A Simple Method for Subspace Estimation with Corrupted Columns

Viktor Larsson
Lund University
viktorl@maths.lth.se

Carl Olsson
Lund University
calle@maths.lth.se

Fredrik Kahl
Chalmers University of Technology
fredrik.kahl@chalmers.se

Abstract

This paper presents a simple and effective way of solving the robust subspace estimation problem where the corruptions are column-wise. The method we present can handle a large class of robust loss functions and is simple to implement. It is based on Iteratively Reweighted Least Squares (IRLS) and works in an iterative manner by solving a weighted least-squares rank-constrained problem in every iteration. By considering the special case of column-wise loss functions, we show that each such surrogate problem admits a closed form solution. Unlike many other approaches to subspace estimation, we make no relaxation of the low-rank constraint and our method is guaranteed to produce a subspace estimate with the correct dimension.

Subspace estimation is a core problem for several applications in computer vision. We empirically demonstrate the performance of our method and compare it to several other techniques for subspace estimation. Experimental results are given for both synthetic and real image data including the following applications: linear shape basis estimation, plane fitting and non-rigid structure from motion.

1. Introduction

Subspace estimation problems appear as a subtask in many seemingly different applications, such as motion segmentation, structure from motion and face recognition. The problem can be posed as finding a low-rank matrix approximation of a given measurement matrix. Naturally, the problem has been investigated in depth. Recent work has focused on robustness to outliers and noise, leading to new robustified versions of standard techniques, for instance, Robust Principal Component Analysis (RPCA) [6]. In this work, we present a simple and yet effective method for robust rank approximation, which is capable of handling a large class of robust loss functions and which is guaranteed to find a solution of a given rank. Hence, no relaxation of the rank constraint is required.

There are essentially two approaches to robust rank approximation. The first one is based on relaxing the original hard, non-convex problem to a convex optimization problem, e.g., [7, 6, 27, 21, 28]. Many of these methods simply replace the rank function with the nuclear norm. In [7], it is shown that for their model, the method comes with theoretical performance guarantees of recovering the correct solution provided the low-rank matrix satisfies so called incoherence conditions and that the outliers are sparse and uniformly distributed. However, for many problems these assumptions are not in general fulfilled and the method risks breaking down to such violations. Further, it has been noted that when the dimension of the underlying subspace is known, the performance of nuclear norm based methods is worse compared with other approaches that ensure the correct subspace dimension [5]. This leads us to a second class of approaches that do not apply relaxation, and instead work directly with the original non-convex problem. Bilinear formulations are common which make sure that the recovered solution has the correct rank [1, 17, 12, 24]. Often, alternating optimization techniques are applied which are known to be sensitive to the local minima problem, and hence they require a good initialization.

We will follow the second class of approaches and work directly with the rank constraint. Our method is based on Iteratively Reweighted Least Squares (IRLS) and this by itself is not new for robust subspace estimation. In [19, 18], IRLS is analyzed and used for minimizing the nuclear norm relaxation of the rank function. It is shown that this can lead to efficient and convergent algorithms for solving the convex program, but it still suffers from the disadvantages of relaxation. We will experimentally compare our method to those that use nuclear norm relaxation for similar problem formulations. IRLS has also been used in a bilinear formulation of the robust subspace estimation problem [1]. Here, each step in the IRLS is solved by a heuristic optimization technique by iteratively solving a series of approximate problems called *surrogates*. However, there is no guarantee of convergence for the surrogate problems and as a non-convex problem is solved, there is a risk of getting trapped in a local minimum.

The work most similar to ours is perhaps [11], where IRLS is also used to solve robust subspace estimation problems. In [11] the loss functions considered are robust to element-wise corruptions. While this is a more general error model it comes with the drawback that the update step becomes much more difficult to solve. The authors in [11] propose an alternating minimization scheme to solve the weighted least squares problem which has no optimality guarantees. In contrast, this work considers the special case of column-wise outliers. For this case we show that the update has a closed form solution which is globally optimal. Further, as there is a closed form solution, the method becomes very simple to implement and analyze. We prove that under mild conditions the objective function decreases in each iteration. The conditions are satisfied for a large family of robust loss functions which includes the column-wise ℓ_1-loss and Huber-loss. The proposed method has in experiments been observed to converge very quickly and in other applications IRLS based methods have been shown to converge exponentially fast (linear convergence) when close to the optimum [10].

In summary, we propose an IRLS method for robust subspace estimation which is very simple to implement and contains no tuning parameters except for the choice of robust loss function. We can handle a large class of robust loss functions and the solution is guaranteed to have (at most) a pre-defined rank. Each iteration decreases the objective function but as the rank problem is non-convex, we cannot be certain that the computed solution is globally optimal. It has been tested on a number of applications in computer vision, including linear shape basis estimation, plane fitting and non-rigid structure from motion and we show competitive performance in both speed and solution quality.

Notation. For a matrix X, the notation X_k refers to the k-th column of X. The Frobenius norm of X is given by $\|X\|_F^2 = \sum_k \|X_k\|^2 = \sum_{i,k} |X_{ik}|^2$, whereas the $\ell_{2,1}$-norm is defined as $\|X\|_{2,1} = \sum_k \|X_k\|$. The nuclear norm is defined as the sum of singular values of X, i.e., $\|X\|_* = \sum_k \sigma_k(X)$. The pseudo-inverse is denoted $X^\dagger$.

2. Robust Subspace Estimation Problems

In this paper we are mainly interested in low-rank approximation problems where the data term is robust to incorrect columns. The robust minimization problems we consider are of the form

$$\min_X \sum_k \varphi(\|X_k - M_k\|) \quad \text{s.t.} \quad X \in \mathcal{C} \tag{1}$$

where $\mathcal{C}$ is some constraint set and $\varphi : \mathbb{R}^+ \to \mathbb{R}^+$ is some robust loss. The constraint set $\mathcal{C}$ is typically used to restrict the columns of X to have some simple structure, e.g. belong to a low-dimensional subspace. Some examples of constraint sets $\mathcal{C}$ which are tractable are shown in Table 1 but other choices are possible in the framework as well.

