The Astrophysical Journal, 852:104 (16pp), 2018 January 10 https://doi.org/10.3847/1538-4357/aaa1f2
© 2018. The American Astronomical Society. All rights reserved.

Non-negative Matrix Factorization: Robust Extraction of Extended Structures

Bin Ren (任彬)[1], Laurent Pueyo[2], Guangtun Ben Zhu[1,5], John Debes[2], and Gaspard Duchêne[3,4]
[1] Department of Physics and Astronomy, Johns Hopkins University, Baltimore, MD 21218, USA; ren@jhu.edu, guangtun.ben.zhu@gmail.com
[2] Space Telescope Science Institute (STScI), Baltimore, MD 21218, USA; pueyo@stsci.edu
[3] Astronomy Department, University of California, Berkeley, CA 94720, USA
[4] Université Grenoble Alpes, CNRS, IPAG, F-38000 Grenoble, France
Received 2017 September 7; revised 2017 December 1; accepted 2017 December 12; published 2018 January 11

Abstract

We apply the vectorized non-negative matrix factorization (NMF) method to the post-processing of the direct imaging data of exoplanetary systems such as circumstellar disks. NMF is an iterative approach, which first creates a nonorthogonal and non-negative basis of components using the given reference images and then models a target with the components. The constructed model is then rescaled with a factor to compensate for the contribution from the disks. We compare NMF with existing methods (classical reference differential imaging method, and the Karhunen–Loève image projection algorithm) using synthetic circumstellar disks and demonstrate the superiority of NMF: with no need of prior selection of references, NMF not only can detect fainter circumstellar disks but also better preserves their morphology and does not require forward modeling. As an application to a well-known disk example, we process the archival *Hubble Space Telescope* STIS coronagraphic observations of HD 181327 with different methods and compare them, and NMF is able to extract some circumstellar materials inside the primary ring for the first time. In an appendix, we mathematically investigate the stability of NMF components during the iteration and the linearity of NMF modeling.

Key words: protoplanetary disks – stars: imaging – stars: individual (HD 181327) – techniques: image processing

1. Introduction

Detection and characterization of circumstellar disks and exoplanets with direct imaging rely on excellent design of telescope instruments and state-of-the-art post-processing of the observations. Well-designed instruments are able to stabilize the temporally varying noise in telescope exposures, creating quasi-static features (Perrin et al. 2004, 2008; Hinkley et al. 2007; Soummer et al. 2007; Traub & Oppenheimer 2010). These quasi-static features (speckles), together with stellar point-spread function (PSF), can be empirically modeled and removed by post-processing techniques, revealing circumstellar disks and exoplanets around stars (e.g., Marois et al. 2006; Lafrenière et al. 2007, 2009; Soummer et al. 2012).

Post-processing techniques have been evolving in the past decade. When the stellar PSF and the speckles are stable and do not have significant variation over time, they can be removed by subtracting the image of a reference star (i.e., reference-star differential imaging [RDI]). This classical RDI method has been extensively used, unveiling revolutionary results, especially for bright disks (e.g., Smith & Terrile 1984; Grady et al. 2010; Debes et al. 2013, 2017; Schneider et al. 2014, 2016). When the speckles do vary, especially in the ground-based telescopes, several techniques have been proposed to resolve this (e.g., angular differential imaging [ADI], Marois et al. 2006; spectral differential imaging [SDI], Biller et al. 2004).

Current successful removal of both stellar PSF and quasi-static noises is based on the utilization of innovative statistical methods (Mawet et al. 2012). Using a large sample of references, there are two widely known advanced post-processing algorithms: one is the Locally Optimized Combination of Images (LOCI) algorithm (Lafrenière et al. 2007), and the other is the principal-component-analysis-based Karhunen–Loève Image Projection (KLIP) algorithm (Amara & Quanz 2012; Soummer et al. 2012). Both LOCI and KLIP have yielded a lot of new discoveries even with archival data (e.g., Lafrenière et al. 2009; Soummer et al. 2012, 2014; Choquet et al. 2014, 2016; Mazoyer et al. 2014, 2016); however, the aggressive PSF subtraction of LOCI biases the photometry and astrometry of the results (Marois et al. 2010; Pueyo et al. 2012), the projection onto the eigenimages of KLIP reduces the flux from the astrophysical objects, and forward modeling has to be performed (Soummer et al. 2012; Choquet et al. 2016, 2017). Most forward-modeling works are designed for exoplanets (Pueyo 2016), and attempts have been made for circumstellar disks (e.g., Esposito et al. 2014; Milli et al. 2014; Wahhaj et al. 2015; Follette et al. 2017); however, for some disks that are irregularly shaped (see canonical examples in Grady et al. 2013; Dong et al. 2016), forward modeling does not work satisfactorily since the morphology of disks is buried in current aggressive post-processing procedures (Follette et al. 2017).

An accurate recovery of disk morphology can help us understand the disk properties in several ways. First, we can retrieve their surface brightness profiles and reveal possible asymmetric structures and the traces of complex dynamical structures (spiral arms, jets, clumps, etc.; see Dong et al. 2015a, 2016, for some canonical examples). Second, we will be able to study the evolution of them on short timescales (see Debes et al. 2017, for the yearly evolution of the TW Hydrae disk). Third, the morphology of disks in current simulation results also indicates the possible existence of unseen planets that are perturbing the circumstellar disk structure and create observable signatures (e.g., Rodigas et al. 2014; Lee & Chiang 2016; Nesvold et al. 2016; Dong & Fung 2017).

[5] Hubble Fellow.

The aim of post-processing is to detect and characterize point sources and extended circumstellar disks. Point-source searching methods have gained much progress in recent years (SDI, ADI, etc.); however, all these algorithms perform poorly for extended disks, and it is hard, if not nearly impossible, to fully calibrate the disks in that context. With an eye toward more robust and well-calibrated PSF subtraction for disk studies, we study in this paper the non-negative matrix factorization (NMF) method. The early work of NMF was carried out by Paatero & Tapper (1994) and became well known after Lee & Seung (2001), who provided update rules to guarantee convergence through iteration. In Blanton & Roweis (2007), they came up with update rules that can handle nonuniform and missing data, especially for astronomical spectroscopic observations; their method was improved by Zhu (2016) in a vectorized form, which is adopted in this paper because of its excellent parallel computational efficiency. An early attempt with NMF on high-contrast imaging has been performed by Gomez Gonzales et al. (2017); in this paper we study the method in detail.

In post-processing, the steps of NMF are similar to KLIP: construct components from given references first, and then model any target with the components. Unlike KLIP, NMF does not remove the mean of every image, and it keeps all the entries non-negative, which consequently constructs a non-orthogonal component basis. When modeling a target, non-negative coefficients for the components are obtained. In order to perfectly recover the morphology of the astrophysical signals such as circumstellar disks, the NMF subtraction results do need forward modeling; however, we are able to show that, to the first order, this can be performed with a simple search in a one-dimensional space.

The structure of this paper is as follows: Section 2 discusses the limitation of current methods and explains the mechanism of NMF; Section 3 is composed of the post-processing results of various methods with modeled disks and the application to a classical example of HD 181327, thus making the NMF method stand out among current ones; and Section 4 summarizes the general performance of NMF and discusses its significance to the field. Appendix A provides a list of symbols used in this paper; Appendix B shows the update rules proposed by Zhu (2016), as well as the adjustments made for direct imaging data; Appendix C investigates the stability of NMF components during their construction; Appendix D presents how to model a target with the NMF components; and Appendix E describes our procedure to correct for oversubtraction.

2. Methods

To illustrate the limitation of current methods and compare them with NMF, we took data from the *HST* Space Telescope Imaging Spectrograph (STIS) coronagraphic observations of HD 38393 (γ Leporis; Proposal ID: 14426;[6] PI: J. Debes) and aligned the centers of the stars using a radon-transform-based center-determination method described in Pueyo et al. (2015). We release an improved center-determination code, `centerRadon`,[7] which is based on radon transform, performs line integrals on specific parameter spaces and selected regions, and is at least 2 orders of magnitude faster than the classical radon transform center-determination code for our STIS data.

We cut the aligned exposures into 87×87 pixel arrays,[8] corresponding with a 4.41×4.41 arcsec^2 field of view. The original data contain 810 0.2 s exposures of HD 38393 at nine different telescope orientations. In our simulation, all the data are normalized to have flux units of mJy arcsec^{-2}, and to save computational time, we use only 81 exposures by selecting just nine exposures at each orientation.

This section is organized as follows: Section 2.1 discusses the limitation of current modeling methods, and Section 2.2 explains the NMF method in detail.

2.1. Limitation of Current Methods

Current post-processing methods are excellent in finding disks; however, they are limited and have several disadvantages. We start our simulation with a synthetic face-on disk generated by `MCFOST`, a radiative transfer software to model circumstellar disks with given physical parameters of the disks (Pinte et al. 2006, 2009). For the exposures of HD 38393 at one orientation, the synthetic disk was injected onto them after being added with Poisson noise to simulate real observations, therefore creating the *target* images, while the exposures from the other eight orientations are treated as the *references*.

The targets and references are then used for post-processing. In order to perform the RDI technique, we compute the pixel-wise median of the references, creating an empirical model of the stellar PSF and speckles, and subtract the model from the disk-injected images. For KLIP, we construct the components from the references, model the targets with the components, and then remove the KLIP model from the targets. This procedure is executed for all nine orientations, and then the final result is calculated from the pixel-wise median of the 81 subtracted targets.

The subtraction results with the synthetic face-on disk are shown in Figure 1. The classical RDI subtraction method can barely recover the morphology of the disk; KLIP recovers the general morphology of the disk but loses the details (e.g., the radial profile, especially the radial slope of the disk): it is overfitting the disk and thus reduces the flux (Choquet et al. 2014; Ren et al. 2017). For the radial profiles of the recovered disks, although the error bars of the classical RDI result agree with that of the disk model, this consistency is not convincing —the error bar for a given radial separation is calculated from the standard deviation of the pixels located within ± 0.5 pixels[9] of that radial location: large azimuthal variations result in large standard deviations. Although the error bars of the KLIP result are smaller, there is a systematic downward bias, and more importantly, the slope of the KLIP result is not consistent with that of the disk model, and forward modeling has to be performed to reveal the true slope.

In order to accurately retrieve the morphology and photometry of the planets or disks simultaneously, forward modeling has been adopted; however, not only does this procedure assume a *prior* morphology model of the objects (e.g., Soummer et al. 2012; Wahhaj et al. 2015;

[6] http://archive.stsci.edu/proposal_search.php?mission=hst&id=14426

[7] https://github.com/seawander/centerRadon

[8] Note: the dimension of all the images in this paper is 87×87 pixels (4.41×4.41 arcsec^2) unless otherwise specified.

[9] The STIS pixel size corresponds with 0.05078 arcsec (Riley 2017).

Pueyo 2016), but it is also time-consuming to iteratively recover them and break the degeneracy (e.g., Choquet et al. 2016, 2017).

2.2. Non-negative Matrix Factorization

The methods to faithfully recover the objects have been evolving. On one hand, new methods have been proposed and studied to minimize oversubtraction: Pueyo et al. (2012) focus on the positive coefficients for the LOCI method and substantially improve the characterization quality of point-source spectra. On the other hand, forward modeling is introduced as a correction method for the reduced data: Wahhaj et al. (2015) assume a prior model of the disks for the LOCI subtraction; Pueyo (2016) takes the instrumental PSF to characterize point sources with the KLIP method. Current forward-modeling attempts are best optimized for planet characterizations, while for the disks, assumptions of disk morphology have to be made; however, these assumptions may not represent the true morphology. In this paper, we aim to circumvent the forward-modeling difficulties by studying a new method—NMF.

NMF decomposes a matrix to the product of two non-negative ones (Paatero & Tapper 1994; Lee & Seung 2001), which has been evolving in the past decade to adjust for astrophysical problems (Blanton & Roweis 2007; Zhu 2016). Inspired by the Pueyo et al. (2012) work of adopting positive coefficients, we study NMF because of its non-negativity, which is well suited for astrophysical direct imaging observations. The previous applications of NMF are to one-dimensional astrophysical spectra; for two-dimensional images with a significantly larger amount of data and therefore escalating the computational cost, we make the following adjustments: on the one hand, we flatten every image into one dimension to maximize the utilization of currently available tools, and on the other hand, we adopt the vectorized NMF technique (Zhu 2016; `NonnegMFPy`[10]) and implement parallel computation with multiple cores.

