Dr. Ing. Stephan Wenger
wenger@cg.cg.tu-bs.de
Computer Graphics Lab, TU Braunschweig

Prof. Dr. Ing. Marcus Magnor
magnor@cs.cg.tu-bs.de
Computer Graphics Lab, TU Braunschweig

A Sparse Reconstruction Algorithm for Multi-Frequency Radio Images

Technical Report 2014-11-20

November 26, 2014

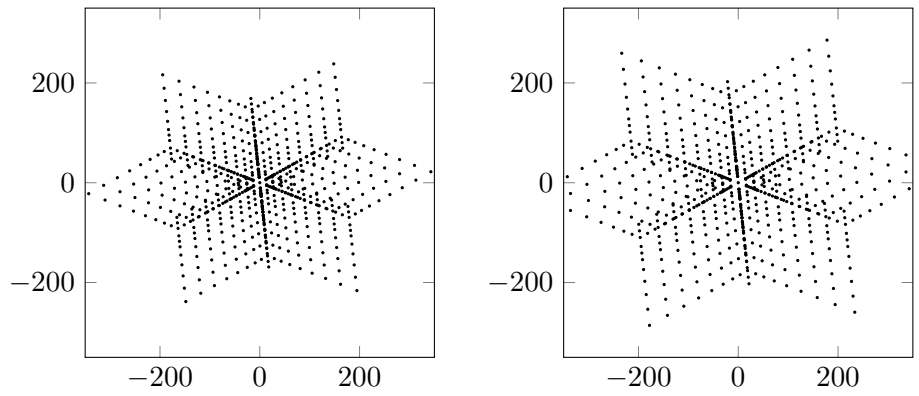
Computer Graphics Lab, TU Braunschweig

http://www.digibib.tu-bs.de/?docid=00058241 28/11/2014

Abstract

In radio interferometry, every pair of antennas in an array defines one sampling point in the Fourier domain of the sky image. By combining information from different wavelengths, sample coverage—and therefore reconstruction quality—can be increased. However, the images at different wavelengths can be dramatically dissimilar; this fact must be taken into account when reconstructing multi-frequency images. In this paper, we present a novel reconstruction algorithm based on the assumption that the spectrum is continuous. In contrast to prior work, we allow for sparse deviations from this assumption: this allows, for example, for accurate reconstruction of line spectra superimposed on a continuum. Using simulated measurements on synthetic multi-frequency images, we show that the proposed approach provides significant improvements over a comparable method based solely on a continuity assumption.

http://www.digibib.tu-bs.de/?docid=00058241 28/11/2014



(a) 100 MHz (b) 120 MHz

Figure 1: Sampling pattern of the VLA "D" configuration at different frequencies in units of the respective wavelength.

1 Introduction

Radio interferometers consist of a set of radio antennas that record radiation from the sky at radio wavelengths [PSB89]. The correlation, or *visibility*, between any two antennas at a given wavelength corresponds to a single point in the Fourier transform of the sky image. The location of this point is given by the vectorial distance between the antennas in units of the wavelength. Modern devices can capture such visibilities for thousands of different wavelengths at once, resulting in separate sparse sampling patterns for each wavelength that are scaled with respect to each other, Figure 1. When data from different wavelengths is combined, the joint sampling pattern exhibits a straight line of sampling points for each pair of antennas; however, each such line contains information about the source at different wavelengths, and a multi-frequency reconstruction algorithm has to account for possible spectral variations within the source.

The reconstruction of artifact-free images from sparse samples in the Fourier domain is essentially a deconvolution problem. As such, it requires assumptions about the structure of the source; traditionally, it is often assumed that at any given wavelength, the source is a sparse superposition of point sources [Hög74] or similar basis functions [Cor08].

Multi-frequency radio observations have been used for several decades [ABS84] since the structure of the spectrum, often characterized by the *spectral index*, provides valuable information about the nature of the source. However, for a long time the images for each wavelength were reconstructed independently.

More recently, algorithms have been developed that make use of the re-

http://www.digibib.tu-bs.de/?docid=00058241 28/11/2014

dundance between images at different wavelengths using a method called *multi-frequency synthesis* [Con91; SW94]. Although sources can look vastly different at different frequencies due to emission and absorption lines and strongly frequency-dependent effects like thermal and synchrotron emission, the spectral behavior of a single pixel often corresponds roughly to a continuous function with few parameters, such as a power law. This assumption can be used to increase sample coverage by combining the scaled sampling patterns from multiple wavelengths. This works well for continuum sources by assuming continuous spectral structure, for example, using Taylor polynomials [Rau12; RC11]. However, the approach is not usable for sources containing structures that vary quickly with frequency, for example, in narrow band observations containing emission or absorption lines superimposed on a continuous spectrum.