For the robust loss φ the two choices we will consider are $\varphi(x) = x$, which gives us the $\ell_{2,1}$-norm and

$$\varphi(x) = \begin{cases} \frac{1}{2}x^2 & \text{if } x \leq \delta \\ \delta x - \frac{1}{2}\delta^2 & \text{if } x \geq \delta \end{cases}$$

which gives the well known Huber-loss. In the experiments throughout the paper we will use the Huber loss except for the experiment in Section 5.1 where the $\ell_{2,1}$-loss is used. The derivations for the IRLS method are performed for a general loss function φ. We only require that the function $\psi(x) := \varphi(\sqrt{x})$ is concave and differentiable (except at a finite number of points), which is the case for the previously mentioned loss functions.

Problem	Constraint set
Subspace estimation	$\mathcal{C} = \{Z \mid \text{rank}(Z) = k\}$
Affine space estimation	$\mathcal{C} = \{Z + t\mathbb{1}^T \mid \text{rank}(Z) = k, t \in \mathbb{R}^n\}$
Dictionary selection	$\mathcal{C} = \{AZ \mid \text{rank}(Z) = k\}$
Affine Non-rigid SfM[1]	$\mathcal{C} = \{RZ + t\mathbb{1}^T \mid \text{rank}(Z) = k\}$

Table 1. Some examples of possible constraints in the framework.

3. Iteratively Reweighted Least Squares with Constraints

To minimize (1) we apply IRLS which has been extensively used in the optimization literature for the approximation of robust norms, e.g. [10, 16]. In IRLS the idea is to solve a sequence of weighted least squares problems which converge to the true cost. In each iteration we solve problems of the following form

$$\min_X \sum_k w_k^2 \|X_k - M_k\|^2 \quad \text{s.t.} \quad X \in \mathcal{C} \tag{2}$$

where $w_k \in \mathbb{R}^+$ are the sample specific weights. This can be rewritten in matrix form as

$$\min_X \|(X - M)W\|_F^2 \quad \text{s.t.} \quad X \in \mathcal{C} \tag{3}$$

where W is a diagonal matrix containing the weights. The weights are selected by requiring that the derivatives of $\varphi(r_k)$ and $w_k^2 r_k^2$ with respect to r_k are equal, where $r_k = \|X_k - M_k\|$, $k = 1, \dots, N$, which implies that

$$w_k^2 = \frac{\varphi'(r_k)}{2r_k}. \tag{4}$$

In order to be able to employ our IRLS approach we will require that the weighted projection $\mathcal{P}_\mathcal{C}^W$ onto the constraint set $\mathcal{C}$ (in Frobenius norm), i.e.,

$$\mathcal{P}_\mathcal{C}^W(Z) = \underset{X \in \mathcal{C}}{\text{argmin}} \|(X - Z)W\|_F^2, \tag{5}$$

can be computed. Given the projection operator the solution to (3) is given by $X^\star = \mathcal{P}_\mathcal{C}^W(M)$. The complete algorithm is summarized in Algortithm 1.

4. Convergence Analysis

Next we show that the sequence of minimizers generated by the IRLS method gives decreasing objective values. Since the objective values are bounded from below they must converge.

[1] We assume known cameras, i.e., R,t are known.

while *not converged* **do**

$r_k = \|X_k - M_k\|$

$W_{kk} = (\varphi'(r_k)/(2r_k))^{1/2}$

$X = \mathcal{P}_{\mathcal{C}}^W(M)$

end

Algorithm 1: Constrained IRLS.

Theorem 1. *Let X^t denote the sequence of minimizers generated from Algorithm 1. Then the function values $f(X^t)$ satisfy*

$$f(X^{t+1}) \leq f(X^t)$$

and the sequence $f(X^t)$ converges.

Proof. The proof is easily obtained using a very elegant argument from [16, 2]. Using the previously defined $\psi(x) := \varphi(\sqrt{x})$ we can rewrite the optimization problem

$$\min_X \sum_k \psi(\|X_k - M_k\|^2) \quad \text{s.t.} \quad X \in \mathcal{C}. \tag{6}$$

Let $r_{k,t}$ denote the k-th residual at iteration t and similarly for $w_{k,t}$. Since we solve the weighted least squares problem optimally on the constraint set $\mathcal{C}$ we must have

$$\sum_k w_{k,t}^2 r_{k,t+1}^2 - \sum_k w_{k,t}^2 r_{k,t}^2 \leq 0. \tag{7}$$

Since ψ is assumed to be concave, it follows that

$$\psi(r_{k,t+1}^2) - \psi(r_{k,t}^2) \leq \psi'(r_{k,t}^2)(r_{k,t+1}^2 - r_{k,t}^2). \tag{8}$$

Note that using (4), it follows that $\psi'(r_{k,t}^2) = w_{k,t}^2$. From (7) we then get

$$\sum_k \psi(r_{k,t+1}^2) - \sum_k \psi(r_{k,t}^2) \leq 0. \tag{9}$$

So the function values are decreasing in each iteration and since they are bounded from below they must converge. □

4.1. Convergence Rate

In this section we empirically evaluate the convergence rate of the proposed approach for the robust subspace estimation problem on synthetic data. In the subspace estimation problem we have the constraint set

$$\mathcal{C} = \{X \mid \operatorname{rank}(X) = r\}. \tag{10}$$

The set of rank r matrices can be parameterized by two thin matrices as $X = AB$ where $A \in \mathbb{R}^{m\times r}$, $B \in \mathbb{R}^{r\times n}$. In this parametrization (1) can be rewritten as

$$\min_{A,B} \sum_k \varphi(\|AB_k - M_k\|). \tag{11}$$

We compare our IRLS approach with the standard Conjugate Gradient method applied to this parametrization. We compare with using both the Fletcher-Reeves [13] and Polak-Ribière [20] update rules.