The NMF application to imaging data is composed of two steps: constructing the basis of components with the reference images (Section 2.2.1), and modeling any new target with the component basis (Section 2.2.2) and then correcting for the overfitting with a scaling factor (Section 2.2.3). To reduce the reader's effort in coding, we release our realization,[11] which is also available in the `pyKLIP` package (Wang et al. 2015).

2.2.1. Component Construction

The first step of NMF is to approximate the *reference* matrix R, with the product of two non-negative matrices: the coefficient matrix W, and the component matrix H, i.e.,

$$R \approx WH, \tag{1}$$

by minimizing their Euclidean distances; see Appendix A for detailed definition of symbols. The approximation of Equation (1) is guaranteed to converge with iteration rules in Zhu (2016) using

$$W^{(k+1)} = W^{(k)} \circ \frac{RH^{(k)T}}{W^{(k)}H^{(k)}H^{(k)T}}, \tag{2}$$

$$H^{(k+1)} = H^{(k)} \circ \frac{W^{(k)T}R}{W^{(k)T}W^{(k)}H^{(k)}}, \tag{3}$$

with random initializations. In the above equations, the circle $\circ$ and fraction bar[12] $\frac{(\cdots)}{(\cdots)}$ denote element-wise multiplication and division for matrices, the superscripts enclosed with $^{(\cdot)}$ denote iteration steps, and the superscript ^{T} stands for matrix transpose. For astronomical data, a weighting function V, which is usually the element-wise variance (i.e., the square of the uncertainties) of R, is applied to weigh the contribution from different pixels and take care of heteroscedastic data (Blanton & Roweis 2007; Zhu 2016); see Appendix B for the adaptation for our STIS imaging data.

The connection between NMF and previous statistical methods can be illustrated using Equation (2): we can cross out the W terms on the right-hand side,[13] and we get $W = \frac{RH^T}{HH^T}$, which stands for the projection of vector R onto vector H. This expression is in essence performing least-squares estimation as in KLIP, where the inversion of the covariance matrix of the components is required (i.e., the inverse of HH^T); however, the covariance matrices are often poorly conditioned for inversion. Intuitively, NMF returns a non-negative approximation of the matrix inverse through iteration.

In the KLIP method, the importance of the components is ranked based on the magnitude of their corresponding eigenvalues. For NMF, we rank them by constructing the components *sequentially*: with n components constructed, we construct the $(n+1)$th component using the n previously constructed ones, i.e., in our construction, we only randomize the initialization of the $(n+1)$th component, while the first n components are initialized with their previously constructed values. See Appendix C for detailed expression and derivation. This construction method not only ranks the components but also is essential for the linearity in target modeling in the next subsection.

2.2.2. Target Modeling ("Projection")

The sequential construction of NMF components is the foundation of this paper. First, as illustrated in Appendix C, the components remains stable in this setup. Second, the stability of the components guarantees a linear separation of the disk signal from the stellar signals (Appendix D). Third and most importantly, the linearity of the target modeling process calls for our attempt to *circumvent forward modeling with a scaling factor*, as illustrated in Appendix E.

[10] https://github.com/guangtunbenzhu/NonnegMFPy
[11] https://github.com/seawander/nmf_imaging

[12] Note: all the fraction bars in this paper are element-wise division of matrices unless otherwise specified.
[13] Note: this operation is for demonstration purposes only; it is not mathematically practical.

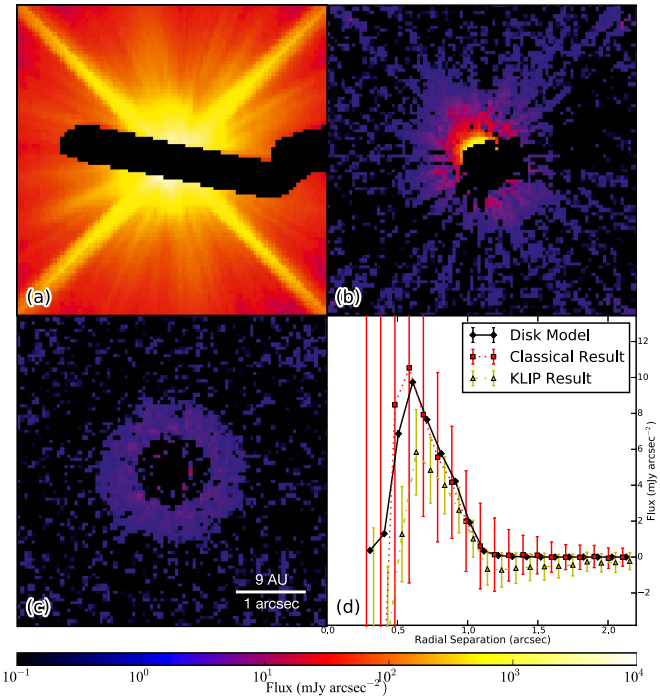

Figure 1. Demonstration of limitations of current PSF subtraction methods with a synthetic face-on circumstellar disk with integrated flux ratio $F_{\rm disk}/F_{\rm star} = 7.4 \times 10^{-6}$. (a) *HST*-STIS exposure of HD 38393 added with a synthetic face-on disk. The disk is barely seen because of its faintness, and the central dark region is the coronagraphic BAR5 mask of STIS. (b) Classical RDI subtraction result; the northeastern region shows an overluminosity that does not belong to the disk. (c) Subtraction result with KLIP; the disk is seen, but its flux is reduced and morphology modified. (d) Radial profiles of the subtraction results and the disk model; although that of the classical method agrees with the disk model, it is not trustworthy because the error bars are calculated from the large standard deviation of pixels at the same radial separation. KLIP not only is unable to recover the radial profile or even change the slope but also induces unphysically negative pixels from its systematic bias. See Figure 8 for the disk model and the subtraction result with NMF.

With the basis of NMF components constructed sequentially, the next step is to model the targets with the components. For a flattened target T, we now minimize $||T - \omega H||^2$ with iteration rule[14]

$$\omega^{(k+1)} = \omega^{(k)} \circ \frac{TH^T}{\omega^{(k)} HH^T}, \tag{4}$$

where ω is the $1 \times n$ coefficient matrix for the target and H is the NMF components constructed in the previous paragraph. This expression is essentially performing a least-squares approximation as in KLIP, but the coefficients are smaller in magnitude (Appendix D). A more detailed expression taking the weighting function into account is given by Zhu (2016), and the adaptation to our STIS data is shown in Appendix B. When the above process converges, the NMF model of the target can be represented by

$$T_{\rm NMF} = \omega H. \tag{5}$$

[14] This rule is the one-dimensional case of Equation (2).

2.2.3. Disk Retrieval via "Forward Modeling"

With sequentially constructed components, the target modeling procedure is able to linearly separate the circumstellar disk from the others, as illustrated in Equation (51). To the first order, we have

$$T_{\rm NMF} = D_{\rm NMF} + S_{\rm NMF}, \tag{6}$$

where the subscript $_{\rm NMF}$ means performing the NMF modeling result for the stellar signal (S) or disk signal (D) alone. In addition, when we *sequentially* model the target, if the disk does not resemble any NMF component, then the first component, which explains ~90% of the residual noise (shown in Section 3.1.1), will always dominate the modeling for the disk—the captured morphology of the disk is just another copy of the NMF model of the stellar PSF and the speckles (shown in Section 3.1.2), i.e.,

$$D_{\rm NMF} \approx \alpha S_{\rm NMF}, \tag{7}$$

where α is a positive number.

The way to correct for the contribution from the disk is to introduce a scaling factor $\hat{f}$, which satisfies

$$\hat{f}\, T_{\rm NMF} = S_{\rm NMF}, \tag{8}$$

and then when we subtract the scaled NMF model of the target from the raw exposure, we will have the disk image:

$$T - \hat{f}\, T_{\rm NMF} = T - S_{\rm NMF} = D. \tag{9}$$

Ideally, we will solve for $\hat{f} = 1/(1 + \alpha)$; however, since α is not known, we have to find $\hat{f}$ empirically.

We introduce the best factor finding (BFF) procedure in Appendix E as our attempt to *circumvent forward modeling* with a simple scaling factor, which finds the factor corresponding with the minimum background noise. To illustrate the efficiency of BFF, we show the reduction results with different scaling factors in Figures 2 and 3. When we do not know the existence of the astrophysical signal (i.e., the disk) a priori, the residual variance dependence on scaling factor agrees consistently with the dependence of the Euclidean distances between the reduction results (D_f) and the true model (D) on the scaling factor. This consistency has been observed for synthetic disks at different inclination angles in our simulation, which is not shown in this paper to avoid redundancy of figures.

We could use multiple scaling factors to rescale every overfitting component, and BFF will work for components that are affecting the overall morphology. Given the sparseness of the NMF coefficients (ω_D, Section 3.1.2), this is easily achievable with a grid search. However, we only focus on the first component, since the BFF procedure is trying to optimize the whole field of view, while the components of higher order usually do not have such influence in our simulations.

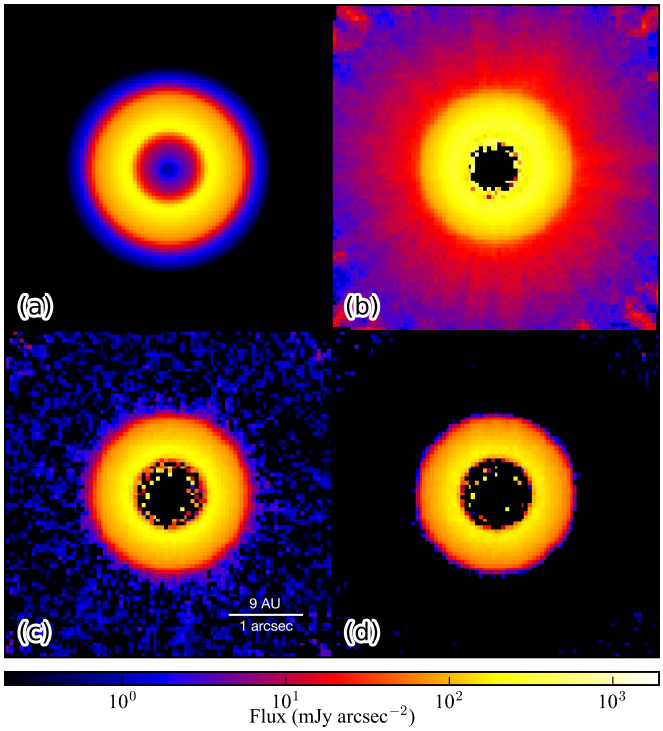

Figure 2. Illustration of the scaling factor for the disk model in Figure 5. (a) Face-on disk model created by MCFOST. (b) Scaled reduced disk with $f = 0.930$; there are PSF residuals since the scaling factor is smaller than the optimum one. (c) Scaled reduced disk with $f = \hat{f} = 0.982$, the best disk corresponding with the optimal scaling factor obtained from the BFF procedure. (d) Reduced disk with no correction (i.e., $f = 1$); this is an oversubtracted disk, the disk flux is reduced, and the outskirts of the disk are all negative. See Figure 3 for a comparison of the radial profiles.

