

## A Blind Source Separation Technique for Document Restoration\*

Antonio Boccuto<sup>†</sup>, Ivan Gerace<sup>†</sup>, and Valentina Giorgetti<sup>†‡</sup>

**Abstract.** We examine an instance of the blind source separation problem, namely, the reconstruction of digital documents degraded by bleed-through and show-through effects. We introduce a nonstationary, locally linear data model and a solution approach based on the assumption of cross-correlated ideal sources. In order to solve the ill-posed local linear problem, we impose that the sum of all rows of the mixture matrix is equal to one, and we assume that the ideal sources are nonnegative and with an estimated level of overlapping (i.e., estimated cross-correlation). The solutions we obtain are related to a factorization of the covariance matrix of the data, which allows the given constraints to be satisfied at best. Our experimental results confirm the effectiveness of the method we propose.

**Key words.** blind source separation, digital document restoration, nonnegative minimization, correlated component analysis

**AMS subject classifications.** 65D18, 68U10

**DOI.** 10.1137/18M1188793

**1. Introduction.** In this paper we consider a *blind source separation* (BSS) problem. This problem has been an active research topic in signal processing since the end of the last century and has several applications in different fields. Examples arise in the cocktail party problem, in which individual speech signals are extracted from a number of observed signals, each representing a different mixture of them; in image classification and change detection (see also [4]); in determining the structure of the buildings in thermographic images for seismic engineering (see also [12, 13]); and in the problem of estimating the *cosmic microwave background* from galactic and extragalactic emissions (see also [18, 32]).

In particular, we study the digital reconstruction of *degraded written documents*. Several causes can lead to their deterioration: weathering, seeping of ink, humidity, powder, mold, and light transmission can all cause the progressive degradation and decay of the paper and the ink of written records.

\*Received by the editors May 21, 2018; accepted for publication (in revised form) March 29, 2019; published electronically June 20, 2019.

<http://www.siam.org/journals/siims/12-2/M118879.html>

**Funding:** The work of the authors was partially supported by the University of Perugia, the G.N.A.M.P.A. (Italian National Group of Mathematical Analysis, Probability and Applications), and the G.N.C.S. (Italian National Group of Scientific Computing), and by the projects “Ricerca di Base 2017” (“Metodi di Teoria dell’Approssimazione e di Analisi Reale per problemi di approssimazione ed applicazioni”) and “Ricerca di Base 2018” (“Metodi di Teoria dell’Approssimazione, Analisi Reale, Analisi Nonlineare e loro applicazioni”).

<sup>†</sup>Laboratorio di Matematica Computazionale “Sauro Tulipani,” Dipartimento di Matematica e Informatica, Università degli Studi di Perugia, Via L. Vanvitelli, 1, I-06123 Perugia, Italy ([antonio.boccuto@unipg.it](mailto:antonio.boccuto@unipg.it), [ivan.gerace@unipg.it](mailto:ivan.gerace@unipg.it)).

<sup>‡</sup>Dipartimento di Matematica e Informatica “Ulisse Dini,” viale G. B. Morgagni, 67/a, I-50134 Firenze, Italy ([valentina.giorgetti@unifi.it](mailto:valentina.giorgetti@unifi.it)).

Here we deal with *show-through* and *bleed-through* effects. *Show-through* is a front-to-back interference, mainly due to the scanning process and the transparency of the paper, which causes the text in the verso side of the document to appear also in the recto side (and vice versa). *Bleed-through* is an intrinsic front-to-back physical deterioration due to ink seeping and produces an effect similar to that of show-through. The physical model for the show-through distortion, an effect due to the scanning process, is very complex and takes into account the features of the paper, the transmittance parameters, the reflectance of the verso, and the spreading of light in the paper. In [43], a simplified, yet still complex, mathematical model was first constructed and then further approximated so to become tractable. This model describes the observed recto and verso images as mixtures of the two uncorrupted texts (the ideal images, or sources). Unfortunately, the situation is typically worse when bleed-through is considered, as the latter is determined by chemical transformation of the material and ink seeping through the paper fiber (see also [24, 39, 51, 53]). Moreover, the bleed-through pattern is, in ancient documents, often stronger than show-through. Furthermore, this process is in general nonstationary, due to unpredictable causes such as inhomogeneity of the paper, accidental humidity in some parts of the sheet, or higher pressure in the writing process on the verso. In view of all the considerations above, it is not easy to formulate a general and comprehensive mathematical model for generic recto-to-verso interference in ancient documents.

In this paper, in order to obtain an accurate algorithm with low computational cost, we propose a nonstationary and locally linear approximation to the physical model. We divide the domain of an image into disjoint subdomains and assume that the model is linear on each of them and varies smoothly in the global domain.

We consider a classical linear and stationary recto-verso model (see also [10, 25, 26, 27, 48]) developed for this purpose and are concerned with the problem of estimating both the ideal source images of the recto and the verso of the document and the mixture matrix producing the bleed-through or show-through effects. This problem is ill-posed in the sense of Hadamard (see also [22]). In fact, as the estimated mixture matrix varies, the corresponding estimated sources are in general different, and thus infinitely many solutions exist. Many techniques to solve this ill-posed inverse problem have been proposed in the literature. Among them, the *independent component analysis* (ICA) methods are based on the assumption that the sources are mutually independent (see also [14]). The best-known ICA technique is the so-called FastICA (see also [25, 26, 27, 30, 33]), which by means of a fixed point iteration finds an orthogonal rotation of the prewhitened data that maximizes a measure of non-Gaussianity of the rotated components. The FastICA algorithm is a parameter-free and extremely fast procedure, but ICA is not a viable approach in our setting, as for the problem we consider there is a clear correlation among the sources. On the other hand, several techniques for ill-posed inverse problems require that the estimated sources are only mutually uncorrelated. In this case, the estimated sources are determined via a linear transformation of the data, which is obtained by imposing either an orthogonality condition, as in *principal component analysis* (see also [10, 47, 48]), or an orthonormality condition, as in *whitening* and *symmetric whitening (SW)* techniques (see also [10, 47, 48]). These approaches all require only a single and very fast processing step. In [10, 48] it is observed that the results obtained by means of

the SW method are substantially equivalent to those produced by an ICA technique in the symmetric mixing case.

Here we assume that the sum of all rows of the mixture matrix is equal to one, since we expect the color of the background of the source to be the same as that of the data. In our setting, we change the variables of the data so that high and low light intensities correspond to presence and absence of text in the document, respectively, and we impose a nonnegativity constraint on the estimated sources (see also [9, 11, 21, 37]). We define the *overlapping matrix* of both the observed data and the ideal sources, a quantity related to the cross-correlation between the signals. From the overlapping matrix we can deduce the *overlapping level*, which measures the similarity between the front and the back of the document.

In order to obtain an accurate estimate of the sources, it is necessary to determine a correct source overlapping level. To this aim, we propose the following iterative procedure. At each iteration, given the current source overlapping level, we estimate the mixture matrix that produces the sources with the lowest possible source overlapping level among those having light intensity in the desired range. This mixture matrix is computed by means of a suitable symmetric factorization of the data overlapping matrix. We then use the estimated sources to update the source overlapping level and iterate the procedure until a fixed point is reached. At the fixed point, the corresponding source overlapping level is the smallest one that allows us to estimate the ideal recto and verso sides with the desired properties. We consider this level an adequate estimate of the ideal source overlapping level. Thus, by means of this technique, we can estimate not only the ideal sources and the mixture matrix but also the source overlapping level, a value that indicates the correlation between the ideal sources. Therefore, our method can be classified as a *correlated component analysis* (CCA) technique (see also [3, 38, 44, 45]). We refer to this method as the *minimum amount of text overlapping in document separation* (MATODS) algorithm. Similarly to the FastICA technique, the MATODS algorithm is a parameter-free and extremely fast procedure. We use the MATODS algorithm to solve the nonstationary and locally linear model we propose, and in particular we present an extension of this technique that fits this model, which we call the *not invariant for translation MATODS* (NIT-MATODS) algorithm.

In section 2 we present the features of the physical problem and in section 3 we describe some of the approximated mathematical models proposed in the literature. In particular, in subsection 3.1 we present the nonstationary and locally linear model we consider. In section 4 we analyze the nature of this new model and formulate some constraints to reduce its ill-posedness. In section 5 we develop the MATODS algorithm to deal with the local linear problem, and in section 6 we discuss the NIT-MATODS algorithm for working with the new model. In section 7 we compare experimentally the MATODS algorithm with other fast and unsupervised methods existing in the literature and show how the NIT-MATODS algorithm performs in restoring real ancient documents.

**2. The physical problem.** Bleed-through is a complex physical phenomenon that involves several parameters, such as the properties of the paper, the distribution of the paper fibers, and the quality and thickness of the ink. From a physical point of view, the bleed-through phenomenon is a diffusion process of ink through paper [24, 53]. Many similar phenomena,

like the seepage of water or oil through soil, are usually described by diffusion models [39, 51]. In general, to consider a model at microscale is computationally unfeasible, because of the sheer number of variables that are involved; thus one typically considers an equivalent model at macroscopic scale, which describes with a good approximation the average behavior of the microscale phenomenon in question.

Various authors have proposed different mathematical models to approximate the physical phenomena of bleed-through and show-through when the images of both the recto and the verso side of the document are available. The variational model proposed in [17] works with an estimated background, that is, the gray level of unprinted/unwritten paper. An anisotropic diffusion model is given in [16], and an invertible nonlinear model that considers the halftoning process of printers is considered in [1]. Other methods use just a single-side observation and reduce the problem to a segmentation one (see also [52]). In [43], a physical analysis of the show-through effect produced by a scanner in a digital image of a document is performed, and a model which takes into account the reflection, transmission, and scattering parameters of the paper is developed. Although extremely simplified with respect to the physical phenomenon, this model is still quite complex, so much so that some approximations are needed to make it tractable. A generalized version of this model for bleed-through removal is discussed in [34, 41]. When the character of the side under examination is sufficiently dark, it does not change, independently of the degradation coming from the opposite side.

**3. Approximated mathematical models.** We represent a gray level image as a vector belonging to  $\mathbb{R}^{n^2}$ , whose elements are the light intensity (which varies between 0 and 255) of the pixels, taken in lexicographic order. We consider a document as a pair of images that represent its sides, the front (*recto*) and the back (*verso*). In particular, we denote by  $\hat{x}_r \in [0, 255]^{n^2}$  the front image of the observed document and by  $\hat{x}_v \in [0, 255]^{n^2}$  the associated back image. Here, we assume that the recto data  $\hat{x}_r$  and the verso data  $\hat{x}_v$  are already spatially registered, that is, the pixel positions of the recto and of the verso of the document correspond, if we do a horizontal flip of the verso. However, the problem of registration of documents is an open and challenging issue (see, for instance, [5, 15, 23, 42, 54]). We denote the observed document by  $\hat{x} = [\hat{x}_r \ \hat{x}_v] \in [0, 255]^{n^2 \times 2}$  and the source ideal document by  $\hat{s} = [\hat{s}_r \ \hat{s}_v] \in [0, 255]^{n^2 \times 2}$ . The blind separation problem in document restoration amounts to estimating the ideal document from the observed document, without knowing the parameters underlying the back-to-front and front-to-back interference related to the model.

The nonlinear model proposed in [34, 35, 36, 41] is

$$(3.1) \quad \begin{aligned} \hat{x}_r(i) &= \hat{s}_r(i) e^{q_v(1-\hat{s}_v(i))}, \\ \hat{x}_v(i) &= \hat{s}_v(i) e^{q_r(1-\hat{s}_r(i))}, \quad i = 1, \dots, n^2, \end{aligned}$$

where  $q_v, q_b \in \mathbb{R}^+$  are the *interference levels* that affect the intensity values from the recto to the verso and vice versa, respectively. Some nonlinear models that assume that the interference levels depend on the location are proposed in [20, 29, 49]. This assumption makes the model nonstationary, that is, not translation invariant. In particular the nonlinear model proposed in [49] is



**Figure 1.** Detail of the document in Figure 10 with a horizontally flipped verso.

$$(3.2) \quad \begin{aligned} \hat{x}_r(i) &= \hat{s}_r(i) \left( \frac{\hat{s}_v(i))}{255} \right)^{q_v(i)}, \\ \hat{x}_v(i) &= \hat{s}_v(i) \left( \frac{\hat{s}_r(i))}{255} \right)^{q_r(i)}, \quad i = 1, \dots, n^2. \end{aligned}$$

The hypothesis of nonstationarity is significant, as in real ancient documents the level of interference varies highly from pixel to pixel. Figure 1 shows a detail of an ancient document, where the verso has been horizontally mirrored for the reader's convenience. From this figure it is evident that an ink infiltration law independent of the position cannot be determined in general. The algorithms proposed in [20, 49] for the resolution of the related inverse problem are fast heuristics. In order to obtain more precise results, a computationally more expensive regularized problem should be investigated (see also [19, 47]).

