

Bilinear Factorization via Augmented Lagrange Multipliers^{*}

Alessio Del Bue¹, João Xavier², Lourdes Agapito³, and Marco Paladini³

¹ Istituto Italiano di Tecnologia, Via Morego 30, 16163 Genova, Italy

² ISR - Instituto Superior Técnico, Av. Rovisco Pais, Lisbon, Portugal

³ Queen Mary University of London, Mile end road, E1 4NS London, UK

Abstract. This paper presents a unified approach to solve different bilinear factorization problems in Computer Vision in the presence of missing data in the measurements. The problem is formulated as a constrained optimization problem where one of the factors is constrained to lie on a specific manifold. To achieve this, we introduce an equivalent reformulation of the bilinear factorization problem. This reformulation decouples the core bilinear aspect from the manifold specificity. We then tackle the resulting constrained optimization problem with Bilinear factorization via Augmented Lagrange Multipliers (BALM). The mechanics of our algorithm are such that only a projector onto the manifold constraint is needed. That is the strength and the novelty of our approach: it can handle seamlessly different Computer Vision problems. We present experiments and results for two popular factorization problems: Non-rigid Structure from Motion and Photometric Stereo.

1 Introduction

Many inference problems in Computer Vision fit the form of bilinear problems since often observations are influenced by two independent factors where each can be described by a linear model. For instance, in photometric stereo [2] the shape of the object and the light source direction interact bilinearly to influence the image intensity. In rigid structure from motion [16] the 3D shape of the object is pre-multiplied by the camera matrix to determine its image coordinates. In facial tracking the problem of separating head pose and facial expression can also be defined as a bilinear problem [1]. In non-rigid structure from motion [4] the 2D coordinates of features arise from a bilinear relation between the camera matrix and the time varying shape. The interaction between two factors has also been generalised to several problems [15] in a learning context. In all these

^{*} This research has received funding from the European Research Council under the European Community's Seventh Framework Programme (FP7/2007-2013) ERC Starting Grant agreement 204871-HUMANIS. This work was also partially supported by FCT, under ISR/IST plurianual funding (POSC program, FEDER), grant SIPM-PTDC/EEA-ACR/73749/2006, grant MODI-PTDC/EEA-ACR/72201/2006. We thank J. Buenaposada for providing an implementation of [2].

problems the objective is to make inferences about both factors – the goal is their simultaneous estimation.

In this paper, we present a unified approach to solve different bilinear factorization problems in computer vision. The problem is formulated as a constrained optimization problem where one of the factors is constrained to lie on a specific manifold. Our key observation is that the main difference between different factorization problems is the manifold on which the solution lies. Thus, intuitively, it should be possible to construct a unified optimization framework in which a change of the manifold constraint just implies replacing an inner module of the algorithm (as opposed to an overall redesign of the optimization method from scratch). In this paper, we propose such a modular approach. To achieve this, we start by introducing an equivalent reformulation of the bilinear factorization problem. In loose terms, the reformulation decouples the core bilinear aspect from the manifold specificity. We then tackle the resulting constrained optimization problem via an Augmented Lagrange Multipliers (ALM) iterative algorithm. The mechanics of our algorithm are such that only a projector onto the manifold constraint is needed. That is the strength and the novelty of our approach: this framework can handle seamlessly different computer vision factorization problems. What will differ in each case is the projector of the solution onto the correct manifold.

In our experiments we show that we are able to deal with high percentages of missing data which has the practical implication that our approach can be used on data coming from real and not just controlled scenarios. We illustrate our unified approach by applying it to solve two popular computer vision problems: Non-rigid Structure-from-Motion (NRSfM) and Photometric Stereo (PS). To the best of our knowledge, this paper constitutes the first attempt to propose a unified optimisation framework for large scale bilinear factorization problems with given manifold constraints on one of the factors and provide a practical algorithm that can deal with missing data in the measurements.

2 Related Work

Bilinear models appear frequently in Computer Vision. However, it is in the area of Structure from Motion (SfM) that most of the efforts dedicated to solve this problem have come from. The wealth of research in this area is such that we cannot give an exhaustive review of the literature. Instead we will focus on describing what we believe are the two most important threads of research to solve the problem of low-rank matrix factorization in the case of missing data.

One line of research that dominates the literature are approaches that perform alternation of closed-form solutions to solve for the two factors of the matrix. The first of these approaches to solve the problem of missing data was proposed by Wiberg [18]. Since then many different solutions have been put forward. Buchanan and Fitzgibbon [5] provide a comprehensive review of these methods while proposing their own alternative approach. Their Damped Newton algorithm provides faster and more accurate solutions than standard alternation approaches. The common property of all these methods is that they only solve the

low-rank matrix factorization problem without imposing manifold constraints. The constraints are applied afterwards, once the low-rank matrix has been estimated. Crucially, the constraints are not imposed during the minimization.

On the other hand, a relatively recent set of algorithms have attempted to solve the problem by including explicitly the non-linear constraints given by the specific problem structure in the low-rank minimization. Marques and Costeira [12] introduced the concept of *motion manifold* in rigid SfM to obtain motion matrices that exactly satisfy the camera constraints. Similarly, Paladini et al. [13] propose an alternation algorithm associated with an optimal projector onto the *motion manifold* of non-rigid shapes. The practical implication of their algorithm is that it can deal with very high percentages of missing data. Shaji *et al.* [14] also propose to solve a non-linear optimisation problem directly on the product manifold of the Special Euclidean Group claiming better results than [5] in a rigid real sequence.

