for galaxies, especially marginally resolved ones, which constitute the vast majority of galaxies in deep surveys, the assumption is appropriate. In addition, even large, extended galaxies can be thought of as conglomerations of components (e.g. bulges, discs, bars, star forming regions) for which the assumptions above hold at least approximately.[4] For instance, the popular bulge-disc decomposition for galaxies (e.g. MacArthur et al., 2003) is justified by this interpretation. We note that the assumption of linear superposition implies that components do not interact, which is correct only for transparent emitters. Absorption, e.g. by dust in the galaxy, can be approximated by allowing negative values in the source spectrum, but substantial opacities cannot fully be

---

[4] In the literature the terms "source separation" and "component separation" are often used interchangeably. To better reflect the hierarchical nature of astrophysical scenes, we will define the term "Source" as a collection of co-centered components that belong to the same astrophysical object.
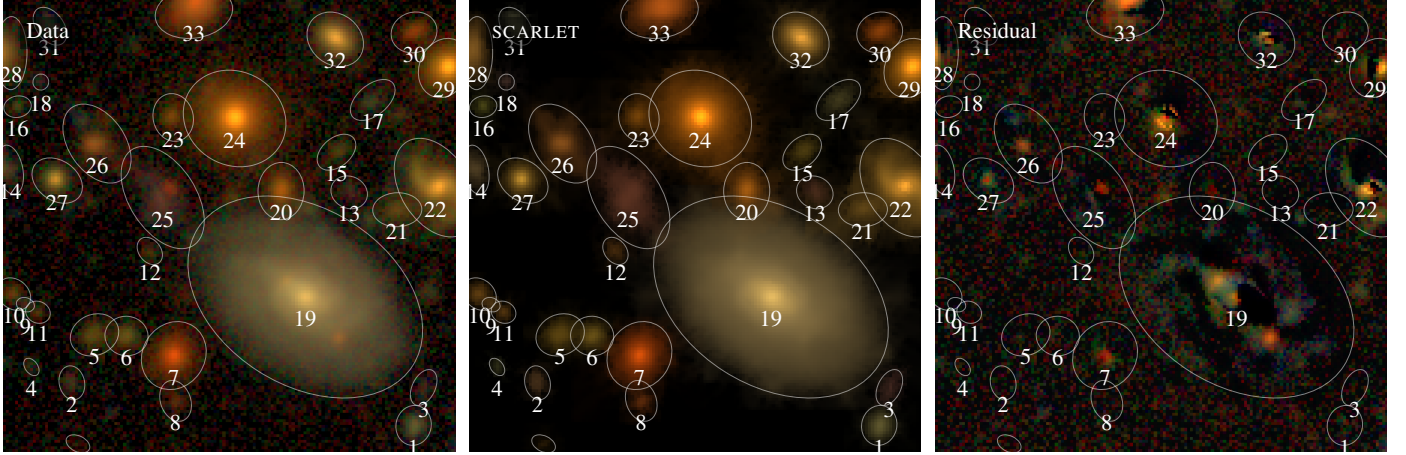
Figure 1: *Left:* False-color image of *grizy* coadd images Y from the HSC UltraDeep COSMOS data release, shown with an arcsinh stretch. The scene spans 25×25 arcsec². Object detections and ellipse fits are performed by SExtractor on the detection image (sum of the five coadds). *Center:* scarlet model AS of the scene with single-component sources for each detection. *Right:* The residual Y − AS reveals the presence of additional sources or color variations within detected sources.

modeled because the effect depends on the amount of absorbing material *and* the intensity of the background radiation.

The assumptions above appear to lend themselves to a parametric modeling framework, where one assumes to know the shapes of the components and potentially their intrinsic spectrum, exploiting quite tight relations between colors and morphologies exhibited by galaxies in the late universe (e.g. Conselice, 1997; Ball et al., 2008). While drastically reducing the number of optimization parameters, we are critical of this approach for two reasons: First, in the translation of an intrinsic, restframe spectrum to the observed broadband colors one needs to take the galaxy's redshift into account, which is equivalent to estimating a photometric redshift as part of the deblending process. If the redshifting prescription is incorrect, e.g. because of a limited library of spectra or the evolution of those spectra with redshift, it would affect the properties of the deblended components—not only their recovered spectra but also their shapes. Second, at the stage in the analysis pipelines of large astronomical surveys where we envision the deblender to operate, it will not necessarily be established what kind of sources are in the scene; in other words, a suitable parameterization is probably not known. This is most evident when looking at the star-galaxy distinction: for stars, three parameters are sufficient (two centroid coordinates and one amplitude), while even simple galaxy models need at least one more parameter (the size). Model-fitting under those conditions can be done by transdimensional sampling (e.g. Green, 1995), but the computational costs are likely too high for large-volume data sets.

Because of these concerns, we seek to characterize the scenes without making questionable astrophysical assumptions, which means describing colors in the observed frame and morphologies in the free-form space of image pixels.

2.1. Non-negative Matrix Factorization

We assume that an astronomical scene Y that we seek to analyze is organized in the form of a multi-band image cube of aligned images in B bands, each of which is suitably flattened

to have a total number of N pixels. Our previous assumptions give rise to a multi-band model M as a sum of a finite number of components K,

$$ \mathsf{M} = \sum_{k=1}^{K} \mathsf{A}_k^\top \times \mathsf{S}_k = \mathsf{AS}, \tag{1} $$

where $\mathsf{A}_k \in \mathbb{R}^B$ is the amplitude of component k across all bands, i.e. its spectral energy distribution (SED), and $\mathsf{S}_k \in \mathbb{R}^N$ is the spatial shape of that component. By arranging the A_k as columns of $\mathsf{A} \in \mathbb{R}^{B \times K}$ and S_k as the rows of $\mathsf{S} \in \mathbb{R}^{K \times N}$, we have a model M that is given by the product of two matrix factors.

With a homoscedastic Gaussian error model, which is appropriate for most extragalactic images in the optical, the likelihood function is

$$ f(\mathsf{A}, \mathsf{S}) = \frac{1}{2} \|\mathsf{Y} - \mathsf{AS}\|_2^2 \tag{2} $$

where $\|.\|_2$ denotes the element-wise L_2 (Frobenius) norm.[5] In its simplest form, the Non-negative Matrix Factorization (NMF Paatero & Tapper, 1994) then amounts to fitting A and S such that they minimize f and obey the non-negativity constraint, which is given by the indicator function of the set of non-negative matrix elements:

$$ g_+(\mathsf{X}) = \begin{cases} 0 & \text{if } \mathsf{X}_{mn} \geq 0 \ \forall m, n \\ \infty & \text{else.} \end{cases} \tag{3} $$

In other words, one seeks to minimize $f(\mathsf{A}, \mathsf{S}) + g_+(\mathsf{A}) + g_+(\mathsf{S})$. The classical way of solving this NMF problem is known as "multiplicative updates" (Lee & Seung, 2001), which suffers from poor convergence if the constraints strongly work against

---

[5] The objective function in Equation 2 is insufficient for dealing with realistic multi-band data, which generally exhibit correlated noise and correlated signals, the former from warping the images to rectify any astrometric distortion, the latter because of the blurring from the point spread function. We will work out in Section 2.4 and Section 2.5 how to deal with both effects.

the minimum of the objective function. Moreover, in its simplest form the NMF is hampered by a degeneracy that stems from the transformation $(A, S) \rightarrow (AQ, Q^{-1}S)$ with an arbitrary invertible matrix Q, which means power can be shifted from A to S or vice versa, hampering convergence to a well-defined solution. To compensate one can introduce another constraint on the norms of either A or S. The most natural normalization for astronomical scenes is to require that each component's SED sums up to unity. We therefore seek to minimize $f(A, S) + g_+(A) + g_{unity}(A) + g_+(S)$, where g_{unity} is another indicator function of the set of matrices in which the columns are properly normalized.

We do so by using a two-block variant of the "proximal gradient method". In short, a closed smooth proper convex function $f(\mathbf{x})$ under a convex constraint function $g(\mathbf{x})$ with $\mathbf{x} \in \mathbb{R}^n$ can be minimized with the following sequence:

$$\mathbf{x}^{it+1} \leftarrow \text{prox}_{\lambda^{it} g}\left(\mathbf{x}^{it} - \lambda^{it} \nabla f(\mathbf{x}^{it})\right). \tag{4}$$

The step size λ must be chosen from $(0, 2/L]$, where L is the Lipschitz constant of ∇f. The function g is not directly evaluated but accessed through its *proximal operator*,

$$\text{prox}_{\lambda g}(\mathbf{x}) \equiv \underset{\mathbf{v}}{\text{argmin}}\left\{g(\mathbf{v}) + \frac{1}{2\lambda}\|\mathbf{x} - \mathbf{v}\|_2^2\right\}. \tag{5}$$

The proximal operator of g effectively hides the infinite values and the non-differentiability of its underlying function. For instance, the proximal operator of the indicator function I_C of a non-empty set C is the Euclidean projection operator onto the set—i.e. it yields $\mathbf{v} \in C$ that is nearest to $\mathbf{x}$—and the scaling parameter λ is irrelevant. In particular,

$$\text{prox}_+(\mathbf{x}) = (x_1^+, \ldots, x_n^+) \text{ with } x_i^+ \equiv \max(0, x_i)$$
$$\text{prox}_{unity}(\mathbf{x}) = \|\mathbf{x}\|_1^{-1} \mathbf{x}, \tag{6}$$

For many other classes of constraint functions proximal operators have analytic forms, which means that the minimization of Equation 5 does not have to be carried out. For more details on proximal methods, we refer the reader to Combettes & Pesquet (2011) and Parikh et al. (2014).

As long as there is only one constraint per variable, we can interleave update steps of the form of Equation 4 that minimize $f(A|S)$ and those that minimize $f(S|A)$, effectively a proximal implementation of a coordinate descent method, as shown in Algorithm 1. Note that we merged $g_+(A) + g_{unity}(A)$ into one proximal operator $\text{prox}_{unity+} \equiv \text{prox}_+\left(\text{prox}_{unity}\right)$, which can be done because the two constraints commute, i.e. they do not interfere. This reduces the problem to calculating the gradients

$$\nabla_A f(A, S) = -(Y - AS)S^\top$$
$$\nabla_S f(A, S) = -A^\top(Y - AS) \tag{7}$$

and their Lipschitz constants

$$L_A = \left\|SS^\top\right\|_s \text{ and } L_S = \left\|A^\top A\right\|_s, \tag{8}$$

where $\|.\|_s$ refers to the spectral norm. The sequence in Algorithm 1 approaches a fixed point that is also a local minimum of

---

Algorithm 1 Proximal Gradient NMF

Two matrix factors A and S are independently updated with a sequence of gradient steps followed by projections to obey the constraints.

