

SUBSPACE PROJECTION MATRIX COMPLETION ON GRASSMANN MANIFOLD

Xinyue Shen Yuantao Gu*

State Key Laboratory on Microwave and Digital Communications
Tsinghua National Laboratory for Information Science and Technology
Department of Electronic Engineering, Tsinghua University, Beijing 100084, CHINA

ABSTRACT

In this paper, we work on the problem of subspace estimation from random downsamplings of its projection matrix. An optimization problem on the Grassmann manifold is formulated for projection matrix completion, and an iterative gradient descend line-search algorithm on the Grassmann manifold (GGDLS) is proposed to solve such optimization problem. The convergence of the proposed algorithm has theoretical guarantee, and numerical experiments verify that the required sampling number for successful recovery with probability 1 is $2s(N - s)$ in the noiseless cases. Compared with some reference algorithms, in the noiseless scenario, the proposed algorithm is very time efficient, and the required sampling number is rather small for successful recovery. In the noisy scenario, the proposed GGDLS is remarkably robust against the noise both under high measurement SNR and low measurement SNR.

Index Terms— Matrix completion, Subspace projection matrix, Subspace estimation, Optimization on Grassmann manifold

1. INTRODUCTION

For a linear subspace in an ambient Euclidean space, its projection matrix is the matrix by which a vector in the ambient space has to multiply when projected onto such subspace. Since subspace projection matrix has a one to one correspondence with subspace, given that some entries of a projection matrix are randomly sampled, we are able to estimate the subspace by the completion of its projection matrix. Subspace estimation has been a concerning problem in signal processing and computer vision. In the problems of face recognition [1], motion segmentation [2], and visual tracking [3], it is found that the objects are in subspaces with much lower dimension than the ambient space.

While there have existed many solving strategies for matrix completion [4, 5], most of them are for general low rank matrices. The number of random samples has to be larger than $s(2N - s)$, in which $N \times N$ is the size of the matrix and s is its rank. The subspace projection matrix has rather specific structure. It is not only semi-definite, but also only has eigenvalues 1 and 0. Thus, if the structure of the projection matrix is fully explored, its completion is allowed to have less complexity.

Considering the fact that a subspace projection matrix is equivalent to its corresponding subspace, one may find it natural to connect subspace projection matrix to the Grassmann manifold $\text{Gr}(\mathbb{R}^N, s)$ denoted as $\text{Gr}_{N,s}$, which is the set of all s -dimensional subspaces

in \mathbb{R}^N . In fact, there are methods solving matrix problems on an abstract manifolds [6–9]. We will review some of them in section 4.

In this paper, we work in the circumstance in which only partial entries of the projection matrix of a subspace \mathcal{X}^* can be randomly sampled. The aim is to reconstruct the projection matrix and to estimate \mathcal{X}^* accordingly. Since a subspace with fixed dimension is on a Grassmann manifold, the reconstruction of the projection matrix is indeed the problem of finding a point on the manifold $\text{Gr}_{N,s}$ such that the sampled entries are satisfied.

The main contribution of this work is that we propose a method of projection matrix completion (GGDLS) based on the line-search algorithm on a manifold, and achieve subspace estimation accordingly. In the noiseless circumstance, the number of samples can be reduced to around $2s(N - s)$ in the numerical experiments, while the algorithm is still very efficient with low computational complexity. When the measurement noise is considered, the proposed GGDLS is remarkably robust both under high measurement SNR and low measurement SNR.

2. PRELIMINARY

Preceding the description of our proposed method of projection matrix completion, we first have a brief review on the Grassmann manifold and the line-search algorithm on a manifold.