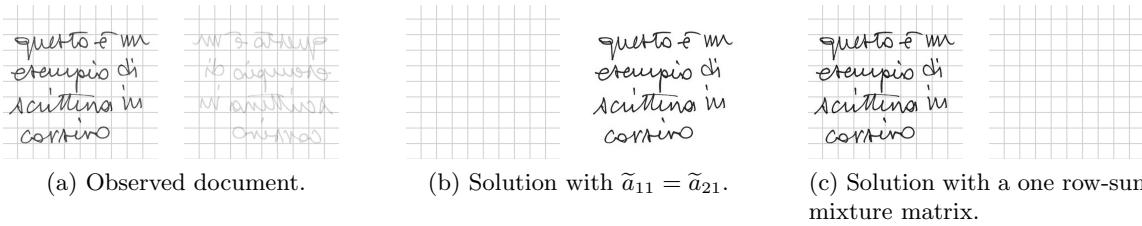
**3.1. A nonstationary locally linear model.** In this paper, in order to obtain an accurate algorithm with low computational cost, we propose a new **nonstationary** and **locally linear** model. Namely, we partition the domain of the document, a set of pixels of size  $n \times n$ , into  $(n/\nu)^2$  disjoint subdomains of size  $\nu \times \nu$ . On each subdomain we approximate the problem by means of the classical linear model (see also [10, 25, 26, 27, 46, 48])

$$(3.3) \quad \hat{x}^T = A \hat{s}^T,$$

where the symbol  $\cdot^T$  denotes the transpose operator of a matrix,  $\hat{x} \in [0, 255]^{\nu^2}$  is the observed document in the involved subdomain,  $\hat{s} \in [0, 255]^{\nu^2}$  is the ideal document, and  $A \in \mathbb{R}^{2 \times 2}$  is called the *mixture matrix*. We assume that the entries of the matrix  $A$  vary smoothly with respect to the corresponding entries in the mixture matrices in the adjacent subdomains. The **size** of the subdomains should be chosen taking into account both the accuracy of the model (for small subdomain dimensions) and the computational cost for its resolution (for large dimensions).

In the next sections we focus on the linear problem related to (3.3), while in section 6 we use the results obtained for the resolution of the linear problem to solve the nonstationary model proposed here.

**4. Analysis of the linear problem.** In this section we discuss the problem of estimating both the ideal sources and the mixture matrix from the observed data using the linear equation (3.3), which is a BSS problem (see also [10, 47]). If we have an invertible estimate  $\tilde{A}$  of  $A$ , then an estimate of  $s$  is



**Figure 2.** One row-sum mixture matrix reconstruction.

$$(4.1) \quad \tilde{s}^T = \tilde{A}^{-1} \tilde{x}^T.$$

Since there are infinitely many choices of  $\tilde{A}$ , our problem admits infinitely many solutions and is ill-posed in the sense of Hadamard. Even if we assume that  $\tilde{A}$  and  $\tilde{s}$  are nonnegative matrices, the problem is NP-hard (see [50]) and ill-posed (see [22]). To overcome this, it is necessary to impose some constraints on the solutions.

Since the color of the paper is the same for each part of the document, we assume that the value of the source background, that is, the gray level of unprinted/unwritten paper, is the same as the background of the data. This value corresponds to the light intensity of the paper on which the document is written. In order to satisfy this requirement, we assume that  $A$  is a *one row-sum matrix*, that is,

$$(4.2) \quad a_{11} + a_{12} = a_{21} + a_{22} = 1.$$

In [6] we prove that if  $(\tilde{A}, \tilde{s})$  is a solution to the linear model in (3.3), with  $\tilde{A}$  nonsingular and such that  $\tilde{a}_{22} \neq \tilde{a}_{12}$  and  $\tilde{a}_{11} \neq \tilde{a}_{21}$ , then there exist  $t_1, t_2 \neq 0$  and a one row-sum matrix  $\bar{A}$ , such that  $(\bar{A}, \bar{s})$  is a solution of (3.3), with  $\bar{s} = [t_1 \tilde{s}_r \ t_2 \tilde{s}_v]$ . In other words, for a given estimation of the solution, it is sufficient to multiply the estimated sources by some given nonzero parameters in order to obtain a solution with a one row-sum estimated mixture matrix. In [6] we prove that if  $(\tilde{A}, \tilde{s})$  is a solution to the linear model in (3.3), and  $\tilde{A}$  is singular, then there exists  $t \in \mathbb{R}$  such that  $x_v = tx_r$ . In subsection 5.2 we show how to find a more realistic solution with a one row-sum mixture matrix in this case. When  $(\tilde{A}, \tilde{s})$  is a solution with  $\tilde{a}_{11} = \tilde{a}_{21}$  or  $\tilde{a}_{12} = \tilde{a}_{22}$ , in [6] we show how to calculate a more realistic solution with a one row-sum mixture matrix. Indeed, if  $\tilde{a}_{11} = \tilde{a}_{21}$  or  $\tilde{a}_{12} = \tilde{a}_{22}$ , then one of the estimated sources usually corresponds to the common background of the recto and the verso, which is a pattern that is equally present on the two sides of the document. An example of this case is shown in Figure 2. Therefore, requiring that the mixture matrix is one row-sum is not a restriction.

A remarkable feature of our approach is that a high light intensity indicates the presence of meaningful information (for example, a letter or a figure), whereas a low light intensity corresponds to the absence thereof. Since the background is usually lighter in color while text or figures are darker, we apply the change of variables

$$(4.3) \quad x = m E - \hat{x}, \quad s = m E - \hat{s},$$

where  $E \in \mathbb{R}^{n^2 \times 2}$  is the matrix such that

$$(4.4) \quad e_{i,j} = 1 \text{ for each } i = 1, 2, \dots, n^2 \text{ and } j = 1, 2,$$

and  $m$  is the maximum between the light intensity on the two sides of the document. Note that, since we deal with paper documents, we assume that this maximum is achieved on the background. In view of (4.3), the values of the light intensity corresponding to the background are equal to 0, while the pixels containing information have positive light intensity values no greater than  $m$ . Motivated by the physical interpretation of these values, we impose that the estimated sources have to satisfy this property, as the values zero and  $m$  correspond to the background color and the black color, respectively. Since  $A$  is a one row-sum matrix, we get

$$(4.5) \quad E^T = A E^T,$$

and hence from (3.3) and (4.5) we obtain

$$(4.6) \quad x^T = m E^T - A \tilde{s}^T = m A E^T - A \tilde{s}^T = A (m E^T - \tilde{s}^T) = A s^T.$$

Here we define the following  $2 \times 2$  *data overlapping matrix* of the observed data:

$$(4.7) \quad C = \begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{bmatrix} = x^T x = \begin{bmatrix} x_r^T \cdot x_r & x_r^T \cdot x_v \\ x_v^T \cdot x_r & x_v^T \cdot x_v \end{bmatrix}.$$

This matrix, when  $x_r$  and  $x_v$  have zero mean, corresponds to the data covariance matrix. The matrix  $C$  tells how much the text on the front overlaps with that on the back. Indeed in our case, since  $x$  is nonnegative, the data overlapping matrix is always nonnegative, and it is diagonal if and only if there is no overlapping text from the recto to the verso of the document. In particular we refer to the entries  $d = c_{12} = c_{21}$  as the *data overlapping level*.

The *source overlapping matrix* can be defined similarly as

$$P = \begin{bmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{bmatrix} = s^T s = \begin{bmatrix} s_r^T \cdot s_r & s_r^T \cdot s_v \\ s_v^T \cdot s_r & s_v^T \cdot s_v \end{bmatrix}.$$

It is easy to see that the matrices  $C$  and  $P$  are symmetric and positive semidefinite. We refer to the value

$$(4.8) \quad k = p_{12} = p_{21} = s_r^T \cdot s_v$$

as the *source overlapping level*. Since we assume that the text of the recto of the document partially overlaps with that on the verso, the estimation of the level  $k$  plays an important role in the design of the technique we propose. In fact, we claim that a correct estimation of  $k$  leads to more accurate estimates of the original sources.

**5. A new technique for solving the linear problem.** We consider the two cases of singular and nonsingular data overlapping matrices separately. We treat the latter now, while the former will be dealt with in subsection 5.2.

We would like to estimate not only the ideal sources  $s_r$  and  $s_v$  and the mixture matrix  $A$ , but also the source overlapping level  $k$ . Since in our algorithm we impose a nonnegativity constraint on the estimated sources  $\tilde{s}_r$  and  $\tilde{s}_v$ , the corresponding value of  $k$  represents the level of overlapping of the recto of the source document with its verso or, equivalently, the portion of text of the estimated front source that is disjoint from that of the estimated back source. The value of  $k$  is different from zero, in general, thus the method we propose can be classified as a CCA technique (see also [3, 38, 44, 45]).

We define a *symmetric factorization* of a symmetric and positive definite matrix  $H \in \mathbb{R}^{n \times n}$  as an identity of the type  $H = ZZ^T$ , where  $Z \in \mathbb{R}^{n \times n}$ . Observe that, given an orthogonal matrix  $Q \in \mathbb{R}^{n \times n}$  and a symmetric factorization of the type  $H = ZZ^T$ , then  $ZQ(ZQ)^T$  is also a symmetric factorization of  $H$ . Moreover, if we consider any two symmetric factorizations  $H = Z_1Z_1^T$  and  $H = Z_2Z_2^T$ , then there is an orthogonal matrix  $Q \in \mathbb{R}^{n \times n}$  such that  $Z_1 = Z_2Q$  (see also [6]).

In the  $2 \times 2$  case, the set of the orthogonal matrices is the union of all rotations and reflections in  $\mathbb{R}^2$ , which are expressed as

$$(5.1) \quad Q^1(\theta) = \begin{bmatrix} \sin \theta & -\cos \theta \\ \cos \theta & \sin \theta \end{bmatrix} \quad \text{and} \quad Q^{-1}(\theta) = \begin{bmatrix} \sin \theta & \cos \theta \\ \cos \theta & -\sin \theta \end{bmatrix},$$

respectively, as  $\theta$  varies in  $]0, 2\pi]$ . As  $C = C^{1/2}(C^{1/2})^T = C^{1/2}C^{1/2}$  is a symmetric factorization of  $C$ , then all possible factorizations of  $C$  are given by

$$(5.2) \quad Z^{(\iota)}(\theta) = C^{1/2}Q^{(\iota)}(\theta) = \begin{bmatrix} \rho_{11} & \rho_{12} \\ \rho_{21} & \rho_{22} \end{bmatrix} Q^{(\iota)}(\theta) = \begin{bmatrix} z_{11}^{(\iota)}(\theta) & z_{12}^{(\iota)}(\theta) \\ z_{21}^{(\iota)}(\theta) & z_{22}^{(\iota)}(\theta) \end{bmatrix},$$

where  $\theta \in ]0, 2\pi]$  and  $\iota \in \{-1, 1\}$ . In particular, we have that

$$(5.3) \quad z_{11}^{(1)}(\theta) = z_{11}^{(-1)}(\theta), \quad z_{12}^{(1)}(\theta) = -z_{12}^{(-1)}(\theta), \quad z_{21}^{(1)}(\theta) = z_{21}^{(-1)}(\theta), \quad z_{22}^{(1)}(\theta) = -z_{22}^{(-1)}(\theta).$$