However, all these approaches are tailored to specific problems. Therefore, for different manifold constraints an overall redesign of the optimization method would be needed. The purpose of our work is to present a generic approach that is not problem dependent. In similar spirit, Chandraker and Kriegman [6] have proposed a globally optimal bilinear fitting approach for general Computer Vision problems. The key contribution of their approach is that they can prove convergence to a global minimiser using a branch and bound approach. However, the main drawback is that they are restricted to simple bilinear problems where the number of variables in one of the sets must be very small (for instance just 9 variables in one of their examples).

Our Bilinear factorization via Augmented Lagrange Multipliers (BALM) is designed to deal with large-scale optimisation problems with the inclusion of non-linear constraints. Our approach is not the first one to adopt the Augmented Lagrangian Multipliers (ALM) framework in the Computer Vision or related contexts. In perspective 3D reconstruction [11] ALM was used to enforce constraints on the perspective depths. In [10] ALM is successfully employed as a single matrix imputation algorithm which can deal with large scale problems.

3 Problem Statement

We denote by $Y \in \mathbb{R}^{n \times m}$ the measurement matrix. In this paper, we consider the general case of missing data. We let the finite set $\mathcal{O} := \{(i, j) : Y_{ij} \text{ is observed}\}$ enumerate the indices of the entries of Y which are available. The bilinear factorization problem we address is the following constrained optimization problem:

$$\begin{aligned} & \text{minimize} \quad \sum_{(i,j) \in \mathcal{O}} (Y_{ij} - s_i^\top m_j)^2 \\ & \text{subject to} \quad M_i \in \mathcal{M}, \quad i = 1, \dots, f, \end{aligned} \quad (1)$$

where s_i^\top denotes the i th row of the matrix $S \in \mathbb{R}^{n \times r}$ and m_j denotes the j th column of the matrix $M = [M_1 \cdots M_i \cdots M_f] \in \mathbb{R}^{r \times m}$, $M_i \in \mathbb{R}^{r \times p}$.

The variable to optimize in (1) is (S, M) . Here, f is the number of frames and we consider throughout the paper that $n \geq r \geq p$. For instance in the structure

from motion problem S would be the 3D structure and M the camera matrices, in photometric stereo S would be the lighting parameters and M the surface normals and albedo.

In words, problem (1) consists in finding the best rank r factorization SM of Y , given the available entries enumerated by \mathcal{O} and subject to the constraints on M . More precisely, each submatrix $M_i \in \mathbb{R}^{r \times p}$ of M must belong to the manifold $\mathcal{M} \subset \mathbb{R}^{r \times p}$. Our aim in this paper is to construct an algorithm to solve problem (1) which takes advantage of the fact that the projector onto \mathcal{M} is available (easily implementable). That is, we assume that, for a given $A \in \mathbb{R}^{r \times p}$, it is known how to solve the projection problem onto \mathcal{M}

$$\begin{aligned} & \text{minimize } \|A - X\|^2, \\ & \text{subject to } X \in \mathcal{M} \end{aligned} \tag{2}$$

where $\|X\|$ denotes the Frobenius norm of X . In the sequel, we let $p_{\mathcal{M}}(A)$ denote a solution of (2).

Problem reformulation. Let us define a new set of variables $z := \{Z_{ij} : (i, j) \notin \mathcal{O}\}$. Think of them as representing the non-observed entries of Y . We can introduce these variables in (1) and obtain the following equivalent optimization problem

$$\begin{aligned} & \text{minimize } \|Y(z) - SM\|^2 \\ & \text{subject to } M_i \in \mathcal{M}, \quad i = 1, \dots, f, \end{aligned} \tag{3}$$

where the (i, j) entry of the matrix $Y(z)$ is defined as

$$(Y(z))_{ij} := \begin{cases} Y_{ij}, & \text{if } (i, j) \in \mathcal{O} \\ Z_{ij}, & \text{if } (i, j) \notin \mathcal{O} \end{cases}.$$

In words, $Y(z)$ is just Y where we fill-in the missing entries with z . Note that the variable to optimize in (3) is (z, S, M) . Problem (3) is equivalent to (1) because once we fix (S, M) in (3) and minimize over z we fall back into (1). Finally, we clone M into a new variable $N = [N_1 \cdots N_i \cdots N_f] \in \mathbb{R}^{r \times m}$, $N_i \in \mathbb{R}^{r \times p}$, and transfer the manifold constraint to the latter. By doing so, we roughly separate the bilinear issue from the manifold restriction. This is our final reformulation:

$$\begin{aligned} & \text{minimize } \|Y(z) - SM\|^2 \\ & \text{subject to } M_i = N_i, \quad i = 1, \dots, f \\ & \quad N_i \in \mathcal{M}, \quad i = 1, \dots, f. \end{aligned} \tag{4}$$

The variable to optimize in (4) is (z, S, M, N) .

