



## Joint filtering-estimation of Stokes vector images based on a non-local means approach

Sylvain Faisan, Christian Heinrich, François Rousseau, Alex Lallement, Jihad Zallat

### ► To cite this version:

Sylvain Faisan, Christian Heinrich, François Rousseau, Alex Lallement, Jihad Zallat. Joint filtering-estimation of Stokes vector images based on a non-local means approach. Journal of the Optical Society of America. A Optics, Image Science, and Vision, 2012, 29 (9), pp.2028-2037. 10.1364/JOSAA.29.002028 . hal-00727659

**HAL Id: hal-00727659**

**<https://hal.science/hal-00727659>**

Submitted on 4 Sep 2012

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Joint filtering-estimation of Stokes vector images based on a non-local means approach

S. Faisan<sup>1,\*</sup>, C. Heinrich<sup>1</sup>, F. Rousseau<sup>1</sup>, A. Lallement<sup>1</sup>, and J. Zallat<sup>1</sup>

<sup>1</sup>*LSIIT - UMR 7005 CNRS, University of Strasbourg  
Pôle API, Boulevard Sébastien Brant, B.P. 10413, 67412 Illkirch Cedex, France*

*\*Corresponding author: faisan@unistra.fr*

Conventional estimation techniques of Stokes images from observed radiance images through different polarization filters suffer from noise contamination that hampers correct interpretation or even leads to unphysical estimated signatures. This paper presents an efficient restoration technique based on non-local means, permitting accurate estimation of smoothly variable polarization signatures in the Stokes image while preserving sharp transitions. The method is assessed on simulated data as well as on real images. © 2012 Optical Society of America

*OCIS codes:* 100.3020, 100.3190, 120.5410.

## 1. Introduction

Polarization imaging enhances classical intensity imaging techniques by providing access to extra parameters that describe the vector nature of light. This is achieved by placing controlled polarization filters prior to the camera. This has been widely used for many applications ranging from medical imaging to surface inspection and remote sensing. In this paper, we focus on the restoration of polarization encoded images from raw radiance measurements.

The Stokes-Mueller formalism is a coherent framework based on a linear observation model that provides access to polarization parameters from raw radiance acquisitions. In polarization imaging, this observation model is extended to the entire image where classical pixel-wise data reduction approaches are generally used [1,2]. Usually, the measurement system is calibrated so as to reduce the impact of noise and to reduce system errors. However, even for a well calibrated imaging polarimeter, errors and noise may be amplified and transferred from raw acquisitions to the estimated polarization parameter images. This reduces also the

accuracy of the estimated quantities (Stokes vectors) that may violate physical admissibility constraints. A general and robust approach is thus needed to correctly restore the polarization signatures while preserving their integrity. Such algorithms can be used as a valuable tool for analyzing the physical contents of observed scenes in the context of polarization imaging.

Bayesian approaches providing robust estimates of polarization channels with regard to image noise have been introduced. The Bayesian framework yields neat solutions to the polarimetric data reduction problem for the case of piecewise constant signatures [3, 4]. The case of smoothly varying signatures has also been addressed, for example by Valenzuela *et al.* [5] and by Sfikas *et al.* [6]. However, the work of Valenzuela does not account for physical admissibility constraints, and Sfikas' algorithm requires a pre-processing step that may blur the edges.

To circumvent these limitations, we propose here to adapt a very efficient denoising algorithm, the non-local means (NLM) algorithm [7], to the estimation of Stokes vectors. The core idea of the denoising algorithm is to exploit the possible repetitive pattern or texture within the image to improve denoising. Basically, the original image is decomposed into a collection of small sub-images called patches. By considering image patches as image features, the NLM algorithm can be viewed as a regression technique adapted to image denoising: the restored patches are computed as a weighted average of noisy patches. Buades *et al.* have shown that, for 2D natural images, the NLM filter outperforms popular denoising methods such as the Rudin-Osher-Fatemi Total Variation minimization scheme [8] or wavelet-based techniques [9]. Since then, this non-local strategy has been studied and applied to several image processing applications such as non-local regularization functionals in the context of inverse problems [10–13]. The interested reader may consult the recent review by Katkovnik *et al.* [14].

From the NLM algorithm, we derive in this article a joint filtering-estimation procedure suitable for Stokes images: the noise is filtered while estimating physically admissible Stokes vectors at each pixel location. Moreover, we have shown that the proposed joint filtering-estimation procedure can be equivalently achieved sequentially by a two step method: a filtering stage based on the NLM approach, which is widely recognized for its edge-preserving smoothing properties, followed by an estimation step ensuring physical admissibility. The contribution of this article is also methodological, by the introduction of NLM weights different from those found in the literature. This new weight structure allows the NLM algorithm to adapt to the content of the image, thus improving the robustness of the approach and the quality of the results.

The article is organized as follows: section 2 states the problem of estimating Stokes vectors from polarized radiance measurements. The proposed approach is presented in Section 3.

Section 4 deals with the application of the method to simulated and real measurements. Conclusions are drawn in section 5.

## 2. Problem statement

A Stokes imaging polarimeter yields data linked to Stokes vectors, where one  $4 \times 1$  Stokes vector is attached to each pixel in the image. These data correspond to the outgoing radiance of a scene, captured by  $K$  ( $K \geq 4$ ) independent probing states of a Polarization State Analyzer (PSA) placed in front of a camera. At each pixel location  $\mathbf{x}$ , the  $K$  intensity measurements  $\mathbf{I}$  are related to the Stokes vector  $\mathbf{S}$  by the linear equation:

$$\mathbf{I}(\mathbf{x}) \simeq \mathbf{P} \mathbf{S}(\mathbf{x}),$$

where  $\mathbf{P}$  is a matrix of size  $K \times 4$  called Polarization Measurement Matrix (PMM) whose rows constitute the  $K$  probing states. Systematic calibration errors as well as CCD noises are accounted for in the same way as in [2, 3], which leads to the following observation model:

$$\mathbf{I}(\mathbf{x}) = \mathbf{P} \mathbf{S}(\mathbf{x}) + \mathbf{n}(\mathbf{x}),$$

where  $\mathbf{n}$  encompasses all noise terms. The problem is now to estimate  $\mathbf{S}$  from observations  $\mathbf{I}$  while ensuring the physical admissibility

$$S_0 \geq \sqrt{S_1^2 + S_2^2 + S_3^2}$$

of the retrieved Stokes vector attached to each pixel.

The classical approach relying on the PSA pseudo-inverse matrix does not take the admissibility constraints into account. We can remedy this problem by replacing the pseudo-inverse estimate by the following constrained optimization:

$$\hat{\mathbf{S}}(\mathbf{x}) = \arg \min_{\mathbf{S} \in \mathbf{B}} \|\mathbf{I}(\mathbf{x}) - \mathbf{P} \cdot \mathbf{S}(\mathbf{x})\|^2, \quad (1)$$

where  $\mathbf{B}$  is the set of admissible Stokes vectors. Note that the pseudo-inverse approach corresponds to optimizing the criterion of Eq. 1 without considering the admissibility constraints.

## 3. Joint filtering-estimation of Stokes vectors

In order to reduce the effects of noise, we propose to take advantage of spatial information by using non-local filtering. Section 3.A addresses non-local filtering for the canonical one-channel image case. Section 3.B introduces some improvements to the standard method. Section 3.C generalizes the algorithm to handle multi-channel Stokes image restoration.

### 3.A. Non-local means filtering

Let consider the following additive model for one-channel noisy images:

$$I(\mathbf{x}) = f(\mathbf{x}) + n(\mathbf{x}), \quad (2)$$

where  $f$  is the ideal image,  $n$  is a zero mean noise with unknown variance  $\sigma^2$ ,  $I$  is the observed image and  $\mathbf{x}$  is a pixel location. The purpose of a denoising algorithm is to estimate the image  $f$  based on the observed image  $I$ .

Let us consider a weighted graph  $w$  linking the pixels  $\mathbf{x} \in \Omega$  and  $\mathbf{y} \in \Omega$  of the input image  $I$  with a weight  $w(\mathbf{x}, \mathbf{y})$ .  $\Omega$  is the image domain, *i.e.*, the support of the digital image ( $\Omega \subset \mathbb{Z}^2$ ). The weighted graph  $w$  is a representation of non-local similarities within the input image  $I$ . In [7], the non-local graph  $w$  is used for denoising purpose using a neighborhood averaging strategy:

$$\forall \mathbf{x} \in \Omega, \quad I_{nlm}(\mathbf{x}) = \frac{\sum_{\mathbf{y} \in \Omega} w(\mathbf{x}, \mathbf{y}) I(\mathbf{y})}{\sum_{\mathbf{y} \in \Omega} w(\mathbf{x}, \mathbf{y})}, \quad (3)$$

where  $I_{nlm}$  is a denoised version of  $I$ . The weights of the graph are estimated by considering image patches that are local features at each pixel. An image patch  $P_{\mathbf{x}}$  of width  $(2k + 1)$  centered around pixel  $\mathbf{x}$  is defined as follows:

$$P_{\mathbf{x}} = \{I(\mathbf{x} + \boldsymbol{\tau}), \boldsymbol{\tau} \in \llbracket -k, k \rrbracket^2\}, \quad (4)$$

where  $\llbracket -k, k \rrbracket$  denotes the set  $\{-k, -(k-1), \dots, (k-1), k\}$  of integers.

