Iterative Projections for Signal Identification on Manifolds: Global Recovery Guarantees

Parikshit Shah and Venkat Chandrasekaran

***Abstract*— We introduce an algorithm known as Manifold Iterative Projection to solve the problem of recovering an unknown high-dimensional signal contained in a low-dimensional sub-manifold from a few linear measurements. The algorithm *provably* and *robustly* recovers any unknown signal on the manifold, provided the measurement operator is benign with respect to the manifold. A variant of the algorithm provably *tracks* slowly time-varying signals on the manifold. Our results are intimately related to, and indeed rely on, the existence of stable embeddings of manifolds via linear maps.**

I. INTRODUCTION

Making inferences about an unknown signal given limited and inaccurate information is a frequently encountered problem in machine learning. In many applications of interest, a significant difficulty arises as the number of measurements available about a signal may be far smaller than the ambient dimension of the signal. Fortunately, many real-world signals have many fewer degrees of freedom than their ambient dimension. Exploiting such low-dimensional structure in the signal of interest is crucial in order to enable reliable recovery. In this paper we propose an efficient computational framework to recover signals living on low-dimensional manifolds given a small number of noisy *linear* measurements. We propose an algorithm that that is *guaranteed* to recover *any* unknown signal on a given manifold, provided the measurement operator is benign (in a certain formal sense) with respect to the manifold.

Our algorithm consists of two simple steps performed in alternation: a gradient step that is easily computed, and a manifold projection step. For many practical manifold-modeled signals of interests, the manifold projection step is also tractable to compute. The main contributions of this paper are (a) a proof that under fairly weak conditions on the operator specifying the linear measurements, our iterative algorithm is guaranteed to recover the unknown signal, (b) robustness of the algorithm with respect to noise and imprecise computations, and (c) an adaptive version of the algorithm that provably *tracks* slowly time-varying signals on the manifold. These results are applicable to manifold-modeled signals such as orthogonal matrices (describing rotations), subspaces (described as points on the Grassmannian), and partial isometries (points on the Stiefel manifols), which arise in many applications in signal processing and computer vision. Our results also include as special cases structured signals such as sparse vectors and low-rank matrices (although these are more accurately specified as finite unions of manifolds).

Unlike previous methods for learning structured signals (e.g., sparse vectors, low-rank matrices), our approach does not rely on convex optimization. Indeed an appealing property of our procedure is that the complexity is explicitly governed by the tractability of the manifold projection step. Further the analysis of the algorithm is simple, and the bounds on the number of measurements required for recovery depend naturally on the underlying geometry of the manifold. In fact, viewed more broadly, this paper shows that under certain conditions efficient local algorithms can be used to compute globally optimal solutions to nonconvex optimization problems.

Our algorithm builds upon previous work for recovering sparse vectors from linear measurements via iterative hard thresholding [5], and the extension to recovering low-rank matrices from linear measurements via singular value projection [7]. Thus, our paper shows that manifold-modeled signals beyond sparse vectors and low-rank matrices can also be recovered from limited and inaccurate linear information. Indeed the key requirements are only that projection onto the underlying family of signals (modeled here as a manifold, but this can be generalized) is tractable, and that this family has "nice" geometric properties (see the technical results for more details).

This paper is organized as follows: Section II gives a brief background on the relevant concepts from differential geometry, Section III describes the main results, and Section IV discusses the application of our main theorem to specific examples. In Section V, we illustrate our results via some numerical experiments.

II. Preliminaries

To begin with we formally describe our problem setup. We focus on signals lying in a Riemannian submanifold $\mathcal{M} \subseteq \mathbb{R}^n$. We are given measurements via a linear map $A : \mathbb{R}^n \to \mathbb{R}^m$ of some unknown signal $x^* \in \mathcal{M}$ of interest:

$$y = Ax^*.$$

The objective is to recover x^* given the information y. We also consider variants of this question (i) when the measurements are corrupted by additive noise, i.e., we have $y = Ax^* + n$, (ii) the task of projection onto the manifold is difficult and one has access to only *approximate* projections and (iii) when we have a time-varying sequence of signals $x^*[t] \in \mathcal{M}$ with a corresponding sequence of measurements $y[t] = Ax^*[t]$. The objective in the first two variations is to recover an estimate of $\hat{x}$ that lies in $\mathcal{M}$ and is close to x^*. The goal in the third variant is to track the sequence of signals $x^*[t]$.

It is clear that our main problem is ill-posed in full generality as there may be many signals in $\mathcal{M}$ that are consistent with the given measurements y. This is because the linear map A has a nontrivial nullspace. Therefore an important requirement of the operator A is that it satisfy a *restricted isometry* property with respect to the set $\mathcal{M}$. Consequently, we can ensure that there exists a unique signal in $\mathcal{M}$ that is consistent with the information y.

