Optimal Dictionaries for Sparse Solutions of Multi-frame Blind Deconvolution

B. R. Hunt
Pacific Defense Solutions/Integrity Applications Inc.

Keith T. Knox
Air Force Research Laboratory

ABSTRACT

Abstract: Sparse representation of data has grown rapidly in signal processing. The benefits of sparse regularization are economy of representation of many different varieties of data, as well as control of difficult aspects of inverse problems, e.g., regularization of ill-conditioned inverse problems. Herein we represent atmospheric turbulence point-spread-functions by training optimal overcomplete dictionaries from atmospheric turbulence data. Implications for blind- deconvolution of turbulent images are discussed. The application of sparse dictionaries is demonstrated by the employment of sparse PSF representations to formulate a multi-frame blind deconvolution (MFBD) algorithm. We present results of the gain in MFBD image reconstruction by simulations of turbulent atmospheric images and the reconstruction of the corresponding images with the sparse PSF MFBD algorithm.

1. INTRODUCTION

Atmospheric turbulence remains a problem for a variety of requirements in the imaging of distant objects, e.g., astronomical sources. This has led to the development of methods for mitigating the effects of turbulence using *blind deconvolution techniques*, which require no knowledge of the instantaneous point-spread-function (PSF) that is present in the degradation of an image frame acquired by short image exposure times [1-4]. A recurring and successful theme in blind deconvolution techniques is the development and application of various mathematical models for the properties of the atmospheric turbulence mechanisms that are associated with the formation and acquisition of imagery degraded by turbulence. Another recurring theme is the formulation of the blind deconvolution solution as an optimization problem. An example of the degree of sophistication that is associated with these methods is the modeling of atmospheric turbulence in the entrance pupil of an optical system by phase screens, using the basis functions of the Zernike polynomials [4]. In these methods the optimization problem is to search the space of coefficients of the Zernike polynomials that are assumed (for a circular aperture) to describe the instantaneous phase screen, which represents the distortion of a plane wave propagating from an object through layers of the atmosphere containing variations in the index of refraction that are caused by thermal and kinetic induced turbulence. Making the problem one of "blind" deconvolution is that there is no known description for the object, so that the space of possible pixel amplitudes of the unknown object is also simultaneously searched with the space of Zernike representation coefficients. We here use the term "searched" in the sense of standard optimization methods, i.e., the "search" is guided by embedding the various Zernike coefficients and object pixel amplitudes into a merit or error function, and the minimization of this merit function is achieved by methods known since the time of Newton and Gauss, e.g., line search in the merit function, gradient of the merit function, the Hessian matrix, etc.

One of the most significant developments in signal processing in the last decade is the application of sparse, redundant and overcomplete representations of data and systems [5]. The core of these methods is the application of signal models that are more general than the orthogonal functions exemplified by the sine/cosine bases of Fourier theory. For example, the common image data compression algorithm known as JPEG was derived from the experimental discovery that the Discrete Cosine Transform functions make a basis set that gives a parsimonious representation of a signal. This was superseded by the introduction of the Wavelet Transform in the JPEG2000 algorithm, which is more efficient in compression because the Wavelet Transform is more adaptive to specific characteristics of the data. The better results obtained with the Wavelet Transform is now known to be the result of the Wavelet Transform having the characteristics of inducing sparsity that is very specifically "tuned" to the actual behavior of the data, rather than abstractly derived from the Fourier series properties that are actually based on the solution to a specific class of differential equations.

An important concern for sparse and redundant methods is the construction of *dictionaries*, i.e., collections of basis functions to compose (i.e., synthesize) any arbitrary function. A number of studies have shown that the data being

analyzed and processed by sparse and redundant representations can be used to construct custom dictionaries that are tuned to, and optimal for, a chosen task [6]. The use of a custom dictionary has been applied to the problem of blind deconvolution [7]. However, the optimal dictionary construction of [7] was based on the application of quasi-maximum-likelihood (QML) methods for a single frame. In the following we step beyond the single frame QML methods of [7] and describe a method for using atmospheric PSFs to directly derive sparse and redundant dictionary elements from multiple short exposure image frames.