2.1. Grassmann Manifold

The Grassmann manifold $\text{Gr}_{N,s}$ has a formal definition as a quotient manifold

$$\text{Gr}_{N,s} := \frac{\mathcal{O}(N)}{\mathcal{O}(s) \times \mathcal{O}(N - s)}, \quad (1)$$

where $\mathcal{O}(N)$ is the $N \times N$ orthogonal matrix group. It is a compact differentiable manifold with dimension $d = s(N - s)$. According to the definition in (1), a point $\mathcal{X} \in \text{Gr}_{N,s}$ can be represented as a matrix X such that the columns of X span the subspace \mathcal{X} and $X^T X = I_s$. In this paper, we refer to such X as its corresponding subspace $\mathcal{X} \in \text{Gr}_{N,s}$.

The tangent space of $\text{Gr}_{N,s}$ at a point X is

$$T_X \text{Gr}_{N,s} = \{\xi \in \mathbb{R}^{N \times s} : X^T \xi = 0\}.$$

For any tangent vectors $\eta, \xi \in T_X \text{Gr}_{N,s}$, the metric on $\text{Gr}_{N,s}$ induced by the Euclidean metric in \mathbb{R}^N is

$$\langle \eta, \xi \rangle_X := \text{trace} \left((X^T X)^{-1} \eta \xi^T \right), \quad (2)$$

*This work was partially supported by the National Program on Key Basic Research Project (973 Program 2013CB329201) and the National Natural Science Foundation of China (NSFC 61371137). The corresponding author is Yuantao Gu (gyt@tsinghua.edu.cn).

and the geodesic distance induced by such metric is

$$d_g(X, Y) := \sum_{i=1}^s \theta_i^2,$$

where θ_i are the principal angles. There are several other distances on $\text{Gr}_{N,s}$, among which the projection distance is defined by the projection matrix

$$d_p(X, Y) := \frac{1}{2} \|XX' - YY'\|_F^2 = \sum_{i=1}^s \sin^2 \theta_i.$$

From its definition, we know that the projection metric tends to be less sensitive than the geodesic metric as θ_i becomes large, but approximates the geodesic metric well when all of the principal angles are small. In this work, we use d_p to evaluate the distance between the target subspace and the estimated subspace.

2.2. General framework of Line-search Algorithm on Manifold

The line-search algorithm on Riemannian manifold \mathcal{M} is the generalization of the iterative line-search method in an Euclidean space. In the k^{th} iteration, it consists of selecting a tangent vector η_{k+1} to \mathcal{M} at X_k , which is the result of the previous iteration, and performing a search along a curve $\gamma(t)$ in \mathcal{M} such that $\gamma'(0) = \eta_{k+1}$.

In order to find the descend direction of f on the tangent space at a point X , we need the notion of the gradient of f with respect to \mathcal{M} .

Definition 1. [10] *Given a smooth scalar field f on a Riemannian manifold \mathcal{M} , the gradient of f at a point $X \in \mathcal{M}$ denoted as $\text{grad}f(X)$ is defined as the unique element of $T_X\mathcal{M}$ such that*

$$\langle \text{grad}f(X), \xi \rangle_X = \nabla f(X)[\xi], \quad \forall \xi \in T_X\mathcal{M}.$$

Any direction $\xi \in T_X\mathcal{M}$ such that

$$\langle \xi, \text{grad}f(X) \rangle_X < 0$$

is a gradient related descend direction. Having obtained a descend direction on the tangent space, we need to find the corresponding curve on the manifold along which the solution is searched. Thus, the retraction map is defined to retract the descend direction from the tangent space back onto the manifold.

Definition 2. [10] *A retraction on a manifold \mathcal{M} is a smooth mapping R from the tangent bundle $T\mathcal{M}$ onto \mathcal{M} with the following properties.*

1. $R_X(0) = X$,
2. $\nabla R_X(0) = \text{id}_{T_X\mathcal{M}}$.

From the above definition, if a line-search algorithm on a compact manifold satisfies several conditions on the choice of descend direction and descend step size, the following theorem guarantees that such algorithm is able to converge.

