DoG-HiT: A novel VLBI multiscale imaging approach

H. Müller, and A. P. Lobanov

Max-Planck-Institut für Radioastronomie, Auf dem Hügel 69, Bonn 53121, Germany
e-mail: hmueller@mpifr-bonn.mpg.de; alobanov@mpifr-bonn.mpg.de

Received 1 February 2022 / Accepted 14 June 2022

ABSTRACT

Context. Reconstructing images from very long baseline interferometry (VLBI) data with a sparse sampling of the Fourier domain (uv-coverage) constitutes an ill-posed deconvolution problem. It requires application of robust algorithms, maximizing the information extraction from all of the sampled spatial scales, and minimizing the influence of the unsampled scales on image quality.

Aims. We develop a new multiscale wavelet deconvolution algorithm, DoG-HiT, for imaging sparsely sampled interferometric data, which combines the difference of Gaussian (DoG) wavelets and hard image thresholding (HiT). Based on DoG-HiT, we propose a multistep imaging pipeline for analysis of interferometric data.

Methods. DoG-HiT applies the compressed sensing approach to imaging by employing a flexible DoG wavelet dictionary, which is designed to adapt smoothly to the uv-coverage. It uses closure properties as data fidelity terms only, initially, and performs nonconvex, nonsmooth optimization by an amplitude-conserving and total-flux-conserving, hard thresholding splitting. DoG-HiT calculates a multiresolution support as a side product. The final reconstruction is refined through self-calibration loops and imaging with amplitude and phase information applied for the multiresolution support only.

Results. We demonstrate the stability of DoG-HiT, and benchmark its performance against image reconstructions made with the CLEAN and regularized maximum-likelihood (RML) methods using synthetic data. The comparison shows that DoG-HiT matches the super-resolution achieved by the RML reconstructions and surpasses the sensitivity to extended emission reached by CLEAN.

Conclusions. The application of regularized maximum likelihood methods, outfitted with flexible multiscale wavelet dictionaries, to imaging of interferometric data, matches the performance of state-of-the-art convex optimization imaging algorithms and requires fewer prior and user-defined constraints.

Key words. techniques: interferometric – techniques: image processing – techniques: high angular resolution – methods: numerical – galaxies: jets – galaxies: nuclei

1. Introduction

In very long baseline interferometry (VLBI), signals recorded at individual radio antennas are combined (correlated) in order to sample angular scales inversely proportional to pairwise antenna separations projected onto the plane of the incoming wavefront. Described by the van Cittert–Zernike theorem, the correlation product (visibility) of the signals recorded at two antennas at a given time is given by a single spectral harmonic corresponding to a single spatial frequency of the Fourier transform of the observed sky brightness distribution (see Thompson et al. 1994). From a complete sampling of spatial frequencies, the true image could be revealed by the inverse Fourier transform. However, the practical limitations on the number of antennas, observing bandwidth, and observing time often result in a situation where VLBI data provide only sparse sampling (uv-coverage) of the spatial frequencies (or “Fourier domain”), below the Nyquist–Shannon sampling rate.

The development of powerful imaging algorithms such as CLEAN (Högbohm 1974) and their decade long successful application in VLBI studies demonstrated that a reliable reconstruction of the true sky brightness distribution is still possible under a strong assumption that the sky brightness distribution is compressible as a sum of point sources. CLEAN and its many variants (e.g., Clark 1980; Schwab 1984) work well not only for compact structures but also for extended emission. CLEAN is still broadly used, mainly because it is practical. However, frontline VLBI applications, such as millimeter- or space-VLBI, demand better imaging tools that would alleviate the known limitations of CLEAN and provide super-resolution, multiscalar decompositions, and a high dynamic range.

Multiresolution imaging routines based on the greedy matching pursuit procedure, inherent to CLEAN, have been developed for decades now (Wakker & Schwarz 1988; Starck et al. 1994; Bhatnagar & Cornwell 2004; Cornwell 2008; Rau & Cornwell 2011). These studies build upon the great success of compressed sensing theory (e.g., Candès et al. 2006; Donoho 2006), namely that an image can be sparsely represented in a suitable set of basis functions (atoms). Even the CLEAN algorithm (sparsity in pixel basis; Lannes et al. 1997) and total variation regularization methods (sparsity of the Haar wavelet) could be understood in this way.

Imaging algorithms based on wavelets attract close attention from the astronomy community because they stand out as extremely helpful in the analysis and compression of image features on multiple scales (Starck & Murtagh 2006; Starck et al. 2015; Mertens & Lobanov 2015; Line et al. 2020). Both extended emission features and small-scale structures are well compressible with wavelets. Moreover, wavelets of varying scales are sensitive to different ranges of visibilities, allowing the user to incorporate information about the radially distributed positions of gaps in the uv-coverage in the imaging procedure. Hence,
sparsity in the wavelet domain is a strong and interesting image prior for the radio aperture synthesis imaging problem.

The past five decades have seen an ongoing development of regularized maximum likelihood (RML) methods for interferometric imaging, in particular with the development of image entropy regularizers such as the maximum entropy method (MEM; e.g., Frieden 1972; Narayan & Nityananda 1986; Wiaux et al. 2009; Li et al. 2011; Garsden et al. 2015; Thiébaut & Young 2017). The RML methods have been applied particularly extensively for imaging with the Event Horizon Telescope (EHT; e.g., Ikeda et al. 2016; Akiyama et al. 2017b,a; Chael et al. 2018; Event Horizon Telescope Collaboration 2019). In a typical RML application, the image is recovered by simultaneously minimizing a data fidelity term, which measures the proximity of the recovered solution to the true data, namely visibilities and/or closure properties, and a set of regularization terms, which measure the feasibility of the recovered solution. It has been demonstrated that $l_1$-penalty terms promote sparsity in the image domain. Hence, RML algorithms and the progress in convex optimization (Beck & Teboulle 2009; Combettes & Pesquet 2009) provide a powerful framework for respecting sparsity during image deconvolution.

However, the deficiencies of $uv$-coverages, inherent to such interferometric instruments as the EHT or the space VLBI mission RadioAstron, pose additional challenges. Compressed sensing approaches applied to data from such arrays are capable, in principle, of recovering the significant structure of the target (achieving small data fidelity terms), while suppressing any additional noise-induced image features and sidelobes (achieving small penalty terms). However, in VLBI observations the sidelobes and the true image structure often become comparable in their magnitudes. The suppression of structure due to image sparsity affects the recovered data significantly. A more advanced treatment of image features, in other words a more advanced differentiation between observed emission and noisy structures induced by $uv$-gap, and an amplitude conserving optimization strategy are needed. Furthermore, an unsupervised approach for blind imaging is desired.

Random and systematic noise factors in the final image can be induced at various steps of the analysis. In particular, errors resulting from $uv$-coverage deficiencies and antenna-based noise factors (calibration issues, thermal noise) depend on the location of the trace of the antenna pair in the $uv$-plane. Hence, these errors are scale and direction dependent. We need a novel algorithm that can deal with this, in other words one that can automatically decompose noisy features from signal features. This is a task that is suitable for wavelets in the first instance, since they decompose the image into a sequence of scales. Direction-dependent information is more difficult to compress and will not be addressed in this paper.

In this paper, we present a new multiscale wavelet imaging algorithm built upon the compressed sensing approach. Our method goes beyond standard sparsity, promoting imaging algorithms by applying a more stringent separation of significant image features from noise contributions, by using an adaptive wavelet dictionary and suppressing the noise-induced artifacts in a novel amplitude-conserving hard thresholding algorithm. This algorithm is well suited for dealing with high-level sidelobes such as those typically found in the data from the EHT or space-VLBI observations.

An important feature of the algorithm is that the initial selection of the scales in the wavelet dictionary is derived from the $uv$-coverage of observations, and not from any assumptions about the structure of the target source. We utilize current state-of-the-art optimization algorithms to solve the resulting RML minimization problem, but amend the imaging pipeline by a hard thresholding sparsity term based on the multiresolution support, which allows us to retain necessary image information while suppressing noisy scales. We deal with potential residual calibration deficiencies of the data by first using only the gain-invariant closure quantities for imaging and then, after identifying and suppressing noise contributions, imaging the full data with an optimized, fixed multiresolution support. The resulting objective functional for minimization is not convex and not smooth, which requires employing nonconvex and nonsmooth optimization strategies. We present a final imaging pipeline that is immediately applicable to VLBI data. This imaging pipeline requires considerably fewer parameters than typical RML pipelines, thereby presenting a viable step toward a more unsupervised imaging approach.

We test our pipeline routine on test images that were recently used to verify the modern generation of RML image routines (Tiede et al. 2020). For incomplete $uv$-coverages, our algorithm performs better than the canonical CLEAN algorithm and its multiscale variants, owing to the flexibility of the dictionary (allowing it to be adapted to a specific $uv$-coverage of the array), the sparse representation of astronomical images in the wavelet basis (compared to the representation with the CLEAN or MS-CLEAN components), and the correct treatment of scale-dependent noise properties.

2. Theory

This section summarizes the relevant theory and background for different aspects of the new algorithm. We focus primarily on application of wavelets for deconvolution in aperture synthesis and on specific aspects of the optimization procedures applied to sparsely sampled data.

2.1. Aperture synthesis

In interferometric observations, every antenna in the array records the electromagnetic field of an incoherent sky brightness distribution $I(x,y)$, where $x$ and $y$ are angular coordinates on the sky. Following the van Cittert–Zernike theorem, the cross-correlation between the signals recorded by two antennas over a baseline $(u,v)$ (spatial frequencies in units of wavelengths) is given by the Fourier transform of $I(x,y)$ at this baseline:

$$\mathcal{V}(u,v) = \int \int e^{-2\pi i xu} I(x,y) dx dy,$$

where $\mathcal{V}$ is the complex visibility. This relation holds under assumptions of a flat wavefront and a small field-of-view approximation. Every antenna pair, at a fixed time, gives rise to a specific baseline. The projection of a baseline on a plane orthogonal to the direction to the target shifts smoothly by time due to Earth rotation describing the typical elliptical traces in $uv$-coverages. However, due to the small number of antennas in VLBI arrays, the coverage of measurements in the $uv$-domain remains sparse. In particular, gaps in the $uv$-coverage introduce sidelobes and artefacts in the recovered image. When inverting the Fourier transform (to produce $\hat{I}$), the result can be written as a convolution:

$$\hat{I} = B \ast I,$$

where $\hat{I}$ is the dirty image, meaning the inverse Fourier transform of the (tapered) observed visibilities, and $B$ is the dirty
beam, meaning the inverse Fourier transform of the (tapered) projection onto measured baselines in the Fourier codomain.