2. TRAINING A SPARSE DICTIONARY

Mathematical models of atmospheric turbulence are usually associated with describing the overall ensemble properties of a turbulent environment, without description of any single PSF realization of turbulence. This is not useful if it is desired to create a sparse and redundant dictionary with elements derived from actual turbulent PSFs. Because a collection of atmospheric turbulent PSFs can be obtained by imaging a point source over an optical path, it is of interest to determine if PSF data can be the source for constructing sparse and redundant dictionaries that are optimal for a specific set of conditions of turbulent imaging.

To demonstrate the construction of optimal dictionaries for atmospheric PSFs a set of 200 atmospheric turbulent PSFs was simulated. The simulation used wave-optics propagation of a point source from an altitude of 25,000 meters through a five-layer turbulence model. The turbulent wavefronts were propagated to sea level, with each of the five layers being 5,000 meters in thickness. The turbulence properties were derived from the Maui4 model, which is based on measurements of mid-Pacific conditions on the island of Maui, Hawaii [8]. The turbulent PSFs were obtained for each of the 200 simulations as images, of a remote point source, in the focal plane of a telescope of 50 cm pupil diameter. In order to characterize the simulated turbulence, the average of all 200 PSFs was computed. The Fourier transform of the average of the PSFs was then computed, making an estimate of the long-term average atmospheric OTF of the simulated turbulence. Because there are analytical formulations of the power spectrum of the long-term average OTF of atmospheric turbulence, comparisons of the long-term analytical OTF with the computed values for the simulated OTFs indicated that the ratio of telescope diameter to the simulated turbulence corresponded to a D/r_0 of approximately 5, as indicated in Fig. 1. This indicates a mild degree of turbulence. It should be noted that this initial mild level of turbulence was a deliberate choice, so that concerns in developing the related MFBD algorithm would not be masked or conflated with difficult turbulence conditions. We also simulated much greater levels of turbulence, i.e., turbulence with D/r_0 values of 33.

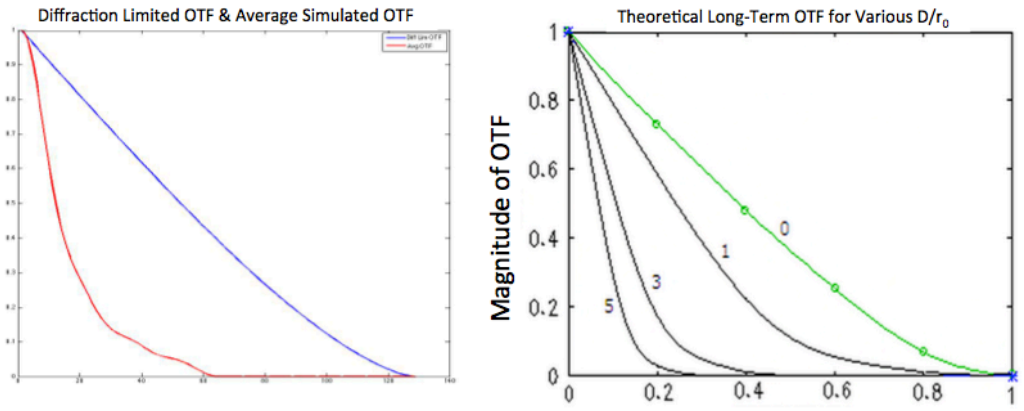

Figure 1: Average Simulated Atmospheric OTFs compared to theoretical OTFs for various D/r_0

Next the collection of 200 atmospheric turbulent PSFs was split into two different sets, each set corresponding to a total of 100 PSFs. The first set was used as the input to a sparse dictionary training algorithm, known as the K-SVD algorithm (MATLAB code for this algorithm is available from the sources disclosed in the text by Elad [9]). For input to the K-SVD algorithm the turbulent PSF images were decomposed into sliding and overlapping blocks of 8 x 8 pixels. The dictionary algorithm was then parameterized to produce a dictionary of 256 elements, each element being 8 x 8 pixels, i.e., a set of small blocks that can be used to construct a representation of any PSF acquired under