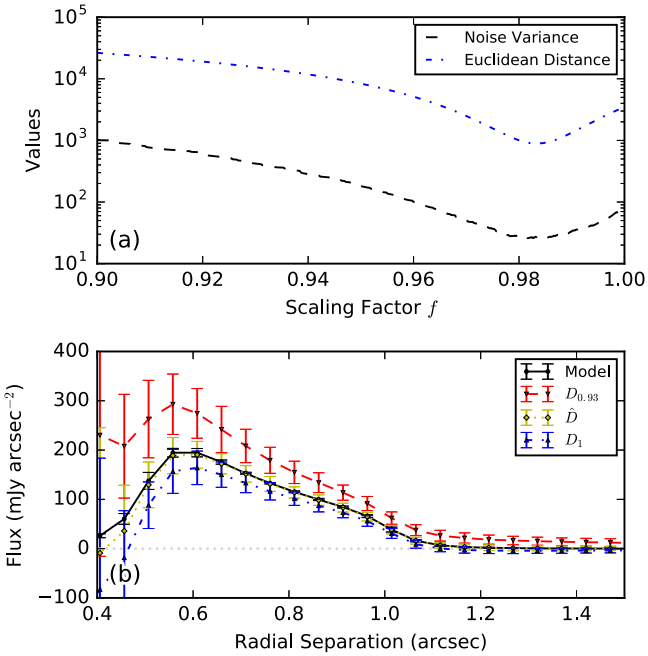

Figure 3. (a) The curve for Euclidean distance between the scaled disks and the MCFOST model in Figures 5 and 2 (blue dot-dashed line) is consistent with the curve for the $\sigma_f^{(\mathrm{conv})}$(dashed black line). This is the demonstration of the effectiveness of the BFF procedure. (b) Radial profiles for the model (black solid line), and scaled disks with three different scaling levels. When $f = 0.930 < \hat{f}$, the radial profile (black solid line) is moving upward relative to that of the model; when $f = 0.982 = \hat{f}$, its radial profile (yellow dotted line) agrees with that of the model; for $f = 1 > \hat{f}$, the diskless pixels are all negative (blue dot-dashed line; compare with the gray dotted horizontal line). See Figure 6 for the results from other methods.

3. Comparison and Application

In the previous section we have demonstrated the limitation of current methods and the mechanism of NMF; in this section, we aim to demonstrate the ability of NMF in direct imaging using specific examples. We first compare the statistical properties of NMF and KLIP, as well as the intermediate steps of them, in Section 3.1, and then focus on the post-processing results: we compare the NMF results with those of the classical RDI and KLIP subtraction methods using synthetic disks in Section 3.2, and then in Section 3.3 we focus on a well-known example as a sanity check of NMF—applying the method to the *HST*-STIS coronagraphic imaging observations of HD 181327.

3.1. NMF versus KLIP: Statistical Properties and Intermediate Steps

In this subsection, we aim to address the statistical differences and intermediate steps between NMF and KLIP and investigate why the non-negativity of NMF can yield better results. Noise and disk signal are the two constituents in a target image; however, they are always correlated with each other, and separating them is the goal of all post-processing efforts. In the target modeling process, we aim to maximize noise removal and minimize disk capturing. We therefore compare NMF with KLIP in these two aspects.

3.1.1. Noise Removal

Removing the quasi-static noise from the observations is the most fundamental procedure in post-processing. With the 81 STIS images of HD 38393, we calculate the fractional residual variance (FRV) curves in the following way (Figure 1 of Soummer et al. 2012): for each image, we cumulatively increase the number of components and model it with KLIP or NMF, and then we subtract the model from the image to obtain the residual image. Then the FRV is calculated by dividing the variance of the residual image by that of the original image. The comparison is shown in Figure 4.

The FRV curves for KLIP decrease very fast at first, indicating that KLIP is removing the quasi-static noise, then reach a plateau for a long time, and drop again when almost all the components are used. The plateau exists when KLIP is not removing the *noise*, and it might even be trying to capture the *disk signal* if anything is fitted during the plateau. When the curves drop again, KLIP is starting to remove the *random noise* that should not be removed using any method.

The FRV curves for NMF decrease relatively slowly; however, it is gradually capturing the quasi-static noise and then converges to a higher level than KLIP when all the components are used. The higher noise level indicates that the random noise is kept, and it is also preserving

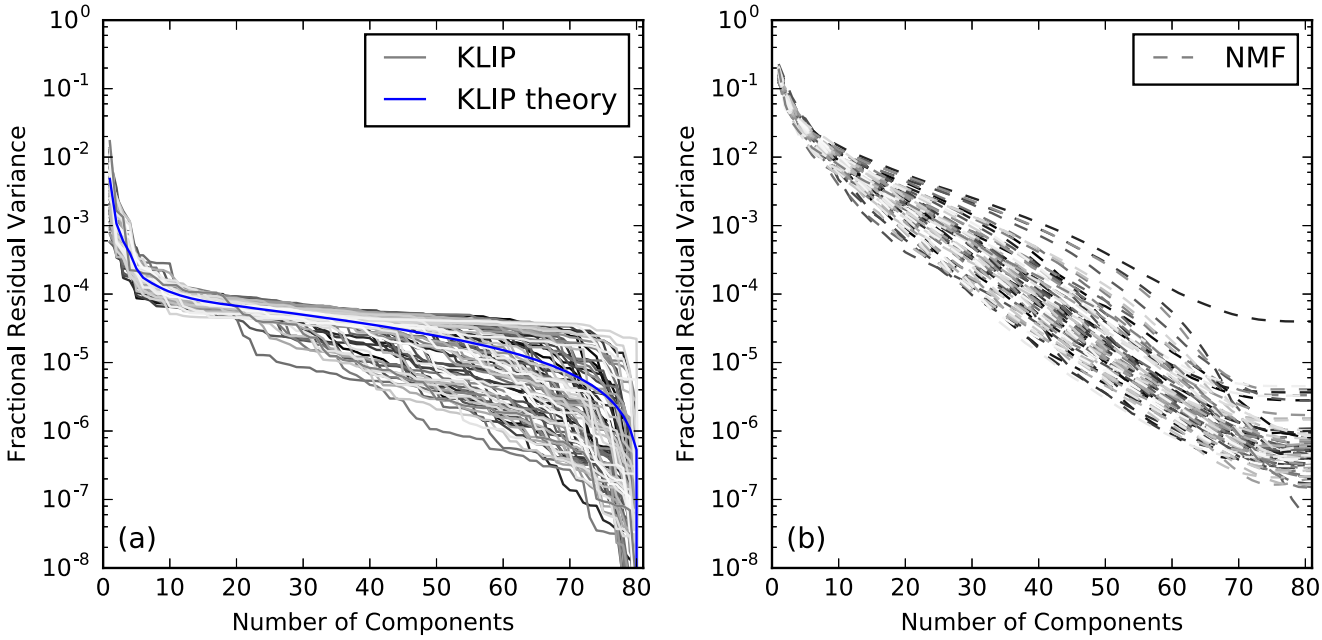

Figure 4. FRV as a function of the number of components. (a) FRV plots for KLIP; the gray solid lines are for individual images, and the blue solid line is the theoretical curve (as in Figure 1 of Soummer et al. 2012). The existence of the plateau from $n \approx 10$ to $n \approx 60$ indicates that KLIP is not efficiently capturing the noises. (b) FRV plots for NMF; the gray dashed lines are for individual images. There is no plateau in the NMF reduction, indicating that it continues noise capturing when we increase the number of components. The comparison between KLIP and NMF projections at similar FRV levels is shown in Figure 5. Note: the FRV trends of KLIP and NMF are not limited to the STIS data studied in this paper and should be applicable to all other instruments.

the difference between the target and the component basis.

3.1.2. Disk Capturing

Disk signal is coupled with the stellar PSF and the quasi-static speckles, and it is likely to be removed together with them in post-processing: this is when overfitting happens, and it is why current post-processing methods need forward modeling to compensate for that oversubtraction. In terms of disk capturing, we assume that the stellar PSF and speckles are perfectly removed, and then we project only a synthetic face-on MCFOST disk onto the components to study the target modeling process: the less disk is captured the better, since the disk signal in principle shall not be classified as the stellar PSF or speckles that should be modeled by the components.

The meaning of projection is different for the two methods: for KLIP, the projection process is directly performing a dot product between the target and the components, while for NMF, its "projection" is an iterative approach, which finds a non-negative combination of the NMF components to model the target as in Equation (4). In this paper, we do *not* distinguish the two processes in words, but they are *not* identical with respect to different methods.

We compare their intermediate modeling step of the disk in Figure 5 under similar residual noise levels ($n = 40$ components when FRV $\approx 10^{-4}$, and with normalized KLIP and NMF components in Figure 4).

KLIP is an "efficient" disk capturing method, and therefore it gives rise to oversubtraction, which requires forward modeling to compensate for that. Although the morphology of the disk does not resemble any KLIP component, the disk is captured as a result from direct linear projection. This provides evidence of KLIP falling into the regime of overfitting: a fraction of the disk is classified as the stellar PSF or speckle noise, and then it is removed from the target image.

NMF is "inefficient" in disk capturing, causing less oversubtraction, and is thus preferred in post-processing. Although the NMF target modeling process is *in essence* performing linear projection, the projection coefficients are sparse and have smaller magnitudes than direct projections. As is shown in Equation (34), the NMF modeling is less overfitting than the direct projection methods like KLIP.

The sparsity of these coefficients inspires the "forward modeling" for NMF in Section 2.2.3: we are able to accomplish this by rescaling the NMF model of a target with a simple factor, which is obtained from the BFF procedure as demonstrated in Appendix E.

3.2. Synthetic Disk Models

To compare NMF with current post-processing methods, we first synthesized three circumstellar disks with MCFOST at different inclination angles and brightness levels. In this paper, we do not aim to fit any physical parameters of the disks as in the previous MCFOST applications. However, the disks are still physically motivated and composed of silicates with grain size ranging from 0.2 to 2000 μm with a power-law index of 3.5. The disks are rings with a flaring index of 1.125 spanning from 0.5 to 1.0 arcsec. Our MCFOST disk models are synthesized at 0.6 μm and convolved with the STIS TinyTim PSF (Krist et al. 2011),[15] to simulate the *HST*-STIS response: at this wavelength, the incident photons