In order to obtain a joint estimation of the mixture matrix, the source matrices, and the source overlapping level, we use an iterative algorithm. At the  $l$ th step we assume that

$$(5.4) \quad C = x^T x = A s^T s A^T = A \tilde{P} A^T,$$

where  $\tilde{P}$  is a symmetric and positive definite estimate of the source overlapping matrix  $P$ . In  $\tilde{P}$  we set

$$(5.5) \quad \tilde{p}_{12} = \tilde{p}_{21} = k^{(l)},$$

where  $k^{(l)}$  is the estimate of the source overlapping level obtained at the  $(l-1)$ th step (we assume that  $k^{(0)} = 0$ ). Note that for the moment we do not assign a value to  $\tilde{p}_{11}$  and  $\tilde{p}_{22}$ , as they will be determined later by imposing that the estimated mixture matrix is one row-sum. Let

$$(5.6) \quad \tilde{P} = YY^T$$

be a symmetric factorization, where  $Y$  is a nonsingular matrix that by (5.5) satisfies

$$(5.7) \quad y_{11} y_{21} + y_{12} y_{22} = k^{(l)}.$$

By virtue of (5.4) and (5.6), it holds that

$$C = AYY^T A^T = AY(AY)^T,$$

that is,  $AY$  realizes a factorization of  $C$ . For any given choice of  $\theta \in ]0, 2\pi]$  and  $\iota \in \{-1, 1\}$ , we define an estimate  $\tilde{A}^{(\iota)}(\theta)$  of the mixture matrix  $A$  as a matrix such that  $\tilde{A}^{(\iota)}(\theta) = Z^{(\iota)}(\theta)Y^{-1}$ , where  $Z^{(\iota)}(\theta)$  is as in (5.2). We get that

$$(5.8) \quad \begin{aligned} a_{11}^{(\iota)}(\theta) &= \frac{z_{11}^{(\iota)}(\theta)y_{22} - z_{12}^{(\iota)}(\theta)y_{21}}{y_{11}y_{22} - y_{21}y_{12}}, & a_{12}^{(\iota)}(\theta) &= \frac{z_{12}^{(\iota)}(\theta)y_{11} - z_{11}^{(\iota)}(\theta)y_{12}}{y_{11}y_{22} - y_{21}y_{12}}, \\ a_{21}^{(\iota)}(\theta) &= \frac{z_{21}^{(\iota)}(\theta)y_{22} - z_{22}^{(\iota)}(\theta)y_{21}}{y_{11}y_{22} - y_{21}y_{12}}, & a_{22}^{(\iota)}(\theta) &= \frac{z_{22}^{(\iota)}(\theta)y_{11} - z_{21}^{(\iota)}(\theta)y_{12}}{y_{11}y_{22} - y_{21}y_{12}}, \end{aligned}$$

and by imposing that  $\tilde{A}^{(\iota)}(\theta)$  satisfies the one row-sum condition in (4.2), we have that

$$(5.9) \quad \begin{aligned} z_{11}^{(\iota)}(\theta)y_{22} - z_{12}^{(\iota)}(\theta)y_{21} + z_{12}^{(\iota)}(\theta)y_{11} - z_{11}^{(\iota)}(\theta)y_{12} &= y_{11}y_{22} - y_{21}y_{12}, \\ z_{21}^{(\iota)}(\theta)y_{22} - z_{22}^{(\iota)}(\theta)y_{21} + z_{22}^{(\iota)}(\theta)y_{11} - z_{21}^{(\iota)}(\theta)y_{12} &= y_{11}y_{22} - y_{21}y_{12}. \end{aligned}$$

Thus, the matrix  $Y$  has to satisfy the conditions in (5.7) and (5.9). The nonlinear system given by (5.7) and (5.9) has infinitely many solutions. For the sake of convenience, we choose the solution

$$(5.10) \quad \begin{aligned} y_{11} &= \frac{\det C - k^{(l)}(z_{11}^{(\iota)}(\theta) - z_{21}^{(\iota)}(\theta))^2}{(z_{22}^{(\iota)}(\theta) - z_{12}^{(\iota)}(\theta))\det Z^{(\iota)}(\theta)}, & y_{12} &= k^{(l)} \frac{z_{11}^{(\iota)}(\theta) - z_{21}^{(\iota)}(\theta)}{\det Z^{(\iota)}(\theta)}, \\ y_{21} &= 0, & y_{22} &= \frac{\det Z^{(\iota)}(\theta)}{z_{11}^{(\iota)}(\theta) - z_{21}^{(\iota)}(\theta)}. \end{aligned}$$

This choice has several desirable consequences. First, from (5.3) and (5.8) we get that  $\tilde{A}^{(1)}(\theta) = \tilde{A}^{(-1)}(\theta)$  for all  $\theta \in ]0, 2\pi]$ . Moreover, from (5.1) and (5.2) we deduce that  $Z(\theta) = -Z(\theta + \pi)$  for  $\theta \in ]0, \pi]$ ; thus from (5.8) and (5.10) we can conclude that

$$(5.11) \quad \tilde{A}(\theta) = \tilde{A}(\theta + \pi)$$

for all  $\theta \in ]0, \pi]$ .

Therefore, in the reminder we consider only the case  $\iota = 1$ , pose  $\tilde{A}(\theta) = \tilde{A}^{(1)}(\theta)$  and  $Z(\theta) = Z^{(1)}(\theta)$  for all  $\theta \in ]0, \pi]$ , and in general consider only the values of  $\theta$  belonging to the interval  $]0, \pi]$ .

Recall that  $Y$  must be nonsingular, as  $Y$  realizes a symmetric factorization of the nonsingular matrix  $P$ . In [6] we show that if

$$(5.12) \quad k^{(l)} < k_{sup} = \frac{\det C}{(\rho_{11} - \rho_{21})^2 + (\rho_{12} - \rho_{22})^2},$$

where  $\rho_{i,j}$ , for  $i, j = 1, 2$ , are the entries of the matrix  $C^{1/2}$  in (5.2), then  $Y$  is nonsingular for all  $\theta \in ]0, \pi]$ . We refer to  $k_{sup}$  in (5.12) as the *source overlapping level upper bound*.

Moreover, the equations in (5.10) are well defined if  $z_{11}(\theta) \neq z_{21}(\theta)$  and  $z_{12}(\theta) \neq z_{22}(\theta)$ . In [6] we show that  $z_{11}(\theta) = z_{21}(\theta)$  or  $z_{12}(\theta) = z_{22}(\theta)$  when  $\theta$  assumes the values  $\varphi + t\frac{\pi}{2}$ , with  $t \in \mathbb{Z}$  and

$$(5.13) \quad \varphi = \begin{cases} \arctan \left( \frac{\rho_{22} - \rho_{12}}{\rho_{11} - \rho_{21}} \right) & \text{if } \rho_{11} \neq \rho_{21}, \\ \frac{\pi}{2} & \text{if } \rho_{11} = \rho_{21}. \end{cases}$$

In subsection 5.1, in formulating the minimization algorithm, we show how to avoid these values.

For any  $\theta \in [\varphi, \varphi + \frac{\pi}{2}[\cup]\varphi + \frac{\pi}{2}, \varphi + \pi[$ , from (4.6) we deduce that an estimate of the ideal sources  $s$  is given by

$$(5.14) \quad \tilde{s}(\theta)^T = [\tilde{s}_r(\theta) \quad \tilde{s}_v(\theta)]^T = \tilde{A}^{-1}(\theta)x^T,$$

which combined with the fact that  $\tilde{A}^{-1}(\theta) = \tilde{A}^1(\theta) = Z^{(1)}(\theta)Y^{-1}$  and (5.9) gives

$$(5.15) \quad \begin{aligned} \tilde{s}_r(\theta) &= \left( z_{22}(\theta) \frac{\det C - k^{(l)}(z_{11}(\theta) - z_{21}(\theta))^2}{(z_{22}(\theta) - z_{12}(\theta)) \det C} - z_{21}(\theta) \frac{k^{(l)}(z_{11}(\theta) - z_{21}(\theta))}{\det C} \right) x_r \\ &\quad + \left( -z_{12}(\theta) \frac{\det C - k^{(l)}(z_{11}(\theta) - z_{21}(\theta))^2}{(z_{22}(\theta) - z_{12}(\theta)) \det C} + z_{11}(\theta) \frac{k^{(l)}(z_{11}(\theta) - z_{21}(\theta))}{\det C} \right) x_v, \\ \tilde{s}_v(\theta) &= -\frac{z_{21}(\theta)}{z_{11}(\theta) - z_{21}(\theta)} x_r + \frac{z_{11}(\theta)}{z_{11}(\theta) - z_{21}(\theta)} x_v. \end{aligned}$$

As we assumed that our estimated sources have intensity values between 0 and  $m$ , we take the orthogonal projection of the estimate  $s_i^{(l)}(\theta)$  on the space  $[0, m]^{\nu^2 \times 2}$  with respect to the **Frobenius norm**. Namely, we apply to the estimate of the sources the function that maps a vector  $s \in \mathbb{R}^{\nu^2}$  to the  $\nu^2$ -dimensional vector  $\tau(s)$ , whose elements are

$$(5.16) \quad (\tau(s))_i = \begin{cases} 0 & \text{if } s_i \leq 0, \\ s_i & \text{if } 0 < s_i \leq m, \\ m & \text{if } s_i > m, \end{cases} \quad i = 1, \dots, \nu^2.$$

By this transformation, the projections of the estimated source images  $\tau(\tilde{s}_{r,i}^{(l)}(\theta))$  and  $\tau(\tilde{s}_{v,i}^{(l)}(\theta))$  are guaranteed to be nonnegative (see also [9, 11, 21, 37]). From now on, we consider the projections above as the new source estimates. Thus, the estimated source overlapping level is a nonnegative value, and it is zero if and only if there is no overlapping text from the recto to the verso of the estimated source document. Hence, among the possible values of  $\theta$  in  $[\varphi, \varphi + \frac{\pi}{2}[\cup]\varphi + \frac{\pi}{2}, \varphi + \pi[$ , we find a value  $\tilde{\theta}$  that minimizes the *objective function*

$$(5.17) \quad g(k^{(l)}, \theta, C) = \tau(\tilde{s}_r(\theta))^T \cdot \tau(\tilde{s}_v(\theta)).$$

Note that, from (5.11) and (5.14), it follows that the function  $g$  is periodic in the variable  $\theta$  with period  $\pi$ . Then we set

$$(5.18) \quad k^{(l+1)} = g(k^{(l)}, \tilde{\theta}, C),$$

and we repeat this process until we find an index  $l$  such that  $k^{(l+1)} = k^{(l)}$ . In [6] we prove that if

$$(5.19) \quad \tau(\tilde{s}_r(\tilde{\theta})) = \tilde{s}_r(\tilde{\theta}) \quad \text{and} \quad \tau(\tilde{s}_v(\tilde{\theta})) = \tilde{s}_v(\tilde{\theta}),$$

then the condition (5.18) holds. In this case the estimated solution  $\tilde{s}(\tilde{\theta})$  belongs to the space  $[0, m]^{\nu^2 \times 2}$ , as required. We note that in all the experiments we performed, when a fixed point was reached the condition (5.19) was always satisfied.

The steps of the algorithm described in this section can be summarized as follows:

