

Spectrally Optimal Factorization of Incomplete Matrices *

Pedro M. Q. Aguiar Marko Stosic
João M. F. Xavier

Institute for Systems and Robotics / IST, Lisboa, Portugal

aguiar@isr.ist.utl.pt

Abstract

From the recovery of structure from motion to the separation of style and content, many problems in computer vision have been successfully approached by using bilinear models. The reason for the success of these models is that a globally optimal decomposition is easily obtained from the Singular Value Decomposition (SVD) of the observation matrix. However, in practice, the observation matrix is often incomplete, the SVD can not be used, and only sub-optimal solutions are available. The majority of these solutions are based on iterative local refinements of a given cost function, and lack any guarantee of convergence to the global optimum. In this paper, we propose a globally optimal solution, for particular patterns of missing entries. To achieve this goal, we re-formulate the problem as the minimization of the spectral norm of the matrix of residuals, i.e., we seek the completion of the observation matrix such that the largest singular value of its difference to a low rank matrix is the smallest possible. The class of patterns of missing entries we deal with is known as the Young diagram, which includes, as particular cases, many relevant situations, such as the missing of an entire submatrix. We describe experiments that illustrate how our globally optimal solution has impact in practice.

1. Introduction

Various large-scale nonlinear problems in computer vision have been addressed by exploiting bilinear models. Examples include the recovery of rigid structure from motion [18, 16], inference of non-rigid models [4], estimation of image motion [12], modeling images obtained with different lighting [3], photometric stereo [2], object recognition [19], and the separation of style and content in images [7]. Usually, a matrix that collects the observations is expressed as the product of two matrices collecting (functions of) the

parameters to estimate. This bilinear relation codes a constraint on the rank of the observation matrix. The advantage of this formulation is that the bulk of the effort needed to compute the solution of the large-scale nonlinear estimation problem is done by computing the rank constrained matrix that best matches the observation matrix (thus, its optimal bilinear decomposition), which is easily obtained from its Singular Value Decomposition (SVD).

In practice, it is often the case that not all the entries of the observation matrix are available. For example, when recovering 3D structure from motion, the observation matrix collects 2D trajectories of (projections of) feature points in the image plane. Naturally, due to self-occlusion and limited field-of-view, feature points are usually not visible in all images available, and the observation matrix is incomplete. This prevents the simple use of the SVD and motivated researchers to address the problem of computing a rank constrained matrix from a partial observation of its entries, i.e., the problem of factorization with missing data.

Current approaches to the factorization of incomplete matrices are sub-optimal. Earlier examples are procedures to combine the constraints that arise from the observed submatrices of the original matrix [18, 13]. More recently, researchers attempted to develop algorithms to minimize the nonlinear cost function that measures the residual of the approximation. These algorithms are iterative, local, thus sensitive to the initialization, i.e., there is not guarantee of convergence to the global optimum. Examples include two-step iterative algorithms, e.g., alternating between the computation of the column and row spaces, [9, 10], non-linear damped Newton optimization of the matrix factors [5], and other non-linear optimization strategies, such as optimizing with respect to one of the factors the cost resulting from replacing the closed-form estimate of the other factor [15]. In opposition to these iterative approaches, some authors proposed the so-called batch algorithms, that address the minimization of an approximation of the residual, to simplify the optimization and avoid local minima [14, 17].

In this paper, we propose an optimal solution for the problem of factorizing an incomplete observation matrix,

*To appear in CVPR'08, Proc. of IEEE Conf. on Computer Vision and Pattern Recognition, Anchorage, Alaska, June 2008.

when the pattern of missing entries has a specific format. We show that the problem can be seen as the minimization of the spectral norm of the matrix of residuals and propose a globally optimal approach to this optimization problem. Our algorithm is not iterative; it rather computes the optimal completion of the observation matrix by filling, in a sequential way, its unknown entries. At each step, the optimal value for one missing is simply obtained in closed form as the solution of a quadratic equation. The class of patterns of missing entries we deal with is known as the Young diagram, which includes, as particular cases, many relevant situations, such as the missing of an entire submatrix.

2. Minimizing the error spectral norm

Motivated by the applications referred above, we address the problem of computing a rank deficient matrix, say a $m \times n$ matrix of rank at most $r < \min(m, n)$, from (noisy) observations of a subset of its entries.

The most common formulation for the problem of finding the rank deficient matrix $\widehat{\mathbf{W}}$ that best matches an incomplete observation matrix \mathbf{W} is (see, e.g., [9, 5])

$$\widehat{\mathbf{W}} = \arg \min_{\widetilde{\mathbf{W}} \in \mathcal{S}_r} \left\| (\mathbf{W} - \widetilde{\mathbf{W}}) \odot \mathbf{M} \right\|, \quad (1)$$

where: \mathbf{M} is a $m \times n$ binary matrix that codes which entries of \mathbf{W} were observed, *i.e.*, is a mask defined by $M_{ij} := 1$ if \mathbf{W}_{ij} was observed, and $M_{ij} := 0$, if otherwise; \odot denotes the Schur-Hadamard (or elementwise) product; and \mathcal{S}_r is the set of matrices with rank less or equal to r . Usually, $\|\cdot\|$ stands for the Frobenius norm, such that, if all the entries of \mathbf{W} are observed, *i.e.*, if $\mathbf{M} = \mathbf{1}_{m \times n}$, the solution of (1) is easily obtained from the SVD of \mathbf{W} . However, when \mathbf{W} is incomplete, there is not known solution for (1) and only suboptimal methods have been proposed.

