

# A Convex Approach to Low Rank Matrix Approximation with Missing Data

Carl Olsson and Magnus Oskarsson

Centre for Mathematical Sciences  
Lund University, Lund, Sweden  
`{calle,magnuso}@maths.lth.se`

**Abstract.** Many computer vision problems can be formulated as low rank bilinear minimization problems. One reason for the success of these problems is that they can be efficiently solved using singular value decomposition. However this approach fails if the measurement matrix contains missing data.

In this paper we propose a new method for estimating missing data. Our approach is similar to that of  $L_1$  approximation schemes that have been successfully used for recovering sparse solutions of under-determined linear systems. We use the nuclear norm to formulate a convex approximation of the missing data problem. The method has been tested on real and synthetic images with promising results.

## 1 Bilinear Models and Factorization

Bilinear models have been applied successfully to several computer vision problems such as structure from motion [1–3], nonrigid 3D reconstruction [4, 5], articulated motion [6], photometric stereo [7] and many other. In the typical application, the observations of the system are collected in a measurement matrix which (ideally) is known to be of low rank due to the bilinearity of the model. The successful application of these models is mostly due to the fact that if the entire measurement matrix is known, singular value decomposition (SVD) can be used to find a low rank factorization of the matrix.

In practice, it is rarely the case that all the measurements are known. Problems with occlusion and tracking failure lead to missing data. In this case SVD can not be employed, which motivates the search for methods that can handle incomplete data.

To our knowledge there is, as of yet, no method that can solve this problem optimally. One approach is to use iterative local methods. A typical example is to use a two step procedure. Here the parameters of the model are divided into two groups where each one is chosen such that the model is linear when the other group is fixed. The optimization can then be performed by alternating the optimization over the two groups [8]. Other local approaches such as non-linear Newton methods have also been applied [9]. There are however no guarantee of convergence and therefore these methods are in need of good initialization. This

is typically done with a batch algorithm (e.g. [1]) which usually optimizes some algebraic criterion.

In this paper we propose a different approach. Since the original problem is difficult to solve due to its non convexity we derive a simple convex approximation. Our solution is independent of initialization, however batch algorithms can still be used to strengthen the approximation. Further more, since our program is convex it is easy to extend it to other error measures or to include prior information.

## 2 Low Rank Approximations and the Nuclear Norm

In this section we will present the nuclear norm. It has previously been used in applications such as image compression, system identification and similar problems that can be stated as low rank approximation problems (see [10–12]). The theory largely parallels that of  $L_1$  approximation (see [13–15]) which has been used successfully in various applications.

Let  $M$  the matrix with entries  $m_{ij}$  containing the measurements. The typical problem of finding a low rank matrix  $X$  that describes the data well can be posed as

$$\min_X \|X - M\|_F^2 \quad (1)$$

$$\text{s.t rank}(X) \leq r, \quad (2)$$

where  $\|\cdot\|_F$  denotes the Frobenius norm, and  $r$  is the given rank. This problem can be solved optimally with SVD even though the rank constraint is highly non-convex (see [16]). The SVD-approach does however not extend to the case when the measurement matrix is incomplete. Let  $W$  be a matrix with entries  $w_{ij} = 1$  if the value of  $m_{ij}$  has been observed and zeros otherwise. Note that the values of  $W$  can also be chosen to represent weights modeling the confidence of the measurements. The new problem can be formulated as

$$\min_X \|W \odot (X - M)\|_F^2 \quad (3)$$

$$\text{s.t rank}(X) \leq r \quad (4)$$

where  $\odot$  denotes element-wise multiplication. In this case SVD can not be directly applied since the whole matrix  $M$  is not known. Various approaches for estimating the missing data exist and the most simple one (which is commonly used for initializing different iterative methods) is simply to let the missing entries be zeros. In terms of optimization this corresponds to finding the minimum Frobenius norm solution  $X$  such that  $W \odot (X - M) = 0$ . In effect what we are minimizing is

$$\|X\|_F^2 = \sum_{i=1}^m \sigma_i(X)^2, \quad (5)$$

where  $\sigma_i(X)$  is the  $i$ 'th largest singular value of the  $m \times n$  matrix  $X$ . It is easy to see that this function penalizes larger values proportionally more than

small values (see figure 1). Hence, this function favors solutions with many small singular values as opposed to a small number of large singular values, which is exactly the opposite of what we want.

