2-14-2020

Restricted Isometric Projections for Differentiable Manifolds and Applications

Vasile Pop
University of South Florida

Restricted Isometric Projections for Differentiable Manifolds and Applications

by

Vasile Pop

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy in Mathematics
Department of Mathematics & Statistics
College of Arts and Sciences
University of South Florida

Major Professor: Razvan Teodorescu, Ph.D.
Iuliana Teodorescu, Ph.D.
Sherwin Kouchekian, Ph.D.
Mohamed Elhamdadi, Ph.D.

Date of Approval:
March 06, 2020

Keywords: Sparsity, Varieties, Low Dimensional Models, Random Matrices,
Covariant Exterior Differential

Dedication

To my wife Maria and my son Adelin

Acknowledgments

First and foremost I would like to express my gratitude to my advisor Dr. Razvan Teodorescu for his extensive support and guidance in my studies and research through my entire graduate program at the University of South Florida. His direction, patience, encouragement and immense knowledge made this dissertation and the research that lead to it possible.

A special thanks to Dr. Iuliana Teodorescu for her massive help, many valuable insights and contribution to this thesis.

I would also like to express my sincere appreciation to Dr. Sherwin Kouchekian and Dr. Mohamed Elhamdadi for been part of my defense committee but also for their knowledge, profession and engaging coursework I took throughout my graduate studies.

I extend my thanks to Dr. Andrei Barbos for agreeing to chair my defense. In addition to my committee, I would like to thank the rest of the faculty and staff of the Department of Mathematics and Statistics at USF for their support throughout my studies.

Most importantly, I am deeply grateful to my amazing wife Maria and my wonderful son Adelin for theirs unending love, support, and encouragement.

Table of Contents

List of Tables

Abstract

The restricted isometry property (RIP) is at the center of important developments in compressive sensing. In $\mathbb{R}^N$, RIP establishes the success of sparse recovery via basis pursuit for measurement matrices with small restricted isometry constants $\delta_{2s} < 1/3$. A weaker condition, $\delta_{2s} < 0.6246$, is actually sufficient to guarantee stable and robust recovery of all s-sparse vectors via l_1-minimization. In infinite Hilbert spaces, a random linear map satisfies a general RIP with high probability and allow recovering and extending many known compressive sampling results. This thesis extends the known restricted isometric projection of sparse datasets of vectors Σ embedded in the Euclidean spaces $\mathbb{R}^N$ down into low-dimensional subspaces $\mathbb{R}^m, m \ll N$, to Riemannian manifolds $(\mathcal{M}, g)$, of manifold dimension m, with Riemannian metric g equivalent to the induced metric from the embedding space $\mathbb{R}^N$. This will establish a higher-dimensional version of the Fisher-Kolmogorov test for comparing populations in usual statistical analysis, allowing to develop an inference procedure analogous to Generalized Linear Models in the usual case ($\#\Sigma \gg N$).

Chapter 1
Introduction

The digital revolution is driving nowadays the development and deployment of new, more sophisticated, robust and flexible sensing processing systems with ever increasing resolution. The theoretical foundation of this revolution is the Shannon/Nyquist sampling theorem, which states that the sampling rate of a continuous-time signal must be at least twice its highest frequency in order to ensure reconstruction. By this results images, signals, videos, and other data can be exactly recovered from Nyquist rate set of uniformly spaced samples. Unfortunately, in many applications, the resulting Nyquist rate is so high that samples must be compressed in order to store or transmit them. Despite extraordinary advances in computational power, the acquisition and processing of signals in applications may simply be too costly [25], or even physically impossible, and continues to pose a tremendous challenge. Compressive sensing (CS) provides an alternative to the Shannon/Nyquist sampling theorem when the signal under acquisition is known to be sparse or compressible [1][2]. Sparsity allows for dramatically "under-sampled" signals to be captured and manipulated using a very small amount of data. This is the fundamental idea behind compressive sensing.

While this idea has only recently gained significant attention in the signal processing community, there has been work done in this direction. In 1795, Prony proposed an algorithm for the estimation of parameters associated with a small number of complex exponentials sampled in the presence of noise [26]. In the early 1900s, Carathéodory showed that a positive linear combination of any k sinusoids is uniquely determined by its value at $t = 0$ and at any other $2k$ points in time [27][28], which represents far fewer samples than the number of Nyquist-rate samples when k is small and the range of possible frequencies is large. More recent, in the early 2000s, Blu, Marziliano, and Vetterli [29] showed that signals that are governed by only k parameters can be sampled and recovered from just $2k$ samples.

Building on these results and on the work of Candès, Romberg, Tao [3][5] and Donoho [30], who showed that a signal having a sparse representation can be recovered exactly from a small set of linear, nonadaptive compressive measurements, CS has emerged as a new framework for signal acquisition and sensor design. To mention just a few, areas that have benefited immediately from this basic discovery are image recovery and compression algorithms. While the research field of CS draws from a variety of other areas such as random matrix theory, approximation theory, numerical linear algebra, Banach space theory, and convex optimization, the combined efforts of mathematicians, computer scientists, and engineers have led to many application such as medical imaging, radar, microscopy and civilian and military surveillance.

Chapter 2

Introduction to Compressed Sensing

In mathematical terms, a classical sparse recovery problem aims to recover a vector $\mathbf{x} \in \mathbb{R}^{\mathbf{N}}$ from linear and underdetermined measurements $\mathbf{Ax} = \mathbf{y}$, where $\mathbf{A} \in \mathbb{C}^{\mathbf{m}\times\mathbf{N}}$ models the linear measurement (information) process with N being much larger than m. Looking closer, the standard compressive sensing problem essentially identifies two questions which are not entirely independent. Which matrices $\mathbf{A} \in \mathbb{C}^{\mathbf{m}\times\mathbf{N}}$ are suitable and what are efficient reconstruction algorithms which recover the original signal $\mathbf{x}$ from measurements $\mathbf{y}$?

Even though there are several tractable strategies to solve the standard compressive sensing problem, in this chapter we focus on the basis pursuit (also called l_1-minimization) strategy which consists of solving a convex optimization problem, more precisely, to find a minimizer for

$$\min_{\mathbf{z}\in\mathbb{R}^{\mathbf{N}}} \|\mathbf{z}\|_{\mathbf{1}} \text{ subject to } \mathbf{Az} = \mathbf{y} \text{ where } \|\cdot\|_1 \text{ is } l_1 - \text{ norm} \quad (P_1)$$

2.1 Sparsity and Compressibility

Sparsity has been exploited in statistics and learning theory as a method for avoiding overfitting [14] and figures prominently in the theory of statistical estimation and model selection [15]. The notions of *sparsity* and *compressibility* are at the core of compressive sensing.

Definition 2.1.1. A vector $\mathbf{x} \in \mathbb{C}^{\mathbf{N}}$ is called s - sparse if it has at most s nonzero entries, i.e., if

$$\|\mathbf{x}\|_{\mathbf{0}} := card(\{j : x_j \neq 0, j = 1 \cdots, N\}) \leq s$$

In applications, sparsity can be a hard constraint to impose, therefore we encounter vectors that are not exactly s - sparse but compressible in the sense that they are well approximated by sparse ones. Informally, a vector $\mathbf{x} \in \mathbb{C}^{\mathbf{N}}$ is *compressible* if the error of its best s-term approximation $\sigma_s(\mathbf{x})_p$ decays quickly in s.

Definition 2.1.2. The l_p-error of best s-term approximation to a vector $\mathbf{x} \in \mathbb{C}^N$ is defined by

$$\sigma_s(\mathbf{x})_p := \inf\{\|\mathbf{x} - \mathbf{z}\|_p, \mathbf{z} \in \mathbb{C}^N \text{ is } s\text{-sparse}\} \text{ where } p > 0$$

Note that sparsity is a highly nonlinear model: given a pair of s-sparse signals, a linear combination of two signals will in general no longer be s-spare. Sparsity is only a model and may not be the best fit for all applications, therefore extensions of sparsity are: block sparsity [19], join sparsity [18] and tree sparsity [20].

2.2 Design Sensing Matrices

Compressive sensing is not fitted for arbitrary matrices. It is still an open problem to construct explicit matrices which are provably optimal, but a breakthrough is achieved by resorting to random matrices. Examples of random matrices are *Gaussian* matrices whose columns consist of independent random variables following a standard normal distribution and *Bernoulli* matrices whose columns are independent random variables taking the values $+1$ and -1 with equal probability. We now list a number of desirable conditions that a sensing matrix $\mathbf{A}$ should have to guaranty recovery of sparse vectors.

Definition 2.2.1. A matrix $\mathbf{A} \in \mathbb{K}^{\mathbf{m}\times\mathbf{N}}$ (where $\mathbb{K}$ is $\mathbb{R}$ or $\mathbb{C}$) is said to satisfy the *null space property* relative to a set $S \subset [N]$ if

$$\|\mathbf{v_S}\|_\mathbf{1} < \|\mathbf{v}_{\overline{\mathbf{S}}}\|_1 \text{ for all } \mathbf{v} \in Ker\mathbf{A} - \{0\}$$

where $\mathbf{v}_S$ is the restriction of $\mathbf{v}$ on the indices in S. A matrix $\mathbf{A} \in \mathbb{K}^{\mathbf{m}\times\mathbf{N}}$ is said to satisfy the *null space property of order* s if it satisfies the null space property relative to any set $S \subset [N]$ with $card(S) \leq s$.

Slightly strengthened versions of the null space property are needed to reconstruct scheme with respect to sparsity defect (stable null space property) or a scheme affected by error (robust null space property).

Definition 2.2.2. A matrix $\mathbf{A} \in \mathbb{C}^{m\times N}$ is said to satisfy the *stable null space property* with constant $0 < \rho < 1$ relative to a set $S \subset [N]$ if

$$\|\mathbf{v}_S\|_1 \leq \rho\|\mathbf{v}_{\overline{S}}\|_1 \text{ for all } \mathbf{v} \in Ker\mathbf{A}$$

A matrix $\mathbf{A} \in \mathbb{C}^{m\times N}$ is said to satisfy the *stable null property of order* s with constant $0 < \rho < 1$ if it satisfies the stable null space property with constant $0 < \rho < 1$ relative to any set $S \subset [N]$ with $card(S) \leq s$.

Definition 2.2.3. The matrix $\mathbf{A} \in \mathbb{C}^{m\times N}$ is said to satisfy the *robust null space property* (with respect to $\|\cdot\|$) with constants $0 < \rho < 1$ and $\tau > 0$ relative to a set $S \subset [N]$ if

$$\|\mathbf{v}_S\|_1 \leq \rho\|\mathbf{v}_{\overline{S}}\|_1 + \tau\|\mathbf{A}\mathbf{v}\| \text{ for all } \mathbf{v} \in \mathbb{C}^N.$$

A It is said to satisfy the *robust null space property of order* s with constants $0 < \rho < 1$ and $\tau > 0$ if it satisfies the robust null space property with constants ρ, τ relative to any set $S \subset [N]$ with $card(S) \leq s$.

It has been proved in [6] (Chapter 4) that *null space property*, *stable null space property* and *robust null space property* are necessary and sufficient conditions for extract recovery of sparse vectors via basis pursuit program (P_1).

Since the null space property is not easily verifiable by direct computation, *coherence* is a much simple concept and preferable to use to assess the quality of a measure matrix [16]. In general, the smaller the coherence, the better the performance of sparse recovery algorithms.

Definition 2.2.4. Let $\mathbf{A} \in \mathbb{C}^{m\times N}$ be a matrix with l_2 normalized columns $\mathbf{a}_1, \cdots, \mathbf{a}_N$, i.e., $\|\mathbf{a}_i\|_2 = 1$ for all $i \in [N]$. The *coherence* $\mu = \mu(A)$ of the matrix $\mathbf{A}$ is defined as

$$\mu := \max_{1\leq i\neq j\leq N} |\langle a_i, a_j\rangle|$$

A general concept of l_1 - *coherence function* is defined, which incorporates the usual coherence as the particular value $s = 1$ of its argument.

Definition 2.2.5. Let $\mathbf{A} \in \mathbb{C}^{m\times N}$ be a matrix with l_2 - normalized columns $\mathbf{a}_1, \cdots, \mathbf{a}_N$. The l_1 - *coherence function* μ_1 of the matrix $\mathbf{A}$ is defined for $s \in [N-1]$ by

$$\mu_1(s) := \max_{i\in[N]} \max \Big\{ \sum_{j\in S} |\langle \mathbf{a}_i, \mathbf{a}_j\rangle|, S \subset [N], card(S) = s, i \notin S \Big\}$$

Remark 2.2.6. Note that coherence and the l_1 - coherence function are invariant under multiplication on the left by unitary matrix $\mathbf{U}$ i.e columns of $\mathbf{UA}$ are l_2 - normalized vectors $\mathbf{Ua}_1, \cdots, \mathbf{Ua}_N$ and $\langle \mathbf{Ua}_i, \mathbf{Ua}_j\rangle = \langle \mathbf{a}_i, \mathbf{a}_j\rangle$. Also using the Cauchy-Schwarz inequality $|\langle \mathbf{a}_i, \mathbf{a}_j\rangle| \leq \|\mathbf{a}_i\|_2 \cdot \|\mathbf{a}_j\|_2$ it is clear that the coherence matrix is bounded from above, $\mu \leq 1$.

What about lower bounds for the coherence and l_1 - coherence function of a matrix $\mathbf{A} \in \mathbb{C}^{m\times N}$ with $m < N$? What are examples of matrices with an almost minimal coherence? A matrix which achieves the coherence lower bound is called *equiangular tight frame* and the coherence lower bound is known as the *Welch bound*.

Definition 2.2.7. A system of l_2 - normalized vectors $(\mathbf{a}_1, \cdots, \mathbf{a}_N)$ in $\mathbb{K}^m$ is called *equiangular* if there is a constant $c \geq 0$ such that

$$|\langle \mathbf{a}_i, \mathbf{a}_j\rangle| = c \text{ for all } i, j \in [N], i \neq j$$

Definition 2.2.8. A system of vectors $(\mathbf{a}_1, \cdots, \mathbf{a}_N)$ in $\mathbb{K}^m$ is called a *tight frame* if there exists a constant $\lambda > 0$ such that one of the following equivalent conditions holds:

(a) $\|\mathbf{x}\|_2^2 = \lambda \sum_{j=1}^{N} |\langle \mathbf{x}, \mathbf{a}_j\rangle|^2$ for all $\mathbf{x} \in \mathbb{K}^m$

(b) $\mathbf{x} = \lambda \sum_{j=1}^{N} \langle \mathbf{x}, \mathbf{a}_j\rangle \mathbf{a}_j$ for all $\mathbf{x} \in \mathbb{K}^m$

(c) $\mathbf{AA}^* = \frac{1}{\lambda}\mathbf{Id}_m$ where $\mathbf{A}$ is the matrix with columns $\mathbf{a}_1, \cdots, \mathbf{a}_N$.

It is now possible to prove [6] that coherence of a matrix is always in the range $\mu(\mathbf{A}) \in \left[\sqrt{\frac{N-m}{m(N-1)}}, 1\right]$

Theorem 2.2.9. *The coherence of a matrix $\mathbf{A} \in \mathbb{K}^{m\times N}$ with l_2 - normalized columns satisfies*

$$\mu \geq \sqrt{\frac{N-m}{m(N-1)}} \tag{2.1}$$

Equality holds if and only if the columns $\mathbf{a}_1, \cdots, \mathbf{a}_N$ *of the matrix* $\mathbf{A}$ *form an equiangular tight frame.*

The Welch bound can be extended to the l_1 - coherence function for small values of its argument.