The graph weights  $w(\mathbf{x}, \mathbf{y})$  are computed as [16]:

$$w(\mathbf{x}, \mathbf{y}) = \phi \left( \frac{d_P(\mathbf{x}, \mathbf{y})^2}{2N\beta\hat{\sigma}^2} \right), \quad (5)$$

where  $d_P(\mathbf{x}, \mathbf{y})$  is an intensity-based distance between patches  $P_{\mathbf{x}}$  and  $P_{\mathbf{y}}$  (we use here the Euclidean norm  $\|\cdot\|_2$ );  $N$  is the number of pixels of a patch ( $N = (2k + 1)^2$ );  $\phi$  is a kernel function measuring the similarity between two patches, which was chosen by Buades *et al.* [7] as  $\phi(x) = e^{-x}$ ;  $\beta$  is a smoothing parameter and  $\hat{\sigma}$  is the estimated standard deviation of the noise. With the assumption of Gaussian noise,  $\beta$  can be set to 1 (see [7] for theoretical justifications) and the standard deviation of noise is estimated via pseudo-residuals  $\epsilon_{\mathbf{x}}$  as defined in [17]. For each pixel  $\mathbf{x}$  of  $I$ , let us define:

$$\epsilon_{\mathbf{x}} = \sqrt{\frac{4}{5}} \left( I(\mathbf{x}) - \frac{1}{4} \sum_{\mathbf{y} \in \mathcal{N}(\mathbf{x})} I(\mathbf{y}) \right), \quad (6)$$

where  $\mathcal{N}(\mathbf{x})$  is the 4-neighborhood at pixel  $\mathbf{x}$ . The standard deviation of noise is estimated using the median absolute deviation:

$$\hat{\sigma} = 1.4826 \text{ med}_i(|\epsilon_i - \text{med}_j(\epsilon_j)|). \quad (7)$$

Note that  $\hat{\sigma}$  is a key parameter that reflects the equilibrium between noise reduction and image destruction. Indeed, this parameter significantly changes the value of the similarities between patches. Basically, the algorithm may not remove details whose intensity range is higher than the estimated noise level of the input image.

In order to decrease the computation time, instead of using the entire image domain  $\Omega$  as the search window for each pixel as in [7], a smaller search window, denoted  $\Omega_{\mathbf{x}}$ , is generally considered. As a consequence, the numerator and the denominator of Eq. 3 contain  $\text{card}(\Omega_{\mathbf{x}})$  terms instead of  $\text{card}(\Omega)$  terms.

### 3.B. Adaptive NLM

The key point of NLM-based denoising techniques is the definition of the weighted graph. We aim at providing a more flexible NLM algorithm, that adapts to the content of the image. In this regard, our contribution concerns three important parameters of the graph  $w$ : the function  $\phi$ , the patch size and the patch shape. In practice, at each pixel  $\mathbf{x}$ , the size of patch that will be used for the computation of  $\{w(\mathbf{x}, \mathbf{y})\}_{\mathbf{y} \in \Omega_{\mathbf{x}}}$  is first determined. The shape of each patch is then determined for each pair  $\{\mathbf{x}, \mathbf{y}\}$ . Finally, small weights are set to 0 in order to discard dissimilar patches.

#### 3.B.1. Function $\phi$

While the original version of the NLM algorithm [7] uses  $\phi(x) = e^{-x}$  as a kernel, the use of kernels with finite support may yield better results [18]. Indeed, the exponential form of  $\phi$  used in [7] assigns positive weights even to very dissimilar patches. Although the weights of these patches are usually very small, the denoising process may not be optimal due to all these non-negligible small contributions.

In this work, the traditional kernel function  $\phi(x) = e^{-x}$  has been used, but we propose to remove the contributions of the outlier patches using statistical arguments. Let us recall that if  $X_1, \dots, X_N$  are independent, standard normal random variables, then the sum  $\sum_{i=1}^N X_i^2$  of their squares is distributed according to the chi-square distribution with  $N$  degrees of freedom. So, by assuming that two patches  $P_{\mathbf{x}}$  and  $P_{\mathbf{y}}$  are associated to the same uniform area and that the noise is white Gaussian with variance  $\sigma^2$ , we have:

$$N.\beta.\frac{d_P(\mathbf{x}, \mathbf{y})^2}{2N\beta\sigma^2} \sim \chi^2(N). \quad (8)$$

We use this property to efficiently define a threshold for outlier removal.

Given a p-value  $p$ , a threshold  $t$  can be estimated such that the probability of random value defined in Eq. 8 exceeding  $t$  is equal to  $p$ . Since  $\phi$  is a decreasing function on  $\mathbb{R}^+$ , then  $P(w(\mathbf{x}, \mathbf{y}) < \phi(t/N.\beta)) = p$ . The patches with weights  $w(\mathbf{x}, \mathbf{y})$  less than  $t_0 = \phi(t/N.\beta)$  are considered as outliers, and their weights are set to 0. In the following,  $p$  is set to 0.05. Such a

thresholding procedure is very efficient to discard irrelevant (i.e. dissimilar) patches during the denoising process.

### 3.B.2. Size and shape of patches

Buades *et al.* proposed to use square patches in the original NLM algorithm version. The parameter  $k$  is generally chosen equal to 1, 2 or 3, which leads to patches of size  $3 \times 3$ ,  $5 \times 5$  or  $7 \times 7$ . The half-width  $k$  is usually considered as a global parameter. However, using a global setting may not lead to an optimal denoising procedure since the information redundancy principle on which the NLM algorithm relies may be scale-dependent. Seen as a problem of multi-scale image representation, the estimation of an optimal patch size appears to be also related to the estimation of an optimal patch shape. While the square shape has the advantage of keeping the algorithm simple, the overall denoising performance might not be optimal around high contrasted edges. To deal with this issue, one possibility is to consider anisotropic supports. In [14], Katkovnik *et al.* propose a star-shaped neighborhood which is approximated by a sectorial decomposition. The use of such a decomposition allows different adaptive scales for different directions. In [15], Deledalle *et al.* also propose to use families of arbitrary shapes (disks, half-pies and quarter-pies, bands) to improve the standard NLM performance.

In this work, we propose an adaptive technique to modify the shape of the patches during the estimation of the weights  $w(\mathbf{x}, \mathbf{y})$ . The patch shape is locally estimated in a non-parametric fashion for each pair  $(\mathbf{x}, \mathbf{y})$  of pixels. As shown in Fig. 1, the patch support  $S$  is decomposed into two disjoint parts, denoted  $S_1$  and  $S_2$ , such that  $S = S_1 \cup S_2$ .  $S_1$  is the center square fully contained in  $S$ , and  $S_2$  is the one-pixel wide peripheral border. In order to keep a coherent local descriptor, all pixels of  $S_1$  are used in the computation of  $w(\mathbf{x}, \mathbf{y})$ . To build variable shapes, pixels of  $S_2$  considered as possible outliers do not take part in the computation of  $w(\mathbf{x}, \mathbf{y})$ . Combining such a patch partitioning with a data driven approach allows to handle locally various shapes in a simple and efficient way. In practice, a fourth of the pixels of  $S_2$  (corresponding to the most dissimilar pixels between the two patches  $P_{\mathbf{x}}$  and  $P_{\mathbf{y}}$ ) are removed for the computation of  $w(\mathbf{x}, \mathbf{y})$ . Note that the aim of the previous subsection was to discard irrelevant patches, that is, to set  $w(\mathbf{x}, \mathbf{y})$  to zero for some couples, whereas the purpose is here to discard some pixels for the computation of  $w(\mathbf{x}, \mathbf{y})$ .

The distance between patches can be expressed as follows:

$$d_P(\mathbf{x}, \mathbf{y})^2 = \min_S \sum_{\boldsymbol{\tau} \in \llbracket -k, k \rrbracket^2} S_{(\mathbf{x}, \mathbf{y})}(\boldsymbol{\tau}) (I(\mathbf{x} + \boldsymbol{\tau}) - I(\mathbf{y} + \boldsymbol{\tau}))^2, \quad (9)$$

where  $S_{(\mathbf{x}, \mathbf{y})}(\boldsymbol{\tau})$  defines the shape of patches  $P_{\mathbf{x}}$  and  $P_{\mathbf{y}}$  ( $S_{(\mathbf{x}, \mathbf{y})}(\boldsymbol{\tau})$  equals 1 for all pixels except for a fourth of the pixels of  $S_2$  for which  $S_{(\mathbf{x}, \mathbf{y})}(\boldsymbol{\tau})$  is set to 0). The computation of Eq. 9 is straightforward using classical sorting algorithms.

