BEYOND UNION OF SUBSPACES: SUBSPACE PURSUIT ON GRASSMANN MANIFOLD FOR DATA REPRESENTATION

*Xinyue Shen** *Hamid Krim*† *Yuantao Gu**

*Tsinghua National Laboratory for Information Science and Technology
Department of Electronic Engineering, Tsinghua University, Beijing 100084, CHINA
†Electrical and Computer Engineering Department, North Carolina State University
Raleigh, NC 27695, U.S.A.

ABSTRACT

Discovering the underlying structure of a high-dimensional signal or *big data* has always been a challenging topic, and has become harder to tackle especially when the observations are exposed to arbitrary sparse perturbations. In this paper, built on the model of a union of subspaces (UoS) with sparse outliers and inspired by a basis pursuit strategy, we exploit the fundamental structure of a Grassmann manifold, and propose a new technique of pursuing the subspaces systematically by solving a non-convex optimization problem using the alternating direction method of multipliers. This problem as noted is further complicated by non-convex constraints on the Grassmann manifold, as well as the bilinearity in the penalty caused by the subspace bases and coefficients. Nevertheless, numerical experiments verify that the proposed algorithm, which provides elegant solutions to the sub-problems in each step, is able to de-couple the subspaces and pursue each of them under time-efficient parallel computation.

***Index Terms*—** robust subspace pursuit, union of subspaces (UoS), Grassmann manifold constrained optimization, non-convex ADMM, parallel computation

1. INTRODUCTION

Recovering information or an underlying signal from noisy measurements has always been at the center of many signal processing problems. Many functional bases approaches have been proposed over the last two decades starting with a multiscale and nonlinear approximations of signals [1], including other robust reconstruction approaches [2], and more recent similarly inspired techniques [3, 4]. One approach which particularly caught the attention of researchers was that of a representation pursuit, such as the Matching Pursuit [5] and the Basis Pursuit [6], on account of their adaptive representation and hence recovery and of their high efficiency.

When the structure of the data deviates from the perennial vector subspace assumption, e.g., a curved space-manifold, or sets of subspaces, the recovery becomes challenging. Such subspace structures have also recently witnessed a resurgence, as in the recent works on subspace clustering [7–9] which show a particularly interesting and a promising potential of sparse models. In [7], a low-rank representation (LRR) also recovers subspace structures from sample-specific corruptions by jointly pursuing the lowest-rank representation of all data. The contaminated samples are sparse among all sampled data. Applications of such critical and important problem are available in signal processing, computer vision, and data mining [10]. Robust principal component pursuit is particularly successful in recovering low dimensional structures of high dimensional data from arbitrary sparse errors [11].

From another aspect, works in retrieving structural data, such as dictionary [12], low-rank matrix [13], and subspace [14–16], via optimization problems on manifolds have unveiled interesting geometry insights. In works [17–19], subspace projection matrix recovery from noisy partial information, as a subspace estimation problem, has been studied via non-convex optimization on a Grassmann manifold, and algorithms and theoretical recovery conditions are established by further exploiting the manifold structure.

Moreover, successful applications of sparse models in computer vision and machine learning [20] [21] have increasingly hinted at a more general model, namely that the underlying structure of high dimensional data looks more like *a union of subspaces* (UoS) [8, 9, 22–24] rather than *one low dimensional subspace*. In [22], the natural question about the feasibility of such an approach in high dimensional data modeling where the union of subspaces is further impacted by sparse errors, is addressed. This problem is intrinsically difficult, since the underlying subspace structure is also corrupted by unknown errors, which may lead to unreliable measurement of distance among data samples, and make data deviate from the original subspaces.

In this paper, built on the UoS with sparse outliers data model and inspired by a basis pursuit strategy, we exploit the fundamental structure of a Grassmann manifold, and propose a new technique where the subspaces of a UoS are systematically pursued by solving non-convex optimization problem on a Grassmann manifold. This problem as noted is further complicated by non-convex constraints as called for by the Grassmann manifold, as well as the bilinearity in the penalty caused by the subspace bases and coefficients. Nevertheless, the proposed algorithm gives an elegant solution to the Grassmann manifold-constraint part, and decouples the subspaces and admits parallel computation.

The following text is organized as follows. In Section 2, the background is reviewed and the problem is formulated as an optimization functional. In Section 3, the derivation is conducted to solve this optimization problem. In Section 4, we illustrate this approach by numerical experiments, and finally draw some concluding remarks in Section 5.