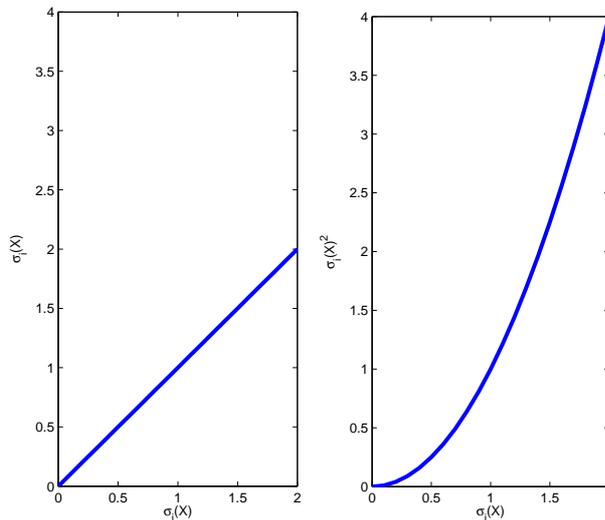


Fig. 1: Comparison between the Frobenius norm and the nuclear norm, showing on the left:  $\sigma_i(X)$  and on the right:  $\sigma_i(X)^2$ .

Since we cannot minimize the rank function directly, because of its non-convexity, we will use the so called nuclear norm which is given by

$$\|X\|_* = \sum_{i=1}^m \sigma_i(X). \quad (6)$$

The nuclear norm can also be seen as the dual norm of the operator norm  $\|\cdot\|_2$ , that is

$$\|X\|_* = \max_{\|Y\|_2 \leq 1} \langle X, Y \rangle \quad (7)$$

where the inner product is defined by  $\langle X, Y \rangle = \text{tr}(X^T Y)$ , see [10]. By the above characterization it is easy to see that  $\|X\|_*$  is convex, since a maximum of functions linear in  $X$  is always convex (see [17]).

The connection between the rank function and the nuclear norm can be seen via the following inequality (see [16]), which holds for any matrix of at most rank  $r$

$$\|X\|_* \leq \sqrt{r} \|X\|_F. \quad (8)$$

In fact it turns out that the nuclear norm is the convex envelope of the rank function on the set  $\{X; \|X\|_F \leq 1\}$  (see [17]). In view of (8) we can try to solve

the following program

$$\min_X \|W \odot (X - M)\|_F^2 \quad (9)$$

$$\text{s.t } \|X\|_*^2 - r\|X\|_F^2 \leq 0. \quad (10)$$

The Lagrangian of this problem is

$$L(X, \mu) = \mu(\|X\|_*^2 - r\|X\|_F^2) + \|W \odot (X - M)\|_F^2, \quad (11)$$

with the dual problem

$$\max_{\mu > 0} \min_X L(X, \mu). \quad (12)$$

The inner minimization is however not convex if  $\mu$  is not zero. Therefore we are forced to approximate this program by dropping the non convex term  $-r\|X\|_F^2$ , yielding the program

$$\min_X \mu\|X\|_*^2 + \|W \odot (X - M)\|_F^2. \quad (13)$$

which is familiar from the  $L_1$ -approximation setting (see [13–15]). Note that it does not make any difference whether we penalize with the term  $\|X\|_*$  or  $\|X\|_*^2$ , it just results in a different  $\mu$ .

