Rank Minimization with Structured Data Patterns

Viktor Larsson[1], Carl Olsson[1], Erik Bylow[1], and Fredrik Kahl[1,2]

[1] Centre for Mathematical Sciences
Lund University, Sweden
[2] Department of Signals and Systems
Chalmers University of Technology, Sweden

Abstract. The problem of finding a low rank approximation of a given measurement matrix is of key interest in computer vision. If all the elements of the measurement matrix are available, the problem can be solved using factorization. However, in the case of missing data no satisfactory solution exists. Recent approaches replace the rank term with the weaker (but convex) nuclear norm. In this paper we show that this heuristic works poorly on problems where the locations of the missing entries are highly correlated and structured which is a common situation in many applications.

Our main contribution is the derivation of a much stronger convex relaxation that takes into account not only the rank function but also the data. We propose an algorithm which uses this relaxation to solve the rank approximation problem on matrices where the given measurements can be organized into overlapping blocks without missing data. The algorithm is computationally efficient and we have applied it to several classical problems including structure from motion and linear shape basis estimation. We demonstrate on both real and synthetic data that it outperforms state-of-the-art alternatives. [1]

1 Factorization Methods and Convex Relaxations

The ability to find the best fixed rank approximation of a matrix has been proven useful in applications such as structure from motion, photometric stereo and optical flow [5,23,9,3,10]. The rank of the approximating matrix typically describes the complexity of the solution. For example, in non-rigid structure from motion the rank measures the number of basis elements needed to describe the point motions [5]. Therefore a good optimization criterion consists of a trade-off between rank and residual errors, leading to formulations of the type

$$\min_X \mu \operatorname{rank}(X) + \|X - M\|_F^2, \tag{1}$$

where M is a matrix of measurements. The above problem can be solved using the singular value decomposition (SVD) followed by truncation of the singular

[1] This work has been funded by the Swedish Research Council (grants no. 2012-4213 and 2012-4215) and the Crafoord Foundation.

D. Fleet et al. (Eds.): ECCV 2014, Part III, LNCS 8691, pp. 250–265, 2014.

values, but the strategy is limited to problems without missing data and outliers. The issue of outliers has received a lot of attention lately. In [8,21,24] the more robust ℓ_1 norm is used. While experimental results indicate robustness to suboptimal solutions, these methods are still local in that updates are computed using a local approximation of the objective function. To handle missing data [2] replaces the Frobenious norm with the spectral norm. It is shown that if the missing data forms a Young pattern the globally optimal solution can be computed. For general problems [2] proposes an alternation over Young patterns.

A recent heuristic that has been shown to work well is to replace the rank function with the nuclear norm [19,7,9,18,1]. This leads to formulations of the type

$$\min_X \mu\|X\|_* + \|W \odot (X - M)\|_F^2, \tag{2}$$

where W is a matrix with entry $W_{ij} = 1$ if M_{ij} has been observed, $W_{ij} = 0$ otherwise and $\odot$ denotes elementwise multiplication. It can be shown [19,7] that if the location of the missing entries are random then problem (2) provides the correct low rank solution. However, in many applications such as structure from motion, missing entries are highly correlated. When a scene point cannot be tracked any longer (either due to occlusion or appearance change) it rarely appears again. This gives rise to certain sparsity patterns (often diagonally dominant) in W which are not random.

Figure 1 shows the results of a synthetic experiment. Here we generated a 100×100 matrix $A = UV^T$ of rank 3 by randomly selecting 100×3 matrices U and V with elements selected from a Gaussian distribution (with mean 0 and standard deviation 1). To generate the measurement matrix M we then added Gaussian noise with standard deviation 0.05 to each entry in A.

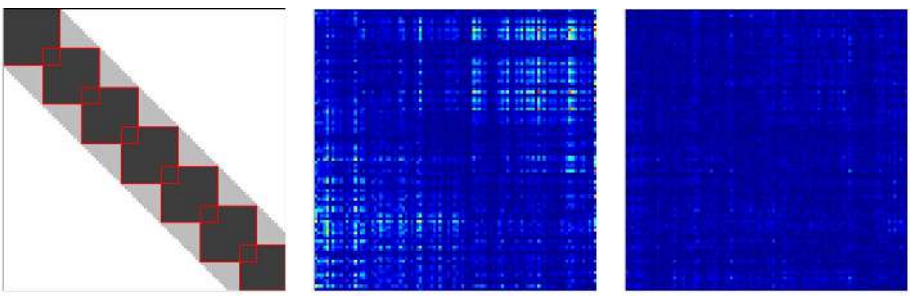

Fig. 1. *Left*: The pattern of missing data (red lines show the blocks used by our algorithm). Available measurements are located close to the diagonal. *Middle*: Absolute errors of the (rank 3) solution obtained using (2)[2]. *Right*: Absolute errors of the (rank 3) solution obtained using the proposed approach.

[2] μ was chosen so that the first three singular values accounted for 99% of the solution.

In an effort to strengthen the nuclear norm formulation, [1] adds prior information using the generalized trace norm. The formulation is related to (2) by a reweighing of the data term and can incorporate knowledge such as smoothness priors. The availability of such information can improve the estimation, however the formulation still uses the nuclear norm for regularization. On a high level our approach is similar to [1] in that it also attempts to find a stronger convex relaxation by considering not just the rank function but also the data. However, in contrast to [1] we do not add priors to the problem but simply use the information in the available measurements.