[15] http://www.stsci.edu/hst/observatory/focus/TinyTim

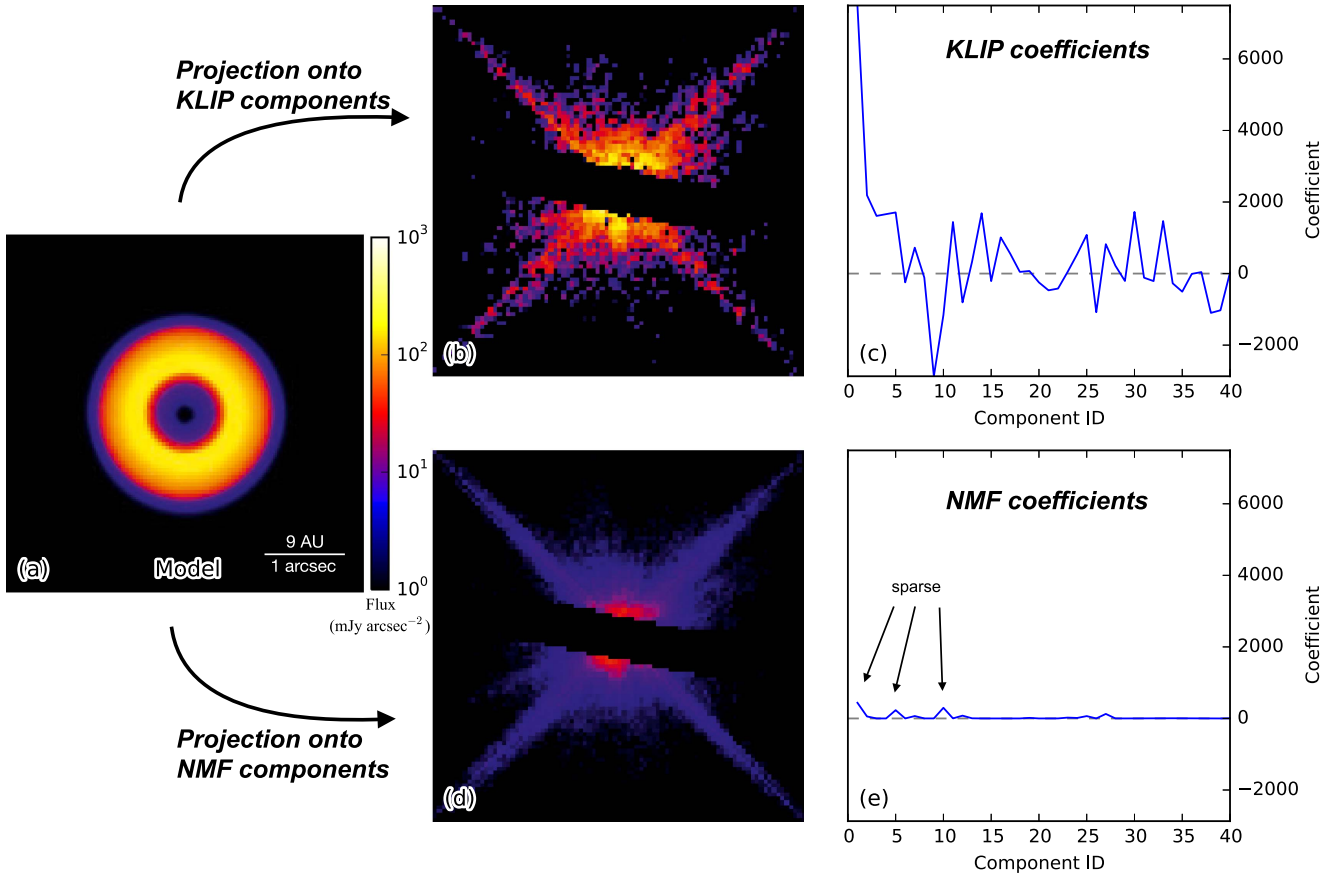

Figure 5. Comparison between KLIP and NMF projections using a synthetic face-on MCFOST disk model. (a) Disk model. (b) Projection of the disk model onto the KLIP components; the central circularly shaped structure is the result from overfitting. (c) Coefficients of each component in KLIP modeling. (d) "Projection" of the model onto the NMF components. (e) Coefficients of each component in NMF modeling in panel (d): the fact that both the components and the coefficients are non-negative reduces the likelihood of overfitting, as shown in Equation (34). Note: the central dark regions in panels (b) and (d) are the coronagraphic occulting mask at the STIS BAR5 position, and the images are in the same scale.

from the host star are scattered by the disks and then received by the telescopes.

We reduced the synthetic disks in the same manner as in Section 2.1 with the classical RDI, KLIP, and NMF methods. The face-on disk spanning from 0.5 to 1.0 arcsec in Figure 5 is adopted as the initial model, which is then tilted at 45° and 75°; to investigate the performance of the three methods at different contrasts, we dim the disks in Figure 6 by dividing by factors of 10, 20, and 50, which is equivalent to reducing disk mass, having $F_{\rm disk}/F_{\rm star}$ range from $\sim 10^{-4}$ to $\sim 10^{-6}$. The morphology results are shown in Figures 6–9, respectively.

3.2.1. Morphology

Morphology is the most direct evidence for the spatial distribution of the circumstellar materials, which should be recovered to the maximum extent. From the morphology results for different disks and brightness levels in Figures 6–9, we compare the three methods as follows.

Classical subtraction is only able to recover the morphology of the face-on disk in the brightest cases ($F_{\rm disk}/F_{\rm star} \sim 10^{-4}$); it cannot recover the dimmer ones.

KLIP is able to recover the general morphology of the disks in all cases; however, the disk fluxes are reduced owing to its systematic overfitting bias (Section 3.1.2). Therefore, forward modeling is needed to recover the real morphology of the disks. Moreover, for detailed structures, especially when the dynamic range is large, KLIP will lose the relatively faint information owing to its mean subtraction (see the vanishment of the far side in the lower regions of the tilted disks images in Figures 6–9).

NMF outperforms other methods not only in recovering the morphology of the disks but also in recovering the faint structure: the faint far sides of the tilted disks are recovered in the results.

When the disks are too faint (Figure 9), the reduced results are dominated by random noise, and none of the methods but NMF could marginally recover the morphology of the disks.

For a quantitative comparison of the three methods, we compute the χ^2 values using the initial model and the results in this subsection. To better illustrate the relative goodness of recovery, we plot the χ^2 ratios between different methods and the classical RDI method in Figure 10. NMF performs better than the classical method in this way; when

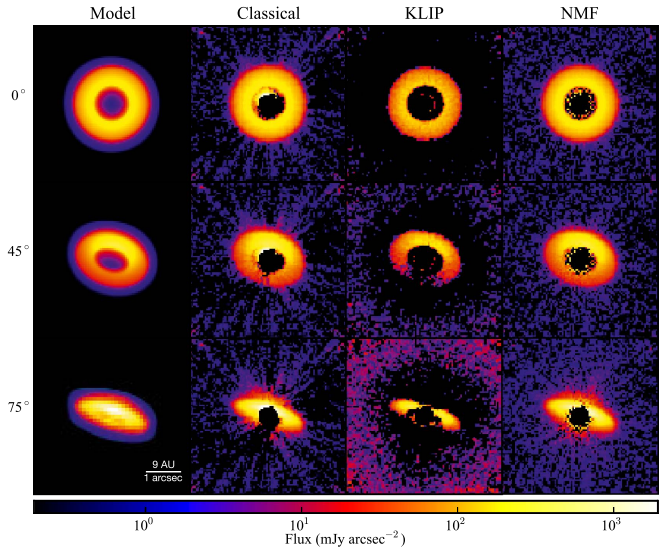

Figure 6. Initial models created by MCFOST at three different inclinations: morphology of disks reduced by different methods. From top to bottom, the disks is tilted at 0°, 45°, and 75° (going from face-on to nearly edge-on) with $F_{\rm disk}/F_{\rm star} = (1.5, 0.9, 1.9) \times 10^{-4}$, respectively. First column: models; second column: classical subtraction results; third column: KLIP subtraction results; fourth column: NMF subtraction results. Both KLIP and NMF recover the geometries better than the classical method, and the dark halo around the KLIP images rises from its oversubtraction.

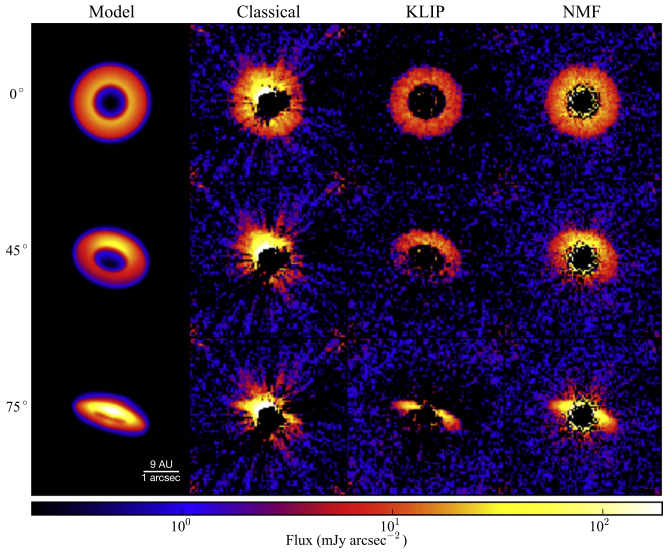

Figure 7. Initial models dimmed by a factor of 10: morphology of disks reduced by different methods for different inclination angles with $F_{\rm disk}/F_{\rm star} = (1.5, 0.9, 1.9) \times 10^{-5}$ from top to bottom. The classical method is working poorly, and NMF works better than KLIP in the sense of recovering faint signals (far side of the tilted disks).

comparing with KLIP, NMF is able reach lower or similar levels, demonstrating its competence in disk retrieval. In the cases when the NMF χ^2 values are slightly larger than that of KLIP, it is from the fact that KLIP is overfitting the random noise, which in principle should not be fitted by any method, rather than KLIP having a better matching to the disk model.

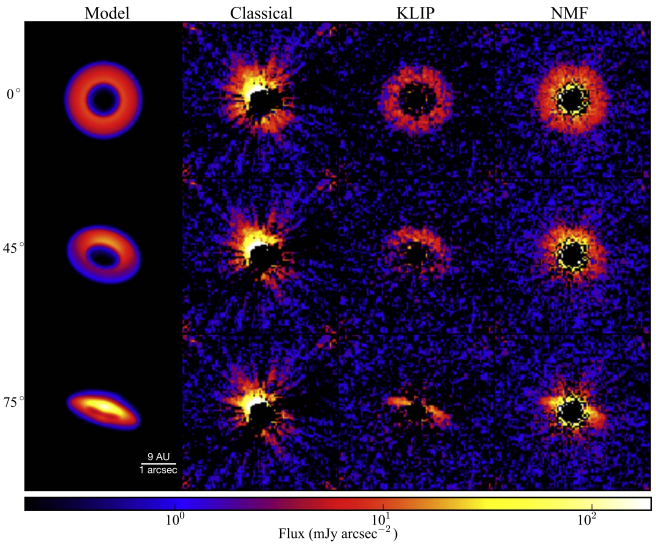

Figure 8. Initial models dimmed by a factor of 20: morphology of disks reduced by different methods for different inclination angles with $F_{\rm disk}/F_{\rm star} = (7.4, 4.8, 9.0) \times 10^{-6}$, respectively. The classical method is not working. Both KLIP and NMF recover the geometries; however, NMF preserves the morphology and flux better than KLIP.

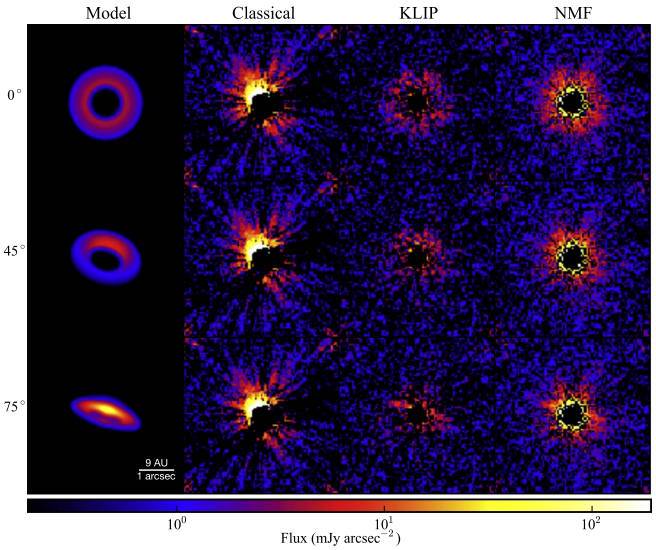

Figure 9. Initial models dimmed by a factor of 50: morphology of disks reduced by different methods for different inclination angles with $F_{\rm disk}/F_{\rm star} = (3.1, 1.9, 3.7) \times 10^{-6}$. The disks are too faint in this case; none of the methods could recover the flux of the disks properly, but NMF is still able to marginally recover the morphology.

3.2.2. Radial Profile

For a face-on disk, its radial profile informs us of the spatial distribution of the amount of materials, which should be recovered faithfully. We compare the three methods from the radial profiles shown in Figure 11 for the recovered face-on disks in Figure 7.

Classical subtraction seems to be able to recover the radial profile of the face-on disk at first glance, but it has large uncertainties. This is because we calculate the uncertainties from the standard deviation of pixels at the same radial

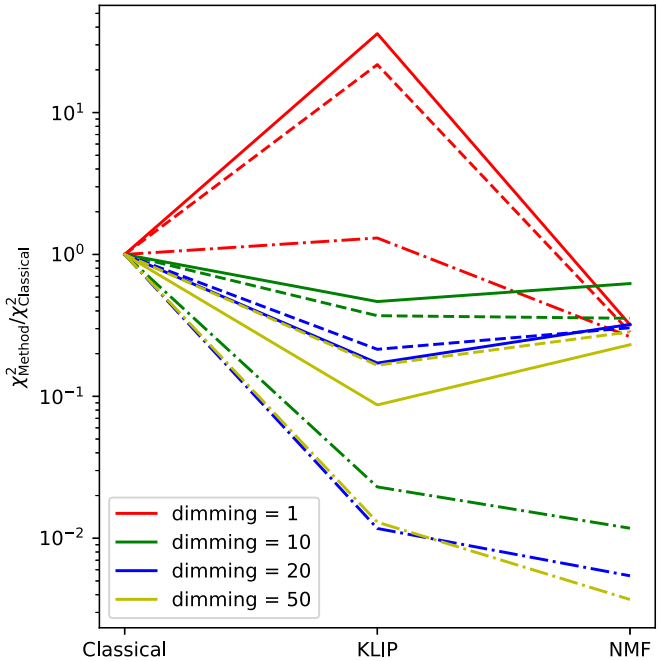

Figure 10. $\chi^2_{\rm Method}/\chi^2_{\rm Classical}$ ratios for different methods and dimming levels in Figures 6–9. The solid lines are for the face-on disk, the dashed lines are for the disks tilted at 45°, and the dot-dashed lines are for the disks tilted at 75°. NMF is able to perform better than the classical method in the χ^2 sense; in comparison with KLIP, NMF is able to reach lower or similar levels of χ^2 values.