Theorem 2.2.10. *The* l_1 *- coherence function of a matrix* $\mathbf{A} \in \mathbb{K}^{m \times N}$ *with* l_2 *normalized columns satisfies*

$$\mu_1(s) \geq s\sqrt{\frac{N-m}{m(N-1)}} \text{ whenever } s < \sqrt{N-1} \tag{2.2}$$

Equality holds if and only if the columns $\mathbf{a}_1, \cdots, \mathbf{a}_N$ *of the matrix* $\mathbf{A}$ *form an equiangular tight frame.*

Remark 2.2.11. Note that when $N \gg m$ the lower bound is approximately $\mu(\mathbf{A}) \geq 1/\sqrt{m}$.

We claimed that the performance of sparse recovery algorithms is enhanced by a small coherence. The next Theorem [6] guarantees the exact recovery of every s-sparse vector via basis pursuit when the measurement matrix has a coherence $\mu < 1/(2s-1)$.

Theorem 2.2.12. *Let* $\mathbf{A} \in \mathbb{C}^{m \times N}$ *be a matrix with* l_2*-normalized columns. If*

$$\mu_1(s) + \mu_1(s-1) < 1,$$

then every s*-sparse vector* $\mathbf{x} \in \mathbb{C}^N$ *is exactly recovered from the measurement vector* $\mathbf{y} = \mathbf{A}\mathbf{x}$ *via basis pursuit.*

Even though null space property is both necessary and sufficient condition to guarantee recovery of sparse vectors, when the measurements are contaminated with noise it will be useful to consider somewhat stronger conditions. Also, the lower bound on the coherence in Theorem 2.2.10 limits recovery algorithms to rather small sparsity levels. To overcome these limitations, in [17], Candès and Tao introduced the Restricted Isometry Property (RIP) on matrices $\mathbf{A}$, also known as the uniform uncertainty principle, and established its important role in compressed sensing.

Definition 2.2.13. The $s-$th restricted isometry constant $\delta_s = \delta_s(\mathbf{A})$ of a matrix $\mathbf{A} \in \mathbb{C}^{m \times N}$ is a positive real number $\delta \geq 0$ such that

$$(1-\delta)\|\mathbf{x}\|_2^2 \leq \|\mathbf{A}\mathbf{x}\|_2^2 \leq (1+\delta)\|\mathbf{x}\|_2^2 \tag{2.3}$$

for all s-sparse vectors $\mathbf{x} \in \mathbb{C}^N$. We say that $\mathbf{A}$ satisfies the *restricted isometry property* if δ_s is small for reasonably large s.

Just like for coherence, small restricted isometry constants are desired. In case $N \geq Cm$ it has been proven in [6] that the restricted isometry constant must satisfy $\delta_s \geq c\sqrt{s/m}$, which is reminiscent of the Welch bound $\mu \geq c'/\sqrt{m}$, for $s = 2$.

The success of sparse recovery via basis pursuit for measurement matrices with small restricted isometry constants is guaranteed by the condition $\delta_{2s} < 1/3$. Weakening this condition to $\delta_{2s} < 0.6246$ is actually sufficient to guarantee stable and robust recovery of all s-sparse vectors via l_1-minimization.

Theorem 2.2.14. *Suppose that the $2s-$th restricted isometry constant of the matrix $\mathbf{A} \in \mathbb{C}^{m\times N}$ satisfies*

$$\delta_{2s} < \frac{1}{3}.$$

Then every s-sparse vector $\mathbf{x} \in \mathbb{C}^{m\times N}$ is the unique solution of

$$\min_{\mathbf{z}\in\mathbb{C}^N} \|\mathbf{z}\|_1 \text{ subject to } \mathbf{A}\mathbf{z} = \mathbf{A}\mathbf{x}.$$

Theorem 2.2.15. *Suppose that the $2s-$th restricted isometry constant of the matrix $\mathbf{A} \in \mathbb{C}^{m\times N}$ satisfies*

$$\delta_{2s} < \frac{4}{\sqrt{41}} \simeq 0.6246.$$

Then, for any $\mathbf{x} \in \mathbb{C}^N$ and $\mathbf{y} \in \mathbb{C}^m$ with $\|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2 \leq \eta$, a solution $\mathrm{x}^\#$ of

$$\min_{\mathbf{x}\in\mathbb{C}^N} \|\mathbf{z}\| \text{ subject to } \|\mathbf{A}\mathbf{z} - \mathbf{y}\|_2 \leq \eta$$

approximates the vector $\mathbf{x}$ with errors

$$\|\mathbf{x} - \mathbf{x}^\#\|_1 \leq C\sigma_s(\mathbf{x})_1 + D\sqrt{s}\eta,$$

$$\|\mathbf{x} - \mathbf{x}^\#\|_2 \leq \frac{C}{\sqrt{s}}\sigma_s(\mathbf{x})_1 + D\eta,$$

where the constants $C, D > 0$ depend only on σ_{2s}.

2.3 Measurements bounds

Research in CS has also focused on reducing the number of measurements m (as a function of N and s). We can now consider how many measurements m are necessary to achieve the RIP. Ignoring the impact of δ_{2s} and taking in consideration only the dimensions of the problem (N, m, and s) led to a simple lower bound for m [7].

Theorem 2.3.1. *Let* $\mathbf{A}$ *be an* $m \times N$ *matrix that satisfies the RIP of order* $2s$ *with constant* $\delta_{2s} \in \left(0, \frac{1}{2}\right]$. *Then*

$$m \geq Cs \log\left(\frac{N}{s}\right),$$

where $C = 1/2 \log(\sqrt{24} + 1) \simeq 0.28$.

2.4 Sensing matrix constructions

Now that we have defined the relevant properties of a matrix A in the context of CS, we turn to the question of how to construct matrices that satisfy these properties. We have already seen that, in general, an equiangular tight frame achieves the coherence lower bound [21]. Similarly, there are known matrices of size $m \times m^2$ that achieve Welch bound (the coherence lower bound) $\mu(\mathbf{A}) = 1/\sqrt{m}$, such as the Gabor frame generated from the Alltop sequence [22]. It was proved that is possible to deterministically construct matrices of size $m \times N$ that satisfy the RIP of order s, but in real world settings these constructions would lead to an unacceptably large requirement on m [23][24]. Fortunately, these limitations can be overcome by randomizing the matrix construction. Random matrices will satisfy the RIP with high probability if the entries are chosen according to a Gaussian, Bernoulli, or more generally any sub-Gaussian distribution [6]. Theorem 5.65 in [7] states that if a matrix $\mathbf{A}$ is chosen according to a sub-Gaussian distribution with $m = O(s \log(N/s)/\delta_{2s}^2)$ then $\mathbf{A}$ will satisfy the RIP of order $2s$ with probability at least $1 - 2exp(-c_1 \delta 2 s^2 m)$. The most significant benefits of using random matrices is met in practice where we are often more interested in recovering sparse signals with respect to some basis Φ, thus we require the product $A\Phi$ to satisfy the RIP. When $\mathbf{A}$ is chosen randomly we do not have to explicitly take Φ into account.

As noted earlier, CS sparse recovery of vectors has so far been predominantly treated in discrete settings in some finite-dimensional space (usually $\mathbb{R}^N$ or $\mathbb{C}^N$) whose nonzero entries with respect to a particular basis are few in comparison to N. Although the standard CS framework assumes that $\mathbf{x}$ is a finite-length vector with a discrete-valued index (such as time or space), in practice we will often be interested in designing measurement systems for acquiring continuously indexed signals such as continuous-time signals or images. It is sometimes possible to extend this model to continuously indexed signals using an intermediate discrete representation.

Most real world signals are analog, or continuous time, and thus are modeled more precisely in infinite dimensional function spaces. Any finite dimensional model may not be suited to such problems. For example, much of finite compressing sensing theory revolves around Gaussian and other random matrices, which have immense value. However, in real applications we do not always have the option to choose $\mathbf{A}$ as we want. Instead, physical properties of the sensing process, as well as constraints to its practical implementability determine the sensing matrix. Thus, the typical sensing matrix is neither Gaussian nor Bernoulli, but rather a matrix with a very specific structure. Also, the standard sparsity has to be extended to include a much richer class of signals and to encode broader data models, including continuous time signals.

Although this issue has been quite widely recognized [7][8][10], few attempts were made so far to extend the existing finite compressing sensing theory to infinite dimensional models. Admirable progress has been done on how to derive compressive sensing theory for a variety of structured sensing matrices that arise in applications [9]. Extending CS to infinite-dimensional spaces is necessary in order to properly apply it in an analog setting [33], or to explore connections with the sampling of signals with finite rate of innovation (signals that have a finite number of degrees of freedom per unit of time) [34]. CS in infinite-dimensional spaces is also needed in machine learning for the development of efficient methods to compute information in a concise way, to feed learning algorithms that work on densities, to extract features or to uncover underlying structures [35][36].

In the next two chapters we discuss two theories of compressed sensing in high-dimensional (or infinite-dimensional) spaces from a few linear measurements. Chapter 3 presents a frame-

work based on a discretization procedure using a ”Low-Dimensional Model” [31], and Chapter 4 presents a general theory of compressed sensing for analog signals, formulated as an infinite dimensional optimization problem [32].

Chapter 3
Compressed Sensing in Hilbert Spaces

A general view of signal recovery problems in classical finite or infinite-dimensional spaces is to consider observations $\mathbf{y}$ modeled as

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{e} \tag{3.1}$$

where $\mathbf{x} \in \mathcal{H}, \mathbf{y} \in \mathcal{F}$ and $\mathcal{H}, \mathcal{F}$ are Hilbert spaces of finite or infinite dimensions, $\mathbf{A}$ is a linear map which models a finite number of measurements m and ϵ is a noise such that $\|e\|_{\mathcal{F}}$ is bounded. If no further prior information on the signal is available, recovering $\mathbf{x}$ from $\mathbf{y}$ is generally not possible. To overcome of the ill-posed nature of such problems the authors in [31] use a "low complexity" hypothesis on $\mathbf{x}$ (either $\mathbf{x}$ belongs to a "low-dimensional model set" Σ or can be described by few parameters). The objective is to address the following question: given a model set Σ and a regularizer f, what RIP condition on $\mathbf{A}$ is sufficient to ensure that solving the optimization problem

$$x^* \in argmin_{\mathbf{z}\in\mathcal{H}} f(\mathbf{z}) s.t. \|\mathbf{A}\mathbf{z} - (\mathbf{A}\mathbf{x} + \mathbf{e})\|_{\mathcal{F}} \leq \epsilon \tag{3.2}$$

yields an optimal decoder? An instance-optimal decoder provide exact recovery of vectors in $\mathbf{x} \in \Sigma$ in the noisless setting, but also stable and robust when noise or error are present.

3.1 Restricted Isometry Property (RIP)

We have seen that in classical CS many efficient instance-optimal algorithms relies on properties of the sensing matrix such as the restricted isometry property. One can extend the notion of RIP of a linear operator $\mathbf{A}$ on a model set $\Sigma - \Sigma := \{\mathbf{x} - \mathbf{x}' : \mathbf{x}, \mathbf{x}' \in \Sigma\}$ which is called *secant set* of Σ.

Definition 3.1.1. (RIP) The linear operator $\mathbf{A} : \mathcal{H} \to \mathcal{F}$ satisfies the RIP on the secant set $\Sigma - \Sigma$ with constant δ if for all $\mathbf{x} \in \Sigma - \Sigma$,

$$(1 - \delta)\|\mathbf{x}\|_{\mathcal{H}}^2 \leq \|\mathbf{A}\mathbf{x}\|_{\mathcal{F}}^2 \leq (1 + \delta)\|\mathbf{x}\|_{\mathcal{H}}^2$$

where $\|\cdot\|_{\mathcal{F}}$ and $\|\cdot\|_{\mathcal{H}}$ are Euclidean norms of $\mathcal{F}$ and $\mathcal{H}$.

3.2 Low-dimensional models

Given a set $\Sigma \subset \mathcal{H}$ the normalized secant set of Σ is defined as

$$\mathcal{S} = \mathcal{S}(\Sigma) := \left\{ \mathbf{z} = \frac{\mathbf{y}}{\|\mathbf{y}\|_{\mathcal{H}}} : \mathbf{y} \in (\Sigma - \Sigma) - \{0\} \right\}$$

Intrisic dimension of $\mathcal{S}$ is measured using the upper box-counting dimension which is linked to the notion of covering number. A comprehensive list of definitions of dimensions is presented in [38].

Remark 3.2.1. There are examples of sets for which no linear embedding to a finite dimensional space is possible even though their dimension is finite, therefore it is not possible to construct a linear map which satisfies the RIP.

Definition 3.2.2. (Covering number). Let $\alpha > 0$ and $S \subset \mathcal{H}$. The covering number $N(\mathcal{S},\ \alpha)$ of $\mathcal{S}$ is the minimum number of closed balls (with respect to the norm $\|\cdot\|_{\mathcal{H}}$) of radius α, with centers in S, needed to cover $\mathcal{S}$.

Definition 3.2.3. (Upper box-counting dimension). The upper box-counting dimension of $\mathcal{S}$ is

$$boxdim(\mathcal{S}) := \limsup_{\alpha \to 0} \log[N(\mathcal{S},\ \alpha)] / \log[1/\alpha].$$

A model Σ is *low-dimensional* if $boxdim(\mathcal{S}(\Sigma))$ is small compared to the ambient dimension of the Hilbert space $\mathcal{H}$ (which may be infinite).

Already known dimensions [6]: for k-sparse vectors, $boxdim(\mathcal{S}(\Sigma)) = k$ and for $n \times n$ matrices of rank lower than r, $boxdim(\mathcal{S}(\Sigma)) = rn$.

3.3 Dimension reduction with random linear operators

Now that the notion of dimension of a model Σ and the RIP property of operator $\mathbf{A}$ has been defined, based on work in [38] the authors in [31] present a generic construction of a dimension-reducing linear operator $\mathbf{A} : \mathcal{H} \to \mathbb{R}^m$ that satisfies the RIP on $\Sigma - \Sigma$. Intuitively, the operator $\mathbf{A}$

is built in two steps: first project the signal in Σ onto a subspace $H \subset \mathcal{H}$ of finite (but potentially large) dimension space then reduce the dimension by a multiplication with a random matrix (see Table 1).

Table 1 Strategy for Dimension Reduction.

Low Dimension		**Finite Dimension**		**Low Dimension**
$\Sigma \subset \mathcal{H}$	Projection $\rightarrow$	Linear Space $H \subset \mathcal{H}$	Reduction $\rightarrow$	Linear Space $= \mathbb{R}^m$
$boxdim(\mathcal{S}(\Sigma)) < \infty$		$dim(H) = d < \infty$		$m \ll d$

Projection on a Finite-Dimensional Subspace Step

Given $\mathcal{S} \subset \mathcal{H}$ such that $boxdim(\mathcal{S}) < \infty$ and $0 < \alpha < 1$, there always exists a finite-dimensional subspace $H \subset \mathcal{H}$ that approximates all vectors in $\mathcal{S}$ with precision α i.e.

$$(1-\alpha)\|\mathbf{x}\|_{\mathcal{H}} \leq \|P_H \mathbf{x}\|_{\mathcal{H}} \leq \|\mathbf{x}\|_{\mathcal{H}}$$

for all $\mathbf{x} \in \Sigma - \Sigma$ where P_H denotes the orthogonal projection onto H [38].