The motivation for replacing the rank function with the nuclear norm in (2) is that it is the convex envelope of the rank function on the set $\{X; \sigma_{max}(X) \leq 1\}$. That is, it is the tightest possible convex relaxation on this set. The constraint $\sigma_{max}(X) \leq 1$ is however not present in (2). In fact, the convex envelope of an unconstrained rank function is the zero function. In this paper we show that a significantly more accurate convex envelope can be computed if a solution estimate X_0 is available. Specifically, we compute the convex envelope of

$$f(X) = \mu \operatorname{rank}(X) + \|X - X_0\|_F^2. \tag{3}$$

We will refer to this function as the **localized rank function**. Compared to the nuclear norm constraint $\sigma_{max}(X) \leq 1$, the term $\|X - X_0\|_F^2$ effectively translates the feasible region to a neighborhood around X_0 instead of the trivial solution 0. Therefore our convex envelope can penalize smaller singular values harder than larger ones, while the nuclear norm tries to force all singular values to be zero.

The work most similar to ours is perhaps [13] where the convex envelope of a vector version of our problem (the cardinality function and the ℓ_2-norm) is computed. In contrast, here we are interested in the matrix setting. Furthermore, in [13] the ℓ_2-norm is centered around the origin. In our work the translation is an essential component required to avoid penalizing large singular values.

Minimizing the convex envelope of (3) gives the same result as $\min f(X)$. The significant advantage of using the envelope instead is that it is convex and therefore can be combined with other convex constraints and functions. We propose to utilize it for problems with missing data patterns as exemplified in Figure 1. More specifically, our convex relaxation is applied to sub-blocks of the matrix with no missing entries rather than the nuclear norm of the entire matrix X. In effect, this can be seen as minimizing the rank of each sub-block separately and due to the convexity of our approximation, it is possible to enforce that the sub-blocks agree on their overlap. Furthermore, we show that under mild assumptions it possible to extract a solution to the full matrix X that has rank equal to the largest rank of the sub-blocks. We present an ADMM [4] based approach for obtaining a solution which only requires to compute SVDs of the sub-blocks rather than the whole matrix resulting in an efficient implementation. In summary, we derive tight convex relaxations for a class of problems with a large amount of missing entries that outperform state-of-the-art methods.

Notation. Throughout the paper we use $\sigma_i(X)$ to denote the ith singular value of a matrix X. The vector of all singular values is denoted $\boldsymbol{\sigma}(X)$. With some abuse of notation we write the SVD of X as $U\,\mathrm{diag}(\boldsymbol{\sigma}(X))V^T$. For ease of notation we do not explicitly indicate the dependence of U and V on X. The scalar product is defined as $\langle X, Y\rangle = \mathrm{tr}(X^TY)$, where tr is the trace function, and the Frobenius norm $\|X\|_F = \sqrt{\langle X, X\rangle} = \sqrt{\sum \sigma_i^2(X)}$. Truncation at zero is denoted $[a]_+$, that is, $[a]_+ = 0$ if $a < 0$ and a otherwise.

2 The Convex Envelope

In this section we show that it is possible to compute a closed form expression for the convex envelope of the localized rank function (3).

Finding the envelope of f is equivalent to computing $f^{**} = (f^*)^*$, where

$$f^*(Y) = \max_X \ \langle Y, X\rangle - f(X), \tag{4}$$

is the Fenchel conjugate of f [20]. By completing squares, $f^*(Y)$ can be written

$$f^*(Y) = \max_k \max_{\mathrm{rank}(X)=k} \left\|\frac{1}{2}Y + X_0\right\|_F^2 - \|X_0\|_F^2 - \left\|X - (\frac{1}{2}Y + X_0)\right\|_F^2 - \mu k. \tag{5}$$

Note that the first two terms are independent of X and can be considered as constants in the maximization over X and k. In addition k is fixed in the inner maximization. For a fixed k, the maximizing X is given by the best rank k approximation of $\frac{1}{2}Y + X_0$ which can be obtained via an SVD of $\frac{1}{2}Y + X_0$ by setting all singular values but the k largest to zero. Inserting into (5) we get

$$f^*(Y) = \max_k \left\|\frac{1}{2}Y + X_0\right\|_F^2 - \|X_0\|_F^2 - \sum_{i=k+1}^{n} \sigma_i^2(\frac{1}{2}Y + X_0) - \mu k \tag{6}$$

$$= \max_k \left\|\frac{1}{2}Y + X_0\right\|_F^2 - \|X_0\|_F^2 - \sum_{i=k+1}^{n} \sigma_i^2(\frac{1}{2}Y + X_0) - \sum_{i=1}^{k} \mu. \tag{7}$$

To select the best k we note that the largest value is achieved when k fulfills

$$\sigma_k^2(\frac{1}{2}Y + X_0) \geq \mu \geq \sigma_{k+1}^2(\frac{1}{2}Y + X_0). \tag{8}$$

For the maximizing k the last two sums can be written

$$\sum_{i=k+1}^{n} \sigma_i^2(\frac{1}{2}Y + X_0) + \sum_{i=1}^{k} \mu = \sum_{i=1}^{n} \min\left(\mu,\ \sigma_i^2(\frac{1}{2}Y + X_0)\right). \tag{9}$$

Therefore we get the conjugate function

$$f^*(Y) = \left\|\frac{1}{2}Y + X_0\right\|_F^2 - \|X_0\|_F^2 - \sum_{i=1}^{n} \min\left(\mu,\ \sigma_i^2(\frac{1}{2}Y + X_0)\right). \tag{10}$$

We next proceed to compute the bi-conjugate $f^{**}(X) = \max_Y \ \langle X, Y\rangle - f^*(Y)$. To simplify computations we change variables to $Z = \frac{1}{2}Y + X_0$ and maximize over Z instead. We get

$$f^{**}(X) = \max_Z 2\langle X, Z - X_0\rangle - \|Z\|_F^2 + \|X_0\|_F^2 + \sum_{i=1}^n \min(\mu,\ \sigma_i^2(Z)). \tag{11}$$