Definition 1: Let $\mathcal{M} \subseteq \mathbb{R}^n$ be a Riemannian manifold and $A \in \mathbb{R}^{n\times m}$. Then A satisfies the *Restricted Isometry Property* with respect to $\mathcal{M}$ with constant $\delta_{\mathcal{M}} \in [0,1)$ if for all $x_1, x_2 \in \mathcal{M}$

$$(1-\delta_{\mathcal{M}})\|x_1-x_2\|^2 \le \|A(x_1-x_2)\|^2 \le (1+\delta_{\mathcal{M}})\|x_1-x_2\|^2. \tag{1}$$

The concept of restricted isometry has played a fundamental role in the analysis of the recovery of sparse vectors [4] and low-rank matrices [10], and we exploit in our paper a generalization of this property to manifolds. In order to prove that linear maps $A : \mathbb{R}^n \to \mathbb{R}^m$ with low dimensional image m satisfy a restricted isometry condition with respect to a manifold $\mathcal{M}$, we require that the manifold satisfy certain regularity conditions. Specifically let V denote the volume of a manifold (which bounds the "size" of the manifold), let $\frac{1}{\tau}$ denote its condition number (which controls the twisting or curvature of the manifold), and let R denote the covering regularity (which controls the ease with which a manifold can be covered by charts). We omit the precise definitions of these concepts here (see [3] for more details), but note that the quantity $C_{\mathcal{M}} = \frac{VR}{\tau}$ is the key regularity parameter summarizing all these preceding quantities.

Theorem 1: [3] Let $\mathcal{M}$ be a compact k-dimensional Riemannian submanifold of $\mathbb{R}^n$ having manifold regularity $C_{\mathcal{M}}$. Fix $0 < \epsilon < 1$ and $0 < \rho < 1$. Let $A : \mathbb{R}^n \to \mathbb{R}^m$ be a random orthoprojector * with

$$m = \mathcal{O}\left(\frac{k\log(nC_{\mathcal{M}}\epsilon^{-1})\log(\rho^{-1})}{\epsilon^2}\right).$$

If $m \le n$ then with probability at least $1-\rho$ the following statement holds: For every pair of points $x_1, x_2 \in \mathcal{M}$,

$$(1-\epsilon)\|x_1-x_2\|^2 \le \|A(x_1-x_2)\|^2 \le (1+\epsilon)\|x_1-x_2\|^2.$$

Thus the dimension of the manifold $\mathcal{M}$, the dimension of the ambient space, and the regularity of $\mathcal{M}$ control the number of measurements that suffice for A to behave as an isometry restricted to $\mathcal{M}$.

Given a manifold $\mathcal{M}$, we define $\mathcal{P}_{\mathcal{M}}(\cdot)$ to be the Euclidean projection operator on to $\mathcal{M}$ so that $z = \mathcal{P}_{\mathcal{M}}(y)$ is a minimizer of the Euclidean distance between y and $\mathcal{M}$.

III. Main Results

In this section we present the main results related to the recovery of unknown signals on a manifold. We present four results in this section. The basic result is regarding the recovery of an unknown signal on a manifold via an algorithm called MIP. Three other interesting variations of the same result are also presented. These variations describe stability of MIP under various scenarios related to the presence of noise, imprecise projections and time-varying nature of the underlying signal. The proofs of some of the main results are included in a separate section at the end. Some proofs are omitted due to space constraints, the proofs of these results are very similar. The proof techniques closely parallel the proof techniques for

*A random orthoprojector may be obtained in standard ways from random matrices. The matrix A must be a suitably scaled orthoprojector [6].

results related to recovery of sparse [5] and low-rank [7] signals.

A. Signal Recovery in the Noiseless Case

We begin with the problem of recovering an unknown signal $x^* \in \mathcal{M}$ from a few random linear measurements of the form $y = Ax^*$. The signal is static, the measurements are noise-free, and it is possible to compute projections onto $\mathcal{M}$ exactly via a projection operator $\mathcal{P}_{\mathcal{M}}$. The algorithm described below takes as input A, y such that $y = Ax^*$, and produces as output a candidate solution x_T. Consider the following algorithm, which we call Manifold Iterative Projection (MIP).

Algorithm: MIP

---

0. Initialization: $z = 0$, $x_0 = \mathcal{P}_{\mathcal{M}}(z)$.

While $k < T$, repeat:
1. $z_{k+1} = x_k - \eta \left(A^T(Ax_k - y) \right)$
2. $x_{k+1} = \mathcal{P}_{\mathcal{M}}(z_{k+1})$
3. Increment k.

Theorem 2: Let $x^* \in \mathcal{M}$ be an unknown signal on the manifold $\mathcal{M}$. Let A be a measurement operator satisfying (1) with $\delta_{\mathcal{M}} < \frac{1}{3}$, and let $y = Ax^*$. Let $\eta = 1/(1 + \delta_{\mathcal{M}})$. Then the algorithm MIP converges, i.e. $x_T \to x^*$ as $T \to \infty$.

Moreover, the algorithm produces $x_T \in \mathcal{M}$ such that $\|x_T - x^*\|^2 \leq \frac{\epsilon}{1-\delta_{\mathcal{M}}}$ and $\|Ax_T - y\|^2 \leq \epsilon$ in

$$T = \left\lceil \frac{1}{\log\left(\frac{(1-\delta_{\mathcal{M}})}{2\delta_{\mathcal{M}}}\right)} \log\left(\frac{\|y\|^2}{2\epsilon}\right) \right\rceil$$

iterations.