4 The BALM Algorithm

The main difficulty in the constrained optimization problem (4) are the equality constrains $M_i = N_i$. We propose to handle them through an augmented Lagrangian approach, see [9,3] for details on this optimization technique. In our

context, the augmented Lagrangian corresponding to (4) is given by

$$L_\sigma(z, S, M, N; R) = \|Y(z) - SM\|^2 - \sum_{i=1}^f \text{tr} (R_i^\top (M_i - N_i)) + \frac{\sigma}{2} \sum_{i=1}^f \|M_i - N_i\|^2.$$

where $\sigma > 0$ is the weight of the penalty term and $R_i, i = 1, \dots, f$, denote Lagrange multipliers. We let $R = [R_1 \cdots R_f]$. The optimization problem (4) can then be tackled by our Bilinear factorization via Augmented Lagrange Multipliers (BALM) algorithm detailed in Algorithm 1.

Algorithm 1. Bilinear factorization via Augmented Lagrange Multipliers (BALM)

- 1: set $k = 0$ and $\epsilon_{\text{best}} = +\infty$
- 2: initialize $\sigma^{(0)}, R^{(0)}, \gamma > 1$ and $0 < \eta < 1$
- 3: initialize $z^{(0)}, S^{(0)}$ and $M^{(0)}$
- 4: **repeat**
- 5: solve

$$\left(z^{(k+1)}, S^{(k+1)}, M^{(k+1)}, N^{(k+1)} \right) = \underset{\text{subject to } N_i \in \mathcal{M}, \quad i = 1, \dots, f,}{\text{argmin}} \quad L_{\sigma^{(k)}}(z, S, M, N; R^{(k)}) \quad (5)$$

using the iterative Gauss-Seidel scheme described in **Algorithm 2**

- 6: compute $\epsilon = \left\| M^{(k+1)} - N^{(k+1)} \right\|^2$
 - 7: **if** $\epsilon < \eta \epsilon_{\text{best}}$
 - 8: $R^{(k+1)} = R^{(k)} - \sigma^{(k)} \left(M^{(k+1)} - N^{(k+1)} \right)$
 - 9: $\sigma^{(k+1)} = \sigma^{(k)}$
 - 10: $\epsilon_{\text{best}} = \epsilon$
 - 10: **else**
 - 10: $R^{(k+1)} = R^{(k)}$
 - 11: $\sigma^{(k+1)} = \gamma \sigma^{(k)}$
 - 12: **endif**
 - 13: update $k \leftarrow k + 1$
 - 14: **until** some stopping criterion
-

Regarding the initialization of the BALM algorithm, we used $\sigma^{(0)} = 50, R^{(0)} = 0, \gamma = 5$ and $\eta = 1/2$ in all our computer experiments. With respect to $z^{(0)}, S^{(0)}$ and $M^{(0)}$, we feel that there is no universally good method, that is, the structure of \mathcal{M} must taken into account. We discuss the initialization $(z^{(0)}, S^{(0)}, M^{(0)})$ for several examples in the experimental section of this paper.

Clearly, solving the inner problem (5) at each iteration of the BALM method is the main computational step. Note that in (5) the optimization variable is (z, S, M, N) ($\sigma^{(k)}$ and $R^{(k)}$ are constants). To tackle (5) we propose an iterative Gauss-Seidel scheme which is described in Algorithm 2. We now show that each of the subproblems (6), (7) and (8) inside the Gauss-Seidel scheme are easily solvable.

Algorithm 2. Iterative Gauss Seidel scheme to solve for (5)

1: set $l = 0$ and choose L_{\max}
 2: set $z^{[0]} = z^{(k)}$, $S^{[0]} = S^{(k)}$ and $M^{[0]} = M^{(k)}$

3: **repeat**

4: solve

$$N^{[l+1]} = \operatorname{argmin}_{N_i \in \mathcal{M}, i=1, \dots, f} L_{\sigma^{(k)}} \left(z^{[l]}, S^{[l]}, M^{[l]}, N; R^{(k)} \right) \quad (6)$$

5: solve

$$\left(S^{[l+1]}, M^{[l+1]} \right) = \operatorname{argmin}_{S, M} L_{\sigma^{(k)}} \left(z^{[l]}, S, M, N^{[l+1]}; R^{(k)} \right) \quad (7)$$

6: solve

$$z^{[l+1]} = \operatorname{argmin}_z L_{\sigma^{(k)}} \left(z, S^{[l+1]}, M^{[l+1]}, N^{[l+1]}; R^{(k)} \right) \quad (8)$$

7: update $l \leftarrow l + 1$

8: **until** $l = L_{\max}$

9: set $S^{(k+1)} = S^{[L_{\max}]}$, $M^{(k+1)} = M^{[L_{\max}]}$ and $N^{(k+1)} = N^{[L_{\max}]}$

4.1 Solving for (6)

It is straightforward to see (details omitted) that (6) decouples into f projections onto the manifold of constraints \mathcal{M} . More precisely, if we partition

$$N^{[l+1]} = \left[N_1^{[l+1]} \dots N_i^{[l+1]} \dots N_f^{[l+1]} \right] \in \mathbb{R}^{r \times m}, \quad N_i^{[l+1]} \in \mathbb{R}^{r \times p},$$

the solution of (6) is given by

$$N_i^{[l+1]} = p_{\mathcal{M}} \left(M_i^{[l]} - \frac{1}{\sigma^{(k)}} R_i^{(k)} \right), \quad i = 1, \dots, f.$$

We recall that $p_{\mathcal{M}}$ stands for the projector onto \mathcal{M} , see (2), which we assume is available. This is the only part of the algorithm where the constraint manifold \mathcal{M} plays a role. Thus, replacing \mathcal{M} amounts to replace the projector $p_{\mathcal{M}}$. This is the modularity which we alluded to previously.