Recently, reconstruction algorithms based on compressed sensing have been introduced [Suk09; Wen+10a; Wen+10b; Wia+09; WPV10]. They directly make use of the common sparsity of radio images in the image domain or some transform domain. In theory, the compressed sensing approach can be generalized to comprise more versatile *regularized optimization methods* that minimize a plausibility function, or *regularizer*, subject to constraints that incorporate the observational data. For example, time-dependent observations can be reconstructed by employing a regularizer that favors solutions in which only a few pixels exhibit temporal variation [WRM13].

In this paper, we present a regularized optimization method that allows for the reconstruction of both wide and narrow band multi-frequency radio images. It is based on the assumption that the spectrum is a superposition of a smooth component and sparse local deviations. In contrast to prior approaches relying purely on a continuity assumption, this allows, for example, for accurate recovery of sources containing emission and absorption lines. A central component of our algorithm consists in finding a sparse decomposition of the reconstructed image in an overcomplete basis that contains both delta functions and the basis functions of Chebyshev polynomials.

2 Mathematical background

Assume that a radio interferometer has observed the sky at N_f different frequencies f_i. We aim to recover N_f images of $N_\mathrm{x} \times N_\mathrm{y}$ pixels resolution; these images are combined into a signal vector $\mathbf{y} = \left(\mathbf{y_1}^\mathsf{T}, \ldots, \mathbf{y_{N_\mathrm{f}}}^\mathsf{T}\right)^\mathsf{T}$, where $\mathbf{y}_i$ is the vector containing all pixels from the image for frequency f_i.

The measurement process for each frequency f_i can be written as $\mathbf{b_i} = M_i \mathbf{y_i}$, where each M_i is a sampled (non-equispaced) Fourier transform and $\mathbf{b_i}$ is the vector of measured visibilities for this wavelength. Combining the visibilities for all frequencies into a vector, this yields $\mathbf{b} = \left(\mathbf{b_1}^\mathsf{T}, \ldots, \mathbf{b_{N_\mathrm{f}}}^\mathsf{T}\right)^\mathsf{T} =$

http://www.digibib.tu-bs.de/?docid=00058241 28/11/2014

$M\mathbf{y}$, where the measurement matrix M is given as

$$M = \begin{pmatrix} M_1 & 0 & \cdots & 0 \\ 0 & M_2 & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & M_{N_\mathrm{f}} \end{pmatrix} . \tag{1}$$

For reconstruction, we find a sparse decomposition $\mathbf{x}$ of the signal vector $\mathbf{y}$ in an overcomplete dictionary D. The signal vector can then be recovered from this sparse vector via the relation $\mathbf{y} = D\mathbf{x}$.

The dictionary D contains basis functions for the *spectral* dimension of the signal; these are the same for each pixel. D is therefore the Kronecker product of some per-pixel dictionary D_p and an $N_\mathrm{x} \times N_\mathrm{y}$ identity matrix, $D = D_\mathrm{p} \otimes \mathbb{I}_{N_\mathrm{x} \times N_\mathrm{y}}$.

The per-pixel dictionary D_p contains only the basis functions for the spectrum of a single pixel. We choose to assume that the spectrum consists of a smooth continuous part and sparse line features. This can be represented by writing $D_\mathrm{p} = (D_\mathrm{s}, D_\mathrm{c})$, where D_s is an $N_\mathrm{f} \times N_\mathrm{f}$ identity matrix and

$$D_\mathrm{c} = \begin{pmatrix} T_0(-1) & \cdots & T_{N_\mathrm{c}-1}(-1) \\ \vdots & & \vdots \\ T_0(1) & \cdots & T_{N_\mathrm{c}-1}(1) \end{pmatrix} \tag{2}$$

contains the first N_c Chebyshev polynomial basis functions: $T_i(x)$ is the value of the i^{th} Chebyshev polynomial (of the first kind) at position x, and x runs smoothly from -1 to 1 over all N_f rows.

3 Algorithm

Our algorithm minimizes a regularizer $f(\mathbf{x})$ subject to constraints given by the observed data. We follow the common approach to relax the constraints by replacing them with a squared ℓ_2 norm data term, which is essentially equivalent to χ^2 minimization. The optimization problem then reads

$$\arg\min_{\mathbf{x}} \frac{1}{2} \left\| A\mathbf{x} - b \right\|_2^2 + f(\mathbf{x}) \, , \tag{3}$$

where $A = MD$, and from which the desired signal vector $\mathbf{y}$ can be recovered via $\mathbf{y} = D\mathbf{x}$.

Equation 3 is minimized using a *proximal algorithm* that alternatingly performs a gradient descent step on the data term and a proximal projection on the regularizer. This method has the advantage of allowing non-differentiable regularizers (including, for example, indicator functions to enforce hard constraints), and makes it easy to replace regularizers for different

http://www.digibib.tu-bs.de/?docid=00058241 28/11/2014

applications. Our implementation follows [BT09], which is documented in detail in [WRM13].

