

Removal of Spectro-Polarimetric Fringes by 2D PCA

R. CASINI¹ AND W. LI¹

¹*High Altitude Observatory, National Center for Atmospheric Research[†]*
P. O. Box 3000, Boulder, CO 80307-3000, USA

ABSTRACT

We investigate the application of 2-dimensional Principal Component Analysis (2D PCA) to the problem of removal of polarization fringes from spectro-polarimetric data sets. We show how the transformation of the PCA basis through a series of carefully chosen rotations allows to confine polarization fringes (and other stationary instrumental effects) to a reduced set of basis “vectors”, which at the same time are largely devoid of the spectral signal from the observed target. It is possible to devise algorithms for the determination of the optimal series of rotations of the PCA basis, thus opening the possibility of automating the procedure of de-fringing of spectro-polarimetric data sets. We compare the performance of the proposed method with the more traditional Fourier filtering of Stokes spectra.

1. INTRODUCTION

The advent of new instrumentation for the investigation of solar phenomena—in particular, with regard to the short- and long-term variability of the Sun’s magnetic activity, and its impact on the near-Earth environment—has put unprecedented demands on the conception of data management plans and infrastructures, which aim to deliver reliable, science-ready data products to the community in a timely fashion.

The U.S. community is getting ready for the completion and first light of the 4-m Daniel K. Inouye Solar Telescope (DKIST; Tritschler et al. 2016) in early 2020, with its suite of complex post-focus instruments, most of which will have polarimetric capabilities, in order to detect the subtle signatures of the Sun’s magnetic field in its light spectrum. The 1.6-m Goode Solar Telescope (GST; Cao et al. 2010) has already started to reveal the complexity of the solar spectrum when unprecedented higher spatial and temporal resolutions are attained in solar observations. Other countries are pursuing similar endeavors, such as India’s 2-m National Large Solar Telescope (NLST; Hasan et al. 2010), the 4-m European Solar Telescope (EST; Collados et al. 2013), and the 8-m annular-mirror Chinese Giant Solar Telescope (CGST; Liu, Deng, & Ji 2014).

Beyond the intrinsic complexity of the theoretical problem of how polarized radiation is formed and transported through the solar atmosphere (e.g., Stenflo 1994; Landi Degl’Innocenti & Landolfi 2004; Casini & Landi Degl’Innocenti 2008; Trujillo Bueno 2010, for a description of the various mechanisms at play), the need to detect and confidently measure very small levels of polarization (down to 0.01%

Corresponding author: R. Casini
casini@ucar.edu

[†]The National Center for Atmospheric Research is sponsored by the National Science Foundation.

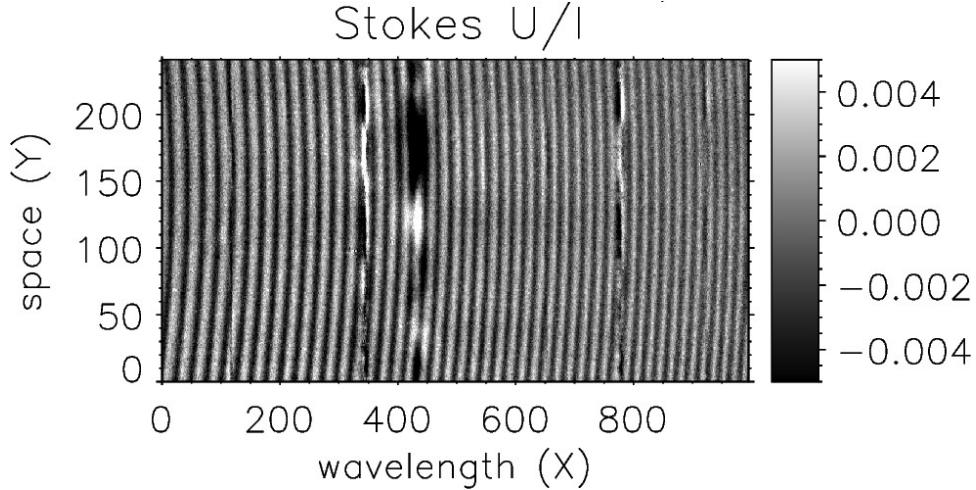


Figure 1. Example of a polarized spectrum (in this case, the fractional linear polarization, Stokes U/I) of the solar radiation around the 1085 nm. The three dominant features from the left are: the photospheric line of Si I at 1082.7 nm ($X \simeq 350$); the chromospheric triplet of He I around 1083 nm ($X \simeq 430$); and the photospheric line of Si I at 1084.4 nm ($X \simeq 780$). We note the presence of strong polarization fringes across the entire spectrum.

of the intensity, for the most demanding scientific applications) puts very hard requirements on the identification and removal of instrumental artifacts from the detected signals.

One recurrent issue in spectro-polarimetric instruments is the appearance of polarization “fringes” that overlap with the actual spectral signal from the observed target. These fringes are interference patterns that are produced by the presence of optical elements in the system of the telescope and the instrument having varying phase retardance properties (e.g., around their optical axis). Such components include polarization modulators, polarizing beam-splitters, and any optical element where parallel optical interfaces may occur (e.g., interference filters, detector windows). These fringes have the appearance of more or less regular 2-dimensional (2D) patterns, preferentially arranged along the spectral dimension of the data (see Figure 1). We refer to review studies of polarized fringes (Lites 1991; Semel 2003; Clarke 2004) for a thorough description of this phenomenon, and to recent work by Harrington et al. (2017) on the use of Berreman calculus for the modeling of fringes in polarimetric instrumentation.

The most common fringe correction methods used in spectro-polarimetric data reduction are Fourier filtering and derivations from it, such as wavelet analysis (Rojo & Harrington 2006). However, Fourier filtering only works satisfactorily when the frequency domains of the fringes and of the target spectral signal (in the spectral and/or spatial domains) are clearly separated, and it is severely limited when the fringe pattern’s period or amplitude vary over the image. Wavelet analysis is a powerful method for removing smoothly varying fringes in flat-field images. However, it becomes difficult to control in the presence of strong spectral signals from the observed target.

In this paper, we extend previous work on the identification and isolation of polarization fringes by Principal Component Analysis (PCA; Pearson 1901; Jolliffe 2002). Casini, Judge, & Schad (2012) considered the implementation and performance of a 2D PCA algorithm based on the method described by Yang et al. (2004; hereafter, Y-PCA). In that approach, the spectral data gets “contracted” over the spatial dimension before performing PCA decomposition. As a result, the PCA basis of

eigenfeatures consists of spectro-polarimetric profiles (rather than 2D images similar to Figure 1), which are akin to spatial averages of the principal components (PCs) of the data. One advantage of the Y-PCA approach is the fast convergence of the singular-value series, which in theory allows to limit the number of PCs used for the data reconstruction to a small set of the lowest order basis vectors. A beneficial side effect is that unwanted, high-order spatial modulations of the signals that tend to cancel out when spatially averaged, such as detector noise, get confined to the lower end of the ordered set of PCs, so a low-pass cut of this set is a very efficient way to remove these unwanted features from the data (see Casini, Judge, & Schad 2012). Not surprisingly, these benefits come at a cost. It may often be hard to tell fringes and spectral signals apart (with the exception of the most dominant fringes), and the compression of spatial information in the Y-PCA algorithm causes high-order spatial variations of the spectral signals to be distributed over many PCs, which therefore must be included for the reconstruction of the signal, if such spatial variations are to be preserved, e.g., in high resolution observations.

In this work, we consider the application to the polarization fringe problem of the original 2D PCA algorithm of Turk & Pentland (1991; hereafter, TP-PCA), which has traditionally been applied to the problem of face recognition. The possibility of using this approach to fringe removal was briefly touched upon in the conclusions of Casini, Judge, & Schad (2012). One downside of the algorithm pointed out in that paper is that the singular value series shows a much slower convergence than in the Y-PCA method, and the dimensionality of the basis of eigenfeatures corresponds to the number of images in the data set to be analyzed, which can in some cases be a small number (even just one). In practice, this implies that often one must keep the entire set of PCs in order to reconstruct the data without significant loss of information, especially for small sets of images.

On the other hand, the advantage of the TP-PCA method is that the contribution of the fringes to the data is immediately recognizable in the set of eigenfeatures, which in principle should make the removal of such contribution an easier task than in the Y-PCA method. However, since most or all of the PCs must be included in the reconstruction in order to preserve the original spatial and spectral information of the data, no practical benefit can come from this, unless the fringes somehow are confined to just a few eigenfeatures essentially devoid of spectral signal, which then can be dropped from the data reconstruction without any risk of signal loss.

The purpose of this paper is to apply a simple manipulation technique of the PCA basis, in order to optimally attain the isolation of the fringes in a few basis vectors. This technique relies on the intrinsic orthogonality of the PCA basis, and uses rotations on the 2D subspaces generated by all possible pairs of eigenfeatures in the basis, so that the corresponding transformed 2D subspace has one of the two basis vectors clean of fringes (as far as possible). By iteratively processing the entire basis, it is possible to *automatically* confine the fringes to a minimal set of transformed basis vectors. Because this transformation of the PCA basis is constructed as a sequence of isometries, the process naturally preserves the orthonormality of the original PCA basis after rotation.

The idea of rotations applied to the PCA basis in order to bring out particular physical properties of the original data set is not new, and it has also been applied to the problem of spectro-polarimetric inversion of photospheric lines (Skumanich & López Ariste 2002). The technique presented here can indifferently be applied to both the TP-PCA and Y-PCA approaches. In the case of the Y-PCA method, this manipulation technique can significantly improve the separation of fringes and spectral

signals in the set of basis vectors, compared to the direct application of the method as presented in Casini, Judge, & Schad (2012).

In Section 2, we present the mathematical concepts and assumptions for the separation of spectral signals and unwanted instrument artifacts in a spectro-polarimetric data set. In Section 3, we show some applications of the basis manipulation technique to the PCA of Stokes data for both the TP-PCA and Y-PCA approaches. The rotation algorithm is summarized in App. B.

2. FORMULATION OF THE PROBLEM

We consider a 2D spectro-polarimetric data set, consisting of an ensemble of N “Stokes images”, having spectral wavelength (or frequency) as the x coordinate, and the spatial position along the slit of a grating-based spectrograph as the y coordinate. Alternatively, the data may come from an imaging spectro-polarimeter, where the spectral signal is acquired through wavelength scanning of a tunable filter instrument (e.g., using a Fabry-Perot interferometer; Cavallini 2006). The data can therefore be represented by the set $\mathcal{D} \equiv \{\mathcal{D}_i(x, y)\}_{i=1,\dots,N}$.