Aperture synthesis imaging is the problem of recovering the true distribution \( I(x,y) \) from a discrete, sparse set of observed visibilities. This procedure could also be understood as a deconvolution problem, see Eq. (2). The incomplete u-v-coverage introduces direction- and scale-dependent sidelobe patterns in the dirty image and the dirty beam. Deconvolution, in this case, becomes an ill-posed inverse problem. In particular, the solution to the imaging problem in Eq. (2) is, strictly speaking, not unique as Fourier harmonics are missing from the observation (historically called the "invisible distributions"). A successful deconvolution method must be able to identify and categorize these invisible distributions and minimize their impact on the restored image.

Image restoration is further complicated by the variable thermal noise and signal-to-noise ratio (S/N) of visibility measurements. The visibility S/N is systematically reduced at long baselines. As the antenna sensitivity enters the reconstruction at specific scales and directions, determined by the position of the baseline corresponding to a given antenna pair, the noise becomes scale and direction dependent.

Various calibration issues also need to be addressed during image restoration. Systematic direction-independent calibration errors can be factorized into multiplicative station-based gains \( g_i \) (where the index \( i \) denotes the antenna in the array), affecting the relation between the observed visibilities \( V_{ij} \) and the true visibilities \( V'_{ij} \):

\[
V_{ij} = g_i g_j^* V'_{ij} + N_{ij},
\]

where \( N_{ij} \) denotes thermal noise on the baseline. In particular, phase information is typically only available after a calibration by an ad hoc initial model. In standard imaging approaches (e.g., in CLEAN), the problem of calibration is typically addressed through a hybrid imaging approach. In this case, an initial image is first produced using the a priori set of instrumental gains, and then the gain terms are solved for, as in Eq. (3), in order to enforce consistency with the current image guess (with the solution typically obtained by a gradient descent approach). With this forward modeling approach, RML methods can work directly on the closure quantities or a mix of data products in order to reduce the influence of calibration errors on the reconstruction. Regularization and missing information are dealt with by simultaneously minimizing a penalization term, which promotes the desired image features (i.e., sparsity, smoothness, or small entropy). The resulting minimization problem is then solved by standard numerical optimization algorithms, such as a gradient descent algorithm.

A major advantage of the work presented in this paper is the use of novel basis functions (i.e., wavelets). We will discuss them in more detail in Sect. 2.3. These wavelets are extended and allow a more thorough analysis of the u-v-coverage of the observations. The basis functions used in the (MS-)CLEAN and RML methods do not typically offer this kind of analysis. The standard CLEAN (Högbom 1974) algorithm models the image as a set of delta functions. Its multiscale variants use some version of truncated Gaussian functions (see the discussions in Cornwell 2008). RML methods utilize pixel grids.

2.3. Wavelets

The continuous wavelet transform (CWT) could be understood as an extension of the Fourier transform (Stark et al. 2015), in which the Fourier decomposition in the frequency domain is amended by a windowing of the measurement domain with a specially designed analyzing wavelet function. In the definition of Grossmann et al. (1989), the CWT related to an analyzing wavelet \( \Phi(t) \) operates in one dimension on the space of square integrable functions so that

\[
I \mapsto W(a,b) = \frac{1}{\sqrt{a}} \int I(t) \Phi^* \left( \frac{t-b}{a} \right) dt = I \ast \tilde{\Phi}_a(b),
\]

where \( \tilde{\Phi}_a(t) = \frac{1}{\sqrt{a}} \Phi^{\ast} (\frac{t}{a}) \), \( a \) is the scale parameter, and \( b \) is the position parameter. Hence, the CWT effectively performs a number of convolutions with dilated versions of the analyzing wavelet \( \Phi \). There are different choices for analyzing wavelet...
functions, including the Morlet wavelet (Goupillaud et al. 1984; Couplin et al. 1992), the Haar wavelet (Stollnitz et al. 1994), the Mexican-hat wavelets (Murenzi 1989), and discrete versions (e.g., see Mallat 1989; Starck et al. 2015, and references therein).

In this work we are using difference of Gaussian (DoG) wavelets that are commonly applied to approximate Mexican-hat wavelets (Gonzalez & Woods 2006; Assirati et al. 2014):

\[
\Phi_{\text{DoG}}^{\sigma_0,\sigma_1}(x, y) = \frac{1}{2\pi\sigma_1^2} \exp\left(-\frac{r(x, y)^2}{2\sigma_1^2}\right) - \frac{1}{2\pi\sigma_2^2} \exp\left(-\frac{r(x, y)^2}{2\sigma_2^2}\right) = G_{\sigma_1} - G_{\sigma_2},
\]

where necessarily \(\sigma_1 \leq \sigma_2\), and \(G_{\sigma_j}\) denotes a Gaussian with standard deviation \(\sigma_j\).

Wavelets in the image domain (convolution) directly translate to masks in the Fourier domain (pointwise multiplication):

\[
F\Phi_{\text{DoG}}^{\sigma_0,\sigma_1}(u, v) \propto \exp\left(-2\pi^2\sigma_1^2g(u, v)^2\right) - \exp\left(-2\pi^2\sigma_2^2g(u, v)^2\right),
\]

where \(g(u, v)\) denotes the radius in Fourier domain.

Of special interest for image compression is the discrete wavelet transform, in particular the à trous wavelet transform (also called the starlet transform). In a nutshell, the à trous wavelet transform aims to compute a sequence of smoothing scales \(c_j\) by convolving the image with a discretized smoothing kernel dilated by \(2^j\) pixels, where \(j\) labels the scale and ranges from \(0\) up to a final smoothing scale \(J\). Wavelet scales are defined as the difference of two smoothing scales:

\[
\omega_j = c_j - c_{j+1}.
\]

The last smoothing scale \(c_J\) is added to the set of wavelet scales resulting in the set: \([\omega_0, \omega_1, ..., \omega_{J-1}, c_J]\). This set decomposes the initial image into subbands \(\omega_j\), each of them containing information on spatial scales from \(2^j\rho\) to \(2^{j+1}\rho\), where \(\rho\) is the smallest scale in the image, namely the width of the smoothing kernel (which is often chosen to be close to the pixel scale). The set is complete in the sense that the image at the limiting resolution \(c_0\) can be recovered by summing all scales:

\[
c_0 = \sum_j \omega_j + c_J.
\]

The à trous wavelet transform has a wide range of applications and was successfully applied to radio interferometry previously (e.g., Li et al. 2011; Garsden et al. 2015). However, the à trous wavelet decomposition, by construction, allows only previously (e.g., Li et al. 2011; Garsden et al. 2015). However, even if \(\sigma_0\) is chosen to be close to the pixel scale).

Therefore, we propose constructing a continuous wavelet decomposition out of DoG wavelets in the same way as the à trous wavelet transform was constructed out of a discretized smoothing kernel. We select an ascending sequence of widths \(\sigma_0 \leq \sigma_1 \leq \ldots \leq \sigma_J\) and compute the smoothing scales \(c_j\) by convolution with Gaussians that have widths \(\sigma_j\), namely \(c_j = I \ast G_{\sigma_j}\). The wavelet scales \(\omega_j\) are then set by

\[
\omega_j = c_j - c_{j+1} = I \ast \Phi_{\text{DoG}}^{\sigma_j,\sigma_{j+1}},
\]

which sufficiently approximates the Mexican hat wavelet scales.

We call a set of basis functions in compressed sensing a “dictionary”, while the basis functions themselves are called “atoms” of the dictionary. The term dictionary is also used for the linear mapping that evaluates a coefficient array of these atoms (Starck et al. 2015). The set of DoG wavelet functions \(\Phi_{\text{DoG}}^{\sigma_j,\sigma_{j+1}}\), together with the last smoothing scale \(G_{\sigma_J}\), builds a multiscale dictionary \(\Gamma\):

\[
\Gamma : (I_0, I_1, I_2, ..., I_J) \mapsto \sum_{j=0}^{J-1} \Phi_{\text{DoG}}^{\sigma_j,\sigma_{j+1}} \ast I_j + G_{\sigma_J} \ast I_J.
\]

The atoms of the dictionary \(\Gamma\) are the wavelets \(\Phi_{\text{DoG}}^{\sigma_j,\sigma_{j+1}}\) and \(G_{\sigma_J}\). By construction, see also Eq. (10), all atoms in the dictionary sum to \(G_{\sigma_J}\), which (given that \(\sigma_0\) should be chosen very small, i.e., \(G_{\sigma_0}\) is a delta peak at the pixel scale) indicates that the dictionary \(\Gamma\) has full rank.

Another crucial property of the dictionary \(\Gamma\) is that the integral of the atoms \(\Phi_{\text{DoG}}^{\sigma_j,\sigma_{j+1}}\) is vanishing. Hence, only the final smoothing scale \(G_{\sigma_J}\) transports the total flux in the image.

The subbands \(I_j\) hold the information of the image at a respective scale described by \(\sigma_j\) and \(\sigma_{j+1}\). We will denote the collection of subbands of an image \(I\) by \(\mathcal{F} = \{I_0, \ldots, I_J\}\) for the rest of the paper. However, even if \(I = \Gamma(I_0, I_1, ..., I_J)\) holds, it is usually \(I_j \neq \omega_j\) due to the nonorthogonality of the DoG wavelet functions. Nevertheless, \(\omega_j\) should provide a reasonable initial guess if one tries to find an array \(\mathcal{F} = \{I_0, I_1, ..., I_J\}\) that satisfies \(I = \Gamma \mathcal{F}\).