the atmospheric conditions that are simulated in the creation of the complete set of 200 PSFs. Fig. 2 displays the 256 elements of the dictionary as 8 x 8 pixel images. Note that individual dictionary elements have a preponderance of simple shapes resembling a wide variety of compact functions. The close juxtaposition of alternating light and dark elements in the dictionary elements shows positive / negative weighting of corresponding image regions. This creates spatial differentiation, with a variety of directions and spacing. The presence of this spatial differentiation in the dictionary is strong evidence of the sparse properties induced by training on the PSFs. It is well known that the favorable sparse properties of the Wavelet Transform are due to the differentiation effects of the construction of wavelets by high-pass and low-pass filters. Further, there is a frequently used image sparsity method, known as total variation, which explicitly computes spatial derivatives [10]. Any single turbulent PSF is composed by an appropriate weighted summation of these dictionary elements, thus using the spatial derivative aspects of the dictionary to impose sparsity in the PSF representations over this dictionary.

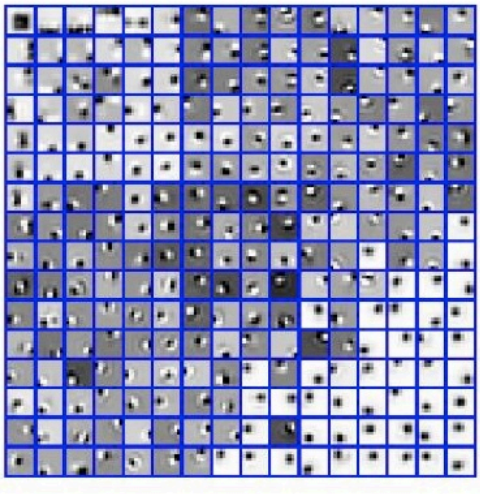

Fig. 2: Dictionary Trained from PSFs

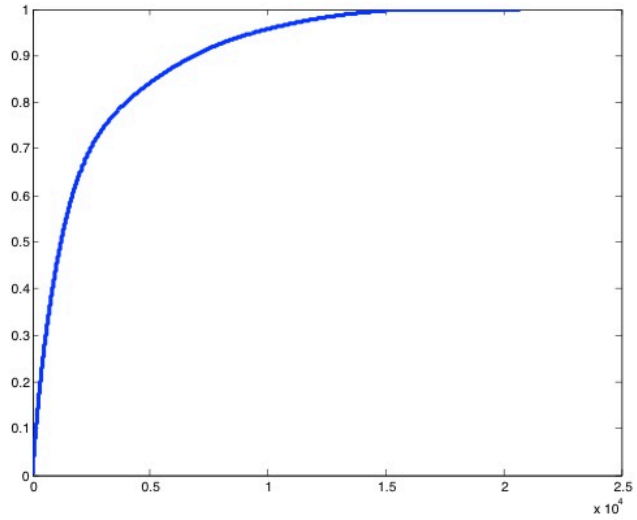

Fig. 3: Cumulative Distribution of Coefficients

To demonstrate the sparsity from using the dictionary in Fig. 2 a point source image was transformed by dictionary of Fig. 2 using eq. (1), as described in the immediate next section. A plot of the cumulative distribution of total amplitude of the transformed samples from the dictionary encoding is seen in Fig. 3. Of significance in Fig. 3 is that only 20% of the transformed values contain 90% of the amplitude, a demonstration of sparsity even for the "broadband" properties of a point source. This is a sharp contrast to the fact that the amplitude distribution of a point source in the Fourier basis set contains 20% of the amplitude in 20% of the components, i.e., the amplitude Fourier distribution of a point source is equal in amplitude in all components. Thus, there is significant sparsity induced by the custom trained dictionary.

A further test was the reconstruction of all the second set of 100 PSFs using the inverse transformation of eq. (2), in the section below, and the dictionary trained from the first set of PSFs. The second set of PSFs was recovered by the dictionary of the first set of PSFs with a signal-to-noise ratio of 39.1 Db, i.e., the residual power of the difference between the recovered second set of PSFs and the original second set was 39.1 Db less than the power of the original second set. It is important to note that the representation of the second set of PSFs was achieved using only the dictionary trained on the second set. Thus, the reconstruction of the second set of PSFs was independent of any literal specific PSF, and only the commonality of PSF statistics in both sets of PSFs was responsible for the highly accurate reconstruction of the second set from the custom dictionary of the first set.