The first three terms can, by completing squares, be simplified into $\|X - X_0\|_F^2 - \|Z - X\|_F^2$. Furthermore, since $\|X - X_0\|_F^2$ does not depend on Z we get

$$f^{**}(X) = \|X - X_0\|_F^2 + \max_Z \left(\sum_{i=1}^n \min\left(\mu,\ \sigma_i^2(Z)\right) - \|Z - X\|_F^2 \right). \tag{12}$$

The sum in (12) only depends on the singular values of Z and is therefore unitarily invariant. We also note that $-\|Z - X\|_F^2 = -\|Z\|_F^2 + 2\langle Z, X\rangle - \|X\|_F^2$. The term $\|Z\|_F^2$ is unitarily invariant and by von Neumann's trace inequality, we know that $\langle Z, X\rangle \leq \sum_{i=1}^n \sigma_i(Z)\sigma_i(X)$. Equality is achieved if X and Z have SVDs with the same U and V. Hence, for Z to maximize the sum in (12), its SVD should be of the form $Z = U \operatorname{diag}(\boldsymbol{\sigma}(Z))V^T$ if $X = U \operatorname{diag}(\boldsymbol{\sigma}(X))V^T$.

What is left now is to determine the singular values of Z and because both terms containing $\sigma_i(Z)$ are separable, we can consider one at a time. Hence we need to solve $\max_{\sigma_i(Z)} \ \min(\mu, \sigma_i^2(Z)) - (\sigma_i(Z) - \sigma_i(X))^2$. There are two cases:

(i) If $\sigma_i(Z) \leq \sqrt{\mu}$ then the problem simplifies to $\max_{\sigma_i(Z)} \ \sigma_i^2(Z) - (\sigma_i(Z) - \sigma_i(X))^2$. Expanding the square, this objective function is $2\sigma_i(X)\sigma_i(Z) - \sigma_i^2(X)$. The singular value $\sigma_i(X)$ is positive and the optimal choice is therefore $\sigma_i(Z) = \sqrt{\mu}$.

(ii) If $\sigma_i(Z) \geq \sqrt{\mu}$ we need to solve $\max_{\sigma_i(Z)} \ \mu - (\sigma_i(Z) - \sigma_i(X))^2$. The maximum is clearly achieved in $\sigma_i(Z) = \sigma_i(X)$.

Summarizing the two cases, the optimal Z has $\sigma_i(Z) = \max(\sqrt{\mu}, \sigma_i(X))$.

We now insert Z into (12) to find the expression for the bi-conjugate. Note that $\min\left(\mu,\ \sigma_i^2(Z)\right) = \min\left(\mu,\ \max(\mu, \sigma_i^2(X))\right) = \mu$, and that $\|Z - X\|_F^2 = \sum_{i=1}^n \left[\sqrt{\mu} - \sigma_i(X)\right]_+^2$. Hence, the convex envelope is given by

$$f^{**}(X) = \mathcal{R}_\mu(X) + \|X - X_0\|_F^2, \tag{13}$$

where

$$\mathcal{R}_\mu(X) = \sum_{i=1}^n \left(\mu - [\sqrt{\mu} - \sigma_i(X)]_+^2\right). \tag{14}$$

In [22] the authors propose a rank regularizer which for some parameter choices is equivalent to $\mathcal{R}_\mu$. However, they make no connection to the convex envelope of f and simply minimize it in a non-convex framework.

Figure 2 shows a one dimensional version of (13). To the left is the term $\mu - \left[\sqrt{\mu} - \sigma\right]_+^2$ which is in itself not convex. For singular values larger than $\sqrt{\mu}$

it gives a constant penalty. When the quadratic term σ^2 is added the result is a convex penalty, see the middle graph in Figure 2. For $\sigma < \sqrt{\mu}$ the function has a linear shape (red dashed curve) similar to the nuclear norm, while for $\sigma \geq \sqrt{\mu}$ it behaves like the quadratic function $\mu + \sigma^2$. Note that the one dimensional version of f is identical to $\mu + \sigma^2$ everywhere except for $\sigma = 0$. In the right image we plotted the graphs of $\mu - \left[\sqrt{\mu} - \sigma\right]_+^2 + (\sigma - \sigma_0)^2$ for $\sigma_0 = 0, 1, 2$. If σ_0 is large enough the function will not try to force σ to be zero.

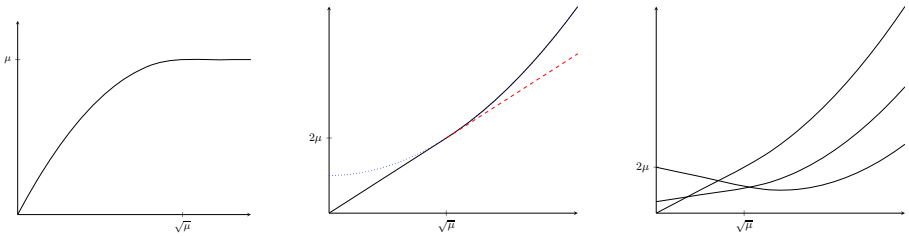

Fig. 2. One dimensional visualizations of (13) for $\mu = 2$. Left: The graph of $\mu - \left[\sqrt{\mu} - \sigma\right]_+^2$. Middle: The graph of $\mu - \left[\sqrt{\mu} - \sigma\right]_+^2 + \sigma^2$. If μ is large its shape resembles the nuclear norm. Right: The graphs of $\mu - \left[\sqrt{\mu} - \sigma\right]_+^2 + (\sigma - \sigma_0)^2$ for $\sigma_0 = 0, 1, 2$.

3 A Block Decomposition Approach

Next we present our approach for solving the missing data problem. The idea is to try to enforce low rank of sub-blocks of the matrix where no measurements are missing using our convex relaxation.