2.4. Sparsity-promoting regularization

We apply sparsity-promoting regularization in the generalized Tikhonov framework:

\[
\hat{I} = \arg\min_{I} \|S(F \mathcal{I}, V) + a R(\mathcal{I})\|,
\]

where \(S\) is the data fidelity term, which measures the proximity between the recovered visibilities \(F \mathcal{I}\) and the observed visibility data, \(V\). The term \(F\) denotes mapping of the image intensity onto the visibilities, meaning it computes a tapered and weighted projection of the Fourier transform of \(x\) on a discrete and fixed sampling. The term \(R\) denotes the regularization term, which measures the feasibility of the guess \(\mathcal{I}\). The parameter \(a\) controls the bias between both terms. The final recovered image solution is then:

\[
\hat{I} = \Gamma \mathcal{F}.
\]

The data fidelity terms used for this paper are introduced as follows. Let \(\mathcal{V} = FT \mathcal{I}\) denote the visibility data predicted from the current guess. We quantify the proximity between the predicted and measured visibilities by the effective \(\chi^2\)-distance between them.

\[
S_{\text{vis}}(\mathcal{V}, V) = \frac{1}{N_{\text{vis}}} \sum_{i=1}^{N_{\text{vis}}} \left| V_i - V_i^* \right|^2 / \Sigma_i^2,
\]

where \(N_{\text{vis}}\) is the number of visibilities and \(\Sigma\), the estimated thermal noise of a given visibility. This \(\chi^2\) corresponds directly to a log-likelihood, given uncorrelated Gaussian thermal noise on the different baselines. In addition to this, we also use similar distances defined for three additional quantities: (a) the distance between the measured and predicted visibility amplitudes,

\[
S_{\text{amp}}(\mathcal{V}, V) = \frac{1}{N_{\text{vis}}} \sum_{i=1}^{N_{\text{vis}}} \left| V_i - V_i^* \right|^2 / \Sigma_i^2;
\]

A&A 666, A137 (2022)
where \( N_{\text{cph}} \) is the number of closure phase combinations, \( \Sigma_{\text{cph,i}} \) is the noise on a closure phase \( \Psi_i(V) \), and \( \Psi_j(Y') \) denotes the respective closure phase computed from the array of predicted visibilities, \( Y' \); and finally, (c) the distance between the measured and predicted closure amplitudes,

\[
S_{\text{cph}}(Y', V) = \frac{1}{N_{\text{cph}}} \sum_{i=1}^{N_{\text{cph}}} \frac{|\Psi_i(Y') - \Psi_i(V)|^2}{\Sigma_{\text{cph},i}^2},
\]

with similar conventions as for the closure phases. We would like to note here that Eqs. (17) and (18) are only approximate \( \Psi \)-functionals, incorporating a total flux constraint, such that

\[
R = \sum_{i=1}^{N_{\text{sub}}} \frac{|\Psi_i(Y') - \Psi_i(V)|^2}{\Sigma_{\text{cph},i}^2}.
\]

(b) the distance between the measured and predicted closure phases,

\[
S_{\text{cph}}(Y', V) = \frac{1}{N_{\text{cph}}} \sum_{i=1}^{N_{\text{cph}}} \frac{|\ln A_i(Y') - \ln A_i(V)|^2}{\Sigma_{\text{cph},i}^2},
\]

the noise on a closure phase \( \Psi_i(V) \), and \( \Psi_j(Y') \) denotes the respective closure phase computed from the array of predicted visibilities, \( Y' \); and finally, (c) the distance between the measured and predicted closure amplitudes,

\[
S_{\text{cph,ij}}(Y', V) = \frac{1}{N_{\text{cph,ij}}} \sum_{i=1}^{N_{\text{cph,ij}}} \frac{|\ln A_i(Y') - \ln A_i(V)|^2}{\Sigma_{\text{cph,ij}}^2},
\]

with similar conventions as for the closure phases. We would like to note that Eqs. (17) and (18) are only approximate expressions for the correct log-likelihoods for closure products (e.g., Blackburn et al. 2020; Arras et al. 2022). These approximations and the combinations of them are applied in Sect. 4 for the analysis of test data.

It is known that sparsity is promoted by convex pseudonorm functionals as a regularization term (Starck et al. 2015), for example by a term of the form:

\[
R = \sum_{i=1}^{N_{\text{sub}}} \frac{|\Psi_i(Y') - \Psi_i(V)|^2}{\Sigma_{\text{cph},i}^2},
\]

and \( \Sigma_{\text{cph,ij}} \) is well defined (i.e., there is a unique single-value minimum). For a convex, proper and lower semicontinuous functional, such as the right hand side of Eq. (22), the zero element is in the subdifferential of the functional at the point of the minimum. Hence, \( \hat{s} := \text{prox}_{\tau H}(z) \) satisfies:

\[
z - \hat{s} \in \tau \partial H[\hat{s}],
\]

The power of proximal operators comes from their fixed-point property. It follows directly from Eqs. (22) and (23), that:

\[
\hat{s} \in \arg\min_{s \in X} \left\{ H(s) + \frac{1}{2\tau} \|s - z\|_X^2 \right\}.
\]

 Independently of \( \tau \geq 0 \). For a sketch of the proof see Appendix A.

Hence, we can solve the minimization in Eq. (13) by fixed-point iterating the proximity operator. This procedure is exact, in the sense that convergence proofs are available (e.g., Martinet 1972). For the combination of a smooth term (data-fidelity term) and a nonsmooth term (penalty term), we use a two-step splitting minimization strategy consisting of a gradient descent step for the data fidelity term and one proximity step for the penalty term (Combettes & Pesquet 2009). The forward–backward splitting algorithm is outlined in its general framework in Table 1. The two-step splitting is realized during the last step of the algorithm, when the current guess is updated by a proximal step and a gradient descent step.

Interestingly, despite being derived in the context of convex optimization, there are also local convergence proofs available for the cases when \( S \) and \( R \) are not convex, but the penalty term remains lower semicontinuous and proper, and satisfies the technical Kurdyka–Łojasiewicz property (e.g., see Attouch et al. 2013; Ochs et al. 2014; Xiao et al. 2015; Boţ et al. 2016; Liang et al. 2016 or Bao et al. 2016 for a connection to wavelets). This is of special interest for radio aperture synthesis as the data fidelity terms \( S_{\text{amp}}, S_{\text{cph}}, \) and \( S_{\text{cph,ij}} \) are indeed not convex. Local convergence to a steady point is known, and under some circumstances even global convergence could be proven (compare the discussion in Liang et al. 2016).
Table 1. Forward–backward splitting for the minimization of \( S + R \).

| Input: \( S, R : \mathbb{X} \rightarrow \mathbb{R} \) (convex) |
| Input: \( \text{grad} \ S \) is L-Lipschitz continuous |

Step size: \( \tau \in (0, 2/L) \), typical choice: \( \tau = 1/\| \text{grad} \( S \) \|^2 \)  
Initial guess: \( x_0 \in \mathbb{X} \)  
while \( i = 0, 1, 2, \ldots \) do  
\[ x_{i+1} = \text{prox}_{\tau \lambda}(x_i - \tau \text{grad} \( S \)(x_i)) \]

given a reasonable initial guess, local minima could be avoided (Starck et al. 2015).

One may wonder if the difficulty with convex nonsmooth penalty functionals is now just transported to the probably troublesome minimization problem in the definition of the proximity operator in Eq. (22). But the proximity operator is known for a large number of examples and the computation is often not more time-consuming than one Landweber iteration. For example, for the \( l_0 \)-functional, the proximity operator is (e.g., Starck et al. 2015):

\[
\text{prox}_{\tau \lambda}(z) = \begin{cases} 
\hat{z} & |z| > \sqrt{2} \tau \\
\text{sign}(z)[0,z] & |z| = 0 \\
0 & |z| < \sqrt{2} \tau 
\end{cases}
\]  

(25)

where the signum and absolute value are meant to be evaluated pointwise. This not always a single value, since the \( l_0 \) norm is not convex. The proximal point operator of characteristic functions is the projection on the support of the characteristic function; in the case of multiresolution support, this is the function that nullifies all coefficients outside the multiresolution support, and in the case of the total flux, it is the function that projects the current guess to the guess with the correct total flux.

3. Pipeline

3.1. Outline

We use the same notations as in the previous subsections: \( V \) are the observed visibilities, \( f \) the prior compact total flux, \( \Gamma \) the dictionary of composed of DoG wavelets, and \( F \) the linear mapping of the image intensity to the tapered visibilities. The core of our imaging method concerns solving the following optimization problem:

\[
\tilde{\mathcal{J}} \in \arg\min_{\mathcal{J}} \left[ S_{\text{cph}}(F \mathcal{T} \mathcal{J}, V) + S_{\text{cl}}(F \mathcal{T} \mathcal{J}, V) + \alpha \cdot R_{\lambda}(\mathcal{J}) + R_{\text{flux}}(\mathcal{J}, f) \right],
\]

(26)

where we choose the maximum of the corresponding DoG wavelet function as weights \( \omega \). We have only one regularization parameter \( \alpha \) that controls the amount of suppression by hard thresholding. We like to emphasize the main motivations behind this optimization problem.

Firstly, we use the more flexible DoG dictionary here, see Eq. (12). This allows us to adapt the dictionary to the \( \omega \)-coverage by separating scales that are well covered by observations from those that are less accurately constrained by observations. This will allow us to better suppress the signal from the latter scales.

Secondly, we initially use the closure properties as the data fidelity term, as a measure to reduce the effect of possible antenna-based calibration errors. Chael et al. (2018) demonstrated that this information is sufficient to recover the image when using strong regularization priors. In later imaging rounds, that is after several self-calibration steps, we also start to include amplitude and phase information.