We start by noting that the optimization in (1) is equivalent to the following one, where we also make explicit the optimal completion of the observation matrix:

$$(\widehat{\mathbf{W}}, \widehat{\mathbf{X}}) = \arg \min_{\widetilde{\mathbf{W}} \in \mathcal{S}_r, \mathbf{X} \in \mathcal{S}_W} \left\| \mathbf{X} - \widetilde{\mathbf{W}} \right\|. \quad (2)$$

Here, \mathcal{S}_W is the set of all matrices that have the entries corresponding to the ones that were observed in \mathbf{W} , equal to the observation values, *i.e.*, $\mathbf{X} \in \mathcal{S}_W$, if and only if, $\forall_{ij: M_{ij}=1}, \mathbf{X}_{ij} = \mathbf{W}_{ij}$; thus, \mathcal{S}_W collects all possible completions of the incomplete observation matrix \mathbf{W} . The matrix $\widehat{\mathbf{X}}$ resulting from (2) is the best completion of \mathbf{W} , in terms of minimizing its difference to a rank r matrix.

Figure 1 represents the scenario in the space of $m \times n$ matrices, seen as $\mathbb{R}^{m \times n}$. The set \mathcal{S}_W is represented as an (hyper-) plane, because the matrices \mathbf{X} that are possible completions of \mathbf{W} live in an affine space. In fact, note that

any completion \mathbf{X} can be written as the sum of a fixed term with a linear combination of a subset of basis matrices, *e.g.*,

$$\mathbf{X} = \mathbf{W} \odot \mathbf{M} + \sum_{i,j: M_{ij}=0} \alpha(i,j) \mathbf{E}^{ij}, \quad (3)$$

where: the matrices $\{\mathbf{E}^{ij}\}$ are the versors of the $\mathbb{R}^{m \times n}$ coordinate system, *i.e.*, each \mathbf{E}^{ij} is a matrix with 1 at entry (i, j) and zeros at all other entries; and the scalars $\{\alpha(i, j)\}$ are defined by $\alpha(i, j) := \mathbf{X}_{ij}$. In Figure 1, we also represent the set \mathcal{S}_r , composed by the $m \times n$ matrices that have rank less or equal to r , and the solution of (2), given by the pair of matrices $\widehat{\mathbf{W}} \in \mathcal{S}_r$ and $\widehat{\mathbf{X}} \in \mathcal{S}_W$ that are the closest to each other.

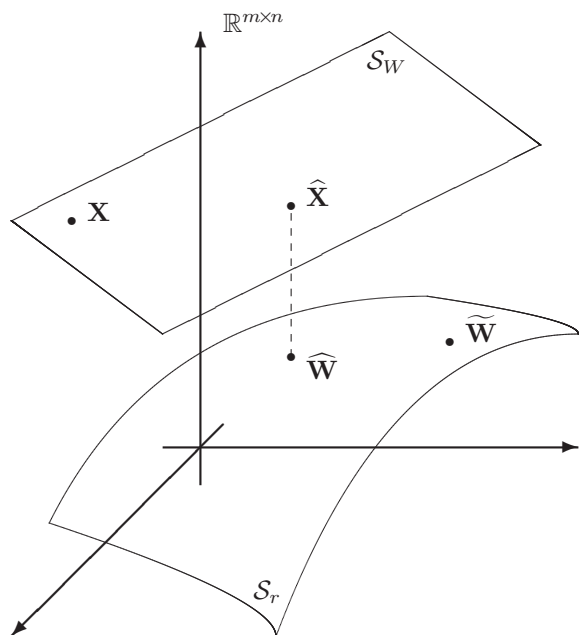


Figure 1. The space of $n \times m$ matrices, with \mathcal{S}_r , the set of matrices of rank at most r , and \mathcal{S}_W , the set of possible completions of the (incomplete) observation matrix \mathbf{W} . We seek the closest pair, *i.e.*, the matrices $\widehat{\mathbf{W}} \in \mathcal{S}_r$ and $\widehat{\mathbf{X}} \in \mathcal{S}_W$ that solve (2).

As it is clearly illustrated by Figure 1, finding the best completion matrix in \mathcal{S}_W corresponds to minimizing its distance to the set of rank constrained matrices, \mathcal{S}_r . In fact, $\widehat{\mathbf{X}}$ in (2) is equivalently written as

$$\widehat{\mathbf{X}} = \arg \min_{\mathbf{X} \in \mathcal{S}_W} d_{\mathcal{S}_r}(\mathbf{X}), \quad (4)$$

where

$$d_{\mathcal{S}_r}(\mathbf{X}) = \min_{\widetilde{\mathbf{W}} \in \mathcal{S}_r} \left\| \mathbf{X} - \widetilde{\mathbf{W}} \right\| \quad (5)$$

denotes the distance from an arbitrary matrix \mathbf{X} to the set \mathcal{S}_r . The advantage of writing the problem as we did in (4)

is that the minimizer of the distance (5) is easily obtained in closed form, *i.e.*, without any search, for several choices of the underlying norm. In fact, from the SVD of a matrix \mathbf{X} ,

$$\mathbf{X} = \mathbf{U}\Sigma\mathbf{V}^T = \sum_{i=1}^N \mathbf{u}_i \sigma_i \mathbf{v}_i^T, \quad (6)$$

where $N = \min\{m, n\}$ and the singular values are non-increasingly ordered, the rank r matrix $\widetilde{\mathbf{W}}^* \in \mathcal{S}_r$ that is closer to \mathbf{X} (with respect to several norms) is simply given by, see, *e.g.*, [8],

$$\widetilde{\mathbf{W}}^* = \sum_{i=1}^r \mathbf{u}_i \sigma_i \mathbf{v}_i^T, \quad (7)$$

and the approximation error matrix is:

$$\mathbf{X} - \widetilde{\mathbf{W}}^* = \sum_{i=r+1}^N \mathbf{u}_i \sigma_i \mathbf{v}_i^T. \quad (8)$$

We propose to use in (2) (thus, in (5)) the spectral norm. The spectral norm of a matrix is given by its largest singular value:

$$\|\mathbf{A}\| := \max_i \sigma_i(\mathbf{A}) = \sigma_1(\mathbf{A}). \quad (9)$$

The spectral norm is popular because it is the induced matrix norm, when the matrix is seen as an operator between Euclidean vector spaces [8]. It has been used in several fields, *e.g.*, control theory [6]. Using this norm in (2) makes sense: we will be seeking the rank deficient matrix that minimizes the largest singular value of its difference to a matrix that collects the observed entries. Furthermore, we will see that, unlike the Frobenius norm, the use of the spectral norm enables us to find the globally optimal solution of (2), for particular (although relevant) patterns of missing entries.