Remark This theorem says that x_T converges to the true signal x^* with a convergence rate of $O(\log(\frac{1}{\epsilon}))$. This problem may be viewed as minimizing a quadratic cost $\|Ax - y\|^2$ over the manifold $\mathcal{M}$. The algorithm MIP may be viewed as a first order projected gradient method. It is interesting that one can obtain an exponential rate of convergence in our situation, where one is minimizing a convex function over a manifestly non-convex set $\mathcal{M}$. Interestingly, restricted isometry replaces convexity as a key property in the proof.

B. Signal Recovery in the Noisy Case

The result in Theorem 2 deals with an idealized situation where the measurement process $Ax^* = y$ is noise-free. In many engineering applications, it would be desirable to allow for a moderate amount of noise. The next result shows that the algorithm MIP is robust to additive noise.

Theorem 3: Let $x^* \in \mathcal{M}$ be an unknown signal on the manifold $\mathcal{M}$. Let A be a measurement operator satisfying (1) with $\delta_{\mathcal{M}} < 1/3$, and let $y = Ax^* + e$. Let $\eta = 1/(1+\delta_{\mathcal{M}})$. Let $C = \frac{1+\delta_{\mathcal{M}}}{1-3\delta_{\mathcal{M}}}$. Then the algorithm MIP produces $x_T \in \mathcal{M}$ such that

$$\|x_T - x^*\|^2 \leq \frac{C^2 \frac{\|e\|^2}{2} + \epsilon}{1 - \delta_{\mathcal{M}}}$$
$$\|Ax^* - y\|^2 \leq C^2 \frac{\|e\|^2}{2} + \epsilon,$$

in

$$T = \left\lceil \frac{1}{\log\left(\frac{1}{D}\right)} \log\left(\frac{\|y\|^2}{(C^2)\|e\|^2 + \epsilon}\right) \right\rceil$$

iterations, where $D < 1$ is a constant that depends only on $\delta_{\mathcal{M}}$.

Theorem 3 says that the algorithm MIP is robust to the presence of noise in the measurement process. As one increases the number of iterations, $\epsilon \to 0$, and x_T approaches a neighborhood of x^*. The size of this neighborhood is determined by the level of noise $\|e\|^2$, and on the constant $\delta_{\mathcal{M}}$. The constant C that appears in the bound (and also D that appears in the running time) is of moderate size.

C. Approximate Projections

It is often not possible to perform exact projections when dealing with manifolds. Rather, a projection may be computed via a separate numerical optimization procedure that minimizes the distance between the given point and the manifold. This computation may yeild only an approximately optimal solution. In this section we show that MIP is robust with respect to inexact and approximate projections.

Definition 2: Let $y \in \mathbb{R}^m$. We say that a point $z \in \mathcal{M}$ is an ϵ_P-approximate projection if

$$\|z - y\|^2 \leq \|\mathcal{P}_{\mathcal{M}}(y) - y\|^2 + \epsilon_P. \quad (2)$$

We denote an ϵ_P-approximate projection of $y \in \mathbb{R}^n$ onto $\mathcal{M}$ by $\mathcal{P}_{\mathcal{M}}^{\epsilon_P}(y)$.

Remark Note that for a manifold, neither a projection, nor an ϵ_P-projection need be unique. The notation $\mathcal{P}_{\mathcal{M}}(y)$ and $\mathcal{P}_{\mathcal{M}}^{\epsilon_P}(y)$ may thus not be uniquely defined, but are simply meant to refer to any points which are actual/approximate minimizers of the Euclidean distance from y to $\mathcal{M}$.

Let a measurement operator A with measurements $y \in \mathbb{R}^m$ such that $y = Ax^*$. Consider the following algorithm, which we call Manifold Iterative Projection-Approximate (MIP$_{approx}$).

Algorithm: MIP$_{approx}$

0. Initialization: $z = 0$, $x_0 = \mathcal{P}_{\mathcal{M}}^{\epsilon_P}(z)$.

While $k < T$, repeat:
1. $z_{k+1} = x_k - \eta\left(A^T(Ax_k - y_k)\right)$
2. $x_{k+1} = \mathcal{P}_{\mathcal{M}}^{\epsilon_P}(z_{k+1})$
3. Increment k.

Theorem 4: Let $x^* \in \mathcal{M}$ be an unknown signal on the manifold. Let A be a measurement operator satisfying (1) with $\delta_{\mathcal{M}} < \frac{1}{3}$, and let $y = Ax^*$. Let $\eta = 1/(1+\delta_{\mathcal{M}})$. Let $\epsilon_P{}' = \frac{(1+\delta_{\mathcal{M}})(1-\delta_{\mathcal{M}})}{1-3\delta_{\mathcal{M}}}\epsilon_P$ and $\epsilon > \epsilon_P{}'$.Then the algorithm MIP$_{approx}$ produces $x_T \in \mathcal{M}$ such that $\|x_T - x^*\|^2 \leq \frac{\epsilon}{1-\delta_{\mathcal{M}}}$ and $\|Ax_T - y\|^2 \leq \epsilon$ in

$$T = \left\lceil \frac{1}{\log\left(\frac{(1-\delta_{\mathcal{M}})}{2\delta_{\mathcal{M}}}\right)} \log\left(\frac{\|y\|^2}{2(\epsilon - \epsilon_P{}')}\right) \right\rceil$$

iterations.