Thirdly, we use hard thresholding (\( l_0 \) pseudonorm regularization), which promotes sparsity. In the few works addressing multiscale imaging for radio aperture synthesis (Li et al. 2011; Carrillo et al. 2012, 2014; Garsden et al. 2015; Girard et al. 2015; Onose et al. 2016, 2017; Mouri Sardarabadi et al. 2016; Pratley et al. 2018), often the \( l_1 \)-norm is used as a convex approximation to \( R_{\lambda} \). This is standard for sparsity-promoting inverse problems (e.g., Starck et al. 2015). However, the \( l_1 \)-norm suppresses both the image features and noisy structures. As it is important to preserve the amplitude on the well-covered scales, we resort to using the nonconvex \( l_0 \)-pseudonorm as penalization. We weight the \( l_0 \) pseudonorms by the maximal peak of the corresponding DoG wavelet basis function. This is done to prevent the scale selection having a strong effect on the choice of the best regularization parameter. In principle these weighting parameters could be considered as free regularization parameters as well. However, to meet our requirement of constructing an algorithm that is as unbiased and data-driven as possible, we restrict them in this work to the choice that seems most reasonable.

It should be noted that \( S_{\text{cph}}, S_{\text{cl}}, \) and \( R_{\lambda} \) are invariant against rescaling the coefficients \( x \) (atoms) by a scale factor \( \lambda \in \mathbb{R} \). To select the most feasible solution along this line, we select the one that matches the prior compact total flux.

There are more possible regularization terms available, for example the total variation or the total squared variation terms that are applied for the EHT imaging (Event Horizon Telescope Collaboration 2019). However, finding suitable weighting parameters for the different data terms and penalty terms is somewhat unintuitive for such different types of regularizations. This task often requires large parameter surveys with feasible synthetic data. We aim to find a largely unsupervised algorithm with only a few free parameters.

Our optimization problem differs significantly from previous multiscales RML imaging approaches (e.g., Li et al. 2011; Carrillo et al. 2012, 2014; Garsden et al. 2015; Girard et al. 2015; Onose et al. 2016, 2017; Mouri Sardarabadi et al. 2016; Cai et al. 2018a,b; Pratley et al. 2018). They used the starlet transform as a dictionary (which we replaced by the DoG dictionary), the distance of observed and predicted visibilities as the data fidelity term (which we replaced by closure properties), and \( l_1 \) penalty terms (which we replaced by \( l_0 \) penalization).

Nevertheless, our algorithm shares some similarities with RML reconstructions. The unpenalized minimization of the data fidelity terms yields a high-resolving reconstruction that fits the observed data points with a fidelity that is (too) high, but that provides clearly unphysical, highly oscillating fits of the visibilities in the gaps of the \( \omega \)-coverage. Total variation and total squared variation penalization effectively smooth the recovered model to a reasonable extent, where the amount of smoothing is controlled by the trade-off between the data fidelity term and the penalization term. We achieve a similar effect by modeling the brightness density distribution with (as few as possible) smooth, extended basis functions.

3.2. Pipeline

The data fidelity terms \( S_{\text{cph}} \) and \( S_{\text{cl}} \), and the regularization term \( R_{\lambda} \), are not convex. Therefore, the minimization problem stated in Eq. (26), strictly speaking, may not have a single-value minimum. Therefore, a careful imaging pipeline helping global convergence is needed. It should also be noted that the representation of the image in wavelet scales is an overcomplete...
representation. Due to the resulting large arrays, computation could be slow. Computation time can, in principle, be reduced when starting from a reasonable initial guess instead of a flat image or a Gaussian prior.

On the other hand, the solution of Eq. (26) returns an adequate calibration image \( \hat{I} = \Gamma \hat{x} \) and computes, on the fly, a multiresolution support \( M \) (all the pixels that are unequal to zero). Further imaging rounds, including the self-calibrated visibilities, allow the solution to vary only in the multiresolution support, and thus could sharpen the image further while respecting the sparsity assumption thanks to the multiresolution support. This approach is realized within the following imaging pipeline:

The first imaging round consists of “single scalar flux constraining imaging”: We minimize the term:
\[
\hat{I}_1 \in \arg\min_{I} S_{\text{amp}}(FI, V) + S_{\text{cph}}(FI, V) + S_{\text{cla}}(FI, V) + R_{\text{wul}}(I, f),
\]
where no dictionary is involved. We do this by the fast minimization method available in the scipy package. In fact, this imaging round is similar to the first imaging round with the eithim imaging package for Event Horizon Telescope Collaboration (2019). This imaging round is used for finding a reasonable initial guess in order to reduce the overall computation time. We convolve the result with the instrument beam (to avoid local minima) and only use a few iterations, meaning it is an incomplete decomposition. Finally, we have to find a wavelet coefficient array \( \hat{\mathcal{J}}^1 \) that satisfies \( \hat{I}_1 = \Gamma \hat{\mathcal{J}}^1 \). To satisfy Eq. (10), we copy the intensity \( \hat{I}_1 \) in every scale \( \hat{\mathcal{J}}^1 \), where \( j \) denotes the scale in use.

The second imaging round consists of “multiscalar closure property hard thresholding imaging”: This imaging round is the heart of the new algorithm. We solve Eq. (26) using a forward–backward splitting approach. We start from the initial guess \( \hat{\mathcal{J}}^1 \) computed in the first imaging round and compute a scale discrete guess in order to minimize Eq. (26). We start from the largest scales only (setting all other subbands to zero), successively adding smaller scales and larger thresholds. We stop at the scale at which the functional (26) is minimal, in other words at the smoothing when accuracy of the fit and sparsity penalization balance. Lastly, we reestimate the thresholds for each scale individually starting from the smallest scales. We then minimize, starting from this initial guess, the functional with a forward–backward splitting strategy. We will explain this forward–backward splitting minimization strategy in Sect. 3.3. An outline of the round 2 imaging algorithm is presented in Table 2.

The third imaging round consists of “multiresolution imaging with visibility amplitudes”: We self-calibrate the data with the image guess derived in the second image round. Moreover, we compute the multiresolution support \( M \) from the result \( \hat{\mathcal{J}}^2 \) of the second imaging round, that is to say, we choose all nonzero elements of the multiscalar coefficient array \( \hat{\mathcal{J}}^2 \) as multiresolution support. We now solve the problem:
\[
\hat{\mathcal{J}}^2 \in \arg\min_{\mathcal{J}} S_{\text{amp}}(FI, \mathcal{F} V) + S_{\text{cph}}(FI, \mathcal{F} V) + S_{\text{cla}}(FI, \mathcal{F} V) + R_{\text{wul}}(\mathcal{F} \mathcal{\hat{J}}, M),
\]
This is solved by a simple gradient descent algorithm starting from the initial guess \( \hat{\mathcal{J}}^3 \) in which only the gradient with respect to the coefficient in the multiresolution support is computed.

The fourth imaging round is “multiresolution imaging with full visibilities”: after another self-calibration step, we solve the imaging problem:
\[
\mathcal{J}^4 \in \arg\min_{\mathcal{J}} S_{\text{vis}}(FI, \mathcal{F} V) + R_{\text{wul}}(\mathcal{F} \mathcal{J}, M),
\]
by a gradient descent algorithm that only varies coefficients in the multiresolution support analog to the third imaging round.

For the last, fifth imaging round we do “single scalar visibility imaging”: We set all pixels with negative flux to zero flux and increase the match to the observed visibilities by a gradient descent algorithm that minimizes \( S_{\text{vis}}(FI, V) \) in the pixel scale starting from \( \hat{F} = I^5 \).

The last three imaging rounds (in particular round 5) are optional and only refine the reconstruction. This will be discussed in our demonstration on synthetic data in Sect. 4.

### 3.3. Minimization algorithm

We now discuss the minimization algorithm used to minimize Eq. (26). All other imaging rounds are based on smooth gradient descent imaging algorithms (rounds 3–5) or a smooth Newton-type minimization (round 1). But Eq. (26) is neither convex nor smooth. However, the data fidelity terms are smooth with Lipschitz continuous derivatives and the \( l_0 \) pseudonorm is proper and lower-semicontinuous, and satisfies the Kurdyka–Lojasiewicz property (e.g., Liang et al. 2016). Thus, the forward–backward splitting algorithm 1 remains applicable, see our discussion in Sect. 2.5. Additionally, we recall that \( S_{\text{cph}}, S_{\text{cla}} \) and \( R_{\text{wul}} \) are invariant against rescaling the coefficient array by a scalar factor \( \alpha \). We therefore propose the following iterative scheme: we first minimize \( S_{\text{cph}}(FT_x, V) + S_{\text{cla}}(FT_x, V) + \alpha \cdot R_{\text{wul}}(x) \) by a fixed number of forward–backward splitting iterations, and then we rescale the coefficient array by a scale factor, such that \( I = FT_x \) has a total flux matching the prior compact flux (leaving the data fidelity terms and regularization terms unaffected). Then we proceed with our forward–backward splitting algorithm, performing rescaling again, and so on. The complete procedure is outlined in Table 2. The needed proximal operator for the \( l_0 \) pseudonorm is computed in Eq. (25).

Iterative reweighted \( l_1 \)-regularization proposed by Candès et al. (2007) provides an alternative approach to solving optimization problems with nonconvex \( l_0 \)-terms and is more common than our forward–backward scheme. However, our rescaling approach to match the total flux would affect the reweighting step of the reweighted \( l_1 \)-regularization method. Thus, it would introduce an additional layer of complexity in solving the optimization problem. This would fail our requirement of a preferably simple imaging algorithm with a small number of parameters to specify.

### 3.4. Selection of scales

Our DoG-wavelet dictionary is flexible in the sense that the Gaussian widths could be chosen to adapt to the \( \omega \)-coverage. Hence, the selection of scales is data driven (e.g., by the \( \omega \)-coverage) and should be performed automatically. In this section we discuss the automatic scale–width selection and outline the key points of this approach.