The columns of $M \in \mathbb{R}^{100\times 1000}$ are sampled from a randomly generated subspace of $\mathbb{R}^{100}$ and corrupted with small Gaussian noise with zero mean. Then 25% of the columns are replaced by Gaussian noise to simulate outliers. Figure 1 shows the function values plotted against iterations for one of the instances. Table 2 shows the average number of iterations required until convergence over 1000 instances. For both methods we consider the method converged when the relative error (to the optimal function value) is smaller than 10^{-6}.

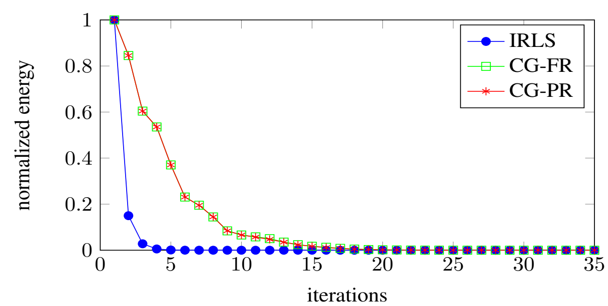

Figure 1. Convergence comparison with Conjugate Gradient. The function values have been shifted to the interval $[0, 1]$ for presentation. *CG-FR* corresponds to the β update rule by Fletcher-Reeves [13] and *CG-PR* to the rule from Polak-Ribière [20]. Note that the IRLS approach converges to the optimum within the first few iterations and that the two conjugate gradients methods have very similar performance.

	IRLS	CG-FR	CG-PR
Mean	7.2	28.7	24.2
Median	5.0	22.0	22.0

Table 2. The average number of iterations for convergence. The average time for each iteration was 0.013 seconds for IRLS and 0.034 seconds for the CG methods. Note that CG methods are using exact line search by numerical optimization in MATLAB.

4.2. Robust Initialization

Since IRLS is a local optimization method having a good initialization is essential. Here we only consider the subspace estimation problem, $\mathcal{C} = \{X \mid \operatorname{rank}(X) = r_0\}$ but the approach generalizes to the other constraints in Table 1. To find a good initialization we propose a random sampling approach. We randomly select r columns and and take these as a basis for the column space. Let $B \in \mathbb{R}^{n\times r_0}$ be the matrix containing the samples. Then the problem reduces to

$$\min_{Z} \sum_{k} \varphi(\|BZ_k - M_k\|) \tag{12}$$

which is separable in the columns of Z, so we can solve for each Z_k separately. We iterate this procedure and keep the solution with best objective value.

5. Experimental Evaluation

5.1. Comparison to Convex Relaxation

A popular heuristic for solving rank approximation problems is to replace the non-convex rank constraint with a nuclear norm penalty term. Due to the convexity of the nuclear norm the resulting problem can be solved using conventional convex optimization methods. On the downside the nuclear norm penalizes all singular values equally in contrast to the desired regularization where only the smaller singular values are penalized. In this section we present an experiment where we compared our method to the convex relaxation approach used in [27, 21]. In these works the authors consider the following version of Robust Principal Component Analysis (RPCA)

$$\min_{X} \quad \lambda \|X\|_* + \|X - M\|_{2,1}, \tag{13}$$

as a convex substitute for

$$\min_{X} \quad \lambda \operatorname{rank}(X) + \|X - M\|_{2,1}. \tag{14}$$

In the experiment we generated 1000 points on a 10-dimensional subspace in $\mathbb{R}^{100}$. The points were corrupted by noise and 25% of the points were also heavily corrupted and became outliers. We solved (13) for varying values of the parameter λ using ADMM [3]. Note that the trailing singular values of the minimizers to (13) will generally be small but non-zero. To determine the rank we counted the number of singular values which were larger than $0.01\sigma_1(M)$. Then to find a true low-rank solution we truncated the remaining singular values. We also compared to simply projecting the measurement matrix to the correct rank using SVD. The result can be seen in Figure 2. Note that the uncorrupted measurement matrix has rank 10. Different error metrics for the rank 10 solutions can be seen in Table 3. The runtimes for the rank 10 solutions were 0.068s for IRLS and 1.056s for RPCA-(2,1).

5.2. Robust Shape Model Estimation

In this section we consider the problem of robust linear shape model estimation. We assume that we are given some samples M_k in $\mathbb{R}^n$ which we want to describe by a low-dimensional shape basis B, i.e.,

$$M_k = BZ_k \quad B \in \mathbb{R}^{n\times r}, Z_k \in \mathbb{R}^r$$

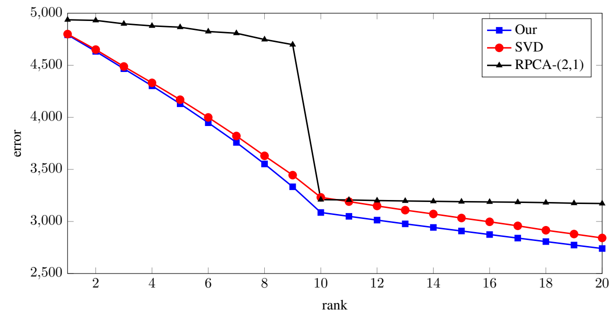

Figure 2. The error against the noisy data measured on all samples in the $\ell_{2,1}$-norm plotted against rank. Note that there is no value of λ for which the error is as low as for the IRLS method.

	Our	SVD	RPCA-(2,1)
Noisy data (all samples)	**3086**	3232	3209
Noisy data (inliers)	**706**	888	779
Ground truth (inliers)	**252**	594	390

Table 3. The errors for the rank 10 solutions. The first row corresponds to Figure 2. The second is the error w.r.t. the noisy data measured only on the inliers. Finally the third row is the error w.r.t. the ground truth data measured only on the inliers. All the errors are measured in the $\ell_{2,1}$-norm.

where r is some small integer. If we assume that the samples are corrupted by Gaussian noise on the samples, then the maximum likelihood estimate is given by

$$\min_{B,Z} \|BZ - M\|_F^2 = \min_{\operatorname{rank}(X)=r} \|X - M\|_F^2. \tag{15}$$

If some outlier samples are present in the data, then the squared error can distort the solution heavily and we instead consider the robustified version

$$\min_{\operatorname{rank}(X)=r} \sum_{k} \varphi(\|X_k - M_k\|). \tag{16}$$

We consider the hand dataset from [23]. The dataset contains 40 images of hands in various poses. For each image 56 points along the edge of the hand are given. See Figure 3 for some examples.