4. SPARSITY IN MULTI-FRAME BLIND DECONVOLUTION

The employment of sparse representations in multi-frame blind deconvolution (MFBD) has been used for several years, although not explicitly understood or acknowledged as such. We refer here to the application of Zernike polynomials, which are an efficient (i.e., sparse) way to parameterize the various aberrations of a waveform, in the development of multi-frame blind deconvolution algorithms. For example, see [11]. The Zernike polynomials

expand the phase screens of the wavefront deviations in a set of circular symmetric functions, which is much more efficient, and thereby sparse, than using a point-by-point description of the wavefront deviation at every position in the pupil of the optical system. The requirement for modeling and description of atmospheric turbulence by a corresponding sparse representation of the turbulent PSFs shifts the emphasis for description of atmospheric turbulence from the pupil plane to the image plane, because the PSFs are a phenomenon present in the image plane. Thus, we chose to construct a MFBD algorithm on the basis of computations solely in the image plane.

A recent publication demonstrated reconstruction of turbulent images with a MFBD algorithm operating in the image plane. The algorithm described in [12] is also an incremental algorithm, i.e., it does not collect all turbulent images into one large batch, with the MFBD algorithm required to operate simultaneously on all the image data. Instead, the first image is processed producing a result that is crude. However, the result of the first MFBD processing step is combined with processing of the next image. This use of the previous result with another image in sequence means that the algorithm can operate on each new turbulent image frame without waiting for all turbulent image frames to be present. This incremental or "on-line" approach was first pointed-out in another MFBD algorithm that also operated solely in the image plane without pupil plane processes [13].

The transformation of a signal into the space spanned by a basis function dictionary is a simple linear operation [6]:

$$p_d = \mathrm{D}p \tag{1},$$

where D is a matrix containing the dictionary elements, p is a vector of signal samples, and p_d is the signal vector projected into the dictionary space. This representation is broad and encompasses the representation of signals in the familiar orthogonal Fourier basis (where D is a matrix of samples from the complex exponential function), wavelet transforms, and the more general non-orthogonal bases of sparse representations. The transformation back to the original signal space is performed by:

$$p_r = \mathrm{D}^{\mathrm{T}} p_d \tag{2}.$$

In an orthonormal dictionary the vector pr will be equal to the original vector p, i.e., the representation is unique. In the case of the sparse and redundant dictionaries, such as trained by the K-SVD algorithm we have used in this work, the representation is not unique. It is necessary, therefore, to have means to recover the best representation over a specific set of dictionary atoms. A number of so-called "Greedy Algorithms" have been developed to do this recovery over a set of dictionary elements. We used the Orthogonal Matching Pursuit (OMP) algorithm [14].

We modified the image domain MFBD algorithm of Hirsch *et al* [13] to use sparse representation of PSFs from the dictionary trained from the first 100 PSFs, as described in the previous section. The MFBD algorithm of Hirsch *et al* consists of two distinct processes. The first process is a series of iterations on the image of each frame, with the input to each iteration being the result from the previous iteration and a new unprocessed frame. These iterations adjust the PSF from a nominal starting value to decrease the error between the new input frame and the back-substitution of the PSF and the previous frame in a forward model. This is a minimizing optimization that uses an approximation for non-negative least squares [13, see eq, (2) –(10)]. At the end of these iterations the PSF estimate has been adjusted to the best fit with the new image data frame. Next there is a similar minimization process using the newly computed PSF to adjust for best fit and minimization of the back-substitution error.

Out modification of the Hirsch *et al* algorithm was minor in computational impact. During the iterations on the PSF the sequence of iterations was periodically interrupted and the PSF computed up to that point in the PSF iterations was sparsified by the application of the OMP algorithm with the trained PSF dictionary. This forces the PSF computation to maintain sparse properties. Computational experiments demonstrated that it was not necessary to frequently sparsify during the iterations on the PSF. Good results were obtained when only a small percentage of PSF iterations was sparsified, e.g., 5-10% of total PSF iterations were sparsified.

