Accurate Optimization of Weighted Nuclear Norm for Non-Rigid Structure from Motion

N.B. When citing this work, cite the original published paper.

(article starts on next page)

Accurate Optimization of Weighted Nuclear Norm for Non-Rigid Structure from Motion

José Pedro Iglesias[1], Carl Olsson[1,2], and Marcus Valtonen Örnhag[2]

[1] Chalmers University of Technology, Sweden
[2] Lund University, Sweden

Abstract. Fitting a matrix of a given rank to data in a least squares sense can be done very effectively using 2nd order methods such as Levenberg-Marquardt by explicitly optimizing over a bilinear parameterization of the matrix. In contrast, when applying more general singular value penalties, such as weighted nuclear norm priors, direct optimization over the elements of the matrix is typically used. Due to non-differentiability of the resulting objective function, first order subgradient or splitting methods are predominantly used. While these offer rapid iterations it is well known that they become inefficent near the minimum due to zig-zagging and in practice one is therefore often forced to settle for an approximate solution.

In this paper we show that more accurate results can in many cases be achieved with 2nd order methods. Our main result shows how to construct bilinear formulations, for a general class of regularizers including weighted nuclear norm penalties, that are provably equivalent to the original problems. With these formulations the regularizing function becomes twice differentiable and 2nd order methods can be applied. We show experimentally, on a number of structure from motion problems, that our approach outperforms state-of-the-art methods.[1]

1 Introduction

Matrix recovery problems of the form

$$\min_X f(\boldsymbol{\sigma}(X)) + \|\mathcal{A}X - b\|^2, \tag{1}$$

where $\mathcal{A}$ is a linear operator and $\boldsymbol{\sigma}(X) = (\sigma_1(X), \sigma_2(X), ...)$ are the singular values of X, are frequently occurring in computer vision. Applications range from high level 3D reconstruction problems to low level pixel manipulations $[39, 7, 46, 17, 3, 16, 42, 12, 31, 23, 18]$. In structure from motion (SfM) the most

common approaches enforce a given low rank r without additionally penalizing non-zero singular values [39, 7, 20] (a special case of (1) by letting f assign zero if fewer than r singular values are non-zero and infinity otherwise).

Since the rank of a matrix X is bounded by to the number of columns/rows in a bilinear parameterization $X = BC^T$, the resulting optimization problem can be written $\min_{B,C} \|\mathcal{A}(BC^T) - b\|^2$. This gives a smooth objective function and can therefore be optimized using 2nd order methods. In SfM problems, where the main interest is the extraction of camera matrices from B and 3D points from C, this is typically the preferred option [8]. In a series of recent papers Hong *et al.* showed that optimization with the VarPro algorithm is remarkably robust to local minima converging to accurate solutions [20–22]. In [24] they further showed how uncalibrated rigid SfM with a proper perspective projection can be solved within a factorization framework. On the downside, typically the iterations are costly since (even when the Schur complement trick is used) 2nd order methods require an inversion of a relatively large hessian matrix, which may hinder application when suitable sparsity patterns are not present.

For low level vision problems such as denoising and inpainting, eg. [31, 23, 18], the main interest is to recover the elements of X and not the factorization. In this context more general regularization terms that also consider the size of the singular values are often used. Since the singular values are non-differentiable functions of the elements in X first order methods are usually employed. The simplest option is perhaps a splitting methods such as ADMM [6] since the proximal operator $\arg\min_X f(\sigma(X)) + \|X - X_0\|^2$, can often be computed in closed form [23, 18, 31, 14, 27]. Alternatively, subgradient methods can be used to handle the non-differentiability of the regularization term [12].

It is well known that while first order methods have rapid iterations and make large improvements the first couple of iterations they have a tendency to converge slowly when approaching the optimum. For example, [6] recommends to use ADMM when a solution in the vicinity of the optimal point is acceptable, but suggests to switch to a higher order method when high accuracy is desired. For low level vision problems where success is not dependent on achieving an exact factorization of a particular size, first order methods may therefore be suitable. In contrast, in the context of SfM, having roughly estimated elements in X causes the obtained factorization B, C to be of a much larger size than necessary yielding poor reconstructions with too many deformation modes.

In this paper we aim to extend the class of methods that can be optimized using bilinear parameterization allowing accurate estimation of a low rank factorization from a general class of regularization terms. While our theory is applicable for many objectives we focus on weighted nuclear norm penalties since these have been successfully used in SfM applications. We show that these can be optimized with 2nd order methods which significantly increases the accuracy of the reconstruction. We further show that with these improvements the model of Hong *et al.* [24] can be extended to handle non-rigid reconstruction with a proper perspective model, as opposed to the orthographic projection model adopted by other factorization based approaches, e.g. [27, 14, 17, 46].

1.1   Related Work and Contributions

Minimization directly over X has been made popular since the problem is convex when f is convex and absolutely symmetric, that is, $f(|x|) = f(x)$ and $f(\Pi x) = f(x)$, where Π is any permutation [29]. Convex penalties are however of limited interest since they generally prefer solutions with many small non-zero singular values to those with few large ones. A notable exception is the nuclear norm [15, 35, 33, 10, 11] which penalizes the sum of the singular values. Under the RIP assumption [35] exact or approximate low rank matrix recovery can then be guaranteed [35, 11]. On the other hand, since the nuclear norm penalizes large singular values, it suffers from a shrinking bias [9, 12, 28].

An alternative approach that unifies bilinear parameterization with regularization approaches is based on the observation [35] that the nuclear norm $\|X\|_*$ of a matrix X can be expressed as $\|X\|_* = \min_{BC^T = X} \frac{\|B\|_F^2 + \|C\|_F^2}{2}$. Thus when $f(\boldsymbol{\sigma}(X)) = \mu \sum_i \sigma_i(X)$, where μ is a scalar controlling the strength of the regularization, optimization of (1) can be formulated as

$$\min_{B,C} \mu \frac{\|B\|_F^2 + \|C\|_F^2}{2} + \|\mathcal{A}BC^T - b\|^2. \tag{2}$$

Optimizing directly over the factors has the advantages that the number of variables is much smaller and the objective function is two times differentiable so second order methods can be employed. While (2) is non-convex because of the bilinear terms, the convexity of the nuclear norm can still be used to show that any local minimizer B,C with $\text{rank}(BC^T) < k$, where k is the number of columns in B and C, is globally optimal [2, 19]. The formulation (2) was for vision problems in [9]. In practice it was observed that the shrinking bias of the nuclear norm makes it too weak to enforce a low rank when the data is noisy. Therefore, a "continuation" approach where the size of the factorization is gradually reduced was proposed. While this yields solutions with lower rank, the optimality guarantees no longer apply. Bach $et\ al.$ [2] showed that

$$\|X\|_{s,t} := \min_{X=BC^T} \sum_{i=1}^k \frac{\|B_i\|_s^2 + \|C_i\|_t^2}{2}, \tag{3}$$

where B_i,C_i are the ith columns of B and C respectively, is convex for any choice of vector norms $\|\cdot\|_s$ and $\|\cdot\|_t$. In [19] it was shown that a more general class of 2-homogeneous factor penalties result in a convex regularization similar to (3). The property that a local minimizer B, C with $\text{rank}(BC^T) < k$, is global is also extended to this case. Still, because of convexity, it is clear that these formulations will suffer from a similar shrinking bias as the nuclear norm.