Let R_i and C_i, $i = 1, .., K$ be a subset of row and column indices for each block. By $\mathcal{P}_i : \mathbb{R}^{m \times n} \mapsto \mathbb{R}^{|R_i| \times |C_i|}$ we will mean the (linear) operator that extracts the elements with indices in $R_i \times C_i$ and forms a sub-matrix of size $|R_i| \times |C_i|$. If M is our (partially filled) measurement matrix, then the submatrix $\mathcal{P}_i(M)$ has no missing values. We seek to minimize the non-convex function

$$f(X) = \sum_{i=1}^{K} \mu_i \operatorname{rank}(\mathcal{P}_i(X)) + \|\mathcal{P}_i(X) - \mathcal{P}_i(M)\|_F^2, \tag{15}$$

by replacing it with the convex envelopes of the localized rank function (13),

$$f_{\mathcal{R}}(X) = \sum_{i=1}^{K} \mathcal{R}_{\mu_i}(\mathcal{P}_i(X)) + \|\mathcal{P}_i(X) - \mathcal{P}_i(M)\|_F^2. \tag{16}$$

Note that we do not explicitly enforce that the rank of X is penalized, but instead this is accomplished via the rank penalization of the sub-matrices. In the next subsection, we shall see that the rank of the full matrix X and the rank of its sub-matrices $\mathcal{P}_i(X)$ are strongly related. Also note that since the

blocks have overlaps this formulation counts some residuals more than others. It is possible to add a term $\|W^r \odot (X - M)\|_F^2$ to ensure that each residual error is counted equally many times, but we have experimentally found that this makes little difference and we therefore ignore this term for ease of presentation.

3.1 Constructing the Full Matrix X

Let us now assume that we have solved the convex relaxation of problem (16). However, as only the sub-matrices of X are involved in the optimization, we still need to find the complete matrix X. Let $r_{\max}$ denote the maximum rank over all sub-matrices. We shall show the following result:

Lemma 1. *Let X_1 and X_2 be two given matrices with overlap matrix X_{22} as depicted in Figure 3, and let $r_1 = \operatorname{rank}(X_1)$ and $r_2 = \operatorname{rank}(X_2)$. Suppose that $\operatorname{rank}(X_{22}) = \min(r_1, r_2)$, then there exists a matrix X with $\operatorname{rank}(X) = \max(r_1, r_2)$. Additionally if $\operatorname{rank}(X_{22}) = r_1 = r_2$ then X is unique.*

Proof. We will assume (w.l.o.g) that $r_2 \leq r_1$, and look at the block X_2. The overlap X_{22} is of rank r_2 so there are r_2 linearly independent columns in $[X_{22}^T \; X_{32}^T]^T$ and rows in $[X_{22} \; X_{23}]$. Now the rank of X_2 is r_2 and we can find coefficient matrices C_1 and C_2 such that

$$\begin{bmatrix} X_{23} \\ X_{33} \end{bmatrix} = \begin{bmatrix} X_{22} \\ X_{32} \end{bmatrix} C_1 \quad \text{and} \quad \begin{bmatrix} X_{32} \; X_{33} \end{bmatrix} = C_2 \begin{bmatrix} X_{22} \; X_{23} \end{bmatrix}. \tag{17}$$

We therefore set $X_{13} := X_{12}C_1$ and $X_{31} := C_2 X_{21}$. To determine the rank of the resulting X we first look at the number of linearly independent columns. By construction, the columns $[X_{13}^T \; X_{23}^T \; X_{33}^T]^T$ are linear combinations of the other columns, and similarly, the rows $[X_{31} \; X_{32} \; X_{33}]$ are linear combinations of the other rows. Hence, the number of linear independent columns (or rows) has not increased. Therefore X has the same rank as X_1. The uniqueness in the case of $r_1 = r_2$ can easily be proven by contradiction. □

We now present a simple way to extend the solution beyond the blocks. The completion of two blocks is in practice performed by finding rank $r_{\max}$ factorizations of X_1 and X_2 using SVD (see Figure 3),

$$X_1 = U_1 V_1^T \quad \text{and} \quad X_2 = U_2 V_2^T \quad \text{where} \quad U_k \in \mathbb{R}^{m_k \times r}, V_k \in \mathbb{R}^{n_k \times r}. \tag{18}$$

The low rank factorizations are however not unique because for any invertible H, $U_1 V_1^T = (U_1 H)(H^{-1} V_1^T)$. To find the unknown H we consider the block $X_{22} = \hat{U}_1 \hat{V}_1^T = \hat{U}_2 \hat{V}_2^T$, where $\hat{U}_i \hat{V}_i^T$ is the restriction of the $U_i V_i^T$ to X_{22}. Then

$$\hat{U}_1 = \hat{U}_2 H \quad \text{and} \quad \hat{V}_1^T = H^{-1} \hat{V}_2^T \implies H \hat{V}_1^T = \hat{V}_2^T, \tag{19}$$

which we solve in a least squares sense. In this way we iteratively combine the sub-blocks. Other approaches, such as nullspace matching methods used in [17] and [12], are also possible. Note however, that in each iteration we only compute the SVD of the new (smaller) block allowing efficient implementation.

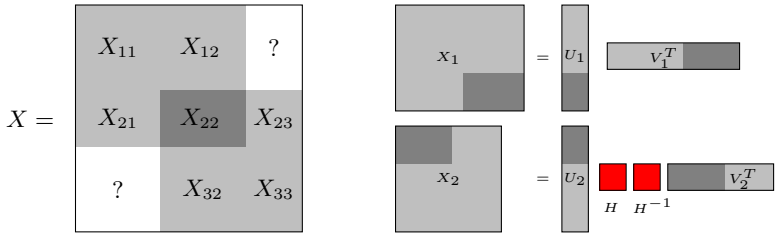

Fig. 3. *Left:* The matrix X contains two overlapping blocks X_1 and X_2. The goal is to fill in the missing entries X_{13} and X_{31} such that rank(X) is kept to a minimum. *Right:* The low-rank factorizations of the two blocks X_1 and X_2. The overlap is marked in both the blocks and the factorizations.