Theorem 1. [10] *Let $\{X_k\}$ be an infinite sequence of iterates generated by a line-search method on a compact Riemannian manifold. If in each iteration, the descend direction η_k on the tangent space is chosen as a gradient related direction, and the step size t_k is chosen according to the Armijo method such that*

$$f(X_k) - f(X_{k+1}) \geq c(f(X_k) - f(R_{X_k}(t_k\eta_k))), \quad (3)$$

in which $c \in (0, 1)$, then

$$\lim_{k \rightarrow \infty} \|\text{grad}f(X_k)\| = 0.$$

Provided that a well-defined retraction mapping and a sequence of gradient related directions can be numerically computed, such general line-search framework on compact non-linear manifold is an efficient solver of minimization problem.

3. SUBSPACE PROJECTION MATRIX COMPLETION

3.1. Problem Formulation

Supposed that \mathcal{X}^* is an s dimensional subspace in \mathbb{R}^N , and $s \ll N$. $\mathbf{x}_1, \dots, \mathbf{x}_s$ are s points such that

$$\mathcal{X}^* = \text{span}\{\mathbf{x}_1, \dots, \mathbf{x}_s\},$$

and that $X^* = [\mathbf{x}_1, \dots, \mathbf{x}_s]$ satisfies

$$(X^*)^T X^* = I_s,$$

so the projection matrix of \mathcal{X}^* is

$$P_{\mathcal{X}^*} = X^*(X^*)^T \in \mathbb{R}^{N \times N}.$$

The measurement is $Y = \mathcal{A}(P_{\mathcal{X}^*})$, in which

$$\mathcal{A} : \mathbb{R}^{N \times N} \rightarrow \mathbb{R}^M$$

is a mapping defined as

$$\mathcal{A}(P_{\mathcal{X}})_i = [P_{\mathcal{X}}]_{\Omega_i}, \quad i = 1, 2, \dots, M,$$

where $\Omega \subset \{1, 2, \dots, N\} \times \{1, 2, \dots, N\}$ is a subset of the matrix indices which are uniformly randomly chosen, and $|\Omega| = M$. Thus, Y selects M entries of $P_{\mathcal{X}^*}$ at random with $M \ll N^2$.

The aim is to reconstruct the projection matrix $P_{\mathcal{X}^*}$ from Y , so it is a problem of projection matrix completion. Since a subspace has a one to one correspondence with its projection matrix, the subspace \mathcal{X}^* is also estimated by the completion of its projection matrix $P_{\mathcal{X}^*}$.

According to the fact that

$$P_{\mathcal{X}} = XX^T, \quad X \in \text{Gr}_{N,s},$$

we propose the following optimization problem to achieve its completion

$$\min_{X \in \text{Gr}_{N,s}} f(X) = \frac{1}{2} \|\mathcal{A}(XX^T) - Y\|_F^2. \quad (4)$$

Note that f is a well-defined cost function on the Grassmann manifold $\text{Gr}_{N,s}$, because

$$f(X) = f(XQ), \quad \forall Q \in \mathbb{R}^{s \times s} \quad QQ^T = I_s,$$

which means that f is invariant with different choice of the orthonormal basis of the subspace.

3.2. The Proposed Solution

In order to complete a subspace projection matrix, we need to solve an optimization problem on the Grassmann manifold $\text{Gr}_{N,s}$. As shown in problem (4), the cost function is differentiable, so we propose an iterative line-search method on $\text{Gr}_{N,s}$ called Grassmann manifold gradient descend line-search algorithm (GGDLS).

The framework of our method is in Algorithm 1. In the first step,

Table 1. Algorithm 1: Grassmann Manifold Gradient Descend Line Search Algorithm (GGDLS)

Require: Grassmann manifold $\text{Gr}_{N,s}$;
continuously differentiable cost function f ;
retraction R from $T\text{Gr}_{N,s}$ to $\text{Gr}_{N,s}$;
scalars for the Armijo step size $\alpha > 0, \beta, \sigma \in (0, 1)$;

Input: Initial iterate $X_0 \in \text{Gr}_{N,s}$;

Output: Sequence of iterates $\{X_k\}$.