The dataset is outlier free so to evaluate the robustness of the method we corrupt the first five images by shrinking one finger in each image. Using the proposed method we find a five dimensional shape basis for the hands. We compare with the result of solving (15) using SVD which is standard in Active Shapes [8]. Figure 4 shows how well each sample can be represented in each shape basis. Note that the first five samples correspond to the outliers. The robust objective function allows for solutions where a few samples have high error. Some qualitative results can be seen in Figures 5 and 6.

To evaluate the performance we measured the error on the 35 inlier samples. Table 4 shows the errors in Frobe-

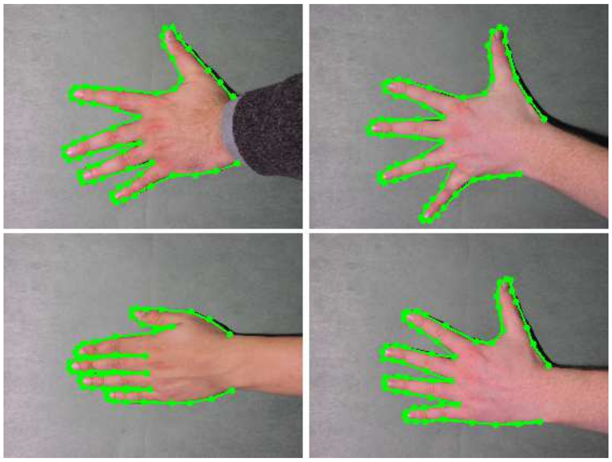

Figure 3. Example images from the hand dataset [23].

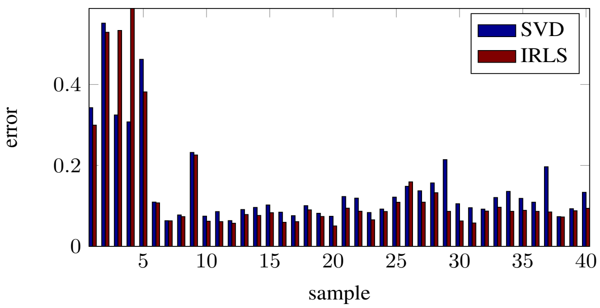

Figure 4. Reconstruction error for the individual samples, i.e., $\|X_k - M_k\|$.

nius norm for the inlier samples. For comparison we also include the results from minimizing the error in the elementwise ℓ_1 norm. The minimization is performed by using the augmented lagrangian method which was proposed in [15] for solving the problem

$$\min_{Z,U,V} \|M - Z\|_1 \quad \text{s.t.} \quad Z = UV^T \tag{17}$$

The table also includes the optimal estimates found by using only unknown the inliers.

	Optimal	IRLS	ALM-ℓ_1	SVD
Error	0.5268	0.5513	0.5787	0.6967

Table 4. Error in Frobenius norm measured only on the inliers. *Optimal* is found by performing SVD on only the inliers.

5.3. Affine Subspace Estimation

Next we consider the problem of estimating an affine subspace

$$\min_{Z,t} \sum_k \varphi(\|Z_k + t - M_k\|) \quad \text{s.t.} \quad \text{rank}(Z) = d. \tag{18}$$

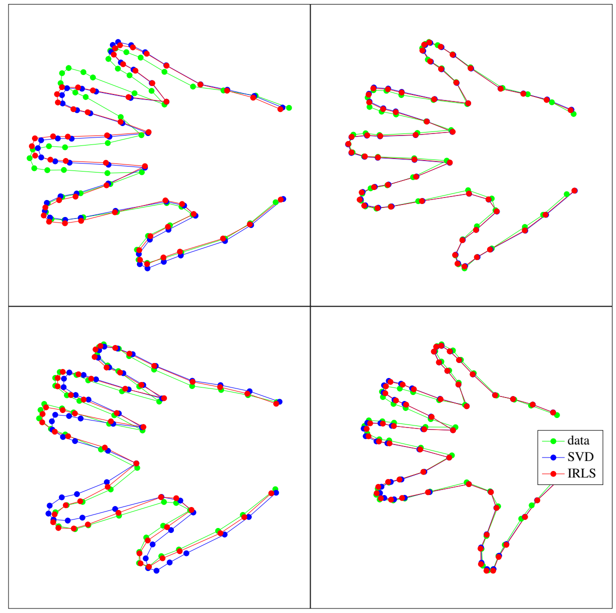

Figure 5. Example of result for some inlier samples. The input data is shown in green, the output of our method in red and the output from SVD in blue.

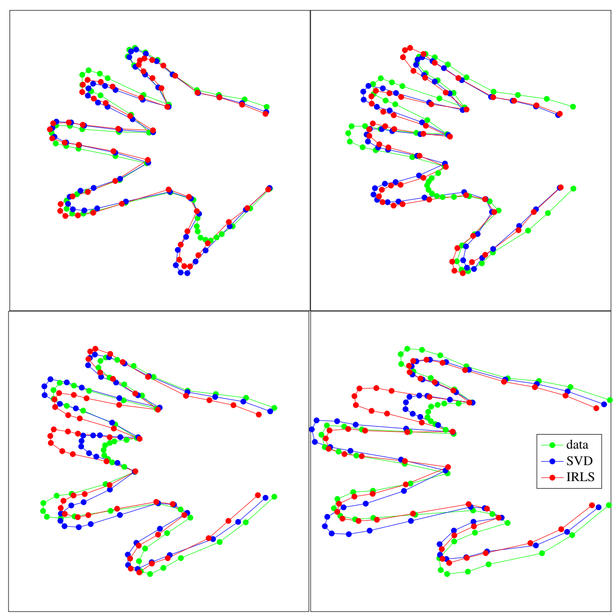

Figure 6. Example of result for some outlier samples. The input data is shown in green, the output of our method in red and the output from SVD in blue.