To compute the gradient of the data term, $A^\mathsf{T}(A\mathbf{x} - \mathbf{b})$, one requires an efficient implementation of the linear operator $A^\mathsf{T}A = D^\mathsf{T}M^\mathsf{T}MD = D^\mathsf{T}CD$, where C describes a convolution of the image for each frequency with the corresponding point spread function (PSF), the Fourier transform of the sampling pattern. We implement this convolution as a multiplication in a *discrete* Fourier domain for efficiency reasons; this process does not cause any loss of accuracy in the gradient computation. The remaining matrix-vector product $A^\mathsf{T}\mathbf{b}$ can be precomputed and does not significantly influence the runtime performance of the algorithm.

4 Regularizers

The optimization problem stated in Equation 3 requires the choice of a suitable regularizer $f(\mathbf{x})$. We choose a regularizer that combines a term to enforce spatial sparsity of the pixel intensities—independent of frequency—with a term that enforces sparsity of any deviations from the smoothness assumption. For comparison, we also implement an algorithm that is based purely on smoothness of the spectrum and sparsity in the pixel domain by simply choosing $\tau_1 = 0$ and $D_\mathrm{p} = D_\mathrm{c}$.

The proposed regularizer takes the form

$$f(\mathbf{x}) = \tau_{1,\infty} \left\|\mathbf{x}\right\|_{1,\infty} + \tau_1 \left\|\mathbf{x}_\mathrm{p}\right\|_1 \ , \tag{4}$$

where $\left\|\mathbf{x}\right\|_{1,\infty}$ is the sum over all pixels in $\mathbf{x}$ of the ℓ_∞ norm (the maximum amplitude) of the spectrum at that pixel. $\mathbf{x}_\mathrm{p}$ contains only the sparse deviation components of $\mathbf{x}$ (as opposed to the smooth Chebyshev components $\mathbf{x}_\mathrm{c}$). The $\ell_{1,\infty}$ term, $\left\|\mathbf{x}\right\|_{1,\infty}$, encourages *group sparsity* across the spectral domain, so that pixels tend to be either active across multiple frequencies or completely inactive. The ℓ_1 term, $\left\|\mathbf{x}_\mathrm{p}\right\|_1$, penalizes deviations from the smooth polynomial and keeps those deviations sparse.

The optimization algorithm requires computing the *proximal mapping* of $f(\mathbf{x})$,

$$p_f(\mathbf{x}) = \arg\min_{\mathbf{w}} \frac{1}{2} \left\|\mathbf{w} - \mathbf{x}\right\|_2^2 + f(\mathbf{w}) \ . \tag{5}$$

We approximate this function by applying the proximal mappings for the ℓ_1 term and the $\ell_{1,\infty}$ term successively.

The ℓ_1 term depends only on a subset $\mathbf{x}_\mathrm{p}$ of the components of $\mathbf{x}$, and the corresponding proximal mapping is applied to these components only. It has the closed-form solution

$$p_{\beta\left\|\cdot\right\|_1}(\mathbf{x}) = \arg\min_{\mathbf{w}} \frac{1}{2} \left\|\mathbf{w} - \mathbf{x}\right\|^2 + \beta \left\|\mathbf{w}\right\|_1 \tag{6}$$

$$= \max\left(|\mathbf{x}| - \beta, 0\right) \operatorname{sign} \mathbf{x} \ , \tag{7}$$

http://www.digibib.tu-bs.de/?docid=00058241 28/11/2014

where $|\mathbf{x}|$ and $\operatorname{sign}\mathbf{x}$ act element-wise on the vector $\mathbf{x}$. The proximal mapping of the $\ell_{1,\infty}$ norm is computed by applying the proximal mapping of the ℓ_∞ norm to the complete spectrum for each pixel. Our implementation of the proximal mapping of the ℓ_∞ norm is based on [Duc+08].

5 Results

As a test case for our algorithm, we create an artificial narrow band data cube of 32×32 pixel images at 100 equispaced frequencies f_i from 100 MHz to 100.3 MHz. For each pixel with index j, we choose a random spectral index α_j from the interval $[-1,-2]$ to simulate thermal emission. To create some spatial coherence as would be expected in a real-world dataset, the image of spectral indicies is smoothed by a Gaussian kernel with standard deviation $\sigma_\alpha = 5\,\mathrm{px}$. An initial continuous-spectrum data cube is then computed as $y_{ij} = {f_i}^{\alpha_j}$. The complete data cube is subsequently normalized by its mean value to provide a reasonable fixed intensity scale.

In a second step, random line features are superimposed on the data cube by adding Gaussians at 10 random locations (x, y) and frequencies f. The standard deviations along the two spatial axes are chosen uniformly at random from the interval $[0, 5\,\mathrm{px}]$. The standard deviation along the spectral axis is given by the thermal line width $f\sqrt{\frac{k_\mathrm{B}T}{mc^2}}$, where we choose $T = 10000\,\mathrm{K}$ and m equal to the mass of a proton. The line amplitudes are chosen uniformly at random from the interval $[-2, 2]$ and scaled with the value of the continuous-spectrum data cube at (x, y, f).