One way of reducing shrinking bias is to use penalties that are constant for large singular values. Shang $et\ al.$ [37] showed that penalization with the Schatten semi-norms $\|X\|_q = \sqrt[q]{\sum_{i=1}^N \sigma_i(X)^q}$, for $q = 1/2$ and $2/3$, can be achieved using a convex penalty on the factors B and C. A generalization to general values of q is given in [44]. An algorithm that address a general class of penalties for

symmetric matrices is presented in [26]. In [40] it was shown that if f is given by $f(\boldsymbol{\sigma}(X)) = \sum_i g(\sigma_i(X))$, where g is differentiable, concave and non-decreasing then (1) can be optimized using 2nd order methods such as Levenberg-Marquart or VarPro. This is achieved by re-parameterizing the matrix X using a bilinear factorization $X = BC^T$ and optimizing

$$\min_{B,C} f(\boldsymbol{\gamma}(B,C)) + \|\mathcal{A}(BC^T) + b\|^2. \tag{4}$$

Here $\boldsymbol{\gamma}(B,C) = (\gamma_1(B,C), \gamma_2(B,C), ...)$ and $\gamma_i(B,C) = \frac{\|B_i\|^2 + \|C_i\|^2}{2}$. In contrast to the singular value $\sigma_i(X)$ the function $\gamma_i(B,C)$ is smooth which allows optimization with second order methods. It is shown in [40] that if X^* is optimal in (1) then the factorization $B = L\sqrt{\Sigma}, C = R\sqrt{\Sigma}$, where $X^* = L\Sigma R^T$ is the SVD of X^*, is optimal in (4). (Here we assume that L is $m \times r$, Σ is $r \times r$ and R is $n \times r$, with $\operatorname{rank}(X) = r$.) Note also that this choice gives $\gamma_i(B,C) = \sigma_i(X^*)$.

A less restrictive way of reducing bias is to re-weight the nuclear norm and use $f(\boldsymbol{\sigma}(X)) = \sum_i a_i \sigma_i(X)$ [23, 18, 27]. Assigning low weights to the first (largest) singular values allows accurate matrix recovery. In addition the weights can be used to regularize the size of the non-zero singular values which has been shown to be an additional useful prior in NRSfM [27]. Note however that the singular values are always ordered in non-increasing order. Therefore, while the function is linear in the singular values it is in fact non-convex and non-differentiable in the elements of X whenever the singular values are not distinct (typically the case in low rank recovery).

In this paper we show that this type of penalties allow optimization with $\boldsymbol{\gamma}(B,C)$ instead of $\boldsymbol{\sigma}(X)$. In particular we study the optimization problem

$$\min_{B,C} f(\boldsymbol{\gamma}(B,C)) \tag{5}$$
$$\text{s.t.} \quad BC^T = X, \tag{6}$$

and its constraint set for a fixed X. We characterize the extreme-points of the feasible set using permutation matrices and give conditions on f that ensure that the optimal solution is of the form $\boldsymbol{\gamma}(B,C) = \Pi\boldsymbol{\sigma}(X)$, where Π is a permutation. For the weighted nuclear norm $f(\boldsymbol{\sigma}(X)) = a^T \boldsymbol{\sigma}(X)$ we show that if the elements of a are non-decreasing the optimal solution has $\boldsymbol{\gamma}(B^*, C^*) = \boldsymbol{\sigma}(X)$. A simple consequence of this result is that

$$\min_{B,C} a^T \boldsymbol{\gamma}(B,C)) + \|\mathcal{A}(BC^T) - b\|^2 \tag{7}$$

is equivalent to $\min_X a^T \boldsymbol{\sigma}(X) + \|\mathcal{A}X - b\|^2$. While the latter is non-differentiable the former is smooth and can be minimized efficiently with second order methods.

Our experimental evaluation confirms that this approach outperforms current first order methods in terms of accuracy as can be expected. On the other hand first order methods make large improvments the first coupler of iterations and therefore we combine the two approaches. We start out with a simple ADMM implementation and switch to our second order approach when only minor progress is being made. Note however that since the original formulation is non-convex

local minima can exist. In addition bilinear parameterization introduces additional stationary points that are not present in the original X parameterization. One such example is $(B, C) = (0, 0)$, where all gradients vanish. Still our experiments show that the combination of these methods often converge to a good solution from random initialization.

2   Bilinear Parameterization Penalties

In this section we will derive a dependence between the singular values $\sigma_i(X)$ and the $\gamma_i(B, C)$, when $BC^T = X$. For ease of notation we will suppress the dependence on X and (B, C) since this will be clear from the context. Let X have the SVD $X = R\Sigma L^T$, $B = R\sqrt{\Sigma}$ and $C = L\sqrt{\Sigma}$. We will study other potential factorizations $X = \hat{B}\hat{C}^T$ using $\hat{B} = BV$, $\hat{C} = CH$ and $VH^T = I_{r \times r}$. In this section we will further assume that V is a square $r \times r$ matrix and therefore H^T is its inverse. (We will generalize the results to the rectangular case in Section 3).

We begin by noting that $\gamma_j = \frac{\|\hat{B}_j\|^2 + \|\hat{C}_j\|^2}{2} = \frac{\|BV_j\|^2 + \|CH_j\|^2}{2}$, where V_j and H_j are columns j of V and H respectively. We have $\|\hat{B}V_j\|^2 = V_j^T B^T B V_j = V_j^T \Sigma V_j = \|\sqrt{\Sigma}V_j\|^2$, and similarly $\|CH_j\|^2 = \|\sqrt{\Sigma}H_j\|^2$ and therefore $\gamma_j = \frac{\|\sqrt{\Sigma}V_j\|^2 + \|\sqrt{\Sigma}H_j\|^2}{2}$. This gives $\gamma_j = \left(\frac{\sigma_1(v_{1j}^2 + h_{1j}^2) + \sigma_2(v_{2j}^2 + h_{2j}^2) + \ldots + \sigma_r(v_{rj}^2 + h_{rj}^2)}{2} \right)$, or in matrix form

$$
\begin{pmatrix} \gamma_1 \\ \gamma_2 \\ \vdots \\ \gamma_r \end{pmatrix} = \frac{1}{2} \underbrace{\begin{pmatrix} v_{11}^2 & v_{21}^2 & \ldots & v_{r1}^2 \\ v_{12}^2 & v_{22}^2 & \ldots & v_{r2}^2 \\ \vdots & \vdots & \ddots & \vdots \\ v_{1r}^2 & v_{2r}^2 & \ldots & v_{rr}^2 \end{pmatrix}}_{=V^T \odot V^T} \begin{pmatrix} \sigma_1 \\ \sigma_2 \\ \vdots \\ \sigma_r \end{pmatrix} + \frac{1}{2} \underbrace{\begin{pmatrix} h_{11}^2 & h_{21}^2 & \ldots & h_{r1}^2 \\ h_{12}^2 & h_{22}^2 & \ldots & h_{r2}^2 \\ \vdots & \vdots & \ddots & \vdots \\ h_{1r}^2 & h_{2r}^2 & \ldots & h_{rr}^2 \end{pmatrix}}_{=H^T \odot H^T} \begin{pmatrix} \sigma_1 \\ \sigma_2 \\ \vdots \\ \sigma_r \end{pmatrix}. \tag{8}
$$