1: **procedure** NMF-PGM(A^0, S^0)
2:      **for** it = 0, 1, ... **do**
3:          $\lambda_A^{it} \leftarrow \left\|S^{it}S^{it\top}\right\|_s^{-1}$
4:          $A^{it+1} \leftarrow \text{prox}_{unity+}\left(A^{it} - \lambda_A^{it}\nabla_A f(A^{it}, S^{it})\right)$
5:          $\lambda_S^{it} \leftarrow \left\|A^{it+1\top}A^{it+1}\right\|_s^{-1}$
6:          $S^{it+1} \leftarrow \text{prox}_+\left(S^{it} - \lambda_S^{it}\nabla_S f(A^{it+1}, S^{it})\right)$

---

f and can therefore be terminated whenever the change between successive iterations falls below a user-defined threshold.

This approach is similar to the one of Rapin et al. (2013) in that the minimization is performed in two blocks and that constraints are expressed by their proximal operators, but we only perform one gradient step instead of seeking the exact solution for every block in *every iteration*. This reduces the number of gradient and constraint evaluations while still yielding a sequence that is guaranteed to converge (Lin, 2007; Xu & Yin, 2013).

2.2. Constrained Matrix Factorization

For many practical applications we want to be able to impose additional constraints besides (or instead of) non-negativity. We will give a list of alternative constraints in Section 2.3, but it is useful to think of constraints more generally as logarithms of priors on the solution. In addition, we are often forced to impose multiple simultaneous constraints to deal with the large number of unknowns in Equation 1, especially when the number of components K is larger than the number of bands B. To reflect these more general conditions, we favor the term "Constrained Matrix Factorization" (CMF) instead of NMF.

With the direct projection method of Equation 4, we can only accommodate any one proximal operator per variable, even though non-interacting constraints can be daisy-chained to act like a single one, as we have done with prox_{unity+}. One could generalize this approach by the method of Alternating Projections, also known as Projection onto Convex Sets (Bauschke & Borwein, 1996), but that still requires that the proximal operators act directly on the variables A and S. Often, it is beneficial to express constraints in a transformed domain, for instance demand that the solution be sparse in that domain (e.g. Rapin et al., 2014). We therefore prefer the Alternating Direction Method of Multipliers (ADMM, Gabay & Mercier, 1976; Glowinski & Marroco, 1975; Eckstein & Bertsekas, 1992; Boyd et al., 2011). In its basic form it seeks to

$$\underset{\mathbf{x}}{\text{minimize }} f(\mathbf{x}) + g(L\mathbf{x}) \tag{9}$$

which allows us to introduce an arbitrary matrix L. For this work L will be a gradient or symmetry operator, but it could as well be a Fourier or wavelet transform, etc. The previous equation can be re-written in "consensus form"

$$\begin{aligned}\text{minimize} \quad & f(\mathbf{x}) + g(\mathbf{z}) \\ \text{subject to} \quad & L\mathbf{x} - \mathbf{z} = 0,\end{aligned} \tag{10}$$

which suggests an approach that splits the optimization into two separate tasks: one that minimizes f and another that satisfies g by introducing the auxiliary variable $\mathbf{z}$ (Douglas & Rachford, 1956). The updating scheme also introduces the dual variable $\mathbf{u}$, which connects the two tasks, creating the following sequence:

$$
\begin{aligned}
\mathbf{x}^{\text{it}+1} &\leftarrow \text{prox}_{\lambda f}\left(\mathbf{x}^{\text{it}} - \tfrac{\lambda}{\rho}\mathsf{L}^\top\left(\mathsf{L}\mathbf{x}^{\text{it}} - \mathbf{z}^{\text{it}} + \mathbf{u}^{\text{it}}\right)\right) \\
\mathbf{z}^{\text{it}+1} &\leftarrow \text{prox}_{\rho g}\left(\mathsf{L}\mathbf{x}^{\text{it}+1} + \mathbf{u}^{\text{it}}\right) \\
\mathbf{u}^{\text{it}+1} &\leftarrow \mathbf{u}^{\text{it}} + \mathsf{L}\mathbf{x}^{\text{it}+1} - \mathbf{z}^{\text{it}+1},
\end{aligned}
\tag{11}
$$

where the scaling parameters need to satisfy $0 < \lambda \le \rho/\|\mathsf{L}\|_s^2$.[6] Two aspects of this sequence are remarkable. The function f is only accessed through its proximal operator, which means it does not need to be differentiable. In our case, f is differentiable, and we can consider

$$
\text{prox}_{\lambda f}(\mathbf{x}) \equiv \mathbf{x} - \lambda \nabla f(\mathbf{x})
\tag{12}
$$

the corresponding proximal operator of the first-order approximation to f.[7] Second, the linear operator L does not need to be invertible; it and its transpose are only used for dot products, rendering the updates particularly efficient with sparse matrices.

In Moolekamp & Melchior (2018), we have demonstrated that the ADMM can be extended to multiple variables of a function f that is convex in each of those variables by constructing a block-updating sequence analogous to Algorithm 1. For the CMF, A and S (which we will call X_1 and X_2 and enumerate with the index j for the remainder of this subsection) thus have their own set of ADMM auxiliary variables $\mathsf{Z}_1, \mathsf{U}_1, \mathsf{Z}_2, \mathsf{U}_2$. In addition, we have introduced a method that allows each X_j to be subject to multiple simultaneous constraints, one pair $(\mathsf{Z}_{ji}, \mathsf{U}_{ji})$ for each constraint $g_{ji}(\mathsf{X}_j)$ with $i \in \{1, \ldots, M_j\}$. Combined, our new method, dubbed SCARLET, provides the flexibility we seek for deblending astronomical scenes via the CMF. It will

$$
\left\{
\begin{aligned}
&\underset{\mathsf{X}_1, \mathsf{X}_2}{\text{minimize}} \quad f(\mathsf{X}_1, \mathsf{X}_2) + \sum_{j=1}^{2} \sum_{i=1}^{M_j} g_{ji}(\mathsf{Z}_{ji}) \\
&\text{subject to} \quad \mathsf{L}_{ji}\mathsf{X}_j - \mathsf{Z}_{ji} = 0 \ \forall j \in \{1, 2\} \text{ and } i \in \{1, \ldots, M_j\}
\end{aligned}
\right.
\tag{13}
$$

through the Block-Simultaneous Method of Multipliers (bS-DMM Moolekamp & Melchior, 2018).

The core working principle of SCARLET is given in Algorithm 2. The ADMM updates of Equation 11 are performed in lines 10–13, using the gradient updates from Equation 7 in the form of Equation 12 in line 10. The step sizes λ_j in line 8 are given as before by $1/L_j$ with the Lipschitz constants from Equation 8. It is worthwhile pointing out that in the limit of all constraints being satisfied, $\mathsf{Z} = \mathsf{X}$ and $\mathsf{U} = 0$, in which case Algorithm 2 is equivalent to Algorithm 1.

Following Boyd et al. (2011), we implement stopping criteria (on line 14) based on primal residual $\mathsf{P}^{\text{it}+1} = \mathsf{L}\mathsf{X}^{\text{it}+1} - \mathsf{Z}^{\text{it}+1}$ and the dual residual $\mathsf{D}^{\text{it}+1} = \tfrac{1}{\rho}\mathsf{L}^\top\left(\mathsf{Z}^{\text{it}+1} - \mathsf{Z}^{\text{it}}\right)$. These residuals

---

[6] For the $\mathbf{x}$-update, we use a linearized form of the ADMM, which is known as *split inexact Uzawa method* (e.g. Esser et al., 2010; Parikh et al., 2014).

[7] In fact, we can treat the entire projected gradient step of Equation 4 as a first-order approximation of $f + g$, which means we can satisfy a single constraint in the direct domain, such as non-negativity, without having to invoke the auxiliary variables $\mathbf{z}$ and $\mathbf{u}$.

Algorithm 2 Block-SDMM Constrained Matrix Factorization

The matrix factors A and S are treated as two blocks X_j with $j \in \{1, 2\}$ and independently updated with a sequence of gradient steps $\text{prox}_{\lambda_j f, j}$ and jointly optimized with an arbitrary number of constraint functions g_{ji} with $i \in \{1, \ldots, M_j\}$. All constraints are expressed as proximal operators $\text{prox}_{\rho_{ji} g_{ji}}$ and include transformation matrices L_{ji} (cf. Equation 11 for the generic ADMM update sequence). The parameters λ_j and ρ_{ji} are the maximum allowable step sizes in this scheme.