In general H can be construct as following. First construct an α - cover of the normalized secant set $\mathcal{S}$. Let $\mathcal{C}_\alpha$ be the set of center of balls used to cover $\mathcal{S}$. It is now sufficient to take $H_\alpha \subset \mathcal{H}$ be the finite-dimensional linear subspace of $\mathcal{H}$ spanned by $\mathcal{C}_\alpha$ and let $P_{H_\alpha} : \mathcal{H} \rightarrow \mathcal{H}$ be the orthogonal projection onto H_α [38].

Dimension Reduction Step

The goal now is to reduce down the dimension of subspace H to $O(boxdim(\mathcal{S}))$ by using random matrices. Denote by d the dimension of H and let $(e_1, \cdots, e_d)$ be an arbitrary orthonormal basis of H. Identify the projection onto H with the linear operator $P_H : \mathcal{H} \rightarrow \mathbb{R}^d$ that returns the coordinates of the orthogonal projection onto H in the basis $(e_1, \cdots, e_d)$. Next compose P_H with a random matrix $M \in \mathbb{R}^{m \times d}$ to build $\mathbf{A} : \mathcal{H} \rightarrow \mathbb{R}^m$ such that $\mathbf{A} = MP_H$ and $\mathbf{A}$ satisfies RIP property and $m \simeq O(boxdim(\mathcal{S}))$. In order to exhibit operator $\mathbf{A}$ satisfying RIP property with constant δ it will be enough to build a random $M : \mathbb{R}^d \rightarrow \mathbb{R}^m$ satisfying a RIP with small constant δ'. The approach to construct M described in [39] is to independent draw of m identically

distributed random vectors $a_i \in \mathcal{H}$ so that for $\mathbf{x} \in \mathcal{H}$, $M\mathbf{x} := (\langle a_i, \mathbf{x}\rangle)_{i=1}^m$). A convenient way to prove that M satisfy RIP is to choose its probability distribution [6] so that for any vector $\mathbf{x} \in \mathcal{H}$, $\mathbb{E}_M \|M\mathbf{x}\|_2^2 = \mathbf{x}\|_{\mathcal{H}}^2$ and require M to satisfy two concentration inequalities below.

There exists two constants $c_1, c_2 \in (0,\ \infty]$ such that for any fixed $\mathbf{y}, \mathbf{z} \in S(\Sigma) \cup \{0\}$,

$$\mathbb{P}_M\{|h_M(\mathbf{y}) - h_M(\mathbf{z})|\ \geq \lambda \|\mathbf{y} - \mathbf{z}\|_{\mathcal{H}}\} \leq 2e^{-c_1 m\lambda^2} \text{ for } 0 \leq \lambda \leq c_2/c_1 \tag{3.3}$$

$$\mathbb{P}_M\{|h_M(\mathbf{y}) - h_M(\mathbf{z})|\ \geq \lambda \|\mathbf{y} - \mathbf{z}\|_{\mathcal{H}}\} \leq 2e^{-c_2 m\lambda},\ \text{for } \lambda \geq c_2/c_1 \tag{3.4}$$

where

$$h_M : \mathcal{H} \to \mathbb{R}$$

$$x \mapsto\ \|M\mathbf{x}\|_2^2 - \|\mathbf{x}\|_{\mathcal{H}}^2.$$

Assuming that the random operator M satisfies above concentration inequality and the set $\mathcal{S}$ has finite box-counting dimension, the next theorem states that reducing the dimension of vectors in Σ is possible and a number of measurements m of the order of the dimension of the secant set $\mathcal{S}$ is sufficient to be able to recover elements of Σ, from the ambient space $\mathcal{H}$:

Theorem 3.3.1. *Let $M : \mathcal{H} \to \mathbb{R}^m$ be a random linear map that satisfies (3.3) and (3.4). Assume that boxdim (S) $< s$ (there exists $0 < \alpha_{\mathcal{S}} < \frac{1}{2}$ such that $N(S,\ \alpha) \leq \alpha^{-s}$ for all $0 < \alpha < \alpha_{\mathcal{S}}$). Then for any $\xi, \delta_0 \in (0, 1)$, M satisfies the RIP on $\Sigma - \Sigma$ it with constant $\delta \leq \delta_0$ with probability at least $1 - \xi$ provided that*

$$m \geq\ \frac{1}{\delta_0^2} \frac{C}{\min(c_1, c_2)}\ \max\left\{ s \log\left(\frac{1}{\alpha_{\mathcal{S}}}\right),\ \log\left(\frac{6}{\xi}\right)\right\}, \tag{3.5}$$

where $C > 0$ is an absolute constant (independent on the dimension parameters).

3.4 Recovery of Low-Dimensional Models - Unions of Subspaces

Since the linear operators $\mathbf{A}$ built in the previous section preserve low complexity models Σ, we turn to the study of the reconstruction of vectors from Σ using general decoders, with an emphasis on a particular class of convex functions: "atomic norms" [39]. Considering a set $\mathcal{A} \subset \mathcal{H}$, called the set of atoms, the corresponding atomic "norm" is built using the convex hull of $\mathcal{A}$.

Definition 3.4.1. The *convex hull* of a set $\mathcal{A}$ is:

$$conv(\mathcal{A}) := \Big\{ \mathbf{x} = \sum c_i a_i \ : \ a_i \ \in \mathcal{A}, \ c_i \ \in \mathbb{R}_+, \ \sum c_i = 1 \Big\} \tag{3.6}$$

Definition 3.4.2. The *atomic norm* induced by the set $\mathcal{A}$ is defined as:

$$\|x\|_{\mathcal{A}} \ := \inf\{t \in \mathbb{R}_+ \ : \ x \in t \cdot \overline{\text{conv}}(\mathcal{A})\} \tag{3.7}$$

where $\overline{\text{conv}}(\mathcal{A})$ is the closure of conv ($\mathcal{A}$) in $\mathcal{H}$.

This function $\|x\|_{\mathcal{A}}$ is a convex gauge but is not always a norm. This norm is finite only on the set $\mathcal{E}(\mathcal{A}) := \mathbb{R}_+ \cdot \overline{conv}(\mathcal{A}) = \{\mathbf{x} = t \cdot \mathbf{y}, \ t \in \mathbb{R}_+, \ \mathbf{y} \in \overline{conv}(\mathcal{A})\} \subset \mathcal{H}$. It can be extended to $\mathcal{H}$ by setting $\|x\|_{\mathcal{A}} := +\infty$ if $x \notin \mathcal{E}(\mathcal{A})$.

As a running example, a classical refinement of the notion of sparsity called structured sparsity in levels is considered. For model $\Sigma = \Sigma_1 \cdots \Sigma_J$ associated to structured sparsity in levels a similar class of atomic norms called the group norms in levels is defined.

Structured sparsity in levels

Structured sparsity also called group sparsity is the assumption that signals are not only sparse but also supported on a few groups of significant coefficients in a transformed domain [40][41][42]. Medical imaging (MRI) and simultaneous signal and noise sparse modeling are two examples were this model is useful [41]. Formally, consider $(e_i)_{i \in \mathbb{N}}$ an orthonormal basis of Hilbert space $\mathcal{H}$, G a finite collection of non-overlapping finite groups of indexes i.e.

$$G := \{g \subset \mathbb{N} \ \mid \ |g| < \infty, g \cap g^{'} = \emptyset \text{ whenever } g \neq g^{'}\}$$

The restriction of a vector $\mathbf{x} \in \mathcal{H}$ to a group g is

$$\mathbf{x}_g := \sum_{i \in g} \langle \mathbf{x}, e_i \rangle e_i$$

A group support is a subset $T \subset G$ and the restriction of $\mathbf{x}$ to a group support is

$$\mathbf{x}_T := \sum_{g \in T} \mathbf{x}_g$$

The group support of $\mathbf{x} \in \mathcal{H}$ denoted $gsupp(\mathbf{x})$ is the smallest $T \subset G$ such that $\mathbf{x}_T = x$ and the size of the group support of $\mathbf{x}$ denoted $|gsupp(\mathbf{x})|$ is the cardinality of $gsupp(\mathbf{x})$. Given an integer k, the k-group-sparse model is defined as

$$\Sigma_k \ := \{\mathbf{x} \in \mathcal{H},\ |gsupp(\mathbf{x})| \ \leq k\} \tag{3.8}$$

Let d be the size of the biggest group. We have the following covering of $\mathcal{S}(\Sigma_k)$:

$$N(\mathcal{S}(\Sigma_k),\ \alpha) \ \leq \ \left(\frac{c}{\alpha}\right)^{dk} \text{ where } C \text{ is a constant depending on } d. \tag{3.9}$$

Let $\mathcal{H}_j \subset \mathcal{H}, j \in J$ be a collection or orthogonal spaces each equipped with a k_j-group sparse model Σ_j as defined in (3.8) each with its Hilbert basis and its set G_j of groups. Then the *structured sparsity in levels* is associated to the model

$$\Sigma := \left\{\mathbf{x} \in \mathcal{H}, \mathbf{x} = \sum_{j=1}^{J} x_j, x_j \in \Sigma_{k_j}\right\}, \tag{3.10}$$

i.e

$$\Sigma = \Sigma_{k_1} \times \Sigma_{k_2} \times \cdots \times \Sigma_{k_j},$$

and

$$N(\mathcal{S}, \alpha) \leq N(\mathcal{S}(\Sigma_{k_1}, \alpha)) \cdot \ldots \cdot N(\mathcal{S}(\Sigma_{k_j}, \alpha)) \leq \left(\frac{C_1}{\alpha}\right)^{d_1 k_1} \cdot \ldots \cdot \left(\frac{C_j}{\alpha}\right)^{d_j k_j} \tag{3.11}$$

where C_j are constants that are of the order of the dimension of each level times the maximum size of groups d_j in level j.

Atomic norms for group norms in levels

Given $\mathcal{H}_j$ associated to the j-th level, let $\mathcal{S}_j(1) \subset \mathcal{H}_j$ its unit sphere, G_j its set of groups, and $\Sigma_{1,j}$ the associated 1-group sparse model, consider $\mathcal{A}_j := \Sigma_{1,j} \cap \mathcal{S}_j(1)$ the collection of atoms of the j-th level. The corresponding atomic norm is associated to the finite-dimensional space

$$\mathcal{E}(\mathcal{A}_j) = span(\{e_j\}_{i \in \cup_{g \in G_j}})$$

and is defined as

$$\|x\|_{A_j} = \begin{cases} \sum_{g\in G} \|x_g\|_{\mathcal{H}}, & x \in \mathcal{E}(\mathcal{A}) \ ; \\ +\infty, & x \notin \mathcal{E}(\mathcal{A}) \end{cases} \tag{3.12}$$

A natural regularizer for the *structured sparsity in levels* model is defined as follows in $\mathcal{H}_1 \times \ldots \times \mathcal{H}_J$:

$$f_w : (x_l, \cdots x_J) \mapsto w_1\|x_1\|_{\mathcal{A}_1} + \ldots + w_J\|x_J\|_{\mathcal{A}_J} \text{ where } w_j > 0 \tag{3.13}$$

Atomic norm associated to a union of subspace models

Given any union of subspaces $\Sigma \subset \mathcal{H}$, the norm associated to its normalized atoms $\mathcal{A}(\Sigma) := \Sigma \cap \mathcal{S}(1)$ is defied as

$$\|\cdot\|_\Sigma := \|\cdot\|_{\Sigma\cap\mathcal{S}(1)} \tag{3.14}$$

and is sometimes useful as a regularizer to perform recovery (i.e. by choosing $f(z) = \|z\|_\Sigma$ in minimization (3.2)).

3.4.1 Exact recovery of unions of subspaces

Exact recovery of unions of subspaces [45] states that the stability of any decoder of the form (3.2) is guaranteed if $\mathbf{A}$ satisfies a RIP on set $\Sigma - \Sigma$ with a constant $\delta < \delta_\Sigma(f)$ holds, where $\delta_\Sigma(f)$ is a constant that depends only on the regularizer f and the model set Σ (See Appendix A). This recovery is a generalization of the sharp RIP result by Cai and Zhang [44], whose RIP theorem states that $\delta < \sqrt{1/2}$ on $\Sigma - \Sigma$ implies stable recovery of sparse vectors with the l_1-norm, and of low rank matrices with the nuclear norm.

3.4.2 Stable recovery of unions of subspaces

A stable recovery formulation for minimization (3.2) is also established in [45] which is stated below.

Theorem 3.4.3. *Assume that Σ is a union of subspaces. Then, for any continuous linear operator $\mathbf{A}$ on $\mathcal{H}$ that satisfies the RIP on the secant set $\Sigma - \Sigma$ with constant $\delta < \delta_\Sigma(f)$ we have: for all*

$\mathbf{x} \in \Sigma, e \in \mathcal{F}$ *such that* $\|e\|_{\mathcal{F}} \leq \epsilon$*, with* $\mathbf{x}^*$ *the result of minimization (3.2),*

$$\|\mathbf{x}^* - \mathbf{x}\|_{\mathcal{H}} \ \leq C_\Sigma(f,\ \delta) \cdot (\|e\|_{\mathcal{F}} + \epsilon), \tag{3.15}$$

where $C_\Sigma \delta < +\infty$.

The constant $C_\Sigma \delta < +\infty$ is explicit in most classical examples [45].

3.4.3 Robust recovery of unions of subspaces

Regarding robustness to modeling error, it is more convenient to use a norm independent of the measurement operator $\mathbf{A}$. This norm was introduced in [46] as an intermediate tool to measure the distance from a vector $\mathbf{x}$ to the model set Σ. Given a constant C, the A-norm is defined by

$$\|\cdot\|_{A,C} := C \cdot \|A \cdot\|_{\mathcal{F}} + \|\cdot\|_{\mathcal{H}}$$

(here the A refers to the measurement operator, not to be confused with the atomic norm). Following robustness result has been proved by authors in [32] using symmetrized distance with respect to a regularizer f

$$d_f(\mathbf{x}, \Sigma) = \inf_{\tilde{\mathbf{x}} \in \Sigma} \frac{f(\mathbf{x} - \tilde{\mathbf{x}}) - f(\tilde{\mathbf{x}} - \mathbf{x})}{2}.$$

Theorem 3.4.4. *Let* Σ *be union of subspaces. Let* f *be positively homogeneous, non-negative and convex with* $f(\mathbf{x}) < +\infty$ *for* $\mathbf{x} \in \Sigma$*. Consider a continuous linear operator* $\mathbf{A}$ *satisfying the RIP on* $\Sigma - \Sigma$ *with constant* $\delta < \delta_\Sigma(f)$*, and a noise level* $\|e\|_{\mathcal{F}} \leq \epsilon$*. Denote* C_Σ *the constant from Theorem 3.4.3, and assume that for all* $\mathbf{u} \in \mathcal{H}, \|\mathbf{u}\|_{A,C_\Sigma} \leq C_{f,A,\Sigma} \cdot f(\mathbf{u})$ *for some* $C_{f,A,\Sigma} < \infty$*. Then, for all* $\mathbf{x} \in \mathcal{H}, e \in \mathcal{F}$*, such that* $\|e\|_{\mathcal{H}} \leq \eta \leq \epsilon$*, any minimizer* $\mathbf{x}^*$ *of (3.2) satisfies*

$$\|\mathbf{x}^* - \mathbf{x}\|_{\mathcal{H}} \ \leq C_\Sigma \cdot (\|e\|_{\mathcal{F}} + \epsilon) + 2C_{f,A,\Sigma} \cdot d_f(\mathbf{x},\ \Sigma) \tag{3.16}$$

Chapter 4

Compressed Sensing for Analog Signals

In contrary to framework presented in Chapter 3, which adopted a discretization procedure using a low dimensional model, in this chapter the approach of infinite compressed sensing framework presented in [32] analyze an infinite dimensional optimization problem directly. Authors aimed for a true analog setting where the following requirements can be identified: the mathematical model consists of reproducing kernel Hilbert spaces, the index set of general measurements is equipped with a more general metric structure, robust and stable recovery does not impose discretization, and analog sparsity is not linked to a specific discrete set but rather it allows for sparse representations within an uncountably infinite dictionary.