Minimizing (5) over different factorizations is therefore equivalent to solving

$$
\min_{\gamma, M \in \mathcal{S}} f(\boldsymbol{\gamma}), \tag{9}
$$
$$
\text{s.t.} \quad \boldsymbol{\gamma} = M\boldsymbol{\sigma}. \tag{10}
$$

where $\mathcal{S} = \{\frac{1}{2}(V^T \odot V^T + H^T \odot H^T); \ VH^T = I\}$. It is clear that $V = H = \Pi^T$, where Π is any permutation, is feasible in the above problem since permutations are orthogonal. In addition they contain only zeros and ones and therefore it is easy to see that this choice gives $\boldsymbol{\gamma} = \frac{1}{2}(\Pi \odot \Pi + \Pi \odot \Pi)\boldsymbol{\sigma} = \Pi\boldsymbol{\sigma}$. In the next section we will show that these are extreme points of the feasible set, in the sense that they can not be written as convex combinations of other points in the set. Extreme points are important for optimization since the global minimum is guaranteed to be attained (if it exists) in such a point if the objective function has concavity properties. This is for example true if f is quasi-concave, that is, the super-level sets $S_\alpha = \{x \in \mathbb{R}_{\geq 0}^r; f(x) \geq \alpha\}$ are convex. To see this let $x = \lambda x_1 + (1 - \lambda)x_2$, and consider the super-level set S_α where $\alpha = \min(f(x_1), f(x_2))$. Since both $x_1 \in S_\alpha$ and $x_2 \in S_\alpha$ it is clear by convexity that so is x and therefore $f(x) \geq \min(f(x_1), f(x_2))$.

2.1   Extreme Points and Optimality

We now consider the optimization problem (9)-(10) and a convex relaxation of the constraint set. For this purpose we let $\mathcal{D}$ be the set of doubly stochastic matrices $\mathcal{D} = \{M \in \mathbb{R}^{r \times r};\ m_{ij} \geq 0,\ \sum_i m_{ij} = 1,\ \sum_j m_{ij} = 1\}$. Note that if V is orthogonal, and therefore $H = V$, then the row sum $\sum_{j=1}^{r} \frac{v_{ij}^2 + h_{ij}^2}{2}$, and the column sum $\sum_{i=1}^{r} \frac{v_{ij}^2 + h_{ij}^2}{2}$ are both one. Hence such a matrix is in $\mathcal{D}$. To handle non-orthogonal matrices we define the set of superstochastic matrices $\mathcal{S}_W$ as all matrices $M = D + N$, where $D \in \mathcal{D}$ and N is a matrix with non-negative elements. It can be shown that (see Theorem 6 in [4]) that $\mathcal{S} \subset \mathcal{S}_W$. In addition it is easy to see that $\mathcal{S}_W$ is convex since it consists of affine constraints. Therefore the problem

$$\min_{\gamma, M \in \mathcal{S}_W} f(\boldsymbol{\gamma}), \tag{11}$$

$$\text{s.t.} \qquad \boldsymbol{\gamma} = M\boldsymbol{\sigma}. \tag{12}$$

is a relaxation of (9)-(10). Next we show that the two problems have the same minimum if a minimizer to (11)-(12) exists when f is quasi-concave (on $\mathbb{R}_{\geq 0}^r$). As mentioned previously, the minimum (over $\mathcal{S}_W$) is then attained in an extreme point of $\mathcal{S}_W$. We therefore need the following characterization.

Lemma 1. *The extreme points of $\mathcal{S}_W$ are $r \times r$ permutation matrices.*

Proof. First we note that any extreme point of $\mathcal{S}_W$ has to be in $\mathcal{D}$ since if $M = D + N$ with $N \neq 0$ then $M = \frac{1}{2}D + \frac{1}{2}(D + 2N)$, which is a convex combination of two points in $\mathcal{S}_W$. By Birkhoff's Theorem [5] any matrix in $\mathcal{D}$ can be written as a convex combination of permutation matrices.

Since permutation matrices are orthogonal with $0/1$ elements it is clear they can be written $\Pi = \frac{1}{2}(\Pi \odot \Pi + \Pi \odot \Pi)$, with $\Pi\Pi^T = I$. Therefore the extreme points of $\mathcal{S}_W$ are also in $\mathcal{S}$. Hence if the minimum of (11)-(12) is attained, there is an optimal extreme point of $\mathcal{S}_W$ which also solves (9)-(10), and therefore the solution is given by a permutation $V = H = \Pi$.

We conclude this section by giving sufficient conditions for the minimum of (11)-(12) to exist, namely that f is lower semi-continuous and non-decreasing in all of its variables, that is, if $\tilde{\gamma}_i \geq \gamma_i$ for all i then $f(\tilde{\boldsymbol{\gamma}}) \geq f(\boldsymbol{\gamma})$. Since the singular values are all positive it is clear that the elements of $(D + N)\boldsymbol{\sigma}$ are larger than those of $D\boldsymbol{\sigma}$. Hence when f is non-decreasing it is enough to consider minimization over $\mathcal{D}$. We then have a lower semi-continuous objective function on a compact set for which the minimum is known to be attained.

We can now summarize the results of this section in the following theorem:

Theorem 1. *Let f be quasi-concave (and lower semi-continuous) on $\mathbb{R}_{\geq 0}^r$ fulfilling $f(\tilde{\boldsymbol{\gamma}}) \geq f(\boldsymbol{\gamma})$ when $\tilde{\gamma}_i \geq \gamma_i$ for all i. Then there is an optimal $\boldsymbol{\gamma}^*$ of (9)-(10) that is of the form $\boldsymbol{\gamma}^* = \Pi\boldsymbol{\sigma}$ where Π is a permutation.*

3   Non-square Matrices

In the previous section we made the assumption that V and H where square matrices, which corresponds to searching over $\hat{B}$ and $\hat{C}$ consisting of r columns when rank$(X) = r$. In addition since V and H are invertible this means that $\hat{B}$ and $\hat{C}$ have linearly independent columns. In this section we generalize the result from Section 2.1 to rectangular matrices V and H. Therefore we suppose that V and H are non-square of size $r \times p$, $p > r$, with $VH^T = I_{r \times r}$, and consider the slightly modified problem

$$\min_{\gamma, V, H} f(\gamma), \tag{13}$$

$$\gamma = \frac{1}{2}(V^T \odot V^T + H^T \odot H^T)\sigma \tag{14}$$

$$VH^T = I_{r \times r} \tag{15}$$

Note that VH^T do not commute and we therefore only assume that V is a left inverse of H^T. In what follows we show that by adding zeros to the vector σ we can extend V, H into square matrices without changing the objective function.