Finally, the data cube is scaled with a hypothetical single-frequency radio map, Figure 2, that mimics a realistic spatial intensity distribution, and any negative intensities are thresholded to zero. The spectral structure of this synthetic source is exemplified in Figures 3 and 4. On this data cube, we simulate a snapshot measurement with the Very Large Array in the "D" configuration, cf. Figure 1, using a cell size of 10 arcmin.

Using the proposed algorithm, the component vector $\mathbf{x}$ is reconstructed from the simulated data with $N_\mathrm{c} = 3$ Chebyshev basis functions. The optimal set of parameters, $\tau_1 = 2.5$ and $\tau_{1,\infty} = 600$, was found by employing a Nelder-Mead simplex algorithm to minimize the reconstruction error with respect to the ground truth data. Figure 5a shows the relationship between each parameter and the relative reconstruction error when the other parameter is kept at its optimal value. The root mean squared error (RMSE) at the optimum relative to the norm of the ground truth signal is $7.09 \cdot 10^{-2}$. The runtime of this reconstruction (averaged over 20 runs) was $23.9\,\mathrm{s} \pm 2.2\,\mathrm{s}$.

For comparison, we run the same reconstruction with an algorithm based only on smoothness as described in Section 4. To allow for a fair comparison, the optimal parameter, $\tau_{1,\infty} = 325$, was again found through optimization and comparison to the ground truth data. The relative RMSE in this case is

http://www.digibib.tu-bs.de/?docid=00058241 28/11/2014

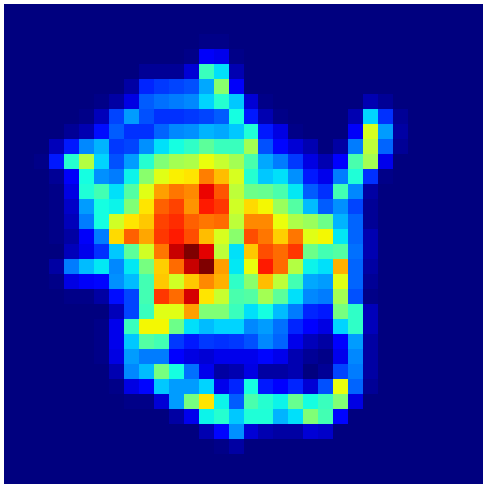


Figure 2: Hypothetical single-frequency radio map used in the experiments.

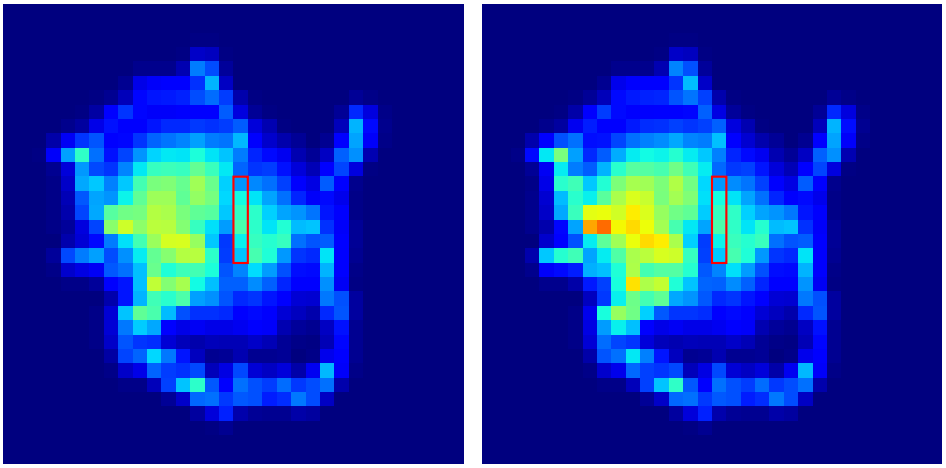
$9.94 \cdot 10^{-2}$, or about 40 percent worse than with the proposed approach. The dependence between $\tau_{1,\infty}$ and the RMSE for the smoothness only approach is plotted in Figure 5b. The runtime of this reconstruction (averaged over 20 runs) was $14.4\,\mathrm{s} \pm 1.9\,\mathrm{s}$. Examples of the reconstructed spectra for both approaches are shown in Figure 4.

To ensure that the applicability of the proposed algorithm is not limited to the narrow band use case, we repeat the experiment for a wide band dataset otherwise similar to the previously presented one, Figure 6. The optimal parameters are $\tau_1 = 4.6$ and $\tau_{1,\infty} = 132$ for the proposed algorithm and $\tau_{1,\infty} = 132$ for the smooth variant. The relationship between parameters and reconstruction error is plotted in Figure 7. The relative RMSE of the proposed approach is $3.80 \cdot 10^{-2}$ at a runtime of $22.8\,\mathrm{s} \pm 1.3\,\mathrm{s}$. The smooth variant yields a relative RMSE of $4.11 \cdot 10^{-2}$ at a runtime of $14.5\,\mathrm{s} \pm 0.5\,\mathrm{s}$, making the error of the prior method about 8 percent worse than that of the proposed approach.