The main idea of this framework is to model signals in reproducing kernel Hilbert spaces as ambient space with the kernel functions being the elements of the representation system. Sparsity of a signal is defined as the minimal number of terms in the kernel functions expansions. Recovery based on total-variation minimization problem (TV -minimization problem - an extension of classical atomic measures) is robust (signal is only approximately sparse) and stable (measurement is corrupted by noise). Two applications of this theory presented in [32] are: recovery of sparse bandlimited functions and functions that have a sparse short-time Fourier transform.

4.1 Reproducing Kernel Hilbert Spaces

Definition 4.1.1. (Reproducing kernel) Let $\mathcal{H}$ be a Hilbert space of $\mathbb{R}$ valued functions defined on a non-empty set X. A function $K : X \times X \to \mathbb{R}$ (K_x) is called a *reproducing kernel* of $\mathcal{H}$ if it satisfies

- $\forall x \in X, K(\cdot, x) \in \mathcal{H}$

- $\forall x \in X, \forall f \in \mathcal{H}, \langle f, K(\cdot, x)\rangle_{\mathcal{H}} = f(x)$ (the *reproducing property*)

In particular, for any $x, y \in X$, $K(x, y) = \langle K(\cdot, x), K(\cdot, y)\rangle_{\mathcal{H}}$

Definition 4.1.2. A Hilber space $\mathcal{H}$ is called Reproducing Kernel Hilbert Space (RKHS) if $\mathcal{H}$ has a reproducing kernel. We say that $\mathcal{H}$ has a unit-norm reproducing kernel, if $K(x, x) = 1$ for all $x \in X$.

Remark 4.1.3. :

- The inner product of $\mathcal{H}$ induces the usual norm and the topology with respect to which $\mathcal{H}$ is complete.

- Since K_x produce the point evaluation of any $f \in \mathcal{H}$ at x via the inner product, it can be deduce that the span of $\{K_x\}_{x \in X}$ is dense in $\mathcal{H}$

- If kernel K is not unit-norm, it can be normalized

$$\tilde{K}(x, y) = \frac{K(x, y)}{\sqrt{K(x, x)K(y, y)}}$$

and

$$\tilde{i} : \mathcal{H} \to \tilde{\mathcal{H}}, \tilde{i}f(x) = \sqrt{K(x, x)}f(x), x \in X$$

is a space isomorphism, where $\tilde{\mathcal{H}}$ is a Hilbert space with unit-norm reproducing kernel $\tilde{K}$.

The weak greedy algorithm described in [48] provides a norm-convergent expansion for each $\tilde{f} \in \tilde{\mathcal{H}}$ in term of at most countable number of kernel function and implicit for each $f \in \mathcal{H}$. It will be convenient to allow also expansions of the form

$$K * \mu \equiv \int_X k_x d\mu(x)$$

where μ is an element of $\mathcal{M}(X)$, the space of regular complex Borel measures of bounded total variation on a locally compact Hausdorff space X.

4.2 Atomic Norms on Reproducing Kernel Hilbert Spaces

The atomic norm was introduced in [39] and possesses a number of favorable properties that are useful for recovering models from limited linear measurements. An extended list of atomic norms obtained by convexifying atomic sets that are of interest in various applications are presented in [39].

Consider all finite disjoint collections $\{E_i\}_{i=1}^n$ of Borel measurable subset of X such that $X = \bigcup_{i=1}^{n} E_i$, then the space $\mathcal{M}(X)$ is equipped with the total-variation norm

$$\|\mu\|_{TV} \equiv |\mu|(X) = \sup \sum_{i=1}^{n} |\mu(E_i)|$$

By the Riesz-Markov-Kakutani theorem, the space of continuous functions vanishing at infinity $\mathcal{C}_0(X)$ is the dual of the normed space $\mathcal{M}(X)$.

Definition 4.2.1. Let X be a locally compact Hausdorff space and $\mathcal{M}(X)$ the space of regular complex Borel measures with bounded total variation. Let $\mathcal{H}$ be a separate reproducing kernel Hilbert space of Borel measurable functions on X with a unit-norm reproducing kernel K. The atomic norm of $\mathcal{H}$ associates with each $f \in \mathcal{H}$ the value

$$\|f\|_A = \inf\{\|\mu\|_{TV} : \mu \in \mathcal{M}(X), K * \mu = f\} \quad (\mathcal{P})$$

Original atomic norm (3.7) from [39] has a more geometric nature but does not appear to be discussed in the context of Hilbert spaces with reproducing kernel. Although this norm [39] and atomic norm ($\mathcal{P}$) are to an extend equivalent [32] and its dual is more commonly known and used.

Proposition 4.2.2. *Given $f \in \mathcal{H}$, then the dual of the atomic norm is*

$$\|f\|_A^* = \sup_{x \in X} |f(x)| = \|f\|_\infty.$$

4.3 Sparse Recovery by Atomic Norm Minimization

Definition 4.3.1. A function $f \in \mathcal{H}$ is called s-sparse if there exist sequences $\{c_n\}_{n=1}^s$ in $\mathbb{C}^s$ and $\{x_n\}_{n=1}^s$ in X^s such that $f = \sum_{n=1}^{s} c_n K_{x_n}$ and for any other expansion $f = \sum_{n=1}^{s'} c_n' K_{s_n'}$ of f we have $s \leq s'$.

To ensure that linear combination of s kernel functions gave an s-sparse function linear, dependencies of the kernel functions need to be ruled out. The key notion to guarantee this property is called *s-HRT property* and is made precise by the following definition [32]. Also s-HRT property leads to the required injectivity of the map between complex measure with support of size at most s and the corresponding expansion in terms of kernel functions.

Definition 4.3.2. A Hilbert space with reproducing kernel has the *s-HRT property*, if for any set of s points $\{x_n\}_{n=1}^s$ in X the corresponding kernel functions form a linearly independent set $\{K_{x_n}\}_{n=1}^s$.

Proposition 4.3.3. *If $\mathcal{H}$ has the $2s$-HRT property, then any $f \in \mathcal{H}$ of the form $f = \sum_{n=1}^{s} c_n K_{x_n}$ with $c \in \mathbb{C}^s$ and $x \in X^s$ is s-sparse and the choice of $c \in \mathbb{C}^s$ and $x \in X^s$ is uniquely determined by f.*

The setting for sparse recovery by atomic norm is as following. Given a signal $f \in \mathcal{H}$, assume that $f = \sum_{n=1}^{s} c_n K_{x_n}$ for some coefficient vector $c \in \mathbb{C}^s$. Let $(\langle f, M_i \rangle)_{i \in I}$ be linear measurements indexed by an at most countable set I and corresponding to a family of vectors $(M_i)_{i \in I}$ in $\mathcal{H}$. The standard assumption used is that $(M_i)_{i \in I}$ is a Bessel family with Bessel bound 1, or equivalently, that the measurement map

$$M : \mathcal{H} \to \mathcal{H}^{'}(= l^2(I)), (Mf)_i = \langle f, M_i \rangle$$

has operator norm $\|M\| = 1$.

Before sparse recovery by atomic norm is stated, it is necessarily to ensure $\|c\|_1 = \|f\|_A$ that is the measure $\mu_* = \sum_{n=1}^{s} c_n \delta_{x_n}$ called *atomic decomposition of f* is a minimizer of the program $(\mathcal{P})$.

Definition 4.3.4. Let X be a locally compact Hausdorff space and $\mathcal{H}$ be be a separable reproducing kernel Hilbert space of Borel measurable functions on X with a unit-norm reproducing kernel K, and $f \in \mathcal{M}$ with $\|f\|_A < \infty$. A minimizer μ_* of $(\mathcal{P})$ is called an atomic decomposition of f (not unique but always exist). Another notation used is $D\mu_* = \int_X K_x d\mu_*(x)$.

The following program is proposed for recovering f from the measurement with $(M_i)_{i \in I}$

$$\min \|g\|_A \text{ subject to } \langle f, M_i \rangle = \langle g, M_i \rangle, i \in I \qquad (\mathcal{P}_A)$$

Theorem 4.3.5. *Let $f \in \mathcal{H}$ be given by the linear combination $f = \sum_{n=1}^{s} c_n K_{x_n}$ with $c \in \mathbb{C}^s$ and $x \in X^s$. Assume that the closed linear span of the measurement vectors $\overline{span}\{M_i, i \in I\} = \mathcal{H}'$ contains a continuous function $\psi \in \mathcal{H} \cap \mathcal{C}_0(X)$ with the properties*

(i) $\|\psi\|_\infty = 1$

(ii) $\psi(x_j) = c_j/|c_j|, j = 1, \cdots, m$

(iii) $|\psi(x)| < 1$ *for* $x \notin \{x_j, j = 1, \cdots m\}$

and additionally that $(K_{x_n})_{n=1}^m$ is linearly independent. Then f is the unique solution of the program $(\mathcal{P}_A)$ and $\mu_0 = \sum_{n=1}^{s} c_n \delta_{x_n}$ is an atomic decomposition of f.

4.4 Stability and Robustness Recovery by Atomic Norm Minimization

Recovery based on minimizing the atomic norm can be achieved even if the model assumption of sparsity is only approximately (stability). Also it provides controllable accuracy for recovery when the measurement is corrupted by noise (robustness).

Let $M : \mathcal{H} \to \mathcal{H}'$ be the measurement operator on $\mathcal{H}$ associated with the Bessel family $(M_i)_{i \in I}$ as described above. Given contaminated measurements $b = Mf + n$, of an element $f \in \mathcal{H}$ with an approximately sparse atomic decomposition

$$f = \sum_{n=1}^{s} c_n K_{x_n} + K * \mu_c$$

in $(K_x)_{x \in X}$ and μ_c is arbitrary small in $\mathcal{M}(X)$, we wish to find conditions under which the following problem approximately recovers f:

$$\min \|g\|_A \text{ subject to } \|Mg - b\| \leq \epsilon \qquad (\mathcal{P}_A^\epsilon)$$

At first glance, it might be tempting to generalize l_1-minimization program

$$\min \|\mathbf{x}\|_1 \text{ subject to } \|M\mathbf{x} - b\|_2 < \epsilon$$

by replacing the l_1-norm with the total variation norm where the minimization is over μ whose support is of size s. It is not possible to prove a statement like this in this setting. If $x \to K_x$ is

continuous then $\|MK_{x'} - MK_x\|_2$ is small for x close to $x^{'}$. Hence, for some contamination n, a multiple of $K_{x'}$ is a solution of $(\mathcal{P}_A^\epsilon)$ but $\|\delta_x - \delta_{x'}\|_{TV} = 2$ for all $x \neq x^{'}$.

Inspired by analysis of the structure of the solution of the Beurling LASSO [48], given by

$$\min_{\mu \in \mathcal{M}(\mathbb{T})} \frac{1}{2}\|\Phi\mu - b\|_2^2 + \lambda\|\mu\|_{TV}$$

where $\Phi : \mathcal{M}(\mathbb{T}) \to L^2([0, 2\pi])$ being a smooth convolution operator and $\mathbb{T}$ is the torus, similar results were proved for $(\mathcal{P}_A^\epsilon)$. Intuitively, in the $(\mathcal{P}_A^\epsilon)$ context it states that for signals having an atomic decomposition with small, finite support, most of the mass of the atomic decompositions of the minimizer of $(\mathcal{P}_A^\epsilon)$ will be concentrated in a set close to the support of an atomic measure. To make the notion of concentration quantitative, the following set of interpolating measurements $\Theta_{x,\omega,\lambda,\delta}$ are defined as following. For a sequence $(x_i)_{i\in\mathbb{N}}$ in metric space X and $\delta > 0$ define $S_\delta = \bigcup_{i\in\mathbb{N}} U_\delta(x_i)$ where $U_\delta(x)$ is the open ball of radius δ centered at x. For a sequence $(\omega_j)_{j\in\mathbb{N}}$ in a reproducing kernel Hilbert space $\mathcal{H}$, and $M : \mathcal{H} \to \mathcal{H}^{'}$ a contraction to another Hilbert space and $\lambda, \delta > 0$ then:

$$\begin{aligned}\Theta_{x,\omega,\lambda,\delta} = \{\nu \in \mathcal{H}' :\ & M^*\nu \in \mathcal{H} \cap \mathcal{C}_0(X) \\ & M^*\nu(x_j) = \omega_j, j \in \mathbb{N} \\ & \|M^*\nu\|_\infty \leq 1 \\ & |M^*\nu(x)| \leq \lambda \text{ for } x \notin S_\delta\}\end{aligned}$$

Theorem 4.4.1. *Let X be a second countable locally compact metric space, $\mathcal{H}$ be a separable reproducing kernel Hilbert space over X with unit-norm kernel K such that $\mathcal{C}_0(X) \cap \mathcal{H}$ is dense in $\mathcal{H}$ and $M : \mathcal{H} \to \mathcal{H}^{'}$ a contraction. Let $f \in \mathcal{H}$ have the form*

$$f = \sum_{i\in\mathbb{N}} c_i K_{x_i} + K * \mu_c$$

and let f_* be a solution of the program given by

$$min\|g\|_A \text{ subject to } \|b - Mg\| \leq \epsilon \quad \mathcal{P}_\epsilon$$

with $\|b - Mf\| \leq \epsilon$. Further, define μ_* through $f_* = D\mu_*$, $\|f_*\|_A = \|\mu_*\|_{TV}$ and for $\delta, \lambda > 0$ set

$$C(\lambda, \delta) := \sup_{|\omega_j=1} \inf\{\|\nu\| : \nu \in \theta_{x,\omega,\lambda,\delta}\}$$

Then we have

$$|\mu_*|(S_\delta) \geq \|f\|_A - \frac{2C(\lambda,\delta)\epsilon + \|\mu_c\|_{TV}}{1-\lambda}. \tag{4.1}$$

Theorem 4.4.2. *Let X be a second countable locally compact metric space, $\mathcal{H}$ be a separable reproducing kernel Hilbert space over X with unit-norm kernel K such that $\mathcal{C}_0(X) \cap \mathcal{H}$ is dense in $\mathcal{H}$. Assume that the map $\mathcal{X} \to \mathcal{H}$, $x \mapsto K_x$ is continuous. Let*

$$f = \sum_{i=1}^{s} c_i K_{x_i} + K * \mu_c$$

and let $(\mu^n) \subseteq \triangle$ ($\triangle$ still denotes the set of finitely supported measures) be a minimizing sequence of the program

$$\inf_{\mu \in \triangle} \|\mu\|_{TV} \text{ subject to } \|b - MK * \mu\| \leq \epsilon, (\mathcal{P}_\triangle)$$

with $\|b - Mf\| \leq \epsilon$. Then there exists a $\delta_0 > 0$ so that, for all $\delta \leq \delta_0$ satisfying the same assumptions as in Theorem 4.4.1. and for all sufficiently large n, the following error bound holds:

$$\|D\mu^n - f\| \leq 2C(\lambda,\,\delta)\epsilon + (C(\lambda,\,\delta)+1)\Big(\epsilon\|f\|_A + \|\mu_c\|_{TV} + \frac{2C(\lambda,\delta)\epsilon + \|\mu_c\|_{TV}}{1-\lambda}\Big) = C_1\epsilon + C_2\|\mu_c\|_{TV}$$

Chapter 5

Extended Restricted Isometric Projections for Riemannian Manifolds

5.1 Background on Exterior Calculus in Euclidean Spaces

Definition 5.1.1. We denote by $\Lambda(\mathbb{R}^n) \equiv \Lambda_0 \oplus \Lambda_1 \oplus \ldots \oplus \Lambda_n$ the graded exterior algebra of $\mathbb{R}^n$, where each $\Lambda_k,\ 0 \leq k \leq n$, is spanned by homogeneous products of order k in differentials $\{dx^j\}$, $j = 1, 2, \ldots, n$.