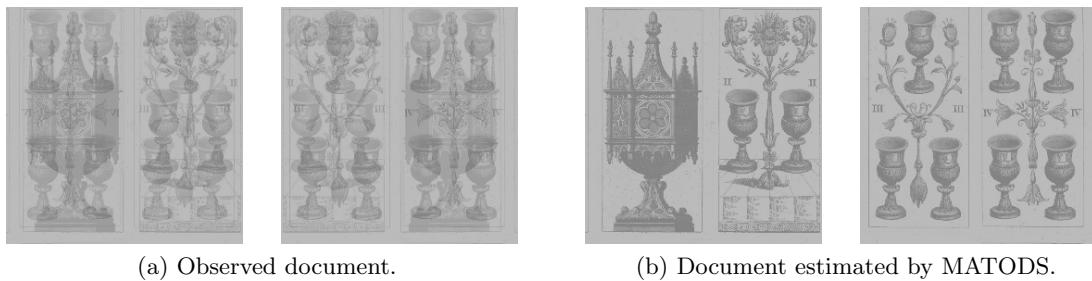
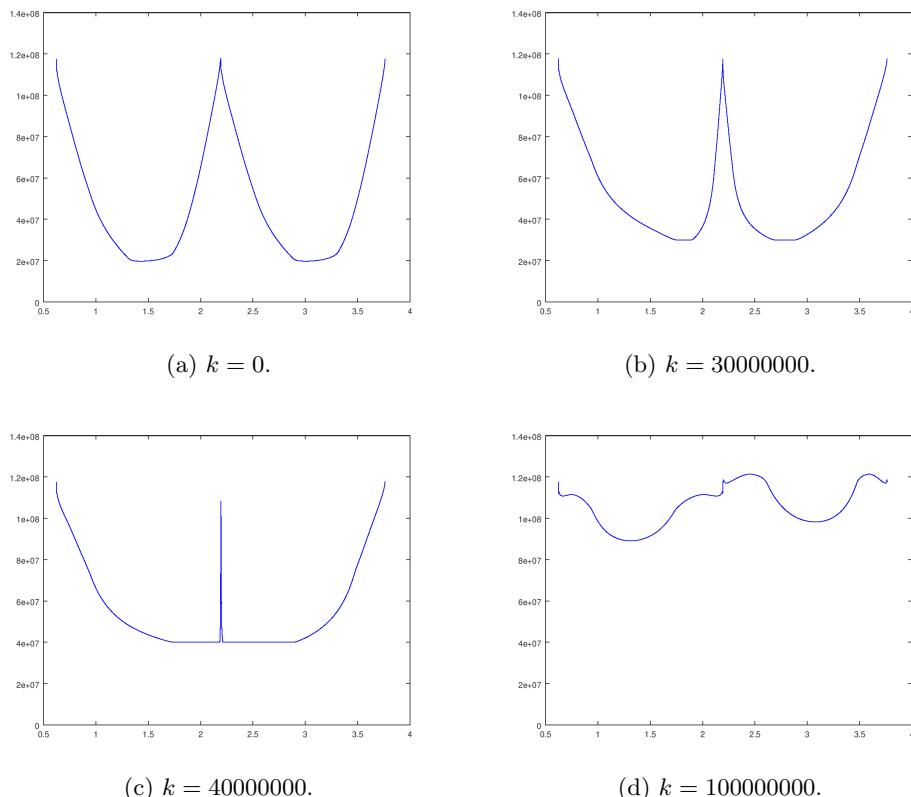
```
function MATODS( $\hat{x}$ )
  Determine the maximum value  $m$  of  $\hat{x}$ ;
   $x = mE - \hat{x}$ ;
   $C = x^T x$ ;
   $k^{(-1)} = -2\varepsilon$ ;
   $k^{(0)} = 0$ ;
   $l = 0$ ;
  while ( $|k^{(l)} - k^{(l-1)}| \geq \varepsilon$ ) do
     $\tilde{\theta} = \arg \min(\mathbf{function} g(k^{(l)}, \cdot, C))$ ;
     $k^{(l+1)} = g(k^{(l)}, \tilde{\theta}, C)$ ;
     $l = l + 1$ ;
  end while
   $Z(\tilde{\theta}) = C^{1/2}Q_1(\tilde{\theta})$ ;
  Compute  $\tilde{s}_r(\tilde{\theta})$  and  $\tilde{s}_v(\tilde{\theta})$  as in (5.15);
  return  $mE - \tau(\tilde{s}(\tilde{\theta}))$ 
```

Here  $\varepsilon$  is a fixed positive real number that represents a suitable tolerance threshold, while the function  $g(\cdot, \cdot, \cdot)$  is computed as follows:

```
function  $g(k, \theta, C)$ 
   $Z(\theta) = C^{1/2}Q_1(\theta)$ ;
  Compute  $\tilde{s}_r(\theta)$  and  $\tilde{s}_v(\theta)$  as in (5.15);
  return  $(\tau(\tilde{s}_r(\theta)))^T \cdot \tau(\tilde{s}_v(\theta))$ 
```

In the next subsection we describe the procedure we use to minimize the objective function  $g$  with respect to the variable  $\theta$ . This method is the MATODS algorithm, which is a parameter-free, and thus unsupervised, technique.

**5.1. The objective function minimization algorithm.** Consider the data document shown in Figure 3(a), for which the source overlapping level upper bound is  $k_{sup} = 40374184.63$  (see (5.12)) and the objective function  $g$  has discontinuities with respect to the variable  $\theta$  at the points  $\varphi + t\frac{\pi}{2}$ , with  $t \in \mathbb{Z}$  and  $\varphi = 0.62377$ . Figure 4 shows the graph of the function  $g(k, \theta, C)$  as  $\theta$  varies. In the plots, we use the overlapping matrix of the document in Figure 3(a) and test four values of  $k$ . The output of the MATODS algorithm is presented in Figure 3(b). The source overlapping level estimated by MATODS is 29670911.87, which is smaller than  $k_{sup}$ . In [6] we prove that if  $k = 0$ , then the objective function  $g$  is periodic of period  $\frac{\pi}{2}$  in the variable  $\theta$ , but this property is not verified for  $k > 0$ , as shown in Figures 4(b)–(d).

**Figure 3.** *MATODS restoration.***Figure 4.** *Graphs of the objective function  $g(k, \cdot, C)$ , where  $C$  is the overlapping matrix given by the document in Figure 3(a).*

Notice that when  $k$  is smaller than the source overlapping level upper bound  $k_{sup}$ , the objective function  $g$  is quasi-convex on each interval which lies between any two successive points of discontinuity. Moreover, on each interval where the objective function is quasi-convex, the local minima are almost identical. As noted in [6], this behavior is typical among objective functions obtained from the documents we considered in our experiments. In [6] we also show that during the execution of the MATODS algorithm, the estimated source overlapping levels  $k^{(l)}$  have increasing values and are always smaller than their source overlapping level upper

bound  $k_{sup}$ . Thus, to find the minimum of the objective function, it is sufficient to find a minimum on an interval that lies between any two successive points of discontinuity, where the objective function is quasi-convex.

In order to minimize  $g$  on an interval where it is quasi-convex, in [6] we consider different algorithms. Some of them are developed specifically for strictly quasi-convex functions and do not rely on derivatives (see also [8, 28, 31]), whereas others are based on the gradient descent and the *Armijo line search* (see also [2, 7]). The algorithm which gives the best performances is a hybrid *successive parabolic interpolation* and *golden section search* (SPI-GSS) (see also [8]).

This algorithm constructs a sequence  $\{\theta^{(h)}\}_{h \in \mathbb{N}}$  such that

$$(5.20) \quad g(k, \theta^{(h)}, C) \geq g(k, \theta^{(h+1)}, C), \quad h = 0, 1, \dots,$$

while at the  $h$ th step we have an uncertainty interval  $[a^{(h)}, b^{(h)}]$ , and we now discuss it in detail.

Let  $\phi = (\sqrt{5} + 1)/2$  be the *golden ratio* or *golden section*, let  $\varphi$  be as in (5.13), let  $\eta \in \mathbb{R}^+$  be small enough, and let  $[a^{(0)} = \varphi + \eta, b^{(0)} = \varphi + \frac{\pi}{2} - \eta]$  be the initial uncertainty interval. The sequence is initialized as

$$\theta^{(0)} = \theta^{(1)} = \theta^{(2)} = a^{(2)} + \frac{b^{(2)} - a^{(2)}}{\phi},$$

which is equivalent to a GSS step (see also [31]).

We will rely on the SPI algorithm (see also [28]), which extends a finite sequence of approximations of the required minimum by adding the minimum of the parabola that interpolates the objective function on the last three terms of that sequence. The main step of the SPI algorithm can be written as

$$\theta^{(h+3)} = \theta^{(h+2)} + \frac{p}{q},$$

where

$$\begin{aligned} p &= (\theta^{(h+2)} - \theta^{(h)})^2(g(k, \theta^{(h+2)}, C) - g(k, \theta^{(h+1)}, C)) \\ &\quad - (\theta^{(h+2)} - \theta^{(h+1)})^2(g(k, \theta^{(h+2)}, C) - g(k, \theta^{(h)}, C)) \end{aligned}$$

and

$$\begin{aligned} q &= 2(\theta^{(h+2)} - \theta^{(h)})(g(k, \theta^{(h+2)}, C) - g(k, \theta^{(h+1)}, C)) \\ &\quad - 2(\theta^{(h+2)} - \theta^{(h+1)})(g(k, \theta^{(h+2)}, C) - g(k, \theta^{(h)}, C)). \end{aligned}$$

If at any point any two of  $\theta^{(h)}$ ,  $\theta^{(h+1)}$ , and  $\theta^{(h+2)}$  coincide, or the parabola degenerates to a line (in which case  $q = 0$ ), or the SPI update is outside the current uncertainty interval  $[a^{(h)}, b^{(h)}]$ , then the step is performed using the GSS technique. The pseudocode of this algorithm is as follows:

```

function SPI-GSS( $k, C$ )
 $h = 0;$ 
 $[a^{(0)}, b^{(0)}] = [\varphi^{(1)} + \eta, \varphi^{(2)} + \pi - \eta];$ 
 $\theta^{(0)} = \theta^{(1)} = \theta^{(2)} = a^{(2)} + (b^{(2)} - a^{(2)})/\phi;$ 
while  $(|\theta^{(h+2)} - \theta^{(h+1)}| > \varepsilon)$  do
     $p = (\theta^{(h+2)} - \theta^{(h)})^2(g(k, \theta^{(h+2)}, C) - g(k, \theta^{(h+1)}, C));$ 
     $p = p - (\theta^{(h+2)} - \theta^{(h+1)})^2(g(k, \theta^{(h+2)}, C) - g(k, \theta^{(h)}, C));$ 
     $q = 2(\theta^{(h+2)} - \theta^{(h)})(g(k, \theta^{(h+2)}, C) - g(k, \theta^{(h+1)}, C));$ 
     $q = q - 2(\theta^{(h+2)} - \theta^{(h+1)})(g(k, \theta^{(h+2)}, C) - g(k, \theta^{(h)}, C));$ 
    if  $((q \neq 0) \text{ and } (\theta^{(h+2)} + p/q \in [a^{(h)}, b^{(h)}]))$  then
         $\theta^{(h+3)} = \theta^{(h+2)} + p/q;$ 
    else
        if  $(\theta^{(h+2)} < (a^{(h+2)} + b^{(h+2)})/2)$  then
             $\theta^{(h+3)} = \theta^{(h+2)} + (b^{(h+2)} - \theta^{(h+2)})/r;$ 
        else
             $\theta^{(h+3)} = \theta^{(h+2)} + (a^{(h+2)} - \theta^{(h+2)})/r;$ 
        end if
    end if
    Compute the new uncertainty interval  $[a^{(h+1)}, b^{(h+1)}]$ ;
    Order  $\{\theta^{(i)}\}_{i=h, \dots, h+3}$  in such a way that (5.20) holds;
     $h = h + 1;$ 
end while
return  $\theta^{(h+2)}$ 

```

Here  $\varepsilon > 0$  is a suitable tolerance threshold.

During the last iterations, the algorithm usually stops choosing the GSS steps and performs only parabolic interpolation steps. Thus, the asymptotic convergence depends only on the SPI algorithm. We recall that the sequence  $\{\theta^{(h)}\}_h$  converges to  $\hat{\theta}$  with *strong order*  $p$  and *asymptotic constant*  $\gamma > 0$  if

$$\lim_{h \rightarrow +\infty} \frac{|\theta^{(h+1)} - \hat{\theta}|}{|\theta^{(h)} - \hat{\theta}|^p} = \gamma$$

and with *weak order*  $p$  if

$$\liminf_{h \rightarrow +\infty} (-\ln |\theta^{(h)} - \hat{\theta}|^p)^{1/h} = p.$$

Note that strong convergence implies weak convergence, but in general the converse is not true.

Let  $n \in \mathbb{N}$ . A function  $f : [a, b] \rightarrow \mathbb{R}$  is *of class*  $LC^n([a, b])$  if its  $n$ th derivative exists and is Lipschitz, that is, if there exists a positive real number  $M_0$  such that

$$\sup_{x, y \in [a, b], |x-y| \leq \delta} |f^{(n)}(x) - f^{(n)}(y)| \leq M_0 \delta$$

for each  $\delta > 0$ . The following result holds.

**Theorem 5.1** (see [8, Theorem 3.7.1]). *Let  $k \geq 0$ ,  $\iota \in \{1, -1\}$ , let  $C \in \mathbb{R}^{2 \times 2}$  be a positive definite matrix, and let  $g(k, \cdot, C)$  be a function of class  $LC^3(N)$ , where  $N$  is a neighborhood*

of its minimum  $\hat{\theta}$ , such that  $g''(k, \hat{\theta}, C) > 0$ . Then the sequence  $\{\theta^{(h)}\}_h$  obtained by the SPI algorithm converges in the neighborhood  $N$  with either strong order  $p \simeq 1.325$  or weak order  $p = ((3 + \sqrt{5})/2)^{1/3} \simeq 1.378$ .