The projection onto an affine subspace of dimension d is given by

$$\mathcal{P}_{\mathcal{C}}^{W}(M) = \underset{Z,t}{\operatorname{argmin}} \left\|(Z + t\mathbb{1}^T - M)W\right\|_F^2 \quad \text{s.t. } \text{rank}(Z) = d. \tag{19}$$

Differentiating w.r.t. t we get

$$(Z+t^\star \mathbb{1}^T - M)W^2\mathbb{1} = 0 \implies t^\star = \frac{-1}{\mathbb{1}^T W^2 \mathbb{1}}(Z-M)W^2\mathbb{1}. \tag{20}$$

The remaining minimization in Z then becomes

$$\min_Z \left\| (Z-M)(I - \frac{1}{\mathbb{1}^T W^2 \mathbb{1}} W^2 \mathbb{1}\mathbb{1}^T) \right\|_F^2 \quad \text{s.t. } \operatorname{rank}(Z) = d. \tag{21}$$

If we denote $A = (I - \frac{1}{\mathbb{1}^T W^2 \mathbb{1}} W^2 \mathbb{1}\mathbb{1}^T)$ we get from the arguments in the appendix that the minimum norm solution is given by

$$Z^\star = \mathcal{P}_d\left(MAV_1\right)\Sigma^{-1}U_1^T \tag{22}$$

where A has the SVD,

$$A = \begin{bmatrix} U_1 & U_2 \end{bmatrix} \begin{bmatrix} \Sigma & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} V_1^T \\ V_2^T \end{bmatrix}. \tag{23}$$

The IRLS algorithm for affine subspace estimation is summarized in Algorithm 2.

while *not converged* **do**
- $r_i = \|X_i - M_i\|$
- $W_{ii} = (\varphi'(r_i)/(2r_i))^{1/2}$
- $A = (I - \frac{1}{\mathbb{1}^T W^2 \mathbb{1}} W^2 \mathbb{1}\mathbb{1}^T)$
- $U_1 \Sigma V_1 = svd(A)$
- $Z = \mathcal{P}_d(MV_1)\Sigma U_1^T$
- $t = \frac{-1}{\mathbb{1}^T W^2 \mathbb{1}}(Z-M)W^2\mathbb{1}$
- $X = (Z + t\mathbb{1}^T)$

end

Algorithm 2: IRLS for affine subspaces.

5.3.1 Experiment: Plane Fitting in 3D

Now to evaluate the effect by the robust initialization we consider the problem of robust 3D plane fitting. The problem can be thought of as an affine subspace estimation problem where we are given points in $\mathbb{R}^3$ and want to fit a two dimensional affine subspace. Note that the problem of 3D plane fitting can in general be solved extremely well by random sampling approaches due to the low dimensionality of the problem and this experiment is only used to illustrate the behavior of the robust initialization.

We generated data for the experiment by sampling 100 points on a random plane in $\mathbb{R}^3$. We then added small noise to these points and for a subset of the points we also added large noise, thus creating some outlier samples. We then projected the points onto a two dimensional affine space using Algorithm 2 initialized both using the standard Total Least Squares estimate and using the robust initialization from Section 4.2. In the experiment the Huber loss was used. We ran the experiment for varying levels of noise and number of outliers. The averaged results can be seen in Figure 7. We can see that for low number of outliers the robust initialization makes no difference.

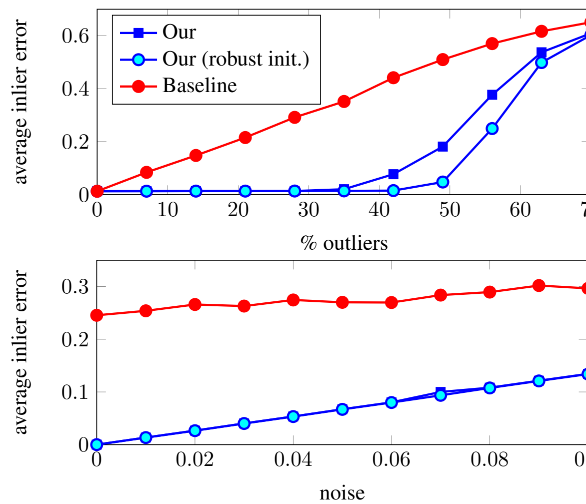

Figure 7. Synthetic experiment with plane fitting in 3D. The graphs show the average inlier error for varying levels of noise and number of outliers. *Top:* The percentage of outliers is varied and the noise level is kept fix at $\sigma = 0.01$. *Bottom:* The noise level is varied and the number of outliers is kept fix at 25%. The baseline is Total Least Squares.

5.4. Non-Rigid Structure from Motion

In Non-Rigid Structure from Motion (NRSfM) we want to reconstruct a dynamic scene from image measurements. We consider the case of known affine cameras[2]. Let $(R_k,\ t_k) \in \mathbb{R}^{2\times 3} \times \mathbb{R}^2, k = 1,..,F$, be the known cameras and construct

$$R = \operatorname{diag}(R_1, ..., R_F), \quad t = \begin{bmatrix} t_1^T & \dots & t_F^T \end{bmatrix}^T. \tag{24}$$

We assume that the measurement matrix $M \in \mathbb{R}^{2F\times N}$ is formed as $M = RX + t\mathbb{1}^T$ where $X \in \mathbb{R}^{3F\times N}$ is the unknown 3D-points in each frame stacked vertically. To regularize the solution we make a linear shape basis assumption which is common in NRSfM [4]. This implies that $\operatorname{rank}(X) \leq 3K$ where K is the number of basis elements.