6 Conclusion

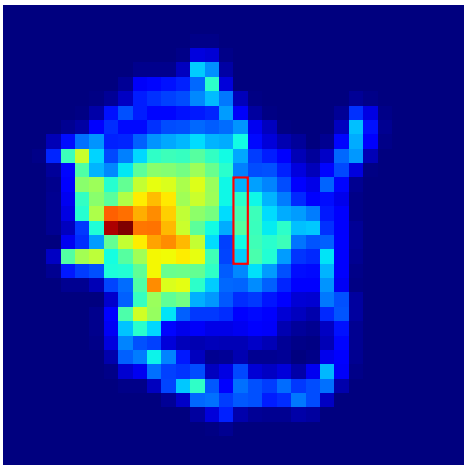
We have presented an algorithm for sparse reconstruction of multi-frequency radio images. In contrast to prior methods based solely on a continuity assumption for the spectrum, we allow for (but penalize) sparse deviations from the smooth background spectrum. Our experiments on a simulated narrow band scenario containing emission and absorption lines superimposed on a continuum show that the proposed approach significantly outperforms purely smoothness-based methods in this setting. This increase in reconstruction quality, however, comes along with a comparable increase in runtime. In a

http://www.digibib.tu-bs.de/?docid=00058241 28/11/2014

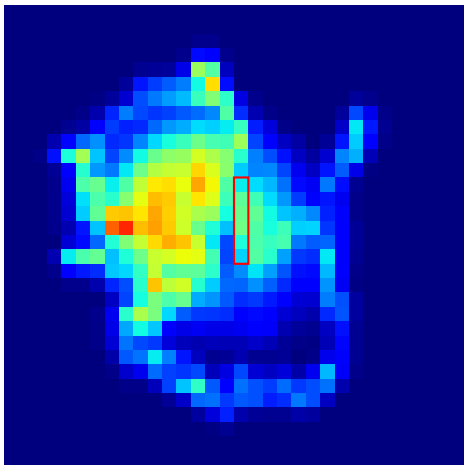


(a) 100 009 091 Hz

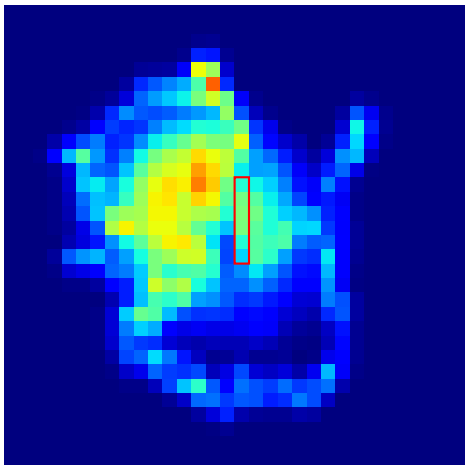
(b) 100 012 121 Hz



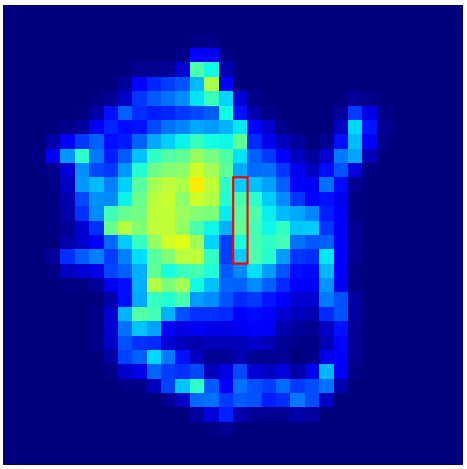
(c) 100 015 152 Hz



(d) 100 018 182 Hz



(e) 100 021 212 Hz



(f) 100 024 242 Hz

Figure 3: Images from the synthetic test data set for different frequencies. Spectra for the highlighted pixels are plotted in Figure 4.