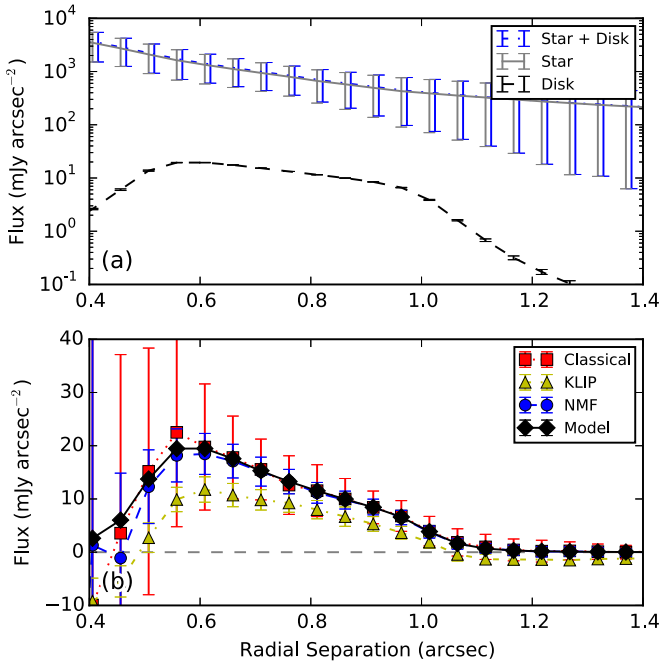

Figure 11. Radial profiles for the face-on disks in Figure 7. (a) Radial profiles of the star, the face-on MCFOST disk model, and the target (star added with disk). Blue dot-dashed line: the target; gray solid line: the star; black dashed line: the disk. The face-on disk is ∼100 times fainter than the stellar PSF wing. (b) Radial profiles of the disk reduced with different methods. Black diamond with solid line: disk model; red square with dotted line: classical subtraction; yellow triangle with dot-dashed line: KLIP result; blue circle with dashed line: NMF result. KLIP is overfitting, therefore introducing unphysical negative pixels in the outskirts (radial separation of more than 1.0 arcsec), but the radial profiles of classical and NMF results are both consistent with the model, whereas NMF performs better with smaller uncertainties.

separations. From the morphology results indicating that classical reduction is not working, this radial profile with large uncertainty is not physically capable for further analysis.

KLIP is not able to recover the radial profile of the face-on disk. This results from the overfitting of the astrophysical signals (as discussed in Section 3.1): not only is KLIP unable to recover the flux correctly, but it is also changing the slope of the radial profile, and forward modeling has to be implemented to recover the distribution. Although the error bars of KLIP are smaller, this is a result from overfitting the signals.

NMF not only recovers the radial profile with no bias but also has small uncertainties. With small uncertainties, NMF is expected to detect fainter structures than any other method is able to. Therefore, the NMF results can be used to perform detailed analysis (e.g., Stark et al. 2014).

In this subsection, we have demonstrated that NMF outperforms current methods with synthetic circumstellar disks both in morphology and in radial profile. In the next subsection, we will apply NMF to a specific case when the classical method works, ensuring the reliability of NMF using a well-characterized disk.

3.3. Application to HST-STIS Observations: HD 181327

Unseen planets are able to perturb the circumstellar disk structure and create observable signatures (e.g., Jang-Condell & Boss 2007; Jang-Condell & Turner 2012, 2013; Dong et al. 2015a, 2015b; Zhu et al. 2015), and faithful recovery of both the morphology and radial profile of circumstellar disks is able to constrain the mass of these hypothetical planets (e.g., Rodigas et al. 2014; Nesvold et al. 2016; Dong & Fung 2017). In this subsection, we aim at checking the effectiveness of NMF with a known circumstellar disk surrounding HD 181327, which ensures the reliability of NMF by comparing with a well-characterized disk.

We obtain all public *HST*-STIS coronagraphic observations available in 2016 December from the *HST* archive,[16] and we focus on the Wedge A0.6 position, then align the images with `centerRadon`, and classify the exposures into two categories as in Ren et al. (2017): *targets* that have infrared (IR) excess in their spectral energy distributions (Chen et al. 2014), where the IR excess is expected to emit from the circumstellar disks, and *references* that do not have IR excess. After constructing the NMF components using the references as described in Section 2.2.1, we center on the observations of HD 181327 (Proposal ID: 12228;[17] PI: G. Schneider), which is located at 51.8 pc and is known to host a relatively bright disk with $F_{\rm disk}/F_{\rm star} = 1.7 \times 10^{-3}$ (Schneider et al. 2014); the disk is composed of a nearly face-on primary ring and a faint asymmetric debris structure to the northwest (Stark et al. 2014).

To compare NMF with classical RDI and KLIP subtractions, we obtain the classical reduction result from Schneider et al. (2014), the KLIP result with the 10% closest-matching references in the L^2 sense (Ren et al. 2017), and the NMF result using the whole reference cube. We deproject and correct for the distance-dependent illumination factor, as in Stark et al. (2014), and show the results in Figure 12: the primary disk is clearly seen in all results, while the KLIP disk is systematically

[16] http://archive.stsci.edu/hst/search.php
[17] https://archive.stsci.edu/proposal_search.php?mission=hst&id=12228

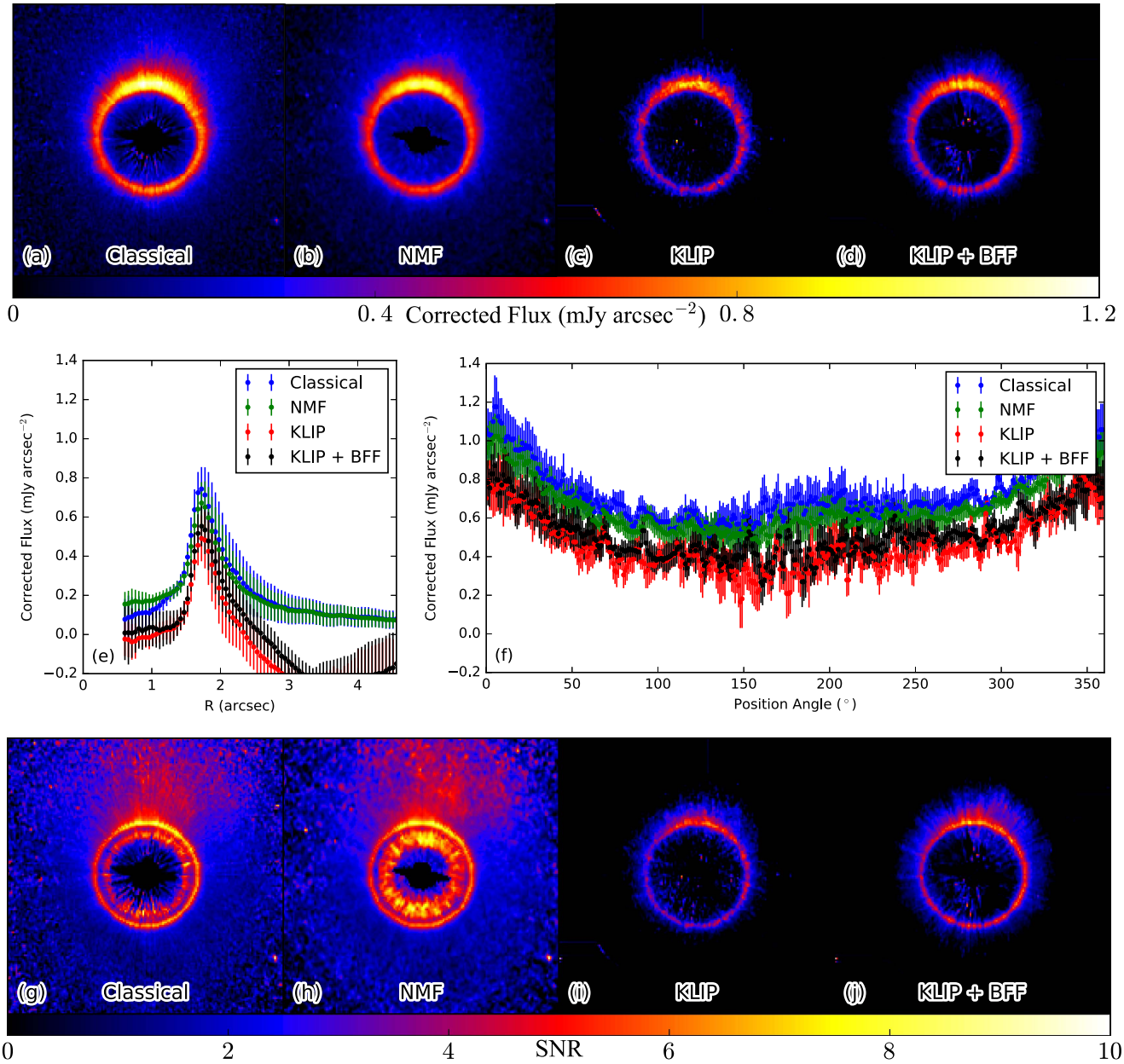

Figure 12. Comparison of HD 181327 STIS disk reduced with classical subtraction, NMF, and KLIP (image dimension: 181 × 181 pixels, 9″.18 × 9″.18). (a) Classical subtraction result of Schneider et al. (2014), deprojected and illumination corrected (i.e., r^2 corrected) as a pseudo-face-on disk using the ellipse parameters as in Stark et al. (2014). (b) NMF subtraction result, corrected in the same way as in panel (a). (c) KLIP subtraction result, corrected in the same way as in panels (a) and (b). (d) KLIP subtraction result, corrected with BFF, then in the same way as in panels (a) and (b). (e) Radial profiles for panels (a)–(d). (f) Azimuthal profiles at the peak of the ring; parameters taken from Stark et al. (2014). Panels (g)–(j) are the signal-to-noise ratio maps for panels (a)–(d), respectively—in the close-in regions (inside the primary ring), NMF is able to reach higher signal-to-noise ratio than the other methods. The results in panels (e) and (f) show that the NMF and classical results are mainly consistent within 1σ. The KLIP results are systematically fainter owing to the overfitting of KLIP; even when corrected with the BFF procedure, the KLIP result is not convincing in either aspect.

dimmer than the other two methods, which is the result from the overfitting of the disk.

In terms of the morphology of the HD 181327 disk, the faint debris at the northwest region is only revealed in the classical and NMF results. Although KLIP is able to extract the primary ring, the debris is buried in the unphysical negative regions. This debris structure is studied in Stark et al. (2014) and is probably caused by a recent catastrophic destruction of an object with mass greater than $0.01M_{\rm Pluto}$. With NMF, we are able to faithfully recover this structure. In situations of fainter disks, NMF is the only method that can extract the disks and retain their morphology.

Focusing on the azimuthal profile for the HD 181327 disk at the peak radial position, the majority of the NMF result agrees with the classical subtraction result both in the values and in the shape within 1σ level, indicating that the NMF result is capable of being studied in the classical way as in Stark et al. (2014), while KLIP is not consistent with the other results in either aspect. We notice that the NMF result is slightly dimmer than the classical result for the HD 181327 primary ring, which

might be caused by the fact that (1) the classical subtraction may not be absolutely correct, or (2) the BFF procedure needs diskless pixels to find the optimum scaling factor, while the faint debris around HD 181327, especially the northwest debris, might be biasing the result. However, we do not aim to argue which result better represents the disk signal, since the major purpose of this subsection is to demonstrate the excellence of NMF on well-characterized bright disks.