1: **procedure** CMF-BSDMM($\mathsf{A}^0, \mathsf{S}^0$)
2:      $\mathsf{X}_1^0, \mathsf{X}_2^0 \leftarrow \mathsf{A}^0, \mathsf{S}^0$
3:      **for** $j = 1, 2$ **do**
4:          $\mathsf{Z}_{ji}^0 \leftarrow \mathsf{L}_{ji}\mathsf{X}_j \ \forall i \in \{1, \ldots, M_j\}$
5:          $\mathsf{U}_{ji}^0 \leftarrow \mathbf{0} \ \forall i \in \{1, \ldots, M_j\}$
6:      **for** it $= 0, 1, \ldots$ **do**
7:          **for** $j = 1, 2$ **do**
8:              $\lambda_j^{\text{it}} \leftarrow j == 1?\left\|\mathsf{X}_2^{\text{it}}\mathsf{X}_2^{\text{it}\top}\right\|_s^{-1} : \left\|\mathsf{X}_1^{\text{it}+1\top}\mathsf{X}_1^{\text{it}+1}\right\|_s^{-1}$
9:              $\rho_{ji}^{\text{it}} \leftarrow 2M_j \lambda_j^{\text{it}+1}\left\|\mathsf{L}_{ji}\right\|_2^2$
10:             $\mathsf{X}_j^{\text{it}+1} \leftarrow \text{prox}_{\lambda_j f, j}\left(\mathsf{X}_j^{\text{it}} - \sum_{i=1}^{M_j} \tfrac{\lambda_j}{\rho_{ji}}\mathsf{L}_{ji}^\top\left(\mathsf{L}_{ji}\mathsf{X}_j^{\text{it}} - \mathsf{Z}_{ji}^{\text{it}} + \mathsf{U}_{ji}^{\text{it}}\right)\right)$
11:             **for** $i = 1, \ldots, M_j$ **do**
12:                 $\mathsf{Z}_{ji}^{\text{it}+1} \leftarrow \text{prox}_{\rho_{ji} g_{ji}}\left(\mathsf{L}_{ji}\mathsf{X}_j^{\text{it}+1} + \mathsf{U}_{ji}^{\text{it}}\right)$
13:                 $\mathsf{U}_{ji}^{\text{it}+1} \leftarrow \mathsf{U}_{ji}^{\text{it}} + \mathsf{L}_{ji}\mathsf{X}_j^{\text{it}+1} - \mathsf{Z}_{ji}^{\text{it}+1}$
14:      **if** $\bigwedge_{ji}\left\{\left\|\mathsf{P}_{ji}^{\text{it}+1}\right\|_2 \le \epsilon^{\text{pri}} \wedge \left\|\mathsf{D}_{ji}^{\text{it}+1}\right\|_2 \le \epsilon^{\text{dual}}\right\}$ **then** break

---

describe how close the current X is to satisfying the constraint in the transformed domain and how much Z changes, respectively. To assess primal and dual feasibility, we require

$$
\begin{aligned}
\left\|\mathsf{P}^{\text{it}+1}\right\|_2 &\le \epsilon^{\text{pri}} \equiv \sqrt{p}\,\epsilon^{\text{abs}} + \epsilon^{\text{rel}} \max\left\{\left\|\mathsf{L}\mathsf{X}^{\text{it}+1}\right\|_2, \left\|\mathsf{Z}^{\text{it}+1}\right\|_2\right\} \text{ and} \\
\left\|\mathsf{D}^{\text{it}+1}\right\|_2 &\le \epsilon^{\text{dual}} \equiv \sqrt{n}\,\epsilon^{\text{abs}} + \epsilon^{\text{rel}}/\rho \left\|\mathsf{L}^\top\mathsf{U}^{\text{it}+1}\right\|_2,
\end{aligned}
\tag{14}
$$

where p and n are the number of elements in Z and X, respectively. The error thresholds ϵ^{abs} and ϵ^{rel} can be set at suitable values, depending on the precision and runtime constraints of the application.

The procedure outlined in Algorithm 2 yields point estimates of the matrix factors. Uncertainties on each factor are currently calculated by linear error propagation based on Equation 2. While those uncertainties provide estimates of the significance of each element of the matrix factors, they do not properly capture the interdependence of the model components or the effects of the constraints on the degeneracies between them. Fully accurate uncertainties are difficult to estimate because of the non-differentiability of the constraint functions. We plan to investigate ways to better characterize the uncertainties of the SCARLET model in forthcoming work.

This concludes our presentation of the general approach we have adopted in SCARLET. A reader who is mostly interested in the astrophysical applications can safely jump to Section 3 and skip details of the formalism below.

2.3. Constraints

The conventional NMF is capable of separating components, but only from the color contrast between objects in the scene. If the color contrast vanishes or the number K of overlapping components gets large, the model in Equation 1 becomes increasingly susceptible to degeneracies, where the mixture model yields a fair representation of the scene but many of the individual components look implausible.

We thus prefer the flexibility of the CMF to impose additional priors or constraints based on existing knowledge or intuition about how SEDs and morphologies are distributed. In addition to non-negativity, which is justified for astronomical sources that emit photons, we will routinely employ symmetry, monotonicity, and sparsity constraints for the galaxy morphologies (see Section 3.2). We currently do not utilize priors on galaxy or stellar SEDs, but we will demonstrate in Section 3.3 how color constraints can help in cases when additional stability of the source extraction is desired.

We want to emphasize that SCARLET is capable of incorporating any prior or constraint that can be expressed as a proximal operator. We therefore think of it as a general-purpose component separation framework that can be optimized for a wide variety of applications.

2.3.1. Symmetry

We take guidance from the SDSS *Photo* deblender (Lupton, in prep.), which, inspired by Doi et al. (1995), constructs templates of the sources that are symmetric under 180° rotations around the peak pixel. We formalize this requirement by stating that

$$\forall k, i : S_{k,i} - S_{k,j} = 0, \tag{15}$$

where i and j are the indices of the symmetric partner pixels,[8] which can be expressed in matrix form,

$$\forall k : L_{\text{symm}} S_k = \mathbf{0}, \tag{16}$$

with a symmetry matrix $L_{\text{symm}} \in \mathbb{R}^{(N/2-1) \times N}$ that has the elements $(1, -1)$ for each of $N/2 - 1$ suitable pixel pairs (the peak pixel is excluded). In this linearly transformed space, the constraint is imposed by the projection operator onto the zero vector, i.e. $\text{prox}_0(L_{\text{symm}} S_k)$. The combination of the transformation matrix and the proximal operator is required for any ADMM-style constraint in Algorithm 2.

Compared to the SDSS *Photo* deblender, which adopts the minimum value for both pixels in the symmetric pair, this operator will converge to the mean of both pixels. This is a direct consequence of the proximal operator being the Euclidean projection operator and the mean statistic having the least squared error.

2.3.2. Monotonicity

The prominent failure mode of the SDSS *Photo* deblender arises from its templates not necessarily declining monotonically from the peak pixel.[9] While resolved images of real galaxies may in detail show non-monotonic behavior, for instance due to star-forming regions in the outer parts of spiral galaxies, we prefer imposing this constraint for the bulk source shape. An area that violates this constraint likely constitutes a different stellar population, potentially with a color different from the bulk model, and should thus be fit by adding another (monotonic) component to the mixture model.

Monotonicity for two-dimensional features is not uniquely defined. We use a form of radial monotonicity, i.e.

$$\forall k, i : \sum_{j \in \mathcal{U}(i)} w_{ji} S_{k,j} - S_{k,i} \geq 0, \tag{17}$$

where $\mathcal{U}(i)$ is the set of neighbor pixels of i that are closer to the peak than i, and the weights are normalized such that $\forall i : \sum_j w_{ji} = 1$. We implement two variants:

1. The neighborhood contains only the nearest neighbor of i in the direction of the peak, $\mathcal{U}(i) = \{NN(i)\}$ and $w_{NN(i),i} = 1$.

2. The neighborhood contains all three neighbor pixels that are closer to the peak, and their weights are given by $w_{ji} = \cos(\phi_j - \phi_i)$, where ϕ_i denotes the angle from the center of pixel i to the center of the peak pixel.

Unsurprisingly, we find the latter variant to yield models that are somewhat softened in azimuthal direction, and they can violate monotonicity along the straight path from pixel i to the peak, which variant 1 would obey.

As before, we rewrite the constraint in matrix form, $\forall k : L_{\text{mono}} S_k \geq 0$, which can be enforced by the non-negativity constraint from Equation 6 in the transformed domain, $\text{prox}_+(L_{\text{mono}} S)$.

This enforcement of monotonicity can be slow to converge in heavily blended regions. The SCARLET package therefore contains a projection operator in the direct domain, $\text{prox}_{\text{mono}}$, which constructs a monotonic version of its argument. While this construction will in general not yield the monotonic solution that is closest to the argument (the operator is thus in fact not a true proximal operator), we have found it to be substantially more robust and efficient in practice.

2.3.3. Sparsity

Sparsity refers to a solution that can be represented with a small number or a low amplitude of coefficients in a particular domain. In the direct, i.e. the pixel, domain, sparsity constraints thus induce solutions with few non-zero pixels and/or compressed pixel amplitudes, which can be useful to reduce the impact of pixel noise. They are introduced by minimizing the objective function and a penalty function based on the norm of the solution, often the ℓ_2 or ℓ_1 norms or the ℓ_0 pseudo-norm, or

---

[8] In order to always have a well-defined partner for each pixel (other than the peak pixel), the peak pixel needs to be in the center of the image, and the number of pixels in the image must be odd in both directions. We achieve this by defining a virtual box around each source that contains all of its pixels, with the peak pixel at the box center.

[9] To address this weakness, the implementation of the SDSS *Photo* deblender in the current HSC software stack imposes a monotonicity constraint on previously smoothed templates.

a combination of them. While the ℓ_2 norm is differentiable, the other two are not. However, both ℓ_0 and ℓ_1 have simple proximal operators, known as "hard thresholding" and "soft thresholding" operators, respectively:

$$\text{prox}_{\lambda \ell_0}(\mathbf{x}) = (x_1^h, \ldots, x_n^h) \text{ with } x_i^h = \begin{cases} 0 & \text{if } |x_i| < \lambda \\ x_i & \text{else} \end{cases}$$

$$\text{prox}_{\lambda \ell_1}(\mathbf{x}) = (x_1^s, \ldots, x_n^s) \text{ with } x_i^s = \text{sign}(x_i) \max\left(0, |x_i| - \lambda\right)$$

(18)

This is the first example of a proximal operator of a function other than an indicator of a closed set. The parameter λ then governs the strength of the penalty compared to the objective function f.

Because SCARLET allows arbitrary transformations of A and S with linear operators L, the sparsity constraints can be applied in the direct domain and/or in the transformed domain.

2.3.4. Other options

We do not attempt to provide a comprehensive list of possible constraints that are expressible by proximal operators. Instead, we discuss several that we found useful and refer the reader to the appendix of Combettes & Pesquet (2011) and section 6 of Parikh et al. (2014) for more details.

Spatially flat components, e.g. for fitting a uniform sky emission, can be realized by projecting onto the mean (see discussion in Section 2.3.1): $\text{prox}_{\text{flat}}(\mathbf{x}) = \frac{1}{n} \sum_i^n x_i \mathbf{1}_n$ with $\mathbf{x} \in \mathbb{R}^n$.

To allow some amount of spatial variability, e.g. for describing Intra-Cluster Light, we find the Maximum Entropy regularization (Frieden, 1972) useful, which fundamentally seeks to find the flattest distribution that is compatible with the data by adding the (negative) Shannon entropy term

$$S = \sum_i^N p_i \log(p_i) \text{ with } p_i = x_i \Big/ \sum x_i \qquad (19)$$

to the objective function f. The proximal operator of S is

$$\text{prox}_{\lambda S}(\mathbf{x}) = \lambda \mathfrak{R}\left[W\left(\frac{1}{\lambda} \exp\left(\frac{\mathbf{x}}{\lambda} - 1\right)\right)\right], \qquad (20)$$

where W denotes the Lambert-W function.