http://www.digibib.tu-bs.de/?docid=00058241 28/11/2014

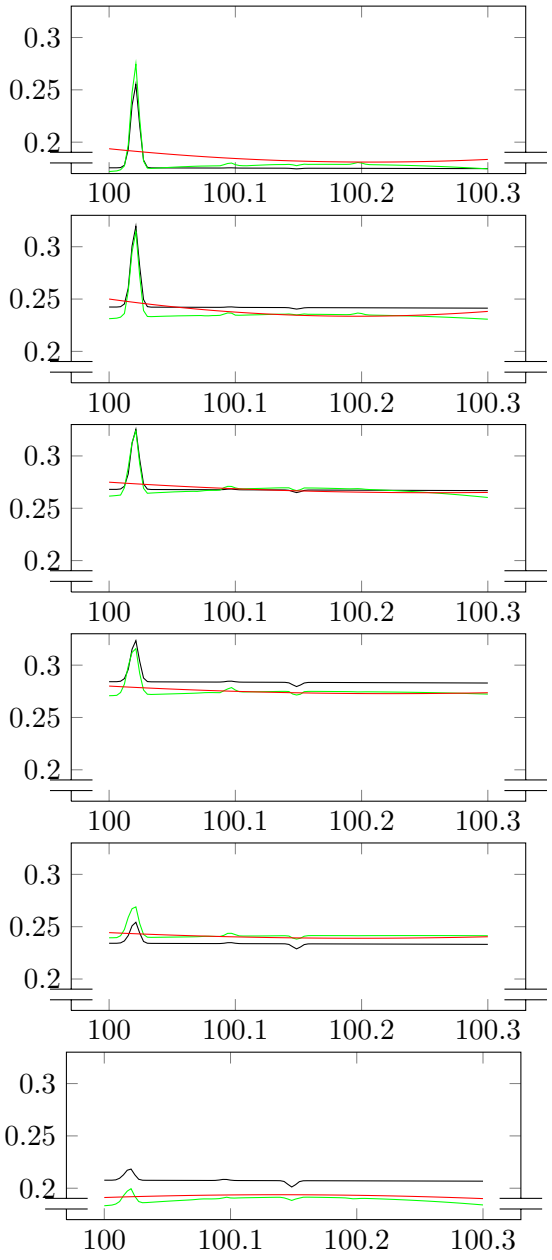
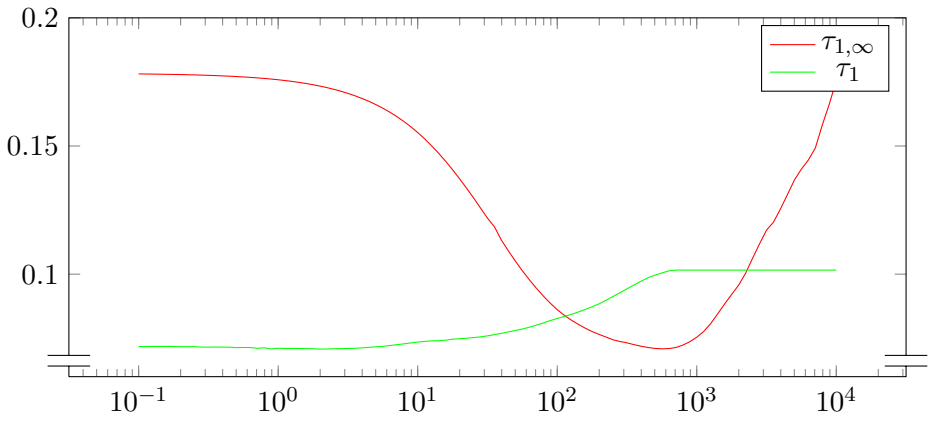
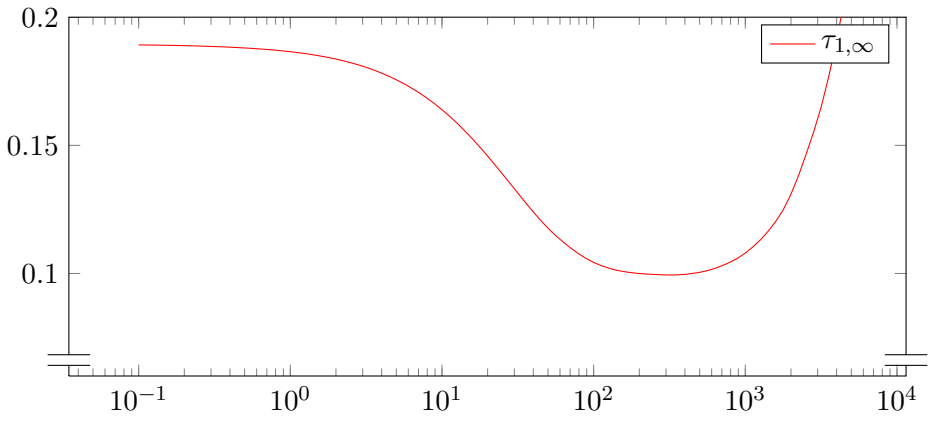


Figure 4: Narrow band spectra for the pixels highlighted in Figure 3, plotted over frequency in MHz: ground truth (black), smoothness only (red), proposed approach (green).

http://www.digibib.tu-bs.de/?docid=00058241 28/11/2014



(a) sparse



(b) smooth

Figure 5: Relative reconstruction errors (in percent) for the smooth and sparse algorithms plotted against different parameter values in a simulated narrow band scenario.