The problem with dropping the non convex part is that (13) is no longer a lower bound on the original problem. Hence (13) does not tell us anything about the global optimum, it can only be used as a heuristic for generating good solutions. An interesting exception is when the entire measurement matrix is known. In this case we can write the Lagrangian as

$$L(x, \mu) = \mu\|X\|_*^2 + (1 - \mu r)\|X\|_F^2 + 2\langle X, M \rangle + \|M\|_F^2. \quad (14)$$

Thus, here  $L$  will be convex if  $0 \leq \mu \leq 1/r$ . Note that if  $\mu = 1/r$  then the term  $\|X\|_F^2$  is completely removed. In fact this offers some insight as to why the problem can be solved exactly when  $M$  is completely known, but we will not pursue this further.

## 2.1 Implementation

In our experiments we use (13) to fill in the missing data of the measurement matrix. If the resulting matrix is not of sufficiently low rank then we use SVD to approximate it. In this way it is possible to use methods such as [5] that work when the entire measurement matrix is known. The program (13) can be implemented in various ways (see [10]). The easiest way (which we use) is to reformulate it as a semidefinite program, and use any standard optimization software to solve it. The semidefinite formulation can be obtained from the dual norm (see equation (7)). Suppose the matrix  $X$  (and  $Y$ ) has size  $m \times n$ , and let  $I_m, I_n$  denote the identity matrices of size  $m \times m$  and  $n \times n$  respectively. That the matrix  $Y$  has operator-norm  $\|Y\|_2 \leq 1$  means that all the eigenvalues of

$Y^T Y$  are smaller than 1, or equivalently that  $I_m - Y^T Y \succeq 0$ . Using the Schur complement [17] and (7) it is now easy to see that minimizing the nuclear norm can be formulated as

$$\min_X \max_Y \operatorname{tr}(Y^T X) \quad (15)$$

$$\begin{bmatrix} I_m & Y \\ Y^T & I_n \end{bmatrix} \succeq 0 \quad (16)$$

Taking the dual of this program, we arrive at the linear semidefinite program

$$\min_{X, Z_{11}, Z_{22}} \operatorname{tr}(Z_{11} + Z_{22}), \quad (17)$$

$$\begin{bmatrix} Z_{11} & \frac{X}{2} \\ \frac{X^T}{2} & Z_{22} \end{bmatrix} \succeq 0. \quad (18)$$

Linear semidefinite programs have been extensively studied in the optimization literature and there are various softwares for solving them. In our experiments we use SeDuMi [18] (which is freely available) but any solver that can handle the semidefinite program and the Frobenius-norm term in (13) will work.

### 3 Experiments

Next we present two simple experiments for evaluating the performance of the approximation. In both experiments we select the observation matrix  $W$  randomly. Not a realistic scenario for most real applications, however we do this since we want to evaluate the performance for different levels of missing data with respect to ground truth. It is possible to strengthen the relaxation by using

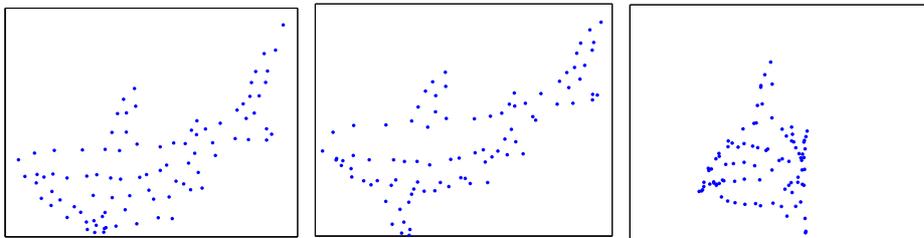


Fig. 2: Three images from the shark sequence.

batch algorithms. However, since we are only interested in the performance of (13) itself we do not do this.