Remark

Notice that if $\epsilon_P = 0$ we recover Theorem 2, as one would expect when it is possible to compute projections exactly.

The above theorem says that MIP is robust with respect to inexact computations of the projection. If ϵ_P-approximate projections are available then the algorithm converges to an $C_1\epsilon_P$-sized neighborhood of the optimal solution, where $C_1 = \frac{(1+\delta_{\mathcal{M}})(1-\delta_{\mathcal{M}})}{1-3\delta_{\mathcal{M}}}$ is a moderate-sized constant.

D. Tracking Slowly Time-Varying Signals

In this section we describe a variant of MIP that adapts to an unknown time-varying signal $x^*[t]$ on the manifold $\mathcal{M}$. The signal is assumed to be slowly time-varying with respect to the computational resources available, i.e. between successive iterations of the MIP algorithm, the signal is allowed to have changed by only a small amount. We consider a discrete time setting where at each time instant, a set of measurements of the form $y[t] = Ax^*[t]$ becomes available. The algorithm is allowed to perform a single iteration per discrete epoch of time. The objective is to eventually track the signal $x^*[t]$ by using the measurements $y[t]$.

More formally let $\mathbb{T} = \mathbb{Z}_{\geq 0}$ denote the temporal index set. We consider time-varying signals denoted by $x[t] \in \mathbb{R}^n$ for all $t \in \mathbb{T}$.

Definition 3: We say that a signal $x[t]$ is slowly time-varying if:

$$\|x[t+1] - x[t]\|^2 \leq \epsilon_\tau \text{ for all } t \tag{3}$$

for some constant ϵ_τ.

Given a measurement operator A and measurements $y[t]$ $\forall t \in \mathbb{T}$ such that $y[t] = Ax^*[t]$. Consider the following algorithm, which we call Temporal Manifold Iterative Projection (TMIP).

Algorithm: TMIP

0. Initialization: $z = 0$, $x[0] = \mathcal{P}_{\mathcal{M}}(z)$.

At time $t+1$:
1. $z[t+1] = x[t] - \eta\left(A^T(Ax[t] - y[t])\right)$
2. $x[t+1] = \mathcal{P}_{\mathcal{M}}(z[t+1])$
3. Increment t.

The following theorem states that the algorithm TMIP successfully tracks the signal $x^*[t]$ from the measurement process $y[t]$. (Recall that at each time $t \in \mathbb{T}$, $y[t] \in \mathbb{R}^m$ has small dimension as compared to the ambient dimension of the true signal $x^*[t]$.)

Theorem 5: Let $x^*[t] \in \mathcal{M}$ for all $t \in \mathbb{T}$ be an unknown time-varying signal on the manifold satisfying (3). Let A be a measurement operator satisfying (1) with $\delta_{\mathcal{M}} < \frac{1}{5}$, and let $y[t] = Ax^*[t]$. Let $\eta = 1/(1+\delta_{\mathcal{M}})$. Then the algorithm TMIP converges to an $\epsilon_\tau{}' = \frac{(1-\delta_{\mathcal{M}})(1+\delta_{\mathcal{M}})}{1-5\delta_{\mathcal{M}}}\epsilon_\tau$ neighborhood of $x[t]$ as $T \to \infty$, i.e. for every $\delta > 0$ there exists a T such that $\|x[t] - x^*[t]\|^2 \leq \epsilon_\tau{}' + \delta$ for all $t \geq T$. Moreover, if $\epsilon > \epsilon_\tau{}'$ the algorithm produces $x_T \in \mathcal{M}$ such that $\|x[T] - x^*[T]\|^2 \leq \frac{\epsilon}{1-\delta_{\mathcal{M}}}$ and $\|A(x[T] - x^*[T])\|^2 \leq \epsilon$

in

$$T = \left\lceil \frac{1}{\log\left(\frac{(1-\delta_{\mathcal{M}})}{4\delta_{\mathcal{M}}}\right)} \log\left(\frac{\|y[0]\|^2}{2(\epsilon - \epsilon_\tau{}')}\right) \right\rceil$$

iterations.

Note that nothing about the dynamics of the signal is known beyond the fact that it is slowly time-varying in the sense of (3). In the absence of knowledge of the dynamics, one cannot hope to estimate $x^*[t]$ closer than ϵ_τ in general. Even if one were to have perfect knowledge of $x^*[t]$ for some t, the state $x^*[t+1]$ could be anywhere in an ϵ_τ neighborhood. Hence, one cannot hope to estimate $x^*[t]$ beyond a resolution of ϵ_τ. Theorem 5 says that TMIP approximates the state to a resolution of $C_2\epsilon_\tau$ for some moderate-sized constant $C_2 = \frac{(1-\delta_{\mathcal{M}})(1+\delta_{\mathcal{M}})}{5\delta_{\mathcal{M}}}$.