*This work was partially supported by National Natural Science Foundation of China (NSFC 61371137, 61571263, 61531166005, 51459003) and Tsinghua University Initiative Scientific Research Program (Grant 2014Z01005). The corresponding author of this paper is Yuantao Gu (gyt@tsinghua.edu.cn).

†The work of HK was in part supported by the following grants, HQ0147-15-C-6008 P002, DE-NA0002576 and W911NF-04-D-0003-0022.

2. BACKGROUND AND PROBLEM FORMULATION

2.1. Grassmann Manifold

Definition 1. *[25] A Grassmann manifold $\mathrm{Gr}_{d,l}$ is the set of l-dimensional linear subspaces of $\mathbb{R}^d$, and is an $l(d-l)$-dimensional compact manifold. It can be formally written as a quotient manifold*

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

where $\mathcal{O}(N)$ is the $N \times N$ orthogonal matrix group. A representative element of $\mathrm{Gr}_{d,l}$ is a $d \times l$ orthonormal matrix $\mathbf{S}$.

The projection distance defined on $\mathrm{Gr}_{d,l}$ is as the following.

$$\forall \mathbf{S}, \mathbf{V} \in \mathrm{Gr}_{d,l},\ d_p(\mathbf{S}, \mathbf{V}) := \frac{1}{2}\|\mathbf{S}\mathbf{S}^{\mathrm{T}} - \mathbf{V}\mathbf{V}^{\mathrm{T}}\|_F^2 = \sum_{i=1}^{l} \sin^2 \theta_i,$$

where $\{\theta_i\}_{i=1}^{l}$ are the principal angles between the two subspaces $\mathcal{S}$ and $\mathcal{V}$. In this work, we use d_p to evaluate the distance between the target subspace and the estimated subspace.

2.2. Problem Statement and Formulation

We are given a $d \times n$ data matrix $\mathbf{X}$ whose structure is a mixture of a noise-free component and perturbation of outliers, and hence admits a decomposition $\mathbf{X} = \mathbf{L} + \mathbf{E}$. The matrix $\mathbf{E}$ is thus a sparse matrix, and the columns of $\mathbf{L}$ belong to k unknown l-dimensional subspaces $\{\mathcal{S}_1, \cdots, \mathcal{S}_k\}$. We denote the orthogonal bases of $\mathcal{S}_j$ by $\mathbf{S}_j \in \mathbb{R}^{d \times l}$, then $\mathbf{L} = \sum_{i=j}^{k} \mathbf{S}_j \boldsymbol{\alpha}_j$, where $\boldsymbol{\alpha}_j \in \mathbb{R}^{l \times n}$. Our aim is to recover $\mathbf{E}$ and $\{\mathcal{S}_1, \cdots, \mathcal{S}_k\}$ from $\mathbf{X}$.

To that end, the recovery of $\mathbf{L}$ in its assumed structure proceeds by a sparse constraint on $\mathbf{E}$, which is in turn, reflected by its ℓ_1 norm (the vector norm) as a convex relaxation of its ℓ_0 norm. Each data point $\mathbf{X}_i$, which is the ith column of $\mathbf{X}$, is supposed to be only represented by the base vectors of its own subspace, i.e., for each i the vector $\boldsymbol{\alpha}^i := [\|\alpha_{i1}\|_2, \cdots, \|\alpha_{ik}\|_2]$ should only have one nonzero element, in which α_{ij} is the ith column of $\boldsymbol{\alpha}_j$. This, in turn, requires a sparse constraint on each $\boldsymbol{\alpha}^i$, i.e., its ℓ_1 norm should also form a penalty function as well.

In light of the above, and towards recovering $\mathbf{L}$ and its associated orthogonal bases and $\mathbf{E}$, we propose to solve the following optimization problem to find $\mathbf{E}$ and $\{\mathcal{S}_1, \cdots, \mathcal{S}_k\}$ from $\mathbf{X}$.

$$\begin{aligned} &\min_{\mathbf{E}, \mathbf{S}_j, \boldsymbol{\alpha}_j, j=1,\cdots,k} \lambda\|\mathbf{E}\|_1 + \sum_{j=1}^{k}\sum_{i=1}^{n} \|\alpha_{ij}\|_2 \\ &\text{s.t. } \mathbf{X} = \mathbf{E} + \sum_{j=1}^{k} \mathbf{S}_j \boldsymbol{\alpha}_j,\ \mathbf{S}_j \in \mathrm{Gr}_{d,l}, j = 1, \cdots, k. \end{aligned} \tag{2}$$