In the first experiment points on a shark are tracked in a sequence of images. The same sequence has been used before, see e.g. [19]. The shark undergoes

a deformation as it moves. In this case the deformation can be described by two shape modes  $S_0$  and  $S_1$ . Figure 2 shows three images from the sequence (with no missing data). To generate the measurement matrix we added noise and randomly selected  $W$  for different levels of missing data. Figure 3 shows the error compared to ground truth when using a one ( $S_0$ ) and a two element basis ( $S_0, S_1$ ) respectively. On the x-axis is the level of missing data and on the y-axis  $\|X - M\|_F / \|M\|_F$  is shown. For lower levels of missing data the two element basis explains most of  $M$ . Here  $M$  is the complete measurement matrix with noise. Note that the remaining error corresponds to the added noise. For missing data levels below 50% the approximation recovers almost exactly the correct matrix (without noise). When the missing data level approaches 70%, the approximation starts to break down. Figure 4 shows the obtained reconstruction when the

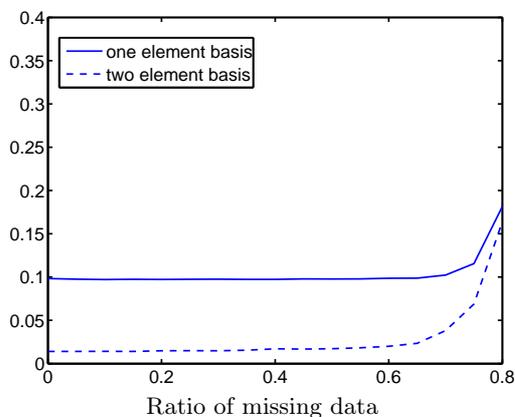


Fig. 3: Reconstruction error for the Shark experiment, for a one and two element basis, as a function of the level of missing data. On the x-axis is the level of missing data and on the y-axis is  $\|X - M\|_F / \|M\|_F$ .

missing data is 40%. Note that we are not claiming to improve the quality of the reconstructions; We are only interested in recovering  $M$ . The reconstructions are just included to illustrate the results. To the upper left is the first shape mode  $S_0$ , and the others are images generated by varying the coefficient corresponding to the second mode  $S_1$  (see [4]). Figure 5 shows the setup for the second experiment. In this case we used real data where all the interest points were tracked through the entire sequence. Hence the full measurement matrix  $M$  with noise is known. As in the previous experiment, we randomly selected the missing data.

Figure 6 shows the error compared to ground truth (i.e.  $\|X - M\|_F / \|M\|_F$ ) when using a basis with one or two elements. In this case the rank of the motion is not known, however the two element basis seems to be sufficient. In this case the approximation starts to break down sooner than for the shark experiment.

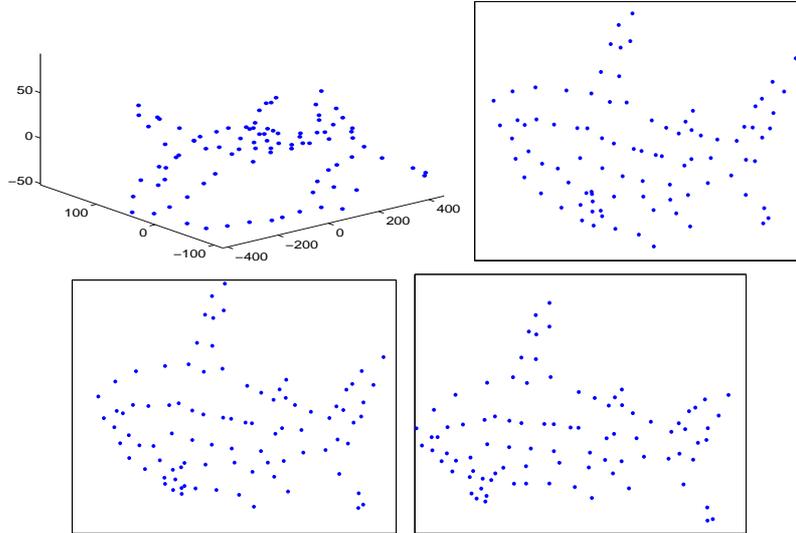


Fig. 4: A 3D-reconstruction of the shark. The first shape mode in 3D and three generated images. The camera is the same for the three images but the coefficient of the second structure mode is varied.