Finally, it is known that the use of large patches leads to high performance denoising in image parts with uniform intensity while small patches are more adapted in contrasted areas. Here, we propose a sequential procedure to automatically choose the local optimal patch size. The underlying idea is to use the largest possible patch. Basically, we start with a large patch size  $(2.k_{max} + 1) \times (2.k_{max} + 1)$  and compute the corresponding weights  $\{w(\mathbf{x}, \mathbf{y})\}_{\mathbf{y} \in \Omega_{\mathbf{x}}}$ . If not enough similar patches are found in the search window  $\Omega_{\mathbf{x}}$ , the patch size is decreased (it is expected to observe more small similar patches than large similar patches) and the set of weights  $\{w(\mathbf{x}, \mathbf{y})\}_{\mathbf{y} \in \Omega_{\mathbf{x}}}$  is recomputed. This procedure is repeated until a minimum number of similar patches is reached or until the minimal size of the patch is reached ( $k_{min} = 1$ ). In practice, the minimum number of similar patches (*i.e.*  $\text{card}(\{P_{\mathbf{y}}, \mathbf{y} \in \Omega_{\mathbf{x}} | w(\mathbf{x}, \mathbf{y}) > 0\})$ ) is set to  $N_0$ . Since the optimal patch size at each pixel is estimated starting from a maximal value of  $2k_{max} + 1$ ,  $k_{max}$  can be arbitrary set to a very high value. However,  $k_{max}$  has to be small to lower the computational cost. For traditional images, a patch of size  $9 \times 9$  ( $k_{max} = 4$ ) is generally sufficient to estimate the similarity between two pixels.

The overall weight computation is described in Alg. 1.

---

**Algorithm 1** Computation of the weights at pixel  $\mathbf{x}$

---

Input: image  $I$ , maximal patch size  $k_{max}$ , search window centered around the point  $\mathbf{x}$ :  $\Omega_x$

Output: weights  $w(\mathbf{x}, .)$

$k_{min} = 1$

**for**  $k = k_{max}$  to  $k_{min}$  **do**

    Estimate the threshold  $t_0$  such that  $P(w(., .) < t_0) = 0.05$ .

**for each**  $\mathbf{y} \in \Omega_x$  **do**

        Compute  $w(\mathbf{x}, \mathbf{y})$  according to Eq. 5 and Eq. 9.

        Set  $w(\mathbf{x}, \mathbf{y})$  to 0 if  $w(\mathbf{x}, \mathbf{y}) < t_0$ .

**end for**

    Set  $w(\mathbf{x}, \mathbf{x})$  to the maximal value of  $\{w(\mathbf{x}, \mathbf{y})\}_{\mathbf{y} \in \Omega_x}$

**if** there are at least  $N_0$  values of  $w(\mathbf{x}, .)$  strictly greater than 0 **then**

        leave the "for" loop

**end if**

**end for**

---

Note that the value of  $N_0$  (the minimum number of similar patches) may have a direct influence on the results. If less than  $N_0$  similar patches are observed, the size of the patch is reduced in order to possibly detect more patches. This is due to the fact that less than  $N_0$  observations are not considered enough to denoise the images efficiently (as an example, in a uniform area corrupted by white noise and under the hypothesis that all weights are equal,



the standard deviation of the noise is expected to be decreased by the factor  $1/\sqrt{N}$  if  $N$  similar patches are considered). Small values of  $N_0$  can also lead to poor denoising properties (large patches will be considered even if a small number of similar patches are found). On the opposite, a high value of  $N_0$  will lead to always consider patches of small size. In practice, the choice of  $N_0$  may vary according to the local geometrical and intensity properties of the considered image. For instance, the number of examples that can be found is expected to be higher for a pixel at the center of a flat region in term of intensity, than for a pixel at the border of a region that is mainly constituted of edges, and that exhibits a smoothly varying intensity signature. Consequently,  $N_0$  has to be high for an image composed of large and uniform regions, and small for an image composed of small regions with varying intensity signatures. In the following,  $N_0$  has been arbitrarily set to 40.

### 3.C. A non-local means filter for the estimation of Stokes vectors

The non-local means filtering approach can be considered as a weighted least squares problem: under the hypothesis that the weights  $w(\mathbf{x}, \cdot)$  have been normalized such that they sum up to 1, we can write:

$$\begin{aligned} \forall \mathbf{x} \in \Omega, \quad I_{nlm}(\mathbf{x}) &= \sum_{\mathbf{y} \in \Omega_{\mathbf{x}}} w(\mathbf{x}, \mathbf{y}) I(\mathbf{y}) \\ &= \arg \min_a \sum_{\mathbf{y} \in \Omega_{\mathbf{x}}} w(\mathbf{x}, \mathbf{y}) (I(\mathbf{y}) - a)^2. \end{aligned} \quad (10)$$

In our application, we have  $K$  ( $K \geq 4$ ) measurement images. The image  $\mathbf{I}$  can also be considered as an image with  $K$  components :  $\mathbf{I} = (I^{(1)}, I^{(2)}, I^{(3)}, \dots, I^{(K)})$ . Denoising  $\mathbf{I}$  can be performed for each channel independently as follows:

$$\begin{aligned} \forall \mathbf{x} \in \Omega, \quad \mathbf{I}_{nlm}(\mathbf{x}) &= \sum_{\mathbf{y} \in \Omega_{\mathbf{x}}} \mathbf{D}_{\mathbf{x}}(\mathbf{y}) \cdot \mathbf{I}(\mathbf{y}) \\ &= \arg \min_{\mathbf{a}} \sum_{\mathbf{y} \in \Omega_{\mathbf{x}}} (\mathbf{I}(\mathbf{y}) - \mathbf{a})^T \cdot \mathbf{D}_{\mathbf{x}}(\mathbf{y}) \cdot (\mathbf{I}(\mathbf{y}) - \mathbf{a}), \end{aligned} \quad (11)$$

where  $\mathbf{D}_{\mathbf{x}}(\mathbf{y})$  is a  $K \times K$  diagonal matrix. The diagonal element  $d_{ii}$  ( $i = 1 \dots K$ ) is the normalized weight between pixel  $\mathbf{x}$  and pixel  $\mathbf{y}$  for the  $i^{th}$  channel (which can be considered as a scalar image). As a consequence,  $\sum_{\mathbf{y} \in \Omega_{\mathbf{x}}} \mathbf{D}_{\mathbf{x}}(\mathbf{y})$  is equal to the identity matrix.

However, our purpose is not to denoise the measurement images, but rather to estimate a denoised version of the Stokes vectors. Eq. 11 becomes:

$$\forall \mathbf{x} \in \Omega, \quad \hat{\mathbf{S}}(\mathbf{x}) = \arg \min_{\mathbf{S}} \sum_{\mathbf{y} \in \Omega_{\mathbf{x}}} (\mathbf{I}(\mathbf{y}) - \mathbf{P} \cdot \mathbf{S})^T \cdot \mathbf{D}_{\mathbf{x}}(\mathbf{y}) \cdot (\mathbf{I}(\mathbf{y}) - \mathbf{P} \cdot \mathbf{S}). \quad (12)$$

Three comments can be made about Eq. 12. The first one is that this equation represents a joint filtering-estimation of the Stokes vectors. Nevertheless, admissibility constraints are not taken into account. This point will be addressed later.

The second comment is that minimizing this criterion is equivalent to minimizing the quadratic reconstruction error between the observed measurements  $\mathbf{I}$  and the predicted ones  $\mathbf{P.S}$  if no filtering is performed, which is equivalent to the pseudo-inverse approach since no admissibility constraint has been defined yet. Indeed, without any filtering, the matrix  $\mathbf{D}_{\mathbf{x}}(\mathbf{y})$  is the null matrix if  $\mathbf{y} \neq \mathbf{x}$ , and the identity matrix otherwise: only data associated to pixel  $\mathbf{x}$  are considered for the estimation of the Stokes vector at this pixel.

The last comment is related to the estimation of  $\mathbf{D}_{\mathbf{x}}(\mathbf{y})$ . Coefficient  $d_{ii}$  is the normalized weight between pixel  $\mathbf{x}$  and pixel  $\mathbf{y}$  for the  $i^{th}$  channel. It is estimated using the approach presented in Sec. 3.B. However, it could be envisaged to consider other matrices  $\mathbf{D}_{\mathbf{x}}(\mathbf{y})$  related to other approaches.