Another difference appears inside the primary ring of the disk. With a large number of references, NMF is able to better model the region near the inner working angle—for the region inside the primary ring in Figure 12, the NMF result has greater signal-to-noise ratio than the classical one: the region inside the primary ring is nonzero at significance levels better than 1σ, which calls for speculation of the existence of some scattered light. The scattering materials might originate from the primary ring and then be dragged inward by the gas in the system (Marino et al. 2016), or by the Poynting–Robertson drag from the radiation of the host star.

4. Summary

In the post-processing of direct imaging data, the most important step is to find the best template of the stellar PSF and the speckles for a target image. Especially for broadband imaging instruments such as *HST*-STIS, due to the response of its filter, an ideal template to the target is a reference star with an identical spectral type. However, even if there is an ideal match, the quasi-static noise caused by either the adaptive optics systems or telescope breathing will change the PSF of both the target and the reference. To capture these quasi-static noises, multiple statistical methods have been proposed, and they are working efficiently. Current advanced post-processing methods do excel in finding circumstellar disks, but their disadvantages are preventing us from studying the detailed morphology of these systems.

To extract disk signals with reference differential imaging, we have demonstrated that NMF is an excellent method for capturing the stellar PSF and speckle noises. In this paper, we first compared NMF with current methods using synthetic faint disks and demonstrated that NMF supersedes current methods both in retrieving disk morphology and in photometry; we then applied NMF to a bright disk whose morphology is well studied with the classical subtraction method, ensuring that NMF is working in the most classical examples.

We propose to use NMF to overcome the limitations of current post-processing methods in extracting signals from circumstellar disks, especially to minimize oversubtraction, thus circumventing the tedious forward-modeling attempts. We summarize the properties of NMF as follows:

NMF does not need reference selection to detect circumstellar disks.[18] For broadband imaging instruments, as long as a reference library with all spectral types is given, NMF will construct the components and then find the best combination of components to model the targets. The NMF component basis can be constructed only once but works for all different targets, until new references are taken, the new basis will converge promptly with the old one used as initialization. This will be the dominant advantage of NMF in current and future big-data astronomy, e.g., the surveying telescopes such as the *Wide Field Infrared Survey Telescope*. This iterative approach will need more computational time,[19] but the gain is excellent as demonstrated in this paper.

NMF can extract disk signals and keep their morphology. The utilization of NMF will enable the study of the detailed structures and morphology of circumstellar disks, especially for the faint ones. With well-constrained disk morphology, we will be able to better study the formation, evolution, and even the planet–disk interaction of planetary systems.

With NMF, we can accomplish two goals for extracting circumstellar disks through post-processing of imaging data in this paper: detecting faint signals scattered from the disks, and recovering the morphology of them. Although our paper utilizes space-based coronagraphic observations for their excellent imaging stability, NMF is capable of capturing the varying stellar PSF and speckles from the ground-based exposures (Bin Ren et al. 2018) opening up a new way to better characterize the circumstellar disks.

The authors thank the anonymous referee for the useful suggestions and comments. This work is based on observations made with the NASA/ESA *Hubble Space Telescope* and obtained from the Hubble Legacy Archive, which is a collaboration between the Space Telescope Science Institute (STScI/NASA), the Space Telescope European Coordinating Facility (ST-ECF/ESA), and the Canadian Astronomy Data Centre (CADC/NRC/CSA). B.R. acknowledges the useful discussions with Christopher Stark, Cheng Zhang, and Jason Wang; the comments and suggestions from François Ménard and Christophe Pinte; the classical subtraction result of HD 181327 provided by Glenn Schneider; and computational resources from the support by Colin Norman and the Maryland Advanced Research Computing Center (MARCC). MARCC is funded by a State of Maryland grant to Johns Hopkins University through the Institute for Data Intensive Engineering and Science (IDIES). G.B.Z. acknowledges support provided by NASA through Hubble Fellowship grant no. *HST*-HF2-51351 awarded by the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Inc., for NASA, under contract NAS 5-26555. G.D. acknowledges funding from the European Commission's seventh Framework Program (contract PERG06-GA-2009-256513) and from Agence Nationale pour la Recherche (ANR) of France under contract ANR-2010-JCJC-0504-01.

Facility: HST (STIS).

Software: `centerRadon`: center-determination code for stellar images; `MCFOST`: Radiative transfer code for circumstellar disk modeling; `NonnegMFPy`: Vectorized non-negative matrix factorization code; `nmf_imaging`: Application of `NonnegMFPy` on high-contrast imaging.

Appendix A
List of Symbols

In this Appendix, we present in Table 1 the list of symbols used in this paper.

[18] We still need the images of reference stars. For further detailed analyses, reference selection is preferred to get better results; otherwise, the spectral types of stars should be evenly sampled.

[19] For 72 images of dimension 87×87 pixels, the component construction time of NMF is $\sim 0.3n$ minutes using four cores of Intel Xeon E5-2680v3 (2.5 GHz), where n is the number of components that are constructed sequentially, while the KLIP process takes less than 5 s to construct all the components with one core.

Table 1
List of Symbols

Symbol	Expression	Dimension	Meaning
$\circ$	$(A \circ B)_{ij} = A_{ij} B_{ij}$		Element-wise (Hadamard) multiplication for matrices A and B of same dimension.
D		$1 \times N_{\rm pix}$	Flattened image of the astrophysical signal (i.e., no stellar information).
$\hat{D}$	$T - \hat{f}\, T_{\rm NMF}$	$1 \times N_{\rm pix}$	Reduced best image of the astrophysical signal (D), obtained from BFF procedure.
D_f	$T - f T_{\rm NMF}$	$1 \times N_{\rm pix}$	Reduced image of the astrophysical signal with scaling factor f.
$D_{\rm NMF}$	$\omega^{(D)} H$	$1 \times N_{\rm pix}$	NMF model of the astrophysical signal (D).
$\delta(\cdot)$			The change of the $(\cdot)$ item after one iteration.
$F_{\rm disk}/F_{\rm star}$			Flux ratio between the disk and the star.
f			Scaling factor, where $0 < f < 1$.
$\hat{f}$			Optimum scaling factor obtained from the BFF procedure, corresponding with $\hat{D}$.
H, $H^{(k)}$, $H^{(k+1)}$	$[h_n^T, \cdots, H_n^T]^T$	$n \times N_{\rm pix}$	NMF component matrix for the reference cube.
h_n, H_i, H_n		$1 \times N_{\rm pix}$	The first, ith, and nth NMF component for the reference cube (R).
$(\cdot)^{(k)}$, $(\cdot)^{(k+1)}$	superscript		Iteration step number.
$\mu_f^{(k)}$			The median of the pixels in D_f at iteration step k.
$N_{\rm pix}$			Number of pixels in each image.
$N_{\rm ref}$			Number of images in the reference cube (R).
n			Number of NMF components.
$o(\cdot)$			Little o notation, meaning $\lvert o(\cdot)\rvert \ll \lvert(\cdot)\rvert$. Vectorized form means element-wise o's.
$p_{A,B}$	$AB^T/(BB^T)$		The projection coefficient of column vector A onto column vector B.
R	$[S_1^T, S_2^T, \cdots, S_{N_{\rm ref}}^T]^T$	$N_{\rm ref} \times N_{\rm pix}$	Reference cube with rows containing flattened references.
$\sigma_f^{(k)}$, $\sigma_f^{(\rm conv)}$			The standard deviation for the pixels in D_f at step k, or when the BFF procedure converges.
S, S_i		$1 \times N_{\rm pix}$	Flattened image of only a star (S); subscript i denotes the ith star.
$S_{\rm NMF}$	$\omega^{(s)} H$	$1 \times N_{\rm pix}$	NMF model of the star only (S), i.e., no other astrophysical signal is added.
$(\cdot)^T$	$(A^T)_{ij} = A_{ji}$		Transpose operator for matrices.
T	$S + D$	$1 \times N_{\rm pix}$	Flattened image of a target.
$T_{\rm NMF}$	ωH	$1 \times N_{\rm pix}$	NMF model of the target (T).
V		$N_{\rm ref} \times N_{\rm pix}$	Variance of each pixel for the reference cube (R).
v		$1 \times N_{\rm pix}$	Variance of each pixel for the target image (T).
W, $W^{(k)}$, $W^{(k+1)}$		$N_{\rm ref} \times n$	NMF coefficient matrix for the reference cube (R).
ω, $\omega^{(T)}$, $\omega^{(k)}$, $\omega^{(k+1)}$		$1 \times n$	NMF coefficient matrix for the target image (T).
$\omega^{(S)}$		$1 \times n$	NMF coefficient matrix for the stellar image (S).
$\omega^{(D)}$		$1 \times n$	NMF coefficient matrix for the astrophysical signal (D).
ω_1, ω_i, ω_n			The 1st, ith, and nth entry of NMF coefficient matrix for the target image (T).

Appendix B
NMF with Weighting Function

The update rules adopted in this paper are summarized here.

1. Rules for component construction with weighting function (Zhu 2016):

$$W^{(k+1)} = W^{(k)} \circ \frac{(V \circ R) H^{(k)T}}{[V \circ W^{(k)} H^{(k)}] H^{(k)T}}, \tag{10}$$

$$H^{(k+1)} = H^{(k)} \circ \frac{W^{(k)T} (V \circ R)}{W^{(k)T} [V \circ W^{(k)} H^{(k)}]}, \tag{11}$$

where R is the reference cube, V is the variance matrix of the reference cube (if V is not given, an empirical $V = R$ is suggested because of Poisson noise), $H^{(\cdot)}$ is the NMF component matrix for the reference cube, and $W^{(\cdot)}$ is the coefficient matrix for the reference cube. In this paper, these weighted update rules are adopted. For the *HST*-STIS images, the variance matrix of the reference cube is obtained from the square of the error extension in the flat-field `FITS` files (i.e., the `ERR` extension); when the exposures are added with simulated disks (D), we have $V = D + \mathrm{ERR}^2$, where ERR denotes the `ERR` extension.

2. Rule for target modeling with weighting function:

$$\omega^{(k+1)} = \omega^{(k)} \circ \frac{(v \circ T) H^T}{[v \circ \omega^{(k)}] H H^T}, \tag{12}$$

where T is the target, v is the variance matrix of the target, H is the NMF components constructed above, and $\omega^{(\cdot)}$ is the coefficient matrix for the target.

Appendix C
Construction and Stability of the Component Basis

As stated in the main text, we propose to use a scaling factor to correct for the capturing of disk signal by NMF (see Section 2.2.3 for the detailed procedure). The use of the scaling factor is based on the facts that (1) the disk captured by NMF resembles that of the stellar exposure (Section 2.2.3), (2) the target modeling process is linear to the first order (Appendix D), and (3) the target modeling linearity relies on the property that the NMF components are stable through iteration when they are constructed *sequentially*, which is illustrated in this section.

In the component construction with the references, there are two ways: set the number of components and run the iteration in Appendix B directly, i.e., *randomly*; or construct the components by starting from one component, and then use

the corresponding coefficient and component matrices to initiate the construction for two components, ..., as in Zhu (2016), i.e., *sequentially*.