To perform non-rigid structure from motion which is robust to outlier tracks we want to solve the problem

$$\min_X \sum_k \varphi(\|RX_k + t - M_k\|) \quad \text{s.t.} \quad \operatorname{rank}(X) = 3K. \tag{25}$$

We now consider the weighted projection operator onto constraint set $\mathcal{C} = \{RZ + t\mathbb{1}^T \mid \operatorname{rank}(Z) = 3K\}$, i.e.,

$$\mathcal{P}_{\mathcal{C}}^W(M) = \operatorname*{argmin}_{\operatorname{rank}(Z)=3K} \left\|(RZ + t\mathbb{1}^T - M)W\right\|_F^2. \tag{26}$$

[2]This step is only a subtask in the complete problem.

Let $R = U \begin{bmatrix} \Sigma & 0 \end{bmatrix} \begin{bmatrix} V_1 & V_2 \end{bmatrix}^T$ be the SVD of R, then by the arguments in the appendix the optimal solution to (26) is given by

$$Z^\star = (V_1 Y_1^\star + V_2 Y_2) W^{-1}, \tag{27}$$

$$Y_1^\star = \Sigma^{-1} \mathcal{P}_{3K} \left(U^T (M - t \mathbb{1}^T) W \right), \tag{28}$$

and Y_2 is any matrix such that $\operatorname{rank}(\begin{bmatrix} Y_1^T & Y_2^T \end{bmatrix}^T) \leq 3K$. Let $Y_1^\star = \bar{U} \bar{\Sigma} \bar{V}^T$ be the thin SVD of $Y_1^\star$. Then any choice of Y_2 which preserves the rank can be written as $Y_2 = \Theta \bar{V}^T$. The choice of Y_2 affects the solution along the nullspace of R, i.e., along the principal axis of the cameras. This has an large effect on the quality of the solution and the minimum norm solution will favor points close to the camera plane. To choose Y_2 we assume a smoothness prior which was suggested in [9]. Let D be the first order finite difference matrix. We then solve the problem

$$\min_{\Theta} \; \| D Z(\Theta) \| = \min_{\Theta} \; \left\| D (V_1 Y_1^\star + V_2 \Theta \bar{V}^T) W^{-1} \right\|. \tag{29}$$

This is a simple least squares problem with the solution

$$\Theta = -(D V_2)^\dagger \left(D V_1 Y_1^\star W^{-1} \right) (\bar{V}^T W^{-1})^\dagger. \tag{30}$$

The algorithm is summarized below.

$U \begin{bmatrix} \Sigma & 0 \end{bmatrix} \begin{bmatrix} V_1^T \\ V_2^T \end{bmatrix} = \text{svd}(R)$

while *not converged* **do**

$r_i = \| R Z_i + t - M_i \|$

$W_{ii} = (\varphi'(r_i)/r_i)^{1/2}$

$Y_1 = \Sigma^{-1} \mathcal{P}_{3K} \left(U^T (M - t \mathbb{1}^T) W \right)$

$\bar{U} \bar{\Sigma} \bar{V}^T = \text{svd}(Y_1)$

$\Theta = -(D V_2)^\dagger \left(D V_1 Y_1^\star W^{-1} \right) (\bar{V}^T W^{-1})^\dagger.$

$Z = (V_1 Y_1 + V_2 \Theta \bar{V}^T) W^{-1}$

end

Algorithm 3: Rank Constrained IRLS for affine NRSfM.

5.4.1 Experiment: CMU Motion Capture Database

We now consider data from the CMU Motion Capture database [25]. The dataset consists of 3D point tracks obtained using a motion capture system. To generate the measurement matrix M we projected the 3D points into synthetic orthographic cameras and added some small Gaussian noise. The orthographic cameras rotated around the subject at ≈ 1 degree per frame. Since we are interested in evaluating the robustness to outliers we also add some tracks consisting of random noise. The sequence we considered[3] consists of 330 points tracked in 740 frames. We added an additional 100 outlier tracks.

[3] Subject 26, Trial 9.

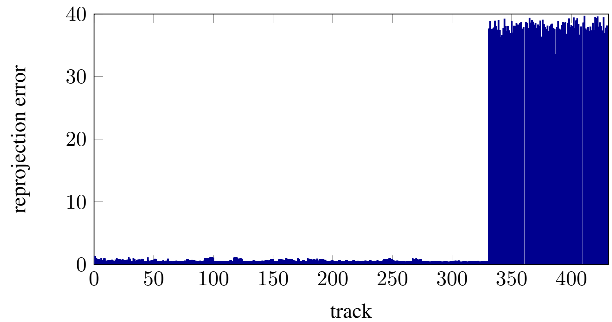

Figure 8. The reprojection errors for the sequence 26-09 in the CMU Motion Capture database. The last 100 tracks are outliers.

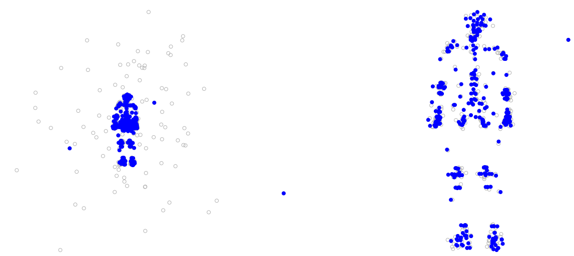

Figure 9. Reprojections in frame 650 of all points (*left*) and only inliers (*right*). The gray circles are the measurements.

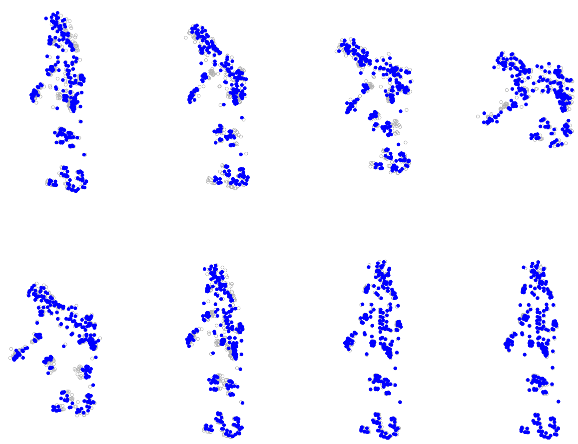

Figure 10. The reconstruction for the sequence 26-09 in the CMU Motion Capture database. The outlier tracks have been removed. The gray circles are the ground truth 3D points.