For $k = 0, 1, 2, \dots$ **do:**

1. Compute the Euclidean gradient ∇f_{X_k} at X_k ;
2. Project ∇f_{X_k} onto the tangent space $T_{X_k}\text{Gr}_{N,s}$ to obtain η_k ;
3. $X_{k+1} = R_{X_k}(t_k \eta_k)$,
where t_k is the Armijo step size;

Until: Stopping criterion satisfied;

the Euclidean gradient ∇f_{X_k} is an $N \times s$ matrix whose entries are computed as

$$\begin{aligned} [\nabla f_X]_{p,q} &= \sum_{j:(p,j) \in \Omega} [X X^T - Y]_{p,j} X_{j,q} \\ &\quad + \sum_{i:(i,p) \in \Omega} [X X^T - Y]_{i,p} X_{i,q}. \end{aligned}$$

In the second step, η_k is the projection of ∇f_{X_k} onto the tangent space of $\text{Gr}_{N,s}$ at X_k

$$\eta_k = (I_N - X_k X_k^T) \nabla f_{X_k}.$$

Now we verify that η_k is the gradient of f on $\text{Gr}_{N,s}$ as defined in Definition 1. For any $\xi \in T_{X_k}\text{Gr}_{N,s}$,

$$\begin{aligned} \langle \eta_k, \xi \rangle_{X_k} &= \text{trace}(\nabla f_{X_k}^T (I_N - X_k X_k^T) \xi) \\ &= \text{trace}(\nabla f_{X_k}^T (\xi - X_k X_k^T \xi)) \\ &= \text{trace}(\nabla f_{X_k}^T \xi) \\ &= \nabla f_{X_k}[\xi], \end{aligned}$$

in which the third equation stands because $X_k^T \xi = 0$ for all $\xi \in T_{X_k}\text{Gr}_{N,s}$.

In the third step of Algorithm 1, the retraction of $t_k \eta_k$ at X_k is the q factor of the QR decomposition of $X_k - t_k \eta_k$. The Armijo step size is $t_k = \alpha \beta^m$ with m being the smallest nonnegative integer such that

$$f(X_k) - f(R_{X_k}(\beta^m \alpha \eta)) \geq -\sigma \langle \eta_k, \beta^m \alpha \eta_k \rangle_{X_k}.$$

Algorithm 1 belongs to the general framework of line-search method on manifold. The merit of such kind of methods is that it avoids a cost function comparatively difficult to solve, such as a non-convex cost function of the matrix, by moving these penalties from the cost function to the constraint set. Although the constraint set is not convex anymore, it becomes a smooth manifold which allows for optimization methods with easy computation and convergence guarantee.

Since $\{\eta_k\}$ is exactly a sequence of gradients of f with respect to $\text{Gr}_{N,s}$, and X_{k+1} satisfies (3), the convergence of Algorithm 1 is guaranteed by Theorem 1 so that the norm of the gradient with respect to $\text{Gr}_{N,s}$ is convergent to 0 given any initial value.

The advantage of our method is that we fully explore the structure of a subspace projection matrix so that the computational complexity is fundamentally reduced. In fact, by the constraint on $\text{Gr}_{N,s}$, we reduce the magnitude of the size of the problem from $N \times N$ to $N \times s$, and the computation in each iteration is easy to implement numerically.

4. PREVIOUS WORKS

There have existed methods solving matrix optimization problems on manifolds. We list a few of them in the following to see the wide use of matrix manifolds as well as the connection between these methods and the proposed method in this work.

In the work [11], a conjugate gradient descend method on Grassmann manifold is proposed for robust subspace estimation in conjunction with the generalized projection based M-Estimator (gpbM). Their problem is to estimate the projection matrix and the intercept vector of an affine subspace from sample points contaminated by noise. They have shown that by optimizing the parameter matrix on the Grassmann manifold, the performance of the gpbM algorithm improves significantly. Different from their view of denoising, the proposed method allows the use of random downsampling so that much fewer samples are needed.