The implementation of 2D PCA according to the TP-PCA method allows to represent each image in the data set as a linear expansion over a basis (i.e., an orthonormal set) of eigenfeatures $\mathbf{e}_j(x, y)$, where $j = 1, \dots, N$,

$$\mathcal{D}_i(x, y) = \sum_{j=1}^N c_{ij} \mathbf{e}_j(x, y) + \bar{\mathcal{D}}(x, y) , \quad i = 1, \dots, N , \quad (1)$$

where $\bar{\mathcal{D}}(x, y)$ is the average image of the data set. The orthonormality condition corresponds to the following definition of inner product on the space \mathfrak{E} generated by the basis $\{\mathbf{e}_j(x, y)\}_{j=1,\dots,N}$,

$$\langle \mathbf{e}_j, \mathbf{e}_k \rangle \equiv \int_{D_{XY}} dx dy \mathbf{e}_j(x, y) \mathbf{e}_k(x, y) = \delta(j, k) , \quad (2)$$

where D_{XY} is the area of the image, and $\delta(j, k)$ is Kronecker’s δ . In the usual implementation of the TP-PCA algorithm (see Turk & Pentland 1991), the basis $\{\mathbf{e}_j(x, y)\}_{j=1,\dots,N}$ is determined via diagonalization of the $N \times N$ covariance matrix $\text{Cov}(\mathcal{D}, \mathcal{D})$. This can be calculated through a simple matrix multiplication,

$$\text{Cov}(\mathcal{D}, \mathcal{D}) = (\mathbf{D} - \bar{\mathbf{D}})^T (\mathbf{D} - \bar{\mathbf{D}}) , \quad (3)$$

where \mathbf{D} is a $(N_X N_Y) \times N$ matrix, whose i -th column corresponds to the $\mathcal{D}_i(x, y)$ array of size $N_X \times N_Y$ reformed into a column vector of length $N_X N_Y$ (see App. A), whereas $\bar{\mathbf{D}}$ is a similarly constructed matrix, where each column is the average of the set of reformed $\mathcal{D}_i(x, y)$ arrays across the map.¹ If we indicate with \mathbf{U} the (orthogonal) matrix of the column eigenvectors of (3), then the normalized basis of eigenfeatures is given by the set (cf. App. A, eq. (A5))

$$\mathbf{e}_j(x, y) = \lambda_j^{-1/2} \sum_i [\mathcal{D}_i(x, y) - \bar{\mathcal{D}}(x, y)] U_{ij} , \quad j = 1, \dots, N , \quad (4)$$

where the λ_j ’s are the corresponding eigenvalues. It is straightforward to verify that this is indeed an orthonormal set (see App. A).

¹ Equation (3) corrects eq. (6) of Casini, Judge, & Schad (2012) in their description of the Turk & Pentland (1991) method.

It is important to remark that, in general, the average image $\bar{\mathcal{D}}(x, y) \notin \mathfrak{E}$, when the basis is determined via diagonalization of the covariance matrix (3). This can also be gleaned directly from eqs. (1) and (4).

In practice, the data will include the desired spectro-polarimetric signal (\mathcal{S}) of the target science diagnostics of the observations alongside with unwanted instrumental signatures, such as detector artifacts and polarization fringes (\mathcal{F}), so that $\mathcal{D}_i(x, y) = \mathcal{S}_i(x, y) + \mathcal{F}_i(x, y)$, for $i = 1, \dots, N$. The problem of de-fringing the data obviously corresponds to the removal of the contribution $\mathcal{F}_i(x, y)$ for each image in the data set. Because the average image $\bar{\mathcal{D}}(x, y)$ will in general contain unwanted instrumental signal—unlike in the problem of face recognition to which the TP-PCA algorithm is usually applied—the inability of the basis $\{\mathbf{e}_j(x, y)\}_{j=1, \dots, N}$ to generate $\bar{\mathcal{D}}(x, y)$ is highly inconvenient. In fact, the need to add back the average image to the reconstructed data, according to the standard algorithm, implies that some of the unwanted instrumental signal is uncontrollably put back into the reconstructed data. For this reason, we adopt a variant of the algorithm, where the basis is determined instead via diagonalization of the correlation matrix (see App. A, eq.(A1))

$$\text{Corr}(\mathcal{D}, \mathcal{D}) = \mathbf{D}^T \mathbf{D} , \quad (5)$$

so that eqs. (1) and (4) become, respectively,

$$\mathcal{D}_i(x, y) = \sum_{j=1}^N c_{ij} \mathbf{e}_j(x, y) , \quad i = 1, \dots, N , \quad (6)$$

$$\mathbf{e}_j(x, y) = \lambda_j^{-1/2} \sum_i \mathcal{D}_i(x, y) U_{ij} , \quad j = 1, \dots, N . \quad (7)$$

Because of the different physical origin of the spectral and instrumental signals (assuming that the instrument behaves as a linear system), the signals are in principle statistically independent, i.e., *uncorrelated*. However, the type and conditions of the observations represented by the (limited) data set \mathcal{D} might be such that this condition of uncorrelation is not fully realized in the data. For the development below, we assume that the set \mathcal{D} indeed satisfies such condition, which is expressed algebraically as

$$\text{Corr}(\mathcal{S}, \mathcal{F}) = 0 . \quad (8)$$

The (i, j) element of this correlation matrix is the inner products $\langle \mathcal{S}_i, \mathcal{F}_j \rangle$ (see App. A), and so the condition (8) of uncorrelation between \mathcal{S} and \mathcal{F} is equivalent to the orthogonality condition between the corresponding spaces, i.e., $\mathcal{F} = \mathcal{S}_\perp$ in \mathcal{D} . This implies that \mathcal{D} is the direct sum of \mathcal{S} and \mathcal{F} , i.e., $\mathcal{D} = \mathcal{S} \oplus \mathcal{F} = \mathcal{S} \oplus \mathcal{S}_\perp$, and so the union of any two bases of \mathcal{S} and \mathcal{F} is also a basis of \mathcal{D} (see, e.g., Birkhoff & Mac Lane 1953, Ch. VII). Therefore, there must exist an orthogonal transformation (isometry) \mathbf{R} of the PCA basis $\{\mathbf{e}_j(x, y)\}_{j=1, \dots, N}$ into the basis $\{\mathbf{\epsilon}_k(x, y)\}_{k=1, \dots, N}$,

$$\mathbf{\epsilon}_k(x, y) = \sum_{j=1}^N R_{kj} \mathbf{e}_j(x, y) , \quad (9)$$

such that \mathcal{S} is generated (after a possible re-ordering of the basis) by the subset $\{\mathbf{\epsilon}_k(x, y)\}_{k=1, \dots, n}$, for some $n < N$, whereas $\mathcal{F} = \mathcal{S}_\perp$ is generated by the complementary subset $\{\mathbf{\epsilon}_k(x, y)\}_{k=n+1, \dots, N}$.

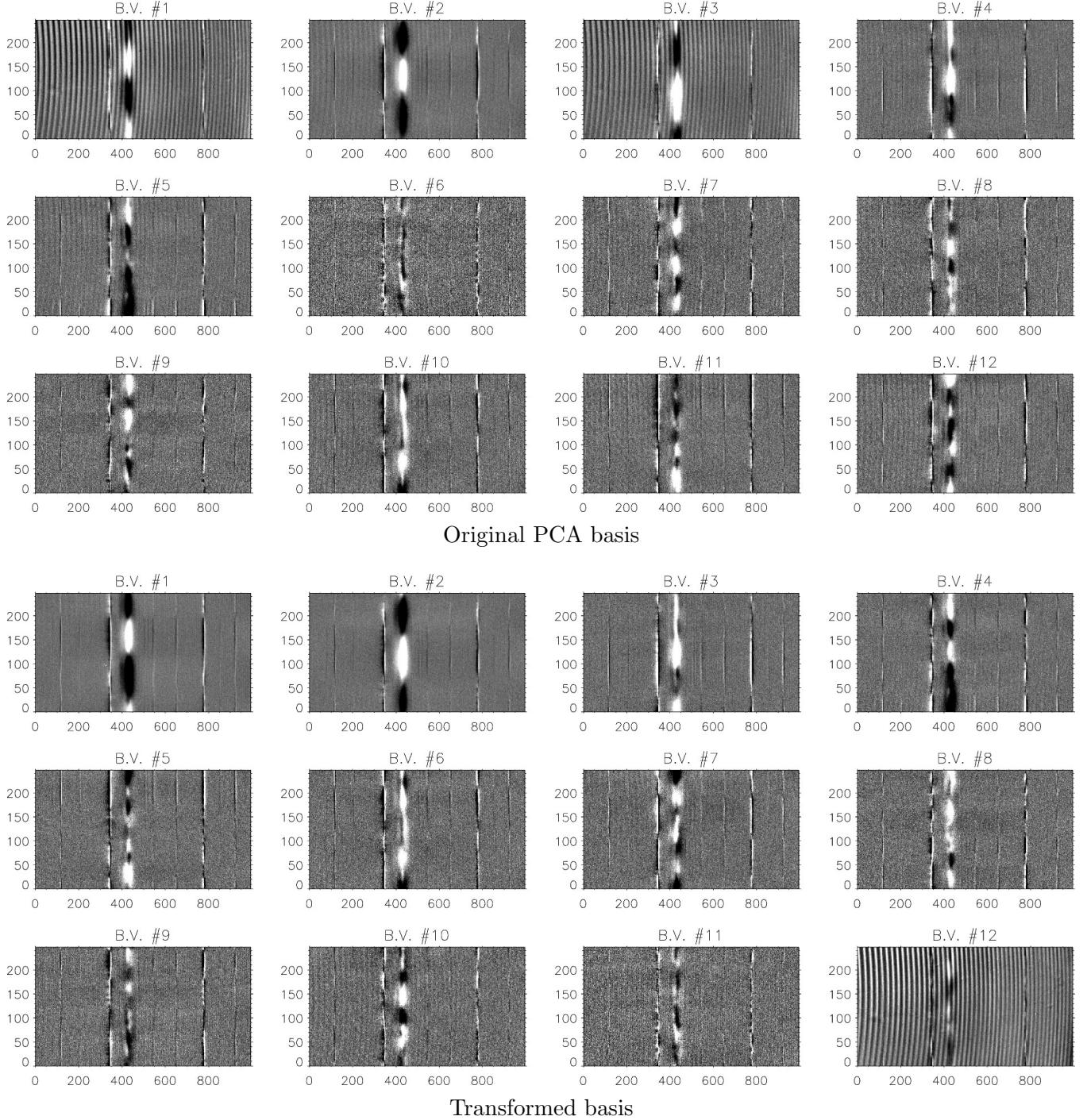


Figure 2. *Top:* the TP-PCA basis $\{e_j(x, y)\}_{j=1, \dots, 12}$ of a Stokes V map with $N = 12$ frames. Note the presence of strong fringes in the dominant basis vectors (B.V.; esp. #1 and #3). The same fringes affect to a minor degree multiple others elements of the basis. *Bottom:* the transformed basis $\{\epsilon_k(x, y)\}_{k=1, \dots, 12}$. Note that the fringes are very well confined to the subspace spanned by the last transformed basis vector. We also note that the residual target signal \mathcal{S} in $\epsilon_{12}(x, y)$ is visibly smaller than the one in both $e_1(x, y)$ and $e_3(x, y)$ of the original basis. Hence, reconstruction of the data set by dropping $\epsilon_{12}(x, y)$ in the PCA reconstruction suppresses the fringes with less science data loss than by using the original basis without $e_1(x, y)$ and $e_3(x, y)$ (see Figure 3).