Note finally that the criterion of Eq. 12 can be expressed more simply and compactly. Since  $\mathbf{D}_{\mathbf{x}}(\mathbf{y})$  is diagonal and symmetrical, the gradient  $\nabla$  of the criterion writes:

$$\begin{aligned}\nabla &= -2.\mathbf{P}^T \sum_{\mathbf{y} \in \Omega_{\mathbf{x}}} \mathbf{D}_{\mathbf{x}}(\mathbf{y}).(\mathbf{I}(\mathbf{y}) - \mathbf{P.S}) \\ &= -2.\mathbf{P}^T (\sum_{\mathbf{y} \in \Omega_{\mathbf{x}}} \mathbf{D}_{\mathbf{x}}(\mathbf{y}).\mathbf{I}(\mathbf{y}) - \sum_{\mathbf{y} \in \Omega_{\mathbf{x}}} \mathbf{D}_{\mathbf{x}}(\mathbf{y})\mathbf{P.S}) \\ &= -2.\mathbf{P}^T (\sum_{\mathbf{y} \in \Omega_{\mathbf{x}}} \mathbf{D}_{\mathbf{x}}(\mathbf{y})\mathbf{I}(\mathbf{y}) - \mathbf{P.S}),\end{aligned}\tag{13}$$

since  $\sum_{\mathbf{y} \in \Omega_{\mathbf{x}}} \mathbf{D}_{\mathbf{x}}(\mathbf{y})$  is equal to the identity matrix. Finally, by definition of  $\mathbf{I}_{nlm}$  (Eq. 11), the gradient of the criterion can be written as:

$$\nabla = -2.\mathbf{P}^T (\mathbf{I}_{nlm}(\mathbf{x}) - \mathbf{P.S}).\tag{14}$$

This gradient is equal to the gradient of the cost function  $||\mathbf{I}_{nlm}(\mathbf{x}) - \mathbf{P.S}||^2$ . The Stokes vector can thus be estimated as:

$$\forall \mathbf{x} \in \Omega, \quad \hat{\mathbf{S}}(\mathbf{x}) = \arg \min_{\mathbf{S}} ||\mathbf{I}_{nlm}(\mathbf{x}) - \mathbf{P.S}||^2,\tag{15}$$

both criteria (Eq. 12, Eq. 15) being equal up to a constant. The criterion of Eq. 15 corresponds to denoising each channel using a NLM approach, and then to estimating the Stokes vector at each pixel separately. For each pixel, the criterion of Eq. 15 can be optimized using the pseudo-inverse solution, but this method suffers from a major limitation: the estimated Stokes vector may not verify the admissibility constraint. To overcome this limitation, the criterion of Eq. 15 is optimized using a constrained optimization procedure (see Appendix 1). Note that solving the criterion of Eq. 15 under the physical admissibility constraint is equivalent to solve the one of Eq. 12 under the same constraint since both criteria are equal to a constant.

Finally, if  $\hat{\mathbf{S}}$  denotes the estimated Stokes vector image, the image  $\hat{\mathbf{I}}$  defined as  $\hat{\mathbf{I}}(\mathbf{x}) = \mathbf{P}.\hat{\mathbf{S}}(\mathbf{x})$  for each  $\mathbf{x}$ , can be considered as the denoised version of  $\mathbf{I}$ . In the case  $k = 4$ , note that the images  $\hat{\mathbf{I}}$  and  $\mathbf{I}_{nlm}$  differ only for the pixels for which the pseudo-inverse solution does not satisfy the admissibility constraints.

## 4. Results

### 4.A. Application to simulated data

#### 4.A.1. Evaluation protocol

In order to evaluate the relevance of the proposed approach, simulations were carried out on simulated data. Methods were assessed considering signal-to-noise ratio. We synthesized a  $256 \times 256$  pixel Stokes image  $\mathbf{S}^{gt}$  composed of two distinct regions: (i) a background with a uniform polarization signature  $\mathbf{S} = [1, 1/\sqrt{3}, 1/\sqrt{3}, 1/\sqrt{3}]^T$ , and (ii) a 100 pixel radius circle with a smoothly varying completely polarized Stokes signature placed in the center of the image (see Fig. 2). The data necessary to carry out our study were generated as described in [2]. The observation model corresponding to an optimal rotating wave-plate Stokes-meter was used to generate the intensity images  $\mathbf{I}^{gt}$  that were degraded by adding white Gaussian noise of variance  $\sigma^2$ . We assume in the following that no systematic errors occur, i.e., that the PMM is error free. Note that the polarization states of the synthetic target lie on the Poincaré sphere and even small noise lead also to non physical solutions if one uses the pseudo-inverse approach. This constitutes a hard test case.

Since the ground truth is known, estimation accuracy can be evaluated by comparing the estimated values (the Stokes vectors  $\hat{\mathbf{S}}$  and the associated intensity values  $\hat{\mathbf{I}}$ ) with the original ones ( $\mathbf{S}^{gt}$  and  $\mathbf{I}^{gt}$ ). The method is first evaluated by comparing the original image  $\mathbf{I}^{gt}$  (noise-free intensity image) with its estimation  $\hat{\mathbf{I}}$  using the peak signal-to-noise ratio (PSNR):

$$PSNR(\mathbf{I}^{gt}, \hat{\mathbf{I}}) = 10 \log_{10} \left( \frac{d^2}{\frac{1}{4P} \sum_{j=1}^4 \alpha_j^2 \sum_{\mathbf{x}} \left( I_j^{gt}(\mathbf{x}) - \hat{I}_j(\mathbf{x}) \right)^2} \right), \quad (16)$$

where  $\alpha_j$  is computed so that the dynamic range of  $\alpha_j \cdot \mathbf{I}_j^{gt}$  is  $d$  (for example 255), and where  $P$  is the number of pixels. Complementary to this PSNR-based assessment, the method is also assessed considering Stokes vectors as:

$$e(\mathbf{S}^{gt}, \hat{\mathbf{S}}) = 100 \sqrt{\frac{1}{P} \sum_{\mathbf{x}} \frac{\|\mathbf{S}^{gt}(\mathbf{x}) - \hat{\mathbf{S}}(\mathbf{x})\|^2}{\|\mathbf{S}^{gt}(\mathbf{x})\|^2}}. \quad (17)$$

Results obtained with the proposed approach are compared with several methods in the next three sections. Methods used for comparison are chosen so as to illustrate (i) the benefit of using spatial information (section 4.A.2), (ii) the relevance of the proposed weight determination procedure (section 4.A.3), and (iii) the benefit of accounting for the admissibility constraints (section 4.A.4).

#### 4.A.2. Benefit of using spatial information

Five methods,  $M_1$ ,  $M_2$ ,  $M_3$ ,  $M_4$ , and  $M_P$ , are evaluated:

- $M_1$ :  $\hat{\mathbf{S}}(\mathbf{x})$  is the pseudo-inverse solution;
- $M_2$ :  $\hat{\mathbf{S}}(\mathbf{x})$  is the pseudo-inverse solution, further orthogonally projected onto the Stokes vector set if the admissibility constraints are not verified;
- $M_3$ :  $\hat{\mathbf{S}}(\mathbf{x})$  is computed pixel-wise (without any spatial filtering) by taking the admissibility constraints into account (see appendix for details);
- $M_4$  is the approach proposed by Sfikas *et al.* [6];
- $M_P$  is the proposed approach with a search window of size  $11 \times 11$ , and a maximal patch size of  $9 \times 9$ .

Note that the methods  $M_1$ ,  $M_2$ , and  $M_3$  do not use any spatial filtering.

The PSNR (Eq. 16) and the Stokes vector estimation error (eq. 17) corresponding to  $M_1$ ,  $M_2$ ,  $M_3$ ,  $M_4$  and  $M_P$  for different values of  $\sigma^2$  (from  $\sigma^2 = 0.001$  to 0.05) are given in Tab. 1.

$\sigma^2$	$M_1$	$M_2$	$M_3$	$M_4$	$M_P$	$\sigma^2$	$M_1$	$M_2$	$M_3$	$M_4$	$M_P$
0.001	24.09	24.17	24.81	30.81	<b>43.49</b>	0.001	5.09	4.94	4.67	2.35	<b>0.54</b>
0.002	21.08	21.16	21.86	28.40	<b>40.33</b>	0.002	7.20	6.98	6.55	3.12	<b>0.79</b>
0.005	17.12	17.21	17.98	25.26	<b>35.85</b>	0.005	11.34	10.95	10.22	4.44	<b>1.31</b>
0.01	14.08	14.15	15.02	23.02	<b>32.31</b>	0.01	16.09	15.47	14.32	5.60	<b>1.95</b>
0.02	11.07	11.13	12.15	20.50	<b>29.06</b>	0.02	22.76	21.73	19.86	7.44	<b>2.86</b>
0.05	7.09	7.14	8.46	16.97	<b>25.24</b>	0.05	36.02	33.86	30.06	10.97	<b>4.35</b>

Table 1. PSNR (left) and Stokes vector estimation error (right) obtained with the five different methods, and for different values of  $\sigma^2$ . Bold values correspond to the best results.

Tab. 1 shows the interest of considering spatial information in the estimation of Stokes vectors. The proposed approach  $M_P$ , and  $M_4$  outperform clearly the methods  $M_1$ ,  $M_2$ , and  $M_3$ . As an example, the PSNR (left part of Tab. 1) obtained with the proposed approach ( $M_P$ ) is at least 15 dB greater than the one obtained with  $M_1$ ,  $M_2$ , or  $M_3$ , and the estimation error (right part of Tab. 1) is decreased of a factor varying from 8 to 10.

We note also that the proposed approach is more efficient than the method  $M_4$  [6]. This is mainly due to the efficiency of the NLM algorithm that has been shown to outclass other denoising methods.

Finally, amongst the methods which do not consider spatial information,  $M_3$  is the one providing the best results. This shows that the traditional way of projecting a vector onto the

Stokes vector set (method  $M_2$ , orthogonal projection) is not an efficient approach, compared to the proposed method which reduces the most the error reconstruction between the observed measurements and the predicted ones, while constraining the solution to be physically admissible.