Definition 5.1.2. The exterior differential $d : \Lambda(\mathbb{R}^n) \rightarrow \Lambda(\mathbb{R}^n)$ is defined by linearity and the Leibniz property:

$$\forall q = q_{i_1 i_2 \ldots i_k} dx^{i_1} \wedge dx^{i_2} \wedge \ldots \wedge dx^{i_k} \in \Lambda_k,$$

$$dq \in \Lambda_{k+1},\ dq = d(q_{i_1 i_2 \ldots i_k}) \wedge dx^{i_1} \wedge dx^{i_2} \wedge \ldots \wedge dx^{i_k},$$

along with the antisymmetry of the wedge product.

Remark 5.1.3. We note that $\Lambda_0 = C^\infty(\mathbb{R}^n)$, and $\Lambda_{n+1} = \emptyset$.

Corollary 5.1.4. *If $p \in \Lambda_k, q \in \Lambda_l$, then*

$$d(p \wedge q) = dp \wedge q + (-1)^k p \wedge dq.$$

Theorem 5.1.5. *The exterior derivative is a nilpotent operator, $d^2 = 0$, so for any component of the graded algebra, Λ_k:*

$$Im(d|_{\Lambda_k}) \subset Ker(d|_{\Lambda_{k+1}}),$$

and the quotient space $H^k \equiv Ker(d|_{\Lambda_k})/Im(d|_{\Lambda_{k-1}})$ is an Abelian group, the $k-$th de Rham cohomology space.

Proof. Let $f \in \Lambda_0$, then $df = \partial_i f dx^i$, and

$$d^2 f = d(\partial_i f) \wedge dx^i = \partial^2_{ji} f dx^j \wedge dx^i = 0, \tag{5.1}$$

since the matrix $(\partial_{ji} f)$ is symmetric and $dx^j \wedge dx^i$ is antisymmetric, and we are computing the trace of their product,

$$Tr(AS) = Tr(SA) = Tr(A^t S^t) = -Tr(AS) \Rightarrow Tr(SA) = 0.$$

For the general case, let $q \in Im(d|_{\Lambda_k}), k \geq 1$, then there exists a $k-1$ form $p = p_{i_1 i_2 \ldots i_{k-1}} dx^{i_1} \wedge \ldots \wedge dx^{i_{k-1}} \in \Lambda_{k-1} : q = dp$, so

$$dq = d^2 p = (d^2 p_{i_1 i_2 \ldots i_{k-1}}) \wedge dx^{i_1} \wedge \ldots \wedge dx^{i_{k-1}} = 0$$

by remark 5.1.3.. Therefore, for any $1 \leq k \leq n$, we have the vector subspaces inclusions

$$Im(d|_{\Lambda_k}) \subset Ker(d|_{\Lambda_{k+1}}) \subset \Lambda_{k+1}$$

The quotient space $H^k \equiv Ker(d|_{\Lambda_k})/Im(d|_{\Lambda_{k-1}})$ defined by the equivalence relation $p \sim q \Leftrightarrow p - q \in Im(d|_{\Lambda_{k-1}})$ and equipped with the addition

$$\widehat{p} + \widehat{q} = \widehat{p+q},$$

is a commutative group, by direct verification. □

Example 1. Find the cohomology spaces of $\mathbb{R}^2$.

Solution. Since $H^0 = Ker(d|_{\Lambda_0})$, it consists of functions $f \in C^\infty(\mathbb{R}^2)$ such that $df = 0$, or constants. Therefore, $H^0 \simeq \mathbb{R}$. To find H^1, take $q = q_1 dx^1 + q_2 dx^2$ such that $dq = 0$, or $\partial_2 q_1 = \partial_1 q_2$. By Green's formula,

$$\oint_\gamma dp = \iint (\partial_1 q_2 - \partial_2 q_1) dx^1 dx^2 = 0$$

for any closed curve $\gamma \in \mathbb{R}^2$, so there exists a function $f \in \Lambda_0$ such that

$$f(x^1, x^2) = \int_{(1,1) \to (x^1, x^2)} q, \quad df = q,$$

where the integration is taken over any curve with endpoints $(1,1), (x^1, x^2)$. Therefore, $q \in Im(d|_{\Lambda_0})$, or $H^1 \simeq \{0\}$.

Example 2. Find the cohomology spaces of $\mathbb{R}^2 \setminus \{(0,0)\}$.

Solution. $H^0 \simeq \mathbb{R}$ by the same argument as in the previous problem. For H^1, note that there are two inequivalent ways of choosing a path $(1,1) \to (x^1, x^2)$, namely circling either clockwise or counterclockwise the origin $(0,0)$. Therefore, $H^1 \simeq \{0,1\} \simeq \mathbb{Z}_2$.

5.2 Induced Metric for Elementary Manifolds Embedded in $\mathbb{R}^n$

Tensor spaces and tensor products

Definition 5.2.1. Let $(V,+)$ be a vector space over the field $\mathbb{K}$, $\dim(V) = n$, V^* its algebraic dual, and $\otimes : \mathbb{K}^n \times \mathbb{K}^n \to \mathbb{K}^{n\times n}$ a bilinear function, such that $Range(\otimes)$ is a vector subspace of $\mathbb{K}^{n\times n}$. We call $\otimes$ a tensor product, and the spaces

$$\mathcal{T}^{(p,q)}(\mathbb{K}) \equiv V^{\otimes^p} \otimes (V^*)^{\otimes^q}$$

the $(p,q)-$tensor spaces over $\mathbb{K}$, with the tensor product $\otimes$.

Example 3. The most natural example for the tensor product is the direct product,

$$e_i \otimes e_j = e_{ij}, \quad \forall\, i,j = 1,2,\ldots,n,$$

and $\{e_i\}_{i=1}^n$ a canonical basis for $\mathbb{K}^n$, that is $(e_i)_j = \delta_{ij}$, while $\{e_{ij}\}_{i,j=1}^n$ is a canonical basis for $\mathbb{K}^{n\times n}$, that is $(e_{ij})_{kl} = \delta_{ik}\delta_{jl}$. Then for any two vectors

$$v = \sum_{i=1}^n v_i e_i,\ w = \sum_{j=1}^n w_j e_j \Rightarrow v \otimes w = \sum_{i=1}^n \sum_{j=1}^n v_i w_j e_{ij}.$$

Theorem 5.2.2. *If $\otimes$ is a tensor product, then $\mathcal{T}^{(p,q)}(\mathbb{K})$ is a vector space over $\mathbb{K}$. If $\{e_i\}_{i=1}^n$ is a basis for V, and $\{f_j\}_{i=1}^n$ is a basis for V^*, then*

$$\mathcal{B} = \left\{e_{i_1} \otimes e_{i_2} \otimes \ldots \otimes e_{i_p} \otimes f_{j_1} \otimes f_{j_2} \otimes \ldots \otimes f_{j_q}\right\}_{i_1,i_2,\ldots,j_1,j_2,\ldots=1}^n$$

is a basis for $\mathcal{T}^{(p,q)}(\mathbb{K})$.

Theorem 5.2.3. *If $\otimes$ is the direct tensor product over $\mathbb{R}^n$, then the functions $\odot, \wedge$ defined as*

$$v \odot w \equiv \frac{1}{2}\left(v \otimes w + w \otimes v\right), \quad v \wedge w \equiv \frac{1}{2}\left(v \otimes w - w \otimes v\right)$$

are also tensor products over $\mathbb{R}^n$, called the symmetric *and* antisymmetric *products, respectively. The corresponding tensor product spaces are, respectively,*

$$\mathbb{R}^n \odot \mathbb{R}^n \simeq \mathbb{R}^{n(n+1)/2}, \quad \mathbb{R}^n \wedge \mathbb{R}^n \simeq \mathbb{R}^{n(n-1)/2}.$$

Also, there is the decomposition

$$\mathbb{R}^n \otimes \mathbb{R}^n = (\mathbb{R}^n \odot \mathbb{R}^n) \oplus (\mathbb{R}^n \wedge \mathbb{R}^n)$$

Example 4. Let $\odot, \wedge$ be the symmetric, and antisymmetric products on $\mathbb{R}^n$. Then for the vectors

$$v = \sum_{i=1}^{n} v_i e_i, \quad w = \sum_{j=1}^{n} w_j e_j,$$

we have

$$v \odot w = \sum_{i \leq j=1}^{n} (v_i w_j + w_i v_j) e_i \odot e_j, \quad v \wedge w = \sum_{1 \leq i < j}^{n} (v_i w_j - w_i v_j) e_i \wedge e_j.$$

Bilinear forms and inner products. Given the isomorphism between spaces V and V^*, it is possible to extend it to identify a subspace of $L(V, V^*)$ with the dual of $\mathcal{T}^{(2,0)}(V)$, or of $\mathcal{T}^{(0,2)}(V)$, or of $\mathcal{T}^{(1,1)}(V)$. We use Einstein's summation convention for repeated indices in the following.

Theorem 5.2.4. *Let Q be an $n \times n$ symmetric, invertible matrix, $\widetilde{Q}$ its inverse, and $\{e_i\}_{i=1}^n$, $\{f^j\}_{j=1}^n$ two bases for $\mathbb{R}^n$. The function $\mathcal{B}_Q : \mathbb{R}^n \to \mathbb{R}^n$ defined by*

$$\mathcal{B}_Q(v) = \lambda, \quad v = x^i e_i, \quad \lambda = y_j f^j,$$

where

$$y_j = Q_{ji} x^i,$$

is a vector space isomorphism. Its inverse, $\mathcal{B}_{\widetilde{Q}}^{-1} : \mathbb{R}^n \to \mathbb{R}^n$, is defined by

$$\mathcal{B}_{\widetilde{Q}}^{-1}(\lambda) = v, \quad v = x^i e_i, \quad \lambda = y_j f^j, \quad x^i = \widetilde{Q}^{ik} y_k.$$

As $V \simeq \mathbb{R}^n \simeq V^*$, this induces an isomorphism $\mathcal{B}_Q : V \to V^*$.

Example 5. If Q is the identity matrix, then $y_j = \delta_{ji}x^i$ is the canonical isomorphism $V \simeq V^*$ (equivalent to vector transposition).

Corollary 5.2.5. *Let $\mathcal{S}(V, V^*)$ be the space of non-degenerate, symmetric bilinear functionals $q : V \otimes V^* \to \mathbb{R}$,*

$$q(\,.\,, \lambda) \in V^*,\ \forall \lambda \in V^*, \quad q(v,\,.\,) \in V,\ \forall v \in V.$$

Then to every element of $\mathcal{S}(V, V^*)$ there corresponds uniquely a non-degenerate, symmetric matrix Q, and therefore an isomorphism $\mathcal{B}_Q : V \to V^*$.

The important consequence of these observations is identifying the basis changes in V, V^* which leave an element of $[\mathcal{T}^{(2,0)}(V)]^*$ unchanged:

Theorem 5.2.6. *Let M be a coordinate change for the space V, and $q \in [\mathcal{T}^{(2,0)}(V)]^*$ a non-degenerate bilinear functional on V,*

$$q(v, w) = Q_{ij}x^i y^j, \quad v = x^i e_i,\ w = y^j e_j.$$

Then the coordinate change $\{x^i\} \to \{\tilde{x}^k\}$, $\{y^i\} \to \{\tilde{y}^l\}$, given by

$$\widetilde{X} = M \cdot X,\ \widetilde{Y} = M \cdot Y,$$

leaves q unchanged iff.

$$M^T \cdot Q \cdot M = Q.$$

Proof. Write

$$q(v, w) = X^T \cdot Q \cdot Y,$$

then we have the condition

$$\widetilde{X}^T \cdot \widetilde{Q} \cdot \widetilde{Y} = X^T \cdot Q \cdot Y,$$

or

$$X^T \cdot M^T \cdot \widetilde{Q} \cdot M \cdot Y = X^T \cdot Q \cdot Y.$$

Bilinearity and the condition $\widetilde{Q} = Q$ now complete the proof. □

Definition 5.2.7. Let $g : \mathbb{R}^n \to Gl(n, \mathbb{R})$, such that $g(x)$ is a continuous function, and $g(x) = (g_{ij}(x))$ is a strictly-positive definite element of the tensor space $\mathcal{T}^{(0,2)}(\mathbb{R}^n)$. Then g is called a metric on $\mathbb{R}^n$.

Remark 5.2.8. The fact that the function g takes its values into the tensor space $\mathcal{T}^{(0,2)}(\mathbb{R}^n)$ means that it has the canonical basis representation

$$g(x) = g_{ij}(x) dx^i \odot dx^j,$$

where $\odot$ denotes the symmetric tensor product, such that under any local change of coordinates $\{x^i\} \to \{y^j\}$, $dy^j = m^j_i dx^i$, the tensor coordinates

$$g_{ij} \to h_{lm}, \quad h_{ij} = \tilde{m}^l_i \tilde{m}^k_j g_{lk},$$

where $\tilde{m}$ is the inverse transformation, $\tilde{m}^l_i m^j_l = \delta^j_i$.

Example 6. The Euclidean norm is given by the choice $g_{ij}(x) = \delta_{ij}$. Under a global coordinate transformation $X \to Y = M \cdot X$ (with M an invertible real matrix), the metric coefficients become

$$h_{ij} = (M^{-1})^l_i (M^{-1})^k_j \delta_{lk},$$

or equivalently the coefficients h_{ij} correspond to the symmetric square matrix $M^{-1} \cdot (M^{-1})^t$.

Definition 5.2.9. A set $\mathcal{M}$ is an elementary differentiable manifold embedded in $\mathbb{R}^n$ if it is the range of a differentiable vector function with full-rank derivative defined on a (parametrization) domain $P \subset \mathbb{R}^m$ of full measure, $F : D \subset \mathbb{R}^m \to \mathbb{R}^n, m < n$, or component-wise

$$\mathcal{M} = \{(F^j(t)), 1 \leq j \leq n, t \in P\}, \ D_t F = (\partial_i F^j)^{1 \leq j \leq n}_{1 \leq i \leq m}, \ \text{rank}\,(D_t F)_{t \in P} = m.$$

The (manifold) dimension of $\mathcal{M}$ is m.