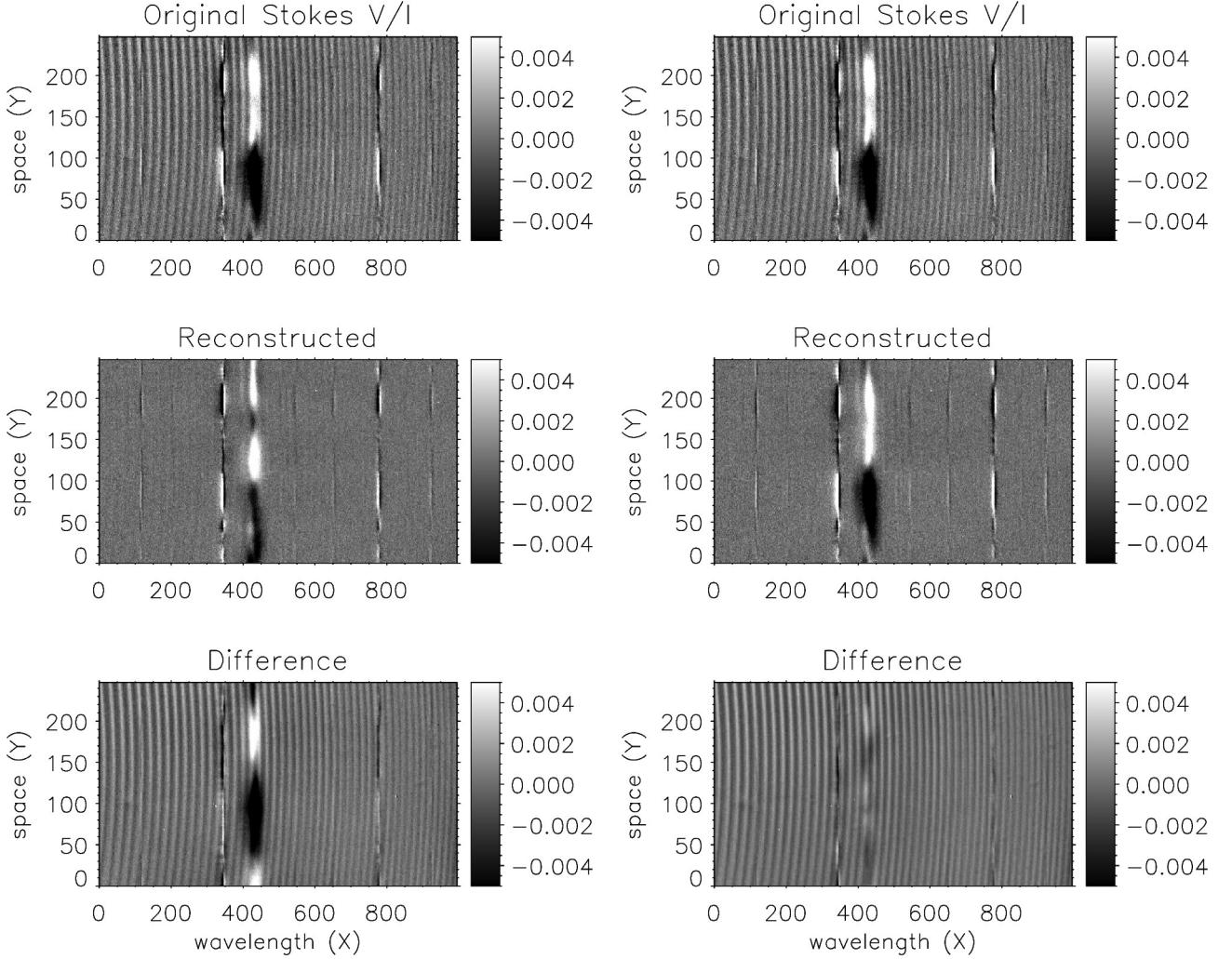


Figure 3. Original (top), PCA-reconstructed (center), and residual of the reconstruction (bottom) for one frame of the same data set used for the creation of the bases of Figure 2. *Left:* using the original TP-PCA basis without $\epsilon_1(x, y)$ and $\epsilon_3(x, y)$. *Right:* using the transformed basis without $\epsilon_{12}(x, y)$.

This allows us to rewrite eq. (6) as

$$\mathcal{D}_i(x, y) = \sum_{k=1}^N \gamma_{ik} \epsilon_k(x, y) , \quad (10)$$

where

$$\gamma_{ik} = \sum_{j=1}^N c_{ij} \mathbf{R}_{kj} = \sum_{j=1}^N c_{ij} \mathbf{R}_{jk}^T . \quad (11)$$

Similarly, for the de-fringed data, we have

$$\mathcal{S}_i(x, y) = \sum_{k=1}^n \gamma_{ik} \epsilon_k(x, y) . \quad (12)$$

The objective of the de-fringing process by 2D PCA is therefore to determine the isometry \mathbf{R} that produces a basis set $\{\epsilon_k(x, y)\}_{k=1, \dots, N}$ with the properties given above. The realizability of this

objective may in practice be hindered by the observing conditions, the nature of the target, and the limitedness of the data set, as these can prevent the full realization of the uncorrelation condition (8) within a specific data set \mathcal{D} .

Despite this practical limitation, it is always possible to devise an algorithm that determines an optimal set of rotations between two any eigenfeatures with the purpose of eliminating the unwanted signal \mathcal{F} from one of them. In App. B, we list the steps of a simple procedure that can easily be automated. By recursively applying that algorithm through the basis $\{\mathbf{e}_j(x, y)\}_{j=1, \dots, N}$ of \mathcal{D} , the unwanted signal can thus be confined to a *minimal* (i.e., no longer reducible) set of transformed basis vectors $\{\mathbf{\epsilon}_k(x, y)\}_{k=n+1, \dots, N}$. Any residual spectral signal in such set practically quantifies the departure of the uncorrelation condition (8) from being fully realized within the data set \mathcal{D} .

It is important to remark that, in general, the transformed basis $\{\mathbf{\epsilon}_k(x, y)\}_{k=1, \dots, N}$ no longer represents a basis of PCs of the data set, as it no longer corresponds to the original matrix \mathbf{U} of eigenvectors of the correlation matrix $\text{Corr}(\mathcal{D}, \mathcal{D})$ via eq. (7). In particular, this implies that the original set of eigenvalues cannot be used to quantify the relative contributions of the transformed basis vectors to the data (see end of Sect. 3.1).

3. EXAMPLES

The concepts presented in the previous section are well demonstrated by Figures 2 and 3. For these, we considered a subset of 12 frames taken from a spectro-polarimetric map of a solar active region in the He I 1083 nm multiplet, observed in 2011 with the Facility Infra-Red Spectro-polarimeter (FIRS; Jaeggli et al. 2010) at the Dunn Solar Telescope (DST) of the National Solar Observatory on Sacramento Peak (NSO/SP; Sunspot, NM). The figures show results of the application of the TP-PCA approach to the Stokes V (circular polarization) signal.

Figure 2 shows the original TP-PCA basis $\{\mathbf{e}_j(x, y)\}_{j=1, \dots, N}$ (top) and the transformed basis $\{\mathbf{\epsilon}_k(x, y)\}_{k=1, \dots, N}$ (bottom), the latter being derived by the recursive application of the algorithm described in App. B. The original basis shows the presence of fringes in several basis vectors, most notably in the eigenfeatures $\mathbf{e}_1(x, y)$ and $\mathbf{e}_3(x, y)$. The determination of the PCA basis relies on the Singular Value Decomposition (SVD) of the correlation matrix, which produces an ordered set of eigenfeatures according to the decreasing importance of their contribution to the data set. Therefore, the presence of strong spectral signals in low-order eigenfeatures is of great concern, since a simple elimination of those eigenfeatures in order to suppress the fringe signal would cause a significant loss of the spectral signal targeted by the science. By applying the optimal rotation algorithm described in App. B, the fringe signal in each eigenfeature is moved down the transformed basis $\{\mathbf{\epsilon}_k(x, y)\}_{k=1, \dots, N}$, until it is completely constrained within the last few basis vectors. Figure 2 demonstrates that, for this specific data set, the rotation algorithm is particularly effective, practically limiting the presence of the unwanted signal \mathcal{F} to just the last basis vector $\mathbf{\epsilon}_{12}(x, y)$.

While the basis transformation generally modifies the “importance” ordering of the original PCA basis, the fundamental result of the transformation is to accomplish the reduction—and ideally, the full suppression—of the targeted spectral signal in the fringe eigenfeatures, essentially realizing the process of orthogonalization of \mathcal{S} and \mathcal{F} within the basis of the data set \mathcal{D} . Of course, perfect orthogonality (and hence, uncorrelation) in the data set is implied only when the transformed basis vectors generating \mathcal{F} are completely devoid of the spectral signal \mathcal{S} . This is certainly not the case for the transformed basis of Figure 2. However, the residual spectral signal in the last basis vector $\mathbf{\epsilon}_{12}(x, y)$ has been significantly reduced.