Remark The results presented in this section consisted of the convergence of the MIP algorithm. We also presented its robustness to noisy measurements, approximate projections, and time-varying nature of the underlying signal. We remark that it is possible to combine all these forms of robustness in a natural way into a single model and obtain an analogous convergence result. While it would lead to more cumbersome notation, only mild modifications to the analysis presented here would be required.

Remark We remark that for the theorems presented in this section, manifold structure is not essential. The results presented here are true for signal recovery over arbitrary sets $V \subseteq \mathbb{R}^n$ via a linear measurement operator A provided A approximately-isometrically embeds V in the sense of (1). Signals with sparse, low-rank and manifold structures are just specific signal classes endowed with this property.

IV. Stylized Applications

In this section, we describe some stylized examples of the recovery problem on manifolds. As pointed out earlier, there are two interesting aspects related to recovery.

1) *The sample complexity:* The number of measurements required for reliable recovery (i.e. with high probability) is precisely the minimal m required to stably embed the manifold with distortion $\delta_{\mathcal{M}} < K$ for some constant K. This constant depends on whether or not the measurements are noise-free and whether the projections are exact or approximate. (For example, in the static, noiseless case with perfect projections, $K = \frac{1}{3}$ as mentioned in Theorem 2). As mentioned earlier, the sample complexity to achieve this constant is roughly $O(k \log(C_{\mathcal{M}} n))$. Computing or bounding $C_{\mathcal{M}}$ precisely involves computing these geometric properties of the specific manifold in question, which may be a challenging task in its own right. We do not dwell on this issue. Rather, we note that for this manifolds considered in this section, we expect $C_{\mathcal{M}}$ to be fairly benign, as is evidenced also by numerical examples.
2) *Algorithmic aspects:* The computational procedure proposed in this paper, namely MIP and its variants, may be viewed as projected gradient schemes. As with all such schemes, there are three main components:
   a) Computing the gradient of the objective function: As noted previously, the recovery problem may be viewed as minimizing the quadratic function $\|Ax - y\|^2$ over the manifold $\mathcal{M}$ (we are guaranteed that the optimal value is 0). In our case the objective function is quadratic, so the gradient computation is straightforward, it is simply $A^T(Ax - y)$.
   b) Step-size: The appropriate step size is characterized via the parameter $\delta_{\mathcal{M}}$; the step size is simply $\frac{1}{1+\delta_{\mathcal{M}}}$.
   c) Projection onto $\mathcal{M}$: This is the most challenging step. In essence the recovery problem on $\mathcal{M}$ reduces to being able to compute projections onto $\mathcal{M}$. In the stylized examples discussed below, we describe specific computational procedures to project $x \in \mathbb{R}^n$ onto $\mathcal{M}$.

A. Manifolds defined by Polynomials

Let $y \in \mathcal{B} \subseteq \mathbb{R}^k$ where $\mathcal{B}$ is the unit ball and $p_i(y)$ for $i = 1, \ldots n$ be multivariate polynomials of degree d. Then the set

$$\mathcal{M} = \{(p_1(y), \ldots, p_n(y)) \in \mathbb{R}^n | \; y \in \mathcal{B}\}$$

defines a k-dimensional compact Riemannian manifold in $\mathbb{R}^n$. Given an unknown $x^* \in \mathcal{M}$ we would like to recover it from a few random measurements. As explained above, a tractable way to compute projections on to $\mathcal{M}$ is the essential algorithmic ingredient.

Note that given a point $z \in \mathbb{R}^n$, computing its projection onto $\mathcal{M}$ is equivalent to solving the following optimization problem:

$$\begin{aligned} &\text{Min.}_{\ \gamma}\ \gamma \\ &\text{subject to: } \sum_{i=1}^{n} (p_i(y) - z_i)^2 - \gamma \geq 0 \ \text{ for all } y \in \mathcal{B}. \end{aligned}$$

The above polynomial optimization problem produces a γ whose value is equal to the squared projection distance between z and $\mathcal{M}$. If one is able to obtain the exact solution to γ, the corresponding minimizer y is typically obtained from the dual. However, in general this polynomial optimization problem is hard to solve. There are natural relaxation schemes which allow one to obtain bounds on γ and also produce approximate minimizers. The well-known approaches replace the non-negativity constraint by a sum-of-squares constraint [9], [8]. The resulting sum-of-squares optimization problem may be solved by semidefinite programming and would yield an approximate projection. Theorem 4 is especially useful in this setting where exact projections are hard but approximate projections are feasible to compute.