The Fourier transform of a two-dimensional DoG wavelet is a ring shaped mask, see Eq. (8). It is reasonable to select the masks such that well-covered regions of the \( \omega \)-space and poorly covered regions are separated. However, for very sparse arrays, there are no really well covered scales. In this situation, our selection should be also driven by the assertion that all the data points belonging to the same antenna pair should be covered in one scale.
We present a sketch of our automatic scale selection in Fig. 1. We unpack the array of \( w \)-distances of the full array, sort it in increasing order (black dots in Fig. 1), and search for jumps between two consecutive data points that exceed a certain threshold. These jumps clearly appear at gaps in the \( w \)-space that are less relevant, as these scales will be suppressed by the algorithm automatically, see Sect. 4.2.

As a demonstration, we apply this procedure to the EHT 2017 array. In Fig. 2, we show our masks and the data points in \( w \)-space. The width information of the scales shown in Fig. 2 is given in Table 3. We also mention in Table 3 which scale is most sensitive to which antenna pair (i.e., what the selection criterion was for this scale). As all DoG wavelets satisfy the zero integral property of wavelets, the only flux-transporting scale is the smoothing scale \( G_{\sigma} \).

The smallest scale in our set has a width of 9.96 \( \mu \)as, which corresponds to 5.02 pixels in our discretization. For the sake of completing our dictionary of wavelet functions so that Eq. (10) remains satisfied, we complete our sets of scales down to the pixel size by adding DoG wavelets according to the widths of one, two, and four pixels. This, however, will turn out to be less relevant, as these scales will be suppressed by the algorithm automatically, see Sect. 4.2.

### 4. Tests with synthetic data

#### 4.1. Testdata

We test our algorithm on the same set of synthetic data that were recently used for testing feature extraction from the EHT data (Tiede et al. 2020). In particular, we use a crescent, a disk, a double Gaussian, and a ring structure.

The crescent is described by the equation (Tiede et al. 2020):

\[
I(r, \theta) = I_0 \left(1 - s \cos(\theta - \xi) \right) \frac{\delta(r - r_0)}{2\pi r_0}.
\]

(30)
of the ehtim package, using the EHT 2017 array at
Telescope Collaboration 2019). We assume phase and gain
correlation times as for the EHT observations (Event Horizon
observe same
We simulate visibility data from the test images with the help
covers, i.e., which antenna pair is mainly covered by these scales.

| Scale | $\sigma_1 \rightarrow \sigma_2$ ($\mu$as) | Main Sensitivity |
|-------|---------------------------------|-----------------|
| 0     | 0.84 $\rightarrow$ 1.69        | Unresolved      |
| 1     | 1.69 $\rightarrow$ 3.37        | Unresolved      |
| 2     | 3.37 $\rightarrow$ 4.23        | Unresolved      |
| 3     | 4.23 $\rightarrow$ 5.78        | PV-SMA/JCMT     |
| 4     | 5.78 $\rightarrow$ 6.66        | Gap             |
| 5     | 6.66 $\rightarrow$ 7.06        | AA/AP-JCMT/SMA, AA/AP-PV |
| 6     | 7.06 $\rightarrow$ 12.18       | AA/AP-SMT, SMT-PV, LMT-PV |
| 7     | 12.18 $\rightarrow$ 14.13      | AA/AP-LMT, JCMT/SMA-LMT |
| 8     | 14.13 $\rightarrow$ 17.55      | JCMT/SMA-SMT    |
| 9     | 17.55 $\rightarrow$ 33.69      | Gap             |
| 10    | 33.69 $\rightarrow$ 39.81      | LMT-SMT         |
| 11    | 39.81                           | AA-AP, JCMT-SMA |

Notes: Different scales are most sensitive either to specific baselines or
the gaps in the $uv$-coverage, i.e., which antenna pair is mainly covered by these scales.

We aim to study the image on a 128x128 pixel grid with
1$\mu$as-pixels. However, to avoid boundary effects in the com-
putation (the largest chosen Gaussian has a FWHM of already
93.75 $\mu$as), we widen the field of view by a factor of two. More-
over, we use 129 pixels instead of 128 pixels to discretize narrow
central Gaussians correctly. We have defined 12 different wavelet
scales. Thus, we are attempting to solve for $12 \times 129 \times 129 \approx 2 \times 10^7$ parameters in the multiscale imaging rounds.

4.2. Imaging pipeline
In this subsection, we use with the crescent image to demon-
strate the stability of our imaging pipeline and present some key
features. We show in Fig. 3 the imaging results obtained from
the crescent test data after different imaging rounds. The image
after the second imaging round is shown in the upper right panel,
and the final image after the fifth imaging round is shown in the
lower right panel. The essential image structure is already recov-
ered after the second imaging round (multiscale imaging with
coverage, i.e., which antenna pair is mainly covered by these scales.

We use $\xi = 180^\circ$, $r_0 = 22 \mu$as, $s = 0.46$, and $b_0 = 0.61$Jy. The
crescent is then convolved with a Gaussian with a full width at
half maximum (FWHM) of 10 $\mu$as. The disk is a disk of diam-
eter 70 $\mu$as. The disk is then convolved with a Gaussian with
FWHM 10 $\mu$as.

The double Gaussian image consists of two Gaussian peaks
with a FWHM of 20 $\mu$as. The first Gaussian is placed at the
origin and has a flux of 0.27Jy. The second Gaussian is placed
30 $\mu$as to the east and 12 $\mu$as to the south. It has a flux of 0.33 Jy.
The ring has radius of 22 $\mu$as and a total flux of 0.61 Jy. The
ring is convolved with a Gaussian with a FWHM of 10 $\mu$as.

We simulate visibility data from the test images with the help
of the ehtim package, using the EHT 2017 array at 229 GHz. We
mimic the observation with the observe_same option, assum-
ing the same systematic noise levels, observation intervals, and
and the final image after the fifth imaging round is shown in the
lower right panel. The essential image structure is already recov-
ered after the second imaging round (multiscale imaging with
scales are sensitive mainly to fine structures, that could only be
self-calibration loops. Nevertheless, the use of the amplitudes
and full visibility data (imaging rounds 3–5, lower panels) refines
the recovered structures, and increases coincidence with
observed visibilities. Moreover, the steady improvement of the
image quality shown in Fig. 3 demonstrates that our amplitude
conserving hard thresholding approach works as intended. We
observe a strong contrast between the ring feature and the inner
depression (due to sparsity) while the amplitude and total flux is
conserved. This would not be available with soft thresholding.

We demonstrate in Fig. 4 that our final image fits the
observed visibilities well. The hard thresholding approach sup-
presses emission that is not significant for fitting the visibilities,
but it does not break the fit to the observed data as soft thresh-
olding would do. In fact, we successfully separated between
significant image structures (fitting the visibilities) and noise
induced features (very small sidelobe level in the final image).

We present the multiscale composition of the image in
more detail in Fig. 5. The panels of Fig. 5 suggest that differ-
ent scales are sensitive to different parts of the final image, for
example an extended Gaussian component (bottom right panel),
the ring feature with a central negative peak to compensate for
this extended emission (middle panels), or the asymmetry of
the crescent (bottom left panel). The final high-resolution and
high-contrast image is only visible when summing all the single-
scale images. Additionally, we present in Fig. 4 the fit to the
data from the single scales only for some selected single scales,
namely the ones that have the largest signal according to Fig. 5.
In the Fourier domain, the various scales are mostly sensitive
to varying parts of the $uv$-coverage, from the short baselines
(scales 9 and 11), over the middle baselines (scale 6), to the
longest baselines (scale 5). The essential image structure is
already recovered after the second imaging round (multiscale imaging with
coverage, i.e., which antenna pair is mainly covered by these scales.

We use $\xi = 180^\circ$, $r_0 = 22 \mu$as, $s = 0.46$, and $b_0 = 0.61$Jy. The
crescent is then convolved with a Gaussian with a full width at
half maximum (FWHM) of 10 $\mu$as. The disk is a disk of diam-
eter 70 $\mu$as. The disk is then convolved with a Gaussian with
FWHM 10 $\mu$as.

The double Gaussian image consists of two Gaussian peaks
with a FWHM of 20 $\mu$as. The first Gaussian is placed at the
origin and has a flux of 0.27Jy. The second Gaussian is placed
30 $\mu$as to the east and 12 $\mu$as to the south. It has a flux of 0.33 Jy.
The ring has radius of 22 $\mu$as and a total flux of 0.61 Jy. The
ring is convolved with a Gaussian with a FWHM of 10 $\mu$as.

We simulate visibility data from the test images with the help
of the ehtim package, using the EHT 2017 array at 229 GHz. We
mimic the observation with the observe_same option, assum-
ing the same systematic noise levels, observation intervals, and
correlation times as for the EHT observations (Event Horizon
Telescope Collaboration 2019). We assume phase and gain
calibration, but add thermal noise.

We aim to study the image on a 128x128 pixel grid with
1$\mu$as-pixels. However, to avoid boundary effects in the com-
putation (the largest chosen Gaussian has a FWHM of already
93.75 $\mu$as), we widen the field of view by a factor of two. More-
over, we use 129 pixels instead of 128 pixels to discretize narrow
central Gaussians correctly. We have defined 12 different wavelet
scales. Thus, we are attempting to solve for $12 \times 129 \times 129 \approx 2 \times 10^7$ parameters in the multiscale imaging rounds.

4.2. Imaging pipeline
In this subsection, we use with the crescent image to demon-
strate the stability of our imaging pipeline and present some key
features. We show in Fig. 3 the imaging results obtained from
the crescent test data after different imaging rounds. The image
after the second imaging round is shown in the upper right panel,
and the final image after the fifth imaging round is shown in the
lower right panel. The essential image structure is already recov-
ered after the second imaging round (multiscale imaging with
coverage, i.e., which antenna pair is mainly covered by these scales.

We use $\xi = 180^\circ$, $r_0 = 22 \mu$as, $s = 0.46$, and $b_0 = 0.61$Jy. The
crescent is then convolved with a Gaussian with a full width at
half maximum (FWHM) of 10 $\mu$as. The disk is a disk of diam-
eter 70 $\mu$as. The disk is then convolved with a Gaussian with
FWHM 10 $\mu$as.