Note that we may assume that V has full row rank since otherwise $X \neq BVH^TC^T$. Let $V^\dagger$ be the Moore-Penrose pseudo inverse and $O_{V_\perp}$ a $(p - r) \times p$ matrix containing a basis for the space orthogonal to the row space of V (and the column space of $V^\dagger$). Since $VH^T = I_{r \times r}$ the matrix H^T is of the form $H^T = V^\dagger + O_{V_\perp}^T K_1$, where K_1 is a $(p-r) \times r$ coefficient matrix. We now want to find matrices $\tilde{V}$ and $\tilde{H}$ such that $\begin{bmatrix} V \\ \tilde{V} \end{bmatrix} \begin{bmatrix} V^\dagger + O_{V_\perp}^T K_1 & \tilde{H}^T \end{bmatrix} = \begin{bmatrix} I_{r \times r} & 0 \\ 0 & I_{(p-r) \times (p-r)} \end{bmatrix}$. To do this we first select $\tilde{H}^T = O_{V_\perp}^T$ since $VO_{V_\perp}^T = 0$. Then we let $\tilde{V} = O_{V_\perp} + K_2 V$, where K_2 is a size $(p - r) \times r$ coefficient matrix, since this gives $\tilde{V}\tilde{H}^T = I_{(p-r) \times (p-r)}$. To determine K_2 we consider $\tilde{V}(V^\dagger + O_{V_\perp}^T K_1) = K_2 I_{r \times r} + I_{(p-r) \times (p-r)} K_1 = K_2 + K_1$. Selecting $K_2 = -K_1$ thus gives square matrices such that $\begin{bmatrix} V \\ \tilde{V} \end{bmatrix} \begin{bmatrix} H^T & \tilde{H}^T \end{bmatrix} = I$. Further letting $\tilde{\Sigma} = \begin{bmatrix} \Sigma & 0 \\ 0 & 0 \end{bmatrix}$ shows that $\|BV_i\| = \|\sqrt{\tilde{\Sigma}} \begin{bmatrix} V_i \\ \tilde{V}_i \end{bmatrix}\|$ and $\|CH_i\| = \|\sqrt{\tilde{\Sigma}} \begin{bmatrix} H_i \\ \tilde{H}_i \end{bmatrix}\|$ and the results of the previous section give that the minimizer of $f(\gamma_1, \gamma_2, ..., \gamma_p)$ is a permutation of the elements in the vector $(\sigma_1, \sigma_2, ..., \sigma_r, 0, ..., 0)$. We therefore have the following result:

Corollary 1. *Let f be quasi-concave (and lower semi-continuous) on $\mathbb{R}_{\geq 0}^p$ fulfilling $f(\tilde{\gamma}) \geq f(\gamma)$ when $\tilde{\gamma}_i \geq \gamma_i$ for all i. Then an optimizer γ^* of (13)-(15) is of the form $\gamma^* = \Pi_{p \times r}\sigma$ where $\Pi_{p \times r}$ contains the first r columns of a $p \times p$ permutation matrix.*

4   Linear Objectives - Weighted Nuclear Norms

We now consider weighted nuclear norm regularization $f(\gamma) = a^T\gamma$. To ensure that the problem is well posed we assume that the elements of a are non-negative.

It is then clear that $f(\tilde{\boldsymbol{\gamma}}) \geq f(\boldsymbol{\gamma})$ when $\tilde{\gamma}_i \geq \gamma_i$. Since linearity implies concavity the results of Sections 2.1 and 3 now show that the minimum of $f(M\boldsymbol{\sigma})$, over $M \in \mathcal{S}$ is attained in $M = \Pi$ for some permutation matrix. To ensure that the bilinear formulation is equivalent to the original one we need to show that the optimum occurs when $\Pi = I$. Suppose that the elements in a are sorted in ascending order $a_1 \leq a_2 \leq ... \leq a_p$. It is easy to see that for Π to give the smallest objective value it should sort the elements of γ so that $\gamma_1 \geq \gamma_2 \geq ... \geq \gamma_p$, which means that $\Pi = I$ and $\boldsymbol{\gamma} = \boldsymbol{\sigma}$. We therefore conclude that minimizing (4) with a linear objective corresponds to regularization with a weighted nuclear norm with non-decreasing weights.

5   Experiments

In this section we start by describing implementation details of our method and then apply it to the problems of low matrix recovery and non-rigid structure recovery. Solving the weighted nuclear norm regularized problem (7) now amounts to minimizing

$$\sum_{i=1}^{p} a_i \frac{\|B_i\|^2 + \|C_i\|^2}{2} + \|\mathcal{A}(BC^T) - b\|^2. \tag{16}$$

Note that the terms in the (16) can be combined into a single norm term by vertically concatenating the vectors B_i and C_i, weighted by $\sqrt{a_i/2}$, with $\mathcal{A}(BC^T) - b$. We define the resulting vector as $r_a := \mathcal{A}_a(BC^T) - b_a$, giving the objective $\|r_a(BC^T)\|^2$, where the subscript reflects the dependence on the weights a. Since the objective is smooth, standard methods such as Levenberg-Marquardt can be applied and Algorithm 1 shows an overview of the method used. Additional information about the algorithm is provided in the supplementary material.

The remainder of this section is organized as follows. The particular form of the data fitting term in (16) when applied to structure from motion is described in Section 5.1. In Section 5.2 we compare the convergence of first and second-order methods, and motivated by the ADMM fast iterations but low accuracy, as opposed to the bilinear parameterization's high accuracy but slower iterations, we combine the two methods by initializing the bilinear parameterization with the ADMM's solution [6, 27] for a non-rigid structure structure recovery problem. Our work focus on the increased accuracy of our method compared to first-order methods, so the comparison of our results with works such as [43, 45, 1, 34] (without the desired regularization term) are not covered.

5.1   Pseudo Object Space Error (pOSE) and Non-Rigid Structure from Motion

To compare the performance of 1st and 2nd order methods, we choose as objective function the Pseudo Object Space Error (pOSE) [24], which consists of a combination of the object space error $\ell_{\text{OSE}} := \sum_{(i,j)\in\Omega} \|P_{i,1:2}\tilde{\mathbf{x}}_j - (\mathbf{p}_{i,3}^T\tilde{\mathbf{x}}_j)\mathbf{m}_{i,j}\|_2^2$ and the affine projection error $\ell_{\text{Affine}} := \sum_{(i,j)\in\Omega} \|P_{i,1:2}\tilde{\mathbf{x}}_j - \mathbf{m}_{i,j}\|_2^2$, where $P_{i,1:2}$

Algorithm 1: Bilinear parameterization of weighted nuclear norm

---

Result: Optimal B,C to (16)

$B = U\Sigma^{\frac{1}{2}}$, $C = V\Sigma^{\frac{1}{2}}$, where $X = U\Sigma V^T$ is the ADMM solution to (16) ;

Choose initial $\alpha > 1$ and λ, and define $z = [\text{vec}(B);\ \text{vec}(C^T)]$;

Compute $error = \|r_a(BC^T)\|^2$;

while *not converged* **do**

    Compute $r = r_a(BC^T)$, and the jacobian J of (16) in terms of z;