For the given Grassmann manifold $\mathrm{Gr}_{d,l}$, our search is for l dimensional subspaces in $\mathbb{R}^d$. A brief and careful inspection of (2) reveals that the bilinearity of $\boldsymbol{\alpha}$ and $\mathbf{S}_j, j = 1, \cdots, k$, term results in the non-convexity of problem (2). This in turn, suggests our adoption of ADMM [26] as a strategy towards a solution. Keep the $\mathbf{S}_j$ constraints in (2), and the augmented Lagrangian is as follows,

$$\begin{aligned} L(\mathbf{E}, \mathbf{S}_j, \boldsymbol{\alpha}_j, \mathbf{Y}) &= \lambda\|\mathbf{E}\|_1 + \sum_{j=1}^{k}\sum_{i=1}^{n} \|\alpha_{ij}\|_2 \\ &+ \langle \mathbf{Y}, \mathbf{X} - \mathbf{E} - \sum_{j=1}^{k} \mathbf{S}_j \boldsymbol{\alpha}_j \rangle + \frac{\mu}{2} \left\| \mathbf{X} - \mathbf{E} - \sum_{j=1}^{k} \mathbf{S}_j \boldsymbol{\alpha}_j \right\|_F^2. \end{aligned}$$

3. SUBSPACE PURSUIT ANALYSIS

Proceeding in a similar spirit as basis pursuit [6], we propose to solve for a basis pursuit on a Grassmann manifold by solving the following optimization problem,

$$\max_{\mathbf{Y}} \min_{\mathbf{E}, \mathbf{S}_j, \boldsymbol{\alpha}_j} L(\mathbf{E}, \mathbf{S}_j, \boldsymbol{\alpha}_j, \mathbf{Y}_i),\ \text{ s.t.} \mathbf{S}_j \in \mathrm{Gr}_{d,l}, j = 1, \cdots, k, \tag{3}$$

which we next propose to independently and iteratively solve over $\mathbf{E}, \boldsymbol{\alpha}_j, \mathbf{S}_j$, $j = 1, \cdots, k$ and $\mathbf{Y}$, where the global iteration number is denoted by t.

3.1. Algorithmic Derivation

We first show that the minimization sub-problem over $\mathbf{E}$ has a closed form solution, when accounting for the decoupled nature of the columns of $\mathbf{E}$. As a result, the problem admits a column-wise parallel solution.

Claim 1: Solution of Problem (3) over $\mathbf{E}$ is solved by

$$\min_{\mathbf{E}_i} \lambda\|\mathbf{E}_i\|_1 - \langle \mathbf{Y}_i^t, \mathbf{E}_i \rangle + \frac{\mu}{2} \left\| \mathbf{E}_i - \mathbf{X}_i + \sum_{j=1}^{k} \mathbf{S}_j^t \alpha_{ij}^t \right\|_2^2. \tag{4}$$

This is readily established by first denoting each column of $\mathbf{E}$ by $\mathbf{E}_i$. Solving the above sub-problems (4) in parallel for $i = 1, \cdots, n$ is equivalent to solving $\min_{\mathbf{E}} L(\mathbf{E}, \mathbf{S}_j^t, \boldsymbol{\alpha}_j^t, \mathbf{Y}^t)$. The solution can be written as the proximal operator of the ℓ_1 norm, whose closed form solution is calculated by soft-thresholding [27].

$$\mathbf{E}_i^{t+1} = \mathrm{prox}_{\frac{\mu}{2\lambda}\|\cdot\|_1} \left(\mathbf{X}_i - \sum_{j=1}^{k} \mathbf{S}_j^t \alpha_{ij}^t + \frac{\mathbf{Y}_i^t}{\mu} \right). \tag{5}$$

The next step in solving the formulated optimization problem (3), is the sub-problem over the coefficient set $\boldsymbol{\alpha}$. Specifically, for each α_{ij}, the iteration is parallel in i but not in j, to state the following.