B. The Stiefel Manifold

An interesting matrix manifold that occurs in engineering is called the Stiefel manifold. Recall that the Stiefel manifold $\mathcal{S}(n,k)$ is given by:

$$\mathcal{S}(n,k) = \left\{ U \in \mathbb{R}^{n \times k} |\ U^T U = I \right\}.$$

Note that in the special case when $k = n$, one obtains the *orthogonal group* $\mathcal{O}(n)$. As noted above, the main computational challenge associated with recovery is computation of projections. Let $W \in \mathbb{R}^{n \times k}$ have a rectangular singular value decomposition $W = U_W \Sigma_W V_W^T$ (so that Σ_W is of size $k \times k$). Then it is straightforward to show that

$$\mathcal{P}_{\mathcal{S}(n,k)} = U_W V_W^T.$$

The Stiefel manifold is a manifold of dimension $nk - \frac{1}{2}k(k+1)$, in a nk-dimensional real space. For fixed k, the dimensions of the manifold and that of the ambient dimension are of the same order of magnitude hence one cannot expect reliable recovery from a very small number of measurements. It may be interesting to consider sub-manifolds of this manifold whose dimensions are vanishingly small as a fraction of the ambient dimension.

An interesting example in this setting would be to consider a sub-manifold of $\mathcal{O}(n)$ whose ambient dimension is a vanishing fraction of the ambient dimension.

We consider the manifold $\mathcal{M} = \bigoplus_{i=1}^{N} \mathcal{O}(k_i)$ given by

$$\mathcal{M} = \left\{ \text{diag}(U_i) |\ U_i^T U_i = I,\ U_i \in \mathbb{R}^{n \times n} \right\}.$$

Note that this is a $\sum_{i=1}^{N} \binom{k_i+1}{2}$-dimensional manifold in a $\left(\sum_{i=1}^{N} k_i\right)^2$-dimensional real space. If we let $M = n$ and $k_i = n$ for all i, then we have a manifold of dimension $O(n^3)$ in a $O(n^4)$-dimensional real space. A sample complexity of $O(n^3 \log(C_{\mathcal{M}} n))$ suffices to reconstruct the unknown orthogonal matrix. Note that the projection step here is again computationally tractable. Given a matrix A of appropriate dimension, it is sufficient to project the i^{th} diagonal block to an orthonormal one and zero the off-diagonal terms.

C. The Grassmannian

Subspace identification is a problem that finds numerous applications in engineering, a prominent example being medical imaging [1]. In a high dimensional setting, it may be desirable to learn an unknown subspace U from a small number of samples. In a similar spirit, the problem of *tracking* an unknown subspace finds numerous applications [2]. These problems of identification and tracking may be viewed in a natural way as the problem of learning an unknown signal on a suitable manifold. The appropriate manifold in question may be viewed abstractly as consisting of all subspaces of $\mathbb{R}^n$, called the *Grassmannian manifold*.

Recall that the Grassmannian manifold $\mathcal{G}(n,k)$ is defined to be the set of all k-dimensional subspaces of $\mathbb{R}^n$. It is a compact Riemannian manifold, and can be equipped with a natural metric. A natural parametrization of the manifold may be obtained via projection maps. Given a subspace $T \subseteq \mathbb{R}^n$, let P_T denote the Euclidean projection map (represented as a $n \times n$ matrix) onto T. Note that projection matrices may be expressed as $P_T = U_T U_T^T$, and the subspace $T \in \mathcal{G}(n,k)$ may be identified with $\text{Im}(U_T)$. Hence, one can identify each subspace with its corresponding projection matrix. This gives the following natural description of the Grassmannian:

$$\mathcal{G}(n,k) = \left\{ P \in \mathbb{R}^{n \times n} |\ P = P^T,\ P^2 = P,\ \text{rank}(P) = k \right\}.$$

In this way the Grassmannian may be viewed as a matrix manifold of dimension $k(n-k)$ in an n^2 dimensional real space. The sample complexity of recovery is $O(k(n-k)\log(C_{\mathcal{M}}n))$. If k is treated as a fixed constant then the sample complexity is $O(n\log(C_{\mathcal{M}}n))$.

To compute projections onto the Grassmannian we note the following. Given a matrix $R \in \mathbb{R}^{n\times n}$ with eigenvalue decomposition $U_R\Sigma_R U_R^T$, its projection on to the Grassmannian is given by

$$\mathcal{P}_{\mathcal{G}(n,k)}(R) = U_R^{(k)} U_R^{(k)^T},$$

where $U_R^{(k)}$ is the matrix of eigenvectors whose columns correspond to the k largest positive eigenvalues (if less than k eigenvalues are positive, then only the columns corresponding to the positive eigenvalues are kept). Note that if R is not symmetric, one may compute its projection onto the Grassmannian by simply projecting its symmetric part $\frac{R+R^T}{2}$.

V. Numerical Experiments

We study the behavior of the algorithm TMIP on a numerical example involving subspace tracking. Consider the dynamics $U(t) = U_0\exp(\mathcal{R}t)$, where $U_0 \in \mathbb{R}^{n\times k}$ is a given matrix on the Stiefel manifold and $\mathbb{R}^{k\times k}$ is a fixed skew-symmetric matrix. Note that $U(t)$ may be viewed as a time-varying subspace. We consider the problem of tracking $P(t) = U(t)U(t)^T$ which essentially corresponds to the subspace tracking problem.