    Update $\tilde{z} = z - (J^T J + \lambda I)^{-1} J^T r$, where $\tilde{z} = [\text{vec}(\tilde{B});\ \text{vec}(\tilde{C}^T)]$;

    if $error > \|r_a(\tilde{B}\tilde{C}^T)\|^2$ **then**

        | Updata $z \leftarrow \tilde{z}$, $\lambda \leftarrow \alpha^{-1}\lambda$, and $error \leftarrow \|\mathcal{A}_a(\tilde{B}\tilde{C}^T) - b_a\|^2$;

    else

        | $\lambda \leftarrow \alpha\lambda$;

    end

end

---

and $\mathbf{p}_{i,3}$ are, respectively, the first two and the third rows of the camera matrix P_i, with $i = 1, \ldots, F$; $\tilde{\mathbf{x}}_j$ is a 3D point in homogeneous coordinates, with $j = 1, \ldots, P$; $\mathbf{m}_{i,j}$ is the 2D observation of the j:th point on the i:th camera; and Ω represents the set of observable data. The pOSE is then given by $\ell_{\text{pOSE}} := (1-\eta)\ell_{\text{OSE}} + \eta\ell_{\text{Affine}}$ where $\eta \in [0, 1]$ balances the weight between the two errors. One of the main properties of pOSE is its wide basin of convergence while keeping a bilinear problem strucuture. The ℓ_{pOSE}, originally designed for rigid SfM, can be extended for the non-rigid case by replacing $P_i\tilde{\mathbf{x}}_j$ by a linear combination of K shape basis, i.e., $\Pi_i\hat{\mathbf{S}}_j$, where $\Pi_i \in \mathbb{R}^{3\times(3K+1)}$ and $\hat{\mathbf{S}}_j \in \mathbb{R}^{3K+1}$ are structured as $\Pi_i = \begin{bmatrix} c_{i,1}R_i & \ldots & c_{i,K}R_i & t_i \end{bmatrix}$ and $\hat{\mathbf{S}}_j = \begin{bmatrix} S_{1,j}^T & \ldots & S_{K,j}^T & 1 \end{bmatrix}^T$. We denote by Π and $\hat{\mathbf{S}}$ the vertical and horizontal concatenations of Π_i and $\hat{\mathbf{S}}_j$, respectively. Note that by construction $\text{rank}(\Pi\hat{\mathbf{S}}) \leq 3K + 1$, and for $K = 1$ we have $\Pi_i\hat{\mathbf{S}}_j = P_i\tilde{\mathbf{x}}_j$, which corresponds to the rigid case.

5.2   Low-Rank Matrix Recovery with pOSE errors

In this section we compare the convergence and accuracy of 1st and 2nd order methods, starting from the same initial guess, for low-rank matrix recovery with pOSE. In this problem, we define $X = \Pi\hat{\mathbf{S}}$ and aim at minimizing

$$\min_X a^T \boldsymbol{\sigma}(X) + \ell_{\text{pOSE}}(X). \tag{17}$$

We apply our method and solve the problem (16) by using the bilinear factorization $X = BC^T$, with $B \in \mathbb{R}^{3F\times r}$, and $C \in \mathbb{R}^{P\times r}$, with $r \geq 3K + 1$. We test the performance of our method in 4 datasets: Door [32], Back [36], Heart [38], Paper [41]. The first one consists of image measurements of a rigid structure with missing data, while the remaining three datasets track points in deformable structures.

For the Door dataset, we apply two different selections of weights on the singular values of X, corresponding to the nuclear norm, i.e., $a_i = \mu_{NN}$, and truncated nuclear norm, i.e., $a_i = 0$, $i = 1, \ldots, 4$ and $a_i = \mu_{TNN}$, $i > 4$. We select $\mu_{NN} = 1.5 \times 10^{-3}$, and $\mu_{TNN} = 1$. For the Back, Heart and Paper datasets, we apply the nuclear norm and a weighted nuclear norm, in which the first four singular values of X are not penalized and the remaining ones are increasingly penalized, i.e., $a_i = 0$, $i = 1, \ldots, 4$ and $a_i = (i-4)\mu_{WNN}$, $i > 4$. We select $\mu_{NN} = 7.5 \times 10^{-4}$, $\mu_{WNN} = 2.25 \times 10^{-3}$. The values of the weights a_i are chosen such that there is a $3K + 1$ rank solution to (17), with $K = 1$ and $K = 2$ for the rigid and non-rigid datasets, respectively.

We compare the bilinear parameterization with three first-order methods commonly used for low-rank matrix recovery: Alternating Direction Method of Multipliers (ADMM) [6], Iteratively Reweighted Nuclear Norm (IRNN) [13], and Accelerated Gradient Descend (AGD) [30]. We also test the methods for two different cases of the ℓ_{pOSE} error, with $\eta = 0.05$ and $\eta = 0.95$, which correspond to the near-perspective and near-affine camera models, respectively. To improve numerical stability of the algorithms, as pre-processing step we normalize the image measurements matrix M by its norm. The methods are initialized with the closed-form solution of the regularization-free problem, i.e., $X = \mathcal{A}^\dagger(b)$. The comparison of the four algorithms in terms of total log-loss over time is shown in Figure 1. The log-loss is used for better visualization purposes. The plots for the IRNN for the nuclear norm are omitted since it demonstrated slow convergence compared to the remaining three methods. A qualitative evaluation of the results on one of the images of the Door dataset for the truncated nuclear norm and near perspective camera model is shown in Figure 2. The qualitative results for the remaining datasets are provided in the supplementary material.

In general, we can observe that first-order methods demonstrate faster initial convergence, mostly due to faster iterations. However when near minima, the convergence rate drops significantly and the methods tend to stall. Contrarily, bilinear parameterization compensates its slower iterations by demonstrating higher accuracy and and reaching solutions with lower energy. This is specially visible for the near perspective camera model, which reinforces the advantages of using a second-order method on image data under perspective projection. To compensate for the slower convergence, we propose the initialization of the bilinear parameterization with the solution obtained with ADMM. In this way, the bilinear parameterization starts near the minimum and performs local refinement to further improve accuracy.