Claim 2: Each such sub-problem entails,

$$\begin{aligned} &\min_{\alpha_{ij}} \|\alpha_{ij}\|_2 - \langle \mathbf{Y}_i^t, \mathbf{S}_j^t \alpha_{ij} \rangle \\ &\quad + \frac{\mu}{2} \left\| \mathbf{X}_i - \mathbf{E}_i^{t+1} - \sum_{p<j} \mathbf{S}_p^t \alpha_{ip}^{t+1} - \sum_{p>j}^{k} \mathbf{S}_p^t \alpha_{ip}^t - \mathbf{S}_j^t \alpha_{ij} \right\|_2^2. \end{aligned} \tag{6}$$

This also admits a closed form solution as

$$\alpha_{ij}^{t+1} = \mathrm{prox}_{\frac{\mu}{2}\|\cdot\|_2} \left(\left(\mathbf{S}_j^t\right)^{\mathrm{T}} \mathbf{v}_{ij} \right), \tag{7}$$

where

$$\mathbf{v}_{ij} = \mathbf{X}_i - \mathbf{E}_i^{t+1} - \sum_{p<j} \mathbf{S}_p^t \alpha_{ip}^{t+1} - \sum_{p>j}^{k} \mathbf{S}_p^t \alpha_{ip}^t + \mathbf{Y}_i^t/\mu.$$

For fixed j, the above can be computed in parallel for $i = 1, \cdots, n$. The third key step in attaining our overall solution, entails solving the minimization problem for each $\mathbf{S}_j \in \mathrm{Gr}_{d,l}$. Such a manifold optimization problem, while in appearance daunting even by way of an iterative on account of the non-convex Grassmann manifold, may in fact be shown to admit a closed form solution as we elaborate

next.

Claim 3: The optimized subspace pursuit over $\mathrm{Gr}_{d,l}$ is achieved by,

$$\begin{aligned}\mathbf{S}_j^{t+1} &= \arg\min_{\mathbf{S}_j\in\mathrm{Gr}_{d,l}} \langle \mathbf{Y}^t, -\mathbf{S}_j\boldsymbol{\alpha}_j^{t+1}\rangle \\ &+\frac{\mu}{2}\left\|\mathbf{X}-\mathbf{E}^{t+1}-\sum_{p<j}\mathbf{S}_p^{t+1}\boldsymbol{\alpha}_p^{t+1}-\sum_{p>j}^{k}\mathbf{S}_p^{t}\boldsymbol{\alpha}_p^{t+1}-\mathbf{S}_j\boldsymbol{\alpha}_j^{t+1}\right\|_F^2 \\ &= \arg\min_{\mathbf{S}_j\in\mathrm{Gr}_{d,l}} \left\langle \mathbf{Y}^t+\mu\Big(\mathbf{X}-\mathbf{E}^{t+1}\right. \\ &\quad \left.-\sum_{p<j}\mathbf{S}_p^{t+1}\boldsymbol{\alpha}_p^{t+1}-\sum_{p>j}^{k}\mathbf{S}_p^{t}\boldsymbol{\alpha}_p^{t+1}\Big), -\mathbf{S}_j\boldsymbol{\alpha}_j^{t+1}\right\rangle \\ &= \arg\min_{\mathbf{S}_j\in\mathrm{Gr}_{d,l}} \left\langle \Big(-\mathbf{Y}^t/\mu-\mathbf{X}+\mathbf{E}^{t+1}\right. \\ &+\left.\sum_{p<j}\mathbf{S}_p^{t+1}\boldsymbol{\alpha}_p^{t+1}+\sum_{p>j}^{k}\mathbf{S}_p^{t}\boldsymbol{\alpha}_p^{t+1}\Big)\cdot(\boldsymbol{\alpha}_j^{t+1})^{\mathrm{T}}, \mathbf{S}_j\right\rangle. \end{aligned} \tag{8}$$