4.2 Solving for (7)

To simplify notation in this subsection, we let $Y = Y(z^{[l]})$, $N = N^{[l+1]}$, $\sigma = \sigma^{(k)}$ and $R = R^{(k)}$. Solving (7) corresponds to solving

$$\operatorname{minimize} \|Y - SM\|^2 + \frac{\sigma}{2} \sum_{i=1}^f \left\| M_i - \left(N_i + \frac{1}{\sigma} R_i \right) \right\|^2.$$

Equivalently, in terms of the new variable $\tilde{S} = \sqrt{\frac{2}{\sigma}}S$, we have

$$\text{minimize } \left\| \tilde{Y} - \tilde{S}M \right\|^2 + \|C - M\|^2 \tag{9}$$

where $\tilde{Y} = \sqrt{\frac{2}{\sigma}}Y$ and $C = N + \frac{1}{\sigma}R$.

Now, any full row rank matrix $M \in \mathbb{R}^{r \times m}$ can be represented as $M = AQ^\top$ where $A \in \mathbb{R}^{r \times r}$ is nonsingular and $Q \in \mathbb{R}^{m \times r}$ is a Stiefel matrix ($Q^\top Q = I_r$). Plugging this representation into (9) produces the optimization problem

$$\begin{aligned} &\text{minimize } \left\| \tilde{Y} - \tilde{S}AQ^\top \right\|^2 + \|C - AQ^\top\|^2 \\ &\text{subject to } Q^\top Q = I_r \end{aligned} \tag{10}$$

with optimization variable $(\tilde{S}, A, Q) \in \mathbb{R}^{n \times r} \times \mathbb{R}^{r \times r} \times \mathbb{R}^{m \times r}$.

Introducing the new variable $\hat{S} = \tilde{S}A$, transforms (10) into

$$\begin{aligned} &\text{minimize } \left\| \begin{bmatrix} \tilde{Y} \\ C \end{bmatrix} - \begin{bmatrix} \hat{S} \\ A \end{bmatrix} Q^\top \right\|^2 \\ &\text{subject to } Q^\top Q = I_r \end{aligned} \tag{11}$$

For a given Q in (11), the optimal (\hat{S}, A) is

$$\begin{bmatrix} \hat{S} \\ A \end{bmatrix} = \begin{bmatrix} \tilde{Y} \\ C \end{bmatrix} Q$$

which, when plugged back into (11), leaves the maximization problem

$$\begin{aligned} &\text{maximize } \text{tr} \left(Q^\top \begin{bmatrix} \tilde{Y} \\ C \end{bmatrix}^\top \begin{bmatrix} \tilde{Y} \\ C \end{bmatrix} Q \right) \\ &\text{subject to } Q^\top Q = I_r \end{aligned} \tag{12}$$

which can be solved through an eigenvalue decomposition ($\text{tr}(X)$ denotes the trace of the square matrix X). Problem (7) optimizes jointly over (S, M) . An alternative approach is to replace (7) by two least-squares problems: one over M (for fixed S) and the other over S (for fixed M).

4.3 Solving for (8)

After solving for $N^{[l+1]}$ and $(S^{[l+1]}, M^{[l+1]})$, problem (8) updates the missing data. The solution of (8) is trivial: we just have to take $Z_{ij}^{[l+1]}$ as the (i, j) th entry of $S^{[l+1]}M^{[l+1]}$ for all $(i, j) \notin \mathcal{O}$.

Algorithm convergence. At best, the BALM algorithm can produce a local minimizer for (1). That is, we do not claim that BALM (algorithm 1) converges to a global minimizer. In fact, even the nonlinear Gauss-Seidel technique (algorithm 2) is not guaranteed to globally solve (5). This is the common situation when dealing with nonconvex problems. See [7] for some convergence results on augmented Lagrangian methods. We now apply our generic BALM algorithm to solve two different bilinear computer vision problems: the Non-Rigid Structure-from-Motion problem (NRSfM) and the Photometric Stereo (PS) problem.

5 Example 1: BALM for Non-rigid SfM

The problem of recovering the non-rigid 3D shape of a deforming object from a monocular video sequence given only 2D correspondences between frames was formulated as a factorization problem by Bregler *et al.* [4] in the case of an orthographic camera. The assumption is that the 3D shape can be represented as a linear combination of a set of d basis shapes with time varying coefficients t_{id} . If the image coordinates are referred to the centroid, the projection of the shape at frame i can be expressed as

$$Y_i = \begin{bmatrix} u_{i1} & \dots & u_{in} \\ v_{i1} & \dots & v_{in} \end{bmatrix}^\top = \left(\sum_{l=1}^d t_{il} B_l \right) Q_i = [B_1 \dots B_d] (t_i \otimes Q_i) = SM_i \quad (13)$$

where Y_i is the $n \times 2$ measurement matrix that contains the 2D coordinates of n image points in frame i , B_l are the basis shapes of size $n \times 3$ and t_{il} are the time varying shape coefficients and Q_i is the projection matrix for frame i , which in the case of an orthographic camera is a 3×2 matrix that encodes the first two columns of a rotation matrix (therefore it is a Stiefel matrix). By stacking all the measurements for all the frames into a single matrix we have

$$Y = [B_1 \dots B_d] [t_1 \otimes Q_1 \dots t_f \otimes Q_f] = S [M_1 \dots M_f] = SM. \quad (14)$$