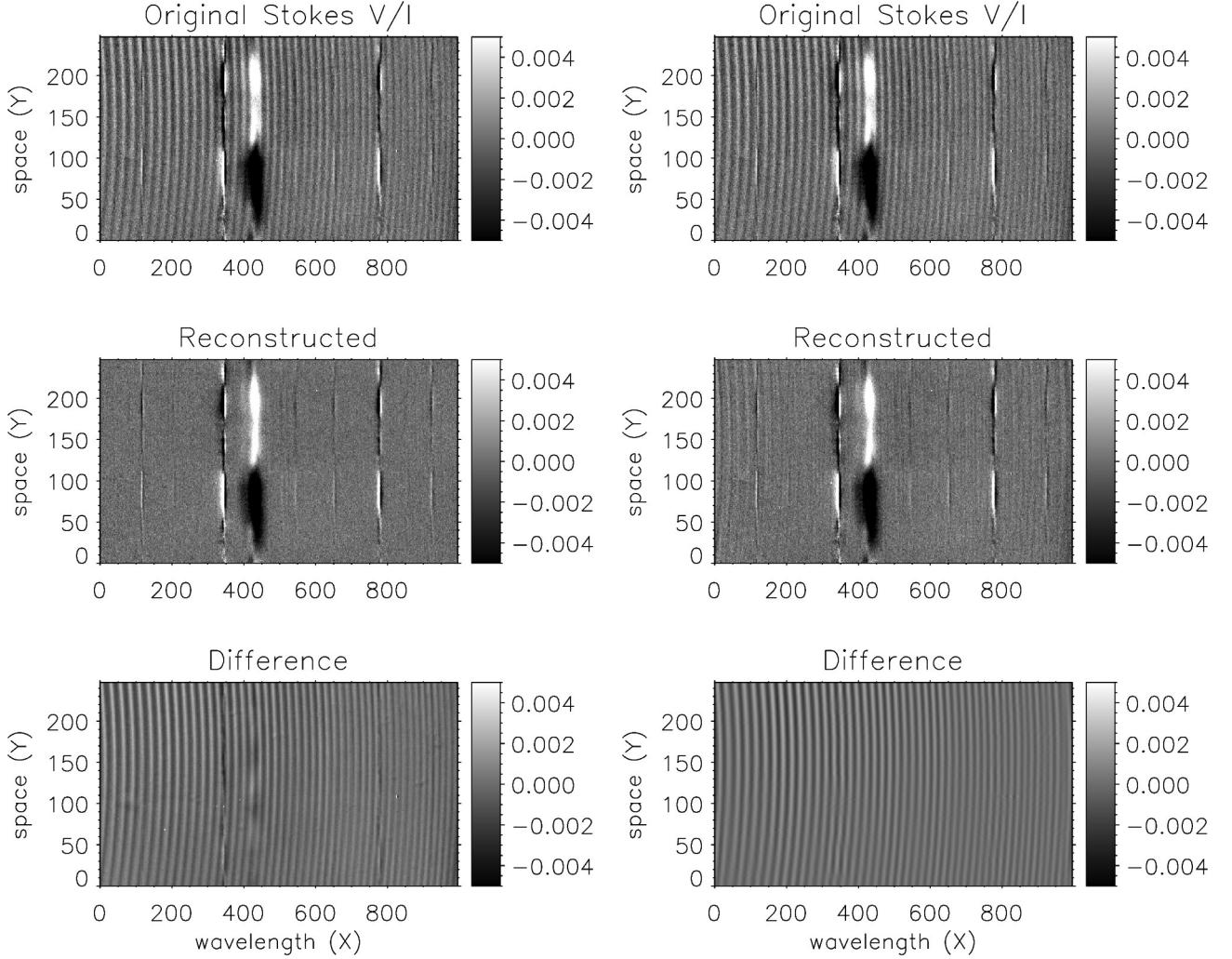


Figure 4. Reconstruction of the same map step as in Figure 3, but using a basis with 32 vectors. *Left:* dropping $\epsilon_{32}(x, y)$ from the reconstruction (similarly to the right side of Figure 3). *Right:* including $\epsilon_{32}(x, y)$ in the reconstruction, but after running a Fourier filtering of that basis vector, in order to remove the confined fringes after the basis transformation.

The above examples were restricted to a small set of Stokes map's frames for practical reasons, and it can be expected that the reduced statistical significance of such a set importantly impacts the realization of the uncorrelation condition (8). In typical science applications, instead, the number of frames is much larger (of order 10^2), and this should favor the realizability of the condition (8), *if the Stokes signals of the observed target are sufficiently varied across the map*. This is well demonstrated by Figure 4 (left), which shows how the quality of fringe removal in the previous example improves when the PCA basis is determined using 32 frames in the map, instead of the 12 used for Figure 3.

When the residual spectral signal in the set of transformed basis vectors containing the fringes is important, one can attempt to recover such signal rather than drop it from the reconstruction together with the unwanted fringes. A direct way for doing this is by Fourier filtering of the fringe component(s) from the relevant basis vectors. Once the fringes are removed from a basis vector, this can be added back to the basis set used for the data reconstruction, rather than being eliminated as done in the previous examples. Figure 4 (right) shows an example of such manipulation of basis

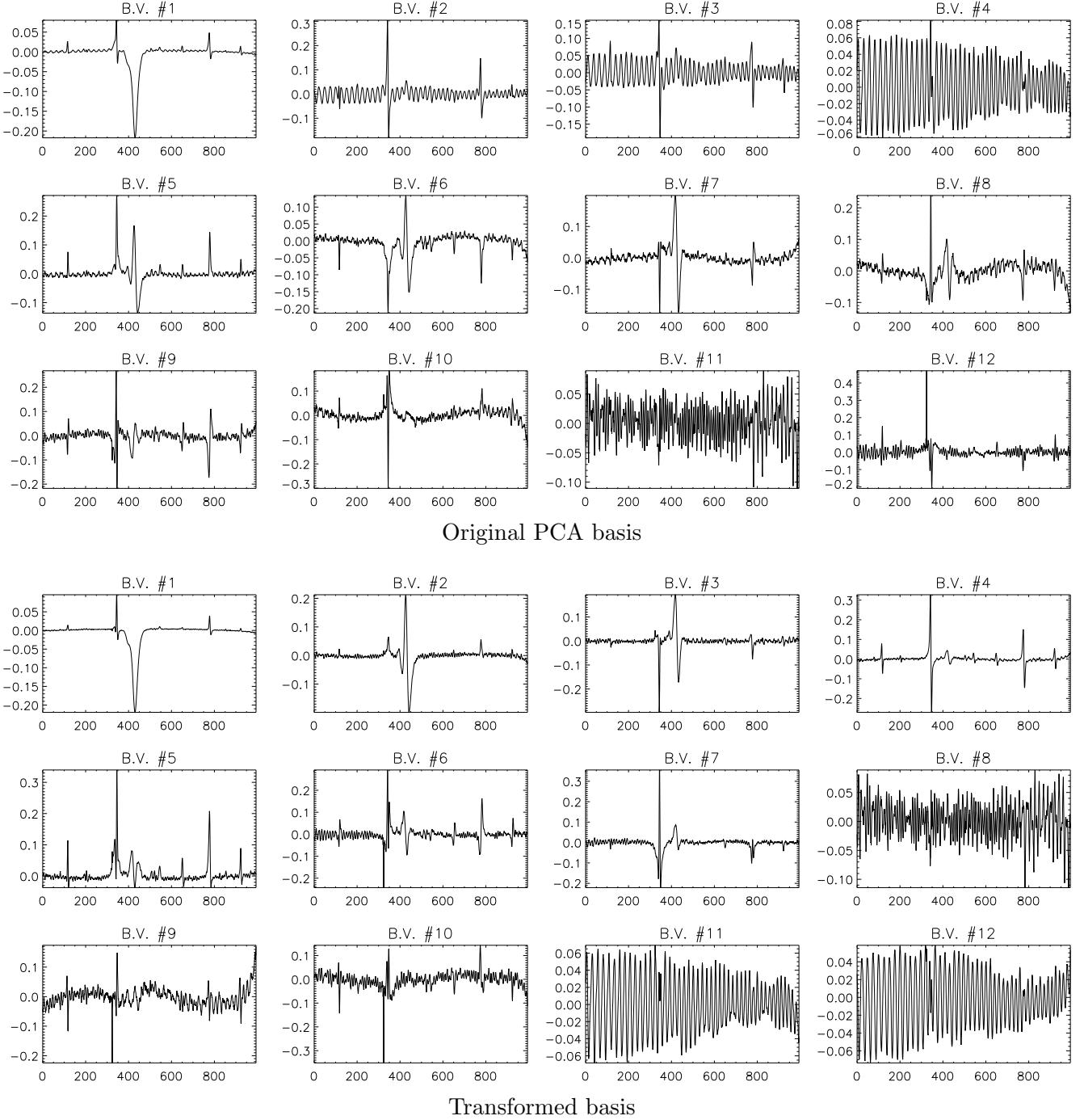


Figure 5. *Top:* the Y-PCA basis $\{e_j(x)\}_{j=1,\dots,12}$ for the same Stokes V map with $N = 32$ frames used for Figure 4. Note the evident presence of fringes in the eigenprofiles 2 to 4. The same fringes affect to a minor degree multiple others elements of the basis, including $e_1(x)$. *Bottom:* the transformed basis $\{\epsilon_k(x)\}_{k=1,\dots,12}$. Note that the fringes have been very well confined to the subspace spanned by the last two basis vectors, and significantly suppressed elsewhere.

elements by Fourier filtering. In the rotated basis of 32 elements, the last two showed visible fringes, particularly $\epsilon_{32}(x, y)$, with non negligible contribution from the targeted spectral signal. Applying