The follow up closed form solution is obtained by first defining the singular value decomposition of

$$\left(\mathbf{E}^{t+1}+\sum_{p<j}\mathbf{S}_p^{t+1}\boldsymbol{\alpha}_p^{t+1}+\sum_{p>j}^{k}\mathbf{S}_p^{t}\boldsymbol{\alpha}_p^{t+1}-\frac{\mathbf{Y}^t}{\mu}-\mathbf{X}\right)\cdot(\boldsymbol{\alpha}_j^{t+1})^{\mathrm{T}}$$

as $\mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^{\mathrm{T}}$, then $\langle\mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^{\mathrm{T}},\mathbf{S}_j\rangle=\langle\mathbf{V}^{\mathrm{T}}\mathbf{S}_j^{\mathrm{T}}\mathbf{U},\boldsymbol{\Sigma}\rangle$. Since each entry of matrix $\mathbf{V}^{\mathrm{T}}\mathbf{S}_j^{\mathrm{T}}\mathbf{U}$ is not larger than 1, and $\boldsymbol{\Sigma}$ is diagonal, we have $\langle\mathbf{V}^{\mathrm{T}}\mathbf{S}_j^{\mathrm{T}}\mathbf{U},\mathbf{S}\rangle\geq\langle-\mathbf{I},\boldsymbol{\Sigma}\rangle$, and the equality holds when

$$\mathbf{S}_j=-\mathbf{U}\mathbf{V}^{\mathrm{T}}.$$

The above calculation for $\mathbf{S}_j^{t+1}$ requiring a singular value decomposition, is not parallel for j. This, as result, may entail a significant computational burden for large scale problems.

Claim 4: Towards lifting such a limitation, we propose to de-couple $\{\mathbf{S}_j\}_{j=1}^k$ by the following update of each α_{ij} at the end of the second step.

$$\alpha_{ij}^{t+1}=\begin{cases}\hat{\alpha}_{ij}^{t+1}, & j=\arg\max_{j=1,\cdots,k}\|\hat{\alpha}_{ij}^{t+1}\|_2,\\ \mathbf{0}, & \text{otherwise},\end{cases} \tag{9}$$

where $\hat{\alpha}_{ij}^{t+1}$ is the output of the second step.

Since each data point in $\mathbf{L}$ should only belong to one subspace, for each fixed i we only preserve the column α_{ij} with the largest ℓ_2 norm, while the others are set to $\mathbf{0}$. For a fixed i, denote the selected j as $j_{(i)}$. The benefit in so doing is that the objective function can be simplified, since if $j=j_{(i)}$, then $\{\alpha_{ip}\}_{p\neq j}$ are all zeros, and if $j\neq j_{(i)}$, then $\alpha_{ij}\mathbf{S}_j$ vanishes. Therefore, (8) is simplified as follows.

$$\begin{aligned}\mathbf{S}_j^{t+1} =& \arg\min_{\mathbf{S}_j\in\mathrm{Gr}_{d,l}} \sum_{i=1}^{n}\langle\mathbf{Y}_i^t,\mathbf{X}_i-\mathbf{E}_i^{t+1}-\mathbf{S}_{j_{(i)}}\alpha_{ij_{(i)}}^{t+1}\rangle \\ &+\frac{\mu}{2}\sum_{i=1}^{n}\|\mathbf{X}_i-\mathbf{E}_i^{t+1}-\mathbf{S}_{j_{(i)}}\alpha_{ij_{(i)}}^{t+1}\|_2^2 \\ =& \arg\min_{\mathbf{S}_j\in\mathrm{Gr}_{d,l}} \sum_{i=1}^{n}\mathbf{1}_{[j_{(i)}=j]}\langle\mu\mathbf{E}_i^{t+1}-\mu\mathbf{X}_i-\mathbf{Y}_i^t,\mathbf{S}_j\alpha_{ij}^{t+1}\rangle \\ =& -\mathbf{U}\mathbf{V}^{\mathrm{T}}, \end{aligned} \tag{10, 11}$$

Table 1. Algorithm

Input:	Data matrix $\mathbf{X}\in\mathbb{R}^{d\times n}$, d, n, k, l, Initial iterate $\mathbf{S}_j^0\in\mathrm{Gr}_{d,l}$, $\boldsymbol{\alpha}_j^0, j=1,\cdots,k, \mathbf{Y}^0$; Algorithm parameters λ,μ;
Output:	$\hat{\mathbf{E}},\hat{\mathbf{S}}_j, j=1,\cdots,k$.
For $t=0,1,2,\cdots$ **do:**	
1.	**For** $i=1,\ldots,n$, **do in parallel:** Update $\mathbf{E}_i^{t+1}$ by (5) ;
2.	**For** $j=1,\ldots,k$, **do:** **For** $i=1,\ldots,n$, **do in parallel:** Update $\hat{\alpha}_{ij}^{t+1}$ by (7) ;
3.	**For** $i=1,\ldots,n$, **do in parallel:** Update α_{ij}^{t+1} by (9) ;
4.	**For** $j=1,\ldots,k$, **do in parallel:** Update $\mathbf{S}_j^{t+1}$ by (10)
5.	Update $\mathbf{Y}^{t+1}$ by (12)
Until:	Stopping criterion satisfied;