Now, we have expressed the measurement matrix as a bilinear interaction between the shape matrix S of size $n \times 3d$ and the motion matrix M of size $3d \times 2f$. This form fits exactly the optimisation problem as presented in Eq. (1). Therefore, in the NRSfM case, the manifold constraint corresponds to $\mathcal{M} = \{t \otimes Q : t \in \mathbb{R}^d, Q \in \mathbb{R}^{3 \times 2}, Q^\top Q = I_2\}$, or in other words, the two rows of the rotation matrix Q^\top must be orthonormal (i.e. it is a Stiefel matrix). To apply our BALM algorithm, we need first to derive the projector onto \mathcal{M} . We now turn to this problem.

5.1 NRSfM Manifold Projector

In [13] Paladini *et al.* derived an exact globally optimal algorithm to project the motion matrices onto the non-rigid motion manifold. However, here, we discuss an alternative which provides an *approximate* projector onto \mathcal{M} . The advantage is that our proposed algorithm still provides accurate estimates while being considerably faster (approximately 100 times in experimental tests).

Let $A \in \mathbb{R}^{3d \times 2}$ be given, and consider the partition $A = [A_1^\top A_2^\top \dots A_d^\top]^\top$, where $A_i \in \mathbb{R}^{3 \times 2}$. We want to compute $p_{\mathcal{M}}(A)$, that is, we want to solve the optimization problem

$$\begin{aligned} & \text{minimize} \quad \|A - t \otimes Q\|^2. \\ & \text{subject to} \quad t \in \mathbb{R}^d \\ & \quad \quad \quad Q^\top Q = I_2 \end{aligned} \quad (15)$$

If $A \in \mathcal{M}$, that is, if $A = t \otimes Q$ for some $t = (t_1, \dots, t_d)^\top \in \mathbb{R}^d$ and Stiefel matrix Q , then we would have the identity $\sum_{i=1}^d A_i A_i^\top = \|t\|^2 Q Q^\top$. That is,

the left-hand side of the equation would reveal the underlying Q (up to a right multiplication by a 2×2 rotation). This motivates the following approach, for a generic A : compute $\mathcal{A} = \sum_{i=1}^d A_i A_i^\top$ and estimate Q as its dominant Stiefel, that is, corresponding to the 2 top eigenvectors of \mathcal{A} . Let \widehat{Q} denotes this Stiefel matrix. Even when $A \in \mathcal{M}$ we do not have, in general, $\widehat{Q} = Q$. Rather, $\widehat{Q} = QR$ for a 2×2 orthogonal matrix R . Thus, we return to problem (15) and we solve for $t \in \mathbb{R}^d$ and R :

$$\begin{aligned} & \text{minimize} \quad \left\| A - t \otimes (\widehat{Q}R) \right\|^2 \\ & \text{subject to} \quad t \in \mathbb{R}^d \\ & \quad \quad \quad R \in \mathbb{R}^{2 \times 2}, R^\top R = I_2 \end{aligned} \tag{16}$$

For a fixed R , the optimal t is $t_i = \frac{1}{2} \text{tr} \left(R^\top \widehat{Q}^\top A_i \right)$ with $i = 1, \dots, d$. Plugging t_i into (16) gives the reduced problem over R

$$\begin{aligned} & \text{maximize} \quad \sum_{i=1}^d (\text{tr} (R^\top T_i))^2 \\ & \text{subject to} \quad R^\top R = I_2 \end{aligned} \tag{17}$$

where $T_i := \widehat{Q}^\top A_i$. Now, a 2×2 rotation matrix R must fall into one of the two following cases $\det(R) = 1$ or $\det(R) = -1$. In both cases the solution is similar. If we have $\det(R) = \pm 1$ we have that $R = \begin{bmatrix} c \mp s & \\ s \pm c \end{bmatrix}$ for some $(c, s) \in \mathbb{R}^2$, $\|(c, s)\| = 1$. Using this representation in (17) yields

$$\begin{aligned} & \text{maximize} \quad [c \ s] \left(\sum_{i=1}^d \begin{bmatrix} T_i(1, 1) + T_i(2, 2) \\ T_i(2, 1) - T_i(1, 2) \end{bmatrix} \begin{bmatrix} T_i(1, 1) \pm T_i(2, 2) \\ T_i(2, 1) \mp T_i(1, 2) \end{bmatrix}^\top \right) \begin{bmatrix} c \\ s \end{bmatrix} \\ & \text{subject to} \quad \|(c, s)\| = 1 \end{aligned} \tag{18}$$

which can be solved by an eigenvalue decomposition. After examining separately the two cases (i.e. \pm), we pick the best.