3.2 Selecting μ

The regularizer $\mathcal{R}_\mu$ in (14) only penalizes singular values less than $\sqrt{\mu}$. This gives a clear way to set how much variation in the data we allow to be explained by measurement noise.

If a fixed rank r solution is desired, then the parameters μ_i in (15) can be found iteratively by solving the problem and then updating each μ_i to be slightly smaller than $\sigma_r^2(\mathcal{P}_i(X))$. This ensures that only singular values less than $\sigma_r(\mathcal{P}_i(X))$ are penalized in each block.

4 An ADMM Implementation

In this section we present a simple approach for computing a solution to our formulation. Working with the sub-blocks of a matrix, the ADMM scheme is a natural choice for implementation. Furthermore, our objective function is convex, so convergence to the optimal value is guaranteed [4].

For each block $\mathcal{P}_i(X)$ we introduce a separate set of variables X_i and enforce consistency via the linear constraints $X_i - \mathcal{P}_i(X) = 0$. We formulate an augmented Lagrangian of (15) as

$$\sum_{i=1}^{K} \left(\mathcal{R}_{\mu_i}(X_i) + \|X_i - \mathcal{P}_i(M)\|_F^2 + \rho \|X_i - \mathcal{P}_i(X) + \Lambda_i\|_F^2 - \rho \|\Lambda_i\|_F^2 \right). \quad (20)$$

At each iteration t of ADMM we solve the subproblems

$$X_i^{t+1} = \arg\min_{X_i} \mathcal{R}_{\mu_i}(X_i) + \|X_i - \mathcal{P}_i(M)\|_F^2 + \rho \|X_i - \mathcal{P}_i(X^t) + \Lambda_i^t\|_F^2, \quad (21)$$

for $i = 1, ..., K$ and

$$X^{t+1} = \arg\min_X \sum_{i=1}^{K} \rho \|X_i^{t+1} - P_i(X) + \Lambda_i^t\|_F^2. \quad (22)$$

Here Λ_i^t, $i = 1, ..., K$ are the scaled dual variables whose updates at iteration t are given by

$$\Lambda_i^{t+1} = \Lambda_i^t + X_i^{t+1} - P_i(X^{t+1}). \tag{23}$$

The second subproblem (22) is a separable least squares problem with closed form solution. In the next section we compute the solutions of (21). Note that these are completely independent of each other and could in principle be solved in parallel.

4.1 The Proximal Operator

Ignoring constants, the objective function in (21) can be written

$$F(X_i) = G(X_i) - 2\langle Y, X_i \rangle, \tag{24}$$

where

$$G(X_i) = \sum_{j=1}^{n} \left(- \left[\sqrt{\mu} - \sigma_j(X_i) \right]_+^2 \right) + (1+\rho)\|X_i\|_F^2, \tag{25}$$

and $Y = \mathcal{P}_i(M) + \rho(\mathcal{P}_i(X^t) - \Lambda_i^t)$. Due to convexity it is sufficient to find X_i such that $0 \in \partial F(X_i)$ to optimize F. Define the function $g : \mathbb{R}^n \mapsto \mathbb{R}$ by

$$g(\boldsymbol{\sigma}) = \sum_{i=1}^{n} g_i(\sigma_i), \tag{26}$$

where $g_i(\sigma) = - \left[\sqrt{\mu} - |\sigma| \right]_+^2 + (1+\rho)\sigma^2$. Then g is an absolutely symmetric convex function and G can be written $G(X) = g \circ \boldsymbol{\sigma}(X)$. If $X = U \operatorname{diag}(\boldsymbol{\sigma}(X)) V^T$ then according to [15], the sub-differentials of G and g are related by

$$\partial G(X) = U \operatorname{diag}(\partial g \circ \boldsymbol{\sigma}(X)) V^T. \tag{27}$$

The function $g(\boldsymbol{\sigma})$ is a sum of one dimensional functions and therefore its sub-differential can easily be computed by considering the components of the sum separately. The functions $g_i(\sigma)$ are differentiable everywhere except in $\sigma = 0$ where $\partial g_i(0) = [-2\sqrt{\mu}, 2\sqrt{\mu}]$. For any other σ we have

$$\frac{\partial g_i}{\partial \sigma} = -2 \operatorname{sgn}(\sigma) \left[\sqrt{\mu} - |\sigma| \right]_+ + 2(1+\rho)\sigma. \tag{28}$$

To solve $0 \in \partial F(X_i)$ we now construct a solution to $2Y \in \partial G(X_i)$. If $Y = U \operatorname{diag}(\boldsymbol{\sigma}(Y)) V^T$ then it can be verified that $X_i = U \operatorname{diag}(\boldsymbol{\sigma}(X_i)) V^T$ where

$$\sigma_i(X) = \begin{cases} \frac{\sigma_i(Y)}{1+\rho} & \text{if } \sigma_i(Y) \geq (1+\rho)\sqrt{\mu} \\ \frac{\sigma_i(Y) - \sqrt{\mu}}{\rho} & \text{if } \sqrt{\mu} \leq \sigma_i(Y) \leq (1+\rho)\sqrt{\mu} \\ 0 & \text{if } \sigma_i(Y) \leq \sqrt{\mu} \end{cases} \tag{29}$$

fulfills all the required constraints.