### 5.3   Non-Rigid Structure Recovery

Consider now that the camera rotations in Π are known (or previously estimated). In this case we have $\Pi\hat{S} = RX + t\mathbb{1}^T$, with $R = \text{blkdiag}(R_1, \ldots, R_F)$ and $t = [t_1^T, \ldots, t_F^T]^T$, where X, the non-rigid structure, and t are the unknowns. It is directly observed that $\text{rank}(\Pi\hat{S}) \leq \text{rank}(RX) + \text{rank}(t\mathbb{1}^T)$, with the later being equal to 1 by construction and independent on K. As consequence, it follows that $\text{rank}(RX) = \text{rank}(X) \leq 3K$, and the rank regularization can be applied
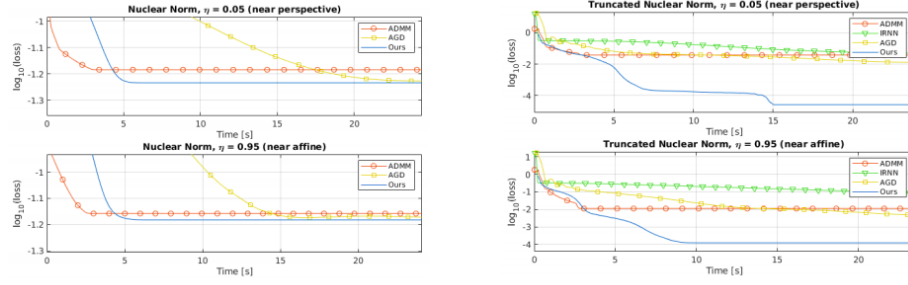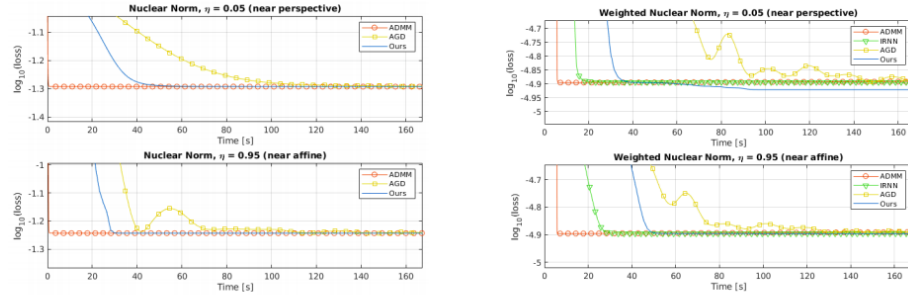
(a) Door Dataset.
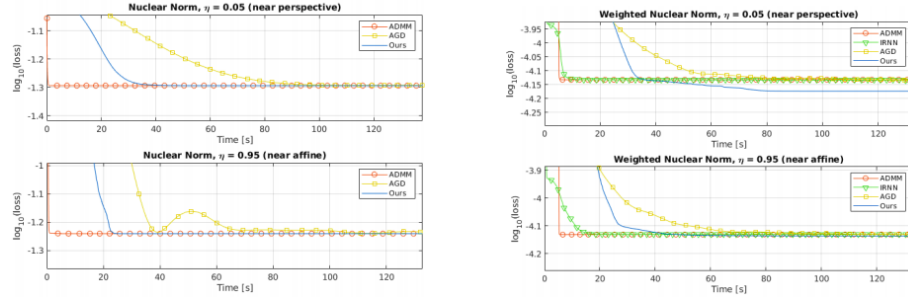
(b) Heart Dataset.
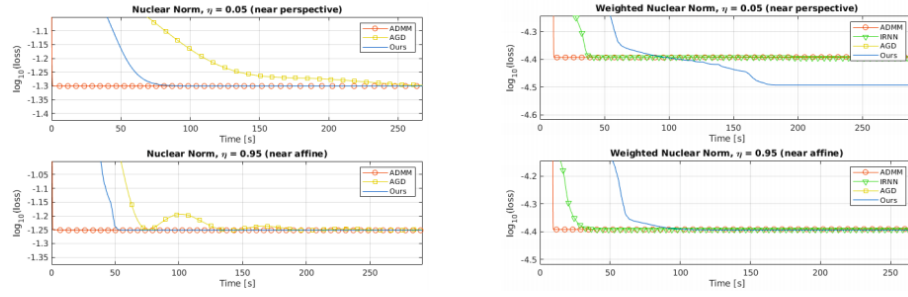
(c) Back Dataset.

(d) Paper Dataset.

Fig. 1. Convergence of the four methods for low-rank matrix recovery on the Door, Heart, Back and Paper datasets.
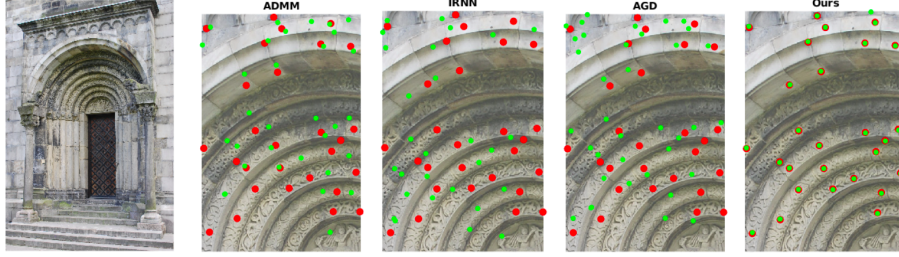
Fig. 2. Evaluation of the four methods for low-rank matrix recovery on one of the images of the Door dataset. The red circles show the target image measurements and the green circles the estimate image points.
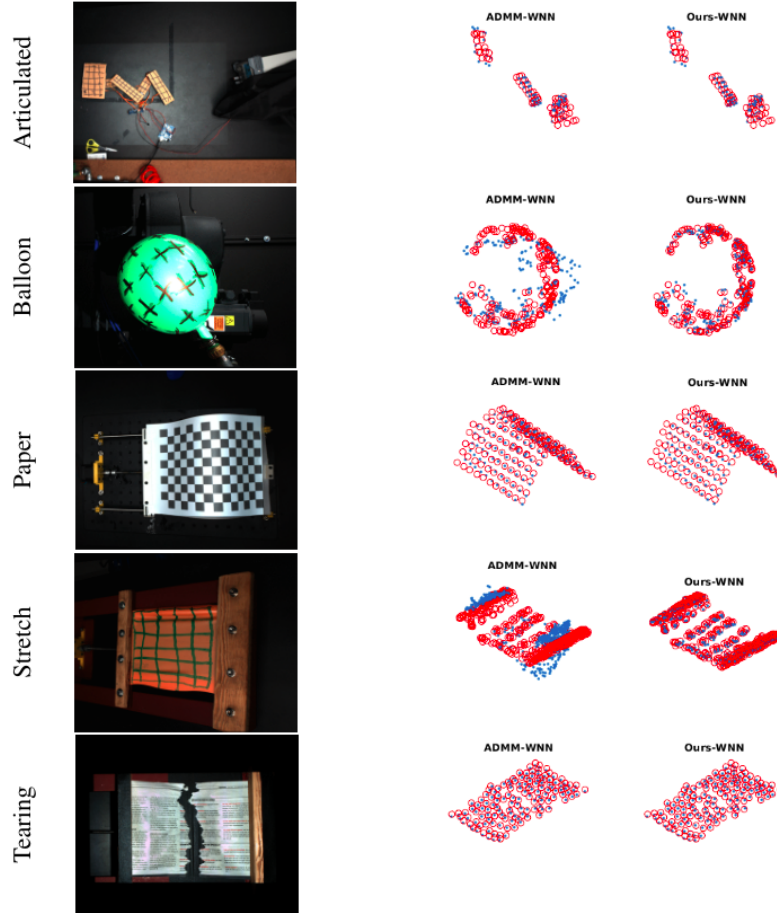

Fig. 3. (Left) Example of the non-rigid objects in the 5 datasets of the NRSfM Challenge. (Right) Estimation (blue) and ground-truth (red) of the non-rigid 3D structure for the two methods with weighted nuclear norm regularization.

on X. A similar problem was studied in [14] but for orthogonal camera models, where the authors propose the rank regularization to be applied on a reshaped version of X, given by $X^{\#} = g^{-1}(X)$, a $F \times 3P$, where the function g performs the permutation on the elements of $X^{\#}$ to obtain X. With this reshaping we have that $\text{rank}(X^{\#}) \leq K$, meaning that we can factorize it as $X^{\#} = BC^T$ with $B \in \mathbb{R}^{F \times K}$ and $C \in \mathbb{R}^{3P \times K}$. The optimization problem then becomes

$$\min_{B,C,t} \sum_{i=1}^{K} a_i \frac{\|B_i\|^2 + \|C_i\|^2}{2} + \ell_{\text{pOSE}}(Rg(BC^T) + t\mathbb{1}^T). \qquad (18)$$