6 Example 2: BALM for Photometric Stereo

Basri et al. [2] derived a bilinear approximation of the image brightness given by luminance variations. This derivation is based on a spherical harmonics representation of lighting variations and it allows to frame PS as a factorization problem with manifold constraints on one of the bilinear factors. Given a set of images of a Lambertian object with varying illumination, it is possible to extract the dense normal to the surface of the object z , the albedo ρ and the lighting directions l . For a 1st order spherical harmonics approximation, the brightness at image pixel j at frame i can be modelled as $Y_{ij} = l_i^\top \rho_j [1 \ z_j^\top]^\top = S_i M_j$, where $l_i \in \mathbb{R}^4$, $\rho_j \in \mathbb{R}$, $z_j \in \mathbb{R}^3$ with $z_j^\top z_j = 1$. A compact matrix form can be obtained for each pixel Y_{ij} as:

$$Y = \begin{bmatrix} Y_{11} & \dots & Y_{1n} \\ \vdots & \ddots & \vdots \\ Y_{f1} & \dots & Y_{fn} \end{bmatrix} = \begin{bmatrix} l_1^\top \\ \vdots \\ l_f^\top \end{bmatrix} \left[\rho_1 \begin{bmatrix} 1 \\ z_1 \end{bmatrix} \dots \rho_n \begin{bmatrix} 1 \\ z_n \end{bmatrix} \right] = SM \tag{19}$$

where a single image i is represented by the vector $Y_i = [y_{i1} \dots y_{in}]$. Manifold constraints are given by the surface normal constraints which implies $M^{4 \times f}$ lying on the manifold defined by: $\mathcal{M} = \{\rho [1 \ z^\top]^\top : \rho \in \mathbb{R}, z \in \mathbb{R}^3, z^\top z = 1\}$. The matrix $S^{n \times 4}$ now contains the collection of lighting directions which which combines the first-order spherical harmonics at each frame i .

6.1 Photometric Stereo Manifold Projector

We now derive the projector onto the manifold. That is, for a given $a \in \mathbb{R}^4$, we show how to solve the associated optimization problem

$$\begin{aligned} & \text{minimize} \quad \|a - \rho [1 \ z^\top]^\top\|^2. \\ & \text{subject to} \quad z^\top z = 1 \end{aligned} \quad (20)$$

The variable to optimize is $(\rho, z) \in \mathbb{R} \times \mathbb{R}^3$. Consider the partition $a = [\alpha \beta^\top]^\top$ with $\alpha \in \mathbb{R}$ and $\beta \in \mathbb{R}^3$. We can rewrite (20) as

$$\begin{aligned} & \text{minimize} \quad \rho^2 - \alpha\rho - \rho\beta^\top z. \\ & \text{subject to} \quad z^\top z = 1 \end{aligned} \quad (21)$$

We denote by (ρ^*, z^*) the solution of (21). If $\beta = 0$ then $\rho^* = \alpha/2$ and z^* can be any unit-norm vector. If $\beta \neq 0$ then optimizing (21) over z (for a fixed ρ) gives

$$z^* = \frac{\rho}{|\rho|} \frac{\beta}{\|\beta\|}. \quad (22)$$

Now, if $\alpha \geq 0$ (respectively $\alpha < 0$) then it is clear that $\rho^* \geq 0$ ($\rho^* < 0$). Inserting this constraint into (21) leaves a quadratic problem with an easy solution: $\rho^* = (\alpha + \|\beta\|)/2$ (respectively $\rho^* = (\alpha - \|\beta\|)/2$).

7 Experiments

7.1 Synthetic Experiments: NRSfM

In our synthetic experiments¹ we used a 3D motion capture sequence showing a deforming face, captured using a VICON system tracking a subject wearing 37 markers. The 3D points were then projected synthetically onto an image sequence 74 frames long using an orthographic camera. To test the performance we computed the relative 3D reconstruction error, which we defined as the Frobenius norm of the difference between the recovered 3D shape S and the ground truth 3D shape S_{GT} , computed as: $\|S - S_{GT}\|/\|S_{GT}\|$. We subtract the centroid of each shape and align them with Procrustes analysis. In the experiments with noise, zero mean additive Gaussian noise was applied with standard deviation $\sigma = n \times s/100$ where n is the noise percentage and s is defined as $\max(Y)$ in pixels. In all experiments the number of basis shapes was fixed to $k = 5$. The trials for each level of noise were averaged over 10 runs.

¹ For additional experiments, videos and the code please check:

<http://www.isr.ist.utl.pt/~adb/the-balm/>

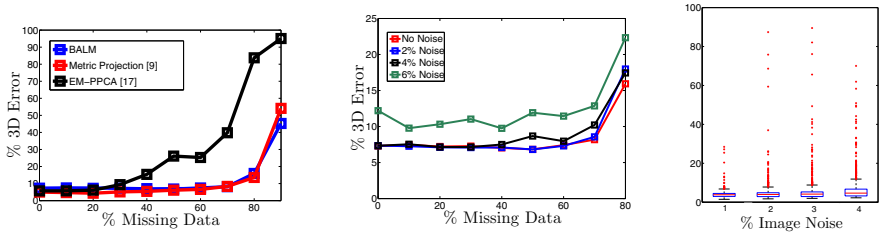


Fig. 1. Synthetic experiment results showing comparison with NRSfM methods (left), robustness of the BALM method with different ratios of missing data and noise (centre) and a boxplot for the convergence of the algorithm for increasing noise levels (right)