Fig. 3 summarizes the interest of the proposed approach. The original image  $\mathbf{S}^{gt}$  (top left), the estimates  $\hat{\mathbf{S}}$  computed using  $\mathbf{I}$  ( $\sigma^2 = 0.01$ ) with the method  $M_1$  (top right), with the method  $M_4$  (bottom left) and with the proposed approach  $M_P$  (bottom right) are presented using the Poincaré sphere representation. Note that it is difficult to observe the homogeneous region (background) in the original image  $\mathbf{S}^{gt}$  with this representation, since it is composed of a unique Stokes vector. Results obtained with  $M_1$  provide vectors which do not belong to the sphere, and it is difficult to distinguish between the two regions (the circle and the background). Results obtained with  $M_4$  and  $M_P$  enable to filter the noise while constraining the solution to be physically admissible. The two regions can then easily be observed. Note also that the images show clearly that the proposed approach enables to achieve a better noise filtering. Note also that Fig. 2 leads to a similar conclusion.

#### 4.A.3. Evaluation of the proposed weight determination procedure

We propose here to compare the proposed weight determination procedure (the size and the shape of the patches are determined according to the procedure of Sec. 3.B) with the traditional one [7]. To this end, we derive four methods ( $T_1$  to  $T_4$ ) that are simplified versions of  $M_P$ , in the sense that the weights are estimated as in [7] (the patch size is a square of fixed size, and the contributions of outlier patches are not removed). In all cases, the search window has been set to a size of  $11 \times 11$ , whereas the patch size has been set to  $3 \times 3$  for  $T_1$ , to  $5 \times 5$  for  $T_2$ , to  $7 \times 7$  for  $T_3$ , and to  $9 \times 9$  for  $T_4$ . Results obtained with these methods are given in Tab. 2.

We can first notice that the proposed approach outperforms the methods  $T_1$ ,  $T_2$ ,  $T_3$ , and  $T_4$  for each experiment. This clearly highlights the interest of the approach which has been defined for the estimation of the weights. Finally, results of Tab. 2 show clearly that the size of the patch may have an impact on the results. In this case, we can observe that neither  $T_1$  nor  $T_4$  enables to achieve the best results (amongst  $T_1$ ,  $T_2$ ,  $T_3$ , and  $T_4$ ), leading to conclude that the size of the patch has to be chosen neither too small nor too large. This clearly highlights the benefit of an adaptive strategy as in the proposed approach.

Moreover, the proposed approach is versatile: other strategies can be envisaged to compute the weights. Indeed, the criterion of Eq. 15 under the physical admissibility constraint is equivalent to the one of Eq. 12 independently of the way the weights are defined. The only constraint is that  $\sum_{\mathbf{y} \in \Omega_{\mathbf{x}}} \mathbf{D}_{\mathbf{x}}(\mathbf{y})$  is the identity matrix (for all  $\mathbf{x}$ ). It is also possible to determine the weights of Eq. 12 so as to obtain a Gaussian or a median filtering. We propose

$\sigma^2$	$T_1$	$T_2$	$T_3$	$T_4$	$M_P$	$\sigma^2$	$T_1$	$T_2$	$T_3$	$T_4$	$M_P$
0.001	40.24	<b>40.68</b>	38.98	37.41	43.49	0.001	0.79	<b>0.76</b>	0.91	1.08	0.54
0.002	37.23	<b>37.57</b>	36.17	35.06	40.33	0.002	1.13	<b>1.08</b>	1.24	1.40	0.79
0.005	33.00	33.24	<b>32.26</b>	31.44	35.85	0.005	1.82	<b>1.74</b>	1.93	2.10	1.31
0.01	29.85	<b>30.41</b>	29.62	28.89	32.31	0.01	2.59	<b>2.41</b>	2.61	2.83	1.95
0.02	26.92	<b>27.80</b>	27.25	26.66	29.06	0.02	3.64	<b>3.25</b>	3.43	3.65	2.86
0.05	23.18	<b>24.54</b>	24.39	24.12	25.24	0.05	5.57	<b>4.70</b>	4.75	4.88	4.35

Table 2. PSNR (left) and Stokes vector estimation error (right) obtained with four simplified versions of the proposed approach ( $T_1$ ,  $T_2$ ,  $T_3$ , and  $T_4$ ), and with the proposed approach for different values of  $\sigma^2$  (see text for details). Bold values correspond to the best results (without considering the proposed approach  $M_P$  which provides always the best results).

to first filter the images with a Gaussian filter (GF) or a median filter (MF), and then to optimize the criterion of Eq. 15 under the physical admissibility constraint (where  $\mathbf{I}_{nlm}$  is now the filtered images). Several values of the Gaussian filter standard deviation (from 0.1 to 3 with a step of 0.1), and several median filter sizes (from a size of  $3 \times 3$  to  $11 \times 11$ ) have been tested. Results are presented in Tab. 3. For each experiment, only the best result in terms of PSNR and of estimation error is given as well as the corresponding setting in parentheses (filter size for MF or standard deviation for GF).

Results of Tab. 3 show clearly that the proposed approach outperforms the Gaussian and the median filtering based approaches. Moreover, since the only difference between the three methods concerns the way the weights are computed, this also shows that the method can benefit from future advances in denoising methods provided that the filtered image can be expressed with Eq. 10, that is, the denoised value at a pixel is a weighted sum of the pixel values of the original image (with the constraint that the weights sum up to 1).

#### 4.A.4. Benefit of accounting for the physical admissibility constraint

We propose to compare the proposed method with the two following approaches:

- NLMPI:  $\hat{\mathbf{S}}(\mathbf{x})$  is the pseudo-inverse solution computed using  $\mathbf{I}_{nlm}$ . This means that  $\hat{\mathbf{S}}(\mathbf{x})$  is not necessarily a Stokes vector since it may not verify the admissibility constraints, but  $\hat{\mathbf{S}}(\mathbf{x})$  is the global minimum of the criterion of Eq. 15. Note that the proposed approach and the NLMPI method give identical results if the pseudo-inverse solution verifies the admissibility constraints.

$\sigma^2$	GF	MF	$M_P$
0.001	27.00 (0.5)	31.04 ( $3 \times 3$ )	43.49
0.002	25.06 (0.6)	29.38 ( $5 \times 5$ )	40.33
0.005	22.86 (0.7)	26.77 ( $5 \times 5$ )	35.85
0.01	21.39 (0.9)	24.57 ( $7 \times 7$ )	32.31
0.02	20.12 (1.2)	22.52 ( $7 \times 7$ )	29.06
0.05	18.61 (1.7)	19.57 ( $7 \times 7$ )	25.24

$\sigma^2$	GF	MF	$M_P$
0.001	3.44 (0.5)	1.98 ( $5 \times 5$ )	0.54
0.002	4.15 (0.6)	2.43 ( $5 \times 5$ )	0.79
0.005	5.27 (0.8)	3.35 ( $5 \times 5$ )	1.31
0.01	6.16 (1.1)	4.28 ( $7 \times 7$ )	1.95
0.02	7.02 (1.4)	5.50 ( $7 \times 7$ )	2.86
0.05	8.29 (1.9)	7.49 ( $9 \times 9$ )	4.35

Table 3. PSNR (left) and Stokes vector estimation error (right) obtained with the Gaussian (GF) and with the median (MF) filtering based approaches, and with the proposed approach ( $M_P$ ) for different values of noise ( $\sigma^2$ ). For each case, only the best result is given as well as the corresponding setting in parentheses (filter size for MF and standard deviation for GF).

- NLMPROJ:  $\hat{\mathbf{S}}(\mathbf{x})$  is the pseudo-inverse solution computed using  $\mathbf{I}_{nlm}$ , further orthogonally projected onto the Stokes vector set if the admissibility constraints are not verified.

Results are presented in Tab. 4.

$\sigma^2$	NLMPI	NLMPROJ	$M_P$
0.001	41.55	41.63	43.49
0.002	38.30	38.36	40.33
0.005	33.97	34.18	35.85
0.01	30.65	30.79	32.31
0.02	27.52	27.63	29.06
0.05	23.77	23.88	25.24

$\sigma^2$	NLMPI	NLMPROJ	$M_P$
0.001	0.67	0.66	0.54
0.002	1.00	0.97	0.79
0.005	1.61	1.57	1.31
0.01	2.36	2.31	1.95
0.02	3.41	3.34	2.86
0.05	5.19	5.08	4.35

Table 4. PSNR (left) and Stokes vector estimation error (right) obtained with two simplified versions of the proposed approach (NLMPI, NLMPROJ), and with the proposed approach for different values of  $\sigma^2$  (see text for details).

The use of the admissibility constraint permits to improve the PSNR and the Stokes vector estimation accuracy (see Tab 4). The constraint acts like a prior that enables to achieve a better estimation. Moreover, the proposed way to account for the admissibility constraint, namely to constrain the solution to be physically admissible while reducing the most the



error reconstruction, outperforms orthogonal projection. Note finally that such a conclusion has already been drawn in Sec. 4.A.2 (methods  $M_1$ ,  $M_2$ , and  $M_3$ ).