The work [12] solves the problem of online identification and tracking of subspaces from incomplete information by analyzing incremental gradient descent on the Grassmann manifold. The number of samples can be reduced to around $10s(N-s)$ in their simulations. Their algorithm is scalable to very high-dimensional data, and is remarkably computationally efficient. However, whether the algorithm converges to the basin around the global minimum is very sensitive to the choice of the step size and the starting point. As a contrast, our method does not suffer from this problem. The numerical simulation indicates that with a random starting point the proposed method is able to converge to a successful recovery given enough samples.

The problem of low-rank matrix completion is also studied in the work [13]. They formulate the problem as a minimization of a differentiable cost function on the Riemannian manifold of matrices with fixed rank. The experiments indicate that their approach scales well for large-scale problem, and the number of samples can be reduced to around $4s(N-s)$. While their constraint is on the manifold of fixed rank matrices, the proposed method is with constraint on the Grassmann manifold so that it is more appropriate for the completion of a projection matrix.

5. NUMERICAL EXPERIMENTS

In this section, we implement the proposed algorithm (GGDLS) for subspace projection matrix completion. The proposed method is compared with other methods on matrix completion. These reference methods are SRF [15], IALM [16]¹, LMaFit [17], and FPCA [18]. Note that all the parameters in these reference algorithms are set as default, and the precondition that a projection matrix is symmetric is utilized in all of them.

In each trail, we first generate a random matrix in $\mathbb{R}^{N \times s}$ whose entries are independent identically distributed Gaussian random variable, and then apply QR decomposition to obtain an $N \times s$ matrix X^* with orthonormal columns. Such X^* represents the target subspace that is to be estimated. The measurement is obtained by randomly choosing M entries from $X^*(X^*)^T$. The error of the estimation

¹IALM is not compared in the third figure, in that its released code returns NaN or Inf when the measurement SNR is 10dB.

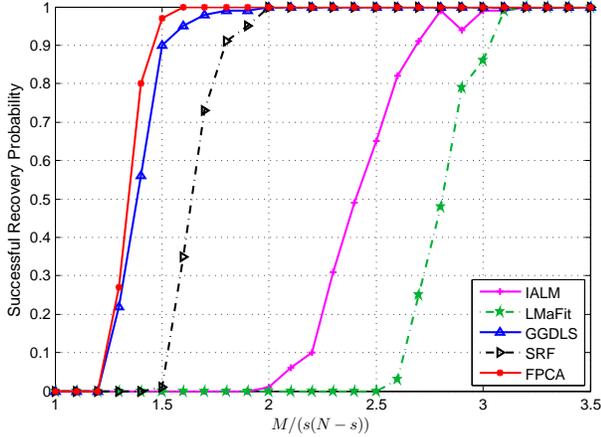


Fig. 1. Probability of successful recovery under various measurement number M . $N = 100$, $s = 10$.

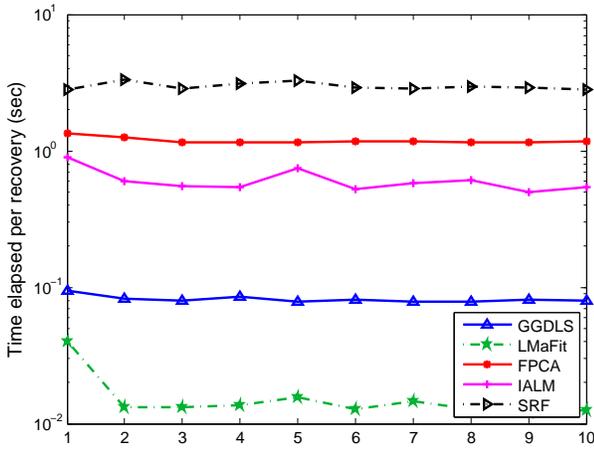


Fig. 2. Time consumption for each algorithm in 10 trails. $N = 100$, $s = 10$. The ratio of $\frac{M}{s(N-s)}$ is chosen as 3.5 so that all recorded recoveries are successful.

is computed as $\|\hat{X}\hat{X}^T - X^*(X^*)^T\|_F$, and a recovery is successful if its error is less than 10^{-2} in the noiseless cases. When the measurement noise is considered in the simulation, each sampling is contaminated by additive Gaussian noise. In all of these trails, $N = 100$ and $s = 10$.