Solving this optimization problem requires small adjustments to be done to the proposed Algorithm 1, which can be consulted in the supplementary material. We apply our methods to the 5 datasets (Articulated, Balloon, Paper, Stretch, Tearing) from the NRSfM Challenge [25]. Each of these datasets include tracks of image points for orthogonal and perspective camera models for six different camera paths (Circle, Flyby, Line, Semi-circle, Tricky, Zigzag), as well as the ground-truth 3D structure for one of the frames. We use the 2D observation for the orthogonal camera model to compute the rotation matrix R, as done in [14], and the ground-truth 3D structure to estimate the intrinsic camera matrix, which is assumed to be fixed during each sequence. The intrinsic camera matrix is used to obtain the calibrated 2D observation of the perspective camera model data. For the nuclear norm (NN), we set $a_i = 1 \times 10^{-3}, i = 1, \ldots, K$. For the weighted nuclear norm (WNN), the weights a are selected similarly to [27] $a_i = \frac{\xi}{\sigma_i(g^{-1}(X_0)) + \gamma}, i = 1, \ldots, K$ where $\xi = 5 \times 10^{-3}$, γ is a small number for numerical stability, and X_0 is the closed-form solution of the objective $\min_X \ell_{\text{pOSE}}(RX)$.

For these datasets we choose $K = 2$ and set the $\eta = 0.05$. As baseline we use the best performing first-order method according to the experiments Section 5.2, ADMM, and apply the method described in Algorithm 1 for local refinement starting from the ADMM's solution. We also try our method for the orthogonal camera model (by setting $\eta = 1$), and compare it with BMM [14] and R-BMM [27], which correspond to ADMM implementations for nuclear norm and weighted nuclear norm, respectively. These methods perform a best rank K approximation to the obtained ADMM solution if $\text{rank}(X^{\#}) > K$ after convergence. We let the ADMM-based methods run until convergence or stalling is achieved for fair comparison. The average log-losses, before and after refinement, obtained on each dataset are shown in Table 1. The average reconstruction errors, in millimeters, on each dataset relatively to the provided ground-truth structure are shown in Table 2. In Figure 3 we also show some qualitative results of the obtained 3D reconstruction of each of the objects in the 5 datasets. More qualitative results are provided in the supplementary material.

The results show that our method is able to achieve lower energies for all datasets comparatively with the ADMM baselines. Similarly to Section 5.2, the difference is more substantial for the perspective model. Furthermore, even though we are not explicitly minimizing the reconstruction error expressed in Table 2, we are able to consistently obtain the lowest reconstruction error for all datasets, sometimes with great improvements compared to the ADMM (see

Table 1. Average log-loss on each of the perspective datasets over the 6 camera paths.

	Method \Dataset	Articulated	Balloon	Paper	Stretch	Tearing
Orthogonal-	BMM [14]	-1.645	-2.267	-1.712	-2.282	-1.453
	Ours-NN	**-1.800**	**-2.352**	**-2.188**	**-2.509**	**-1.634**
	R-BMM [27]	-1.648	-1.979	-1.855	-1.997	-1.522
	Ours-WNN	**-1.648**	**-1.979**	**-1.855**	**-1.997**	**-1.522**
Perspective-	ADMM-NN	-2.221	-2.529	-2.338	-2.395	-1.471
	Ours-NN	**-2.415**	**-2.657**	**-2.560**	**-2.622**	**-2.053**
	ADMM-WNN	-2.455	-2.617	-2.195	-2.651	-1.688
	Ours-WNN	**-2.486**	**-2.931**	**-2.777**	**-2.857**	**-2.103**

Table 2. Average reconstruction errors, in millimeters, on each dataset over the 6 camera paths relatively to the provided ground-truth structure.

	Method \Dataset	Articulated	Balloon	Paper	Stretch	Tearing
Orthogonal	BMM [14]	18.49	10.39	**8.94**	10.02	**14.23**
	Ours-NN	18.31	8.53	10.94	10.67	17.03
	R-BMM [27]	16.00	**7.84**	10.69	**7.53**	16.34
	Ours-WNN	**15.03**	8.05	10.45	9.01	16.20
Perspective	ADMM-NN	16.70	8.05	7.96	6.04	9.40
	Ours-NN	**16.13**	6.48	6.80	6.00	9.31
	ADMM-WNN	18.33	8.95	10.14	8.06	9.28
	Ours-WNN	16.53	**6.27**	**5.68**	**5.93**	**8.42**

Balloon and Stretch in Figure 3). The same does not apply for the orthogonal data, where achieving lower energies did not lead to lower reconstruction errors.

6   Conclusions

In this paper we show that it is possible to optimize a general class of singular value penalties using a bilinear parameterization of the matrix. We show that with this parameterization weighted nuclear norm penalties turn in to smooth objectives that can be accurately solved with 2nd order methods. Our proposed approach starts by using ADMM which rapidly decreases the objective the first couple of iterations and switches to Levenberg-Marquardt when ADMM iterations make little progress. This results in a much more accurate solution and we showed that we were able to extend the recently proposed pOSE [24] to handle non-rigid reconstruction problems.

While 2nd order methods offer increased accuracy, our approach is expensive since iterations require the inversion of a large matrix. Exploring feasible alternatives such as preconditioning and conjugate gradient approaches is an interesting future direction.

Something that we have not discussed is adding constraints on the factors, which is possible since these are present in the optimization. This is very relevant for structure from motion problems and will likely be an fruitful direction to explore.

References

1. Ijaz Akhter, Yaser Sheikh, Sohaib Khan, and Takeo Kanade. Nonrigid structure from motion in trajectory space. In *Proceedings of the 21st International Conference on Neural Information Processing Systems*, NIPS'08, page 41–48, Red Hook, NY, USA, 2008. Curran Associates Inc.
2. Francis Bach. Convex relaxations of structured matrix factorizations. September 2013.
3. Ronen Basri, David Jacobs, and Ira Kemelmacher. Photometric stereo with general, unknown lighting. *International Journal of Computer Vision*, 72(3):239–257, May 2007.
4. Rajendra Bhatia and Tanvi Jain. On symplectic eigenvalues of positive definite matrices. *Journal of Mathematical Physics*, 56(11):112201, 2015.
5. G. Birkhoff. Tres observaciones sobre el algebra lineal. *Universidad Nacional de Tucuman, Revista. Serie A.*, 5:137–151, 1946.
6. Stephen Boyd, Neal Parikh, Eric Chu, Borja Peleato, and Jonathan Eckstein. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Found. Trends Mach. Learn.*, 3(1):1–122, 2011.
7. C. Bregler, A. Hertzmann, and H. Biermann. Recovering non-rigid 3d shape from image streams. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2000.
8. A. M. Buchanan and A. W. Fitzgibbon. Damped newton algorithms for matrix factorization with missing data. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2005.
9. R. Cabral, F. De la Torre, J. P. Costeira, and A. Bernardino. Unifying nuclear norm and bilinear factorization approaches for low-rank matrix decomposition. In *International Conference on Computer Vision (ICCV)*, 2013.
10. Emmanuel J. Candès, Xiaodong Li, Yi Ma, and John Wright. Robust principal component analysis? *J. ACM*, 58(3):11:1–11:37, 2011.
11. Emmanuel J Candès and Benjamin Recht. Exact matrix completion via convex optimization. *Foundations of Computational Mathematics*, 9(6):717–772, 2009.
12. Lu Canyi, Jinhui Tang, Shuicheng Yan, and Zhouchen Lin. Generalized nonconvex nonsmooth low-rank minimization. *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2014.
13. Lu Canyi, Jinhui Tang, Shuicheng Yan, and Zhouchen Lin. Nonconvex nonsmooth low-rank minimization via iteratively reweighted nuclear norm. *IEEE Transactions on Image Processing*, 25, 10 2015.
14. Yuchao Dai, Hongdong Li, and Mingyi He. A simple prior-free method for non-rigid structure-from-motion factorization. *International Journal of Computer Vision*, 107(2):101–122, 2014.
15. Maryam Fazel, Haitham Hindi, and Stephen P Boyd. A rank minimization heuristic with application to minimum order system approximation. In *American Control Conference*, 2001.
16. Ravi Garg, Anastasios Roussos, and Lourdes Agapito. A variational approach to video registration with subspace constraints. *International Journal of Computer Vision*, 104(3):286–314, 2013.
17. Ravi Garg, Anastasios Roussos, and Lourdes de Agapito. Dense variational reconstruction of non-rigid surfaces from monocular video. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2013.