By choosing the matrix L to represent a basis transform, one can fundamentally alter the characteristics of the solution one wants to isolate or penalize. A few examples with applications in data analysis are the Fourier transform (Hassanieh et al., 2012), the wavelet transform (Rapin et al., 2014), or the shapelet transform (Refregier, 2003). In the transformed domain, simple projection operators can then e.g. null undesired frequencies.

Finally, instead of hard constraints one can also introduce prior distributions on the solutions. If the prior distribution $p(\mathbf{x})$ is smooth, the corresponding proximal operator is given by the gradient step of $\log(p)$ analogous to Equation 12; if it has hard edges it can be represented by wrapping the gradient step in a proximal gradient update as in Equation 4.

2.4. Heteroscedastic and correlated errors

Equation 2 does not account for heteroscedastic errors. Blanton & Roweis (2007); Zhu (2016) introduced a simple extensions to heteroscedastic errors with the matrix V of variances of each pixel in Y:

$$f(\mathsf{A}, \mathsf{S}) = \left\| V^{-\frac{1}{2}} \circ (\mathsf{A} \cdot \mathsf{S} - \mathsf{D}) \right\|_2^2, \qquad (21)$$

where $\circ$ denotes the element-wise, or Hadamard, product. It is straightforward to adjust the gradients in Equation 7 but it is not clear how their Lipschitz constants differ from Equation 8.[10] In addition, Equation 21 does not account for correlations in the noise that arise when images are astrometrically rectified.

For that purpose it is necessary to serialize Y, A, and S into vectors $(\mathbf{Y}, \mathbf{A}, \mathbf{S})$ with lengths $B \cdot N$, $B \cdot K$ and $K \cdot N$, respectively. The most convenient vector form $\mathbf{Y} = (Y_1, \ldots, Y_B)$ simply concatenates the images in B bands. The covariance matrix of $\mathbf{Y}$ has block-diagonal form

$$\Sigma_{\mathbf{Y}} = \begin{pmatrix} \Sigma_1 & 0 & \ldots & 0 \\ 0 & \Sigma_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \ldots & \Sigma_B \end{pmatrix}, \qquad (22)$$

with $\Sigma_b \in \mathbb{R}^{N \times N}$ denoting the pixel covariance matrix in band b. We can now rewrite the objective function in a form that uses vectors for Y and one of A or S, with the other one expressed by a sparse block matrix ($\bar{\mathsf{A}}$ or $\bar{\mathsf{S}}$). The objective function is then

$$f(\mathsf{A}, \mathsf{S}) = \frac{1}{2} \left\| (\mathbf{Y} - \bar{\mathsf{S}}\mathbf{A})^{\top} \Sigma_{\mathbf{Y}}^{-1} (\mathbf{Y} - \bar{\mathsf{S}}\mathbf{A}) \right\|_2^2,$$

$$f(\mathsf{A}, \mathsf{S}) = \frac{1}{2} \left\| (\mathbf{Y} - \bar{\mathsf{A}}\mathbf{S})^{\top} \Sigma_{\mathbf{Y}}^{-1} (\mathbf{Y} - \bar{\mathsf{A}}\mathbf{S}) \right\|_2^2$$

(23)

which describes an ordinary vector-valued linear inverse problem. The gradients are

$$\nabla_{\mathsf{A}} f(\mathsf{A}, \mathsf{S}) = -\bar{\mathsf{S}}^{\top} \Sigma_{\mathbf{Y}}^{-1} (\mathbf{Y} - \bar{\mathsf{S}}\mathbf{A})$$

$$\nabla_{\mathsf{S}} f(\mathsf{A}, \mathsf{S}) = -\bar{\mathsf{A}}^{\top} \Sigma_{\mathbf{Y}}^{-1} (\mathbf{Y} - \bar{\mathsf{A}}\mathbf{S})$$

(24)

and the covariance matrices are given by

$$\Sigma_{\mathsf{A}} = \bar{\mathsf{S}}^{\top} \Sigma_{\mathbf{Y}}^{-1} \bar{\mathsf{S}}$$

$$\Sigma_{\mathsf{S}} = \bar{\mathsf{A}}^{\top} \Sigma_{\mathbf{Y}}^{-1} \bar{\mathsf{A}}.$$

(25)

The Lipschitz constants of those gradients, which we need to calculate λ_j in Algorithm 2, are the spectral norms of the covariance matrices. Finally, the block matrices that encode A or S are given by

$$\bar{\mathsf{A}} = \begin{pmatrix} \mathsf{A}_{11}\mathsf{I}_N & \mathsf{A}_{21}\mathsf{I}_N & \ldots & \mathsf{A}_{K1}\mathsf{I}_N \\ \mathsf{A}_{12}\mathsf{I}_N & \mathsf{A}_{22}\mathsf{I}_N & \ldots & \mathsf{A}_{K2}\mathsf{I}_N \\ \vdots & \vdots & \ddots & \vdots \\ \mathsf{A}_{1B}\mathsf{I}_N & \mathsf{A}_{2B}\mathsf{I}_N & \ldots & \mathsf{A}_{KB}\mathsf{I}_N \end{pmatrix} \in \mathbb{R}^{BN \times KN} \qquad (26)$$

---

[10] The two works cited above did not need to compute the respective Lipschitz constants as they solved the NMF by multiplicative updates, not proximal algorithms.

and

$$\bar{\mathsf{S}} = \begin{pmatrix} \mathsf{S}^\top & 0 & \dots & 0 \\ 0 & \mathsf{S}^\top & \dots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \dots & \mathsf{S}^\top \end{pmatrix} \in \mathbb{R}^{BN \times BK}, \qquad (27)$$

each stacking the blocks row-wise over multiple bands.

2.5. PSF (de-)convolution

Incorporating the convolution with the point spread function (PSF) into our model is particularly useful for a deblender because it allows a distinction between overlap that is caused by the physical extent of the sources involved and the one due to atmospheric or instrumental blurring of the images. Knowing the latter reduces the ambiguity caused by the former.

In addition, our model from Equation 1 fits the same morphology to each band, which is only meaningful if the PSF does not change between bands or the images have been PSF-homogenized across all bands. Otherwise, the data will exhibit color variations that our model cannot reproduce, even if the underlying morphological structure of the astronomical source is achromatic like e.g. that of a star.

In the vectorized frame we have introduced in Section 2.4 we express the convolution as

$$\bar{\mathsf{P}} = \begin{pmatrix} \mathsf{P}_1 & 0 & \dots & 0 \\ 0 & \mathsf{P}_2 & \dots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \dots & \mathsf{P}_B \end{pmatrix} \in \mathbb{R}^{BN \times BN}, \qquad (28)$$

with each of the blocks describing the PSF in a given band as a linear operator that couples pixels of S. The model can thus be extended to have the objective functions

$$f(\mathbf{A}, \mathsf{S}) = \frac{1}{2} \left\| (\mathbf{Y} - \bar{\mathsf{P}}\bar{\mathsf{S}}\mathbf{A})^\top \Sigma_{\mathbf{Y}}^{-1} (\mathbf{Y} - \bar{\mathsf{P}}\bar{\mathsf{S}}\mathbf{A}) \right\|_2^2, $$
$$f(\mathbf{A}, \mathbf{S}) = \frac{1}{2} \left\| (\mathbf{Y} - \bar{\mathsf{P}}\bar{\mathbf{A}}\mathbf{S})^\top \Sigma_{\mathbf{Y}}^{-1} (\mathbf{Y} - \bar{\mathsf{P}}\bar{\mathbf{A}}\mathbf{S}) \right\|_2^2 \qquad (29)$$

with gradients

$$\nabla_{\mathbf{A}} f(\mathbf{A}, \mathsf{S}) = - \left(\bar{\mathsf{P}}\bar{\mathsf{S}} \right)^\top \Sigma_{\mathbf{Y}}^{-1} (\mathbf{Y} - \bar{\mathsf{P}}\bar{\mathsf{S}}\mathbf{A})$$
$$\nabla_{\mathbf{S}} f(\mathbf{A}, \mathbf{S}) = - \left(\bar{\mathsf{P}}\bar{\mathbf{A}} \right)^\top \Sigma_{\mathbf{Y}}^{-1} (\mathbf{Y} - \bar{\mathsf{P}}\bar{\mathbf{A}}\mathbf{S}) \qquad (30)$$

and covariance matrices

$$\Sigma_{\mathbf{A}} = \left(\bar{\mathsf{P}}\bar{\mathsf{S}} \right)^\top \Sigma_{\mathbf{Y}}^{-1} \bar{\mathsf{P}}\bar{\mathsf{S}}$$
$$\Sigma_{\mathbf{S}} = \left(\bar{\mathsf{P}}\bar{\mathbf{A}} \right)^\top \Sigma_{\mathbf{Y}}^{-1} \bar{\mathsf{P}}\bar{\mathbf{A}}, \qquad (31)$$

from which we can compute prox_f and the step sizes λ.

This ansatz with the PSF as a linear operator is only valid as long as the PSF images and the deconvolved model remain well sampled.[11] Otherwise, the discretized nature of either will

---

[11] Internal oversampling is an option to relax this requirement, which we will investigate in future work.

lead to convolution artifacts, which may result, e.g., in apparent non-monotonic behavior in the deconvolved morphology even if the source is perfectly monotonic.

We therefore recommend to construct difference kernels for each band, $\mathsf{P}_{\Delta,b}\mathsf{P}_{\text{min}} = \mathsf{P}_b$, such as to perform a partial deconvolution, resulting in an image with a band-independent PSF P_{min}. Requiring Nyquist-sampling for $\mathsf{P}_{\Delta,b}$ and P_{min} suggest that P_{min} should have a width T of about $\frac{1}{2} \min_b T(\mathsf{P}_b)$.