5 Experiments

In this section we evaluate the quality of our relaxation on both synthetic and real data. All experiments were run on a computer running Ubuntu 13.04 with an Intel I7-3930K CPU and 64 GB of RAM. The algorithms were implemented in Matlab and the block update in ADMM was not parallelized.

5.1 Evaluation of the Convex Relaxation

We emprically evaluate the performance of the convex relaxation $f_{\mathcal{R}}(X)$ in (16) for the non-convex objective function $f(X)$ in (15). Note that in the case of a single block $K = 1$, $f_{\mathcal{R}}(X)$ is a tight convex relaxation of $f(X)$. Therefore we test the performance when we have more blocks ($K = 7$).

We generated random 100×100 rank 3 matrices. This was done by sampling $U, V \in \mathbb{R}^{100\times 3}$ from a Gaussian distribution with zero mean and unit variance and then forming the measurement matrix $M = UV^T$. The observation matrix W was chosen to be a band-diagonal matrix with bandwidth 40 similar to the matrix in Figure 1. The blocks were laid out along the diagonal such that their overlap was 6×6 and contained no missing data. We then solved $X^*_{\mathcal{R}} = \arg\min_X f_{\mathcal{R}}(X)$ with $\mu_i = 1$ and varying degrees of Gaussian noise added to M. For each noise level, the test was repeated 1000 times and the results averaged. In the left of Figure 4 we plot the averages of $f(X^*_{\mathcal{R}})$ and $f_{\mathcal{R}}(X^*_{\mathcal{R}})$. Note that if $f(X^*_{\mathcal{R}}) = f_{\mathcal{R}}(X^*_{\mathcal{R}})$ then $X^*_{\mathcal{R}}$ is a global minimizer of f.

For comparison we substituted the rank function for the nuclear norm. The nuclear norm is also convex, hence it can be used in our block decomposition framework. Therefore we did the same experiment with

$$f_N(X) = \sum_{i=1}^{K} \mu_i \|\mathcal{P}_i(X)\|_* + \|\mathcal{P}_i(X) - \mathcal{P}_i(M)\|_F^2. \tag{30}$$

The comparative results can be seen in the left graph of Figure 4. Note that the constraint $\sigma_{\max}(\mathcal{P}_i(X)) \leq 1$ can be violated, so f_N is not necessarily a lower bound on f.

5.2 Comparison to Non-Convex Approaches

Next we compare our methods performance to two state-of-the-art non-convex methods: OptSpace [14] and Truncated Nuclear Norm Regularization (TNNR) [11]. OptSpace is based on local optimization on Grassmann manifolds and in TNNR an energy which penalizes the last $(n-r)$ singular values is minimized. In [11] the authors propose three algorithms. In our experiments TNNR-ADMMAP performed the best and therefore we only include this in our comparison.

The experiment was generated in the same way as in Section 5.1. Both OptSpace and TNNR-ADMMAP try to find the best fixed rank approximation. In contrast our method penalizes a trade-off between the rank and residual

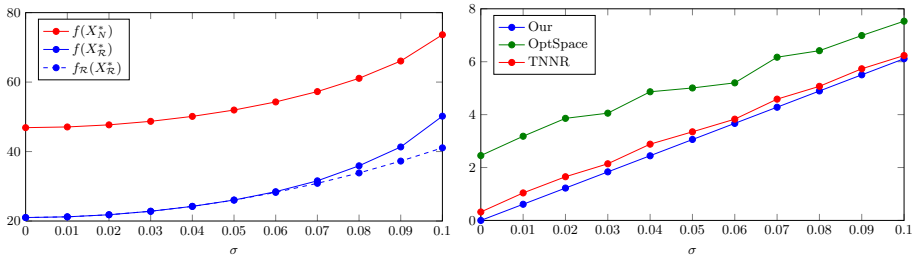

Fig. 4. Evaluation of the proposed formulation on synthetic data with varying noise levels. *Left*: Comparison of our convex relaxation solution $X_{\mathcal{R}}^*$ with the nuclear norm X_N^*. *Right*: The error $\|W \odot (X - M)\|_F$ for varying noise levels.

errors. We therefore iterate our method over μ_i to achieve the correct rank. Typically only one or two iterations are required. To the right in Figure 4, we plot the average values of $\|W \odot (X - M)\|_F$ for the different approaches.

It should be noted that on most problem instances OptSpace and TNNR performed similarly or even slightly better than our approach. The reason for this is that these non-convex approaches try to directly optimize the correct criterion while we use a convex relaxation (and, in addition, our blocks do not cover all the data). Hence, if they converge to the correct solution it will be at least as good as ours. However, as can be seen in Figure 4, local minima raise their average errors.

For the experiment the average run times were 0.88s (our), 20.94s (OptSpace) and 21.44s (TNNR). It is hard to make a fair comparison of different algorithms since it depends on implementations and various tuning parameters. Though we think these running times reflect the fact that we only need to optimize over the blocks and need not keep track of the full matrix while solving our convex problem.

5.3 Real Data

Structure from Motion. We now consider the well-known Oxford dinosaur sequence[3]. The measurement matrix M contains the 2D coordinates of the tracked 3D points. The measurement matrix M will have rank 4 since we do not account for the translation in the affine camera model. In this experiment we consider an outlier free subset consisting of 321 3D points where each point is seen in at least six images. The observation matrix W contains 77% missing data and exhibits the band diagonal structure which is typical for structure from motion problems. The matrix W and the selected blocks for this experiment can be seen in the left graph of Figure 5.