Note that the function  $\tau$  defined in (5.16) is not of class  $C^1$  but can be approximated by the function

$$(5.21) \quad (\bar{\tau}(s))_i = \begin{cases} 0 & \text{if } s_i \leq 0, \\ p_7(s_i) & \text{if } 0 < s_i \leq 1, \\ s_i & \text{if } 1 < s_i \leq m-1, \\ q_7(s_i) & \text{if } m-1 < s_i \leq m, \\ m & \text{if } s_i > m, \end{cases} \quad i = 1, \dots, \nu^2,$$

where

$$\begin{aligned} p_7(x) &= -10x^7 + 36x^6 - 45x^5 + 20x^4, \\ q_7(x) &= m - p_7(m-x), \quad x \in \mathbb{R}, \end{aligned}$$

which is of class  $LC^3$  on  $\mathbb{R}^{n^2}$ . If in (5.17) we replace the mapping  $\tau$  in (5.16) with the function  $\bar{\tau}$  in (5.21), then we obtain that  $g(k, \cdot, C)$  is of class  $LC^3((\varphi^{(1)} + \eta, \varphi^{(2)} + \pi - \eta))$ . Therefore, we are under the hypothesis of Theorem 5.1 and the minimization method has superlinear convergence (see also [6]).

**5.2. The empty page case.** Now we consider the case  $\det C = 0$ . Since  $x_r$  and  $x_v$  are nonnegative vectors, from (4.7) and the Cauchy–Schwarz inequality it follows that there exists  $\zeta > 0$  with  $x_r = \zeta x_v$ . An example is shown in Figure 5(a).

In this case, it is natural to assume that either  $\tilde{s}_r$ , the estimate of  $s_r$ , or  $\tilde{s}_v$ , the estimate of  $s_v$ , is zero, that is, that either the recto or the verso of the ideal source document is an empty page. When  $\zeta \geq 1$ , we assume that  $\tilde{s}_v = 0$ , and we get that  $x_r = \tilde{a}_{11}$ ,  $\tilde{s}_r = \tilde{a}_{21} \tilde{s}_r$ ,  $x_v = \tilde{a}_{21} \tilde{s}_r$ , and  $\zeta = \frac{\tilde{a}_{11}}{\tilde{a}_{21}}$ , where  $\tilde{a}_{11}$  and  $\tilde{a}_{21}$  are estimates of  $a_{11}$  and  $a_{21}$ , respectively. Therefore we obtain

$$\tilde{s}_r = \frac{1}{\tilde{a}_{11}} x_r, \quad \tilde{s}_v = 0 \quad \tilde{A} = \begin{bmatrix} \tilde{a}_{11} & 1 - \tilde{a}_{11} \\ \frac{1}{\zeta} \tilde{a}_{11} & 1 - \frac{1}{\zeta} \tilde{a}_{11} \end{bmatrix},$$

where  $\tilde{a}_{11}$  is arbitrarily chosen in  $]0, 1]$  and  $\tilde{A}$  is an estimate of the mixture matrix  $A$ . If we impose that the matrix  $\tilde{A}$  is symmetric, then we have that  $\tilde{a}_{11} = \frac{\zeta}{\zeta+1}$ . In Figure 5(b) we present a symmetric reconstruction of the document shown in Figure 5(a).



(a) Observed document. (b) Symmetric reconstruction of the document.

**Figure 5.** Document whose recto is a multiple of the verso.

If  $0 < \zeta < 1$ , then we set  $\tilde{s}_r = 0$  and get that  $x_v = \tilde{a}_{12} \tilde{s}_v$ ,  $x_r = \tilde{a}_{22} \tilde{s}_v$ , and  $\zeta = \frac{\tilde{a}_{12}}{\tilde{a}_{22}}$ , where  $\tilde{a}_{12}$  and  $\tilde{a}_{22}$  are estimates of  $a_{12}$  and  $a_{22}$ , respectively. Therefore we obtain

$$\tilde{s}_v = 0, \quad \tilde{s}_v = \frac{1}{\tilde{a}_{22}} x_v, \quad \tilde{A} = \begin{bmatrix} 1 - \zeta \tilde{a}_{22} & \zeta \tilde{a}_{22} \\ 1 - \tilde{a}_{22} & \tilde{a}_{22} \end{bmatrix},$$

where  $\tilde{a}_{22}$  is arbitrarily chosen in  $[0, 1]$ , and by requiring that the estimated mixture matrix  $\tilde{A}$  is symmetric, we obtain that  $\tilde{a}_{22} = \frac{1}{\zeta+1}$ .

Note that, since we consider a locally linear model, it may happen that in a given subdomain at least one of the original documents is empty.

**5.3. Color image case.** An  $n \times n$  color image is usually encoded in the *RGB* space, where *R*, *G*, and *B* indicate red, green, and blue, respectively. We consider every color component of a document as a pair of images, the recto and the verso, and we denote the red, green, and blue data components as

$$\hat{x}_R = [\hat{x}_{rR} \quad \hat{x}_{vR}], \quad \hat{x}_G = [\hat{x}_{rG} \quad \hat{x}_{vG}], \quad \hat{x}_B = [\hat{x}_{rB} \quad \hat{x}_{vB}],$$

respectively, where  $\hat{x}_{rR}, \hat{x}_{rG}, \hat{x}_{rB}, \hat{x}_{vR}, \hat{x}_{vG}, \hat{x}_{vB} \in [0, 255]^{\nu^2}$ . We write the observed color document as

$$\hat{x} = [\hat{x}_{rR} \quad \hat{x}_{vR} \quad \hat{x}_{rG} \quad \hat{x}_{vG} \quad \hat{x}_{rB} \quad \hat{x}_{vB}],$$

which belongs to  $[0, 255]^{\nu^2 \times 6}$ . The source ideal document is given by the matrix

$$\hat{s} = [\hat{s}_{rR} \quad \hat{s}_{vR} \quad \hat{s}_{rG} \quad \hat{s}_{vG} \quad \hat{s}_{rB} \quad \hat{s}_{vB}],$$

where  $\hat{s} \in [0, 255]^{\nu^2 \times 6}$ , and we set

$$\hat{s}_R = [\hat{s}_{rR} \quad \hat{s}_{vR}], \quad \hat{s}_G = [\hat{s}_{rG} \quad \hat{s}_{vG}], \quad \hat{s}_B = [\hat{s}_{rB} \quad \hat{s}_{vB}].$$

The linear model for a color image is  $\hat{x}^T = A \hat{s}^T$ . In this case the mixture matrix  $A \in \mathbb{R}^{6 \times 6}$  is the block matrix

$$A = \begin{bmatrix} A_R & O & O \\ O & A_G & O \\ O & O & A_B \end{bmatrix} \text{ with } A_R = \begin{bmatrix} a_{11}^R & a_{12}^R \\ a_{21}^R & a_{22}^R \end{bmatrix}, A_G = \begin{bmatrix} a_{11}^G & a_{12}^G \\ a_{21}^G & a_{22}^G \end{bmatrix}, A_B = \begin{bmatrix} a_{11}^B & a_{12}^B \\ a_{21}^B & a_{22}^B \end{bmatrix},$$

where  $O \in \mathbb{R}^{2 \times 2}$  is the zero matrix. Thus, we get

$$\hat{x}_R^T = A_R \hat{s}_R^T, \quad \hat{x}_G^T = A_G \hat{s}_R^T, \quad \hat{x}_B^T = A_B \hat{s}_R^T.$$

According to our model, every observed channel is formed by a linear combination of components related to the same channel of the front and the back of the ideal source document, and we can solve the problem independently on each channel with the technique proposed for gray level images.

**6. A new technique for solving the nonstationary problem.** In this section we discuss how to use the MATODS algorithm to solve the nonstationary model proposed in subsection 3.1.

Given a document defined on a domain of dimension  $n \times n$ , in each nonoverlapping subdomain of dimension  $\nu \times \nu$  we model the problem by means of a linear operator, and we assume that these linear operators vary smoothly between adjacent subdomains. For this purpose we solve the problem on an overlapping subimage of size  $\bar{n} \times \bar{n}$ , with  $\bar{n} > \nu$ , using the MATODS algorithm, and then we average the results obtained in each subdomain.

In other words, if  $n$  and  $\bar{n}$  are multiples of  $\nu$ , then we consider the subimages  $\bar{x}^{(p,q)}$ , for  $p, q = 1, \dots, \frac{n-\bar{n}}{\nu}$ , which have fixed size  $\bar{n} \times \bar{n}$ , and solve the linear problem on each subimage. The domain of these subimages is obtained by shifting by  $\nu$  pixels a window of  $\nu \times \nu$  pixels either horizontally or vertically. Finally, we set the light intensity value of every pixel of the estimated source  $\tilde{s}$  to the arithmetic mean of the light intensity value of the estimated subsources to which the pixel belongs.

By this procedure, all pixels lying in a subdomain of dimension  $\nu \times \nu$  belong to the same subimage, and on each subdomain the result is obtained by averaging the  $(\frac{\bar{n}}{\nu})^2$  linear operators. Note that the resulting average is a linear operator, being a linear combination of linear operators. In adjacent subdomains, the reconstruction is the average of linear operators which are almost coincident. Thus, the resulting operators on adjacent domains turn out to be similar, as required.

We now give the pseudocode of the approach discussed thus far:

```

function NIT-MATODS( $x$ )
  Initialize  $\tilde{s}$  as a null matrix;
  for  $p = 1$  to  $n - \bar{n}$  with step  $\nu$  do
    for  $q = 1$  to  $n - \bar{n}$  with step  $\nu$  do
      for  $i = 1$  to  $\bar{n}$  do
        for  $j = 1$  to  $\bar{n}$  do
           $\bar{x}_{r,i,j}^{(p,q)} = x_{r(i+p)\nu+j};$ 
           $\bar{x}_{v,i,j}^{(p,q)} = x_{v(i+p)\nu+j};$ 
        end for
      end for
       $\bar{s}^{(p,q)} = \text{MATODS}(\bar{x}^{(p,q)})$ 
       $\dim_y = \min\{\bar{n}/\nu, \lceil (i+p)/\nu \rceil, \lceil (n+1-i-p)/\nu \rceil\};$ 
       $\dim_x = \min\{\bar{n}/\nu, \lceil (j+q)/\nu \rceil, \lceil (n+1-j-q)/\nu \rceil\};$ 
      for  $i = 1$  to  $\bar{n}$  do
        for  $j = 1$  to  $\bar{n}$  do
           $\tilde{s}_{r,i+p,j+q} = \tilde{s}_{r,i+p,j+q} + \bar{s}_{r,i,j}^{(p,q)} / (\dim_x \cdot \dim_y);$ 
           $\tilde{s}_{v,i+p,j+q} = \tilde{s}_{v,i+p,j+q} + \bar{s}_{v,i,j}^{(p,q)} / (\dim_x \cdot \dim_y);$ 
        end for
      end for
    end for
  end for
  return  $\tilde{s}$ 

```

This method is the NIT-MATODS algorithm. A correct selection of the parameters  $\nu$  and  $\bar{n}$  can improve the quality of the reconstruction. Clearly, a smaller subdomain size  $\nu$  corresponds to a more precise reconstruction at the price of an increase of the overall computational cost of

the algorithm. The choice of the subimage size  $\bar{n}$  is more complex, since as  $\bar{n}$  increases, the total number of subimages decreases, while the degree of smoothness between subdomains increases.

**7. Experimental results.** We implemented both the MATODS and NIT-MATODS algorithms in the C language, and the experiments were run on a Linux machine equipped with a 2.80-GHz processor. First we compared the MATODS algorithm with existing methods for the linear problem, and then we assessed how the NIT-MATODS algorithm performs on real ancient documents, comparing it with fast algorithms that were developed by considering other approximated mathematical models.

For restoring color image documents of dimension  $\nu = 256$ , in most cases the MATODS algorithm requires less than one second (see also [6]); thus we compare it with fast and unsupervised methods, like the FastICA (see also [25, 26, 27, 30, 33]) and the SW (see also [10, 47, 48]) algorithms. We proceeded as follows. First, we generated a synthetic document from a by applying a given mixture matrix to an uncorrupted source document, using the linear model in (3.3). Then we compared the estimated sources with the given source document by means of the *mean squared error* (MSE). In order to compare the FastICA and the SW techniques with our algorithm, at the end of the execution of the FastICA and SW algorithms we transformed the estimated mixture matrices in equivalent one row-sum matrices, as described in section 3, and then we applied an orthogonal projection operator, so that all the results are in the space  $[0, 255]^{\nu^2 \times 6}$ . As source documents we considered the  $256 \times 256$  images in Figure 6.