where $\mathbf{U}$ and $\mathbf{V}$ are from the singular value decomposition of the matrix $\sum_{i=1}^{n}\mathbf{1}_{[j_{(i)}=j]}(\mu\mathbf{E}_i^{t+1}-\mu\mathbf{X}_i-\mathbf{Y}_i^t)(\alpha_{ij}^{t+1})^{\mathrm{T}}$. Therefore, this step can be computed in parallel for each $\mathbf{S}_j$.

The last step in the algorithm proceeds with the dual ascent, and is carried out by

$$\mathbf{Y}^{t+1}=\mathbf{Y}^t+\mu\left(\mathbf{X}-\mathbf{E}^{t+1}-\sum_{j=1}^{k}\mathbf{S}_j^{t+1}\boldsymbol{\alpha}_j^{t+1}\right). \tag{12}$$

For conciseness, the above details are summarized in the following Table 1.

4. NUMERICAL EXPERIMENTS

For such non-convex ADMM algorithm, the convergence is not theoretically insured, and is highly dependent on the parameter μ and λ, and the initial values. In the simulations, we choose the initial values for $\mathbf{S}_j^0$ from the column space of $\mathbf{X}$, and the initial value for $\mathbf{Y}^0$ is the zero matrix. The stopping criterion is either that the number of iterations has reached to a maximum or that the primal residue $\|\mathbf{Y}^{t+1}-\mathbf{Y}^t\|_F\leq\epsilon$. The subspace error is computed by $\sum_{j=1}^{k}\|\mathbf{S}_j^*(\mathbf{S}_j^*)^{\mathrm{T}}-\hat{\mathbf{S}}_j(\hat{\mathbf{S}}_j)^{\mathrm{T}}\|_F^2$, and the $\mathbf{E}$-error is computed as $\|\mathbf{E}^*-\hat{\mathbf{E}}\|_F^2$. For the true values, $\mathbf{E}^*$ is randomly generated with uniformly random non-zero positions and Gaussian distributed random values, and $\mathbf{L}^*$ is randomly generated with orthogonalized Gaussian matrices $\mathbf{S}_j^*$ and Gaussian distributed coefficients $\boldsymbol{\alpha}^*$.

In the following, we will demonstrate numerical results both with fixed μ and with incremental μ respectively.

4.1. Fixed μ

An example of the convergence curves is in Fig. 1, in which there are $n=45$ data points in $\mathbb{R}^{10}$, $k=3$, $l=3$, and the number of non-zero elements in $\mathbf{E}^*$ is 4. From Fig. 1, we can see that, on the one hand, if λ is too small relative to μ, then the recovered $\hat{\mathbf{E}}$ is not sparse enough, so the error is larger. On the other hand, if λ and μ increase at the same time and become too large, then the algorithm becomes hard to converge.

According to the previous inspection, in the next experiment, we test the optimal λ and μ for 100 randomly generated data sets. In

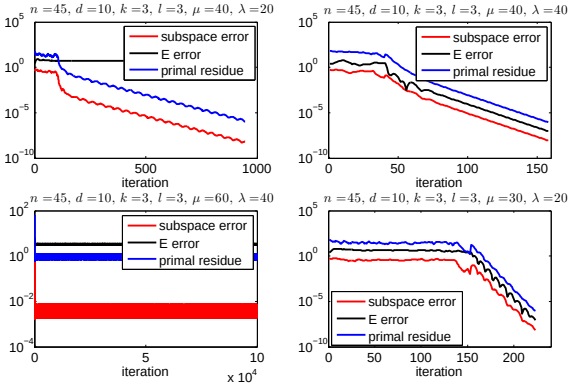

Fig. 1. convergence curves

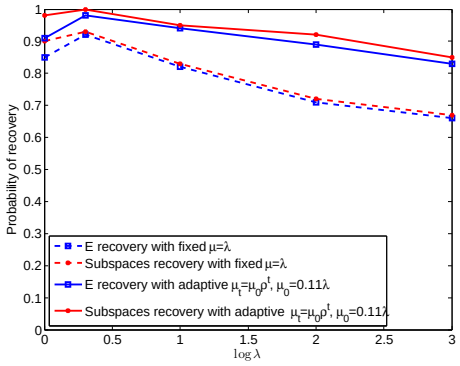

Fig. 2. probability of recovery for random data, $n = 24$, $d = 12$, $l = 2$, $k = 2$