Using the spectral norm in (5), the distance of a matrix \mathbf{X} to the space of matrices of rank at most r , *i.e.*, the norm of the approximation error matrix in (8), is simply given by the $(r+1)^{\text{th}}$ largest singular value of \mathbf{X} :

$$d_{\mathcal{S}_r}(\mathbf{X}) = \|\mathbf{X} - \widetilde{\mathbf{W}}^*\| \quad (10)$$

$$= \sigma_1(\mathbf{X} - \widetilde{\mathbf{W}}^*) \quad (11)$$

$$= \sigma_{r+1}(\mathbf{X}), \quad (12)$$

where: equality (10) uses the optimality of $\widetilde{\mathbf{W}}^*$ in (7); equality (11) is just the definition of spectral norm in (9); and (12) results from the SVD of the error matrix in (8).

Replacing $d_{\mathcal{S}_r}(\mathbf{X})$ in (4) by the result (12), we end up with the optimization problem

$$\widehat{\mathbf{X}} = \arg \min_{\mathbf{X} \in \mathcal{S}_W} \sigma_{r+1}(\mathbf{X}), \quad (13)$$

which we address in the following sections. In words, the solution we seek in (13) is the completion of \mathbf{W} that has the smallest $(r+1)$ singular value. We will derive the globally optimal solution for relevant cases.

3. Optimal solution for one missing entry

We start by studying the case of only one missing entry in the $m \times n$ observation matrix \mathbf{W} , which we write as:

$$\mathbf{W} = \begin{bmatrix} ? & \mathbf{a}^T \\ \mathbf{b} & \mathbf{C} \end{bmatrix}, \quad (14)$$

where the $(n-1) \times 1$ vector \mathbf{a} , the $(m-1) \times 1$ vector \mathbf{b} , and the $(m-1) \times (n-1)$ matrix \mathbf{C} , are known — they contain the observed entries of \mathbf{W} . Although, for simplicity, we have assumed in (14) that the unknown entry is located in the top left position, this entails no loss of generality, since the generic case can be brought to this canonical format, through left and right multiplication of \mathbf{W} by adequate permutation matrices, without changing its singular values.

According to our derivations in the previous section, see expression (13), the single missing entry case leads to the study of the following optimization problem:

$$p^* = \min_{x \in \mathbb{R}} \sigma_{r+1} \left(\begin{bmatrix} x & \mathbf{a}^T \\ \mathbf{b} & \mathbf{C} \end{bmatrix} \right), \quad (15)$$

where $x \in \mathbb{R}$ denotes a candidate value for the unknown entry of \mathbf{W} , thus the matrix inside brackets (\cdot) represents a generic element of the set of possible completions \mathcal{S}_W .

The theorem of interlacing inequalities for singular values, see *e.g.*, [11], provides the two following lower bounds, valid for any choice of $x \in \mathbb{R}$:

$$\sigma_{r+1} \left(\begin{bmatrix} x & \mathbf{a}^T \\ \mathbf{b} & \mathbf{C} \end{bmatrix} \right) \geq \sigma_{r+1}([\mathbf{b} \ \mathbf{C}]), \quad (16)$$

$$\sigma_{r+1} \left(\begin{bmatrix} x & \mathbf{a}^T \\ \mathbf{b} & \mathbf{C} \end{bmatrix} \right) \geq \sigma_{r+1} \left(\begin{bmatrix} \mathbf{a}^T \\ \mathbf{C} \end{bmatrix} \right). \quad (17)$$

These bounds are transmitted to the infimum p^* in (15) as

$$p^* \geq \max \left\{ \sigma_{r+1}([\mathbf{b} \ \mathbf{C}]), \sigma_{r+1} \left(\begin{bmatrix} \mathbf{a}^T \\ \mathbf{C} \end{bmatrix} \right) \right\}. \quad (18)$$

It can be shown that the lower bound in (18) is tight, *i.e.*, that (18) holds as an equality. Moreover, in general, the optimization problem (15) is solvable, *i.e.*, there is a value x^* for the missing entry that realizes the infimum p^* :

$$\sigma_{r+1} \left(\begin{bmatrix} x^* & \mathbf{a}^T \\ \mathbf{b} & \mathbf{C} \end{bmatrix} \right) = p^*. \quad (19)$$

More precisely, the set of values of \mathbf{a} , \mathbf{b} , and \mathbf{C} , that make problem (15) solvable is dense in $\mathbb{R}^{n-1} \times \mathbb{R}^{m-1} \times \mathbb{R}^{(m-1) \times (n-1)}$, in the sense that the closure of that set yields the whole space, see [1] for details.

In this general case, not only the optimization problem (15) is solvable, as it is straightforward to obtain the solution for x^* in (19). In fact, equality (19) implies that the

square of the singular value p^* is an eigenvalue (the $(r+1)$ th largest one) of the matrix

$$\begin{bmatrix} x^* & \mathbf{a}^T \\ \mathbf{b} & \mathbf{C} \end{bmatrix} \begin{bmatrix} x^* & \mathbf{b}^T \\ \mathbf{a} & \mathbf{C}^T \end{bmatrix} = \begin{bmatrix} (x^*)^2 + \mathbf{a}^T \mathbf{a} & x^* \mathbf{b}^T + \mathbf{a}^T \mathbf{C}^T \\ x^* \mathbf{b} + \mathbf{C} \mathbf{a} & \mathbf{b} \mathbf{b}^T + \mathbf{C} \mathbf{C}^T \end{bmatrix}.$$