http://www.digibib.tu-bs.de/?docid=00058241 28/11/2014

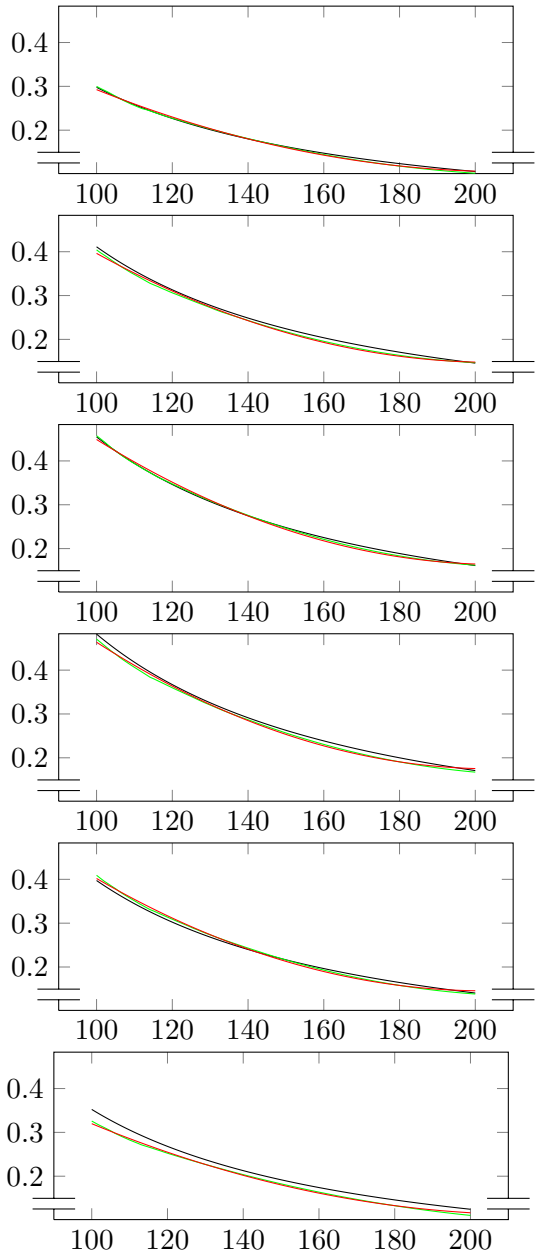
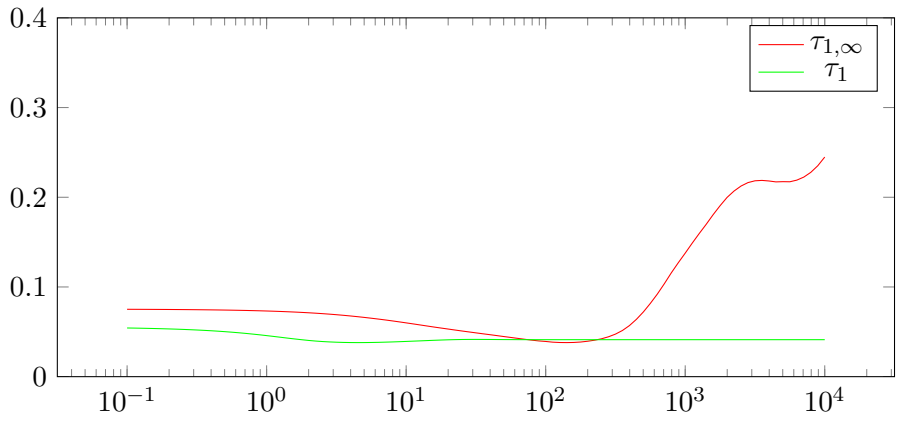
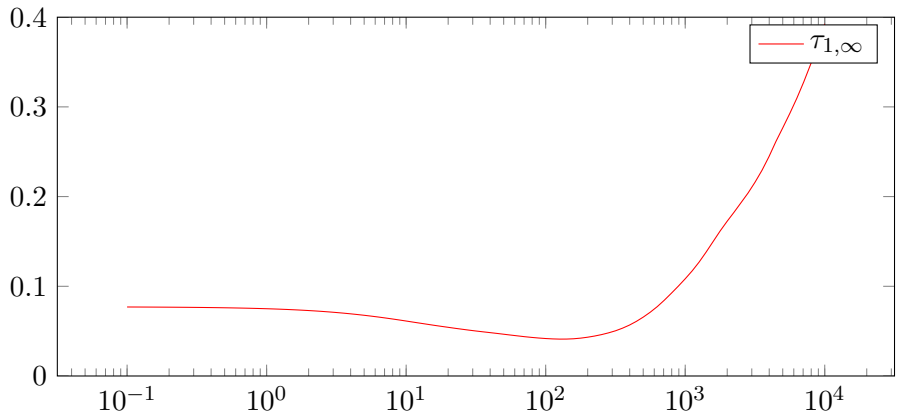


Figure 6: Wide band spectra for the pixels highlighted in Figure 3, plotted over frequency in MHz: ground truth (black), smoothness only (red), proposed approach (green).

http://www.digibib.tu-bs.de/?docid=00058241 28/11/2014



(a) sparse



(b) smooth

Figure 7: Relative reconstruction errors (in percent) for the smooth and sparse algorithms plotted against different parameter values in a simulated wide band scenario.

http://www.digibib.tu-bs.de/?docid=00058241 28/11/2014

wide band setting with no apparent line features, the improvement in reconstruction quality is only marginal at a comparable runtime cost. On the whole, the proposed approach thus provides benefits in both narrow and wide band settings, whereas the optimal trade-off between reconstruction quality and algorithm runtime may depend on the specific application.