each trial, the number of points $n = 24$, the dimension of the ambient space $d = 12$, the number of subspaces $k = 2$, the dimension of each subspace $l = 2$, and the sparsity of $\mathbf{E}^*$ is 3. The simulation results are the dashed curves in Fig. 2. Note that for successful recovery, we mean that the error is less than 0.01. Firstly, the probability of recovering subspaces is higher than that of recovering $\mathbf{E}$. When $\lambda = 1$ the gap is the largest, since λ is too small to induce proper sparsity. Secondly, unlike some other non-convex ADMM algorithms [28], simply increasing μ does not ensure better convergence for the proposed algorithm. We can see that a relatively better choice is $\lambda = \mu = 2$, and the probability of recovery is about 0.93.

4.2. Adaptive μ

For a better convergence behavior, we propose to use adaptive μ during the iterations. To be specific, we set $\mu^{t+1} = \rho^t \mu_0$ and $\rho > 1$.

In the following experiment, the optimal λ and μ_0 are tested for 100 random data, and $\rho = 1.00001$. The data settings are the same as the previous experiment, and the results are the solid curves in Fig. 2. Compared with the dashed curves, the convergence performance is improved, and with $\lambda = 2$ and $\mu_0 = 0.22$, the probability of subspace recovery reaches 1.

For a better understanding of the choice of parameters λ and μ, we test the subspace recovery probability with various λ and μ_0 in Fig. 3. In this experiment, $\rho = 1.00001$, $n = 24$, $d = 12$, $l = 2$, $k = 3$, and the sparsity level of $\mathbf{E}^*$ is 3. Each probability is counted from 100 random trials. From the result, we first see that when $\mu_0 >$

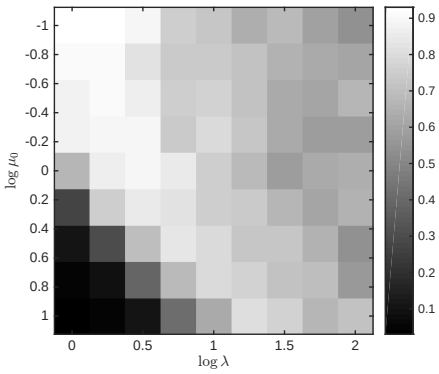

Fig. 3. probability of recovery for random data, $n = 24$, $d = 12$, $l = 2$, $k = 3$

λ the recovery is hardly achieved. Secondly, when $\mu_0 < \lambda$ the probability of recovery generally decreases as λ increases from 1 to 1000. The optimal choice in this experiment is $\lambda = 1$ and $\mu_0 = 0.1$.

5. CONCLUSION

In this paper an optimization problem with constraints on Grassmann manifold is formulated to model the robust subspace pursuit problem of recovering a union of subspaces from a data matrix corrupted by sparse errors. Adopting the strategy of ADMM, the sparse errors, the representation coefficients, and the latent subspaces are iteratively solved in parallel by columns. Because of the non-convex constraint and the bilinearity in cost function, the convergence to a global optimum is difficult to be theoretically guaranteed. However, numerical simulations demonstrate that with elaborately selected parameter set, the union of subspaces can be successfully recovered in a large probability close to unitary.

This work is not exhausted, and there are further works to do on such problem. The optimal choice of λ is to be theoretically analyzed for a given relative sparsity level between $\mathbf{E}$ and the coefficients. For the algorithm, the impact of μ on the convergence behavior is to be theoretically studied.

6. REFERENCES

[1] Stephane Mallat. *A wavelet tour of signal processing: the sparse way*. Academic press, 2008.

[2] Hamid Krim and Irvin C Schick. Minimax description length for signal denoising and optimized representation. *Information Theory, IEEE Transactions on*, 45(3):898–908, 1999.

[3] David L Donoho. Compressed sensing. *Information Theory, IEEE Transactions on*, 52(4):1289–1306, 2006.

[4] Emmanuel J Candès and Terence Tao. Near-optimal signal recovery from random projections: Universal encoding strategies? *Information Theory, IEEE Transactions on*, 52(12):5406–5425, 2006.