#### 4.B. Results on real data

To test the ability of the method to handle real images, we used a well calibrated Stokes polarimeter [19] to image a real scene. The scene consisted of an ensemble of objects (see Fig. 4) lighted with multidirectional partially polarized light. Observations were carried out through a narrow band interferential filter to ensure that the PMM is known with high accuracy. For the sake of conciseness, we present only the degree of polarization (DOP), orientation and ellipticity images that are obtained with the proposed approach (left part of Fig. 4), and the pseudo-inverse solution (right part). We observe that the three estimated characteristics, especially the ellipticity and the orientation, are less noisy with the proposed approach. With respect to the degree of polarization, the same conclusion can be drawn if we look more closely at the local variations of the polarization signature (see for example Fig. 5 which is a zoom in on the foot of the figurine).

More generally, we found that for all cases considered here, our approach performed better than the other methods considered, without having to tune any parameter.

## 5. Conclusion

Polarimetric imaging carries information absent in conventional optical imaging, and as such is of great interest for many applications. However, exploiting polarimetric data requires special care. Admissibility constraints have to be taken into account. Besides, noise stemming from narrow spectral band measurements and contaminating data has to be handled.

Traditional approaches estimate Stokes vectors without considering the admissibility constraints, and without considering spatial information to reduce the impact of noise. We have presented here an algorithm which overcomes these two limitations. The proposed joint filtering-estimation procedure can be equivalently achieved sequentially by a two step method: a filtering stage, and an estimation step. Noise is filtered using a very efficient denoising approach – the non-local means algorithm – and admissibility constraints are ensured in the framework of convex optimization. We assessed the performance of the proposed approach on simulated and real data.

### A. Estimation of the Stokes vector at a pixel

#### A.A. Problem statement and solution

The purpose of this appendix is to show how to compute the Stokes vector at pixel  $\mathbf{x}$  using the intensity of the denoised image  $\mathbf{I}_{nlm}$  at pixel  $\mathbf{x}$ . Since the processing is achieved pixel-wise, the term  $\mathbf{x}$  will not be considered so that  $\mathbf{I}$  denotes here the  $K$  measures associated to



a pixel, and  $\mathbf{S}$  represents a Stokes vector. The aim is then to solve the following optimization problem:

$$\hat{\mathbf{S}} = \arg \min_{\mathbf{S}} \|\mathbf{I} - \mathbf{P} \cdot \mathbf{S}\|^2, \quad (18)$$

or more precisely the following one:

$$\hat{\mathbf{S}} = \arg \min_{\mathbf{S} \in \mathbf{B}} \|\mathbf{I} - \mathbf{P} \cdot \mathbf{S}\|^2, \quad (19)$$

where  $\mathbf{B}$  denotes the set of admissible Stokes vectors.

The criterion of Eq. 18 has a unique minimum which is the pseudo-inverse solution:  $\hat{\mathbf{V}} = \mathbf{P}^\dagger \mathbf{S}$ . However, this solution may not verify the admissibility constraints. In this case ( $\hat{\mathbf{V}} \notin \mathbf{B}$ ), the criterion of Eq. 19 has a unique minimum denoted  $\mathbf{S}^*$  which belongs to the border  $\partial \mathbf{B}$  of  $\mathbf{B}$ . This is a consequence of the fact that the criterion is strictly convex and that  $\mathbf{B}$  is a convex part of  $\mathbb{R}^4$ . Moreover,  $\mathbf{S}^*$  is the unique point of  $\partial \mathbf{B}$  verifying:

- the gradient of the criterion 18 at point  $\mathbf{S}^*$  is orthogonal to  $\partial \mathbf{B}$ ,
- the gradient of the criterion 18 at point  $\mathbf{S}^*$  gives the direction of the interior of  $\mathbf{B}$ .

Note that the two aforementioned points correspond to the Kuhn-Tucker (KT) conditions. Since there is a unique KT point and since it corresponds to a global minimum, the criterion and the inequality constraint are Type I invex [20]. This means that a nonlinear optimization approach such as a sequential quadratic programming (SQP) is ensured to converge to the global minimum. However, in practice, we have observed that when using a SQP algorithm, the convergence rate can be slow, leading to a poor approximation of the solution, or to a time-consuming algorithm (such algorithms have been tested with the optimization toolbox of Matlab). To solve this problem, we propose an algorithm which takes benefit from the fact that  $\partial \mathbf{B}$  is the set of Stokes vectors with a degree of polarization of 1. The proposed algorithm is much simpler than a SQP approach and has shown to be very efficient.

#### A.B. Optimization algorithm

If  $\hat{\mathbf{V}} \in \mathbf{B}$ , then  $\hat{\mathbf{S}} = \hat{\mathbf{V}}$ . Otherwise,  $\hat{\mathbf{S}}$  has to be searched in  $\partial \mathbf{B}$ . Unfortunately, the criterion  $\|\mathbf{I} - \mathbf{P} \mathbf{S}\|^2$  restricted to  $\partial \mathbf{B}$  is not convex and it may have local minima. However, since the general problem (Eq. 19) is convex, local descent methods can be used to estimate  $\hat{\mathbf{S}}$  if the following two strategies are used iteratively:

1. local descent on the border  $\partial \mathbf{B}$ :

$$\min_{S_1, S_2, S_3} \left\| \mathbf{I} - \mathbf{P} \begin{bmatrix} \sqrt{S_1^2 + S_2^2 + S_3^2}, S_1, S_2, S_3 \end{bmatrix}^T \right\|^2; \quad (20)$$

2. local descent on the interior  $\check{\mathbf{B}}$  of  $\mathbf{B}$ :

$$\min_{\mathbf{S} \in \check{\mathbf{B}}} \|\mathbf{I} - \mathbf{PS}\|^2, \quad (21)$$

with

$$\check{\mathbf{B}} = \left\{ \mathbf{S} \mid S_0 > \sqrt{S_1^2 + S_2^2 + S_3^2} \right\}.$$

The algorithm is initialized on  $\partial\mathbf{B}$  (the orthogonal projection of  $\hat{\mathbf{V}}$  onto  $\mathbf{B}$ ) and the strategy 1 is considered. When a minimum is obtained, estimating the gradient of criterion 18 at this points enables us to determine if this minimum is the global one or a local one: it gives the direction of the interior of  $\mathbf{B}$  *iff* the minimum is the global one. If this minimum is a local one, the criterion can be reduced by entering into  $\mathbf{B}$ . The descent is then continued with strategy 2. In this strategy, the border is then ensured to be met since there is no local minimum in  $\check{\mathbf{B}}$ . Then, the strategy 1 can be used again. The algorithm is described more precisely in Alg. 2.

We have observed that the method converges in all cases at the first iteration of the while loop (see Alg. 2): this means that the solution obtained after the first optimization of the criterion of Eq. 20 provides always an estimation of the global minimum. However, instead of setting  $\hat{\mathbf{S}}$  to the projection of  $\hat{\mathbf{V}}$  onto  $\mathbf{B}$  for the initialization (see Alg. 2), we can set  $\hat{\mathbf{S}}$  randomly in  $\partial\mathbf{B}$ . In this case, the algorithm may need two iterations to converge. Note finally that the proposed approach has been made possible thanks to the fact that  $\partial\mathbf{B}$  can be easily parametrized.

## References

1. R. A. Chipman, “Polarimetry,” in the Handbook of Optics, (McGraw-Hill, 1994), chap. 22.
2. J. Zallat, S. Ainouz, and M.-P. Stoll, “Optimal configurations for imaging polarimeters: impact of image noise and systematic errors,” *Journal of Optics A* **8**, 807–814 (2006).
3. J. Zallat and C. Heinrich, “Polarimetric data reduction: a Bayesian approach,” *Optics Express* **15**, 83–96 (2007).
4. J. Zallat, C. Heinrich, and M. Petremand, “A Bayesian approach for polarimetric data reduction: the Mueller imaging case,” *Optics Express* **16**, 7119–7133 (2008).
5. J. Valenzuela and J. Fessler, “Joint reconstruction of Stokes images from polarimetric measurements,” *Journal of the Optical Society of America A* **26**, 962–968 (2009).
6. G. Sfikas, C. Heinrich, J. Zallat, and C. Nikou, “Recovery of polarimetric Stokes images by spatial mixture models,” *Journal of the Optical Society of America A* **28**, 465–474 (2011).

---

**Algorithm 2** Estimation of the Stokes vector at a pixel

---

Input:  $\mathbf{I}$  (the  $K$  measurements),  $\mathbf{P}$

Output:  $\hat{\mathbf{S}}$  (the associated Stokes vector)

Computation of the pseudo-inverse  $\hat{\mathbf{V}}$

**if**  $\hat{\mathbf{V}} \in \mathbf{B}$  **then**

$\hat{\mathbf{S}} = \hat{\mathbf{V}}$

**else**

    Set  $\hat{\mathbf{S}}$  to the orthogonal projection of  $\hat{\mathbf{V}}$  onto  $\mathbf{B}$

    Set *ASolutionHasBeenFound* to 0

**while** *ASolutionHasBeenFound* == 0 **do**

        Estimate  $\hat{\mathbf{S}}$  by optimizing the criterion of Eq. 20 using the current value of  $\hat{\mathbf{S}}$  as a starting point (descent on the border).