The double Gaussian image consists of two Gaussian peaks
with a FWHM of 20 $\mu$as. The first Gaussian is placed at the
origin and has a flux of 0.27Jy. The second Gaussian is placed
30 $\mu$as to the east and 12 $\mu$as to the south. It has a flux of 0.33 Jy.
The ring has radius of 22 $\mu$as and a total flux of 0.61 Jy. The
ring is convolved with a Gaussian with a FWHM of 10 $\mu$as.

We simulate visibility data from the test images with the help
of the ehtim package, using the EHT 2017 array at 229 GHz. We
mimic the observation with the observe_same option, assum-
ing the same systematic noise levels, observation intervals, and
correlation times as for the EHT observations (Event Horizon
Telescope Collaboration 2019). We assume phase and gain
calibration, but add thermal noise.
Fig. 2. Observed uv-coverage (black and red points) of the EHT data array (observation of M 87 on 5 April 2017) and masks defined by the DoG wavelets listed in Table 3 (color maps). The masks are the Fourier transform of the respective wavelets and they define ring-like filters in Fourier domain. The visibilities highlighted by a specific filter are plotted in red.

4.3. Proof of concept

One of the principal ideas of this paper is to define a flexible wavelet dictionary that adapts smoothly to the uv-coverage. We now prove this concept. We present in Fig. 6 a reconstruction with the complete pipeline, with the selection of scales specified in Table 3, and with a coarser grid that would be available, for instance, with the less flexible à trous wavelet transform (right panel): \( \tilde{\Sigma} = [1, 4, 8, 32] \) (in units of 1.98 μas pixels). We used only every second power of two here for demonstration purposes, to enhance the effect of a less fine grid of scales.

The crescent structure is much more robustly recovered with our selection of scales. This is expected, as illustrated by Fig. 5. The smaller scales respond to different aspects of the fine structure of the crescent test image, such as the ring-like emission, the narrow central ring line or the southern emission peak. The larger scales compress the extended emission. The final high-resolution image is only visible by the sum of all these scales. The artificial selection of scales \( \tilde{\Sigma} \) has a less complex separation of scales. The complex conglomerate of multiple structure features has to be compressed in only one or two scales. Due to the coarse gridding of widths in \( \tilde{\Sigma} \), the algorithm is forced to utilize to small scales, which are not able to compensate the bad fitting of the unconstrained minimization. Our automatic scale selection outperforms this rigid choice of scales because of a more suitable smoothing and thresholding due to adaptive steps in the scale selection, and hence a more rigorous compression of structure information.

That said, it should be mentioned again that the wavelet dictionaries are complete regardless of the selection of scales. Hence, theoretically, the same image can be represented by both wavelet dictionaries regardless of the special choice of scales. The dependency of the reconstruction on the selection of scales is induced by the imaging pipeline (we recall that the objective functional is not convex, and thus only convergence to a local minimum can be assured). It is easier to recover the image feature at a specific scale, if this scale is well covered by measurements, which helps global convergence with our imaging pipeline. On the other hand, a deconvolution at a less well covered scale is...
Fig. 3. Imaging results of the crescent at various steps of the imaging pipeline. Upper left: Gaussian prior image. Upper middle: initial guess, result after round 1 blurred by the 20 μas beam. Upper right: after imaging round 2. Bottom left: after imaging round 3. Bottom middle: after imaging round 4. Bottom right: final image after imaging round 5.

Fig. 4. Observed amplitudes (black) and recovered visibilities (yellow) as functions of \( u-v \)-radius for the crescent test data. Moreover, we show the fit of single scales for some selected scales (blue, red, purple, green).

more uncertain and would possibly fail in the reconstruction of some features.

One may ask now whether progressively refining the grid of scales could further increase the accuracy of image restoration. Whilst, in principle, this is expected, it also comes with the cost of increased computation time and requires more complexity. In this regard, our automatic scale selection may be viewed as a viable optimum and data-driven approach.

4.4. Regularization parameter

Our algorithm depends on significantly fewer critical parameters that need to be specified by the user. The user only needs to define the regularization parameter \( \alpha \) that controls the size of the penalty term, in contrast to the RML methods, which require multiple penalty terms (e.g., with MEM, l1, TV, TSV... penalty terms), balanced by the term weightings. All other parameters in DoG-HiT are determined automatically from data; the widths of the DoG dictionary are defined by the automatic procedure described in Sect. 3.4 and the total flux could be identified with the zero-spacing flux, which can be measured or estimated. Parameters corresponding to the numerical minimization methods (stepsize, number of iterations, relative tolerance) have only a minor impact on the final result, as long as convergence is assured. We present a more quantitative analysis of the impact of the regularization parameter \( \alpha \) on the reconstruction in Appendix B. In short, if the regularization parameter is too small, the visibilities are overfitted by a greedy model with a high background level. For higher regularization parameters, the penalty term becomes more important; the background flux level is decreased and the greedy, blobby model becomes more uniform. Thus, the best fit is achieved. On the other hand, if the regularization parameter chosen is too big, the sparsity penalization dominates the objective functional. The hard thresholding suppresses significant image information and the image is badly fitted with a small number of large wavelet scales.

5. Comparison to alternative imaging algorithms

We compare our image reconstruction with the image reconstructions by standard Högbom CLEAN and the RML method available in the ehtim software package. We utilize the weighting of the data terms for RML reconstructions that was used for Event Horizon Telescope Collaboration (2019) and apply their
Fig. 5. Multiresolution image after imaging round 4. Each panel shows the recovered images from one scale only. The scales are computed with the DoG method with the widths defined in Table 3. The images is shown for every scale in increasing order from the upper left to the lower right.

Fig. 6. Reconstruction of the crescent image. Left panel: true image. Middle panel: reconstruction with the selection of scales specified in Table 3. Right panel: reconstruction with scale widths that are a power of two (discrete wavelet transform).

5.1. Qualitative comparison

We show in Fig. 7 our test image reconstructions on the set of test data presented in Sect. 4.1. Our image reconstruction shows a greater resolution than the CLEAN images. Moreover, we achieve a greater contrast between image features and background noise levels than the CLEAN algorithm, (i.e., sharper edges in the recovered images).

Compared to the powerful RML imaging method, our algorithm achieves comparable resolutions. This is surprising, as we probe the observed images with extended basis functions. In particular, we are able to recover some of the fine structure that is not visible in the RML reconstructions. We find the correct crescent-shaped north–south asymmetry in the crescent image, the fine

---

1 Available under https://github.com/eventhorizontelescope/2019-D01-02
2 https://github.com/eventhorizontelescope/2019-D01-02
Fig. 7. Comparison of the reconstructions with various imaging algorithms. We show in the upper line the true images (crescent, disk, double, ring). In the second to fourth lines, we present the image reconstructions with DoG-HiT, RML, and CLEAN, respectively.

ring ridgeline in the ring image, and the correct peak values in the double Gaussian image. Moreover, we find a greater contrast between the ring-like features in the ring and crescent images, and the central depression, compared to the contrast observed in RML image. However, the inner “no-emission” radius is smaller than in the true images with DoG-HiT, while the spherical shape remains better recovered. This region is significantly better recovered by the RML algorithm. Moreover, RML appears to perform better in resolving the ring and crescent features transversely.

Notably, our algorithm also succeeds in the reconstruction of smooth extended emission (e.g., of the disk image). The reconstruction of the disk is quite accurate and comparable to the reconstruction with CLEAN. It does not manifest the greedy image disk features or background emission present in the RML reconstruction. The ring image demonstrates that DoG-HiT is able to fit uniform emission (ring extension) and sharp features (ring edges) simultaneously. The CLEAN reconstruction of the ring lacks the proper reconstruction of the sharp ring edges and the central depression. The RML reconstruction fits the central depression well, but the ring brightness distribution is less homogeneous than in the DoG-HiT reconstruction. In this way DoG-HiT combines the major advantages of RML reconstructions (super-resolving structures) and CLEAN (high dynamic range sensitivity to extended structures), and at the same time reduces the drawbacks of both of these two methods. It should, therefore, be well suited for imaging problems arising in the context of EHT observations, in which the demand for the recovery of information contained on the smallest accessible scales requires simultaneous robust imaging of extended structures (jet). The performance of Dog-HiT under these conditions will be discussed in Sect. 6, using simulated data with a wide range of spatial scales.

5.2. Quantitative comparison

We now compare the various imaging algorithms in a more quantitative way, using a measure of their relative error,

$$\text{err} = \frac{\|\text{recovered solution} - \text{exact image}\|}{\|\text{exact image}\|}. \quad (31)$$
We present the relative errors of the reconstructions in Table 4. The comparison may be somewhat unfair for CLEAN, given the large beam size compared to the size of the structures, but a final convolution with a synthetic point spread function is the common standard in radio astronomy. We present the relative error of the reconstruction both without blurring (as is standard for RML and DoG-HiT) and with blurring by one-half of the beam size and the full beam size (as is standard for CLEAN). The super-resolving DoG-HiT reconstructions worsen with a larger restoring beam, while for CLEAN the opposite is true. DoG-HiT wins the challenge for three of the four test images (crescent, disk, ring) and performs similar to RML for narrow structures (crescent, double). Overall, we can conclude that DoG-HiT is able to achieve a similar precision as current imaging algorithms, but alleviates some of the limitations of both CLEAN (no super-resolution) and RML methods (sensitivity to smooth extended features).

We present in Fig. 8 the residuals of the reconstructions of the ring feature with RML and with DoG-HiT. The residuals for both imaging methods are ring-shaped and spatially correlated, indicating that there is still not recovered structure. However, the histograms of the residuals in the lower panels of Fig. 8 demonstrate a very good reconstruction overall. The pixel residual distribution is well approximated by a narrow Gaussian distribution in both cases. Nevertheless, the residual distribution for DoG-HiT is slightly more narrow and less skewed, which agrees well with the overall slightly smaller relative error listed in Table 4.