Since $\mathcal{R}$ (a skew-symmetric matrix) is in the Lie algebra of the Lie group of orthogonal matrices, it is a standard fact that $U(t)$ is also on the Stiefel manifold for all $t \in \mathbb{R}_{\geq 0}$. It is easy to check that $P(t) = U(t)U(t)^T$ represents the orthogonal projection onto the column span of $U(t)$. We consider the signal $P(t)$ sampled at over $t \in [0,1]$ at discrete intervals of $\Delta t = 10^{-4}$ to obtain the discrete time signal $P[t]$. The signal $P[t]$ thus constructed, constitutes the unknown signal. We let $A : \mathbb{R}^{n^2} \to \mathbb{R}^m$ be a matrix whose entries are distributed as $A_{ij} \sim \mathcal{N}(0, \frac{1}{m})$. The map A acts on $P(t)$ via $y(t) = A\text{vec}(R(t))$, where $\text{vec}(\cdot)$ denotes the standard vectorization of a matrix. U_0 and $\mathcal{R}$ are chosen to be random orthogonal and skew-symmetric matrices respectively. The appropriate discrete time signal $y[t]$ constitute the measurements.

We run the algorithm on the task of tracking a k-dimensinional subspace of $\mathbb{R}^n$ using $m = C_0kn$ random measurements at each time for some constant C_0. The typical behavior of the algorithm (tracking error versus time) is shown in Fig. 1. The tracking error is expressed in terms of the Euclidean (Frobenius) norm between the true and estimated projection matrices. In Fig. 2 we report the probability of success as a function of the number of measurements over a range of numerical experiments.

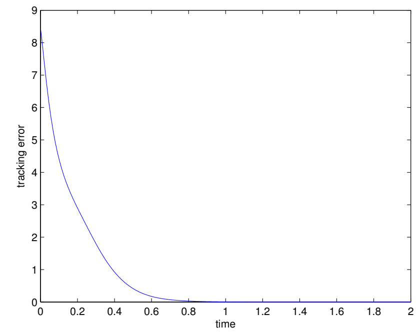

Fig. 1. Tracking error of a signal on the Grassmannian $\mathcal{G}(30,5)$ from $m = 2kn = 300$ measurements.

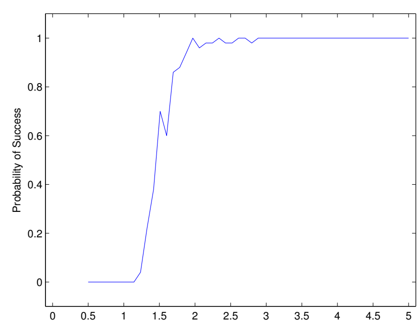

Fig. 2. The probability of success in tracking the signal as a function of the parameter $\frac{m}{kn}$ for $\mathcal{G}(20,5)$.

VI. Proofs

In this section we present the proof of Theorem 5. Due to space constraints we omit the proofs of the other theorems, but note that their proofs are very similar in spirit.

We define $\psi_t(z) := \frac{1}{2}\|Az - y[t]\|^2$.

Lemma 1: Let $\{x[t]\}_{t\in\mathbb{T}}$ be a sequence of iterates produced by the algorithm TMIP. Under the assumptions of Theorem 5,

$$\psi_t(x[t+1]) \leq \frac{2\delta_{\mathcal{M}}}{1-\delta_{\mathcal{M}}}\psi_t(x[t]).$$

Proof: Note that since $\psi_t(z)$ is a quadratic function,

$$\psi_t(z) - \psi_t(x[t]) = \langle \bigtriangledown \psi(x[t]), z - x[t] \rangle + \frac{1}{2} \|A(z - x[t])\|^2$$
$$\leq \langle A^T(Ax[t] - y[t]), z - x[t] \rangle + \frac{1}{2}(1 + \delta_{\mathcal{M}}) \|z - x[t]\|^2.$$

Define $f_t(z) := \langle A^T(Ax[t] - y[t]), z - x[t] \rangle + \frac{1}{2}(1 + \delta_{\mathcal{M}}) \|z - x[t]\|^2$. By completion of squares,

$$f_t(z) = \frac{1}{2}(1 + \delta_{\mathcal{M}}) \|z - w[t+1]]\|^2$$
$$- \frac{1}{2(1 + \delta_{\mathcal{M}})} \|A^T(Ax[t] - y[t])\|^2,$$

where $w[t+1] = x[t] - \frac{1}{1+\delta_{\mathcal{M}}} A^T(Ax[t] - y[t])$. Note that

$$\text{arg min}_{z \in \mathcal{M}} f_t(z) = \mathcal{P}_{\mathcal{M}}(w[t+1]) = x[t+1].$$

Now note that

$$\begin{aligned} f_t(x[t+1]) &\leq f_t(x^*[t]) \\ &= \langle A^T(Ax_k - y), x^*[t] - x[t] \rangle + \\ &\quad \frac{1}{2}(1 + \delta_{\mathcal{M}}) \|x^*[t] - x[t]\|^2 \\ &\leq \langle A^T(Ax[t] - y[t]), x^*[t] - x[t] \rangle + \\ &\quad \frac{1}{2} \frac{1 + \delta_{\mathcal{M}}}{1 - \delta_{\mathcal{M}}} \|A(x^*[t] - x[t])\|^2 \\ &= \psi_t(x^*[t]) - \psi_t(x[t]) + \\ &\quad \frac{\delta_{\mathcal{M}}}{1 - \delta_{\mathcal{M}}} \|A(x^*[t] - x[t])\|^2. \end{aligned}$$