First, we mixed our documents using the mixture matrices

$$(7.1) \quad A_R = A_G = A_B = \begin{pmatrix} 0.7 & 0.3 \\ 0.3 & 0.7 \end{pmatrix}.$$

In Table 1 we report the MSE of the MATODS, FastICA, and SW algorithms with respect to the original documents. Here,

$$\text{MSE}(s_\rho, \tilde{s}_\rho) = \frac{\|s_\rho - \tilde{s}_\rho\|_F^2}{3\nu^2},$$

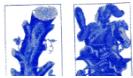
where  $\|\cdot\|_F$  denotes the Frobenius norm,  $\rho \in \{r, v\}$  denotes the recto or the verso of the document, respectively,  $s_\rho$  is the ideal recto or verso, and  $\tilde{s}_\rho$  is an estimate produced by one of the three algorithms we consider. Figure 7(a) shows the mixtures obtained by applying the mask (7.1) to Document 4, and Figures 7(b)–(d) present the result of the MATODS, FastICA, and SW algorithms, respectively. Note that in Table 1, the MATODS algorithm always obtains an error smaller than that of other methods, and in most cases its MSE is negligibly small.

Now we consider the mixture matrices

$$(7.2) \quad A_R = A_G = A_B = \begin{pmatrix} 0.55 & 0.45 \\ 0.45 & 0.55 \end{pmatrix}.$$

The mixtures obtained with these matrices will have the recto very similar to the verso, and the problem becomes more difficult to solve, since the matrices in (7.2) are more ill-conditioned than those in (7.1). The MSEs for the three algorithms are given in Table 2, Figure 8(a) reports the mixtures obtained by applying the mask (7.1) to Document 6, while Figures 8(b)–(d) present the reconstructions obtained by the MATODS, FastICA, and SW algorithms, respectively.

Sud: 2b,  
Est: 7b,  
Nord: Cb,  
Ovest: una scartina a  
piombo a bastoni.



(a) Document 1.

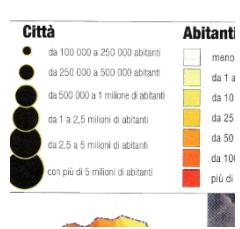
con il Cavallo, per po  
avrebbe scartato denari  
con il cavallo di spade e

Nona presa



rcantile più potente d'Europa

ziario è quello che attualmente  
nomia greca le maggiori poten  
luppo. Ostacolata dalla confor



(b) Document 2.

ne Contraction Mapping the  
tated here, is a simplified ve  
ianach fixed point theorem. S  
ounded modern functional ar  
made outstanding contributio  
theory of topological vector s  
measure theory; integration;  
ets, and orthogonal series. F  
account of Banach's life and  
Kaluza, "Through the Eyes o  
the Life of Stefan Banach" B  
oston, Ma ,1996.

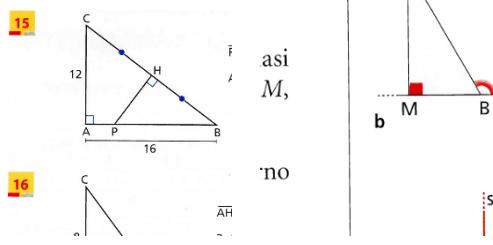
Numerical analysis provides  
foundation for the numerical  
we rely on to solve a multi  
computational problems in si  
on a successful course at Ox  
this book covers a wide rang  
problems from the approxim  
functions and integrals to th  
solution of algebraic, transce  
differential and integral equa  
Throughout the book, particu  
paid to the essential qualities

(c) Document 3.



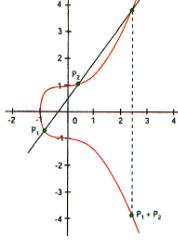
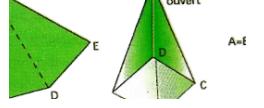
27) Domus di  
28) Ninfeo sen  
29) Sacello di  
30) Ninfeo cor

(d) Document 4.



(e) Document 5.

tagliare an  
ne ed incolin  
nizzionato in  
st'ultima doma  
me il tema della  
non sia solame  
SA, nu  
v: le  
imi  
ste,  
rdi,  
o in  
bro  
ide  
rdy  
sta,  
fet  
ver



(f) Document 6.

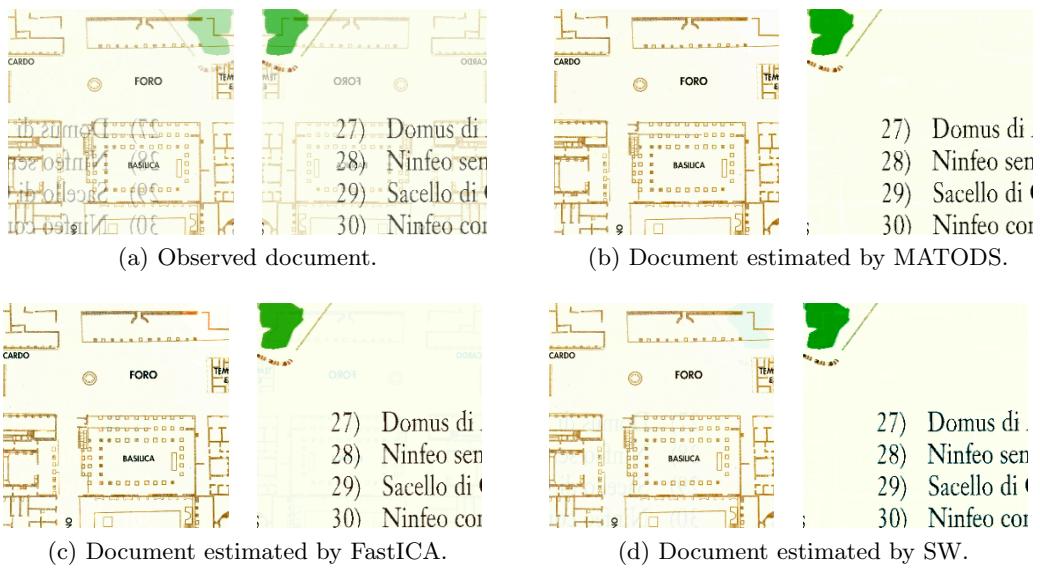
Figure 6. Ideal sources.

Table 1

MSE of the MATODS, FastICA, and SW algorithms using the mixture matrices in (7.1).

Ideal document	MATODS		FastICA		SW	
	MSE recto	MSE verso	MSE recto	MSE verso	MSE recto	MSE verso
1	$1.54 \cdot 10^{-10}$	$4.15 \cdot 10^{-11}$	53.27	1.45	0.95	27.79
2	$3.56 \cdot 10^{-9}$	$6.21 \cdot 10^{-10}$	17.71	12.14	25.74	15.75
3	$6.93 \cdot 10^{-8}$	$9.69 \cdot 10^{-8}$	171.23	30.92	3.69	6.25
4	1.09	6.99	20.62	25.67	8.57	58.14
5	$1.25 \cdot 10^{-5}$	$5.04 \cdot 10^{-12}$	4.46	1.78	5.55	3.44
6	$1.13 \cdot 10^{-9}$	$4.54 \cdot 10^{-11}$	130.81	24.84	159.96	67.11

Again we note that in several cases the reconstructions obtained with MATODS substantially correspond to the ideal document. In this case the data documents have higher overlapping levels; thus FastICA and SW, which force the estimated source overlapping levels to be zero, give results too far from the desired ones. On the other hand, the MATODS algorithm is not affected by the high data overlapping levels, which it estimates correctly. Recall that the MATODS stopping criterion is based on the estimated source overlapping level, whose correct computation requires, in this case, even more accurate source estimates.



**Figure 7.** Results for Document 4 mixed using the matrices in (7.1).

**Table 2**

MSE of the MATODS, FastICA, and SW algorithms using the mixture matrices in (7.2).

Ideal document	MATODS		FastICA		SW	
	MSE recto	MSE verso	MSE recto	MSE verso	MSE recto	MSE verso
1	$5.76 \cdot 10^{-11}$	$5.88 \cdot 10^{-11}$	97.82	27.61	19.33	56.60
2	$4.49 \cdot 10^{-9}$	$8.52 \cdot 10^{-10}$	192.34	105.86	79.44	49.67
3	$6.95 \cdot 10^{-8}$	$9.71 \cdot 10^{-8}$	381.62	331.93	2.88	7.52
4	0.44	4.61	198.80	243.44	28.71	241.31
5	$1.24 \cdot 10^{-5}$	$3.67 \cdot 10^{-12}$	49.58	26.03	19.01	10.96
6	$4.44 \cdot 10^{-10}$	$2.50 \cdot 10^{-11}$	951.51	807.94	628.37	155.55

Now we consider the case of nonsymmetric and partially nonhomogeneous mixture matrices, that is, we assume that  $A_R$ ,  $A_G$ , and  $A_B$  are nonsymmetric matrices and do not always coincide. Let us take

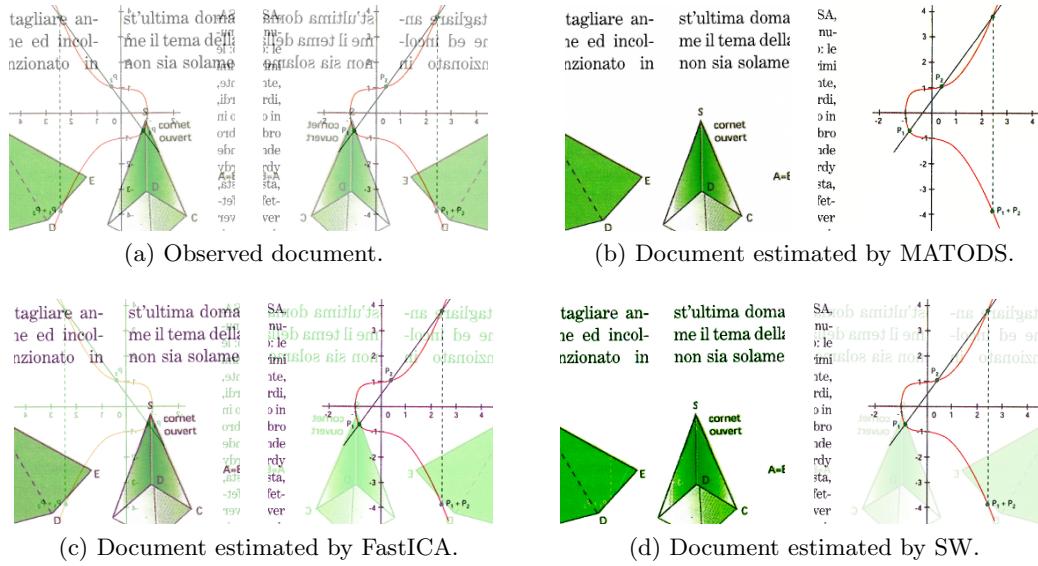
$$(7.3) \quad A_R = A_B = \begin{pmatrix} 0.7 & 0.3 \\ 0.4 & 0.6 \end{pmatrix}, \quad A_G = \begin{pmatrix} 0.6 & 0.4 \\ 0.3 & 0.7 \end{pmatrix}.$$

The MSE of the algorithms we consider is presented in Table 3. In this case, the results obtained using the SW algorithm are not optimal, because the algorithm imposes a symmetry constraint on the estimated mixture matrices. On the other hand, the FastICA algorithm gives better results than in the previous case, because the condition numbers of the matrices in (7.3) are smaller.

Now we illustrate the nonsymmetric and nonhomogeneous case by using the matrices

$$(7.4) \quad A_R = \begin{pmatrix} 0.6 & 0.4 \\ 0.3 & 0.7 \end{pmatrix}, \quad A_G = \begin{pmatrix} 0.7 & 0.3 \\ 0.4 & 0.6 \end{pmatrix}, \quad A_B = \begin{pmatrix} 0.55 & 0.45 \\ 0.4 & 0.6 \end{pmatrix}.$$

The MSE for the three algorithms is reported in Table 4.