When we sequentially construct the components, we can denote w_n and h_n as the coefficient and component matrices that are already constructed in the previous n steps and use w_{n+1} and h_{n+1} as[20] the additional coefficient and component vectors for the additional component, i.e., when $n+1$ components should be calculated. Then the simple update rules become (for simplicity in this section, we only focus on the simple update rules in Equations (2) and (3); the method and procedure in Appendix C should work in principle with the weighted update rules in Appendix B, with the substitutions in Appendix A3 of Blanton & Roweis 2007), for the coefficient matrix,

$$[w_n, w_{n+1}] \leftarrow [w_n, w_{n+1}] \circ \frac{R[h_n^T, h_{n+1}^T]}{[w_n, w_{n+1}]\begin{bmatrix} h_n \\ h_{n+1}\end{bmatrix}[h_n^T, h_{n+1}^T]} = [w_n, w_{n+1}] \circ \frac{[Rh_n^T, Rh_{n+1}^T]}{[w_n h_n + w_{n+1}h_{n+1}][h_n^T, h_{n+1}^T]}, \tag{13}$$

where the leftward-pointing arrow ($\leftarrow$) is a simplified notation of the updating procedure, where the left-hand side is the result in the $(k+1)$th step and the right-hand side contains the results from the kth step. For the component matrix,

$$\begin{bmatrix} h_n \\ h_{n+1}\end{bmatrix} \leftarrow \begin{bmatrix} h_n \\ h_{n+1}\end{bmatrix} \circ \frac{\begin{bmatrix} w_n^T \\ w_{n+1}^T\end{bmatrix} R}{\begin{bmatrix} w_n^T \\ w_{n+1}^T\end{bmatrix}[w_n, w_{n+1}]\begin{bmatrix} h_n \\ h_{n+1}\end{bmatrix}} = \begin{bmatrix} h_n \\ h_{n+1}\end{bmatrix} \circ \frac{\begin{bmatrix} w_n^T R \\ w_{n+1}^T R\end{bmatrix}}{\begin{bmatrix} w_n^T w_n & w_n^T w_{n+1} \\ w_{n+1}^T w_n & w_{n+1}^T w_{n+1}\end{bmatrix}\begin{bmatrix} h_n \\ h_{n+1}\end{bmatrix}}. \tag{14}$$

Focusing on the individual matrices, we have

$$w_n \leftarrow w_n \circ \frac{Rh_n^T}{w_n h_n h_n^T + w_{n+1}h_{n+1}h_n^T} = w_n \circ \frac{Rh_n^T}{w_n h_n h_n^T} \circ \frac{1}{1 + \frac{w_{n+1}h_{n+1}h_n^T}{w_n h_n h_n^T}}, \tag{15}$$

$$h_n \leftarrow h_n \circ \frac{w_n^T R}{w_n^T w_n h_n + w_n^T w_{n+1}h_{n+1}} = h_n \circ \frac{w_n^T R}{w_n^T w_n h_n} \circ \frac{1}{1 + \frac{w_n^T w_{n+1}h_{n+1}}{w_n^T w_n h_n}}, \tag{16}$$

$$w_{n+1} \leftarrow w_{n+1} \circ \frac{Rh_{n+1}^T}{w_n h_n h_{n+1}^T + w_{n+1}h_{n+1}h_{n+1}^T} = w_{n+1} \circ \frac{Rh_{n+1}^T}{w_n h_n h_{n+1}^T} \circ \frac{1}{1 + \frac{w_{n+1}h_{n+1}h_{n+1}^T}{w_n h_n h_{n+1}^T}}, \tag{17}$$

$$h_{n+1} \leftarrow h_{n+1} \circ \frac{w_{n+1}^T R}{w_{n+1}^T w_n h_n + w_{n+1}^T w_{n+1}h_{n+1}} = h_{n+1} \circ \frac{w_{n+1}^T R}{w_{n+1}^T w_n h_n} \circ \frac{1}{1 + \frac{w_{n+1}^T w_{n+1}h_{n+1}}{w_{n+1}^T w_n h_n}}. \tag{18}$$

Given the facts that

(a) when $n = m$, the coefficient matrix w_n and component matrix h_n are already satisfying $R \approx w_n h_n$, and
(b) the new coefficient and component vectors (w_{n+1} and h_{n+1}) are randomly initialized (Zhu 2016): the elements are drawn from a uniform distribution from 0 to 1,

the change of h_n in the *first* iteration of Equation (16) before ($h_n^{\rm old}$) and after ($h_n^{\rm new}$) the inclusion of w_{n+1} and h_{n+1} gives

$$\delta h_n = h_n^{\rm new} - h_n^{\rm old} \tag{19}$$

$$= \left(\frac{1}{1 + \frac{w_n^T w_{n+1}h_{n+1}}{w_n^T w_n h_n}} - 1\right) \circ h_n \circ \frac{w_n^T R}{w_n^T w_n h_n} \tag{20}$$

$$= \left(\frac{1}{1 + \frac{w_n^T w_{n+1}h_{n+1}}{w_n^T w_n h_n}} - 1\right) \circ h_n^{\rm old}. \tag{21}$$

Lemma (*Stability*): For the individual elements ($R_{(\cdot)j}$) in the reference cube, if the $R_{(\cdot)j}$'s are *sufficiently large* (for our purpose, they should have large signal-to-noise ratios), then the update has little impact on the constructed components (i.e., h_n) if we construct the components *sequentially*.

Proof: In high-contrast imaging, if the values of the pixels in the references are large, $R_{(\cdot)j}$ can be represented by

$$R_{(\cdot)j} \gg 1, \tag{22}$$

which, when accompanied with a weighting function as adopted in our paper (see Appendix B, as well as Blanton & Roweis 2007), the pixels should have large signal-to-noise ratios, i.e.,

$$\mathrm{SNR}_{(\cdot)j} = \frac{R_{(\cdot)j}}{\sqrt{V_{(\cdot)j}}} \gg 1. \tag{23}$$

To simplify our derivation, the above representation of signal-to-noise ratio is represented by $R_{(\cdot)j}$ in this section. This simplification is in principle valid following the substitution as in Blanton & Roweis (2007).

Assuming that before the inclusion of the additional component, i.e., the $(m+1)$th component (represented by w_{n+1} and h_{n+1}), the following relationship is already satisfied during previous iterations (for $n = m$):

$$R_{kj} \approx (w_n h_n)_{kj}, \tag{24}$$

[20] Note: the definition of symbols (i.e., w_n, h_n, w_{n+1}, h_{n+1}) here is only valid in this section for simplification (Appendix C), and it is not included in the table of symbols in Appendix A.

then given that w_{n+1} and h_{n+1} are randomly initialized (drawn from a uniform distribution from 0 to 1), we have

$$(w_{n+1}h_{n+1})_{kj} \leqslant 1. \tag{25}$$

Combining Equations (22)–(25), we have

$$\frac{(w_{n+1}h_{n+1})_{kj}}{(w_n h_n)_{kj}} \ll 1, \tag{26}$$

and if this inequality is written in the little o notation (i.e., $|o(x)| \ll |x|$), we have

$$\left(\frac{w_n^T w_{n+1}h_{n+1}}{w_n^T w_n h_n}\right)_{ij} = o(1), \tag{27}$$

where $(w_n^T w_n h_n)_{ij} = \sum_{k=1}^m (w_n)_{ik}(w_n h_n)_{kj}$ is the weighted sum of the pixels at the same position in all the references.

Substituting Equation (27) into Equation (21), we have

$$\delta(h_n)_{ij} = \left(\frac{1}{1+o(1)} - 1\right)(h_n)_{ij}^{\text{old}} \tag{28}$$

$$= \{[1+o(1)]^{-1} - 1\}(h_n)_{ij}^{\text{old}} \tag{29}$$

$$\approxeq -o((h_n)_{ij}^{\text{old}}), \tag{30}$$

and to the first order, we have equality in the above equation. In a vectorized form, we have

$$|\delta h_n| = |-o(h_n)| \ll |h_n|, \tag{31}$$

element-wise, i.e., the addition of an extra component has little impact on the previously constructed components.

Appendix D
Target Modeling

When the NMF components are stable through iteration, as illustrated in Appendix C, we are able to demonstrate the linearity of NMF modeling in this section. Assuming that there are n components chosen to model a target (T), then the ith entry of the coefficient matrix ω ($i = 1, \cdots, n$) in update rule Equation (4) is

$$\omega_i^{(k+1)} = \omega_i^{(k)} \frac{TH_i^T}{\sum_{j=1}^n \omega_j^{(k)} H_j H_i^T} \tag{32}$$

$$= \frac{TH_i^T}{H_i H_i^T} \frac{1}{1 + \sum_{j=1, j\neq i}^n \frac{\omega_j^{(k)}}{\omega_i^{(k)}} \frac{H_j H_i^T}{H_i H_i^T}} \tag{33}$$

$$< \frac{TH_i^T}{H_i H_i^T}, \tag{34}$$

where the superscripts $^{(k)}$ and $^{(k+1)}$ are the iteration numbers, H_i is the ith component of H, and ω_i is the ith entry of the coefficient matrix ω. On the right-hand side of Equation (34), this represents the coefficient of the projection of vector T onto vector H_i.

Inequality (34) arises from the fact that all the components are nonpositive; therefore, the denominator of the second term in Equation (33) is always larger than 1. This is the evidence why NMF is less prone to overfitting—the NMF coefficients always have smaller absolute values than normal projections: since NMF the elements are always non-negative, the normal projection coefficients are always equal to their absolute values.

When Equation (33) converges (i.e., $|\delta\omega_i| \ll |\omega_i|$, or $\delta\omega_i = \omega_i^{(k+1)} - \omega_i^{(k)} = o(\omega_i)$, is satisfied), it will have a form of

$$\omega_i = \frac{TH_i^T}{H_i H_i^T} \frac{1}{1 + \sum_{j=1, j\neq i}^n \frac{\omega_j}{\omega_i} \frac{H_j H_i^T}{H_i H_i^T}} + o(\omega_i); \tag{35}$$

for simplicity, when we replace the projections with definition

$$p_{AB} = \frac{AB^T}{BB^T}, \tag{36}$$

Equation (35) becomes

$$\omega_i = p_{TH_i} \frac{1}{1 + \sum_{j=1, j\neq i}^n \frac{\omega_j}{\omega_i} p_{H_j H_i}} + o(\omega_i)$$

$$= \frac{\omega_i p_{TH_i}}{\sum_{j=1}^n \omega_j p_{H_j H_i}} + o(\omega_i). \tag{37}$$

Since all the ω_i's are non-negative, dividing both sides by ω_i, the above equation becomes

$$\sum_{j=1}^n \omega_j p_{H_j H_i} = p_{TH_i} + o(p_{TH_i}). \tag{38}$$

Given $T = S + D$, and using them as superscripts, we can substitute the equation into Equation (38) and obtain

$$\begin{cases} \sum_{j=1}^n \omega_j^{(T)} p_{H_j H_i} = p_{TH_i} + o(p_{TH_i}) \\ \sum_{j=1}^n \omega_j^{(S)} p_{H_j H_i} = p_{SH_i} + o(p_{SH_i}) \\ \sum_{j=1}^n \omega_j^{(D)} p_{H_j H_i} = p_{DH_i} + o(p_{DH_i}) \end{cases}, \tag{39}$$

and in addition, since

$$p_{TH_i} = \frac{TH_i^T}{H_i H_i^T} = \frac{(S+D)H_i^T}{H_i H_i^T} = p_{SH_i} + p_{DH_i}, \tag{40}$$

we have

$$\sum_{j=1}^n [\omega_j^{(T)} - (\omega_j^{(S)} + \omega_j^{(D)})] p_{H_j H_i} = o(p_{TH_i}). \tag{41}$$

Theorem (*Linearity*): The NMF target modeling process is linear to the first order when the NMF components are created sequentially and stable under iterations (i.e., when Equation (31) in Lemma holds).

Proof: The above equation is equivalent to proving

$$\omega_j^{(T)} = \omega_j^{(S)} + \omega_j^{(D)} + o(\omega_j^{(T)}). \tag{42}$$

Now we prove the above equation by way of induction:

A. $n = 1$, since $p_{h_n h_n} > 0$, for $i = 1$, we have

$$[\omega_1^{(T)} - (\omega_1^{(S)} + \omega_1^{(D)})] p_{h_n h_n} = o(p_{Th_n}), \tag{43}$$

$$\omega_1^{(T)} - (\omega_1^{(S)} + \omega_1^{(D)}) = o(p_{Th_n}/p_{h_n h_n}) = o(\omega_1^{(T)}), \tag{44}$$

Equation (42) holds.