**if** the gradient of Eq. 18 at point  $\hat{\mathbf{S}}$  gives the direction of the exterior of  $\mathbf{B}$  **then**

            Estimate  $\hat{\mathbf{S}}$  by optimizing the criterion of Eq. 21 using  $\hat{\mathbf{S}}$  as a starting point (descent on the interior). For the optimization, a gradient approach is used: a maximal admissible step is computed at each iteration to constrain  $\hat{\mathbf{S}}$  to be a Stokes vector.

        At the end of the procedure,  $\hat{\mathbf{S}}$  is ensured to belong to  $\partial\mathbf{B}$ .

**else**

            set *ASolutionHasBeenFound* to 1.

**end if**

**end while**

**end if**

---

7. A. Buades, B. Coll, and J. Morel, “A review of image denoising algorithms, with a new one,” *Multiscale Modeling & Simulation* **4**, 490–530 (2005).
8. L. Rudin, S. Osher, and E. Fatemi, “Nonlinear total variation based noise removal algorithms,” *Physica D* **60**, 259–268 (1992).
9. R. Coifman and D. Donoho, *Wavelets and Statistics, chapter Translation-invariant denoising* (Springer Verlag, 1995).
10. S. Kinderman, S. Osher, and P. Jones, “Deblurring and denoising of images by nonlocal functionals,” *Multiscale Modeling & Simulation* **4**, 1091–1115 (2005).
11. G. Gilboa and S. Osher, “Nonlocal operators with applications to image processing,” *Multiscale Modeling & Simulation* **7**, 1005–1028 (2008).
12. M. Mignotte, “A non-local regularization strategy for image deconvolution,” *Pattern Recognition Letters* **29**, 2206–2212 (2008).
13. F. Rousseau, “A non-local approach for image super-resolution using intermodality priors,” *Medical Image Analysis* **14**, 594–605 (2010).
14. V. Katkovnik, A. Foi, K. Egiazarian, and J. Astola, “From local kernel to nonlocal multiple-model image denoising,” *International Journal of Computer Vision* **86**, 1–32 (2010).
15. C.-A. Deledalle, V. Duval, and J. Salmon, “Non-local methods with shape-adaptive patches (NLM-SAP),” *Journal of Mathematical Imaging and Vision* pp. 1–18 (2011).
16. P. Coupé, P. Yger, S. Prima, P. Hellier, C. Kervrann, and C. Barillot, “An optimized blockwise nonlocal means denoising filter for 3d magnetic resonance images,” *IEEE Transactions on Medical Imaging* **27**, 425–441 (2008).
17. T. Gasser, L. Sroka, and C. Steinmetz, “Residual variance and residual pattern in non-linear regression,” *Biometrika* **73**, 625–633 (1986).
18. B. Goossens, H. Luong, A. Pizurica, and W. Philips, “An improved non-local means algorithm for image denoising,” in “2008 International Workshop on Local and Non-Local Approximation in Image Processing (LNLA2008),” (Lausanne, Switzerland, August, 2008).
19. J. Zallat, M. Torzynski, and A. Lalllement, “Double-pass self-spectral-calibration of a polarization state analyzer,” *Optics Letters* **37**, 401–403 (2012).
20. M. Hanson, “Invexity and the Kuhn-Tucker theorem,” *Journal of Mathematical Analysis and Applications* **236**, 594–604 (1999).

## List of Figures

1	Example of 3 patches of size 3 (left), 5 (middle), and 7 (left): $S_1$ is in white, and $S_2$ is in gray. . . . .	21
2	Representation of the simulated data. The $i$ -th column is associated to the $i$ -component of the Stokes vector. The first line represents the ground truth, whereas the other lines represent the Stokes vectors that are estimated from the most noisy simulated images ( $\sigma^2=0.05$ ) with the pseudo-inverse solution $M_1$ (second line), the proposed approach $M_P$ (third line), and with $M_4$ (fourth line). . . . .	22
3	Poincaré sphere representation of the original image $\mathbf{S}^{gt}$ (top left), of its estimation from $\mathbf{I}$ ( $\sigma^2 = 0.01$ ) with method $M_1$ (top right), method $M_4$ (bottom left) and the proposed approach $M_P$ (bottom right). . . . .	23
4	Degree of polarization (first line), ellipticity (second line) and orientation (third line) of the Stokes vectors estimated with the proposed approach (left), and by using the pseudo-inverse (right). . . . .	24
5	Degree of polarization of the Stokes vectors estimated with the proposed approach (left) and by using the pseudo-inverse (right). The zoom in on the foot of the figurine reveals the denoising efficiency of the proposed approach. . . .	25

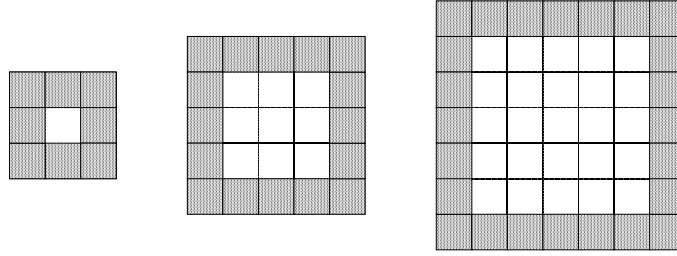


Fig. 1. Example of 3 patches of size 3 (left), 5 (middle), and 7 (left):  $S_1$  is in white, and  $S_2$  is in gray.

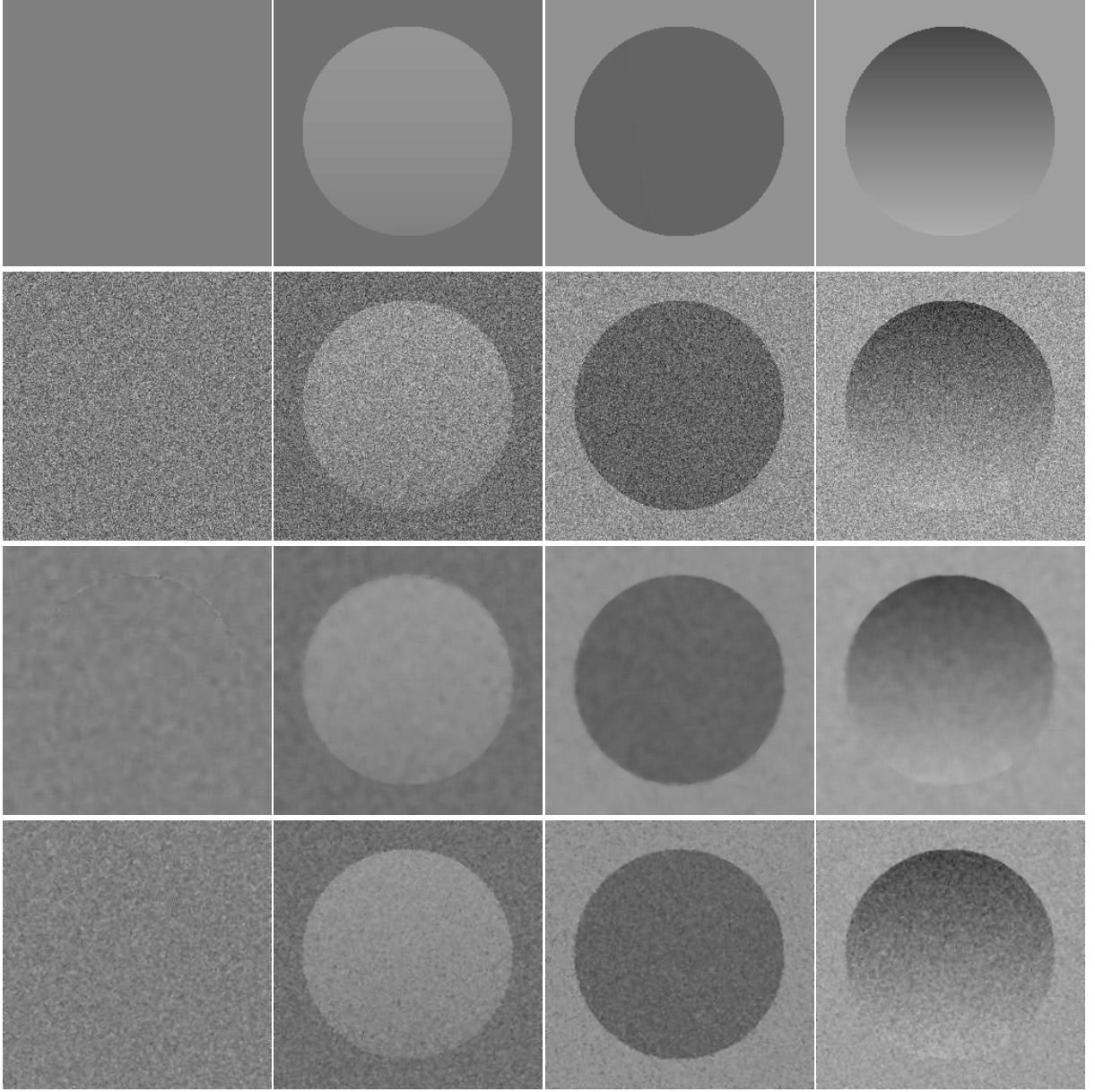


Fig. 2. Representation of the simulated data. The  $i$ -th column is associated to the  $i$ -component of the Stokes vector. The first line represents the ground truth, whereas the other lines represent the Stokes vectors that are estimated from the most noisy simulated images ( $\sigma^2=0.05$ ) with the pseudo-inverse solution  $M_1$  (second line), the proposed approach  $M_P$  (third line), and with  $M_4$  (fourth line).