We believe that this is caused by the fact that the number of points and views in this experiment is less than for the shark experiment, making it more sensitive to missing data. Still the approximation manages to recover the matrix  $M$  well, for noise levels up to 50% without any knowledge other than the low rank assumption.

## 4 Conclusions

In this paper we have presented a heuristic for finding low rank approximations of incomplete measurement matrices. The method is similar to the concept of  $L_1$ -approximation that has been used with success in for example compressed sensing. Since it is based on convex optimization and in particular semidefinite programming, it is possible to add more knowledge in the form of convex constraints to improve the resulting estimation. Experiments indicate that we are able to handle missing data levels of around 50% without resorting to any type of batch algorithm.

In this paper we have merely studied the relaxation itself and it is still an open question how much it is possible to improve the results by combining our method with batch methods.

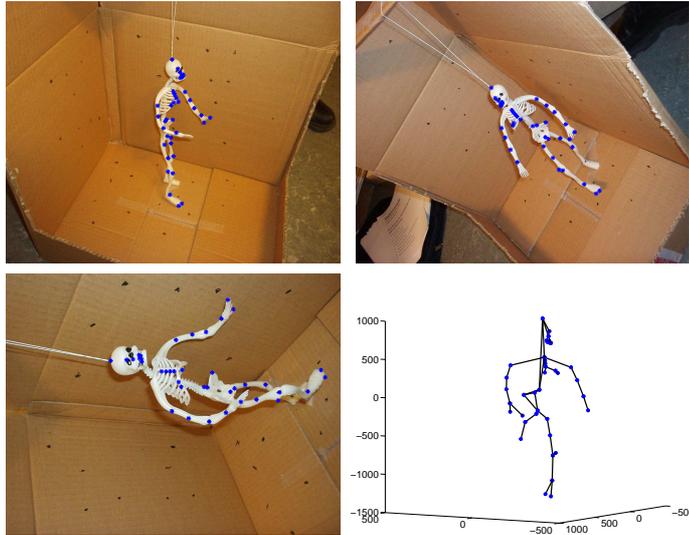


Fig. 5: Three images from the skeleton sequence, with tracked image points, and the 1st mode of reconstructed nonrigid-structure.

## Acknowledgments

This work has been funded by the European Research Council (GlobalVision grant no. 209480), the Swedish Research Council (grant no. 2007-6476) and the Swedish Foundation for Strategic Research (SSF) through the programme Future Research Leaders.

## References

1. Tardif, J., Bartoli, A., Trudeau, M., Guilbert, N., Roy, S.: Algorithms for batch matrix factorization with application to structure-from-motion. In: Int. Conf. on Computer Vision and Pattern Recognition, Minneapolis, USA (2007)
2. Sturm, P., Triggs, B.: A factorization based algorithm for multi-image projective structure and motion. In: European Conference on Computer Vision, Cambridge, UK (1996)
3. Tomasi, C., Kanade, T.: Shape and motion from image streams under orthography: a factorization method. *Int. Journal of Computer Vision* **9**(2) (1992)
4. Bregler, C., Hertzmann, A., Biermann, H.: Recovering non-rigid 3d shape from image streams. In: Int. Conf. on Computer Vision and Pattern Recognition, Hilton Head, SC, USA (2000)
5. Xiao, J., Kanade, T.: A closed form solution to non-rigid shape and motion recovery. *International Journal of Computer Vision* **67**(2) (2006) 233–246
6. Yan, J., Pollefeys, M.: A factorization approach to articulated motion recovery. In: IEEE Conf. on Computer Vision and Pattern Recognition, San Diego, USA (2005)