The result of the probability of successful recovery under various sampling number is demonstrated in figure 1. For each point, 100 times of trails are simulated. The time consumption of these algorithms is shown in figure 2. 10 trails are recorded for each algorithm, and the ratio of $\frac{M}{s(N-s)}$ is chosen as 3.5 so that all recorded recoveries are successful. The robustness of these algorithms is tested in figure 3, in which the ratio of $\frac{M}{s(N-s)}$ is also chosen as 3.5. Under each level of the measurement SNR (MSNR), 100 times of trails are simulated for each algorithm, and the reconstructed SNR (RSNR) is shown.

From the simulation result, it is shown that in the noiseless cases, for successful recovery with probability 1, the required number of samples M for the proposed GGDLS is $2s(N-s)$. Given a fixed

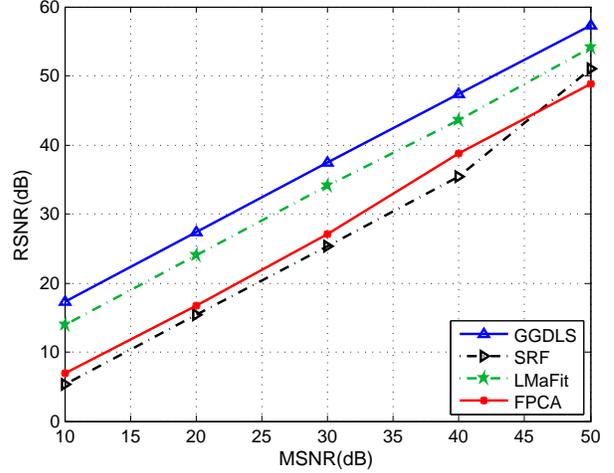


Fig. 3. Reconstructed SNR under various measurement SNR. $N = 100$, $s = 10$. The ratio of $\frac{M}{s(N-s)}$ is chosen as 3.5.

recovery probability, the required M for GGDLS is more than that of FPCA, but the time consumption of the proposed GGDLS is only about 10 percent of that of FPCA. Although the time consumption of LMaFit is less than the proposed GGDLS, the required M for GGDLS is obviously less than that for LMaFit. In the noisy scenario, the reconstruction SNR that the proposed GGDLS is able to achieve is 7dB higher than the measurement SNR both under high MSNR cases and low MSNR cases. Thus, its robustness is the best among all the reference algorithms in this experiment.

It can be concluded that among the algorithms in our experiment, in the noiseless cases, given random samples of a fixed number, the proposed GGDLS is time efficient with rather high probability of successful recovery. In the noisy cases, both when the MSNR is high and low, the proposed GGDLS outperforms the other reference algorithms.

6. CONCLUSION

In this work, we propose a method on projection matrix completion (GGDLS) based on the line-search algorithm on a manifold, and achieve compressive subspace estimation accordingly. The number of samples can be reduced to $2s(N-s)$ in the numerical experiments. Compared with other reference algorithms on matrix completion, in the noiseless cases, given random samples of a fixed number, when both the time consumption and the successful recovery probability are considered, the proposed algorithm exhibits advantage in the completion of projection matrix. In the noisy cases, the proposed GGDLS is remarkably robust against the noise both under high MSNR and low MSNR.

This work is not exhausted, and there are further works to do on such problem. For instance, for our proposed method, the theoretical lower bound on the number of random samplings is to be studied. It should be between $s(N-s)$ and $2s(N-s)$, in that the theoretical degrees of freedom is $s(N-s)$, and our simulations show that $2s(N-s)$ is enough for recovery with probability 1. Also, when the data is contaminated with noise, the robustness of the proposed method is still to be theoretically analyzed.