In Figure 6 we show the resulting image point trajectories of our method versus the nuclear norm approach with missing data, cf. (2). We optimize (2)

[3] `http://www.robots.ox.ac.uk/$\sim$vgg/data/data-mview.html`

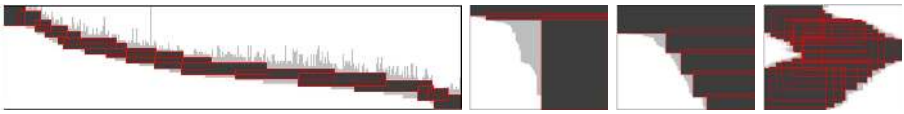

Fig. 5. Structured data patterns of observations (matrix W) and sub-blocks for the *dino*, *book*, *hand* and *banner* sequences.

using ADMM with the shrinkage operator from [6]. For comparison we also show the matrix found by performing a full perspective reconstruction (using bundle adjustment) followed by truncating the reprojections to rank 4. The errors $||W \odot (X - M)||_F$ for the three solutions were 73.2 (our), 1902.5 (nuclear) and 116.2 (perspective). The nuclear norm favors solutions where elements away from the observed data have small values. Therefore its trajectories are stretched towards the origin. Out of the three, the perspective reconstruction captures the turntable motion best. However, its error is higher than our solution and therefore, seen as a low-rank approximation problem, our solution is better. The run times were 14.58s (our) and 18.11s (nuclear).

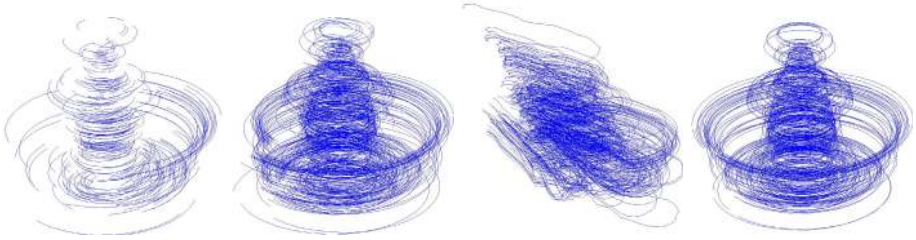

Fig. 6. *From Left to Right*: Observed image point trajectories, results obtained with the proposed algorithm, with the nuclear norm formulation (2) and with perspective reconstruction followed by projecting onto rank 4 using SVD.

Linear Shape Basis Models. Next we consider the linear shape basis model common in non-rigid structure from motion. Let X_f be the 2D- or 3D- coordinates of N tracked points in frame f. The model assumes that in each frame the coordinates are a linear combination of some unknown shape basis vectors, i.e.

$$X_f = \sum_{k=1}^{K} C_{fk} S_k, \quad f = 1, ..., F, \tag{31}$$

where S_k is the shape basis and C is the coefficient matrix. The measurement matrix M consists of the observations of X_f stacked on top of each other. The elements of the observation matrix W indicate if the point was successfully tracked in the particular frame. We search for a solution with as few basis elements as possible by finding a low rank approximation of M.

Fig. 7. Frames 1, 121, 380 and 668 of the *book* sequence. The solution has rank 3.

Fig. 8. Frames 1, 210, 340 and 371 of the *hand* sequence. The solution has rank 5.

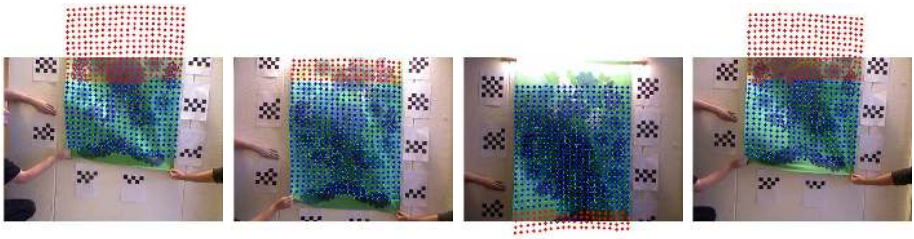

Fig. 9. Frames 70, 155, 357 and 650 of the *banner* sequence. The solution has rank 9.

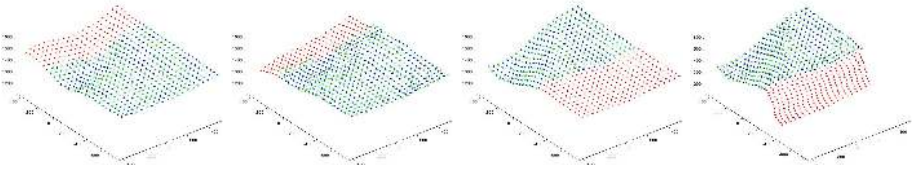

Fig. 10. *From left to right:* Our solution (frame 329), nuclear norm solution (frame 329), our solution (frame 650), nuclear norm solution (frame 650)

We consider three problem instances of varying rank. In the first two (see Figures 7 and 8) we track image points through a video sequence using the standard KLT tracker [16]. Due to occlusion and appearance changes only a subset of the points were successfully tracked throughout the entire sequence. Using (16) we found a low-rank approximation of the measurement matrix. The results can be seen in Figures 7 and 8. Blue dots correspond to the reconstructions of the successfully tracked points. The available measurements for these points are indicated by green crosses. Red dots are the reconstructed point positions for points with no measurements available. The run times for the *book* and *hand* sequences were 34.3s and 35.1s.

In the third problem we used an RGB-D camera to track the 3D-position of a point grid on a moving piece of fabric (see Figure 9). Here missing data is due to limited field of view of the camera and missing depth values. To obtain 3D grid coordinates in a common coordinate system from each video frame we registered the cameras using the patterns visible on the wall. Figures 9 and 10 show the results in 2D and 3D. In Figure 10 we also plot the nuclear norm solution of (2) for comparison. The same tendency to move undetected points towards the origin is visible. The run times for the *banner* sequence were 17.7 min (our) and 2.7 min (nuclear). In this instance we used a large number of blocks. With these blocks and this size of matrix, it is still faster to factorize the entire matrix compared to factorizing all the blocks without parallelization. On average our method spends 0.6s while the nuclear norm approach spends 0.2s per iteration computing SVDs.