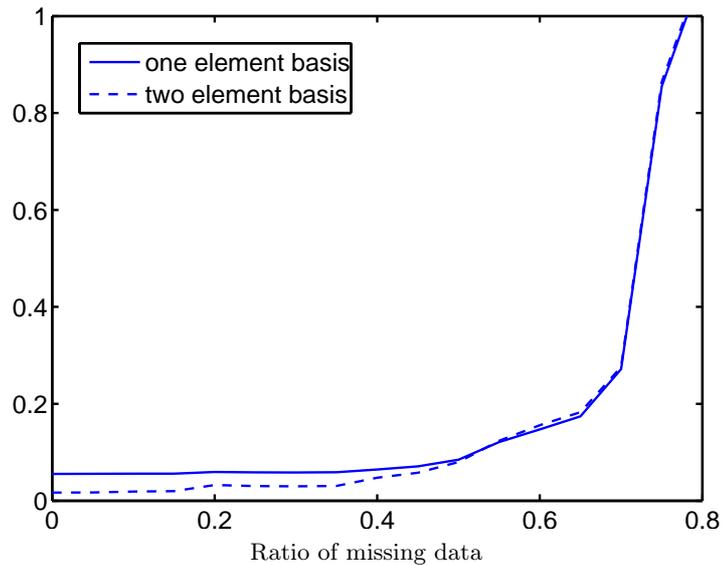


Fig. 6: Reconstruction error for the Skeleton experiment, for a one and two element basis, as a function of the level of missing data. On the y-axis  $\|X - M\|_F / \|M\|_F$  is shown.

7. Basri, R., Jacobs, D., Kemelmacher, I.: Photometric stereo with general, unknown lighting. *Int. Journal of Computer Vision* **72**(3) (2007) 239–257
8. Hartley, R., Schaffalitzky, F.: Powerfactorization: An approach to affine reconstruction with missing and uncertain data. In: *Australia-Japan Advanced Workshop on Computer Vision*, Adelaide, Australia. (2003)
9. Buchanan, A., Fitzgibbon, A.: Damped newton algorithms for matrix factorization with missing data. *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on* **2** (20-25 June 2005) 316–322 vol. 2
10. Recht, B., Fazel, M., Parrilo, P.: Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization. <http://arxiv.org/abs/0706.4138v1> (2007)
11. Fazel, M., Hindi, H., Boyd, S.: A rank minimization heuristic with application to minimum order system identification. *Proceedings of the American Control Conference* (2003)
12. El Ghaoui, L., Gahinet, P.: Rank minimization under lmi constraints: A framework for output feedback problems. *Proceedings of the European Control Conference* (1993)
13. Tropp, J.: Just relax: convex programming methods for identifying sparse signals in noise. *Information Theory, IEEE Transactions on* **52**(3) (March 2006) 1030–1051
14. Donoho, D., Elad, M., Temlyakov, V.: Stable recovery of sparse overcomplete representations in the presence of noise. *Information Theory, IEEE Transactions on* **52**(1) (Jan. 2006) 6–18

15. Candes, E., Romberg, J., Tao, T.: Stable signal recovery from incomplete and inaccurate measurements. *Communications of Pure and Applied Mathematics* **59** (2005) 1207–1223
16. Golub, G., van Loan, C.: *Matrix Computations*. The Johns Hopkins University Press (1996)
17. Boyd, S., Vandenberghe, L.: *Convex Optimization*. Cambridge University Press (2004)
18. Sturm, J.F.: Using sedumi 1.02, a matlab toolbox for optimization over symmetric cones. (1998)
19. Torresani, L., Hertzmann, A., Bregler, C.: Non-rigid structure-from-motion: Estimating shape and motion with hierarchical priors. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **30**(5) (2008)
20. Raiko, T., Ilin, A., Karhunen, J.: Principal component analysis for sparse high-dimensional data. In: 14th International Conference on Neural Information Processing, Kitakyushu, Japan (2007) 566–575