2.6. Translations and centering

In blends, peak positions from the initial detection step may not accurately reflect the true source centers. In fact, the apparent peak positions are often shifted towards neighboring objects. In addition, the combination of symmetry and monotonicity constraints results in the model being strongly affected by the assumed location of each source's center. To avoid spurious dipole residuals from inaccurate center positions, we introduce linear translation operators T_x and $\mathsf{T}_y \in \mathbb{R}^{N \times N}$ to shift the position of each source. These operators encode the translation operation, e.g.

$$\forall i < N : (\mathsf{S}_{k,i}, \mathsf{S}_{k,i+1}) \rightarrow ((1 - dx)\mathsf{S}_{k,i}, dx\,\mathsf{S}_{k,i+1}) \qquad (32)$$

for a shift of component k by an amount dx in the positive horizontal direction, as sparse band-diagonal matrices, i.e. we treat the translation as a linear interpolation between the pixel and its relevant neighbors. This is again feasible only if the source is well sampled, and higher-order interpolation schemes can be implemented if needed.

We introduce the translation in horizontal and vertical directions into the model via $\mathbf{S}' = \bar{\mathsf{T}}\mathbf{S}$, using the block-diagonal matrix

$$\bar{\mathsf{T}} = \begin{pmatrix} \mathsf{T}_1 & 0 & \dots & 0 \\ 0 & \mathsf{T}_2 & \dots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \dots & \mathsf{T}_K \end{pmatrix} \in \mathbb{R}^{KN \times KN}, \qquad (33)$$

where the two-dimensional shift for each component is given by $\mathsf{T}_k = \mathsf{T}_{y,k}\mathsf{T}_{x,k}$. If multiple components belong to the same source, they have identical T_k. The shifted morphology vector $\mathbf{S}'$ is then used instead of $\mathbf{S}$ in Equation 23, Equation 27, and Equation 29. This way, all gradients and Lipschitz constants are correctly computed with the equations given above.

To determine optimal source centers, we create a new set of $\mathsf{T}'_{x,k}$ and $\mathsf{T}'_{y,k}$ matrices, each with an additional shift δ of typically 0.1 pix in either horizontal or vertical direction. We form two difference images for each component,

$$\mathsf{D}_{x,k} = \mathsf{A}_k \times \left(\mathsf{T}'_{x,k} - \mathsf{T}_k \right) \mathsf{S}_k$$
$$\mathsf{D}_{y,k} = \mathsf{A}_k \times \left(\mathsf{T}'_{y,k} - \mathsf{T}_k \right) \mathsf{S}_k, \qquad (34)$$

and combine them into the matrix

$$\bar{\mathsf{D}} = \begin{pmatrix} \mathsf{D}_{x,1} & \mathsf{D}_{y,1} & \mathsf{D}_{x,2} & \mathsf{D}_{y,2} & \dots & \mathsf{D}_{x,K} & \mathsf{D}_{y,K} \end{pmatrix}. \qquad (35)$$

We then state that the current model residuals are a linear combination of the difference images,

$$\mathbf{Y} - \bar{\mathsf{P}}\bar{\mathbf{A}}\mathbf{S} = \bar{\mathsf{D}}\Delta, \qquad (36)$$

an ordinary linear inverse problem for the two-dimensional offsets $\mathbf{\Delta}$, which we solve with a least-squares solver. We perform this secondary minimization every 10-th iteration, so that the updates for S have time to adjust to changes in the centers.

3. Applications

3.1. Scene from the HSC UltraDeep survey

We demonstrate the performance of scarlet in deblending deep ground-based images from HSC. The test case shown in Figure 1 stems from the UltraDeep COSMOS coadd images of the HSC public data release (Aihara et al., 2018b). The data are taken in filters *grizy* to a 5σ point source depth of 27.4, 27.3, 27.0, 26.4, and 25.6, respectively, with average seeing of 0.74 arcsec or better. This example was selected because of its complex configuration with 33 detected sources, which span a wide range of morphologies, sizes, and brightnesses.

We detect the sources in a detection image, defined as the sum across all bands, using the Python package sep[12] (Barbary, 2016), which implements the functionality of SExtractor. The detection threshold was set at 1.2σ above the background, with very sensitive deblending parameters DEBLEND_THRESH=64 and DEBLEND_CONT=3e-4; these settings were chosen to detect the majority of sources in this particular example, without over-deblending the larger galaxies in the image.

3.1.1. Results of scarlet

We specify the CMF model in the simplest form, namely with single-component sources, which means the model cannot describe color variations within a galaxy. We initialize the component morphologies S_k by constructing symmetric, monotonic approximations to the region in the detection image around each identified peak, very similar to the approach the SDSS *Photo* deblender employs. The initial SEDs A_k are set to the color of the data averaged with the spatial weight function S_k. We perform 200 iterations of Algorithm 2, using inverse per-pixel variances as weights, but ignoring any noise correlation from resampling on the coadd image. We also ignore the relatively small PSF variation between bands in the example, thereby performing the source separation in the convolved frame.

As one can see in the central panel of Figure 1, most features of the data are represented well by the scarlet model, despite the limited flexibility of one-component sources. Furthermore, there are two resolved spiral galaxies (IDs 19 and 22), which in detail violate our assumptions of symmetric and monotonic behavior. Nonetheless, the residuals in the right panel of Figure 1 indicate that this simple model can describe most of the source morphologies. Moreover, the residuals clearly reveal that our model could be extended by adding several components to account for previously undetected localized emission, either from unrelated sources (such as the red clumps in the regions of IDs 19 and 25) or distinct stellar populations in detected sources (as in the centers of IDs 24, 29, and 33).

We leave the model refinement based on the analysis of the residuals to future work, and instead compare the deblending results of scarlet, the HSC-SDSS deblender, and SExtractor for several sources in Figure 2. Because SExtractor does not attempt to separate overlapping sources (it merely seeks to find a suitable boundary between them), we show the multi-band data for the pixels belonging to the respective objects according to SExtractor's segmentation map of the detection image.

scarlet performs well on the two large spiral galaxies (IDs 19 and 22), recovering a close approximation of their shape, despite the symmetry and monotonicity constraints. While there will be cases where those constraints are too restrictive, e.g. for tightly wound spirals, these two examples demonstrate that galaxies with a modest amount of internal structure can be modeled well, even with a single component. The combination of constraints is useful for dense groups of objects, e.g. around IDs 20 or 6, where e.g. a symmetry constraint alone would not suffice. ID 6 is well modeled despite being rather faint because the simultaneous use of all available bands helps determine a best-fitting morphology, which in turn is used as a matched filter to determine the SED. However, in IDs 20 or 25 we can see that this process can fail when undetected sources are present in the vicinity of model components. The optimization algorithm of scarlet will reduce the residuals by connecting to undetected emission via a flat "bridge", which is the only option available under the monotonicity constraint. If an undetected source has a different SED, a single-component source has no other option than to adopt the flux-weighted mean SED, which is clearly visible in ID 25. An ℓ_0 sparsity penalty (cf. Section 2.3.3) could suppress those bridges, however its use is problematic because it would bias low the recovery of total fluxes.

3.1.2. Comparison to other deblenders

SExtractor makes no such structural assumptions, but the limitations of the "cookie cutter" approach clearly show with the large spirals IDs 19 and 22, as well as their neighbors, e.g. ID 20. As a result, most objects are too compact, which will lead to low biases on total fluxes and sizes. SExtractor is also prone to merging different sources together, even when they are discontinuous in the segmentation map (IDs 19, 22, 25).

The HSC-SDSS deblender performs surprisingly badly on ID 19 and, by extension, on its neighbors. Because the pixel fluxes are exactly preserved, a failure in one source has strong effects on others, e.g. IDs 20 and 25 pick up the flux from the outskirts of 19. However, the addition of monotonicity, which was not present in the SDSS *Photo* implementation by Lupton (in prep.), helped in the three-in-a-row case of ID 6.

As a summary, we find scarlet to perform well on a complex blended scene of deep HSC imaging, despite its simplifying assumptions of single-component sources and the rigidity of the symmetry and monotonicity constraints. It simultaneously models all sources and is automatically consistent across all bands. A conceptual limitation remains, namely the sensitivity to undetected sources or multiple stellar populations within detected sources, which could be mitigated by adding or increasing penalties such as a strong ℓ_0 sparsity that restricts the support of each component to a minimum number of pixels.
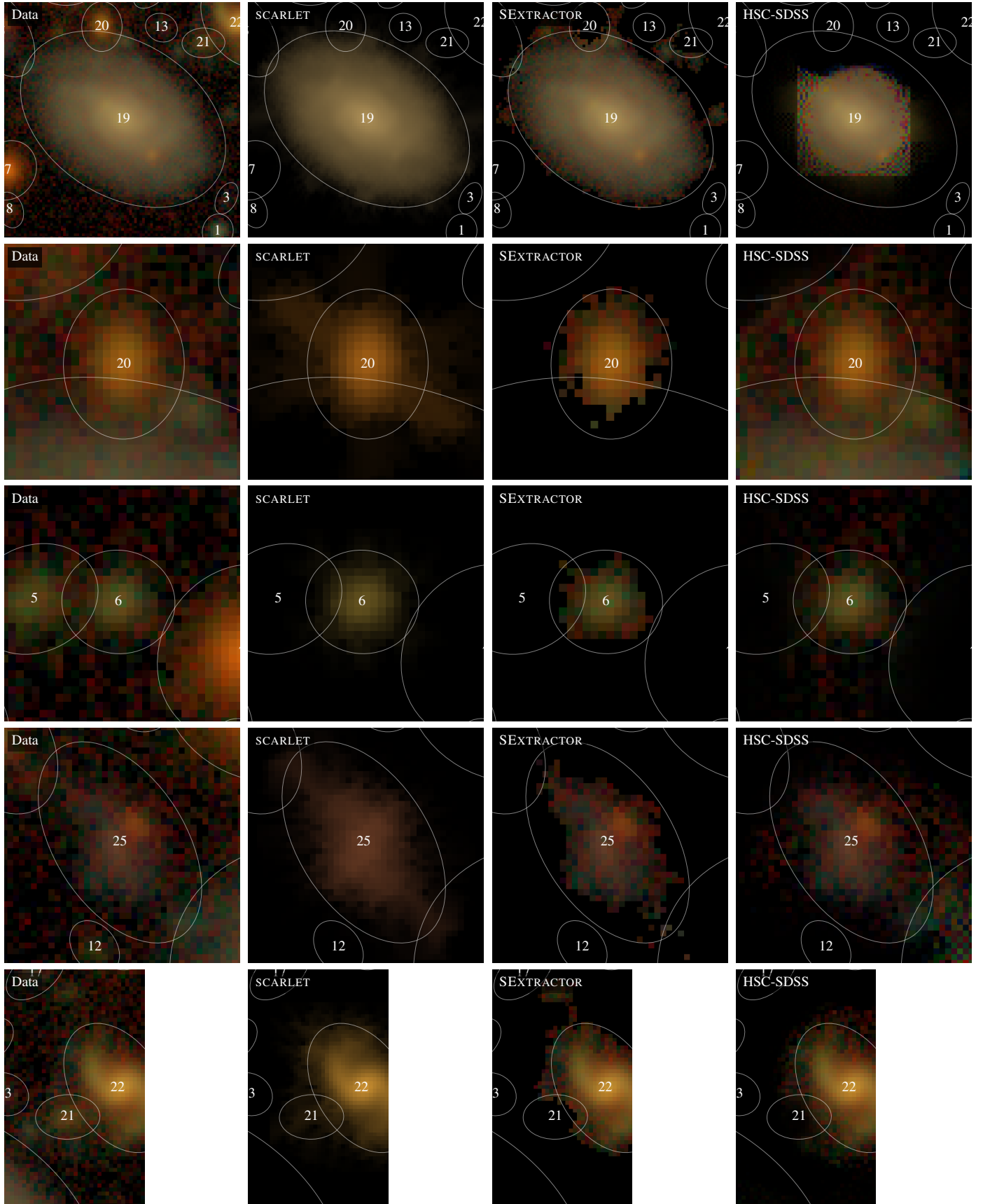
Figure 2: Deblender comparison for several sources (one per row) from Figure 1. *1st column:* HSC *grizy* coadd data. *2nd column*: The corresponding component of the scarlet model. *3rd column:* Data from the 1st column, restricted to the object's pixels in the SExtractor segmentation map of the detection image. *4th column*: HSC-SDSS deblender result for the object, run independently on each band.