Since $(p^*)^2$ is a root of the characteristic polynomial of this matrix, the following equation is satisfied by the unknown entry x^* :

$$\begin{vmatrix} (x^*)^2 + \mathbf{a}^T \mathbf{a} - (p^*)^2 & x^* \mathbf{b}^T + \mathbf{a}^T \mathbf{C}^T \\ x^* \mathbf{b} + \mathbf{C} \mathbf{a} & \mathbf{b} \mathbf{b}^T + \mathbf{C} \mathbf{C}^T - (p^*)^2 \mathbf{I}_{m-1} \end{vmatrix} = 0,$$

where \mathbf{I}_{m-1} stands for the $(m-1) \times (m-1)$ identity matrix and $|\mathbf{Z}|$ represents the determinant of the matrix \mathbf{Z} . Since the determinant is a multilinear operator (*i.e.*, linear in each column separately), the equality above is simply a quadratic equation in the unknown x^* , which is easily solved (in [1], it is further shown that, under the general assumptions of solvability of the optimization problem (19), the solution is unique).

4. Optimal solution for Young diagrams

We now build on the result of the previous section to derive the optimal solution of (13), for a more general pattern of missing entries. More precisely, we will consider that the missing entries follow a format known as the Young diagram. This format corresponds to the having the missing entries arranged in the first n_1 entries of the 1st row, the first n_2 entries of the 2nd row, ..., the first n_k entries of the k th row, such that $n_1 \geq n_2 \geq \dots \geq n_k$. As an example, consider the 5×6 observation matrix

$$\mathbf{W} = \begin{bmatrix} ? & ? & ? & ? & a_{15} & a_{16} \\ ? & ? & ? & a_{24} & a_{25} & a_{26} \\ ? & ? & a_{33} & a_{34} & a_{35} & a_{36} \\ ? & ? & a_{43} & a_{44} & a_{45} & a_{46} \\ a_{51} & a_{52} & a_{53} & a_{54} & a_{55} & a_{56} \end{bmatrix}, \quad (20)$$

where $k = 4$ and $n_1 = 4 \geq n_2 = 3 \geq n_3 = 2 \geq n_4 = 2$. The symbol ? represents a missing entry and the known entries are denoted by a_{ij} . The Young diagram pattern of unknown entries is thus able to represent relevant cases, such as, for example, the missing of an entire submatrix.

The interesting fact about this class of incomplete matrices is that the optimal completion, *i.e.*, the completion minimizing the $(r+1)$ th singular value, can be found in a sequential way, *i.e.*, by solving, for each missing entry, a one-dimensional problem such as the one addressed in the previous section and expressed in (15). Our procedure is the following: we start at the k th row and complete from right to left, then we move to the $(k-1)$ th row and complete from right to left, ..., and so on, until we complete the 1st row. For example, for the matrix in (20), the order by which the

missing entries are filled in is indicated below as $\checkmark_1, \checkmark_2, \dots, \checkmark_{11}$:

$$\begin{bmatrix} \checkmark_{11} & \checkmark_{10} & \checkmark_9 & \checkmark_8 & a_{15} & a_{16} \\ \checkmark_7 & \checkmark_6 & \checkmark_5 & a_{24} & a_{25} & a_{26} \\ \checkmark_4 & \checkmark_3 & a_{33} & a_{34} & a_{35} & a_{36} \\ \checkmark_2 & \checkmark_1 & a_{43} & a_{44} & a_{45} & a_{46} \\ a_{51} & a_{52} & a_{53} & a_{54} & a_{55} & a_{56} \end{bmatrix}. \quad (21)$$

Computing \checkmark_i corresponds to solving (15) for the largest submatrix of \mathbf{W} that has the entry corresponding to \checkmark_i in its top left position. Note that, due to the filling order just described, all the other entries of this submatrix are available: they correspond either to observed entries of \mathbf{W} or to $\checkmark_j, j = 1, 2, \dots, i-1$, which were computed before. For example, computing \checkmark_6 in (21) means solving

$$\min_{x \in \mathbb{R}} \sigma_{r+1} \left(\begin{bmatrix} x & \checkmark_5 & a_{24} & a_{25} & a_{26} \\ \checkmark_3 & a_{33} & a_{34} & a_{35} & a_{36} \\ \checkmark_1 & a_{43} & a_{44} & a_{45} & a_{46} \\ a_{52} & a_{53} & a_{54} & a_{55} & a_{56} \end{bmatrix} \right), \quad (22)$$

where $\checkmark_5, \checkmark_3$ and \checkmark_1 denote entries that have already been computed at this stage.

A simple example provides insight over why the procedure just described yields the optimal completion. Consider the problem of computing the optimal completion of an observation matrix \mathbf{W} with two missing entries, located at the first and second elements of the first row:

$$p^* = \min_{(x, y) \in \mathbb{R}^2} \sigma_{r+1} \left(\begin{bmatrix} x & y & \mathbf{a}^T \\ \mathbf{b}_1 & \mathbf{b}_2 & \mathbf{C} \end{bmatrix} \right). \quad (23)$$

According to the procedure described above, we first solve for the unknown entry y , *i.e.*, we compute y^* such that, according to the derivations of the previous section,

$$\sigma_{r+1} \left(\begin{bmatrix} y^* & \mathbf{a}^T \\ \mathbf{b}_2 & \mathbf{C} \end{bmatrix} \right) = \max \left\{ \sigma_{r+1}([\mathbf{b}_2 \ \mathbf{C}]), \sigma_{r+1} \left(\begin{bmatrix} \mathbf{a}^T \\ \mathbf{C} \end{bmatrix} \right) \right\}. \quad (24)$$

Then, we plug in the solution y^* and solve for the entry x , *i.e.*, we compute x^* such that

$$\sigma_{r+1} \left(\begin{bmatrix} x^* & y^* & \mathbf{a}^T \\ \mathbf{b}_1 & \mathbf{b}_2 & \mathbf{C} \end{bmatrix} \right) = \max \left\{ \sigma_{r+1}([\mathbf{b}_1 \ \mathbf{b}_2 \ \mathbf{C}]), \sigma_{r+1} \left(\begin{bmatrix} y^* & \mathbf{a}^T \\ \mathbf{b}_1 & \mathbf{C} \end{bmatrix} \right) \right\}. \quad (25)$$