B. Assume for $n = m$, Equation (42) holds, and we also have the following equation holds (Equation (41), for

$i = 1,\cdots,m$):

$$\sum_{j=1}^{m}[\omega_j^{(T)} - (\omega_j^{(S)} + \omega_j^{(D)})]p_{H_jH_i} = o(p_{TH_i}). \tag{45}$$

C. For $n = m + 1$, given the fact that the components do not vary to the first order when the number of components increases (Appendix C, Conclusion), for $i = 1,\cdots, m$, Equation (41) becomes

$$\begin{aligned} o(p_{TH_i}) &= \sum_{j=1}^{m+1}\{\omega_j^{(T)} - [\omega_j^{(S)} + \omega_j^{(D)}]\}p_{[H_j+o(H_j)][(H_i+o(H_i)]} \\ &= \sum_{j=1}^{m}\{\omega_j^{(T)} - [\omega_j^{(S)} + \omega_j^{(D)}]\} \\ &\quad \times [p_{H_jH_i} + 2o(p_{H_jH_i}) + o^2(p_{H_jH_i})] \\ &\quad + \{\omega_{m+1}^{(T)} - [\omega_{m+1}^{(S)} + \omega_{m+1}^{(D)}]\} \\ &\quad \times [p_{H_{m+1}H_i} + o(p_{H_{m+1}H_i})] \\ &= o(p_{TH_i}) + \sum_{j=1}^{m}\{\omega_j^{(T)} - [\omega_j^{(S)} + \omega_j^{(D)}]\} \\ &\quad \times [2o(p_{H_jH_i}) + o^2(p_{H_jH_i})] \\ &\quad + \{\omega_{m+1}^{(T)} - [\omega_{m+1}^{(S)} + \omega_{m+1}^{(D)}]\} \\ &\quad \times [p_{H_{m+1}H_i} + o(p_{H_{m+1}H_i})], \end{aligned}$$

where Equation (45) is substituted into Equation (41) in the above derivation. Since $p_{H_{m+1}H_i}$ is a simple number rather than a vector, and Equation (42) holds, by keeping up to the first order, we have

$$\begin{aligned} &\omega_{m+1}^{(T)} - [\omega_{m+1}^{(S)} + \omega_{m+1}^{(D)}] \\ &= \frac{o(p_{TH_i}) - \sum_{j=1}^{m}\{\omega_j^{(T)} - [\omega_j^{(S)} + \omega_j^{(D)}]\}[2o(p_{H_jH_i}) + o^2(p_{H_jH_i})]}{p_{H_{m+1}H_i} + o(p_{H_{m+1}H_i})} \end{aligned} \tag{46}$$

$$= \frac{o(p_{TH_i}) - 2o^2(p_{TH_i}) - o^3(p_{TH_i})}{p_{H_{m+1}H_i} + o(p_{H_{m+1}H_i})} = \frac{o(p_{TH_i})}{p_{H_{m+1}H_i}} \tag{47}$$

$$= o(\omega_{m+1}^{(T)}), \tag{48}$$

which is also true when $i = m + 1$; therefore, the proof of Equation (42) is complete.

Rewriting Equation (42) in vector form, we have

$$\omega^{(T)} = \omega^{(S)} + \omega^{(D)} + o(\omega^{(T)}), \tag{49}$$

and thus

$$T_{\rm NMF} = \omega^{(T)}H = \omega^{(S)}H + \omega^{(D)}H + o(\omega^{(T)}H) \tag{50}$$

$$= S_{\rm NMF} + D_{\rm NMF} + o(T_{\rm NMF}), \tag{51}$$

i.e., to the first order, we can linearly separate the stellar PSF and speckles from the circumstellar disk signal.

Appendix E
The BFF Procedure

We notice that when the optimum scaling factor is in effect (Section 2.2.3), the diskless regions should be well modeled by the NMF model of the target; therefore, the values on these pixels should be small and have a histogram distribution that is symmetric about 0, and consequently, the variation of the noise of the diskless region should be minimized. We thus introduce the BFF procedure as follows to find this factor:

1. For each target (T), construct its NMF model ($T_{\rm NMF}$) with the component basis (H), and then vary the scaling factor (f) from 0 to 1, creating several scaled reduced images ($D_f = T - fT_{\rm NMF}$).
2. For each scaled reduced image (D_f),
   (a) Identification of the background region iteratively: in each iteration (k), find the median ($\mu_f^{(k)}$) and standard deviation ($\sigma_f^{(k)}$) of D_f, remove the pixels having values satisfying condition

   $$\text{value} > \mu_f^{(k)} + 3\sigma_f^{(k)} \quad \text{or} \quad \text{value} < \mu_f^{(k)} - 10\sigma_f^{(k)};$$

   these pixels are treated as nonbackground ones because of their large deviations from the median. Repeat this process until the number of background pixels does not change.
   (b) Calculation of the noise of the diskless region: calculate the standard deviation of the remaining pixels when step (a) converges, and denote it by $\sigma_f^{(\rm conv)}$.
3. The factor corresponding with the minimum standard deviation of D_f of the diskless pixels will be taken as the best one ($\hat{f}$), i.e.,

   $$\hat{f} = \text{argmin}_f\, \sigma_f^{(\rm conv)}.$$

The connection between BFF and the classical optimum scaling factor is that both of them are minimizing the residual noise. In comparison, the classical method minimizes the residual noise along the major diffraction spikes after PSF subtraction (e.g., Schneider et al. 2009). When the diffraction spikes are not stable, especially for ground-based observations, BFF is able to focus on the entire field of view and is more able to minimize the overall difference between the PSF template and the target.

ORCID iDs

Bin Ren (任彬) https://orcid.org/0000-0003-1698-9696
Guangtun Ben Zhu https://orcid.org/0000-0002-7574-8078
John Debes https://orcid.org/0000-0002-1783-8817
Gaspard Duchêne https://orcid.org/0000-0002-5092-6464

References

Amara, A., & Quanz, S. P. 2012, MNRAS, 427, 948
Biller, B. A., Close, L., Lenzen, R., et al. 2004, Proc. SPIE, 5490, 389
Bin Ren, B., Dong, R., Esposito, T. M., et al. 2018, ApJ, submitted
Blanton, M. R., & Roweis, S. 2007, AJ, 133, 734
Chen, C. H., Mittal, T., Kuchner, M., et al. 2014, ApJS, 211, 25
Choquet, É., Milli, J., Wahhaj, Z., et al. 2017, ApJL, 834, L12
Choquet, É., Perrin, M. D., Chen, C. H., et al. 2016, ApJL, 817, L2
Choquet, É, Pueyo, L., Hagan, J. B., et al. 2014, Proc. SPIE, 9143, 914357
Debes, J. H., Jang-Condell, H., Weinberger, A. J., Roberge, A., & Schneider, G. 2013, ApJ, 771, 45
Debes, J. H., Poteet, C. A., Jang-Condell, H., et al. 2017, ApJ, 835, 205
Dong, R., & Fung, J. 2017, ApJ, 835, 146
Dong, R., Zhu, Z., Fung, J., et al. 2016, ApJL, 816, L12
Dong, R., Zhu, Z., Rafikov, R. R., & Stone, J. M. 2015a, ApJL, 809, L5
Dong, R., Zhu, Z., & Whitney, B. 2015b, ApJ, 809, 93
Esposito, T. M., Fitzgerald, M. P., Graham, J. R., & Kalas, P. 2014, ApJ, 780, 25
Follette, K. B., Rameau, J., Dong, R., et al. 2017, AJ, 153, 264

Gomez Gonzalez, C. A., Wertz, O., Absil, O., et al. 2017, AJ, 154, 7
Grady, C. A., Hamaguchi, K., Schneider, G., et al. 2010, ApJ, 719, 1565
Grady, C. A., Muto, T., Hashimoto, J., et al. 2013, ApJ, 762, 48
Hinkley, S., Oppenheimer, B. R., Soummer, R., et al. 2007, ApJ, 654, 633
Jang-Condell, H., & Boss, A. P. 2007, ApJL, 659, L169
Jang-Condell, H., & Turner, N. J. 2012, ApJ, 749, 153
Jang-Condell, H., & Turner, N. J. 2013, ApJ, 772, 34
Krist, J. E., Hook, R. N., & Stoehr, F. 2011, Proc. SPIE, 8127, 81270J
Lafrenière, D., Marois, C., Doyon, R., & Barman, T. 2009, ApJL, 694 L148
Lafrenière, D., Marois, C., Doyon, R., Nadeau, D., & Artigau, É 2007, ApJ, 660, 770
Lee, D. D., & Seung, H. S. 2001, in Advances in Neural Information Processing Systems 13, ed. T. K. Leen, T. G. Dietterich, & V. Tresp (Cambridge, MA: MIT Press), 556
Lee, E. J., & Chiang, E. 2016, ApJ, 827, 125
Marino, S., Matrà, L., Stark, C., et al. 2016, MNRAS, 460, 2933
Marois, C., Lafrenière, D., Doyon, R., Macintosh, B., & Nadeau, D. 2006, ApJ, 641, 556
Marois, C., Macintosh, B., & Véran, J.-P. 2010, Proc. SPIE, 7736, 77361J
Mawet, D., Pueyo, L., Lawson, P., et al. 2012, Proc. SPIE, 8442, 844204
Mazoyer, J., Boccaletti, A., Augereau, J.-C., et al. 2014, A&A, 569, A29
Mazoyer, J., Boccaletti, A., Choquet, É., et al. 2016, ApJ, 818, 150
Milli, J., Lagrange, A.-M., Mawet, D., et al. 2014, A&A, 566, A91
Nesvold, E. R., Naoz, S., Vican, L., & Farr, W. M. 2016, ApJ, 826, 19
Paatero, P., & Tapper, U. 1994, Environmetrics, 5, 111
Perrin, M. D., Graham, J. R., Kalas, P., et al. 2004, Sci, 303, 1345
Perrin, M. D., Graham, J. R., & Lloyd, J. P. 2008, PASP, 120, 555
Pinte, C., Harries, T. J., Min, M., et al. 2009, A&A, 498, 967
Pinte, C., Ménard, F., Duchêne, G., & Bastien, P. 2006, A&A, 459, 797
Pueyo, L. 2016, ApJ, 824, 117
Pueyo, L., Crepp, J. R., Vasisht, G., et al. 2012, ApJS, 199, 6
Pueyo, L., Soummer, R., Hoffmann, J., et al. 2015, ApJ, 803, 31
Ren, B., Pueyo, L., Perrin, M. D., Debes, J. H., & Choquet, E. 2017, Proc. SPIE, 10400, 1040021
Riley, A. 2017, STIS Instrument Handbook, version 16.0 (Baltimore, MD: STScI)
Rodigas, T. J., Malhotra, R., & Hinz, P. M. 2014, ApJ, 780, 65
Schneider, G., Grady, C. A., Hines, D. C., et al. 2014, AJ, 148, 59
Schneider, G., Grady, C. A., Stark, C. C., et al. 2016, AJ, 152, 64
Schneider, G., Weinberger, A. J., Becklin, E. E., Debes, J. H., & Smith, B. A. 2009, AJ, 137, 53
Smith, B. A., & Terrile, R. J. 1984, Sci, 226, 1421
Soummer, R., Ferrari, A., Aime, C., & Jolissaint, L. 2007, ApJ, 669, 642
Soummer, R., Perrin, M. D., Pueyo, L., et al. 2014, ApJL, 786, L23
Soummer, R., Pueyo, L., & Larkin, J. 2012, ApJL, 755, L28
Stark, C. C., Schneider, G., Weinberger, A. J., et al. 2014, ApJ, 789, 58
Traub, W. A., & Oppenheimer, B. R. 2010, in Exoplanets, ed. S. Seager (Tucson, AZ: Univ. Arizona Press), 111
Wahhaj, Z., Cieza, L. A., Mawet, D., et al. 2015, A&A, 581, A24
Wang, J. J., Ruffio, J.-B., De Rosa, R. J., et al. 2015, pyKLIP: PSF Subtraction for Exoplanets and Disks, Astrophysics Source Code Library, ascl:1506.001
Zhu, G. 2016, arXiv:1612.06037
Zhu, Z., Dong, R., Stone, J. M., & Rafikov, R. R. 2015, ApJ, 813, 88