**Figure 8.** Results for Document 6 mixed using the matrices in (7.2).

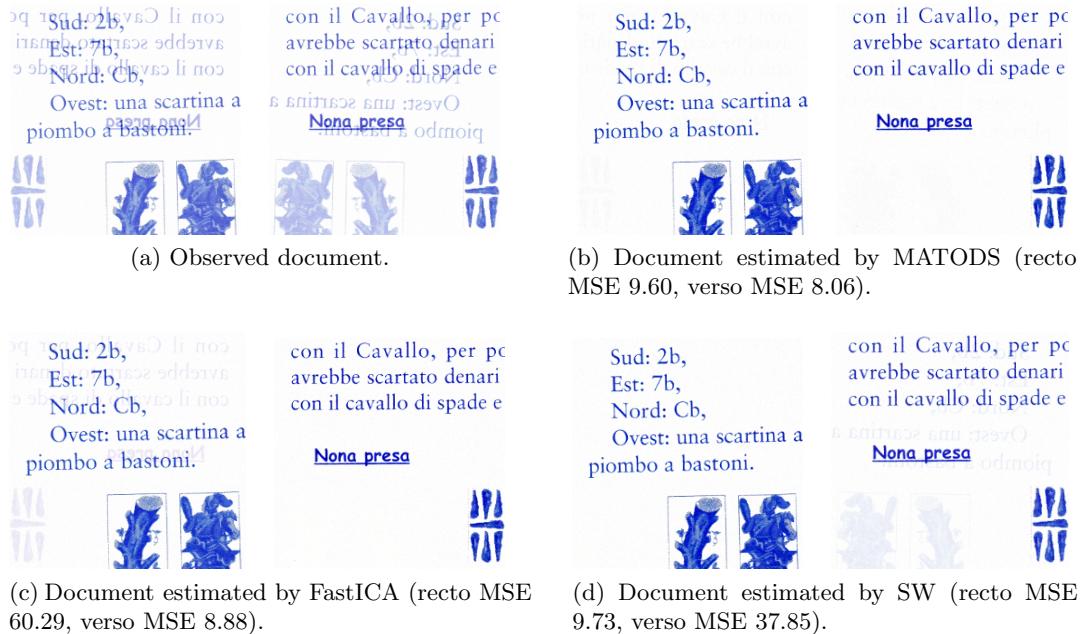
**Table 3**  
MSE of the MATODS, FastICA, and SW algorithms using the mixture matrices in (7.3).

Ideal document	MATODS		FastICA		SW	
	MSE recto	MSE verso	MSE recto	MSE verso	MSE recto	MSE verso
1	$1.28 \cdot 10^{-10}$	$4.56 \cdot 10^{-11}$	43.71	2.27	43.79	54.42
2	$3.51 \cdot 10^{-9}$	$6.07 \cdot 10^{-10}$	19.90	17.64	84.33	50.01
3	$6.95 \cdot 10^{-8}$	$9.71 \cdot 10^{-8}$	175.96	68.05	24.94	16.33
4	0.78	5.93	18.94	33.65	18.37	60.96
5	$1.24 \cdot 10^{-5}$	$3.90 \cdot 10^{-12}$	3.33	2.91	28.21	17.33
6	$1.43 \cdot 10^{-10}$	$1.76 \cdot 10^{-11}$	258.04	96.14	509.88	129.25

**Table 4**  
MSE of the MATODS, FastICA, and SW algorithms using the mixture matrices in (7.4).

Ideal document	MATODS		FastICA		SW	
	MSE recto	MSE verso	MSE recto	MSE verso	MSE recto	MSE verso
1	$1.22 \cdot 10^{-10}$	$4.27 \cdot 10^{-11}$	28.98	0.29	35.03	39.82
2	$3.54 \cdot 10^{-9}$	$6.16 \cdot 10^{-10}$	38.76	18.77	53.57	37.16
3	$6.94 \cdot 10^{-8}$	$9.71 \cdot 10^{-8}$	212.02	89.25	12.08	25.74
4	0.63	5.35	33.75	60.95	25.14	169.47
5	$1.24 \cdot 10^{-5}$	$5.10 \cdot 10^{-12}$	7.05	1.88	10.78	9.01
6	$6.76 \cdot 10^{-10}$	$4.87 \cdot 10^{-11}$	470.29	181.33	202.20	94.64

Note that, in general, the results obtained by the SW and FastICA algorithms are very similar, which confirms what was observed in [48]. Moreover, the estimation of the source overlapping level is very useful for a correct reconstruction of the original sources. Some other examples are given in [6]. In all the cases we examine, the MATODS algorithm obtains the best results.



**Figure 9.** Results for Document 1 mixed using the mixture matrices in (7.1) and corrupted with noise with variance  $\sigma^2 = 4$ .

The MATODS algorithm has been developed to separate linearly mixed components and does not remove deterioration phenomena, like noise in the data. In a noisy document the brightest value does not necessarily coincide with that of the background, and in order to handle noisy data the MATODS algorithm, instead of computing the maximum value of the light intensity, uses the statistical mode of the noisy document, from which it subtracts  $4\sigma^2$ , in order to exclude noise tails. In Figure 9(a) we present Document 1 mixed with the mixture matrices in (7.1) and corrupted with additive white independent Gaussian noise with variance  $\sigma^2 = 4$  and mean zero. In Figure 9(b) we show the result of MATODS, while in Figures 9(c) and (d) we present those of FastICA and SW. Note that the MATODS algorithm separates the sources better than FastICA and SW but does not reduce the noise disturbance.

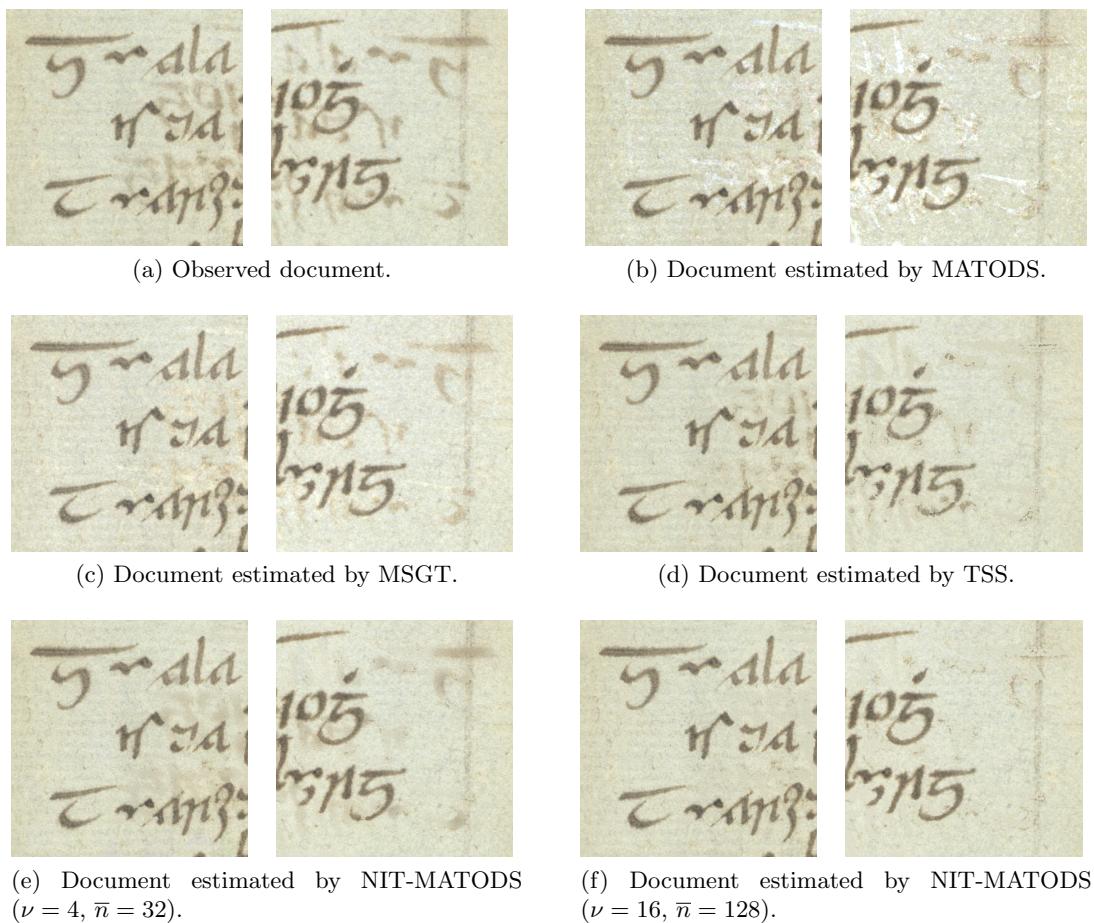
Now we illustrate how the NIT-MATODS algorithm restores **real ancient documents**. In this case, we take some practical measures: we compare the maximum light intensity of the recto with that of the verso, and if the two values do not coincide, we add a constant to the light intensities of the darker image, **in order to reduce the difference between the background color of the two sides**. This is justified by the fact that the color of the paper has to be the same on both sides of the paper. Since images of real documents present noise degradation phenomena, we compute the statistical mode of the document instead of the maximum. We assume that the determinant of the overlapping matrix of the observed data,  $C$ , which corresponds to each channel of every involved subimage, is zero when  $\det(C)/\|C\|_\infty \leq \bar{\varepsilon}$ , where  $\bar{\varepsilon}$  is an accuracy threshold and  $\|\cdot\|_\infty$  is the infinity norm. In the following experiments we deal with documents of size  $n = 512$ , and we set the **subimage dimension  $\bar{n}$  to 128**, while the dimension of the sub-domains is fixed to  $\nu = 16$  pixels. These values were chosen empirically in order to obtain a reasonable trade-off between the quality of the result and the required computational time.

Experimentally we noticed that, as  $\nu$  increases, both the execution time and the quality of the result decrease, whereas as  $\bar{n}$  increases, while remaining below  $n/4$ , both the execution time and the quality of the result increase. Although the choice of the values of  $\nu$  and  $\bar{n}$  is related to the considered data, we experimentally found some values that give good results in the general case. Some reconstructions with different values of  $\nu$  and  $\bar{n}$  are presented in [6]. However, a more exhaustive analysis of the choice of these parameters would be of interest. In the examples presented here, the average execution time of the NIT-MATODS algorithm is 59.72 seconds.

The reconstructions obtained by the NIT-MATODS algorithm are compared with those obtained considering the stationary linear model, the nonlinear stationary model in (3.1), and the nonstationary nonlinear model in (3.2). In particular, for the reconstructions based on the stationary linear model we use the MATODS algorithm, to treat the stationary nonlinear model we use the algorithm proposed in [34], and for the nonstationary nonlinear model we consider the algorithm in [49]. From the initials of the authors, we refer to the algorithm in [34] as the MSGT algorithm and to that in [49] as the TSS algorithm. These methods were chosen because their computational cost is similar to that of the NIT-MATODS algorithm.



**Figure 10.** First ISOS document.

**Figure 11.** Second ISOS document.

The data documents in Figures 10(a)–12(a) are taken from the database created as part of the *Irish Script on Screen* (ISOS) project of the School of Celtic Studies at the Dublin Institute for Advanced Studies, in conjunction with the SIGMEDIA group of the Department of Electrical and Electronic Engineering at Trinity College Dublin (see [40]). This database contains ancient documents affected by bleed-through. The results of MATODS are presented in Figures 10(b)–12(b), those of MSGT in Figures 10(c)–12(c), and those of TSS in Figures 10(d)–12(d). In Figures 10(e)–12(e) we present the reconstruction of the NIT-MATODS algorithm with  $\nu = 4$  and  $\bar{n} = 32$ , in order to show how NIT-MATODS works using nonoptimal parameters. Finally, the results obtained by NIT-MATODS using the optimal parameters  $\nu = 16$  and  $\bar{n} = 128$  are presented in Figures 10(f)–12(f). Other results are shown in [6]. We note that NIT-MATODS improves upon the results of the MSGT and TSS algorithms. This is due to the fact that MSGT and TSS, in order to lower their computational cost, reduce the quality of the reconstructions. In order to obtain more accurate reconstructions from a nonstationary and nonlinear model, a more expensive regularization technique may be adopted (see also [19, 47]).

**Figure 12.** Third ISOS document.

**Conclusions.** In this paper, we dealt with a blind source separation problem in the particular case of document restoration. We examined a nonstationary and locally linear model. To solve the linear problem locally, we considered a *correlated component analysis* technique. In particular, we assumed that the mixture matrix related to the linear model is one row-sum and that the reconstructed documents are nonnegative. Thus we introduced the MATODS algorithm, which iteratively estimates the source overlapping level of the recto and the verso of the document to be restored, from which the ideal sources are estimated. In addition, we proposed a technique for solving the nonstationary and locally linear problem that uses MATODS as a subprocedure. This is the NIT-MATODS algorithm. The experimental results confirm that MATODS works better than some other classical fast and unsupervised algorithms for the linear model. We compared the NIT-MATODS algorithm on real data with algorithms developed to solve other models, showing the effectiveness of our approach. However, we assumed that the background color of the documents to be restored corresponds to the lightest color that is observed and that there is no blur degradation in the data. These assumptions need not be true and may not be satisfied by real documents. An extension of the techniques discussed here that is capable of dealing with these cases will be the subject of future work.