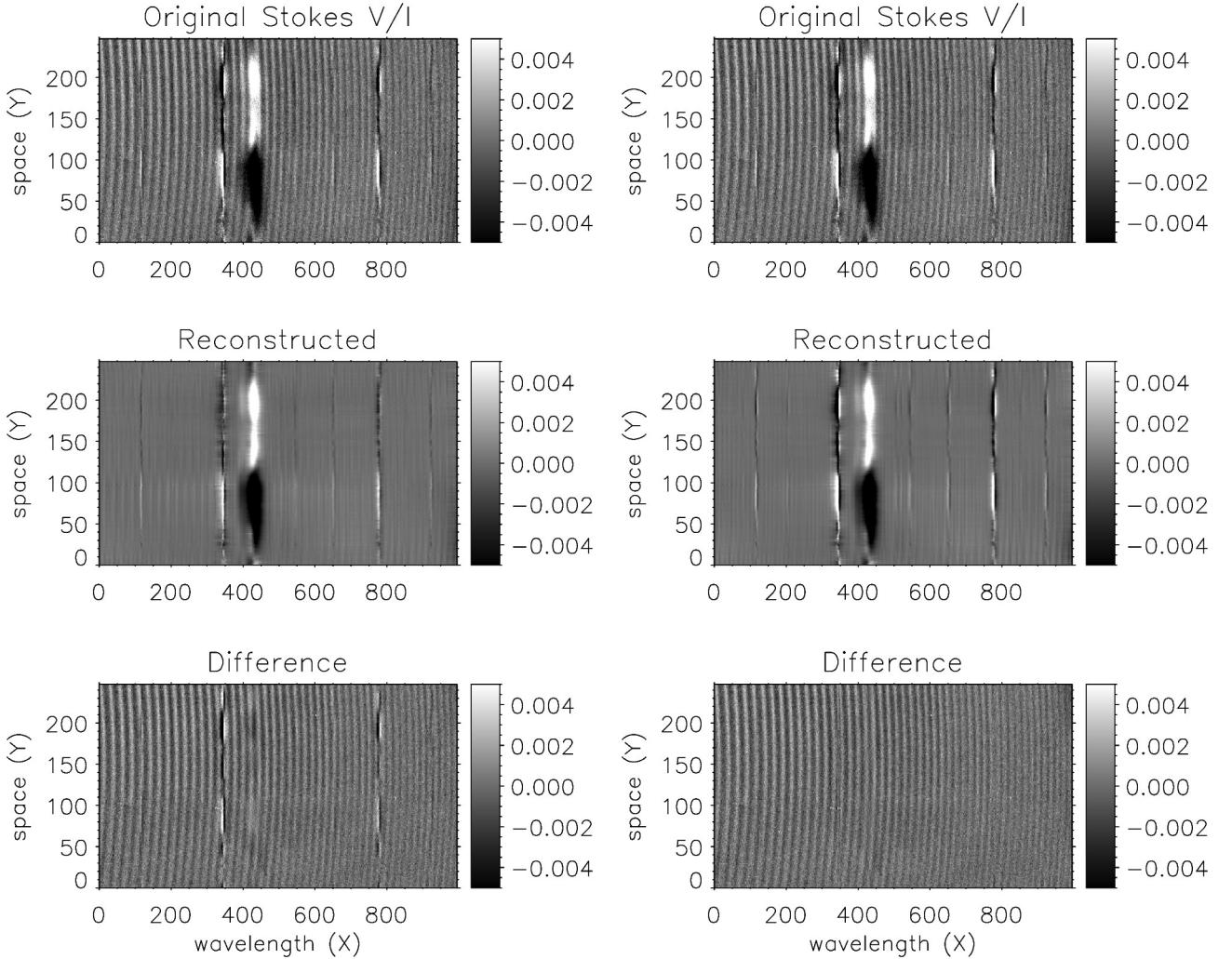


Figure 6. Original (top), Y-PCA reconstructed (center), and residual of the reconstruction (bottom) for a specific frame of the data set used for the creation of the bases of Figure 5. *Left:* using the original Y-PCA basis without $e_2(x)$, $e_3(x)$, and $e_4(x)$. *Right:* using the rotated basis without $\epsilon_{11}(x)$ and $\epsilon_{12}(x)$. We note the significantly reduced loss of spectral signal with the rotated basis (especially for the two Si I lines), despite a comparable suppression of the dominant fringes.

Fourier filtering to the last two basis vectors, that signal can be added back to the reconstruction, producing a residual without practically any science data loss (see right panels of Figure 4).

It is important to observe that the suppression of any signal from a basis vector, e.g., by Fourier filtering, breaks the orthonormality of the basis, which therefore can no longer be used as a projection vector set for the data. However, it is possible to reconstruct the data by using the same coefficient matrix \mathbf{C} for the transformed basis $\{\epsilon_i\}_{i=1,\dots,N}$ prior filtering. This is mathematically equivalent to performing the data reconstruction by excluding the basis vectors affected by the fringes, and then apply the Fourier filtering to the difference image, in order to retrieve the residual target signal from it. The downside of the second approach is that the residual must in principle be Fourier filtered independently for each map step (although it is easy to devise an automated way to do so for an entire map), whereas by directly Fourier filtering the basis set, the residual correction is automatically applied to the entire Stokes map.

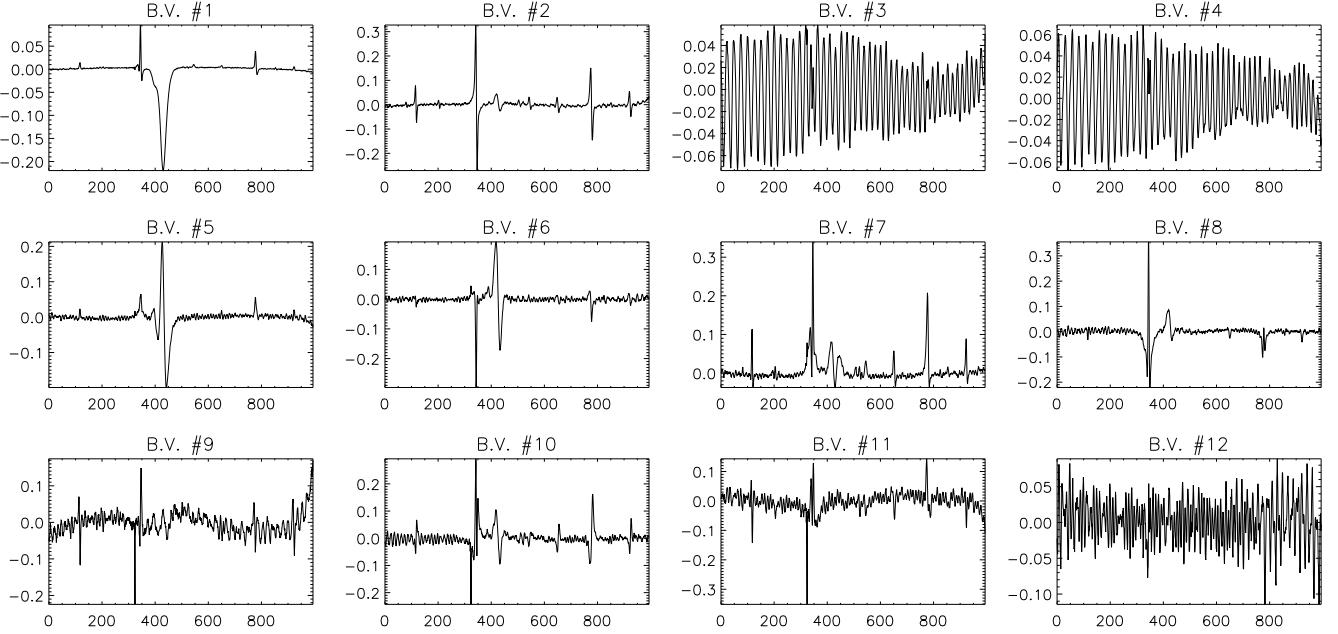


Figure 7. The transformed basis of Figure 5 after reordering of the basis vectors according to the weight of their relative contributions to the map data.

In Sect. 3.2, we compare this *selective* Fourier filtering of fringe PCs or difference images with the more traditional application of Fourier filtering directly to a frame of the Stokes map.

3.1. Application to the Y-PCA method

The technique of orthogonal transformations of the PCA basis, in order to confine the polarization fringes components to an orthogonal subspace of the spectral signal, can also be employed to improve the performance of the Y-PCA method considered by Casini, Judge, & Schad (2012).

In order to test its performance, we employed the same Stokes V map with 32 scan steps used for the results of Figure 4. The Y-PCA algorithm implies the creation of the data correlation matrix (cf. eq. (1) of Casini, Judge, & Schad 2012)

$$\text{Corr}(\mathcal{D}, \mathcal{D}) = \frac{1}{N} \sum_{i=1}^N \mathcal{D}_i(x, y)^T \mathcal{D}_i(x, y) . \quad (13)$$

The above definition implies contraction over the spatial dimension Y , and leads to a $N_X \times N_X$ correlation matrix, with N_X eigenprofiles in the resulting PCA basis.

In the case of the Stokes maps considered in the previous examples, $N_X = 995$, but in Figure 5 we only consider the first 12 basis vectors. The fringes are easily recognizable, and in order to suppress them by using the decomposition of the Stokes map on the original PCA basis (top half of Figure 5), one must drop at least $\mathbf{e}_2(x)$, $\mathbf{e}_3(x)$, and $\mathbf{e}_4(x)$ from the data reconstruction. The result of this operation is shown on the left of Figure 6. Qualitatively, it fares comparably to the TP-PCA reconstruction using the full basis of 32 vectors after rotation (left side of Figure 4)—slightly better for He I 1083 nm, but visibly worse for the two Si I lines. However, using the transformed basis (bottom half of Figure 5), where the fringes are very well confined to the last two basis vectors $\mathbf{e}_{11}(x)$ and $\mathbf{e}_{12}(x)$, the suppression of the fringes in the reconstructed data by dropping those two basis