Replacing (24) into (25) yields

$$\sigma_{r+1} \left(\begin{bmatrix} x^* & y^* & \mathbf{a}^T \\ \mathbf{b}_1 & \mathbf{b}_2 & \mathbf{C} \end{bmatrix} \right) = \max \left\{ \sigma_{r+1}([\mathbf{b}_1 \ \mathbf{b}_2 \ \mathbf{C}]), \sigma_{r+1}([\mathbf{b}_2 \ \mathbf{C}]), \sigma_{r+1} \left(\begin{bmatrix} \mathbf{a}^T \\ \mathbf{C} \end{bmatrix} \right) \right\}. \quad (26)$$

Since any singular value of a matrix is greater or equal than the corresponding one of any of its submatrices, we have $\sigma_{r+1}([\mathbf{b}_2 \ \mathbf{C}]) \leq \sigma_{r+1}([\mathbf{b}_1 \ \mathbf{b}_2 \ \mathbf{C}])$, and equality (26) reduces to

$$\sigma_{r+1} \left(\begin{bmatrix} x^* & y^* & \mathbf{a}^T \\ \mathbf{b}_1 & \mathbf{b}_2 & \mathbf{C} \end{bmatrix} \right) = \max \left\{ \sigma_{r+1}([\mathbf{b}_1 \ \mathbf{b}_2 \ \mathbf{C}]), \sigma_{r+1} \left(\begin{bmatrix} \mathbf{a}^T \\ \mathbf{C} \end{bmatrix} \right) \right\}. \quad (27)$$

Now, remark that, in a similar way of that of the previous section, the theorem of interlacing inequalities for singular values [11] provides the following bound for the optimal p^* of the overall problem (23):

$$p^* \geq \max \left\{ \sigma_{r+1}([\mathbf{b}_1 \ \mathbf{b}_2 \ \mathbf{C}]), \sigma_{r+1} \left(\begin{bmatrix} \mathbf{a}^T \\ \mathbf{C} \end{bmatrix} \right) \right\}. \quad (28)$$

This bound, together with equality (27), proves that the pair (x^*, y^*) , computed according to our sequential procedure, is the optimal completion. Naturally, this simple example can be extended to show, in a similar way, that the global optimality of our procedure holds in general, see [1] for details.

5. Young-wise optimal iterative algorithms

The procedure proposed in the previous section yields the optimal completion when the pattern of missing entries is a Young diagram. However, it also enables the development of (sub-optimal) iterative schemes to deal with fully general patterns of missing entries, by optimizing in a block-wise way.

Consider as an example observation matrices with missing entries grouped in two separate Young diagrams, such as the pattern illustrated by

$$\mathbf{W} = \begin{bmatrix} ? & ? & ? & \dots & ? & \times & \times & \dots & \times \\ ? & ? & ? & \dots & ? & \times & \times & \dots & \times \\ ? & ? & \dots & ? & \times & \times & \times & \dots & \times \\ ? & ? & \dots & ? & \times & \dots & \times & ? & ? \\ ? & \dots & ? & \times & \dots & \times & ? & \dots & ? \\ ? & \dots & ? & \times & \dots & \times & ? & \dots & ? \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ ? & \times & \times & \dots & \times & ? & ? & \dots & ? \\ \times & \times & \dots & \times & ? & ? & ? & \dots & ? \end{bmatrix},$$

where the observed entries are represented by \times and the missing ones by $?$. This type of pattern arises when recovering structure from motion, see *e.g.*, [18, 16, 13, 9, 10, 5, 15], because, in this case, the observation matrix collects trajectories of feature point projections, that in general enter and leave the visible field, during the image sequence to process.

Using the result of the previous section, we propose an iterative scheme to compute the missing entries of observation matrices like the one above, in such a way to minimize

the $(r+1)^{\text{th}}$ singular value: alternate between the closed-form solutions for each of the Young diagrams. This means iterating the two following minimizations:

$$\min \sigma_{r+1} \left(\begin{bmatrix} ? & ? & \dots & ? & \times & \times & \dots & \times \\ ? & ? & \dots & ? & \times & \times & \dots & \times \\ ? & \dots & ? & \times & \times & \times & \dots & \times \\ ? & \dots & ? & \times & \dots & \times & \checkmark^i & \checkmark^i \\ ? & ? & \times & \dots & \times & \checkmark^i & \dots & \checkmark^i \\ ? & ? & \times & \dots & \times & \checkmark^i & \dots & \checkmark^i \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ ? & \times & \dots & \times & \checkmark^i & \checkmark^i & \dots & \checkmark^i \\ \times & \dots & \times & \checkmark^i & \checkmark^i & \checkmark^i & \dots & \checkmark^i \end{bmatrix} \right),$$

$$\min \sigma_{r+1} \left(\begin{bmatrix} \checkmark^{i+1} & \dots & \checkmark^{i+1} & \times & \times & \dots & \times \\ \checkmark^{i+1} & \dots & \checkmark^{i+1} & \times & \times & \dots & \times \\ \checkmark^{i+1} & \checkmark^{i+1} & \times & \times & \times & \dots & \times \\ \checkmark^{i+1} & \checkmark^{i+1} & \times & \dots & \times & ? & ? \\ \checkmark^{i+1} & \times & \dots & \times & ? & \dots & ? \\ \checkmark^{i+1} & \times & \dots & \times & ? & \dots & ? \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \checkmark^{i+1} & \times & \dots & \times & ? & \dots & ? \\ \times & \dots & \times & ? & ? & \dots & ? \end{bmatrix} \right),$$

where \checkmark^i and \checkmark^{i+1} represent generic values for the entries that were computed at the previous iteration. The pattern of missing entries to be completed in the first minimization is precisely the Young diagram studied in the previous section; the pattern of the second one is easily brought to this same format by performing adequate column and row permutations.