5.3. Transverse resolution

We study the transverse resolution of the algorithms with the crescent image in this section. We present in Fig. 9 the profiles of the true (blue) and the recovered crescent images in the north–south direction at central right ascension. We recover the correct double peak structure with north–south asymmetry, both with the RML method and with DoG-HiT. CLEAN is not able to reproduce this fine structure sufficiently. Regarding transverse resolution of the ring features and the central depression, RML and DoG-HiT perform equally well, approximately recovering the correct widths of the Gaussian blurred ring and the correct depth of the central depression. However, DoG-HiT recovers a
zero-flux central depression, which is not captured in the true image. We computed the blurring beam size that maximizes the correlation between the (blurred) true image and the reconstructed images to quantify the resolutions. The largest correlation between the DoG-HiT reconstruction and the true image was achieved if the true image was blurred by a beam with widths 6.1 µas. That means that DoG-HiT was able to reproduce image features down to a resolution of approximately 6 µas. For the RML reconstruction, we found a maximal correlation for a beam of 5.2 µas, similar to DoG-HiT. This resolution is expected due to the reverse taper of 5 µas applied in the ehtim pipeline. For CLEAN we found a width of 19.7 µas, coinciding well with the applied point spread function of the array.

5.4. Simplicity and performance

The five imaging steps presented in Sect. 3.2 may not appear as a simple approach to imaging. However, one should recall that this strategy resembles typical steps in the imaging of interferometric data with CLEAN (imaging in several loops of cleaning and self-calibration). Hence, our lengthy pipeline is not more complex than automatic cleaning scripts.

More importantly, DoG-HiT only takes into account a very limited number of regularization parameters, namely only the prior guess for the total flux and the biasing parameter \( \alpha \), which controls the weight of the hard thresholding regularization term (see Eq. (26)). Apart from these parameters, only solver related choices such as stepsize or relative tolerances need to be specified. This is a first step toward a more unsupervised imaging algorithm in which crucial choices for the imaging procedure (i.e., selection of window or regularization parameters) come directly from data and are not manually selected doing the analysis.

The down side of this simplicity is that DoG-HiT is also less flexible compared to RML methods. RML methods combine different choices of parameters and prior assumptions that could be relatively weighted according to specifications of a certain data set. While it is promising news that wavelet sparsity-promoting algorithms are similar to RML methods, or in some settings, perform better, this conclusion cannot be automatically generalized. In particular, more advanced calibration issues could add an additional layer of complexity to the problem. Furthermore, the hard thresholding method used in DoG-HiT may limit the dynamic range of the reconstruction (i.e., the minimal flux that can be recovered). A more rigorous study of this drawback should be made in subsequent works and applications.

Furthermore, a serious disadvantage of our algorithm is that it presently requires considerably more time and computing resources than the fast RML methods. This is because the image is represented by an overcomplete dictionary of wavelet scales.

6. Physical source model

To demonstrate the performance of the algorithm on structures covering a wider range of spatial scales, we present here a DoG-HiT image reconstruction made from synthetic data from the first ngEHT Analysis Challenge\(^3\), which emulate the black hole shadow and the jet base in M87 as observed with a possible ngEHT configuration (Roelofs et al. 2022). The ngEHT is a planned, but not finally proposed, future global VLBI array, designed to produce real time movies of the dynamics in the extreme vicinity of a black hole and the innermost jet region (Doeleman et al. 2019). The ngEHT will build upon the enormous success of the EHT and will extend the EHT science with higher dynamic ranges, sensitivity, and resolution. It is believed that it will deliver novel groundbreaking results for the formation of jets, accretion physics, and general relativity tests (Doeleman et al. 2019). In particular, the dense \( \mu \)-coverage (including short baselines) and high sensitivity of the ngEHT, compared to the current EHT, allow for the reconstruction of the extended jet emission. The reconstruction and the true image are compared in Fig. 10 (in left column) and logarithmic (middle and right column) scales, with the latter employed to highlight the extended emission. The simulated source structure is taken from a MAD GRMHD simulation of a rapid spinning black hole surrounded by an accretion disk with electron heating from reconnection (Roelofs et al. 2022; Mizuno et al. 2021; Fromm et al. 2022). The simulated visibilities are calculated for a possible template ngEHT configuration at 230 GHz that is used throughout the ngEHT Analysis Challenge (Roelofs et al. 2022) and might be realized in the final concept of the array. It contains the 11 current EHT sites (ALMA, APEX, GLT, IRAM 30 m, JCMT, KP, LMT, NOEMA, SMA, SMT, SPT) and ten additional stations from the list of Raymond et al. (2021, BAR, OVRO, BAJA, NZ, SGO, CAT, GARS, HAY, CNI, GAM). HAY, OVRO, and GAM are 37-, 10-, and 15-m antennas, respectively. All of the remaining additional antennas are assumed to be of 6 m in diameter. The synthetic visibilities are simulated with a 10-s averaging time and with alternating 10-min observation scans and 10-min gaps. The resulting \( \mu \)-coverage is presented in Fig. 11. For more information on the generation of the ground truth image and the synthetic observation, we refer the interested reader to the description of the ngEHT Analysis Challenge available at the link listed in footnote 3.

The DoG-HiT reconstruction in Fig. 10 accurately represents the central ring-like structure. This is expected, judging from the successful reconstructions obtained in Sect. 5.1 on the compact crescent images from a much sparser synthetic observation. In addition to this, Fig. 10 demonstrates that DoG-HiT also reproduces the extended emission from the jet base (middle and right panels) very well. The structural details of the jet base are compressed on much larger scales than the smaller ring feature, and are less bright (hence only visible in logarithmic scale). This result demonstrates the ability of DoG-HiT to work on images with a wide range of spatial scales and a large dynamic range.

\(^3\) Available under https://challenge.ngeht.org/challenge1/
7. Conclusion

In this paper, we presented a novel interferometric imaging algorithm that is capable of adapting to the Fourier domain coverage of observations and is particularly applicable to sparse \( \mu \nu \)-coverages. Our imaging algorithm models the image as a sum of the difference of Gaussians wavelet functions. This wavelet dictionary is more flexible than the usual discrete à trous wavelet transform and allows us to select the scales to adapt to the \( \mu \nu \)-coverage.

We formulate the imaging problem as an optimization problem with an objective functional consisting of the reduced \( \chi^2 \) of the recovered closure properties (closure phase and logarithmic closure amplitudes) and an \( l_0 \)-pseudonorm sparsity term in the wavelet domain. As this objective functional is still invariant against rescaling of the image guess, we also add a total flux constraint. The resulting objective functional is non-smooth and non-convex, but could be solved by an iterative hard thresholding splitting algorithm for which local convergence to a steady point is known. Due to non-convexity, global convergence cannot be assured, but practice shows that local minima could be avoided by proper initial guesses. Our algorithm is an amplitude- and total-flux-conserving algorithm, in contrast to schemes using soft thresholding. Together with a more thorough separation of image features and sidelobes by a flexible wavelet dictionary analysis, this is expected to bring significant improvements in imaging of VLBI data with strongly varying and scale-dependent noise.

We present a complete imaging pipeline ready for application. Our imaging pipeline consists of five imaging rounds, where we refine the initial imaging results from the closure properties in an iterative imaging, self-calibration loop, which uses the amplitude and phase information. We apply, for the first time, a multiresolution constraint for these refinement steps. Moreover, we prove the stability of our pipeline in practice on synthetic data. Comparisons of imaging performance on the synthetic data show that DoG-HiT achieves super-resolution and outperforms CLEAN in the reconstruction of fine structure (super-resolving) and that it comparable to RML methods in terms of accuracy of reconstruction. DoG-HiT succeeds in the reconstruction of smooth extended emission components, where it outperforms RML. It effectively combines the strengths of CLEAN and RML methods, and reduces their specific weaknesses. DoG-HiT should therefore be well suited for application to targets with a wide range of spatial scales, for which it may outperform current RML reconstructions in the context of better recovering smoother emission on large scales. We have demonstrated this capability on a synthetic data set from the first ngEHT challenge, with excellent reconstructions achieved for of both the small-scale inner ring-like structure and the faint, larger-scale emission from the jet base. It should also be noted that the DoG-HiT reconstruction accurately reproduces features with a

---

**Fig. 10.** Reconstruction of synthetic M87 observation with a possible ngEHT array taken from the ngEHT Analysis Challenge (Roelofs et al. 2022). The true image is presented in the upper panels, and the reconstruction with DoG-HiT in the lower panels. The left panels show the ground truth and the recovered images in linear scale, the middle panels in logarithmic scale (i.e., highlighting the extended emission from the jet basis), and the right panels compare the ground truth and the recovered image, both smoothed with a restoring beam of 20 \( \mu \)as.

**Fig. 11.** \( \mu \nu \)-coverage of a synthetic ngEHT observation of M87 at 230 GHz. The \( \mu \nu \)-coverage with the EHT 2017 antennas only is plotted in red. For more details, see Roelofs et al. (2022).
strong contrast between emission and the background. At the same
time, DoG-HiT presently introduces some systematic inaccuracies (e.g., a limited dynamic range) into the reconstruction,
and this needs to be addressed in future works.

Acknowledgments. We thank F. Roelofs, C. Fromm, L. Blackburn, G. Lindahl, A. Raymond, S. Doelerman and the team of the ngEHT Analysis Challenge for pro-
viding their data set and for useful discussions. H.M. received financial support for this research from the International Max Planck Research School (IMPRS) for Astronomy and Astrophysics at the Universities of Bonn and Cologne. Software availability: We will make our imaging pipeline and algorithms publicly available soon as part of the software package MrBeam (https://gitlab.com/
hmullergoe/mrbeam). MrBeam makes use of the publicly available chimt (Chael et al. 2018), regpy (Regpy 2019), and WISE software packages (Mertens & Lobanov 2015).