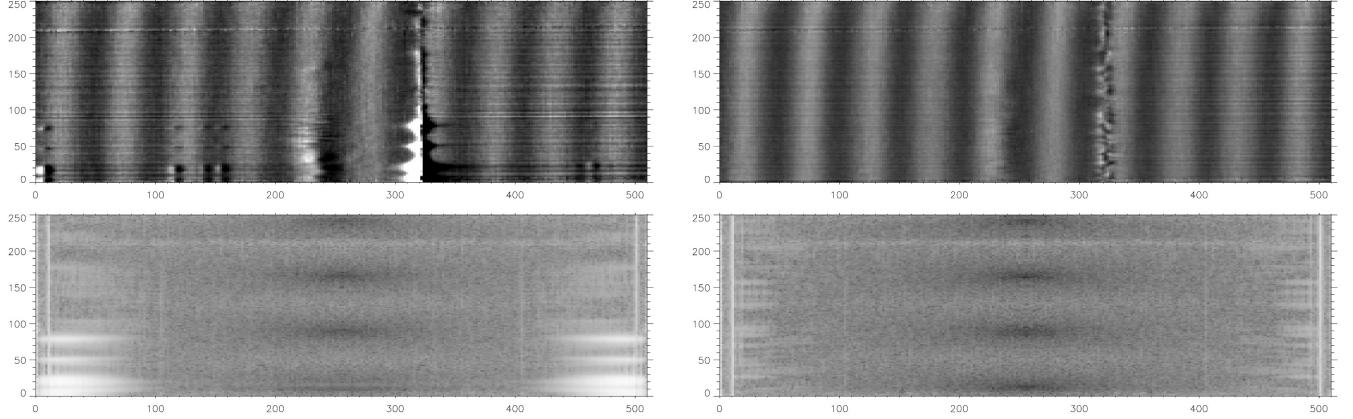


Figure 8. *Left:* A Stokes map frame (top), and its corresponding spectral autocorrelation image (bottom). The frequency range of the dominant fringes is clearly visible as two vertical strips in Fourier space, at around $X = 10$ and $X = 400$, stretching along the entire spatial dimension. Note the significant amplitude overlap of these fringes with the spectral signals in several spatial regions below $Y = 100$. *Right:* The residual of the same frame after PCA reconstruction by excluding the fringe affected basis vectors (top), and its corresponding spectral autocorrelation image (bottom). We note the much improved isolation of the fringe contribution in Fourier space, due to the overall suppression of the spectral signal, which allows for an efficient filtering of the fringes in the residual image.

vectors improves significantly, becoming qualitatively comparable to the TP-PCA reconstruction after Fourier filtering of the fringe-carrying basis vectors (cf. the right sides of Figs. 4 and 6). Incidentally, we note the complete absence of random “noise” in the reconstructed images of Figure 6, because of the low-pass cut implied by considering only the lowest-order 12 basis vectors out of the full set of $N_X = 995$.

This test confirms the results of Casini, Judge, & Schad (2012) about the applicability of the Y-PCA method to the Stokes de-fringing problem. However, depending on the particular data set, the state of polarization, and the instrument properties, having access to different methods for analyzing the data can help optimize the de-fringing process.

Typically, compared to Y-PCA, the TP-PCA method manages to confine the fringe subspace into a smaller number of basis vectors (i.e., \mathcal{F} has a lower dimensionality), which are therefore easier to identify and manage for the data reconstruction, whether it is a matter of excluding them from the transformed basis, or of applying Fourier filtering to them in order to retrieve some residual spectral signal. One downside of both methods, when a basis rotation is applied, is the fact that the basis transformation scrambles the *importance ordering* of the original PCA basis. For the sake of reducing data loss, one always tries to minimize the number of basis vectors of \mathcal{F} that must be further handled, whether by simply dropping them from the data reconstruction or by Fourier filtering. Therefore, lack of information on the relative importance of the contributions of the transformed basis vectors to the data can be a disadvantage, especially for the Y-PCA method, where the dimensionality of \mathcal{F} is typically larger.

A workaround that allows to restore the proper importance ordering of the transformed basis is to use the expression for the eigenvalues λ_i of the correlation matrix of the map data (cf. eq. (A3)),

$$\lambda_i \equiv \lambda(\mathbf{U}_i) = \mathbf{U}_i^T \text{Corr}(\mathcal{D}, \mathcal{D}) \mathbf{U}_i \quad (14)$$

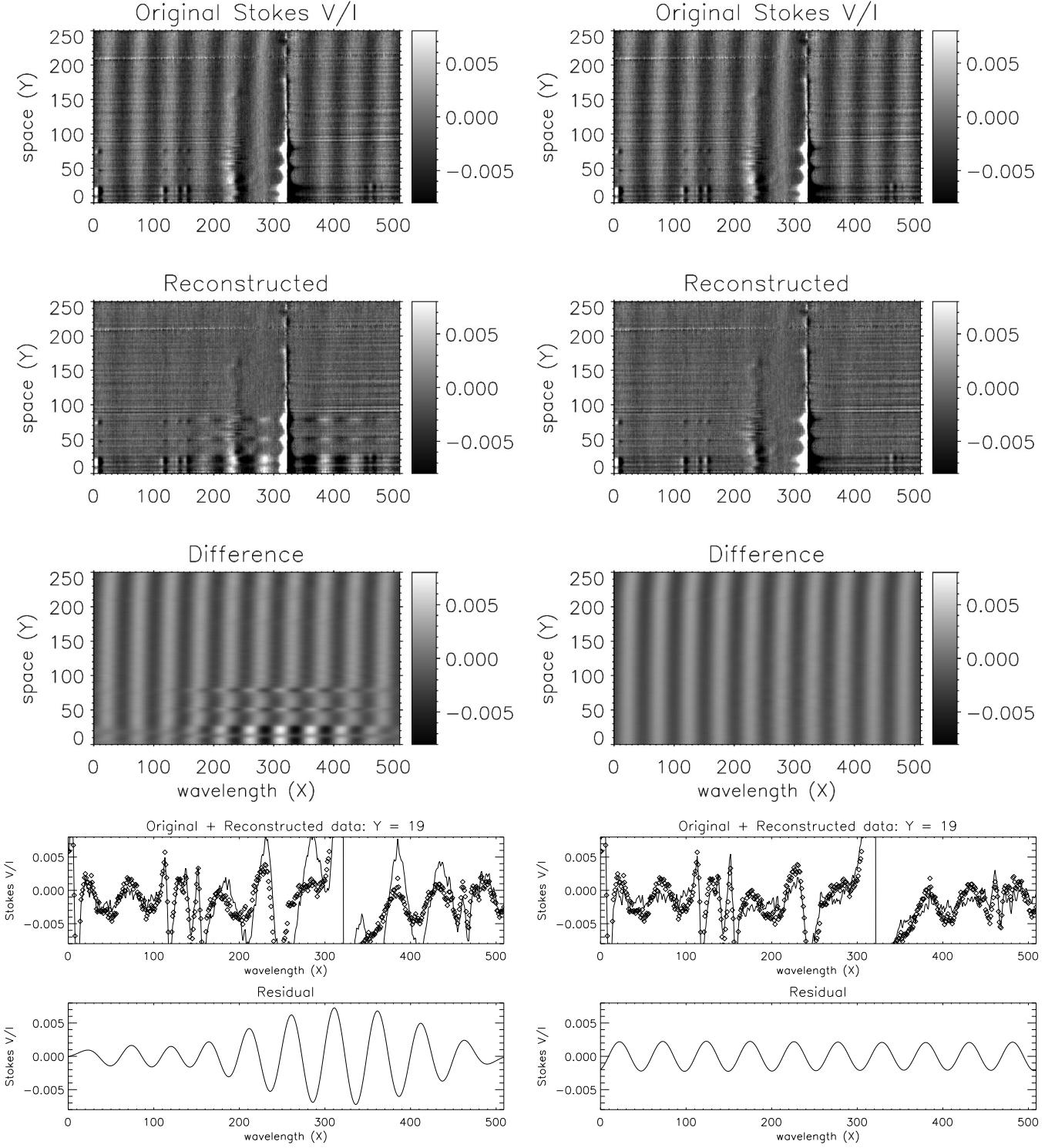


Figure 9. Original (1st row), reconstructed (2nd row), and residual of the reconstruction (3rd row) for one frame of a Stokes V map around the He I 1083 nm line. The map (frequency ordered) was taken with the DST/SPINOR instrument, and contains a total of 100 frames. *Left:* Stokes signal reconstructed after a direct Fourier filtering of the frame. *Right:* Stokes signal reconstructed after a selective Fourier filtering of the contribution from the fringe basis vectors, after a TP-PCA of the full map, and successive transformation of the PCA basis. The bottom two rows show the profiles at $Y = 19$ in the corresponding reconstructed frames. *Top row:* original data (diamonds) and the reconstructed profile (continuous line). *Bottom row:* filtered fringes. We note the significant spurious signal introduced by the direct Fourier filtering of the frame shown on the left.