Due to the global optimality of each of the steps above, this iterative algorithm yields a non-increasing sequence of values for $\sigma_{r+1}(\mathbf{X}_i)$, where $\mathbf{X}_i \in \mathcal{S}_W$ is the completion of \mathbf{W} at iteration i . However, as with any other greedy algorithm, there is no guarantee of convergence to the global minimum, and initialization is a crucial issue. Naturally, for other patterns of missing entries, different block-wise optimal iterative minimization schemes can be used, *e.g.*, alternating among the completion of more than two Young diagrams.

6. Experiments

Since our method to complete rank constrained matrices, with patterns of missing entries in a Young diagram, runs in a finite number of steps and computes the optimal completion, our experiments focused in illustrating its usefulness in practice, rather than in emphasizing its optimality. In subsection 6.1, we illustrate the use of our algorithm in a structure from motion application. Subsection 6.2 addresses robustness to noise. Finally, in subsection 6.3, we

report the behavior of the suboptimal iterative scheme proposed in Section 5.

6.1. Application – structure from motion

We have generated synthetic trajectories of 2D projections of a set of 3D points, according to an affine camera model. To simulate occlusion, *i.e.*, points that disappear along the image sequence, several trajectories were ended at different points, well before the end of the synthetic sequence, thus made incomplete on purpose. Figure 2 illustrates the scenario.

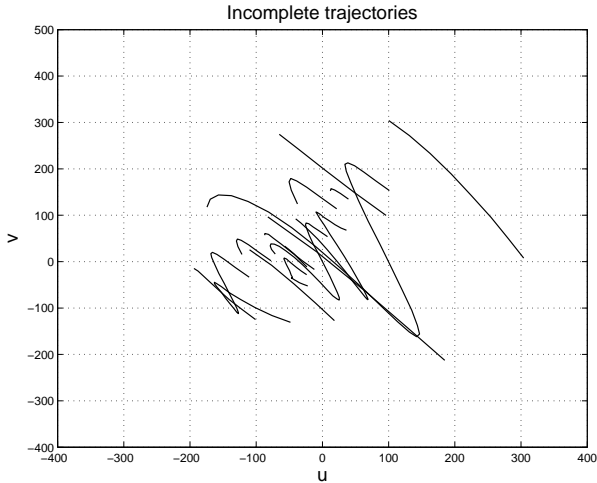


Figure 2. Incomplete trajectories of (projections of) feature points, observed when recovering 3D structure from motion. These 2D coordinates of the feature point projections are collected in an incomplete observation matrix.

Due to the rigidity of the set of 3D points, the observation matrix, which collects the 2D projections, is rank 4 in a noiseless situation, see, *e.g.*, [18], for details. When the trajectories are incomplete, the observation matrix misses several entries and the challenge in robustly recovering rigid structure from motion is computing its optimal completion, according to the rank constraint, see, *e.g.*, [13, 10, 14, 17]. We thus used our approach to complete this kind of rank constrained observation matrices. Note that the patterns of missing entries that model occlusion, *i.e.*, patterns leading to incomplete trajectories such as the ones illustrated in Figure 2, are easily arranged in a Young-diagram format, such as the one illustrated in expression (20), by performing adequate row and/or column permutations.

As discussed in the paper, our method computes the values of the missing entries that globally minimize the spectral norm of the matrix of residuals, or, in this case, the completion that has the 5th singular value, σ_5 , as small as possible. We have tested the proposed completion procedure with several scenarios: number of feature points and number of images from 10 to 100; different patterns of missing

entries, whose percentages varied from 20% to 80%; several sets of 3D point locations, within a sphere that spanned the simulated field of view; and several values for camera parameters. The impact of the noise level is discussed in the following subsection.

In all the experiments, we obtained complete trajectories very similar to the ground truth. Figure 3 illustrates this evidence for the data in Figure 2, which corresponds to a scenario with 30 images, 15 points, and $\simeq 60\%$ missing data. The top image shows the complete trajectories that were recovered by our algorithm. They are visually indistinguishable from the ground truth trajectories, *i.e.*, the trajectories that would be observed if there was no occlusion, which are shown in the bottom image of Figure 3. Naturally, by processing complete trajectories it is possible to recover more accurate 3D models than the ones that would be obtained by using suboptimal strategies to deal with incomplete ones.

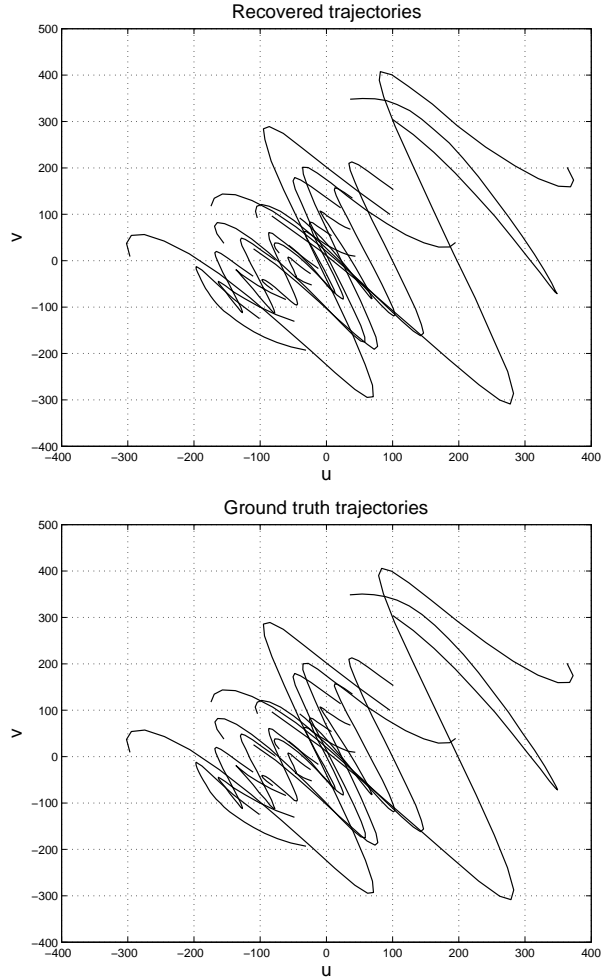


Figure 3. Top image: estimate of the complete trajectories that are partially observed in Figure 2. These were recovered by using our algorithm to compute the optimal completion of the corresponding observation matrix. Bottom image: Ground truth.