References
Akiyama, K., Ikeda, S., Pleau, M., et al. 2017a, AJ, 153, 159
Akiyama, K., Kuramochi, K., Ikeda, S., et al. 2017b, ApJ, 838, 1
Arras, P., Frank, P., Haim, P., et al. 2022, Nat. Astron., 6, 259
Assirati, L., Silva, N. R., Berton, L., Lopes, A. A., & Bruno, O. M. 2014, J. Phys.: Conf. Ser., 490, 012020
Attoueh, H., Bolte, J., & Svaiter, B. 2013, Math. Programm., 137, 91
Bao, C., Dong, B., Hou, L., et al. 2016, Inverse Probl., 32, 115004
Beck, A., & Teboulle, M. 2009, SIAM J. Imaging Sci., 2, 183
Bhatnagar, S., & Cornwell, T. J. 2004, A&A, 426, 747
Blackburn, L., Pesce, D. W., Johnson, M. D., et al. 2020, ApJ, 894, 31
Boj, R., Csetnek, E., & SziZler Csaba, L. 2016, Inverse Probl., 32, 155004
Cai, X., Pereyra, M., & McEwen, J. D. 2018a, MNRAS, 480, 4154
Cai, X., Pereyra, M., & McEwen, J. D. 2018b, MNRAS, 480, 4170
Candes, E., Romberg, J., & Tao, T. 2006, IEEE Trans. Inform. Theory, 52, 489
Candes, E., Wakin, M., & Boyd, S. 2007, J. Fourier Anal. Appl., 14, 877
Carrillo, R. E., McEwen, J. D., & Wiaux, Y. 2012, MNRAS, 426, 1223
Carrillo, R. E., McEwen, J. D., & Wiaux, Y. 2014, MNRAS, 439, 3591
Chael, A. A., Johnson, M. D., Bouman, K. L., et al. 2018, ApJ, 857, 23
Clark, B. G. 1980, A&A, 89, 377
Combettes, P., & Pesquet, J. 2009, Fixed-Point Algorithms for Inverse Problems in Science and Engineering, 49
Cornwell, T. J. 2008, IEEE J. Sel. Top. Signal Process., 2, 793
Coupinot, H., Hecquet, J., Auriere, M., & Futaille, R. 1992, A&A, 259, 701
Doelerman, S., Blackburn, L., Dexter, J., et al. 2019, Bull. Am. Astron. Soc., 51, 256
Donoho, D. 2006, IEEE Trans. Inform. Theory, 52, 128
Event Horizon Telescope Collaboration (Akiyama, K., et al.) 2019, ApJ, 875, L4
Frieden, B. R. 1972, J. Opt. Soc. Am., 62, 511
Fromm, C. M., Cruz-Osorio, A., Mizuno, Y., et al. 2022, A&A, 660, A107
Garsden, H., Girard, J. N., Starck, J. L., et al. 2015, A&A, 575, A90
Girardin, J. N., Garsden, H., Starck, J. L., et al. 2015, J. Instrum., 10, C08013
Gonzalez, R. C., & Woods, R. E. 2006, Digital Image Processing, 3rd edn.
Grove, F., Grossmann, A., & Morlet, J. 1984, Geoexploration, 23, 85
Grossmann, A., Kronland-Martinet, R., & Morlet, J. 1989, In Wavelets: Time-Frequency Methods and Phase Space -1, 2
Högber, J. A. 1974, A&AS, 15, 417
Ikeda, S., Tazaki, F., Akiyama, K., Hada, K., & Honma, M. 2016, PASJ, 68, 45
Lannes, A., Anterrieu, E., & Marechal, P. 1997, A&A, 123, 183
Li, F., Cornell, T. J., & de Hoog, F. 2011, A&A, 528, A31
Liang, J., Fadili, J., & Peyré, G. 2016, in Advances in Neural Information Processing Systems, eds. D. Lee, M. Sugiyama, U. Luxburg, I. Guyon, & R. Garnett, (Curran Associates, Inc.), 29, 4035
Lime, J. L. B., Mitchell, D. A., Pindor, B., et al. 2020, PASA, 37, e027
Mallat, S. G. 1989, IEEE Trans. Pattern Anal. Mach. Intell., 11, 674
Martinet, B. 1972, Comptes Rendus Acad. Sci. Paris, 274, 163
Mertens, F., & Lobanov, A. 2015, A&A, 574, A67
Mizuno, Y., Fromm, C. M., Younsi, Z., et al. 2021, MNRAS, 506, 741
Moeur, J. 1962, Comptes Rendus Hebdomadaires SÉances Acad. Sci., Paris, 255
Mouri Sardarabadi, A., Leshem, A., & van der Veen, A.-J. 2016, A&A, 588, A95
Murenzi, R. 1989, in Wavelets. Time–Frequency Methods and Phase Space, eds. J.-M. Combes, A. Grossmann, & P. Tchamitchian, 239
Narayan, R., & Nityananda, R. 1986, ARA&A, 24, 127
Ochs, P., Chen, Y., Brox, T., & Pock, T. 2014, SIAM J. Imaging Sci., 7, 1388
Onose, A., Carrillo, R. E., Repetti, A., et al. 2016, MNRAS, 462, 4314
Onose, A., Dubachev, A., & Wiaux, Y. 2017, MNRAS, 469, 938
Pratley, L., McEwen, J. D., d’Avezac, M., et al. 2018, MNRAS, 473, 1038
Rau, U., & Cornwell, T. J. 2011, A&A, 532, A71
Raymond, A. W., Palumbo, D., Paine, S. N., et al. 2021, ApJS, 253, 5
Regpy 2019, regpy: Python tools for regularization methods, https://gitlab.com/regpy/regpy
Roelofs, F., et al. 2022, Galaxies, in press
Schwab, R. F. 1984, AJ, 89, 1076
Starck, J. L., & Murtagh, F. 2006, Astronomical Image and Data Analysis (Springer)
Starck, J.-L., Bijaoui, A., Lopez, B., & Perrier, C. 1994, A&A, 283, 349
Starck, J.-L., Murtagh, F., & Fadili, J. 2015, Sparse Image and Signal Processing: Wavelets and Related Geometric Multiscale Analysis, 2nd edn, 1
Stollnitz, E., Derose, T., & Salesin, D. 1994, IEEE Comput. Graph. Applic., 15
Thiébaut, É., & Young, J. 2017, J. Opt. Soc. Am. A, 34, 904
Thompson, A., Moran, J., & Swenson, G. 1994, Interferometry and Synthesis in Radio Astronomy (Krieger Publishing Company)
Tiede, P., Broderick, A. E., & Palumbo, D. C. M. 2020, ApJ, submitted
Wiaux, Y., Jacques, L., Puy, G., Scaife, A. M. M., & Vandergheynst, P. 2009, J. Fourier Anal. Applic., 14, 877
Wakker, B. P., & Schwarz, U. J. 1988, A&A, 200, 312
Wiaux, Y., Jacques, L., Puy, G., Scaife, A. M. M., & Vandergheynst, P. 2009, MNRAS, 395, 1733
Xiao, J., Ng, M. K.-P., & Yang, Y.-F. 2015, IEEE Trans. Image Process., 24, 1587
Appendix A: Fixed point property of proximity operators

Let $\hat{s} \in \text{argmin}_s H(x)$ and let $x \in \mathbb{X}$. Then it is:

$$H(x) + \frac{1}{2\tau} \|s - \hat{s}\|_{\mathbb{X}} \geq H(\hat{s}) + \frac{1}{2\tau} \|\hat{s} - \hat{s}\|_{\mathbb{X}},$$

(A.1)

as $\hat{s}$ is in the argmin of $H$ and $\tau \geq 0$. Vice versa, let $\hat{s}$ be the solution to $\hat{s} = \text{prox}_{\tau H}(\hat{s})$, then it follows, see Eq. (23):

$$0 = \hat{s} - \hat{s} \in \tau \partial H[\hat{s}],$$

(A.2)

which suffices to show that for a convex, proper, and lower semicontinuous functional, $\hat{x}$ is in the argmin of $H$.

Appendix B: Variation of the regularization parameter

In this subsection we discuss the impact of the regularization parameter $\alpha$. We show in Fig. B.1 the reconstruction of the crescent image and of the disk image, with varying regularization parameter $\alpha$. The leftmost panels show the true image, and the panels to the right of that show the unconstrained reconstructions (e.g., $\alpha = 0$). The middle panels show, from left to right, the reconstruction results obtained with increasing values of $\alpha$. In Table B.1 we present the relative precisions (31) of the different reconstructions. The reconstructions are worse for $\alpha$ that is too small or too big. The best-fit value lies somewhere in between.

The reconstructions for very small $\alpha$ show a greedy resolving model that is too fine, which differs significantly from the true image. Moreover, fainter sidelobes and background emission are visible in the reconstruction. These models overfit the observed visibilities, in other words, the observed visibilities are fitted exactly, but the gaps in the uv-coverage are filled by high-oscillating fits.

For intermediate $\alpha$, the reconstruction is best. The unconstrained reconstruction is modeled with few (due to sparsity) extended, smooth wavelet functions. This approach effectively smooths the fit to the visibilities and fills the gaps in the uv-coverage with smooth fits. In this spirit, the sparsity approach in the wavelet basis has a similar effect on the data, such as total variation, total squared variation, and maximum entropy penalizations. Moreover, sidelobes in the image are suppressed by the hard thresholding.

At the other end of the table, the reconstructions worsen again for reconstruction parameters that are too large. In these cases, the penalty term dominates the objective functional in the forward-backward minimization. The image is modeled with too few wavelet scales. The result is a blurry reconstruction. Moreover, the hard thresholding minimization procedure cuts significant fainter features, for instance, the northern emission in the crescent test image (upper most right panel in Fig. B.1).

### Table B.1. Relative error of the DoG-HiT reconstruction with varying assumptions on the regularization parameter $\alpha$.

|        | $10^{-3}$ | $10^{-2}$ | $10^{-1}$ | $10^{0}$ | $10^{1}$ |
|--------|----------|----------|----------|----------|----------|
| crescent | 0.345    | 0.17     | 0.148    | 0.148    | 0.169    | 0.254    |
| disk    | 0.202    | 0.164    | 0.154    | 0.137    | 0.138    | 0.231    |
Fig. B.1. Reconstructions with varying regularization parameter $\alpha$. Leftmost panels: True images. Middle panels from left to right: $\alpha \in \{0, 10^{-3}, 10^{-2}, 10^{-1}, 10^0, 10^1\}$. 