18. Shuhang Gu, Qi Xie, Deyu Meng, Wangmeng Zuo, Xiangchu Feng, and Lei Zhang. Weighted nuclear norm minimization and its applications to low level vision. *International Journal of Computer Vision*, 121, 07 2016.
19. B. D. Haeffele and R. Vidal. Structured low-rank matrix factorization: Global optimality, algorithms, and applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(6):1468–1482, 2020.
20. Je Hyeong Hong and Andrew Fitzgibbon. Secrets of matrix factorization: Approximations, numerics, manifold optimization and random restarts. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015.
21. Je Hyeong Hong, Christopher Zach, Andrew Fitzgibbon, and Roberto Cipolla. Projective bundle adjustment from arbitrary initialization using the variable projection method. In *European Conference on Computer Vision (ECCV)*, 2016.
22. Je Hyeong Hong, Christopher Zach, Andrew Fitzgibbon, and Roberto Cipolla. Projective bundle adjustment from arbitrary initialization using the variable projection method. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.
23. Y. Hu, D. Zhang, J. Ye, X. Li, and X. He. Fast and accurate matrix completion via truncated nuclear norm regularization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(9):2117–2130, 2013.
24. Je Hyeong Hong and Christopher Zach. pose: Pseudo object space error for initialization-free bundle adjustment. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018.
25. Sebastian Hoppe Nesgaard Jensen, Alessio Del Bue, Mads Emil Brix Doest, and Henrik Aanæs. A benchmark and evaluation of non-rigid structure from motion, 2018.
26. Mikhail Krechetov, Jakub Marecek, Yury Maximov, and Martin Takac. Entropy-penalized semidefinite programming. In *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI-19*, pages 1123–1129. International Joint Conferences on Artificial Intelligence Organization, 7 2019.
27. Suryansh Kumar. Non-rigid structure from motion: Prior-free factorization method revisited. In *The IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 51–60, March 2020.
28. Viktor Larsson and Carl Olsson. Convex low rank approximation. *International Journal of Computer Vision*, 120(2):194–214, 2016.
29. Adrian S Lewis. The convex analysis of unitarily invariant matrix functions. *Journal of Convex Analysis*, 2(1):173–183, 1995.
30. Huan Li and Zhouchen Lin. Provable accelerated gradient method for nonconvex low rank optimization. *Machine Learning*, 06 2019.
31. T. H. Oh, Y. W. Tai, J. C. Bazin, H. Kim, and I. S. Kweon. Partial sum minimization of singular values in robust pca: Algorithm and applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38(4):744–758, 2016.
32. Carl Olsson and Olof Enqvist. Stable structure from motion for unordered image collections. In *Proceedings of the Scandinavian Conference on Image Analysis (SCIA)*, pages 524–535, 2011.
33. Samet Oymak, Karthik Mohan, Maryam Fazel, and Babak Hassibi. A simplified approach to recovery conditions for low rank matrices. In *IEEE International Symposium on Information Theory Proceedings (ISIT)*, pages 2318–2322, 2011.
34. M. Paladini, A. Del Bue, M. Stosic, M. Dodig, J. Xavier, and L. Agapito. Factorization for non-rigid and articulated structure using metric projections. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pages 2898–2905, 2009.

35. Benjamin Recht, Maryam Fazel, and Pablo A. Parrilo. Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization. *SIAM Rev.*, 52(3):471–501, August 2010.

36. Chris Russell, Joao Fayad, and Lourdes Agapito. Energy based multiple model fitting for non-rigid structure from motion. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 3009 – 3016, 07 2011.

37. F. Shang, J. Cheng, Y. Liu, Z. Luo, and Z. Lin. Bilinear factor matrix norm minimization for robust pca: Algorithms and applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(9):2066–2080, Sep. 2018.

38. Danail Stoyanov, George P. Mylonas, Fani Deligianni, Ara Darzi, and Guang Zhong Yang. Soft-tissue motion tracking and structure estimation for robotic assisted mis procedures. In James S. Duncan and Guido Gerig, editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2005*, pages 139–146, Berlin, Heidelberg, 2005. Springer Berlin Heidelberg.

39. Carlo Tomasi and Takeo Kanade. Shape and motion from image streams under orthography: A factorization method. *International Journal of Computer Vision*, 9(2):137–154, 1992.

40. Marcus Valtonen Ornhag, Carl Olsson, and Anders Heyden. Bilinear parameterization for differentiable rank-regularization. In *The IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2020.

41. Aydin Varol, Mathieu Salzmann, Engin Tola, and Pascal Fua. Template-free monocular reconstruction of deformable surfaces. In *International Conference on Computer Vision (ICCV)*, pages 1811 – 1818, 11 2009.

42. Naiyan Wang, Tiansheng Yao, Jingdong Wang, and Dit-Yan Yeung. A probabilistic approach to robust matrix factorization. In *European Conferencec on Computer Vision (ECCV)*, 2012.

43. Jing Xiao, Jin-xiang Chai, and Takeo Kanade. A closed-form solution to non-rigid shape and motion recovery. In Tomás Pajdla and Jiří Matas, editors, *Computer Vision - ECCV 2004*, pages 573–587, Berlin, Heidelberg, 2004. Springer Berlin Heidelberg.

44. Chen Xu, Zhouchen Lin, and Hongbin Zha. A unified convex surrogate for the schatten-p norm. In *Proceedings of the Conference on Artificial Intelligence (AAAI)*, 2017.

45. J. Yan and M. Pollefeys. A factorization-based approach for articulated nonrigid shape, motion and kinematic chain recovery from video. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(5):865–877, 2008.

46. Jingyu Yan and M. Pollefeys. A factorization-based approach for articulated non-rigid shape, motion and kinematic chain recovery from video. *IEEE Trans. Pattern Anal. Mach. Intell.*, 30(5):865–877, 2008.