Acknowledgments

This work was supported in part by a Feodor Lynen alumni grant from the Alexander von Humboldt foundation, Germany.

http://www.digibib.tu-bs.de/?docid=00058241 28/11/2014

References

[ABS84] P. Alexander, M. T. Brown, and P. F. Scott. "A multi-frequency radio study of Cygnus A". In: *Monthly Notices of the Royal Astronomical Society* 209.4 (1984), pp. 851–868.

[BT09] A. Beck and M. Teboulle. "A fast iterative shrinkage-thresholding algorithm for linear inverse problems". In: *SIAM J. Imaging Sciences* 2.1 (2009), pp. 183–202.

[Con91] J. E. Conway. "Multi-frequency synthesis". In: *IAU Colloq. 131: Radio Interferometry. Theory, Techniques, and Applications*. Vol. 19. 1991, pp. 171–179.

[Cor08] T. J. Cornwell. "Multiscale CLEAN Deconvolution of Radio Synthesis Images". In: *IEEE J. Selected Topics in Signal Processing* 2 (2008), pp. 793–801.

[Duc+08] J. Duchi, S. Shalev-Shwartz, Y. Singer, and T. Chandra. "Efficient projections onto the ℓ_1-ball for learning in high dimensions". In: *Proc. International Conference on Machine Learning*. 2008, pp. 272–279.

[Hög74] J. A. Högbom. "Aperture Synthesis with a Non-Regular Distribution of Interferometer Baselines". In: *Astronomy and Astrophysics Supplement* 15 (1974), pp. 417–426.

[PSB89] R. A. Perley, F. R. Schwab, and A. H. Bridle, eds. *Synthesis Imaging in Radio Astronomy*. Vol. 6. Conference Series. Astronomical Society of the Pacific, 1989.

[Rau12] U. Rau. "Radio interferometric imaging of spatial structure that varies with time and frequency". In: *Image Reconstruction from Incomplete Data VII*. Vol. 8500. Proc. SPIE Optical Engineering + Applications. 2012, pages.

[RC11] U. Rau and T. J. Cornwell. "A multi-scale multi-frequency deconvolution algorithm for synthesis imaging in radio interferometry". In: *Astronomy & Astrophysics* 532 (2011), A71.

[Suk09] A. Suksmono. "Deconvolution of VLBI images based on compressive sensing". In: *Proc. International Conference on Electrical Engineering and Informatics*. Vol. 1. 2009, pp. 110–116.

[SW94] R. J. Sault and M. H. Wieringa. "Multi-frequency synthesis techniques in radio interferometric imaging". In: *Astronomy and Astrophysics Supplement Series* 108 (1994), pp. 585–594.

[Wen+10a] S. Wenger, S. Darabi, P. Sen, K.-H. Glassmeier, and M. Magnor. "Compressed Sensing for Aperture Synthesis Imaging". In: *Proc. IEEE International Conference on Image Processing*. 2010, pp. 1381–1384.

http://www.digibib.tu-bs.de/?docid=00058241 28/11/2014

[Wen+10b] S. Wenger, M. Magnor, Y. Pihlström, S. Bhatnagar, and U. Rau. "SparseRI: A Compressed Sensing Framework for Aperture Synthesis Imaging in Radio Astronomy". In: *Publications of the Astronomical Society of the Pacific* 122.897 (2010), pp. 1367–1374.

[Wia+09] Y. Wiaux, L. Jacques, G. Puy, A. Scaife, and P. Vandergheynst. "Compressed sensing imaging techniques for radio interferometry". In: *Monthly Notices of the Royal Astronomical Society* 395 (2009), pp. 1733–1742.

[WPV10] Y. Wiaux, G. Puy, and P. Vandergheynst. "Compressed sensing reconstruction of a string signal from interferometric observations of the cosmic microwave background". In: *Monthly Notices of the Royal Astronomical Society* 402.4 (2010), pp. 2626–2636.

[WRM13] S. Wenger, U. Rau, and M. Magnor. "A Group Sparsity Imaging Algorithm for Transient Radio Sources". In: *Astronomy and Computing* 1 (2013), pp. 40–45.

http://www.digibib.tu-bs.de/?docid=00058241 28/11/2014