Figure 8 shows the reprojection errors obtained using the proposed algorithm. Due to the robust formulation the outlier tracks are able to have a high reprojection error without degrading the rest of the solution. Note that without the rank constraint every track would have had zero reprojection error. Figure 9 shows the reprojections of a single frame with and without the points indicated as outliers and Figure 10 shows the 3D reconstruction together with the ground truth. Note that two of the outlier tracks were incorrectly regarded as inliers. This is due to them being close enough to the subspace to be indistinguishable from a true inlier.

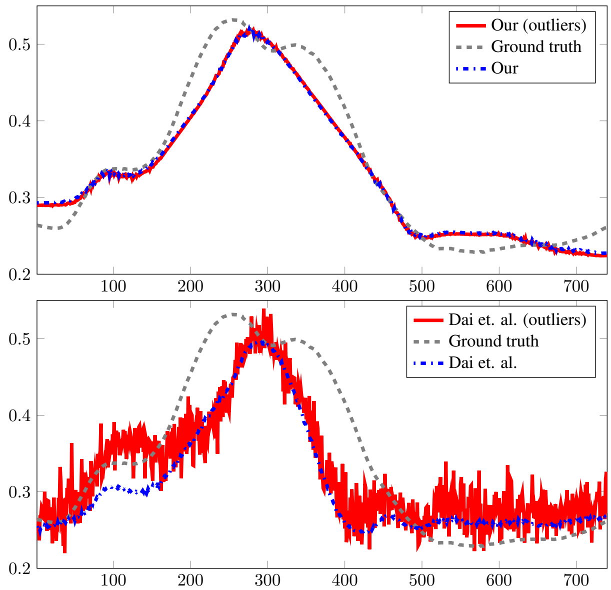

Figure 11. The x-coordinate for a single point across all frames for the different reconstructions.

	3D error		Time
	No outliers	Added outliers	
Our	0.0968	0.1228	27.1s
Dai et al. [9]	0.1405	0.4257	231.8s

Table 5. Normalized mean 3D error measured only on the inliers with $K = 3$ and the total running time.

We also compared our results to the non-robust method proposed by Dai et al. [9]. In their method they estimate the cameras and structure separately. Here we assume known cameras so we only use the structure estimation part of their algorithm together with the ground truth cameras. In [9] they find the structure by solving the convex problem

$$\min_{X} \ \|X^{\#}\|_* \quad \text{s.t.} \quad RX = \mathcal{P}_{3K}\left(M - t\mathbb{1}^T\right) \tag{31}$$

where $X^{\#}$ denotes the stacked matrix where each row corresponds to a frame, i.e., $X \in \mathbb{R}^{3F\times N}$ and $X^{\#} \in \mathbb{R}^{F\times 3N}$. After solving the optimization problem they project the stacked matrix to rank K to ensure a valid shape basis factorization exists.

To evaluate the reconstructions we measure the normalized mean 3D error given by

$$e_{3D} = \frac{1}{\sigma F N}\sum_{f}^{F}\sum_{n}^{N} e_{fn}, \quad \sigma = \frac{1}{3F}\sum_{f}^{F}(\sigma_{fx}+\sigma_{fy}+\sigma_{fz}), \tag{32}$$

where σ_{fx}, σ_{fy} and σ_{fz} denote the standard deviation of the x, y, and z-coordinates in ground truth shape. This is the same error metric that was used in [9]. Table 5 shows the errors for the experiment with and without outliers. When outliers are added the quality of solution is heavily degraded. Figure 11 shows the x-coordinate of a single point in all frames for the two methods.

6. Conclusions

We have shown that several subspace estimation problems can be robustly performed using IRLS. The proposed method is competitive both in terms of solution accuracy and the running times compared to other approaches. The method can simply be implemented in a few lines of Matlab code and have few if any tuning parameters.

A. Rank Constrained Least Squares

In this section we give a short review of how to estimate a rank constrained matrix under the Frobenius norm. The approach seems to have first appeared in [22] and has later been used in [14, 26]. We consider optimization problems of the type

$$\min_{X} \|AXB - C\|_F^2 \quad \text{s.t.} \quad \operatorname{rank}(X) = d. \tag{33}$$

If the matrices A and B are invertible the solution is easily found using

$$X^\star = A^{-1}\mathcal{P}_d\left(C\right)B^{-1}, \tag{34}$$

where $\mathcal{P}_d\left(C\right)$ is projection onto the rank d matrices (in the Frobenius norm sense). For the general case we let

$$A = \begin{bmatrix} U_{A1} & U_{A2} \end{bmatrix} \begin{bmatrix} \Sigma_A & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} V_{A1}^T \\ V_{A2}^T \end{bmatrix}, \tag{35}$$

and

$$B = \begin{bmatrix} U_{B1} & U_{B2} \end{bmatrix} \begin{bmatrix} \Sigma_B & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} V_{B1}^T \\ V_{B2}^T \end{bmatrix}, \tag{36}$$

be the singular value decompositions of A and B respectively. Inserting into the objective function of (33) we see after some manipulations that it reduces to

$$\left\|\Sigma_A Y_{11}\Sigma_B - U_{A1}^T M V_{B1}\right\|_F^2, \tag{37}$$

where

$$Y = \begin{bmatrix} Y_{11} & Y_{12} \\ Y_{21} & Y_{22} \end{bmatrix} = \begin{bmatrix} V_{A1}^T \\ V_{A2}^T \end{bmatrix} X \begin{bmatrix} U_{B1} & U_{B2} \end{bmatrix}. \tag{38}$$

Note that the matrices Σ_A and Σ_B are of full rank and therefore the optimal Y_{11} is given by

$$Y_{11} = \Sigma_A^{-1}\mathcal{P}_d\left(U_{A1}^T M V_{B1}\right)\Sigma_B^{-1}. \tag{39}$$

The minimum norm solution is obtained by setting the remaining blocks of Y to zero. For the general set of solutions we see that $Y_{21} = \Theta_r Y_{11}$, that is, the rows of Y_{21} must be linear combinations of the rows of Y_{11}. A similar argument for the other blocks gives the general solution to (33) as

$$X^\star = \begin{bmatrix} V_{A1} & V_{A2} \end{bmatrix} \begin{bmatrix} Y_{11} & Y_{11}\Theta_c \\ \Theta_r Y_{11} & \Theta_r Y_{11}\Theta_c \end{bmatrix} \begin{bmatrix} U_{B1}^T \\ U_{B2}^T \end{bmatrix}, \tag{40}$$

for any choice of Θ_r and Θ_c.