5. RESULTS

In the following we some results from the simulations and the sparse PSF algorithm for MFBD. Fig. 4 is typical of

two different simulated turbulent images. The object is the Hubble Space Telescope (HST). As stated above, the dictionary training used the first 100 of the total of the simulated PSFs. The second set of 100 PSFs was used to simulate 100 images of the HST acquired in atmospheric turbulence, each of dimension 256 x 256 pixels. The noise present in each of each of these two images is low, both with Signal-to-Noise Ratio (SNR) of 18.5 Db. The left image is as acquired through the mild turbulence, D/r_0 of approximately 5. The left image shows a more extreme example of turbulence, with the D/r_0 value estimated (as described above by comparison to analytical calculations of D/r_0) to be approximately 33. Certainly, in the second image the major features and structures of the HST are all but obliterated and the strength of the turbulence is made vivid.

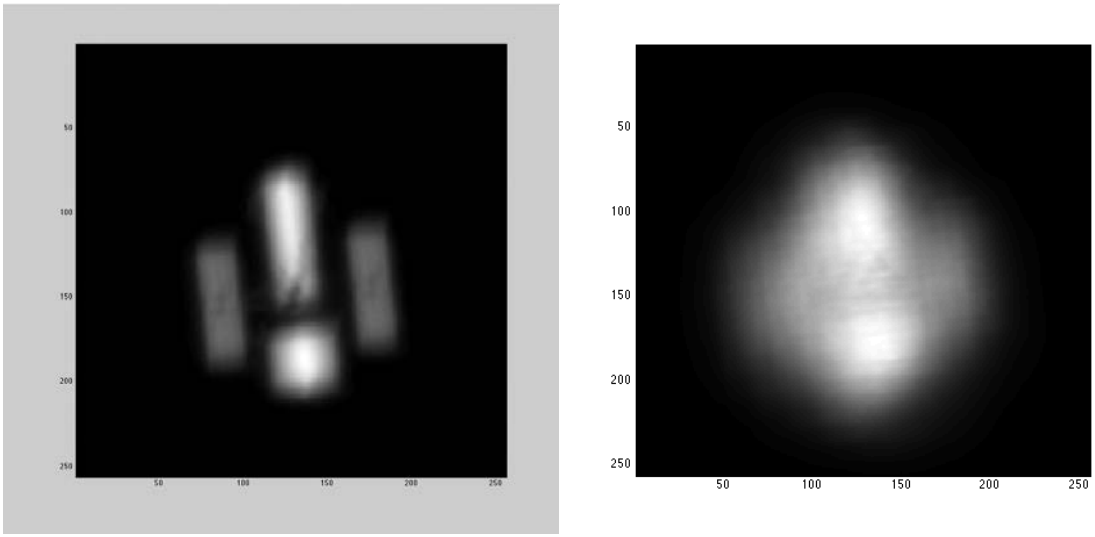

Fig. 4: HST simulated in mild turbulence (left) and in strong turbulence (right)

In Fig. 5 we see the results of the sparse PSF MFBD algorithm applied to images simulated with the second set of 100 PSFs in the case of mild turbulence. The image on the left is the simulated diffraction limited image of the HST, and the image on the right is the reconstructed image HST from the 100 simulated images in mild turbulence. The resulting detail and feature sharpness comes close to the diffraction-limited. HST image.