6.2. Robustness to noise

Our algorithm computes the matrix completion that globally minimizes the spectral norm of the matrix of residuals. However, the reader may wonder if this completion is close to the one minimizing the Frobenius norm, which is the cost function that arises when the problem is formulated as a maximum likelihood estimation, under white gaussian noise. To make this comparison, since there is not a globally optimal algorithm to minimize the Frobenius norm, we performed several experiments that consisted in running our algorithm with white gaussian noisy data and then providing its output as an initialization to an iterative (locally) optimal refinement of the Frobenius norm, *e.g.* [9, 10].

The first evidence we got is that the refinement algorithms, when initialized by our spectrally optimal completion, converged to the global minimum of the Frobenius norm, in few iterations, at every run (although we do not know this minimum *a priori*, we may guess it was correctly computed by comparing the final value of different runs with the same data and making sure the error was small enough when compared to the noise level). Naturally, this did not happen when the refinement was initialized in a different way, *e.g.*, with a random guess, which lead frequently to local minima of the Frobenius norm.

Furthermore, we noticed that the completion provided by our algorithm consisted in a very good approximation to the one minimizing the Frobenius norm. Figure 4, obtained by performing 1000 runs with noisy observations of a 40×30 rank 4 matrix with 50% missing entries, plots the approximation errors (mean square error of the entries of the estimated rank constrained matrix), as functions of the noise standard deviation. We used signal-to-noise ratios typical of what can be expected in structure from motion (the values of the matrix entries are in $[-100, 100]$ and the noise standard deviation ranges from 0 to 2).

We conclude that, although our completion was derived as being optimal according to the minimization of the spectral norm of the residuals, *i.e.*, the minimization of σ_5 in the case of the experiments of Figure 4, it is simultaneously a very good approximation to the maximum likelihood estimate, in a white gaussian noise observation model. In fact, the entry-wise mean square error of our completion (solid line in the plot of Figure 4) is almost indistinguishable from the one obtained by refining the estimate through the locally optimal refinement of the Frobenius norm (dashed line).

6.3. Young-wise iterative optimization

We now illustrate the behavior that can be expected of iterative algorithms based on the block-wise decomposition suggested in Section 5. Inspired by the structure from motion application, we used matrices with patterns of missing entries composed by two Young diagrams, such as the one

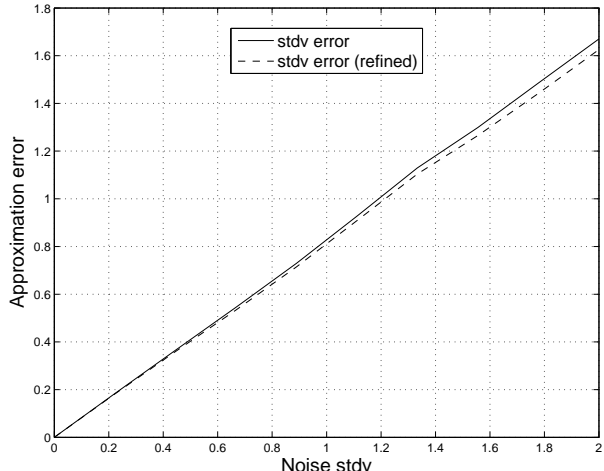


Figure 4. Mean square error of the estimated rank deficient matrix, from incomplete noisy observations. The (global) spectrally optimal solution provided by our algorithm (solid line) is almost indistinguishable from the one that (locally) minimizes the mean square error (dashed line).

illustrated in Section 5. We then performed the alternate completion that, at each step, minimized σ_5 with respect to one of the Young-diagram blocks of missing entries. The plots in Figure 5 represent the evolution of the cost, *i.e.*, the spectral norm of the residuals, *i.e.*, σ_5 , across the iterations, for two typical situations.

The plots in Figure 5 show that, as expected, the spectral norm of the residuals, measured by σ_5 , decreases along the iterations. However, one of two very distinct situations may happen: while, in the top plot, the spectral norm of the residuals stabilizes at a high value, in the bottom one, it converges to a very small error (note that, in both plots, the ordinate at which the small circle is placed corresponds to the true value of the 5th singular value of the (complete, noiseless) matrix).

Naturally, this behavior is due to the only block-wise optimality of the alternate completion. In fact, although each step computes the optimal completion for the corresponding block, the overall algorithm may lead to a local minimum of the spectral norm, as it happens in the top plot of Figure 5. Depending on the initialization, the overall algorithm may also converge to the global optimum, as in the bottom plot. In all our experiments, we randomly initialized the process, *i.e.*, we just filled the unknown entries in one of the Young diagrams with random guesses.

7. Conclusion

We have presented an efficient approach to the computation of the optimal factorization of an incomplete observation matrix. The only restriction is that the pattern of missing entries must be described by a so-called Young diagram.