Example 7. The upper half-circle is a manifold of dimension 1, embedded into $\mathbb{R}^2$ by the vector function

$$F : (0, \pi) \to \mathbb{R}^2, \ F(\theta) = (\cos\theta, \sin\theta).$$

Example 8. The right hemisphere is a manifold of dimension 2, embedded into $\mathbb{R}^3$ by the vector function

$$F : (0, \pi) \times (0, \pi) \to \mathbb{R}^3,\ F(\theta, \phi) = (\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta).$$

Definition 5.2.10. Let $\mathcal{M}$ be an elementary manifold of dimension m embedded into $\mathbb{R}^n$ by the vector function F, with full-rank derivative DF given by the matrix-valued function $(D_i^j)_{1\leq i\leq m}^{1\leq j\leq n}$, $dF^j = D_i^j dt^i$. If the embedding space has metric $g(x)$ given by the matrix $G \in Gl(n, \mathbb{R})$ of tensor coefficients g_{ij}, then the induced metric on $\mathcal{M}$ is given by the (matrix of) coefficients

$$h_{ij}(t)dt^i \odot dt^j,\ H = D^t \cdot G \cdot D \in Gl(m, \mathbb{R}).$$

Theorem 5.2.11. *The induced metric is well-defined, as under any coordinate change $dt^i = m_j^i du^j$, the coefficient matrix changes as given by Remark* (5.2.8)*.*

Example 9. The induced metric on a circular arc embedded into $\mathbb{R}^2$ with canonical Euclidean metric $\delta_{ij}dx^i \odot dx^j$ is given by the tensor-valued function

$$g(\theta) = d\theta \odot d\theta.$$

Example 10. The induced metric on the hemisphere embedded into $\mathbb{R}^3$ with canonical Euclidean metric $\delta_{ij}dx^i \odot dx^j$ is given by the tensor-valued function

$$g(\theta, \phi) = d\theta \odot d\theta + \sin^2(\theta)d\phi \odot d\phi.$$

Example 11. Compute the arclength of the curve of endpoints $(\theta, \phi_1), (\theta, \phi_2)$, given by the parametrization $(\theta(t), \phi(t)) = (\theta, \phi_1 + (\phi_2 - \phi_1)t), t \in [0, 1]$.

Solution. The arclength of the curve given is

$$\ell = \int_0^1 \sqrt{\left(\frac{d\theta}{dt}\right)^2 + \sin^2\theta\left(\frac{d\phi}{dt}\right)^2}\,dt = \int_{\phi_1}^{\phi_2} \sin(\theta)d\phi = \sin(\theta)(\phi_2 - \phi_1).$$

Remark 5.2.12. Arclength is independent of the parametrization used (under differomorphic reparametrization). Therefore, it would have been equally valid to choose the parametrization given by the coordinate ϕ.

5.3 Differentiable Varieties as Generalized Elementary Manifolds

Recall that an elementary manifold can be understood as the image of an injective differentiable function, embedded in Euclidean space, $\vec{F}(D) \subset \mathbb{R}^n,\ D \subset \mathbb{R}^m$. The inverse function $\vec{F}^{-1}$ is well defined on $\vec{F}(D)$, and serves as a coordinate map for the points on the manifold. We now generalize this construction by requiring the existence of such maps only locally.

Definition 5.3.1. A set $\mathcal{M}$ is a differentiable variety of real dimension m if for every point $p \in \mathcal{M}$, there exist an open set $p \in U \subset \mathcal{M}$ and a homeomorphism $f_U : U \to \mathbb{R}^m$ (called local coordinate chart), such that if U_i, U_j, U_k are three such sets and f_i, f_j, f_k their respective charts, then the transition function

$$\varphi_{ij} \equiv f_i \circ f_j^{-1} : f_j(U_i \cap U_j) \to f_i(U_i \cap U_j)$$

is a diffeomorphism, and $\varphi_{ij} \circ \varphi_{jk} = \varphi_{ik}$ on $U_i \cap U_j \cap U_k$.

Example 12. For the circle S^1, we can define local charts via stereographic projection from the points $N \equiv (0, 1)$ and $S \equiv (0, -1)$. Let $U_N \equiv S^1 \setminus \{N\}$, $U_S \equiv S^1 \setminus \{S\}$, and

$$f_N(x, y) = \frac{x}{1 - y}, \quad f_S(x, y) = \frac{x}{1 + y}$$

It can checked by elementary geometry that these differentiable functions are indeed bijections, and that

$$f_N(U_N) = f_S(U_S) = \mathbb{R}, \quad f_N(S) = f_S(N) = 0,$$

and moreover $f_N(U_N \cap U_S) = f_S(U_N \cap U_S) = \mathbb{R} \setminus \{0\}$, such that $\forall t \in \mathbb{R} \setminus \{0\}$,

$$f_N^{-1}(t) = \left(\frac{2t}{t^2 + 1}, \frac{t^2 - 1}{t^2 + 1} \right), \quad (f_S \circ f_N^{-1})(t) = \frac{1}{t}.$$

Since $t \to 1/t$ is a diffeomorphism on $\mathbb{R} \setminus \{0\}$, all the conditions are satisfied.

Definition 5.3.2. Let $\mathcal{M}, \mathcal{N}$ be differentiable varieties, and $F : \mathcal{M} \to \mathcal{N}$ a bijective function. Then F is a (manifold) diffeomorphism if for any local charts $f : U \subset \mathcal{M}$, $g : V \subset \mathcal{N}$, the function $g \circ F \circ f^{-1}$ is a diffeomorphism on Euclidean spaces. We then say that $\mathcal{M}, \mathcal{N}$ are diffeomorphic.

Corollary 5.3.3. *It follows from the definitions that if $\mathcal{M}, \mathcal{N}$ are diffeomorphic, then they have the same dimension.*

Example 13. Let $F : S^1 \to S^1$, given by the reflexion with respect to the $x-$axis, or $F(x, y) = (x, -y)$ in the embedding space $\mathbb{R}^2$. Obviously, $F(N) = S, F(S) = N$, and for $t \neq 0$,

$$(f_N \circ F \circ f_N^{-1})(t) = (f_S \circ F \circ f_S^{-1})(t) = \frac{1}{t},\ (f_N \circ F \circ f_S^{-1})(t) = (f_S \circ F \circ f_N^{-1})(t) = t,$$

which are diffeomorphisms on $\mathbb{R} \setminus \{0\}$. The last two composition maps can be used at points N, S respectively, as well.

5.3.1 Tangent and Cotangent Spaces

Definition 5.3.4. Let $\mathcal{M}$ be an $m-$dimensional differentiable variety, $p \in \mathcal{M}$, and $f : U \to \mathbb{R}^m, p \in U \subset \mathcal{M}$ a local coordinate chart. Let $x_0 = f(p)$, and denote by $\gamma^i(t) = x_0 + te^i$, where $\{e^i\}_{i=1}^m$ is an orthonormal basis, $(e^i)_j = \delta^i_j$, and $t \in \mathbb{R}$. The pre-images $\Gamma^i \equiv f^{-1}(\gamma^i(\mathbb{R}) \cap f(U))$ are curves in U, such that $\cap_i \Gamma^i = \{p\}$, and their tangent vectors at p span the tangent space $T_p\mathcal{M}$:

$$T_p\mathcal{M} \equiv \mathrm{Span}_{\mathbb{R}} \left\{ \frac{d\Gamma^i}{dt}\Big|_{t=0} \right\}$$

Remark 5.3.5. The linear independence of the basis vectors defined above follows from a direct application of chain rule,

$$\frac{d\Gamma^i}{dt}\Big|_{t=0} = D_{x_0}(f^{-1}) \cdot \frac{d\gamma^i}{dt}\Big|_{t=0} = D_{x_0}(f^{-1}) \cdot e^i,$$

and the condition that f^{-1} is a diffeomorphism.

We introduce now a central object of differential geometry, which makes contact with function theory in general. From this point forward, we will be mainly interested to describe the geometric properties of a variety through their action on real-valued smooth functions defined on the variety. To this end, we define the algebra of smooth functions on a differentiable variety.

Definition 5.3.6. Let $\mathcal{M}$ be an $m-$dimensional differentiable variety, $p \in \mathcal{M}$, and $f : U \to \mathbb{R}^m, p \in U \subset \mathcal{M}$ a local coordinate chart. The set

$$\mathcal{A}(\mathcal{M}) \equiv \{h : \mathcal{M} \to \mathbb{R} | h \circ f^{-1} \in C^\infty(f(U))\}$$

is called the algebra of smooth functions on $\mathcal{M}$.

Remark 5.3.7. Smoothness means that any partial derivative, of any order, of the function $h \circ f^{-1}$, is well-defined on $f(U)$. The fact that this set is closed under addition and multiplication follows directly from basic properties of differentiable functions (linearity and Leibniz property of the derivative).

Consider now a function $h \in \mathcal{A}(\mathcal{M})$, and the problem of computing the derivative of the function $(h \circ f^{-1} \circ \gamma)(t)$, where $\gamma = \alpha_i \gamma^i$ is a linear combination of the lines $\{\gamma^i\}$. By chain rule, we obtain

$$\frac{dh}{dt}\Big|_{t=0} = \frac{\partial(h \circ f^{-1})}{\partial x^k} \frac{d\gamma^k}{dt}\Big|_{t=0} = \alpha^k \frac{\partial}{\partial x^k}(h \circ f^{-1})\Big|_{t=0},$$

which shows that there is an obvious vector space isomorphism between $T_p\mathcal{M}$ and Span $\{\partial_k\}, k = 1, 2, \ldots, m$, where $(x^k(q)) \equiv f(q), \forall q \in U$.

Definition 5.3.8. The dual space of $T_p\mathcal{M}$ (understood as a space of linear differential operators), or the space of linear functionals over Span $\{\partial_k\}, k = 1, 2, \ldots, m$ is defined as the cotangent space $T_p^*\mathcal{M}$. The elements of the canonical dual basis are denoted by $\{dx^k\}$, and defined by

$$dx^i(\partial_j) = \delta^i_j,$$

and by linearity.

5.4 Vector Bundles, Sections, and Covariant Exterior Differential

The kind of differentiable variety (called vector bundle) arising in general relativity, and to some extent in gauge theories as well, has a special structure which allows to regard it as a "base manifold" endowed with a local copy of the same vector space, called "fiber". The more subtle aspects of the general definition have to do with the fact that, *a priori*, we do not know if (or how) such a variety is embedded in $\mathbb{R}^N$ or $\mathbb{C}^N$, for some (large) $N \in \mathbb{N}$.

Definition 5.4.1. A differentiable variety $\mathcal{V}$ (called total space) of real dimension $n+m$ is a vector bundle over the differentiable variety $\mathcal{M}$ (base manifold) with fiber $E \simeq \mathbb{R}^n$, and projection map $\pi : \mathcal{V} \to \mathcal{M}$ if π is continuous, surjective, and satisfies the following:

(i) $\forall p \in \mathcal{V}, \exists$ an open set $U : \pi(p) \in U \subset \mathcal{M}$ and a diffeomorphism $\pi^{-1}(U) \to f(U) \times E$, where f is a local coordinate chart at $\pi(p) \in \mathcal{M}$

(ii) if (U_i, f_i) and (U_j, f_j) are two coordinate charts at $\pi(p) \in U_i \cap U_j$, and $(\pi^{-1}(U_i), \varphi_i)$, $(\pi^{-1}(U_j), \varphi_j)$ are coordinate charts at p, then the transition function

$$\phi_{ij} \equiv \varphi_i \circ \varphi_j^{-1} : \varphi_j(\pi^{-1}(U_i) \cap \pi^{-1}(U_j)) \to \varphi_i(\pi^{-1}(U_i) \cap \pi^{-1}(U_j))$$

is a diffeomorphism on $\mathbb{R}^{n+m}$ that has the form $\phi_{ij}(x, e) = (x, f_{ij}e)$, where $f_{ij} \in \mathcal{G} = Gl(n, \mathbb{R})$. The group $\mathcal{G}$ is called the **structural group** of the bundle;

(iii) for three coordinate charts at $\pi(p)$, U_i, U_j, U_k with $U_i \cap U_j \cap U_k \neq \emptyset$, the transition functions satisfy $f_{ij}f_{ji} = \mathbb{I}_{n\times n}$, $f_{ij}f_{jk}f_{ki} = \mathbb{I}_{n\times n}$.

Remark 5.4.2. It is possible to replace the structural group $\mathcal{G}$ in condition (ii) by subgroups of $Gl(n, \mathbb{R})$, such as $SO(n)$ or $SL(n, \mathbb{R})$. Likewise, it is possible to replace $\mathbb{R}^n$ by $\mathbb{C}^n$ throughout, and therefore define complex varieties, with structural groups such as $U(n), SU(n)$, etc.

Definition 5.4.3. A section on a vector bundle $\mathcal{V}$ over the base $\mathcal{M}$ is a function $u : \mathcal{M} \to \mathcal{V}$ such that $(\pi \circ u)(x) = x$, $\forall x \in \mathcal{M}$.

Example 14. Let $\mathcal{V} = \{(e^{i\theta}, ie^{i\theta}(t + t\cos\theta, t\sin\theta)), \theta \in [0, 2\pi], t \in \mathbb{R}\}$. Then $\mathcal{V}$ is a vector bundle over $\mathcal{M} = S^1$, with fiber $\mathbb{R}$. It can be embedded into $\mathbb{R}^3$ as a slanted cylinder over the unit circle. The function

$$u : S^1 \to \mathcal{V},\ u(e^{i\theta}) = (e^{i\theta}, ie^{i\theta}(\cos\theta + \cos^2\theta, \cos\theta\sin\theta))$$

is a section on $\mathcal{V}$.

Definition 5.4.4. Given two vector bundles $\mathcal{V}_1, \mathcal{V}_2$ over the same base $\mathcal{M}$, of fibers E_1, E_2, the tensor sum, tensor product, tensor powers, duals of the two fibers are defined as equivalence classes (under global diffeomorphisms) of the respective vector bundles over $\mathcal{M}$, with fibers $E_1 \oplus E_2, E_1 \otimes E_2, E_1^*$.

Example 15. The tensor sum $\mathcal{V} = \mathbb{R} \oplus \mathbb{R}^n$ is a vector bundle over $\mathbb{R}$, of fiber $\mathbb{R}^n$. The bundle product $\mathcal{V} \otimes \mathcal{V}$ is then defined as

$$\mathcal{V} \otimes \mathcal{V} \equiv \mathbb{R} \oplus (\mathbb{R}^n \otimes \mathbb{R}^n)$$

5.4.1 Classification of Vector Bundles

Definition 5.4.5. Two vector bundles $\mathcal{V}_1, \mathcal{V}_2$ over the same base, and with projection maps π_1, π_2, are isomorphic if there exists an invertible function $F : \mathcal{V}_1 \to \mathcal{V}_2$, such that

$$F(p) = q \Rightarrow \pi_1(p) = \pi_2(q),\ F|_{\pi_1^{-1}(\{x\})} : \pi_1^{-1}(\{x\}) \to \pi_2^{-1}(\{x\})$$

is a vector space isomorphism, and the same for the inverse function F^{-1}.

There is a useful criterion for vector bundle isomorphism, given below.

Theorem 5.4.6. *Two vector bundles with transition functions ϕ_{ij}, ψ_{ij} (with respect to the same covering $\{U_i\}$) are isomorphic if and only if there exist (intertwining) functions $r_i : U_i \to Gl(n, \mathbb{R})$, such that*

$$\psi_{ij} r_j = r_i \phi_{ij}.$$

Remark 5.4.7. We are only concerned with characterizing vector bundles up to bundle isomorphism.

There exists also a useful method to "transplant" the fiber of one manifold onto another base manifold, and create a new vector bundle this way. This is referred to as **inducing**.