On the other hand, an iterative sequence that adds components at the location of significant and coherently colored residuals appears viable and could substantially augment traditional detection methods.

3.2. Simulations

As the problem of deblending astronomical scenes is inter-related with detecting all relevant sources, we seek to decouple these two tasks with the help of simulations where we know the number and location of each source. The simulations also allow us to determine the accuracy of the recovered source properties, namely total fluxes, SEDs, and morphologies. We will first test how well the assumptions of monotonicity and symmetry allow us to recover complex morphologies of isolated sources (Section 3.2.1) and then how that recovery is affected when multiple overlapping objects are present in each scene (Section 3.2.2). Both sets of simulations are created with the GALSIM software package (Rowe et al., 2015).

As metrics to quantify the fidelity of the source modeling and deblending, we determine the fluxes, SEDs, and per-pixel morphologies of deblender results. For the total flux, which for SCARLET is given by $\int d^2x\, S_k$, we simply compute the fractional error compared to the input model. For SEDs and morphologies we compute the correlation coefficient between the true and the observed vectors v,

$$\xi_v = \frac{v_{\text{true}} \cdot v_{\text{obs}}}{\sqrt{v_{\text{true}} \cdot v_{\text{true}}} \cdot \sqrt{v_{\text{obs}} \cdot v_{\text{obs}}}}. \tag{37}$$

The coefficients ξ_{A_k} and ξ_{S_k} measure how well a deblender recovers colors and shapes (including intensity) of component k, respectively. We consider these fidelity measures as a function of true source flux and source "blendedness" (Bosch et al., 2018)

$$\beta = 1 - \frac{S_k \cdot S_k}{S_{\text{all}} \cdot S_k}, \tag{38}$$

where $S_{\text{all}} = \sum_k S_k$, expecting that good results will be harder to achieve for faint sources in very crowded areas.

3.2.1. Individual galaxies

We start by investigating how well SCARLET can reproduce realistic galaxy morphologies. We use a set of galaxies from the COSMOS *HST* sample described in Mandelbaum et al. (2012) that are available with GALSIM [13]. From the full sample of 87,000 galaxies with $i < 25.2$, we further select those with total fluxes between 300 and 2,000 counts that pass most stringent "marginal" cuts. The bright cutoff was chose to remove galaxies with visible blending: many of the brightest galaxies have faint background galaxies or foreground stars in their wings. As we seek to test the recovery of *individual* sources here, those interlopers are contaminants that reduce the correlation measure in Equation 37. They could easily be fit with SCARLET if their existence and positions were known, but for the simulations here
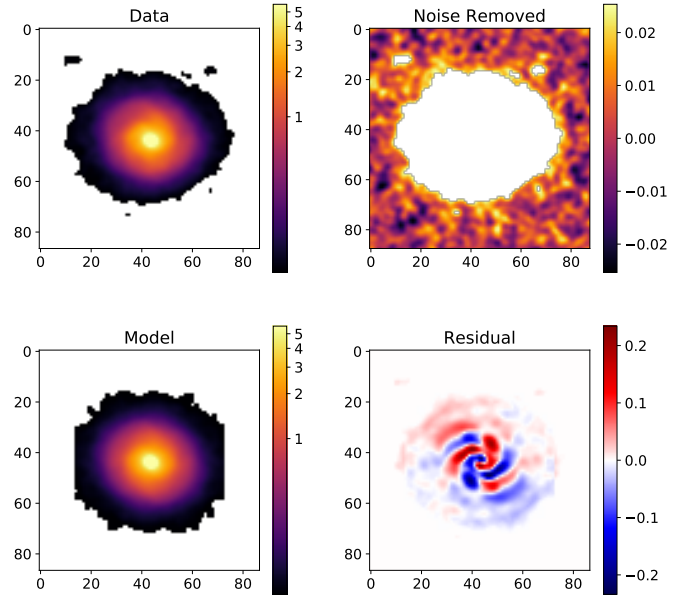
Figure 3: An example galaxy from the COSMOS *HST* dataset in GALSIM. *Top left:* The "truth" image, defined by an original *HST* image from which we removed the noise in the galaxy outskirts. *Top right:* The difference between the original *HST* image and the "truth". *Bottom left:* The model generated by SCARLET with symmetry and monotonicity constraints. *Bottom right:* The residuals (truth − model) reveal spiral structure that is missing in the SCARLET model. The panels on the left use a $\sinh^{-1}$ scaling.

we want to assume that we know the position of every individual source. The lower flux cutoff was imposed to eliminate faint sources, whose outskirts are dominated by correlated noise from the original *HST* observations. To further reduce the effect of correlated noise, we estimate the noise level over the postage stamp and subtract the corresponding value from the entire image, setting all negative pixel values to zero. While this procedure is likely to cut off some flux from the wings of the galaxy, the bulk of the galaxy and especially its internal structure is preserved. The remaining sample comprises 799 galaxies. An example galaxy from this procedure is shown in Figure 3.
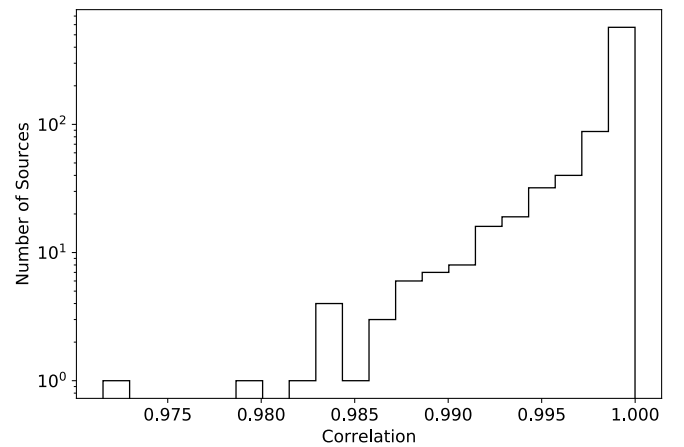
Figure 4: Morphology correlation coefficient (Equation 37) of SCARLET models for 799 isolated galaxies from the COSMOS *HST* sample in GALSIM.

11

We run SCARLET on single-band images of all 799 galaxies. The pixel resolution is set to a typical ground-based value of 0.2 arcsec; no noise is added to the image. We show the resulting model and the residuals for the example galaxy in the bottom panels of Figure 3. For reference, its total flux error is 0.08%, its morphology correlation coefficient $\xi_S = 0.999$. The RMS of the total flux error for the entire sample is 0.39%, which increases for fainter sources, suggesting that some correlated noise still made it into the input images and is suppressed by the monotonicity constraint. The morphology correlation ξ_S is shown as a cumulative distribution function in Figure 4. These results demonstrate that for realistic galaxies, SCARLET recovers morphologies with a very strong correlation to the input morphologies despite the constraints of symmetry and monotonicity; consequently, the estimation of total fluxes is also excellent.

3.2.2. Blended scenes

We now create a set of 10,000 blended scenes in the LSST filters *ugrizy*, with widths and heights varying between 30 and 70 pixels, comprising on average a mix of 20% stars and 80% galaxies, with a common PSF across all bands. The stellar sources have SEDs taken from the Pickles (1998) templates for O5V, B5V, A5V, F5V, K5V, and M5V stars, and each stellar morphology is defined as a 2D Gaussian with $\sigma = 10^{-9}$ pix that is convolved with the PSF in each band. Galaxy sources have SEDs based on four empirical (Coleman et al., 1980, CWW) templates available in GALSIM, randomly perturbed to generate a total of 25 different galaxy SED templates. Their redshifts are randomly chosen between 0 and 2.1. For the galaxy morphologies we use the same 799 galaxies selected in Section 3.2.1. All sources are inserted into the scenes with a power-law distribution in radius from the center, resulting in strongly crowded blends in the middle of the images. Uncorrelated Poisson noise is added to the images, using a sky background brightness at the mean noise level of the *HST* images, so that we can largely hide the truncation of the input galaxy morphologies (cf. Figure 3).

We run SCARLET and the current HSC-SDSS deblender and separate the results for stars and galaxies assuming that larger average brightness and more compact shape should make stars easier to deblend than galaxies. Indeed, we find in Figure 5 that total flux errors are smaller for stars than for galaxies and that SCARLET substantially outperforms the HSC-SDSS deblender. These two trends continue with the correlation coefficients of SEDs in Figure 6 and of morphologies in Figure 7. We understand the remarkable difference between these two deblending strategies, which have very similar constraints on the source morphologies, as a result of fitting a joint model across all bands (SCARLET) or independent models for each band (HSC-SDSS). Our strategy is superior in these test cases because we combine images of all bands, resulting in a morphology that is driven by the band with the highest significance, which is in turn used as an optimal filter to determine the source SED.