The right of Figure 5 shows the data patterns and the selected blocks for the three problems. Note that in the *book* sequence, the blocks leave a relatively large portion of the available measurements unused.

6 Conclusions and Future Work

In this paper we have proposed a method for low-rank approximation of matrices with missing elements. The approach is based on a new convex relaxation - the strongest one possible - of the localized rank function. Unlike the nuclear norm, it is able to avoid penalizing large singular values. Our experiments clearly show the benefits of being able to do so in a convex framework. It should be noted that the presented results are the outputs of our approach without refinement. In cases where the relaxation is not tight, the solution can be used as a starting point for local optimization to obtain even better results.

The proposed method only has to compute the SVD of small sub-matrices, therefore it has potential to tackle large-scale problems. Furthermore, the ADMM approach allows to perform computations in a parallel and distributed manner.

A limitation of the formulation is that in its current form it is sensitive to outliers. The issue has received a lot of attention lately, for example, using the arguably more robust ℓ_1-norm [8,21,24] and it is something that we intend to address in the near future.

In our experiments we have exclusively used rectangular (manually selected) sub-blocks of the measurement matrix. This is however not a limitation. Blocks can be formed from any rows and columns in the matrix. How to select these blocks so as to cover as many measurements as possible and achieve sufficiently large overlaps is still an open problem.

References

1. Angst, R., Zach, C., Pollefeys, M.: The generalized trace-norm and its application to structure-from-motion problems. In: International Conference on Computer Vision (2011)

2. Aquiar, P.M.Q., Stosic, M., Xavier, J.M.F.: Spectrally optimal factorization of incomplete matrices. In: IEEE Conference on Computer Vision and Pattern Recognition (2008)
3. Basri, R., Jacobs, D., Kemelmacher, I.: Photometric stereo with general, unknown lighting. Int. J. Comput. Vision 72(3), 239–257 (2007)
4. Boyd, S., Parikh, N., Chu, E., Peleato, B., Eckstein, J.: Distributed optimization and statistical learning via the alternating direction method of multipliers. Found. Trends Mach. Learn. 3(1), 1–122 (2011)
5. Bregler, C., Hertzmann, A., Biermann, H.: Recovering non-rigid 3d shape from image streams. In: IEEE Conference on Computer Vision and Pattern Recognition (2000)
6. Cai, J.F., Candès, E.J., Shen, Z.: A singular value thresholding algorithm for matrix completion. SIAM J. on Optimization 20(4), 1956–1982 (2010)
7. Candès, E.J., Li, X., Ma, Y., Wright, J.: Robust principal component analysis? J. ACM 58(3), 11:1–11:37 (2011)
8. Eriksson, A., Hengel, A.: Efficient computation of robust weighted low-rank matrix approximations using the L_1 norm. IEEE Trans. Pattern Anal. Mach. Intell. 34(9), 1681–1690 (2012)
9. Garg, R., Roussos, A., de Agapito, L.: Dense variational reconstruction of non-rigid surfaces from monocular video. In: IEEE Conference on Computer Vision and Pattern Recognition (2013)
10. Garg, R., Roussos, A., Agapito, L.: A variational approach to video registration with subspace constraints. Int. J. Comput. Vision 104(3), 286–314 (2013)
11. Hu, Y., Zhang, D., Ye, J., Li, X., He, X.: Fast and accurate matrix completion via truncated nuclear norm regularization. IEEE Trans. Pattern Anal. Mach. Intell. 35(9), 2117–2130 (2013)
12. Jacobs, D.: Linear fitting with missing data: applications to structure-from-motion and to characterizing intensity images. In: IEEE Conference on Computer Vision and Pattern Recognition (1997)
13. Jojic, V., Saria, S., Koller, D.: Convex envelopes of complexity controlling penalties: the case against premature envelopment. In: International Conference on Artificial Intelligence and Statistics (2011)
14. Keshavan, R.H., Montanari, A., Oh, S.: Matrix completion from a few entries. IEEE Trans. Inf. Theory 56(6), 2980–2998 (2010)
15. Lewis, A.S.: The convex analysis of unitarily invariant matrix functions (1995)
16. Lucas, B.D., Kanade, T.: An iterative image registration technique with an application to stereo vision. In: International Joint Conference on Artificial Intelligence (1981)
17. Olsen, S., Bartoli, A.: Implicit non-rigid structure-from-motion with priors. Journal of Mathematical Imaging and Vision 31(2-3), 233–244 (2008)
18. Olsson, C., Oskarsson, M.: A convex approach to low rank matrix approximation with missing data. In: Scandinavian Conference on Image Analysis (2009)
19. Recht, B., Fazel, M., Parrilo, P.A.: Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization. SIAM Rev. 52(3), 471–501 (2010)
20. Rockafellar, R.: Convex Analysis. Princeton University Press (1997)
21. Strelow, D.: General and nested Wiberg minimization. In: IEEE Conference on Computer Vision and Pattern Recognition (2012)
22. Wang, S., Liu, D., Zhang, Z.: Nonconvex relaxation approaches to robust matrix recovery. In: International Joint Conference on Artificial Intelligence (2013)

23. Yan, J., Pollefeys, M.: A factorization-based approach for articulated nonrigid shape, motion and kinematic chain recovery from video. IEEE Trans. Pattern Anal. Mach. Intell. 30(5), 865–877 (2008)
24. Zheng, Y., Liu, G., Sugimoto, S., Yan, S., Okutomi, M.: Practical low-rank matrix approximation under robust L_1-norm. In: IEEE Conference on Computer Vision and Pattern Recognition (2012)