7. REFERENCES

- [1] K.C. Lee, J. Ho, and D. Kriegman. Acquiring linear subspaces for face recognition under variable lighting, May 2005.
- [2] S. Rao, R. Tron, R. Vidal, and Y. Ma. Motion segmentation in the presence of outlying, incomplete, or corrupted trajectories, Oct 2010.
- [3] J. Ho, K.C. Lee, M.H. Yang, and D. Kriegman. Visual tracking using learned linear subspaces. In *Computer Vision and Pattern Recognition, 2004 IEEE Computer Society Conference on*, volume 1, pages I-782–I-789 Vol.1, June 2004.
- [4] E.J. Candès and B. Recht. Exact low-rank matrix completion via convex optimization. In *Communication, Control, and Computing, 2008 46th Annual Allerton Conference on*, pages 806–812, Sept 2008.
- [5] P. Jain, P. Netrapalli, and S. Sanghavi. Low-rank matrix completion using alternating minimization. In *Proceedings of the 45th annual ACM symposium on Theory of computing, 2013*, pages 665–674.
- [6] Y.M. Lui. Advances in matrix manifolds for computer vision. *Image and Vision Computing*, 30(6-7):380 – 388, 2012.
- [7] A. Edelman, T. Arias, and S. Smith. The geometry of algorithms with orthogonality constraints. *SIAM Journal on Matrix Analysis and Applications*, 20(2):303–353, 1998.
- [8] P.A. Absil, R. Mahony, and R. Sepulchre. Riemannian geometry of grassmann manifolds with a view on algorithmic computation. *Acta Applicandae Mathematica*, 80(2):199–220, 2004.
- [9] X. Dong, P. Frossard, P. Vandergheynst, and N. Nefedov. Clustering on multi-layer graphs via subspace analysis on grassmann manifolds. In *Global Conference on Signal and Information Processing, 2013 IEEE*, pages 993–996, Dec 2013.
- [10] P.A. Absil, R. Mahony, and R. Sepulchre. *Optimization Algorithms on Matrix Manifolds*. Princeton University Press, 2008.
- [11] S. Mittal and P. Meer. Conjugate gradient on grassmann manifolds for robust subspace estimation. *Image and Vision Computing*, 30(6-7):417 – 427, 2012.
- [12] L. Balzano, R. Nowak, and B. Recht. Online identification and tracking of subspaces from highly incomplete information. In *Communication, Control, and Computing (Allerton), 2010 48th Annual Allerton Conference on*, pages 704–711, Sept 2010.
- [13] B. Vandereycken. Low-rank matrix completion by riemannian optimization. *SIAM Journal on Optimization*, 23(2):1214–1236, 2013.
- [14] L. Chen and Y. Gu. Robust recovery of low-rank matrices via non-convex optimization. In *Digital Signal Processing, 2014 19th International Conference on*, pages 355–360, Aug 2014.
- [15] H. Ghasemi, M. Malek-Mohammadi, M. Babaie-Zade, and C. Jutten. Srf: Matrix completion based on smoothed rank function. In *Proceedings of ICASSP 2011*, pages 3672–3675.
- [16] Z. Lin, R. Liu, and Z. Su. Linearized alternating direction method with adaptive penalty for low-rank representation. In *Advances in Neural Information Processing Systems 24*, pages 612–620. 2011.
- [17] Z. Wen, W. Yin, and Y. Zhang. Solving a low-rank factorization model for matrix completion by a nonlinear successive over-relaxation algorithm. *Mathematical Programming Computation*, 4(4):333–361, 2012.
- [18] S. Ma, D. Goldfarb, and L. Chen. Fixed point and bregman iterative methods for matrix rank minimization. *Mathematical Programming*, 128(1-2):321–353, 2011.