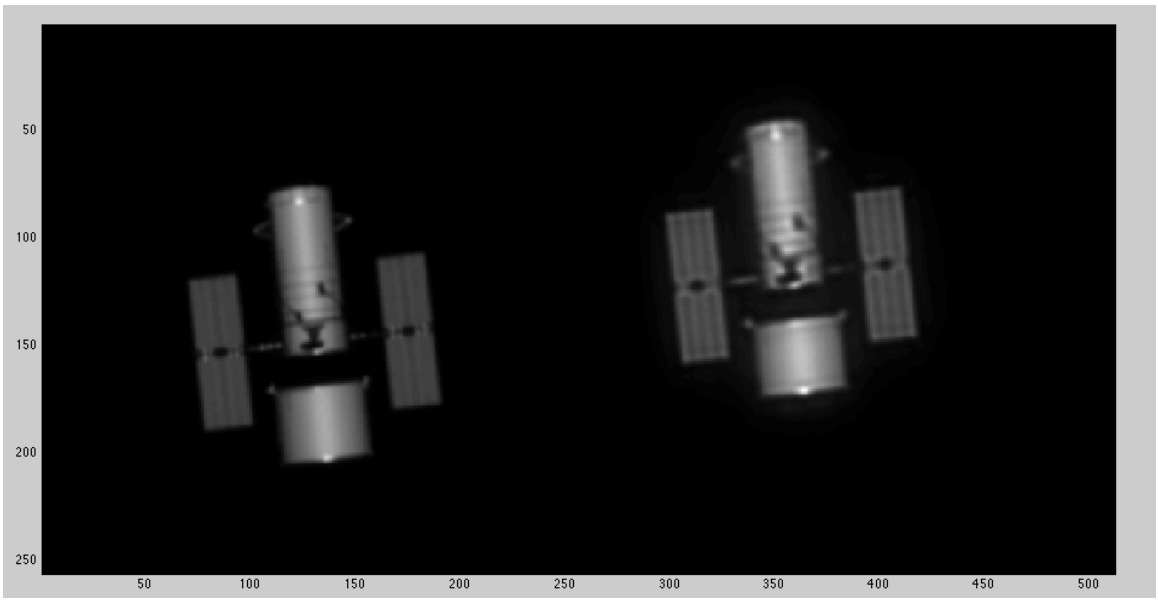

Fig. 5 Diffraction limited HST image (left) and recovered by the sparse PSF MFBD algorithm (right)

The next pair of images, Figure 6, show the reconstruction of the HST image from the case of strong turbulence. The recovery of detail and sharpness is not as satisfactory as in the case of mild turbulence.

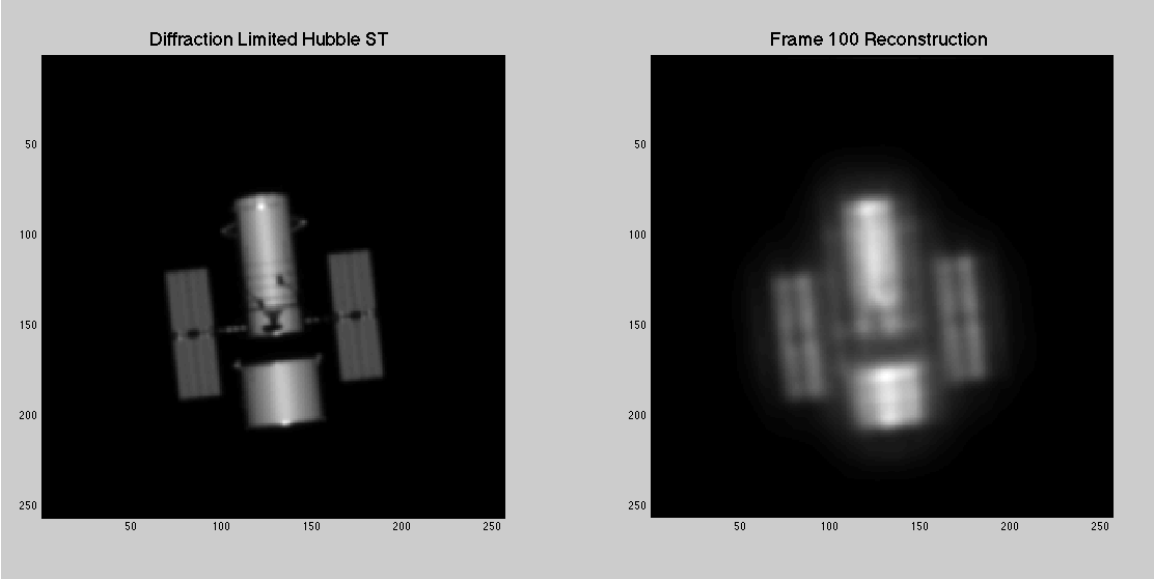

Fig. 6 Diffraction-limited HST image (left) and image (right) from the Sparse MFBD method in strong turbulence

To quantify the beneficial effects of recovery with sparse PSF we computed the SNR between the diffraction-limited HST image and the reconstruction of the HST for two different cases. The MFBD algorithm is constructed so that sparsity can be forced on the PSFs computed through successive iterations of the algorithm. However, the sparsity enforcement can be "turned off", and then the algorithm carries out MFBD exactly as described in Hirsch *et al* [12]. The SNR (in Db) of the reconstruction is calculated by the expression:

$$SNR = \frac{1}{N}\sum_{1}^{N} \frac{\sum (Hubble_{DL})^2}{\sum (Hubble_{DL} - Hubble_{RE})^2} \quad N = \text{number of frames}$$