[5] Stéphane G Mallat and Zhifeng Zhang. Matching pursuits with time-frequency dictionaries. *Signal Processing, IEEE Transactions on*, 41(12):3397–3415, 1993.

[6] Scott Shaobing Chen, David L Donoho, and Michael A Saunders. Atomic decomposition by basis pursuit. *SIAM journal on scientific computing*, 20(1):33–61, 1998.

[7] Guangcan Liu, Zhouchen Lin, Shuicheng Yan, Ju Sun, Yong Yu, and Yi Ma. Robust recovery of subspace structures by low-rank representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(1):171–184, 2013.

[8] Ehsan Elhamifar and René Vidal. Sparse subspace clustering. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 2790–2797. IEEE, 2009.

[9] Mahdi Soltanolkotabi, Ehsan Elhamifar, and Emmanuel Candès. Robust subspace clustering. *arXiv preprint arXiv:1301.2603*, 2013.

[10] Michael Elad. Sparse and redundant representation modeling: What next? *IEEE Signal Processing Letters*, 19:922–928, December 2012.

[11] Emmanuel J. Candès, Xiaodong Li, Yi Ma, and John Wright. Robust principal component analysis? *J. ACM*, 58(3):11:1–11:37, June 2011.

[12] J. Sun, Q. Qu, and J. Wright. Complete dictionary recovery over the sphere. *arXiv preprint arXiv:1504.06785*, 2015.

[13] B. Vandereycken. Low-rank matrix completion by riemannian optimization. *SIAM Journal on Optimization*, 23(2):1214–1236, 2013.

[14] S. Mittal and P. Meer. Conjugate gradient on grassmann manifolds for robust subspace estimation. *Image and Vision Computing*, 30(6-7):417 – 427, 2012.

[15] Laura Balzano, Robert Nowak, and Benjamin 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. IEEE, 2010.

[16] Jun He, Laura Balzano, and John Lui. Online robust subspace tracking from partial information. *arXiv preprint arXiv:1109.3827*, 2011.

[17] Xinyue Shen and Yuantao Gu. Restricted isometry property of subspace projection matrix under random compression. *Signal Processing Letters, IEEE*, 22(9):1326–1330, Sept 2015.

[18] Xinyue Shen and Yuantao Gu. Subspace projection matrix completion on grassmann manifold. In *Acoustics, Speech and Signal Processing (ICASSP), 2015 IEEE International Conference on*, pages 3297–3301, April 2015.

[19] Xinyue Shen and Yuantao Gu. Subspace projection matrix recovery from incomplete information. *http://gu.ee.tsinghua.edu.cn/index.php/component/attachments/download/78*, 2015.

[20] Michael Elad, Mario AT Figueiredo, and Yi Ma. On the role of sparse and redundant representations in image processing. *Proceedings of the IEEE*, 98(6):972–982, 2010.

[21] Ron Rubinstein, Tomer Faktor, and Michael Elad. K-svd dictionary-learning for the analysis sparse model. In *Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on*, pages 5405–5408. IEEE, 2012.

[22] Xiao Bian and Hamid Krim. Robust subspace recovery via bi-sparsity pursuit. *arXiv preprint arXiv:1403.8067*, 2014.

[23] Xiao Bian and Hamid Krim. Bi-sparsity pursuit for robust subspace recovery. In *Image Processing (ICIP), 2015 IEEE International Conference on*, 2015.

[24] Xiao Bian, Hamid Krim, A. Bronstein, and L. Dai. Sparse null space basis pursuit and analysis dictionary learning for high-dimensional data analysis. In *Acoustics, Speech and Signal Processing (ICASSP), 2015 IEEE International Conference on*, 2015.

[25] P.A. Absil, R. Mahony, and R. Sepulchre. *Optimization Algorithms on Matrix Manifolds*. Princeton University Press, 2008.

[26] Stephen Boyd, Neal Parikh, Eric Chu, Borja Peleato, and Jonathan Eckstein. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Found. Trends Mach. Learn.*, 3(1):1–122, January 2011.

[27] Neal Parikh and Stephen Boyd. Proximal algorithms. *Found. Trends Optim.*, 1(3):127–239, January 2014.

[28] Rongjie Lai and Stanley Osher. A splitting method for orthogonality constrained problems. *Journal of Scientific Computing*, 58(2):431–449, 2014.