Theorem 5.4.8. *Let $\mathcal{V}$ be a vector bundle over $\mathcal{M}$, with projection map π, and $f : \mathcal{N} \to \mathcal{M}$ a function between manifolds. Then the induced bundle of $\mathcal{V}$ onto $\mathcal{N}$ is the vector bundle over $\mathcal{N}$ with fiber E' (or projection function π') defined by*

$$(\pi')^{-1}(\{x\}) = E'_x \equiv E_{f(x)} = \pi^{-1}(\{f(x)\}),\ \forall x \in \mathcal{N}.$$

A classification result for vector bundles over compact manifolds is obtained by considering (linear) projector maps. Let $\mathcal{M}$ be a compact (base) manifold, and $P : \mathcal{M} \to \mathbb{C}^{N \times N}$ a projector function of rank $n \leq N$, that is

$$P^2(x) = P(x),\ \operatorname{Tr} P(x) = n,\ \forall x \in \mathcal{M}.$$

The range of this function is a vector subspace of $\mathbb{C}^N$, and it defines the fiber E. More precisely, let G_N^n be the space of all $n-$dimensional subspaces of $\mathbb{C}^N$ (known as a Grassmannian). Any point in $x \in G_N^n$ is an $n-$dimensional subspace of $\mathbb{C}^N$, and we define the fiber to be the subspace itself:

$$\pi^{-1}(\{x\}) \equiv \{x\} = E.$$

Then (G_N^n, E, π) is a vector bundle over G_N^n. The construction is completed by inducing a vector bundle over $\mathcal{M}$ using the function P. We obtain:

Theorem 5.4.9 (Vector bundle classification)**.** *Any vector bundle over a compact base $\mathcal{M}$ and fiber dimension n is isomorphic to a bundle induced over $\mathcal{M}$ by a projector function of rank n, called* **characteristic map***,*

$$P : \mathcal{M} \to G_N^n,$$

for some N large enough. Two bundles are isomorphic if and only if their characteristic maps are homotopic.

5.4.2 Calculus on Vector Bundles

Definition 5.4.10. Let $\mathcal{V}$ be a (real) vector bundle over the $m-$dimensional base manifold $\mathcal{M}$, and fiber E, of dimension n. We denote by $T\mathcal{M}, T^*\mathcal{M}$ the tangent and cotangent bundles over $\mathcal{M}$, by $\Lambda^k(\mathcal{M})$ the bundle of $k-$forms over $\mathcal{M}$, and by $\Lambda(\mathcal{M})$ the tensor sum of all $\Lambda^k(\mathcal{M})$. We denote by

$$\mathcal{S}(E,k) \equiv C^\infty(E \otimes \Lambda^k),\ \mathcal{S}(E) \equiv C^\infty(E \otimes \Lambda)$$

the spaces of smooth sections over $\mathcal{V} \otimes \Lambda^k(\mathcal{M})$ and $\mathcal{V} \otimes \Lambda(\mathcal{M})$, respectively.

Using the vector bundles classification theorem, it is possible, in principle, to embed any bundle over a compact base into a (big enough) Grassmannian bundle, which is in a way a trivial (as in, linear) structure. Then defining differential operators on the bundle of interest reduces to projecting the usual differential operators (on the Grassmannian) back onto our bundle. However, this approach is not practical. We are therefore led to enlarging the notion of exterior differential and tangent vector, in such a way as to ensure they are differential operators acting between the proper spaces. A direct way to achieve this generalization is by defining a **connection form**.

Definition 5.4.11. Let $\mathcal{V}$ be a vector bundle over $\mathcal{M}$, of fiber E, and $\mathcal{V}^*$ its dual fiber. A connection form A is a section on $\mathcal{V} \otimes \mathcal{V}^* \otimes \Lambda^1(\mathcal{M})$, such that

$$\nabla(A) \equiv \mathbb{I}_{n\times n} d + A : \mathcal{S}(E) \to \mathcal{S}(E)$$

is a covariant linear operator.

If $\{e_k\}_{k=1}^n$ forms a basis in the local fiber at $x \in \mathcal{M}$, $E_x = \pi^{-1}(\{x\})$, and we denote by $\{f^j\}_{j=1}^n$ its canonical dual base, then locally a connection form can be written (summation over repeated indices is assumed)

$$A = (A_k)^i_j e_i \otimes f^j \otimes dx^k,$$

so it is equivalent to m matrices A_k of dimension $n \times n$. In particular, its action on the basis vectors is given by

$$Ae_l = (A_k)^i_l e_i \otimes dx^k,$$

where we have used the identity $f^j(e_l) = \delta^j_l$. Using (multi)linearity and the Leibniz property, we can define the action of the operator $\nabla(A)$ on any smooth section in a tensor bundle $\mathcal{S}(\mathcal{T}^{(p,q)}(E))$. In particular, we have the most common cases of homogenous forms:

Theorem 5.4.12. *Let $u \in \mathcal{S}(E, k)$, or $u = u^j_{i_1 i_2 \ldots i_k}(x) e_j \otimes dx^{i_1} \wedge \ldots \wedge dx^{i_k}$, then $\nabla(A)(u) = w \in \mathcal{S}(E, k+1)$, or $w = w^j_{i_1 i_2 \ldots i_{k+1}}(x) e_j \otimes dx^{i_1} \wedge \ldots \wedge dx^{i_{k+1}}$, with*

$$w = du + A \wedge u = [\partial_l u^j_{i_1 i_2 \ldots i_k} + (A_l)^j_i u^i_{i_1 i_2 \ldots i_k}] e_j \otimes dx^l \wedge dx^l \wedge dx^{i_1} \wedge \ldots \wedge dx^{i_k}$$

If $u \in \mathcal{S}(E, 0)$ (a smooth function on the base manifold), then $\nabla u \equiv du$.

Remark 5.4.13. It is important to note that any section is locally a tensor, and therefore its density must transform accordingly under coordinate transformations. Likewise, the coefficients of a connection must transform under coordinate changes (in the fiber) in such a way as to ensure that the covariant differential $\nabla(A)$ is indeed invariant under coordinate transformations.

Theorem 5.4.14. *Let G be a section on $\mathcal{V} \otimes \mathcal{V}^*$, such that $G(x) \in Gl(n, \mathbb{R})$, representing local coordinate changes in the fiber of $\mathcal{V}$. Then the connection form A transforms as*

$$A \to B = GAG^{-1} + dGG^{-1}.$$

Proof. In the basis $\{e_i\}_{i=1}^n$ of the fiber, we have

$$\nabla e_i = (A_k)_i^l e_l dx^k.$$

Likewise, in the new basis $\{\epsilon_j\}_{j=1}^n, \epsilon_j = G_j^i e_i$, we have

$$\nabla \epsilon_i = (B_k)_i^l \epsilon_l dx^k.$$

Therefore,

$$(B_k)_i^l \epsilon_l dx^k = (B_k)_i^l G_l^j e_j dx^k = \nabla(G_i^j e_j) = dG_i^j e_j + (A_k)_i^l G_l^r e_r dx^k$$

In matrix form, $BG = GA + dG$, so by multiplying from the right by G^{-1}, we obtain the result. □

Corollary 5.4.15. *The connection form does not transform covariantly under fiber basis changes.*

5.4.3 Curvature Form, Bianchi Identities, and Torsion of the Connection

The nilpotency of the exterior differential d in exterior calculus on Euclidean spaces suggests to consider similar quantities in a general vector bundle. To that end, we compute the covariant differential (for a given connection A) of the connection form, and of an invariant 1-form, defined below.

Definition 5.4.16. The 2-form Ω defined by

$$\Omega(A) = \nabla A = dA + A \wedge A$$

is called the curvature of the connection A.

Remark 5.4.17. As the connection is a matrix-valued form, the wedge product $A \wedge A$ does not vanish. Instead, we obtain the component representation

$$\Omega = \Omega_{ij} dx^i \wedge dx^j, \quad \Omega_{ij} = \partial_i A_j - \partial_j A_i + [A_i, A_j]$$

Theorem 5.4.18. *Curvature is a global section of $\mathcal{V} \otimes \mathcal{V}^* \otimes \Lambda^2(\mathcal{M})$.*

Proof. Under a basis change, the connection becomes $A \to GAG^{-1} + dGG^{-1}$. Introducing the notation $G(x) \equiv e^{g(x)}$, we compute the new curvature form as

$$\Omega \to d(GAG^{-1} + dg) + (GAG^{-1} + dg) \wedge (GAG^{-1} + dg)$$

We first note that all the terms independent on A vanish identically: $d^2 g = 0$, and

$$dg \wedge dg = \partial_j g dx^j \wedge \partial_k g dx^k = [\partial_j g, \partial_k g] dx^j \wedge dx^k$$

Since $\partial_j g \partial_k g = \partial_j (g \partial_k g) - g\partial^2_{jk} g = \partial^2_{jk}(g^2/2) - g\partial^2_{jk} g$, which is a symmetric tensor, we conclude that $[\partial_j g, \partial_k g] dx^j \wedge dx^k = 0$. We are therefore left evaluating

$$d(GAG^{-1}) + GAG^{-1} \wedge GAG^{-1} + GAG^{-1} \wedge dg + dg \wedge GAG^{-1}$$

From the first and the second term, we obtain

$$GdAG^{-1} + GA \wedge AG^{-1} = G\Omega G^{-1}.$$

The remaining terms can be written, using the rules of exterior differentials,

$$dG \wedge AG^{-1} - GA \wedge dG^{-1} + GAG^{-1} \wedge dg + dg \wedge GAG^{-1}$$

Finally, using $dG = dgG, dG^{-1} = -G^{-1}dg$ (as G, g commute), we obtain

$$2G(dg \wedge A + A \wedge dg)G^{-1} = 0.$$

Thus, the curvature form transforms covariantly, so it is a global section. □

Theorem 5.4.19. *Let $u \in \mathcal{S}(E)$, then*

$$\nabla^2 u = \Omega u.$$

Proof. From the definition,

$$\nabla u = du + A \wedge u, \ \nabla^2 u = d(\nabla u) + A \wedge \nabla u = d(A \wedge u) + A \wedge du + A \wedge A \wedge u$$

Since by the generalized Leibniz property

$$d(A \wedge u) + A \wedge du = dA \wedge u,$$

we obtain the desired result. □

Remark 5.4.20. Notice that in "flat" spaces (Euclidean spaces), the operator d satisfies $d^2 = 0$. Therefore, $\Omega = \nabla^2 \neq 0$ is a measure of "non-flatness".

In view of this result, it is natural to inquire if higher powers of the covariant differential are also non-trivial. The (negative) answer is known as the Bianchi identity.

Theorem 5.4.21 (The Bianchi identity)**.** *For any curvature form,*

$$\nabla\Omega = 0.$$

Proof. If u is a section, then

$$\nabla(\Omega u) = (\nabla\Omega)u + \Omega\nabla u,$$

as Ω is a 2-form. On the other hand,

$$\nabla(\Omega u) = \nabla(\nabla^2 u) = \nabla^2(\nabla u) = \Omega\nabla u.$$

Comparing the two equations, we obtain $\nabla\Omega = 0$. □

5.4.4 Torsion of a Connection Form

Assume that dim $\mathcal{V}$ = 2 dim $\mathcal{M}$, or equivalently that the fiber is isomorphic to the tangent space. For some vector bundles of this type, it is possible to define a global 1-form which realizes this isomorphism:

Definition 5.4.22. If n = dim E = dim $\mathcal{M}$, a fundamental one-form is a form $\theta \in \mathcal{V} \otimes T^*(\mathcal{M})$ such that $\forall x \in \mathcal{M}$, $\theta(x)$ is an isomorphism $T_x\mathcal{M} \to E_x$.

Remark 5.4.23. If a fundamental 1-form exists, it can be represented in a local fiber basis $\{e_k\}_{k=1}^n$ as $\theta = e_i \otimes dx^i$, such that $\theta(\partial_j) = e_j$, mapping the basis elements between the two linear spaces.

Remark 5.4.24. In the special case $\mathcal{V} = T(\mathcal{M})$, the one-form is always well-defined, and acts as the identity map on the tangent space.

Another Bianchi identity is obtained when computing the covariant differential of the fundamental one-form. We present its special form relevant for general relativity:

Theorem 5.4.25. *Assume $\mathcal{V}$ = 2 dim $\mathcal{M}$ and that $\mathcal{V}$ has a fundamental one-form θ. Then*

$$S(\theta) \equiv \nabla\theta = 0 \Leftrightarrow (A_k)^i_j = (A_j)^i_k.$$

Remark 5.4.26. The form $S(\theta)$ is called *torsion* of the connection form A.

Proof. Let $X = X^j\partial_j$ be a section in $T(\mathcal{M})$, then by definition $\theta(X) = X^j e_j$:

$$S(\theta) = \nabla(e_j \otimes dx^j) = (A_k)^i_j e_i \otimes dx^k \wedge dx^j + e_j \otimes (A_k)^j_i dx^i \wedge dx^k$$

Therefore,

$$S(\theta) = [(A_k)^i_j - (A_j)^i_k]e_i \otimes dx^k \wedge dx^j = 0 \Leftrightarrow (A_k)^i_j = (A_j)^i_k.$$

□

5.5 Covariant Derivative Formalism

Definition 5.5.1. Let $\mathcal{V}$ be a vector bundle over $\mathcal{M}$, and A a connection form. If $X = X^j\partial_j$ is a global section on $T(\mathcal{M})$, then we define the operator $i_X : \mathcal{S}(E,k) \to \mathcal{S}(E,k-1)$ by

$$i_X(dx^k) = dx^k(X) = X^k,$$

and (multi)linearity.

Example 16. If $u = f_k dx^k$, then $i_X(u) = f_k X^k$.

Example 17. If $u = f_{ij}dx^i \wedge dx^j$, then $i_X(u) = f_{ij}X^j dx^i - f_{ij}X^i dx^j$.

Remark 5.5.2. Action of i_X is equivalent to **tensor contraction** by X.

Definition 5.5.3. Let $\mathcal{V}$ be a vector bundle over $\mathcal{M}$, and A a connection form. If $X = X^j\partial_j$ is a global section on $T(\mathcal{M})$, then we define the operator $\nabla_X : \mathcal{S}(E,k) \to \mathcal{S}(E,k)$ by

$$\nabla_X \equiv i_X\nabla.$$

Remark 5.5.4. The operator ∇_X is called the **covariant derivative** along X. It depends on both the connection form and the vector field X.

Remark 5.5.5. The covariant derivative operator is the generalization of the directional derivative of vector calculus. Indeed, in Euclidean spaces $\mathbb{R}^n$, we have for a smooth real function f,

$$\nabla = d, \quad \nabla_X f = i_X df = X^j \partial_j f = \vec{X} \cdot \vec{\nabla} f.$$

Using the covariant derivative formalism, it is possible to express all the objects contructed so far in a form which is manifestly covariant. Conversely, any new object expressed solely through covariant derivatives is guaranteed to transform covariantly.

Theorem 5.5.6. *For any section $X \in T(\mathcal{M})$ and smooth function $f \in \mathcal{A}(\mathcal{M})$, the covariant derivative satisfies*

$$\nabla_X f = X(f) = X^k \partial_k f.$$

Theorem 5.5.7. *For any two sections $X, Y \in T(\mathcal{M})$, the curvature form satisfies*

$$\Omega(X, Y) = [\nabla_X, \nabla_Y] - \nabla_{[X,Y]}$$

Theorem 5.5.8. *If $\mathcal{V} = T(\mathcal{M})$, then for any two sections $X, Y \in T(\mathcal{M})$, the torsion form satisfies*

$$S(\theta)(X, Y) = [\nabla_X, \nabla_Y] - [X, Y]$$

All the proofs can be checked by direct computation on the component representation of these objects.

Remark 5.5.9. The covariant derivative inherits all the (multi)linearity and Leibniz properties of the covariant differential operator.

Corollary 5.5.10. *A section $u \in \mathcal{S}(E)$ satisfies $\nabla u = 0$ if and only if for any $X \in T(\mathcal{M}), \nabla_X u = 0$.*

5.6 Riemannian Manifolds

Definition 5.6.1. A Riemannian manifold is a vector bundle $T^*(\mathcal{M}) \odot T^*(\mathcal{M})$ endowed with a positive-definite quadratic form section g, the metric.