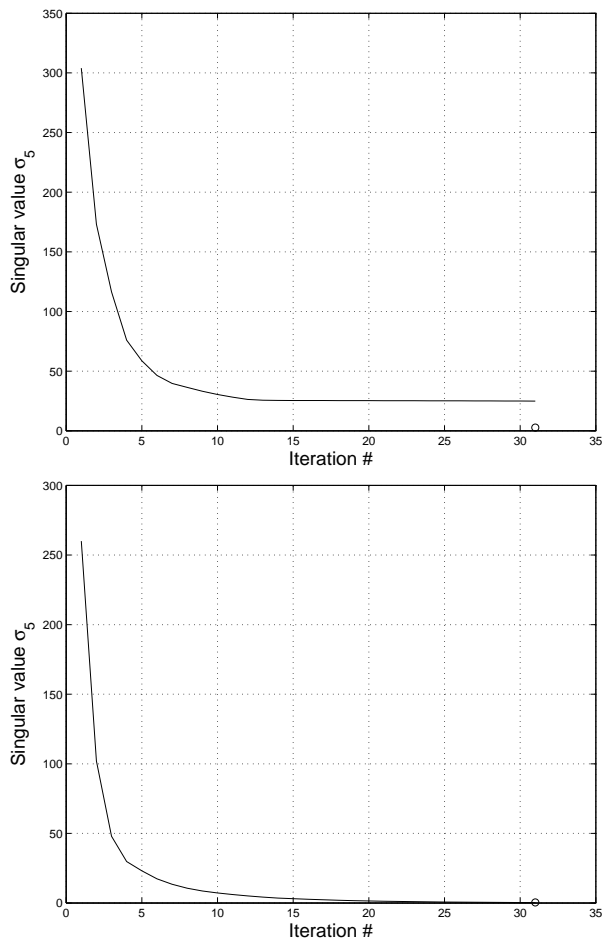


Figure 5. Typical behaviors of block-wise iterative optimization. The plots show the non-increasing evolution of the cost (*i.e.*, the singular value σ_5 of the incomplete matrix) along the iterations. While the top plot illustrates the convergence to a local minimum, the bottom one shows a successful experiment (the circle marks the singular value σ_5 of the noiseless (complete) matrix).

Our algorithm finds the optimal completion, in the sense of minimizing the spectral norm of the residuals, by filling in the unknown entries in a sequential way.

References

- [1] P. Aguiar, M. Stosic, and J. Xavier. On singular values of partially prescribed matrices. ISR/IST Technical Report, 2007. [3](#), [4](#), [5](#)
- [2] R. Basri, D. Jacobs, and I. Kemelmacher. Photometric stereo with general, unknown lighting. *Int. Journal of Computer Vision*, 72(3):239–257, 2007. [1](#)
- [3] P. Belhumeur and D. Kriegman. What is the set of images of an object under all possible lighting conditions? In *Proc. of IEEE Int. Conf. on Computer Vision and Pattern Recognition*, San Francisco, CA, USA, 1996. [1](#)
- [4] C. Bregler, A. Hertzmann, and H. Biermann. Recovering non-rigid 3d shape from image streams. In *Proc. of IEEE Int. Conf. on Computer Vision and Pattern Recognition*, Hilton Head Island, South Carolina, USA, 2000. [1](#)
- [5] A. Buchanan and A. Fitzgibbon. Damped newton algorithms for matrix factorization with missing data. In *Proc. of IEEE Int. Conf. on Computer Vision and Pattern Recognition*, San Diego CA, USA, 2005. [1](#), [2](#), [5](#)
- [6] G. Dullerud and F. Paganini. *A Course in Robust Control Theory: A Convex Approach*. Number 36 in Texts in Applied Mathematics. Springer-Verlag, New York, 2000. [3](#)
- [7] W. Freeman and J. Tenenbaum. Learning bilinear models for two-factor problems in vision. In *IEEE Int. Conference on Computer Vision and Pattern Recognition*, Puerto Rico, USA, 1997. [1](#)
- [8] G. Golub and C. V. Loan. *Matrix Computations*. The Johns Hopkins University Press, 1996. [3](#)
- [9] R. Guerreiro and P. Aguiar. Estimation of rank deficient matrices from partial observations: two-step iterative algorithms. In *Energy Min. Meth. in Computer Vision and Pattern Recognition*, number 2683 in Lecture Notes in Computer Science. Springer-Verlag, 2003. [1](#), [2](#), [5](#), [7](#)
- [10] R. Hartley and F. Schaffalitzky. Powerfactorization: an approach to affine reconstruction with missing and uncertain data. In *Australia-Japan Advanced Workshop on Computer Vision*, Adelaide, Australia, 2003. [1](#), [5](#), [6](#), [7](#)
- [11] R. Horn and C. Johnson. *Matrix Analysis*. Cambridge University Press, Cambridge, UK, 1985. [3](#), [5](#)
- [12] M. Irani. Multi-frame correspondence estimation using subspace constraints. *Int. Journal of Computer Vision*, 48(3):173–194, 2002. [1](#)
- [13] D. Jacobs. Linear fitting with missing data: Applications to structure-from-motion and to characterizing intensity images. In *Proc. of IEEE Int. Conf. on Computer Vision and Pattern Recognition*, 1997. [1](#), [5](#), [6](#)
- [14] D. Martinec and T. Pajdla. 3D reconstruction by fitting low-rank matrices with missing data. In *Proc. of IEEE Int. Conf. on Computer Vision and Pattern Recognition*, San Diego, USA, 2005. [1](#), [6](#)
- [15] T. Okatani and K. Deguchi. On the Wiberg algorithm for matrix factorization in the presence of missing components. *Int. Journal of Computer Vision*, 72(3):329–337, May 2007. [1](#), [5](#)
- [16] P. Sturm and B. Triggs. A factorization based algorithm for multi-image projective structure and motion. In *Proc. of European Conf. on Computer Vision*, Cambridge, UK, 1996. [1](#), [5](#)
- [17] J. Tardif, A. Bartoli, M. Trudeau, N. Guilbert, and S. Roy. Algorithms for batch matrix factorization with application to structure-from-motion. In *IEEE Int. Conf. on Computer Vision and Pattern Recognition*, Minneapolis, USA, 2007. [1](#), [6](#)
- [18] C. Tomasi and T. Kanade. Shape and motion from image streams under orthography: a factorization method. *Int. Journal of Computer Vision*, 9(2), 1992. [1](#), [5](#), [6](#)
- [19] S. Ullman and R. Basri. Recognition by linear combinations of models. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 13(10), 1991. [1](#)