Figure 1 (left) shows a comparison between our proposed BALM algorithm and two state of the art methods: Torresani *et al.*'s [17] EM-PPCA and Metric Projections (MP) [13] in the absence of noise and for increasing levels of missing data. While in the case of full data the performance of the algorithms is comparable, BALM and MP clearly outperform EM-PPCA in all cases of missing data higher than 30%. Notice that BALM, has a similar performance to MP, which is a specific NRSfM algorithm. Another interesting fact is the extreme resilience of BALM to missing data. In Figure 1 (centre) we evaluate BALM's performance with respect to noise in the image measurements of up to 6% and missing data ratios of up to 90% in a combined test. The plot shows robustness to noise even for increasing levels of missing data. We also performed a set of convergence tests, in the full data case, for varying levels of noise to evaluate convergence empirically. In this case we used a synthetically generated 3D shape to ensure a known global minimum. Figure 1 (right) shows a boxplot of the 3D error, on the vertical axis, for 1000 trials of the BALM algorithm for each level of noise and no missing data. The algorithm achieves an overall median error close to zero for all the noise levels. Most of the experiments are between 0% and 20% 3D error with very few local minima reaching higher errors.

Regarding run times, in an experiment with 60% missing data, the convergence time was 16s for BALM, 10s for EM-PPCA [17] and 10m for MP [13]. Although the runtimes for BALM and EM-PPCA are comparable, BALM systematically outperforms EM-PPCA in the case of missing data (see Figure 1 (left)). All implementations are in Matlab. However, EM-PPCA runs with partial MEX code while BALM is not optimised. For runtime evaluation we used a Desktop PC AMD X2 2.6Ghz with 4GB of RAM.

Regarding the initialization of the BALM algorithm, the projection matrices, the mean shape and the missing data tracks are first initialised using [12]. We then used Torresani *et al.*'s initialisation [17] to estimate the configuration weights and the basis shapes given the residual of the first rigid solution.

7.2 Real Data: NRSfM

We tested our method on a real sequence of a cushion being bent. We tracked 90 points and we simulated a missing data ratio of 40% by eliminating data points

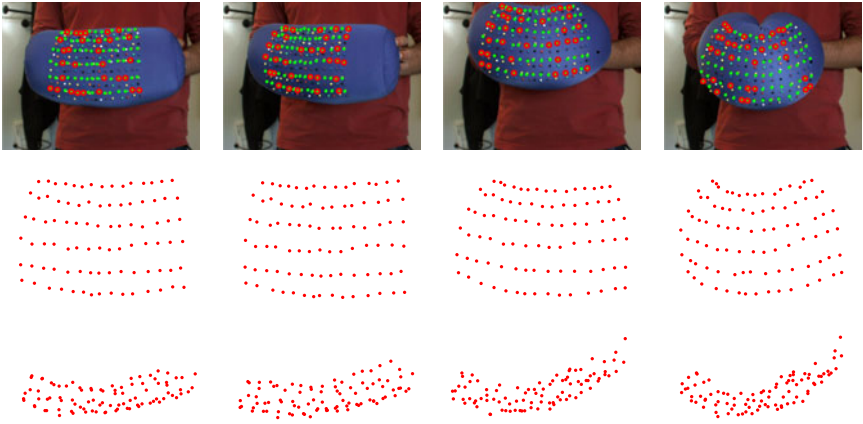


Fig. 2. Cushion sequence with 40% missing data. First row shows four image samples with missing points highlighted with a red circle. Second and third rows show a frontal and side view of the 3D reconstruction using BALM.

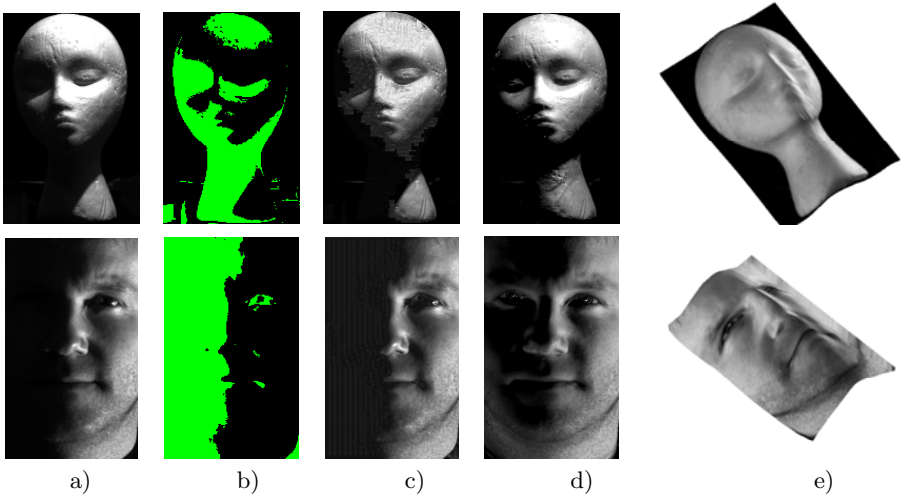


Fig. 3. Real photometric stereo results for the 46-frames long Sculpture sequence with 48% missing data (1st row) and 49-frames long YaleB10 sequence with 31% missing data (2nd row). Each image shows in a) a selected original frame from the sequence; in b) an image mask where green pixels represents missing entries; in c) the image used as initialisation for BALM; in d) the resulting optimised image with a Lambertian model and in e) the reconstructed 3D surface.

manually. Figure 2 shows 4 selected frames and their respective 3D reconstructions (frontal and top view). The bending is clearly observable in the 3D shape plots where BALM shows robustness given the high percentage of missing data.