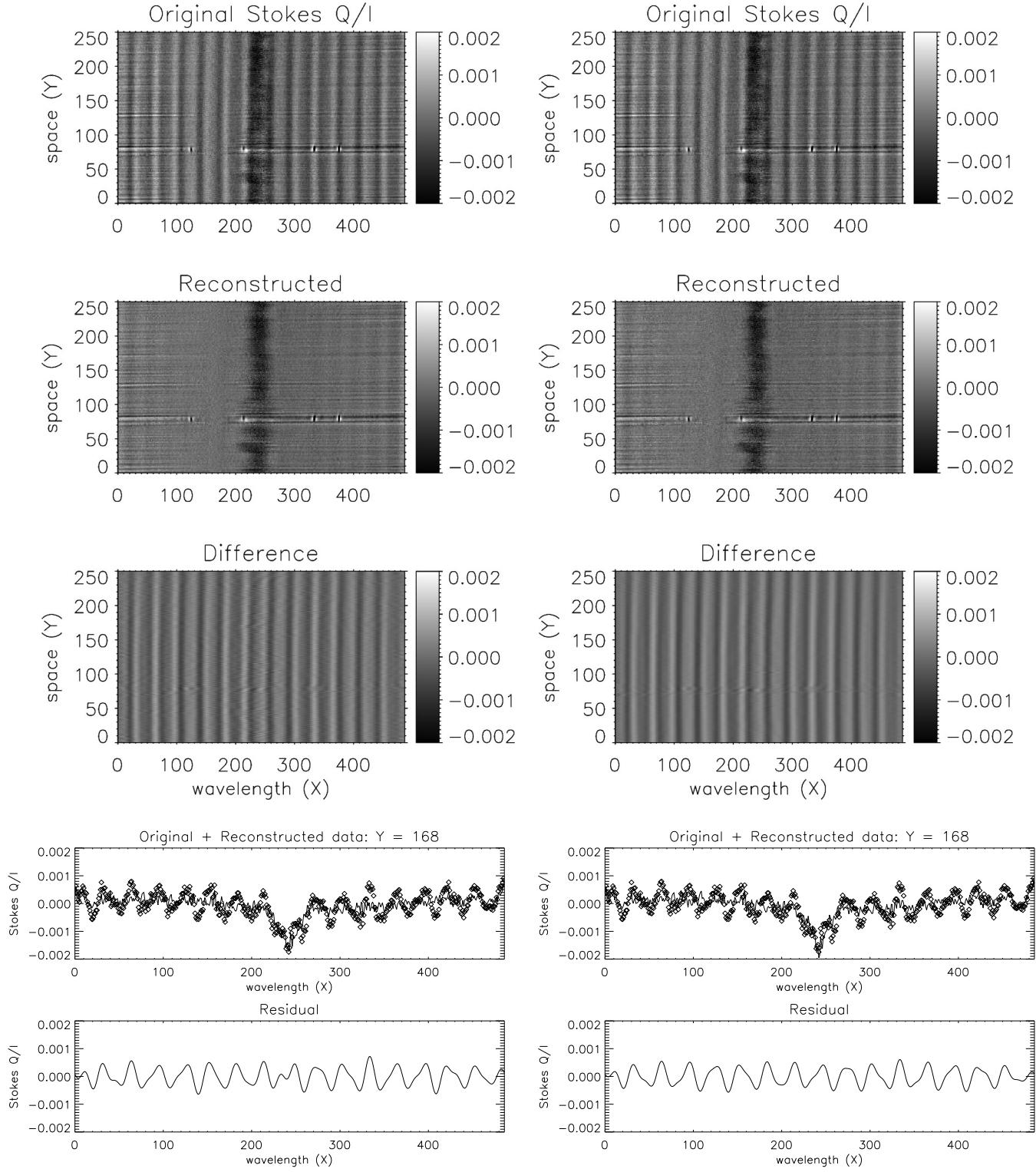


Figure 10. Same as Figure 9 for one frame of a Stokes Q map of the H I H α line at 656 nm. The map (frequency ordered) was taken with the DST/SPINOR instrument, and contains a total of 60 frames. For this data set, direct Fourier filtering (left) produces a very good result, with only minor differences with respect to selective Fourier filtering of the contribution from the dominant fringe basis vectors (in this case e_1 , e_2 , and e_6) shown at the right.

where \mathbf{U}_i is the corresponding (column) eigenvector. Then, if $\tilde{\mathbf{U}}_i = \mathbf{R}\mathbf{U}_i$ is the transformed eigenvector through the same basis transformation of eq. (9), one can use $\lambda(\tilde{\mathbf{U}}_i)$ as an estimate of the weight of the transformed basis vector ϵ_i to the map data. Figure 7 shows the transformed basis of Figure 5 after such reordering.

3.2. Comparison with direct Fourier filtering

Direct Fourier filtering of Stokes map frames is a popular method for tackling the problem of polarization fringe removal, because of its straightforward implementation. One must simply identify the frequency range of the fringe pattern to be removed in the spectral Fourier transform of the map frame, and suppress those frequencies before restoring the map frame by inverse Fourier transform. Fringes that are well separated in frequency space from the targeted spectral signal can easily and effectively be removed by this method. However, in general, some frequency overlap between fringes and signals is to be expected, which inevitably leads to some science data loss.

Figure 8 illustrates a typical example. The top panel on the left shows a frame from a Stokes V map of the He I 1083 nm line region, with a total of 100 steps, whereas the panel right below it shows the corresponding spectral autocorrelation image. The instrument in this case was the Spectro-Polarimeter for INfrared and Optical Regions (SPINOR; Socas-Navarro et al. 2006), also deployed at the NSO/SP DST (note the spectrum is ordered by frequency rather than wavelength). The frequency interval of the fringes is well captured in Fourier space, and it is identified by the two narrow vertical strips located around $X = 10$ and $X = 400$, spanning the full spatial dimension. However, the map frame also contains spectral signal with a frequency range that significantly overlaps with the amplitude of the fringes in Fourier space, predominantly for positions along the spectrograph slit below $Y = 100$. At those slit positions, we can expect that the reconstruction of the signal will be impacted by Fourier filtering. The panels on the right show instead the residual image of a TP-PCA reconstruction, after transformation of the PCA basis, and the exclusion of the last two fringe-carrying basis vectors. We note that the amplitude of the residual spectral signal in Fourier space has been significantly reduced, and that the separation between fringes and spectral signal is much improved, allowing for an efficient filtering of the fringes in the residual image.

Figure 9 demonstrates in practice the effects of direct Fourier filtering in such case. The left side of the figure shows the reconstruction of the same map step in the example of Figure 8 after direct Fourier filtering of the frame. As anticipated, the spatial regions where the spectral signal has significant overlap with the fringes in frequency space suffer the most in the direct application of Fourier filtering. The right side of the figure shows instead how the prior PCA decomposition of the Stokes map (in this case, TP-PCA) allows to attain much better results with Fourier filtering, either by selectively applying the filtering to just the (few) basis vectors visibly affected by fringes, or directly on the difference image after excluding those basis vectors from the data reconstruction (see discussion right before Sect. 3.1). In this example, the full PCA basis of the map with 100 eigenfeatures was priorly transformed in order to confine the fringe contribution to just two basis vectors. The data reconstruction was performed using the complementary set of basis vectors, and then the difference image (corresponding to the top-right panel of Figure 8) was Fourier filtered in order to add back the residual signal.

Figure 10 provides an illuminating counterexample. In this case, we reconstructed a Stokes Q frame from a map of the H I $H\alpha$ line region around 656 nm, with a total of 60 steps, also observed with the SPINOR instrument. Despite the complex fringe pattern in this map, direct Fourier filtering of

the full frame (left) already gives very good results, comparable to what can be obtained with the selective Fourier filtering of the difference image from the TP-PCA reconstruction (right). In fact, in this case the spectral autocorrelation image of the map frame shows a clear separation between the fringe and signal contributions in frequency space, allowing a clean suppression of the fringes with minimal loss of science data. In this case, evidently, a prior transformation of the PCA basis does not necessarily improve the fringe suppression, and in fact the right side of Figure 10 was obtained by Fourier filtering the difference image after reconstructing the data without the eigenfeatures e_1 , e_2 , and e_6 of the original PCA basis. Despite the effectiveness of direct Fourier filtering of the map frame in this particular example, a prior PCA of the full map, which allows to target a minimal set of PCs for the selective filtering of the fringe component, still leads to a visibly cleaner removal of the fringe signal from the map frame. This clearly demonstrates the benefit of running a PCA of a Stokes map, even in cases where the transformation of the PCA basis in order to maximally confine the fringe signal may not be necessary.

4. CONCLUSIONS

We studied and compared the performance of various methods for identifying and removing polarization fringes from spectro-polarimetric data based on 2-dimensional Principal Component Analysis (2D PCA) of Stokes maps. All examples were taken from data sets acquired with slit-based spectro-polarimeters, but the same methods can indistinctly be applied also to observations taken with wavelength-tunable imaging spectro-polarimeters. We analyzed in depth the performance of the Turk & Pentland (1991) algorithm of 2D PCA, which was only briefly touched upon in a previous work by Casini, Judge, & Schad (2012). We compared its performance with that of the Yang et al. (2004) algorithm, which was instead the focus of that previous work.

In this study, we highlighted the improvements in spectro-polarimetric data de-fringing by PCA that can be expected by subjecting the PCA basis of a Stokes map to a series of 2D rotations among its various eigenfeatures. This transformation can be set up in a quantitative and repeatable way to specifically target the fringe signal in all eigenfeatures, in order to “sweep” it across the basis and “dump” it into a minimal set of transformed basis vectors, which determine the dimensionality of the fringe subspace \mathcal{F} in the data set. When perfect orthogonality between this subspace and the subspace \mathcal{S} of the targeted spectral signal is realized in a particular data set, it must be expected that the basis vectors spanning the subspace \mathcal{F} will be perfectly orthogonal (hence, uncorrelated) to the subspace \mathcal{S} .

In most practical cases, this ideal condition can never be achieved, and the problem of handling, and possibly “rescuing” any residual spectral signal of interest in the fringe basis vectors was also considered here. It is found that *selective* Fourier filtering of the fringe basis vectors can greatly improve the performance of 2D PCA de-fringing, as well as providing significantly better results than the more traditional *direct* Fourier filtering of each individual frame of the Stokes map.

We describe one possible algorithm for the implementation of an optimal transformation of the PCA basis for the purpose of isolating the fringes into a minimal set of basis vectors (see App. B). Such algorithmic description lends itself to the development of *automated* procedures for the de-fringing of spectro-polarimetric data sets.

Throughout this study, we elected to work with the *correlation* matrix of the Stokes data set rather than its *covariance* matrix as typically done in most PCA methods. The reason for such choice is that the creation of the covariance matrix of the data implies the subtraction of the average image

from the data. In the traditional application of 2D PCA to face recognition, this average must then be added back for data reconstruction, but in the problem of Stokes data de-fringing this would inevitably imply adding back fringes to the reconstructed data. On the other hand, if the average image is left out, and this contains some significant residual spectral signal of interest, this would be inevitably lost in the reconstructed data.

The use of the covariance matrix (which was considered by Casini, Judge, & Schad 2012) is thus only justified when the average image is devoid of any significant residual of the targeted spectral signal. In those cases, sometimes good data de-fringing can already be attained simply by subtracting the average image from the full data. In case of evolving fringe patterns during the Stokes map scan, the (covariance based) PCA method for de-fringing can complete the process, delivering spectro-polarimetric data sets fully usable for science.

APPENDIX

A. LINEAR ALGEBRA OF THE TP-PCA METHOD

We present the basic linear algebra of the Turk & Pentland (1991) approach to 2D PCA (hereafter, TP-PCA), which is needed to implement the de-fringing method for spectro-polarimetric data presented in this work.

As described in Sect. 2, we consider a Stokes map \mathcal{D} with N steps, $\{\mathcal{D}_i(x, y)\}_{i=1,\dots,N}$, where each map step has N_X wavelength points and N_Y spatial points. For algebraic convenience, we reform each $N_X \times N_Y$ image $\mathcal{D}_i(x, y)$ into a $N_X N_Y$ column vector \mathbf{D}_i , so that the full Stokes map can be represented by the $(N_X N_Y) \times N$ matrix \mathbf{D} with column vectors \mathbf{D}_i .