**Acknowledgments.** We thank the referees and Massimiliano Fasi for their useful suggestions, which allowed us to improve the presentation of the paper. We also thank Bruno Iannazzo for suggesting minimizing over the unitary matrices.

## REFERENCES

- [1] M. S. C. ALMEIDA AND L. B. ALMEIDA, *Nonlinear separation of show-through image mixtures using a physical model trained with ICA*, Signal Process., 92 (2012), pp. 872–884.
- [2] L. ARMIJO, *Minimization of functions having Lipschitz continuous first partial derivatives*, Pacific J. Math., 16 (1966), pp. 1–3.
- [3] A. K. BARROS, *The independence assumption: Dependent component analysis*, in Advances in Independent Component Analysis (M. Girolami, ed.), Springer, New York, 2000, pp. 63–71.
- [4] B. XIAO, F. LI, X. MAO, AND H. JIN, *Study on independent component analysis application in classification and change detection of multispectral images*, Internat. Arch. Photogrammetry Remote Sensing Spatial Inform. Sci., 37 (2008), pp. 871–876.
- [5] G. BIANCO, F. BRUNO, A. TONAZZINI, E. SALERNO, AND E. CONSOLE, *Recto-verso registration, enhancement and segmentation of ancient documents*, in 15th International Conference on Virtual Systems and Multimedia, 2009, pp. 131–136.
- [6] A. BOCCUTO, I. GERACE, AND V. GIORGETTI, *Minimum Amount of Text Overlapping in Document Separation*, viXra:1805.0284, 2018, pp. 1–91.
- [7] A. BOCCUTO, I. GERACE, AND F. MARTINELLI, *Half-quadratic image restoration with a non-parallelism constraint*, J. Math. Imaging Vis., 59 (2017), pp. 270–295.
- [8] R. BRENT, *Algorithms for Minimization without Derivatives*, Prentice-Hall, Englewood Cliffs, NJ, 1973.
- [9] T.-H. CHAN, W.-K. MA, C.-Y. CHI, AND Y. WANG, *A convex analysis framework for blind separation of non-negative sources*, IEEE Trans. Signal Process., 56 (2008), pp. 5120–5134.
- [10] A. CICHOCKI AND S.-I. AMARI, *Adaptive Blind Signal and Image Processing*, John Wiley & Sons, Chichester, UK, 2002.
- [11] A. CICHOCKI, R. ZDUNEK, AND S.-I. AMARI, *New algorithms for non-negative matrix factorization in applications to blind source separation*, in Proceedings of the 2006 IEEE International Conference Acoustics, Speech and Signal Processing, Toulouse, France, 2006, pp. 1–4.
- [12] F. CLUNI, D. COSTARELLI, A. M. MINOTTI, AND G. VINTI, *Applications of sampling Kantorovich operators to thermographic images for seismic engineering*, J. Comput. Anal. Appl., 19 (2015), pp. 602–617.
- [13] F. CLUNI, D. COSTARELLI, A. M. MINOTTI, AND G. VINTI, *Enhancement of thermographic images as tool for structural analysis in earthquake engineering*, NTD & E Internat., 70 (2015), pp. 60–72.
- [14] P. COMON, *Independent component analysis. A new concept?*, Signal Process., 36 (1994), pp. 287–314.
- [15] M. P. DESHMUKH AND U. BHOSLE, *A survey of image registration*, Internat. J. Image Process., 5 (2011), pp. 245–269.
- [16] R. FARRAHI MOGHADDAM AND M. CHERIET, *Low quality document image modeling and enhancement*, Int. J. Document Anal. Recognition, 11 (2009), pp. 183–201.
- [17] R. FARRAHI MOGHADDAM AND M. CHERIET, *A variational approach to degraded document enhancement*, IEEE Trans. Pattern Anal. Machine Intell., 32 (2010), pp. 1347–1361.
- [18] L. FEDELI, I. GERACE, AND F. MARTINELLI, *Unsupervised blind separation and deblurring of mixtures of sources*, in Proceedings of Knowledge-Based Intelligent Information and Engineering Systems, Vietri sul Mare, Italy, Lecture Notes in Comput. Sci. 4694, Springer, New York, 2007, pp. 25–32.
- [19] I. GERACE, F. MARTINELLI, AND A. TONAZZINI, *Restoration of recto-verso archival documents through a regularized nonlinear model*, in Proceedings of 20th European Signal Processing Conference, 2012, pp. 1588–1592.
- [20] I. GERACE, C. PALOMBA, AND A. TONAZZINI, *An inpainting technique based on regularization to remove bleed-through from ancient documents*, in 2016 International Workshop on Computational Intelligence for Multimedia Understanding, 2016, pp. 1–5.

- [21] N. GILLIS, *Successive nonnegative projection algorithm for robust nonnegative blind source separation*, SIAM J. Imaging Sci., 7 (2014), pp. 1420–1450.
- [22] N. GILLIS, *Sparse and unique nonnegative matrix factorization through data preprocessing*, J. Mach. Learn. Res., 13 (2012), pp. 3349–3386.
- [23] L. GOTTESFELD BROWN, *A survey of image registration techniques*, ACM Comput. Surveys, 24 (1992), pp. 325–376.
- [24] S.-W. HUANG, D.-L. WAY, AND Z.-C. SHIH, *Physical-based model of ink diffusion in Chinese paintings*, J. WSCG, 10 (2003), pp. 520–527.
- [25] A. HYVÄRINEN, *Fast and robust fixed-point algorithms for independent component analysis*, IEEE Trans. Neural Networks, 10 (1999), pp. 626–634.
- [26] A. HYVÄRINEN, *The fixed-point algorithm and maximum likelihood estimation for independent component analysis*, Neural Process. Lett., 10 (1999), pp. 1–5.
- [27] A. HYVÄRINEN AND E. OJA, *A fast fixed-point algorithm for independent component analysis*, Neural Comput., 9 (1997), pp. 1483–1492.
- [28] P. JARRATT, *An iterative method for locating turning points*, Comput. J., 10 (1967), pp. 82–84.
- [29] M. R. KHAN, H. IMTIAZ, AND M. K. HASAN, *Show-through correction in scanned images using joint histogram*, Signal Image Video Process., 4 (2010), pp. 337–351.
- [30] A. KHAPARDE, M. MADHAVILATHA, M. B. L. MANASA, AND S. PRADEEP KUMAR, *FastICA algorithm for the separation of mixed images*, WSEAS Trans. Signal Process., 4 (2008), pp. 271–278.
- [31] J. KIEFER, *Sequential minimax search for a maximum*, in Proc. Amer. Math. Soc., 4 (1953), pp. 502–506.
- [32] E. KURUOGLU, L. BEDINI, M. T. PARATORE, E. SALERNO, AND A. TONAZZINI, *Source separation in astrophysical maps using independent factor analysis*, Neural Networks, 16 (2003), pp. 479–491.
- [33] R. K. MALIK AND K. SOLANKI, *FastICA based blind source separation for CT imaging under noise conditions*, Int. J. Adv. Engrg. Technol., 5 (2012), pp. 47–55.
- [34] F. MARTINELLI, E. SALERNO, I. GERACE, AND A. TONAZZINI, *Nonlinear model and constrained ML for removing back-to-front interferences from recto-verso documents*, Pattern Recognition, 45 (2012), pp. 596–605.
- [35] B. OPHIR AND D. MALAH, *Improved cross-talk cancellation in scanned images by adaptive decorrelation*, in Proceedings of the 23rd IEEE Convention of Electrical and Electronics Engineers in Israel, 2004.
- [36] B. OPHIR AND D. MALAH, *Show-through cancellation in scanned images using blind source separation techniques*, in Proceedings of the IEEE International Conference on Image Processing, 2007, pp. 233–236.
- [37] W. S. B. OUEDRAOGO, A. SOLOUMIAC, M. JAIDANE, AND C. JUTTEN, *Non-negative blind source separation algorithm based on minimum aperture simplicial cone*, IEEE Trans. Signal Process., 62 (2014), pp. 376–389.
- [38] S. RICCIARDI, A. BONALDI, P. NATOLI, G. POLENTA, C. BACCIGALUPI, E. SALERNO, K. KAYABOL, L. BEDINI, AND G. DE ZOTTI, *Correlated component analysis for diffuse component separation with error estimation on simulated plank polarization data*, Mon. Not. R. Astron. Soc., 406 (2010), pp. 1644–1658.
- [39] K. ROTH, *Scaling of water flow through porous media and soils*, European J. Soil Sci., 59 (2008), pp. 125–130.
- [40] R. ROWLEY-BROOKE, F. PITIÉ AND A. KOKARAM, *A ground truth bleed-through document image database*, in Theory and Practice of Digital Libraries P. Zaphiris, G. Buchanan, E. Rasmussen and F. Loizides, eds., Lecture Notes in Comput. Sci. 7489, Springer, New York, 2012, pp. 185–196.
- [41] E. SALERNO, F. MARTINELLI AND A. TONAZZINI, *Nonlinear model identification and see-through cancellation from recto-verso data*, Internat. J. Document Anal. Recognition, 16 (2013), pp. 177–187.
- [42] P. SAVINO, L. BEDINI, AND A. TONAZZINI, *Joint non-rigid registration and restoration of recto-verso ancient manuscripts*, in 2016 International Workshop on Computational Intelligence for Multimedia Understanding, 2016, pp. 1–5.
- [43] G. SHARMA, *Show-through cancellation in scans of duplex printed documents*, IEEE Trans. Image Process., 10 (2001), pp. 736–754.
- [44] A. TONAZZINI AND L. BEDINI, *Restoration of recto-verso colour documents using correlated component analysis*, EURASIP J. Adv. Signal Process., 58 (2013), pp. 1–10.

- [45] A. TONAZZINI, L. BEDINI, E. E. KURUOGLU, AND E. SALERNO, *Blind separation of auto-correlated images from noisy mixtures using MRF models*, in Proceedings of the 4th International Symposium on Independent Component Analysis and Blind Source Separation, Nara, Japan, 2003.
- [46] A. TONAZZINI, I. GERACE, AND F. MARTINELLI, *Multichannel blind separation and deconvolution of images for document analysis*, IEEE Trans. Image Process., 19 (2010), pp. 912–925.
- [47] A. TONAZZINI, I. GERACE, AND F. MARTINELLI, *Document image restoration and analysis as separation of mixtures of patterns: From linear to nonlinear models*, in Image Restoration—Fundamentals and Advances, B. K. Gunturk and X. Li, eds., CRC Press, Boca Raton, FL, 2013, pp. 285–310.
- [48] A. TONAZZINI, E. SALERNO, AND L. BEDINI, *Fast correction of bleed-through distortion in greyscale documents by a blind source separation technique*, Int. J. Document Anal., 10 (2007), pp. 17–25.
- [49] A. TONAZZINI, P. SAVINO, AND E. SALERNO, *A non-stationary density model to separate overlapped texts in degraded documents*, SIViP, 9 (Suppl. 1) (2015), pp. S155–S164.
- [50] S. VAVASIS, *On the complexity of nonnegative matrix factorization*, SIAM J. Optim., 20 (2009), pp. 1364–1377.
- [51] H. H. VAZIRI, Y. XIAO, R. ISLAM, AND A. NOURI, *Numerical modeling of seepage-induced sand production in oil and gas reservoirs*, J. Petroleum Sci. Engng., 36 (2002), pp. 71–86.
- [52] C. WOLF, *Document ink bleed-through removal with two hidden Markov random fields and a single observation field*, IEEE Trans Pattern Anal. Mach. Intell., 32 (2010), pp. 431–447.
- [53] Q. ZHANG, Y. SATO, J.-Y. TAKAHASHI, K. MURAOKA, AND N. CHIBA, *Simple cellular automaton-based simulation of ink behaviour and its application to Suibokuga-like 3D rendering of trees*, J. Visual. Comput. Animat., 10 (1999), pp. 27–37.
- [54] B. ZITOVÁ AND J. FLUSSER, *Image registration methods: A survey*, Image Vision Comput., 21 (2003), pp. 977–1000.