Here $Hubble_{DL}$ is the diffraction-limited Hubble image and $Hubble_{RE}$ is the reconstructed image from the 100 simulated atmospheric turbulence degraded images of the HST. The SNR expression computes an SNR that is the ratio of the power in the diffraction limited HST image to the difference between the diffraction-limited image and the reconstructed image. Table 1 shows that in both cases, of mild and strong simulated turbulence, the SNR is greater for the reconstruction with sparsity in the PSFs than for reconstruction by the same algorithm with no sparsity imposed on the PSFs computed in the algorithm. For context, recall that 3 Db represents an increase in SNR of a factor of 2. Thus, the calculated SNR values show an increase in SNR (and equivalent decrease in reconstruction error) corresponding to a factor of approximately 2.8.

Turbulence	Non-Sparse vs. Sparse Reconstruction SNR
Mild	4.5 Db. increase in SNR for Sparse case
Strong	4.4 Db. increase in SNR for Sparse case

Table 1: SNR Improvement for using Sparse PSFs in MFBD

6. CONCLUDING REMARKS

There are two different conclusions to draw from the results presented herein. First, the use of sparsity is of benefit to multi-frame blind deconvolution. This is supported by the fact that the use of Zernike polynomials in many

versions of MFBD is well known. The results shown herein indicate that sparsity achieves a significant reduction in the error of the MFBD reconstruction. This we attribute to connecting the sparse basis functions directly to PSFs of atmospheric turbulence, and by tuning the sparsity with custom dictionaries that characterize the turbulent PSF. Second, the use of sparsity in the manner of custom dictionaries also shifts the MFBD modeling from the pupil plane to the image plane. This offers the opportunity to achieve MFBD in cases where pupil plane phase screen modeling is known to fail, i.e., when the turbulence becomes so strong that amplitude changes are introduced into the pupil plane, and the simple phase screen models of Zernike polynomials is not appropriate. We have not shown the extension to this application yet, but it offers the promise of a new approach for an important problem in imaging through atmospheric turbulence.

6. REFERENCES

[1] J-L Starck, "Deconvolution and blind deconvolution in astronomy" in *Blind Image Deconvolution: Theory and Applications*, P Campisi, K Egiazarian (Eds.), (CRC Press, 2007), Chap. 1

[2] A Cornelio, E Piccolomini, J Nagy, "Constrained numerical optimization methods for blind deconvolution", Numerical Algorithms, **65**, 23-42 (2014)

[3] M Roggemann, B Welsh, *Imaging Through Turbulence*, CRC Press (1996)

[4] C Matson, K Borelli, "Parallelization and automation of a blind deconvolution algorithm", in *Proceedings, High Performance Computing Modernization Program Users Group Conference*, Denver (2006)

[5] M Elad, *Sparse and Redundant Representations*, Springer (2010)

[6] *Ibid*, Chap. 12.

[7] M Bronstein, *et al* "Blind Deconvolution of Images Using Optimal Sparse Representations," *IEEE Transactions on Image Processing*, **14**, 726-735 (2005)

[8] Bradford, L. W. "Maui4: a 24 hour Haleakala turbulence profile," in *Proceedings of the AMOS Technical Conference* (2010)

[9] Elad, *Op.Cit.*

[10] Bruckstein, et al, *Op. Cit*.

[11] Matson, C., *et al*, "Fast and optimal multiframe blind deconvolution algorithm for high-resolution ground-based imaging of space objects," *Applied Optics*, vol. 48, pp 75-92, 2009

[12] Hirsch, M., *et al*, "Online multi-frame blind deconvolution with super-resolution and saturation correction," *Astron. And Astrophys.*, vol. 531/A9, 2011.

[13] Sheppard, D, *et al*, "Iterative multiframe superresolution algorithms for atmospheric-turbulence-degraded imagery," *J Opt. Soc. Am – A*, vol. 15, pp 978-992, 1998.

[14] Elad, *Op.Cit.*