7.3 Real Data: Photometric Stereo

We present results for the BALM factorization using the photometric projector as presented in Section 6.1. The aim is to extract the 3D surface, albedo and luminance parameters with significant occlusions in the input image data. The occlusions are defined as the darkest and brightest (saturated) pixels in the image sequence since in these areas the Lambertian model will not be satisfied. An initialisation for the missing entries in Y is given by the inpainting technique [8] which fills the image holes given the known parts of the image. In such a way we exploit image pixels which may resemble the current illumination in the image. The initialisation for $(S^{(0)}, M^{(0)})$ is then given by a simple rank-4 SVD on the “inpainted” Y . These affine low-rank components are then normalized as in [2] to comply with the equal norm constraints of the spherical harmonics model. Figure 3 shows the results for two image frames selected from the sequence. The first row of Figure 3 (Sculpture sequence) shows that even with large occlusions the initialisation with inpainting copes well if the overall texture of the object is quite homogenous (although errors can still be noticed on the top of the head and the darkest areas). Note that the final optimised image (Figure 3(d)) reveals some further details in the neck area which were hidden in the original frame. The sequence on the second row, taken from the YaleB database sequence, shows an extreme occlusion in which half of face of the the subject is not visible. In this case inpainting clearly fails to provide a usable initialisation but still the reconstructed shape resembles the subject.

8 Conclusions

We have provided a novel and general optimisation framework for a broad range of bilinear problems in Computer Vision with manifold constraints on the space where the data lies. Our results match state of the art methods in NRSfM and show a considerable improvement in performance for the PS problem. Our approach can deal with a number of entries in the data matrix Y in the order of 10^6 and more. This feature, together with the robustness to missing data, render the BALM algorithm a preferable choice for bilinear modelling in large-scale inference scenarios.

References

1. Bascle, B., Blake, A.: Separability of pose and expression in facial tracking and animation. In: Proc. 6th International Conference on Computer Vision, Bombay, India, pp. 323–328 (1998)
2. Basri, R., Jacobs, D., Kemelmacher, I.: Photometric stereo with general, unknown lighting. *International Journal of Computer Vision* 72(3), 239–257 (2007)
3. Bertsekas, D.: *Constrained optimization and Lagrange multiplier methods*. Academic Press, New York (1982)
4. Bregler, C., Hertzmann, A., Biermann, H.: Recovering non-rigid 3D shape from image streams. In: Proc. IEEE Conference on Computer Vision and Pattern Recognition, Hilton Head, South Carolina, June 2000, pp. 690–696 (2000)

5. Buchanan, A.M., Fitzgibbon, A.: Damped newton algorithms for matrix factorization with missing data. In: Proc. IEEE Conference on Computer Vision and Pattern Recognition, San Diego, California, vol. 2, pp. 316–322 (2005)
6. Chandraker, M., Kriegman, D.: Globally optimal bilinear programming for computer vision applications. In: IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2008, pp. 1–8 (2008)
7. Conn, A., Gould, N., Sartenaer, A., Toint, P.: Convergence properties of an augmented Lagrangian algorithm for optimization with a combination of general equality and linear constraints. *SIAM Journal on Optimization* 6(3), 674–703 (1996)
8. Criminisi, A., Pérez, P., Toyama, K.: Region filling and object removal by exemplar-based image inpainting. *IEEE Trans. on Image Processing* 13(9), 1200–1212 (2004)
9. Hestenes, M.: Multiplier and gradient methods. *Journal of Optimization Theory and Applications* 4(5), 303–320 (1969)
10. Lin, Z., Chen, M., Wu, L., Ma, Y.: The Augmented Lagrange Multiplier Method for Exact Recovery of Corrupted Low-Rank Matrices. UIUC Technical Report UILU-ENG-09-2215 (2009)
11. Mai, F., Hung, Y.S.: Augmented lagrangian-based algorithm for projective reconstruction from multiple views with minimization of 2d reprojection error. *Journal of Signal Processing Systems* (2009)
12. Marques, M., Costeira, J.P.: Estimating 3D shape from degenerate sequences with missing data. In: *Computer Vision and Image Understanding* (2008)
13. Paladini, M., Del Bue, A., Stosic, M., Dodig, M., Xavier, J., Agapito, L.: Factorization for Non-Rigid and Articulated Structure using Metric Projections. In: Proc. IEEE Conference on Computer Vision and Pattern Recognition, Miami, Florida, pp. 2898–2905 (2009)
14. Shaji, A., Chandran, S., Suter, D.: Manifold Optimisation for Motion Factorisation. In: 19th International Conference on Pattern Recognition (ICPR 2008), pp. 1–4 (2008)
15. Tenenbaum, J., Freeman, W.: Separating style and content with bilinear models. *Neural Computation* 12(6), 1247–1283 (2000)
16. Tomasi, C., Kanade, T.: Shape and motion from image streams under orthography: A factorization approach. *International Journal of Computer Vision* 9(2) (1992)
17. Torresani, L., Hertzmann, A., Bregler, C.: Non-rigid structure-from-motion: Estimating shape and motion with hierarchical priors. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 878–892 (2008)
18. Wiberg, T.: Computation of principal components when data are missing. In: COMPSTAT 1976, Proc. Comput. Stat., 2nd Symp., Berlin, pp. 229–236 (1976)