References

[1] H. Aanaes, R. Fisker, K. Åström, and J. Carstensen. Robust factorization. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 24(9):1215–1225, 2002. 1

[2] K. Aftab and R. I. Hartley. Convergence of iteratively reweighted least squares to robust m-estimators. In *IEEE Winter Conference on Applications of Computer Vision*, Waikoloa, Hawaii, USA, 2015. 3

[3] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends® in Machine Learning*, 3(1):1–122, 2011. 4

[4] C. Bregler, A. Hertzmann, and H. Biermann. Recovering non-rigid 3D shape from image streams. In *Conf. Computer Vision and Pattern Recognition*, Hilton Head, USA, 2000. 6

[5] R. Cabral, F. D. la Torre, J. P. Costeira, and A. Bernardino. Unifying nuclear norm and bilinear factorization approaches for low-rank matrix decomposition. In *Int. Conf. Computer Vision*, Sydney, Australia, 2013. 1

[6] E. J. Candès, X. Li, Y. Ma, and J. Wright. Robust principal component analysis? *Journal of ACM*, 58(1):1–37, 2009. 1

[7] E. J. Candès and B. Recht. Exact matrix completion via convex optimization. *CoRR*, abs/0805.4471, 2008. 1

[8] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham. Active shape models-their training and application. *Computer vision and image understanding*, 61(1):38–59, 1995. 4

[9] Y. Dai, H. Li, and M. He. A simple prior-free method for non-rigid structure-from-motion factorization. *Int. Journal Computer Vision*, 107(2):101–122, 2014. 7, 8

[10] I. Daubechies, R. Devore, M. Fornasier, and C. S. Güntürk. Iteratively reweighted least squares minimization for sparse recovery. *Comm. Pure Appl. Math*, 63(1):1–38, 2010. 2

[11] F. De La Torre and M. J. Black. A framework for robust subspace learning. *International Journal of Computer Vision*, 54(1-3):117–142, 2003. 1

[12] A. Eriksson and A. Hengel. Efficient computation of robust weighted low-rank matrix approximations using the L_1 norm. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 34(9):1681–1690, 2012. 1

[13] R. Fletcher and C. M. Reeves. Function minimization by conjugate gradients. *The computer journal*, 7(2):149–154, 1964. 3

[14] S. Friedland and A. Torokhti. Generalized rank-constrained matrix approximations. *SIAM Journal on Matrix Analysis and Applications*, 29(2):656–659, 2007. 8

[15] K. Glashoff and M. M. Bronstein. Structure from motion using augmented lagrangian robust factorization. In *3D Imaging, Modeling, Processing, Visualization and Transmission (3DIMPVT), 2012 Second International Conference on*, pages 379–386. IEEE, 2012. 5

[16] R. I. Hartley. Tutorial notes on IRLS. In *European Conf. Computer Vision*, Zurich, Switzerland, 2014. 2, 3

[17] Q. Ke and T. Kanade. Robust L_1 norm factorization in the presence of outliers and missing data by alternative convex programming. In *Conf. Computer Vision and Pattern Recognition*, San Diego, USA, 2005. 1

[18] C. Lu, Z. Lin, and S. Yan. Smoothed low rank and sparse matrix recovery by iteratively reweighted least squares minimization. *Preprint arXiv:1401.7413*, 2014. 1

[19] K. Mohan and M. Fazel. Iterative reweighted algorithms for matrix rank minimization. *Journal of Machine Learning Research*, 13(1):3441–3473, 2012. 1

[20] E. Polak and G. Ribiere. Note sur la convergence de méthodes de directions conjuguées. *ESAIM: Mathematical Modelling and Numerical Analysis-Modélisation Mathématique et Analyse Numérique*, 3(R1):35–43, 1969. 3

[21] W. Siming and L. Zhouchen. Analysis and improvement of low rank representation for subspace segmentation. *arXiv preprint arXiv:1107.1561*, 2011. 1, 4

[22] D. Sondermann. Best approximate solutions to matrix equations under rank restrictions. *Statistical Papers*, 27(1):57–66, 1986. 8

[23] M. B. Stegmann and D. D. Gomez. A brief introduction to statistical shape analysis, 2002. 4, 5

[24] D. Strelow. General and nested Wiberg minimization. In *Conf. Computer Vision and Pattern Recognition*, Providence, RI, USA, 2012. 1

[25] R. Tron and R. Vidal. A benchmark for the comparison of 3-D motion segmentation algorithms. In *Conf. Computer Vision and Pattern Recognition*, Minneapolis, USA, 2007. 7

[26] S. Xiang, Y. Zhu, X. Shen, and J. Ye. Optimal exact least squares rank minimization. In *ACM SIGKDD conf. on Knowledge discovery and data mining*, Beijing, China, 2012. 8

[27] H. Xu, C. Caramanis, and S. Sanghavi. Robust PCA via outlier pursuit. In *Advances in Neural Information Processing Systems*, Vancouver, Canada, 2010. 1, 4

[28] Y. Zheng, G. Liu, S. Sugimoto, S. Yan, and M. Okutomi. Practical low-rank matrix approximation under robust L_1-norm. In *Conf. Computer Vision and Pattern Recognition*, Providence, RI, USA, 2012. 1