We emphasize that these results are achieved with single-component, symmetric, and monotonic source models, even though the true morphologies in the simulations exhibit more complex structure. There will be cases in which those con-
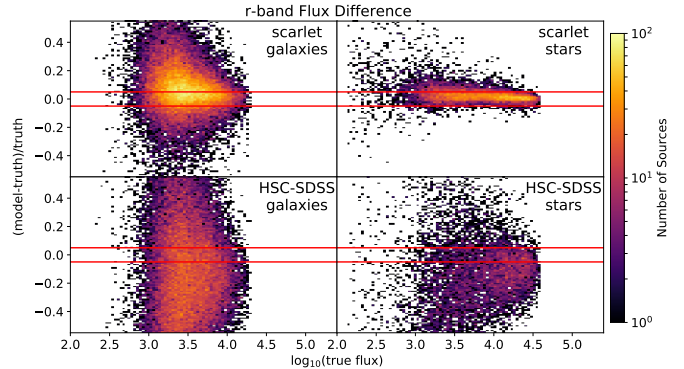
Figure 5: Comparison of the relative error of total fluxes in the *g* between SCARLET (top) and the HSC-SDSS deblender (bottom), for galaxies (left) and stars (right). For reference, the red horizontal lines show 5% relative errors. We only show the results of the *r*-band as the other bands look qualitatively similar.

straints lead to substantial errors, but we believe that they are now demonstrably justified as powerful default configurations for deblending in ground-based wide-field surveys.

We finally note that the runtime of SCARLET in our tests is about 10 ms per band and source, driven almost entirely by the cost to evaluate the likelihood gradients in Equation 30. The computational cost increases substantially if the PSF convolution cannot be expressed efficiently. While slower than e.g. the HSC-SDSS deblender, SCARLET only has to be run once over all images and directly yields measurements of flux, SED, and morphology of the sources. We fully expect source extraction with SCARLET to be a computationally expensive but still feasible task for deep optical surveys.

3.3. Extended Emission Line Regions of AGN

To further demonstrate the capabilities of SCARLET as a flexible component separation framework, we analyze HSC Wide coadd images in *grizy* bands of galaxies with extended emission-line regions (EELRs) created by active galactic nuclei (AGN). Resulting from AGN radiation on inhomogeneous inter-stellar medium, the morphology of EELR is often complex, asymmetric, and non-monotonic. The strong emission lines, e.g., [OIII]λ5007, emitted can be detected and resolved by broad-band imaging as components that may spatially overlap with but have colors distinct from their host galaxies (e.g., Lintott et al., 2009; Keel et al., 2012).

The particular example shown in Figure 8, SDSS J023106-034513, is taken from the EELR sample of Sun et al. (2018), who selected EELR candidates from spectroscopically identified type 2 obscured AGNs. With a high luminosity of $L_{bol} = 10^{45.7}$ erg s^{-1}, this AGN creates extended [OIII]λ5007 emission with a diameter of 60 kpc that is captured in the HSC *i*-band image, see Figure 8.

We seek to separate the EELR from its host galaxy with SCARLET. We first detect all sources in the image, and then assign them to either host, EELR, or one of 3 likely unrelated interlopers. We constrain the morphology of the interlopers with our default combination of non-negativity, monotonicity and symmetry. The host galaxy does not appear symmetric, so
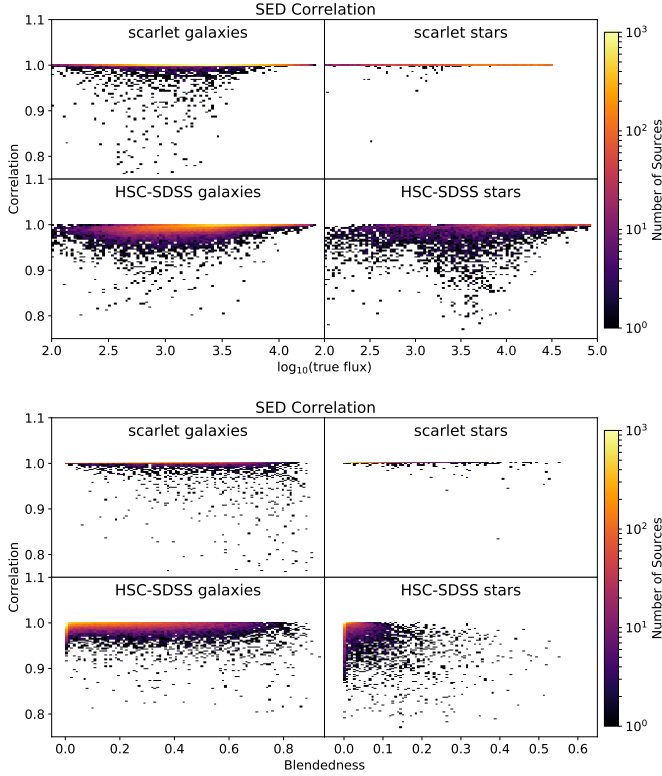
Figure 6: Correlation, as defined by Equation 37, between true SEDs and the SEDs from the HSC-SDSS deblender or scarlet as a function of true source flux (top four panels) and blendedness (Equation 38, bottom four panels.)
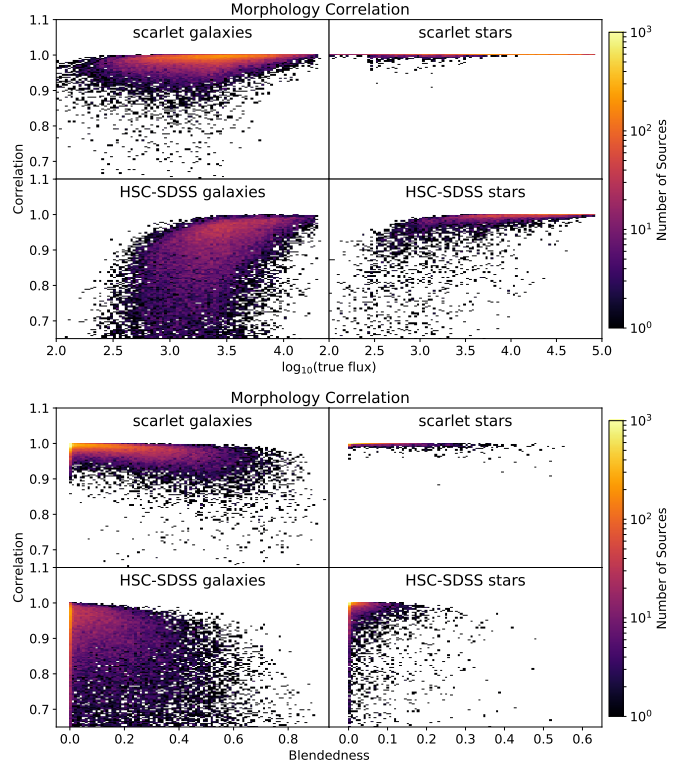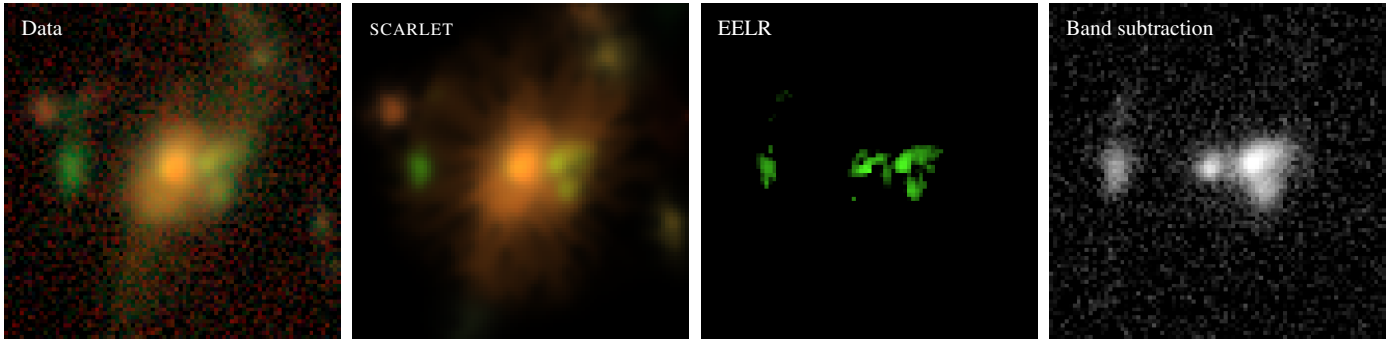

Figure 7: Same as Figure 6 for the correlation between the true source morphology and the morphologies from the HSC-SDSS deblender or scarlet. For HSC-SDSS, the morphology is taken from the *g*-band; other bands look similar.

we remove that constraint[14], while the EELR has no constraints other than non-negativity.

To compensate for the weakening of morphological constraints, we can utilize our knowledge of the EELR broadband SED based on the emission lines in the observed SDSS spectrum. Similarly, by assuming that the host galaxy is responsible for the continuum spectrum without emission lines, we can also predict its broadband SED. The proximal operator for A_k is thus simply a projection onto the precomputed SEDs of EELR or host and mute for any of the interlopers:

$$\text{prox}_{\text{AGN}}(A_k) = \begin{cases} A_{\text{EELR}} & \text{if } k = k_{\text{EELR}} \\ A_{\text{Host}} & \text{if } k = k_{\text{Host}} \\ A_k & \text{else.} \end{cases} \quad (39)$$

We use inverse per-pixel variances as weights and match the observed PSFs between bands by deconvolving to an effective resolution of 0.35 arcsec, using the formalism of Section 2.5. We show the converged, convolved scarlet model in the second panel of Figure 8 and the partially deconvolved EELR component in the third panel. For comparison, in the forth panel, we also show the EELR image obtained via image continuum subtraction from Sun et al. (2018). In this approach, to iso-

late the EELR signal, the galaxy continuum in the *i*-band is removed by subtracting the (scaled) *z*-band image, which contains only the continuum but not line emissions. We can see that the scarlet model of the EELR strongly resembles the continuum-subtraction result, while revealing smaller spatial features of the EELR. A more detailed application of scarlet to EELR extraction is planned for future work.