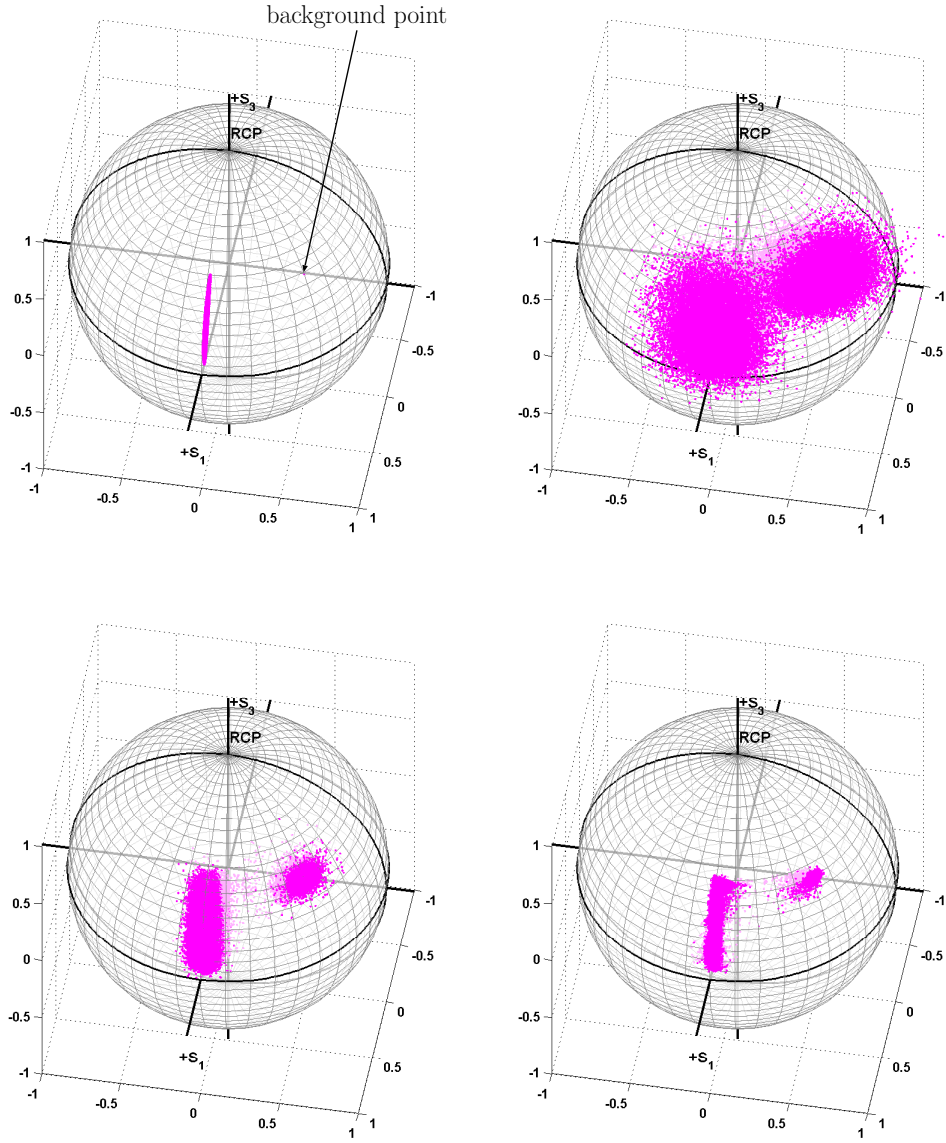


Fig. 3. Poincaré sphere representation of the original image  $\mathbf{S}^{gt}$  (top left), of its estimation from  $\mathbf{I}$  ( $\sigma^2 = 0.01$ ) with method  $M_1$  (top right), method  $M_4$  (bottom left) and the proposed approach  $M_P$  (bottom right).



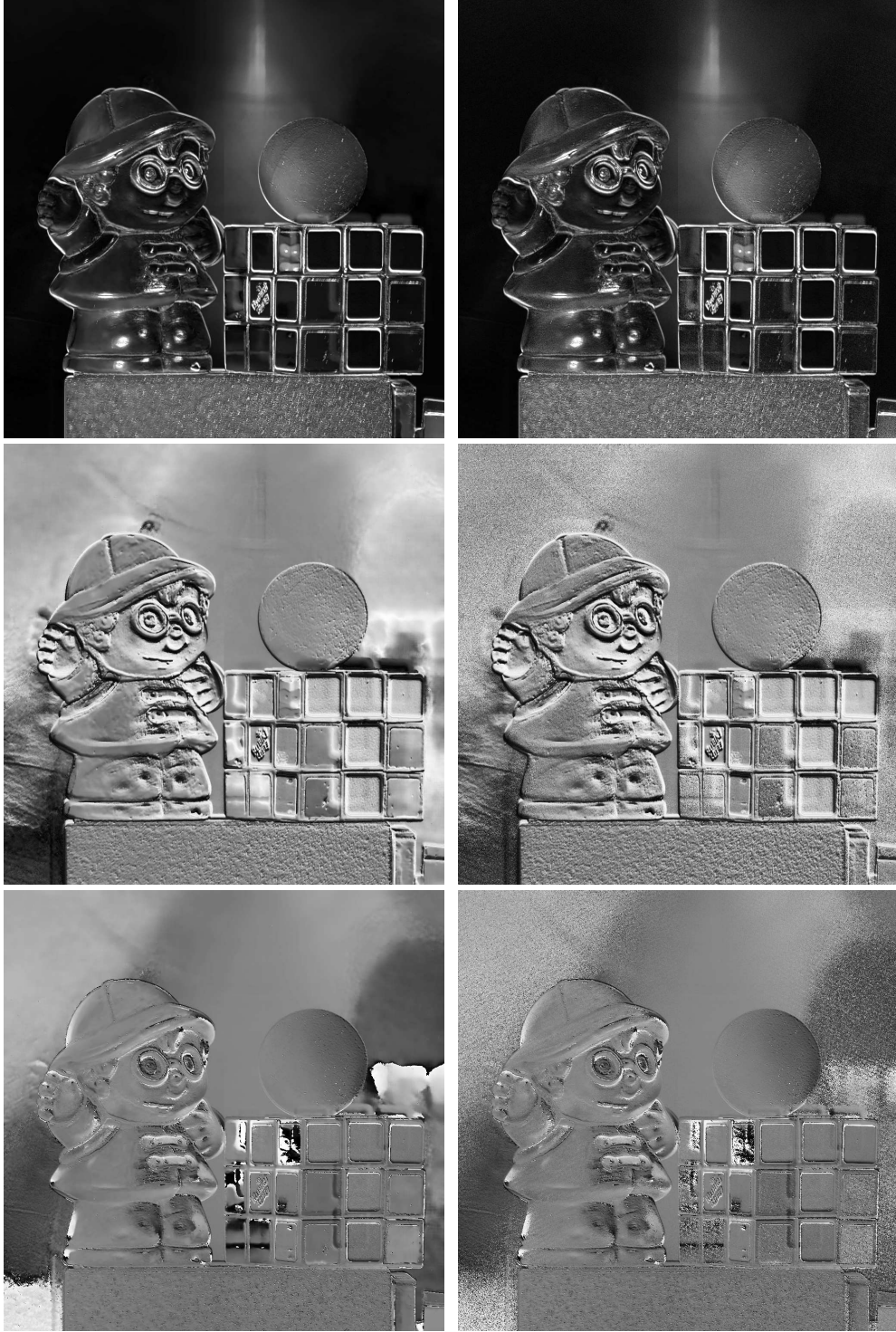


Fig. 4. Degree of polarization (first line), ellipticity (second line) and orientation (third line) of the Stokes vectors estimated with the proposed approach (left), and by using the pseudo-inverse (right).

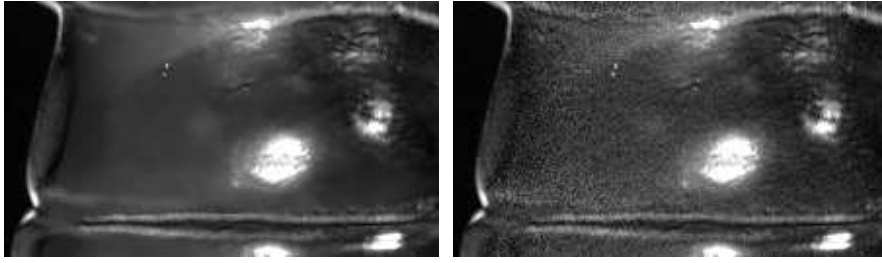


Fig. 5. Degree of polarization of the Stokes vectors estimated with the proposed approach (left) and by using the pseudo-inverse (right). The zoom in on the foot of the figurine reveals the denoising efficiency of the proposed approach.