Hence

$$\psi_t(x[t+1]) - \psi_t(x[t]) \leq f_t(x[t+1]) \leq f_t(x^*[t])$$
$$\leq \psi_t(x^*[t]) - \psi_t(x[t]) + \frac{\delta_{\mathcal{M}}}{1 - \delta_{\mathcal{M}}} \|A(x^*[t] - x[t])\|^2.$$

■

Lemma 2: Let $\{x[t]\}_{t \in \mathbb{T}}$ be a sequence of iterates produced by the algorithm TMIP. Under the assumptions of Theorem 5,

$$\psi_{t+1}(x[t+1]) \leq \frac{4\delta_{\mathcal{M}}}{1 - \delta_{\mathcal{M}}} \psi_t(x[t]) + (1 + \delta_{\mathcal{M}}) \epsilon_\tau.$$

Proof:

$$\begin{aligned} \psi_{t+1}(x[t+1]) &= \frac{1}{2} \|y[t+1] - Ax[t+1]\|^2 \\ &\leq \frac{1}{2} \|Ax[t+1] - y[t]\|^2 + \|y[t] - y[t+1]\|^2 \\ &= 2\psi_t(x[t+1]) + \|A(x^*[t+1] - x^*[t])\|^2 \\ &\leq 2\psi_t(x[t+1]) + (1 + \delta_{\mathcal{M}}) \|x^*[t+1] - x^*[t]\|^2 \\ &\leq \frac{4\delta_{\mathcal{M}}}{1 - \delta_{\mathcal{M}}} \psi_t(x[t]) + (1 + \delta_{\mathcal{M}}) \epsilon_\tau. \end{aligned}$$

■

Proof: [Proof of Theorem 5] We use Lemma 2 and the fact that $\psi_t(x^*[t]) = 0$ for all t to note that,

$$\begin{aligned} \psi_T(x[T]) &\leq \frac{4\delta_{\mathcal{M}}}{1 - \delta_{\mathcal{M}}} \psi_{T-1}(x[T-1]) + \frac{1}{2}(1 + \delta_{\mathcal{M}}) \epsilon_\tau \\ &\leq \left(\frac{4\delta_{\mathcal{M}}}{1 - \delta_{\mathcal{M}}} \right)^T \frac{\|y\|^2}{2} + \frac{(1 + \delta_{\mathcal{M}})(1 - \delta_{\mathcal{M}})}{1 - 5\delta_{\mathcal{M}}} \epsilon_\tau. \end{aligned}$$

The choice of $T = \left\lceil \frac{1}{\log\left(\frac{(1-\delta_{\mathcal{M}})}{4dm}\right)} \log\left(\frac{\|y\|^2}{2(\epsilon - \epsilon_\tau{}')}\right) \right\rceil$ ensures that $2\psi_T(x[T]) \leq \epsilon$ for $\epsilon > \epsilon_\tau{}'$. Furthermore, by (1), $(1 - \delta_{\mathcal{M}}) \|x[T] - x^*[T]\|^2 \leq \|A(x[T] - x^*[T])\|^2 \leq \epsilon$. ■

References

[1] B. Ardekani, J. Kershaw, K. Kashikura, and I. Kanno. Activation detection in functional mri using subspace modeling and maximum likelihood estimation. *IEEE Transactions on Medical Imaging*, 18(2), 1999.

[2] L. Balzano, R. Nowak, and B. Recht. Online identification and tracking of subspaces from highly incomplete information. In *Proceedings of the 48th Annual Allerton Conference*, 2010.

[3] R. G. Baraniuk and M. B. Wakin. Random projections of smooth manifolds. *Foundations of Computational Mathematics*, 9(1):51–77, 2009.

[4] E.J. Candes and T. Tao. Decoding by linear programming. *Information Theory, IEEE Transactions on*, 51(12):4203 – 4215, 2005.

[5] R. Garg and R. Khandekar. Gradient descent with sparsification: An iterative algorithm for sparse recovery with restricted isometry property. In *Proceedings of the 26^{th} International Conference on Machine Learning (ICML)*, 2009.

[6] C. Hegde, M. Wakin, and R. Baraniuk. Random projections for manifold learning. In *Neural Information and Processing Systems*, 2007.

[7] P. Jain, R. Meka, and I. Dhillon. Guaranteed rank minimization via singular value projection. In *Neural Information and Processing Systems*, 2010.

[8] Jean B. Lasserre. Global optimization with polynomials and the problem of moments. *SIAM Journal on Optimization*, 11(3):796–817, 2001.

[9] P. A. Parrilo. *Structured Semidefinite Programs and Semialgebraic Geometry Methods in Robustness and Optimization*. PhD thesis, California Institute of Technology, 2000.

[10] B. Recht, M. Fazel, and P. A. Parrilo. Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization. Technical report, 2007.

[11] Bruce Reznick. Some concrete aspects of Hilbert's 17th problem. In *In Contemporary Mathematics*, pages 251–272. American Mathematical Society, 1996.