4. Summary and Outlook

The employment of multi-channel information is long known to aid source separation schemes. It was therefore expected that multi-band imaging should be beneficial in cases when multiple sources overlap in a celestial scene. To exploit the availability of multi-band imaging from modern sky surveys, we have developed the source separation framework scarlet. It uses a matrix factorization scheme, in which one matrix factor describes the amplitude of all components in each band and the other factor describes their spatial shape. We use the proximal optimization method bSDMM presented by Moolekamp & Melchior (2018) to allow for an arbitrary number of constraints on each matrix factor. The astrophysical interpretation of this model is that the scene is composed of a number of components that each describe one star or distinct stellar population. Galaxies can be modeled as several such components to characterize e.g. bulge and disc-like features.

---

[14]In experiments, we found that the host is better modeled by two co-centered components, similar to a bulge-disk model.

Figure 8: *1st panel:* False-color image of *grizy* coadd images of SDSSJ023106-034513 from the HSC Wide data release, shown with an arcsinh stretch. *2nd panel:* Convolved scarlet model of the scene with a two-component monotonic source for the host galaxy, a free-form source for the EELR, and monotonic and symmetric sources for other detections in the scene. *3rd panel:* Partially deconvolved EELR component. *4th panel:* Result of the band subtraction $i - z$ from Sun et al. (2018), which subtracts the galactic continuum to isolates the EELR emission.

To aid the source separation further, especially in crowded regions with little variation in source colors (such as galaxy cluster cores), we constrain the component morphologies to be monotonic and/or symmetric with respect to the sources peak position. If available, constraints on the SED of the sources, e.g. from stellar SED libraries or high-resolution spectra, are beneficial as well.

We have derived the analytical treatment of correlated pixel noise and of band-dependent PSF convolution, rendering scarlet capable of working on coadd images with variable seeing conditions. In essence, it performs a PSF-matched photometry measurement with an optimally chosen weight function given by the signal-to-noise-weighted mean morphology in all available bands. Applied to an example scene from the HSC Ultra-Deep survey, we find scarlet to robustly recover all detected sources, even if they exhibit complex morphologies. For detailed analyses of extended galaxies one can conceive of additional constraints, such as the Fourier mode extension of Galfit (Peng et al., 2010).

We show with a set of dedicated simulations comprised of stars and realistic galaxy morphologies that scarlet yields an accurate recovery of source fluxes, SEDs, and morphologies, even for prominent levels of blendedness. In our tests it clearly outperforms the HSC-SDSS deblender in each of these measurements and is especially powerful for stars and marginally resolved as well as faint sources.

Because of its non-parametric nature, scarlet exhibits a set of inherent advantages and drawbacks, which we summarize in Table 1. In addition, due to its flexibility it is sensitive to undetected sources or additional components of detected sources. On the other hand, the inspection of the fit residuals appears to clearly reveal their presence. We therefore envision a scheme in which components are added at the location of significant and coherently colored residuals. A decision whether any newly added component is part of a previously detected source or independent of all other sources will require a post-processing step that encodes an interpretation based on the nature of the astronomical scene.

When run in this fashion scarlet breaks a common model of data processing in optical astronomy that consists of three distinct stages: detection, deblending, and characterization. If components are added from an analysis of the residuals and the models are used to infer source properties, it essentially covers aspects of all three stages. While doing so, it makes assumptions about the SEDs and morphologies of the sources as well as the noise and the PSFs in the observations, which need to be valid to produce reliable results. It will therefore be necessary to assess the accuracy of scarlet models for specific science cases. To facilitate inference directly on scarlet sources, we intend to provide an error-estimation scheme, which accurately captures component degeneracies and effects of non-differentiable constraints.

We have intentionally developed scarlet such that it can utilize any constraint or prior that can be expressed as a proximal operator without having to change the algorithm itself, thus rendering it suitable to a wide range of source separation problems in image data. We demonstrated this flexibility with an AGN jet–host galaxy separation in this work and with a hyperspectral unmixing task in Moolekamp & Melchior (2018). It is inherent to the matrix factorization approach that scarlet should work well in applications where high-quality information on the spatial distribution of sources is to be combined with high-quality information on their spectra. An example is the combination of a single high-resolution image from a space-born telescope, such as *HST* or the future *WFIRST* (Spergel et al., 2015), with multi-band imaging of lower spatial resolution from a ground-based telescope. We will pursue this research direction in forthcoming works.

scarlet is implemented in Python with C++ extensions, accessible through pybind11, and is available at `https://github.com/fred3m/scarlet`. Its interface is highly modular and should be easily extendible to different applications. Contributions and requests from the astronomical community are highly encouraged.

14

Table 1: Main advantages and drawbacks of SCARLET.

Feature	Advantage	Disadvantage
Matrix Factorization	Low-rank representation of celestial scenes	Not suited for correlated changes of SED *and* morphology
Non-parametric model	Arbitrary morphologies and SEDs	Requires critical sampling No analytic PSF convolution
Proximal gradient optimization	Wide range of priors and penalties	No analytic error propagation

References

References

Aihara, H., Arimoto, N., Armstrong, R., et al. 2018a, PASJ, 70, S4
Aihara, H., Armstrong, R., Bickerton, S., et al. 2018b, PASJ, 70, S8
Ball, N. M., Loveday, J., & Brunner, R. J. 2008, MNRAS, 383, 907
Barbary, K. 2016, Journal of Open Source Software, 1, 58
Barden, M., Häußler, B., Peng, C. Y., McIntosh, D. H., & Guo, Y. 2012, MNRAS, 422, 449
Bauschke, H. H., & Borwein, J. M. 1996, SIAM review, 38, 367
Bertin, E., & Arnouts, S. 1996, A&AS, 117, 393
Blanton, M. R., & Roweis, S. 2007, AJ, 133, 734
Bosch, J., Armstrong, R., Bickerton, S., et al. 2018, PASJ, 70, S5
Boyd, S., Parikh, N., Chu, E., Peleato, B., & Eckstein, J. 2011, Foundations and Trends® in Machine Learning, 3, 1
Coleman, G. D., Wu, C.-C., & Weedman, D. W. 1980, ApJS, 43, 393
Combettes, P. L., & Pesquet, J.-C. 2011, in Fixed-point algorithms for inverse problems in science and engineering (Springer), 185–212
Connor, T., Donahue, M., Kelson, D. D., et al. 2017, ApJ, 848, 37
Conselice, C. J. 1997, PASP, 109, 1251
Dark Energy Survey Collaboration. 2016, MNRAS, 460, 1270
Dawson, W. A., Schneider, M. D., Tyson, J. A., & Jee, M. J. 2016, ApJ, 816, 11
Doi, M., Fukugita, M., & Okamura, S. 1995, ApJS, 97, 59
Douglas, J., & Rachford, H. H. 1956, Transactions of the American mathematical Society, 82, 421
Drlica-Wagner, A., Sevilla-Noarbe, I., Rykoff, E. S., et al. 2018, ApJS, 235, 33
Eckstein, J., & Bertsekas, D. P. 1992, Mathematical Programming, 55, 293

Esser, E., Zhang, X., & Chan, T. F. 2010, SIAM Journal on Imaging Sciences, 3, 1015
Frieden, B. R. 1972, J. Opt. Soc. Am., 62, 511
Furusawa, H., Koike, M., Takata, T., et al. 2018, PASJ, 70, S3
Gabay, D., & Mercier, B. 1976, Computers & Mathematics with Applications, 2, 17
Glowinski, R., & Marroco, A. 1975, Revue française d'automatique, informatique, recherche opérationnelle. Analyse numérique, 9, 41
Green, P. J. 1995, Biometrika, 82, 711
Hassanieh, H., Indyk, P., Katabi, D., & Price, E. 2012, ArXiv e-prints, arXiv:1201.2501
Ivezic, Z., Tyson, J. A., Abel, B., et al. 2008, ArXiv e-prints, arXiv:0805.2366
Joseph, R., Courbin, F., & Starck, J.-L. 2016, A&A, 589, A2
Keel, W. C., Chojnowski, S. D., Bennert, V. N., et al. 2012, MNRAS, 420, 878
Komiyama, Y., Obuchi, Y., Nakaya, H., et al. 2018, PASJ, 70, S2
Lee, D. D., & Seung, H. S. 2001, in Advances in Neural Information Processing Systems 13, ed. T. K. Leen, T. G. Dietterich, & V. Tresp (MIT Press), 556–562
Lin, C.-J. 2007, Neural computation, 19, 2756
Lintott, C. J., Schawinski, K., Keel, W., et al. 2009, MNRAS, 399, 129
MacArthur, L. A., Courteau, S., & Holtzman, J. A. 2003, ApJ, 582, 689
Mandelbaum, R., Hirata, C. M., Leauthaud, A., Massey, R. J., & Rhodes, J. 2012, MNRAS, 420, 1518
Miyazaki, S., Komiyama, Y., Kawanomoto, S., et al. 2018, PASJ, 70, S1
Moolekamp, F., & Melchior, P. 2018, Optimization and Engineering, arXiv:1708.09066
Paatero, P., & Tapper, U. 1994, Environmetrics, 5, 111
Parikh, N., Boyd, S., et al. 2014, Foundations and Trends® in Optimization, 1, 127
Peng, C. Y., Ho, L. C., Impey, C. D., & Rix, H.-W. 2010, AJ, 139, 2097
Pickles, A. J. 1998, PASP, 110, 863
Rapin, J., Bobin, J., Larue, A., & Starck, J.-L. 2013, IEEE Transactions on Signal Processing, 61, 5620
—. 2014, ArXiv e-prints, arXiv:1407.7691
Refregier, A. 2003, MNRAS, 338, 35
Rix, H.-W., Barden, M., Beckwith, S. V. W., et al. 2004, ApJS, 152, 163
Rowe, B. T. P., Jarvis, M., Mandelbaum, R., et al. 2015, Astronomy and Computing, 10, 121
Samuroff, S., Bridle, S. L., Zuntz, J., et al. 2018, MNRAS, 475, 4524
Spergel, D., Gehrels, N., Baltay, C., et al. 2015, ArXiv e-prints, arXiv:1503.03757
Sun, A.-L., Greene, J. E., Zakamska, N. L., et al. 2018, MNRAS, arXiv:1803.07241
Xu, Y., & Yin, W. 2013, SIAM Journal on imaging sciences, 6, 1758
Zhang, Y., McKay, T. A., Bertin, E., et al. 2015, PASP, 127, 1183
Zhu, G. 2016, ArXiv e-prints, arXiv:1612.06037