In the TP-PCA approach, the $N \times N$ correlation matrix of the Stokes map is computed as

$$\text{Corr}(\mathcal{D}, \mathcal{D}) = \mathbf{D}^T \mathbf{D} , \quad (\text{A1})$$

so its (i, j) element is simply the inner product

$$\langle \mathbf{D}_i, \mathbf{D}_j \rangle \equiv \mathbf{D}_i^T \mathbf{D}_j . \quad (\text{A2})$$

2D PCA involves the diagonalization of (A1), which is typically accomplished via the method of Singular Value Decomposition. In our case, this produces an orthogonal matrix \mathbf{U} of column eigenvectors \mathbf{U}_i with unit norm, and a corresponding, semi-positive definite, diagonal matrix Λ of eigenvalues λ_i , with ordered diagonal elements (i.e., $\lambda_i \geq \lambda_j$ if $i > j$), such that

$$\mathbf{U}^T \text{Corr}(\mathcal{D}, \mathcal{D}) \mathbf{U} = \Lambda . \quad (\text{A3})$$

We can project the data set over the basis of eigenvectors,

$$\mathbf{W} = \mathbf{D} \mathbf{U} , \quad (\text{A4})$$

and notice that the N corresponding column vectors \mathbf{W}_i form an orthogonal set, since $\mathbf{W}^T \mathbf{W} = \mathbf{U}^T \mathbf{D}^T \mathbf{D} \mathbf{U} = \mathbf{U}^T \text{Corr}(\mathcal{D}, \mathcal{D}) \mathbf{U} = \Lambda$. Evidently, this set is not *orthonormal*. However, normalization is immediately attained by letting

$$\mathbf{E} = \mathbf{W} \Lambda^{-1/2} = \mathbf{D} \mathbf{U} \Lambda^{-1/2} . \quad (\text{A5})$$

The N column vectors \mathbf{E}_i of length $N_X N_Y$ can be reformed into N eigenfeatures $\mathbf{e}_i(x, y)$ of dimensions $N_X \times N_Y$. These eigenfeatures are orthonormal according to the inner product (2), and therefore form a basis for the representation of the data set $\{\mathcal{D}_i(x, y)\}_{i=1,\dots,N}$. In fact, by letting

$$\mathbf{D} = \mathbf{E} \mathbf{C}, \quad (\text{A6})$$

from eq. (A5) we see that the $N \times N$ matrix \mathbf{C} of expansion coefficients is uniquely determined as

$$\mathbf{C} = \mathbf{E}^T \mathbf{D} = \Lambda^{1/2} \mathbf{U}^T. \quad (\text{A7})$$

B. THE OPTIMAL ROTATION ALGORITHM

We present a procedure to determine the set of optimal 2D rotations of the PCA basis necessary to confine polarization fringes to a minimal set of basis vectors. The algorithm is described for the TP-PCA method, but can easily be adapted for the Y-PCA method.

Let $\{\mathcal{D}_i(x, y)\}_{i=1,\dots,N}$ be a set of spectro-polarimetric images (a Stokes map), each of size $N_X \times N_Y$. We indicate with $\bar{\mathcal{D}}(x, y)$ the average image of the set. Typically, the fringes are a persistent feature of the data set, while the spectro-polarimetric signal tends to average out to some degree, especially in the case of a spatially complex target. Therefore, the average image $\bar{\mathcal{D}}(x, y)$ tends to show a weaker signal from the target with respect to the fringes. In the ideal case in which the target signal averages out completely across the Stokes map, the simple subtraction of $\bar{\mathcal{D}}(x, y)$ from the data set can be sufficient to effectively remove the fringes from each spectro-polarimetric image of the map. Otherwise, the PCA basis must be transformed in order to minimize the number of basis vectors that span the subspace \mathcal{F} . This can be accomplished with the following procedure:

1. Select a subregion Σ of the average image $\bar{\mathcal{D}}(x, y)$ that is as clear as possible of the target signal. This subregion will be the testing area on which the merit value for the fringe removal by rotation is going to be calculated.
2. Compute the spectral auto-correlation image of the subregion Σ . This image will show clearly identifiable, dominant bands corresponding to the range of spatial frequencies spanned by the fringes in the subregion. The 2D integral of the auto-correlation image over this interval of frequencies Δk and the spatial dimension of the subregion Σ can be used as the merit function of the basis rotation algorithm.
3. For each eigenfeature $\mathbf{e}_i(x, y)$, with $i = 1, \dots, N - 1$, perform all the possible 2D rotations

$$\begin{aligned}\mathbf{e}_i(x, y) &= \cos \theta \mathbf{e}_i(x, y) + \sin \theta \mathbf{e}_j(x, y), \\ \mathbf{e}_j(x, y) &= -\sin \theta \mathbf{e}_i(x, y) + \cos \theta \mathbf{e}_j(x, y),\end{aligned}$$

for $j = i + 1, \dots, N$ and $\theta \in [-\frac{\pi}{2}, +\frac{\pi}{2}]$. Compute the merit function over the subregion Σ of $\mathbf{e}_i(x, y)$, and select the pair $(j, \theta) = (j_M, \theta_M)$ that produces the minimum value.

4. Perform the substitutions $\mathbf{e}_i(x, y) \rightarrow \mathbf{e}_i(x, y)$ and $\mathbf{e}_{j_M}(x, y) \rightarrow \mathbf{e}_{j_M}(x, y)$ in the PCA basis, and repeat steps #3 and #4 through the end of the loop over i .

The same algorithm can be adapted to the rotation of the 1D bases produced in the Y-PCA method. In that case, one must calculate the spectral auto-correlation function of the transformed basis vector $\epsilon_i(x)$, and integrate it over the frequency interval Δk of the fringes in order to compute the merit function of the transformation.

At the end of the transformation procedure, the polarization fringes will have been optimally removed from the basis vectors ϵ_i of lower order, and confined to those of higher order.

However, it is important to clarify the meaning of “optimally” transformed basis in this context. According to the proposed algorithm, each eigenfeature e_i is transformed by combining with only one other eigenfeature e_j of the PCA basis, where $j > i$. The algorithm determines the optimal pairing of eigenfeatures, but generally none of those available in the PCA basis will lead to a perfect cancellation of the unwanted signal in the transformed basis vector ϵ_i . However, when the rotation procedure is completed, the transformed basis $\{\epsilon_i\}_{i=1,\dots,N}$ will contain a new set of basis vector, which may be “more optimal” than the original PCA basis for the purpose of canceling the (residual) unwanted signal from each eigenfeature.² In other words, it is possible to improve the realization of the orthogonality between \mathcal{F} and \mathcal{S} by iterating the rotation algorithm over the transformed basis $\{\epsilon_i\}_{i=1,\dots,N}$.

REFERENCES

- Birkhoff, G., & Mac Lane, S. 1953, A Survey of Modern Algebra (MacMillan: New York)
- Cao, W., Gorceix, N., Coulter, R., Ahn, K., Rimmele, T. R., & Goode, P. R. 2010, AN, 331, 636
- Casini, R., & Landi Degl’Innocenti, E. 2008, Astrophysical Plasmas, in Plasma Polarization Spectroscopy, eds. T. Fujimoto & A. Iwamae (Springer: Berlin), 247
- Casini, R., Judge, P. G., & Schad, T. A. 2012, ApJ, 756, 193
- Cavallini, F. 2006, SoPh, 236, 415
- Clarke, D. 2004, J. Opt. A, 6, 1036
- Collados, M., Bettonvil, F., Cavaller, L., Ermolli, I., Gelly, B., Pérez, A., Socas-Navarro, H., Soltau, D., Volkmer, R., & EST Team 2013, Mem. SAIt, 84, 379
- Harrington, D. M., Snik, F., Keller, C. U., Sueoka, S. R., & van Harten, G. 2010, JATIS, 3, 048001 (DOI:10.1117/1.JATIS.3.4.048001)
- Hasan, S. S., Soltau, D., Kärcher, H., Süß, M., & Berkefeld, T. 2010, AN, 331, 628
- Jaeggli, S. A., Lin, H., Mickey, D. L., Kuhn, J. R., Hegwer, S. L., Rimmele, T. R., and Penn, M. J. 2010, Mem. SAIt, 81, 763
- Jolliffe, I. T. 2002, Principal Component Analysis, 2nd ed. (Springer: New York)
- Landi Degl’Innocenti, E., & Landolfi, M. 2004, Polarization in Spectral Lines (Springer: Dordrecht)
- Lites, B. W. 1991, Fringes in Polarizing Optical Elements, in Solar Polarimetry, Proc. 11th Sacramento Peak Summer Workshop, ed. L. J. November (Sunspot: NSO), 173
- Liu, Z., Deng, Y., & Ji, H. 2014, The Chinese Giant Solar Telescope, in Nature of Prominences and their Role in Space Weather, eds. B. Schmieder, J.-M. Malherbe, & S. T. Wu, Proc. IAU, 300, 349
- Pearson, K. 1901, Phil. Mag., 2, 559
- Rojo, P. M., & Harrington, J. 2006, ApJ, 649, 553
- Semel, M. 2003, A&A, 401, 1
- Skumanich, A. P., & López Ariste, A. 2002, ApJ, 570, 379
- Socas-Navarro, H., Elmore, D., Pietarila, A., Darnell, A., Lites, B. W., Tomczyk, S., & Hegwer, S. 2006, SoPh, 235, 55
- Stenflo, J. O. 1994, Solar Magnetic Fields (Kluwer: Dordrecht)
- Tritschler, A., Rimmele, T. R., Berukoff, S., Casini, R., Kuhn, J. R., Lin, H., Rast, M. P., Matthäi, T., Hildner, P., Schmidt, K., Wignatti, F., Mögler, F., and the DKIST Team 2016, AN, 337, 1064

² By construction, this is indeed the case, since the algorithm Matthäi, T., Hildner, P., Schmidt, K., Wignatti, F., Mögler, F., and the DKIST Team produces a minimal set of transformed basis vectors.

- Trujillo Bueno, J. 2010, Recent Advances in Chromospheric and Coronal Polarization Diagnostics, in Magnetic Coupling between the Interior and Atmosphere of the Sun, Proc. Astrophysics and Space Science, eds. S. S. Hasan & R. J. Rutten (Springer: Berlin), 118
- Turk, M. A., & Pentland, A. P. 1991, JoCN, 3, 71
- Yang, J., Zhang, D., Frangi, A. F., & Yang, J.-Y. 2004, IEEE Trans. PAMI, 26, 131