Example 18. The unit sphere S^2 with the metric induced by the Euclidean norm from the embedding space $\mathbb{R}^3$ is a Riemannian manifold.

Remark 5.6.2. Any metric can be expressed in a local coordinate chart as

$$g = g_{ij} dx^i \odot dx^j,$$

where $\odot$ is the symmetric tensor product, and $\{g_{ij}\}$ is a symmetric tensor.

Given two vector fields $X, Y \in T(\mathcal{M})$, we can write

$$g(X, Y) = g_{ij} X^i Y^j, \quad g(X, X) \geq 0.$$

The metric tensor inverse leads to a dual tensor in the dual bundle:

Definition 5.6.3. Let (g^{ij}) denote the inverse matrix for (g_{ij}) : $g^{ij} g_{jk} = \delta^i_k$. Then the form $g^{ij} \partial_i \odot \partial_j$ is a global section on $T(\mathcal{M}) \otimes T(\mathcal{M})$.

The metric tensor and its inverse allow to define new operations on the tensor bundles of $\mathcal{M}$, namely the **raising and lowering** of indices. Using linearity, we define these operations for a section of $T(\mathcal{M})^{\otimes^p} \otimes (T^*(\mathcal{M}))^{\otimes^q}$, on components, as

$$T^{i_1 \ldots i_p}_{j_1 \ldots j_q} \to g^{i j_1} T^{i_1 \ldots i_p}_{j_1 \ldots j_q}, \quad T^{i_1 \ldots i_p}_{j_1 \ldots j_q} \to g_{j i_1} T^{i_1 \ldots i_p}_{j_1 \ldots j_q}$$

Example 19. Let $g = \text{diag}(1, -1)$ give the Minkowski metric on $\mathbb{R}^2$, then $\partial^i \equiv g^{ij} \partial_j$, or

$$\partial^1 = \frac{\partial}{\partial x^1}, \ \partial^2 = -\frac{\partial}{\partial x^2}.$$

The central question of Riemannian manifolds is: given a metric g, find a connection A such that g remains invariant under the action of the covariant differential $\nabla(A)$. This is very relevant when considering the evaluation

$$\nabla_Z g(X, Y) = (\nabla_Z g)(X, Y) + g(\nabla_Z X, Y) + g(X, \nabla_Z Y),$$

for which only the last two terms would remain, if $\nabla g = 0$. The following constitutes the *fundamental theorem of Riemannian geometry*:

Theorem 5.6.4. *On a Riemannian manifold, there exists a (unique) torsion-free connection* Γ*, called Levi-Civita connection, such that*

$$\nabla(\Gamma)(g) = 0.$$

The coefficients of Γ *are called Cristoffel symbols, and are given by (the lowered indices form):*

$$2\Gamma_{ijk} = \partial_j g_{ki} + \partial_k g_{ij} - \partial_i g_{jk}.$$

Proof. Both $S(\Gamma) = 0$ and $\nabla(g) = 0$ can be checked by direct computation. □

Example 20. Let $\mathcal{M} = S^2$, and $g = \text{diag}(1, \sin^2\theta)$ (the induced metric). Then the inverse metric tensor is $g^{-1} = \text{diag}((1, \sin^{-2}\theta)$. Since the local coordinates are $(x^1, x^2) = (\theta, \phi)$, the only non-zero term corresponds to $\partial_1 g_{22} = 2\sin\theta\cos\theta$, so we obtain

$$\Gamma^1_{22} = -\sin\theta\cos\theta = -\frac{\sin 2\theta}{2},\ \Gamma^2_{12} = \Gamma^2_{21} = \cot\theta.$$

Therefore, the Levi-Civita connection for this metric is

$$\Gamma = \begin{pmatrix} 0 & 0 \\ 0 & \cot\theta \end{pmatrix} d\theta + \begin{pmatrix} 0 & -\sin\theta\cos\theta \\ \cot\theta & 0 \end{pmatrix} d\phi.$$

The curvature tensor (matrix) components reduce to

$$\Omega_{12} = \partial_1\Gamma_2 + [\Gamma_1, \Gamma_2] = \begin{pmatrix} 0 & -\cos(2\theta) \\ -\sin^{-2}\theta & 0 \end{pmatrix} + \begin{pmatrix} 0 & \cos^2\theta \\ \cot^2\theta & 0 \end{pmatrix},$$

$$\Omega = \begin{pmatrix} 0 & \sin^2\theta \\ -1 & 0 \end{pmatrix} d\theta \wedge d\phi$$

Definition 5.6.5. The metric tensor of a Riemannian manifold corresponding to the Levi-Civita connection is called the Riemann tensor, $\mathcal{R} \equiv \Omega(\Gamma)$.

Definition 5.6.6. The contraction of the Riemann tensor

$$R_{ij} \equiv (\mathcal{R}_{ki})^k_j$$

defines the Ricci tensor of the Levi-Civita connection on $\mathcal{M}$.

Remark 5.6.7. Due to symmetry of the Cristoffel symbols, the Ricci tensor is symmetric.

Example 21. For the unit sphere S^2, we obtain

$$R_{11} = (\mathcal{R}_{21})^2_1 = 1,\ R_{22} = (\mathcal{R}_{12})^1_2 = \sin^2\theta$$

Definition 5.6.8. The scalar curvature of a Riemannian manifold is

$$R \equiv R^i_i = g^{ij}R_{ji}$$

Example 22. For the unit sphere S^2, we obtain

$$R^1_1 = 1,\ R^2_2 = \sin^{-2}\theta \cdot \sin^2\theta = 1,\ R = 2.$$

Definition 5.6.9. On a Riemannian manifold of dimension n with the Levi-Civita connection it is possible to define a covariant volume form

$$dV \equiv \sqrt{g}dx^1 \wedge \ldots \wedge dx^n,\ g = \det(g_{ij}).$$

Remark 5.6.10. The volume form is properly defined by bringing the metric to its canonical form (in normal coordinates $\{\xi^i\}$):

$$g_{ij}dx^i \odot dx^j = \delta_{ij}d\xi^i \odot d\xi^j,\ dV = d\xi^1 \wedge \ldots \wedge d\xi^n.$$

Example 23. For the unit sphere S^2, we obtain

$$dV = \sin\theta d\theta \wedge d\phi.$$

5.6.1 Levi-Civita Calculus

It is helpful to first derive some of the formulas often encountered in covariant calculus with the Levi-Civita connection.

Theorem 5.6.11. *Let g be a metric on a Riemannian manifold, and Γ its associated Levi-Civita connection. Denoting by $G = (g_{ij}), G^{-1} = (g^{ij}), g = \det(G)$, and $\lambda = \pm 1$ (to account for both*

hyperbolic and spherical metric cases), we have the following identities, in component and matrix forms:

$$\partial_k g_{ij} = \Gamma^l_{ik} g_{lj} + \Gamma^l_{jk} g_{li} \Leftrightarrow \partial_k G = G\Gamma_k + (\Gamma_k)^t G \tag{5.2}$$

$$\partial_k g^{ij} = -(\Gamma^i_{lk} g^{lj} + \Gamma^j_{lk} g^{li}) \Leftrightarrow \partial_k G^{-1} = -(G^{-1}\Gamma_k + (\Gamma_k)^t G^{-1}) \tag{5.3}$$

$$\frac{\partial \log g}{\partial g_{ij}} = g^{ij} \Leftrightarrow \frac{\partial \log g}{\partial G} = G^{-1} \tag{5.4}$$

$$\frac{\partial \log g}{\partial g^{ij}} = -g_{ij} \Leftrightarrow \frac{\partial \log g}{\partial G^{-1}} = -G \tag{5.5}$$

$$\frac{\partial \sqrt{\lambda g}}{\partial g_{ij}} = \frac{1}{2} g^{ij} \sqrt{\lambda g} \Leftrightarrow \frac{\partial \sqrt{\lambda g}}{\partial G} = \frac{1}{2}\sqrt{\lambda g} G^{-1} \tag{5.6}$$

$$\frac{\partial \sqrt{\lambda g}}{\partial g^{ij}} = -\frac{1}{2} g_{ij} \sqrt{\lambda g} \Leftrightarrow \frac{\partial \sqrt{\lambda g}}{\partial G^{-1}} = -\frac{1}{2}\sqrt{\lambda g} G \tag{5.7}$$

$$\sqrt{\lambda g}\nabla_k X^k = \partial_k(\sqrt{\lambda g} X^k), \quad \partial_k g = 2g\mathrm{Tr}(\Gamma_k), \tag{5.8}$$

where $X^k\partial_k$ is a vector field (section of the tangent bundle).

Example 24. On the 2-dimensional unit sphere with the induced Euclidean metric, we obtain the identities

$$\frac{\partial G}{\partial \theta} = \Gamma_1 G + G\Gamma_1 = 2\Gamma_1 G,$$

$$0 = \frac{\partial G}{\partial \phi} = \Gamma_2^t G + G\Gamma_2$$

as elementary identities for the matrices

$$\Gamma_1 = \begin{pmatrix} 0 & 0 \\ 0 & \cot\theta \end{pmatrix}, \ \Gamma_2 = \begin{pmatrix} 0 & -\sin\theta\cos\theta \\ \cot\theta & 0 \end{pmatrix},$$

and metric matrix

$$G = \begin{pmatrix} 1 & 0 \\ 0 & \sin^2\theta \end{pmatrix}.$$

5.7 Restricted Isometric Projections for Riemannian Manifolds

Given a set of observation vectors Σ embedded in the Euclidean space $\mathbb{R}^N$, where $\#\Sigma < N$, we wish to be able to compare various instances of the restricted isometric projection of Σ on $m-$dimensional linear subspaces of $\mathbb{R}^N$, $m \ll N$, and establish if the set Σ may be associated to a Riemannian manifold $(\mathcal{M}, g)$, of manifold dimension m, with Riemannian metric g equivalent to the induced metric from the embedding space $\mathbb{R}^N$. The purpose behind formulating this question is that of establishing a higher-dimensional version of the Fisher-Kolmogorov test for comparing populations in usual statistical analysis, or (alternatively) to develop an inference procedure analogous to GLM (Generalized Linear Models) in the usual case ($\#\Sigma \gg N$). If successful, the association $\Sigma \to (\mathcal{M}, g)$ would allow to establish an obvious equivalence relation between two distinct sets of vectors Σ_1, Σ_2, once they are associated to the same manifold.

In the following section we formulate the fundamental problem and present a classification criterion.

5.7.1 Generalized Restricted Isometric Projections

In the following, we take positive integers m, n, N to be related by $n < N$, $m \ll N$.

Fundamental problem. Let $\Sigma = \{v_1, v_2, \ldots, v_n\}$ be embedded in the Euclidean space $\mathbb{R}^N$, such that there exists a restricted isometric projection to a hyperplane $\mathcal{H} \simeq \mathbb{R}^m$, with distortion factor $0 < \delta \ll 1$. Is there a Riemannian manifold $(\mathcal{M}, g)$, with $\dim \mathcal{M} = m$, and a point $P \in \mathcal{M}$, such that

$$\mathcal{H} = T_P\mathcal{M},\ ||\,.\,||_{\ell^2(\mathbb{R}^N)} \hookrightarrow g(\,.\,),\ \varphi_P(\Sigma) = \widehat{\Sigma} \subset \mathcal{M}, \tag{5.9}$$

and the matrix of pairwise distances between the elements of $\widehat{\Sigma}$, in the metric g, has distortion $O(\delta)$, where φ_P is the inverse local coordinate chart $\varphi_P : T_P\,\mathcal{M} \to \mathcal{M}$?

Remark 5.7.1. An obvious extension of the problem would only require identifying the manifold $\mathcal{M}$ up to an isometry.

Remark 5.7.2. If a set $\Sigma \subset \mathbb{R}^N$ can be associated with a Riemannian manifold $(\mathcal{M}, g)$ as described

in the Fundamental Problem, then we say that Σ has the *extended restricted isometric property*, and $\widehat{\Sigma}$ is an extension of Σ.

Theorem 5.7.3. *Assume that an arbitrary set of vectors $\Sigma \subset \mathbb{R}^N$ has the* extended restricted isometric property *with projection to $\mathbb{R}^m$. Then the Riemannian manifold $(\mathcal{M}, g)$ is homemorphic to the symmetric space $SO(m+1)/(SO(1) \times SO(m))$.*

Proof. Assume there exists a differentiable function $F : D \to \mathbb{R}^N$, where $D \subset \mathbb{R}^m$ is an open, simply-connected set of full measure, such that F is a diffeomorphism between D and $F(D)$. Then the induced metric on the cotangent space at $p \in F(D)$ is given in parametric form as

$$g \in T_p^* F(D) \odot T_p^* F(D),\ g(V,W) = \sum_{1 \leq i,j \leq m} g_{ij} V^i W^j,$$

where $V, W \in T_p F(D)$ are vectors from the tangent space, and

$$g_{ij} = \Big\langle \frac{\partial F}{\partial t^i}, \frac{\partial F}{\partial t^j} \Big\rangle,$$

with $\{t^i\}_{i=1}^m$ coordinates in D, and $\langle\ ,\ \rangle$ the usual scalar product on $\mathbb{R}^N$.

If the collection of vectors $\Sigma = \{v_1, \ldots, v_n\} \in D$ are the result of projecting the original set of vectors $\Sigma_0 = \{x_1, \ldots, x_n\}$ from $\mathbb{R}^N$ to $\mathbb{R}^m$, with restricted isometry constant δ, by application of compressions A_k, $k = 1, 2, \ldots, n$, where $A_k \in \mathbb{R}^{m \times N}$, then the metric tensor evaluated at the point $p_k = F(v_k) \in F(D)$, $k = 1, 2, \ldots, n$ takes the form

$$g_{ij}(p_k) = \Big\langle \frac{\partial F}{\partial t^i}(v_k), \frac{\partial F}{\partial t^j}(v_k) \Big\rangle,$$

or in matrix form

$$G_k = (DF)^T(v_k) \cdot DF(v_k),$$

where

$$DF \in \mathbb{R}^{N \times m}$$

is the derivative matrix of F, $(DF)^T$ is its transpose, and $G_k \in \mathbb{R}^{m \times m}$ is the metric matrix at p_k. If the diffeomorphism F coincides with the inverse transformation $v_k \to x_k$ when evaluated on Σ, then we obtain the set of matrix conditions

$$A_k \cdot (DF)_k = \mathbb{I}_{m \times m},\ \forall k = 1, 2, \ldots, n.$$

This shows that, if the compression matrices $\{A_k\}_{k=1}^n$ are independent and identically distributed random variables (i.i.d). from the same ensemble of random matrices, then the matrices $(DF)_k$ are also i.i.d. with the distribution given by the generalized inverse of A_k, and therefore the matrices $\{G_k\}_{k=1}^m$ are also i.i.d. covariance matrices, obviously positive-definite, and invertible with probability 1. Therefore,

$$G_k \overset{i.i.d.}{\sim} G, \ \forall k = 1, 2, \ldots, n,$$

where G is a diagonal, positive-definite matrix, and all G_k are obviously in its conjugacy class within $GL(n, \mathbb{R})$.

Denote by $\widehat{G}$ the isotropy group of G, then since the set Σ was chosen arbitrary, we conclude that the Riemannian manifold $F(D)$ has the metric isotropy group $\widehat{G}$ acting transitively, and therefore $F(D)$ must be on open subset of a symmetric space $U/\widehat{G}$. Therefore, we can use the Cartan classification of Riemannian symmetric spaces to distinguish two possible cases: either $F(D)$ has zero curvature, and is therefore an Euclidean space, or it has positive curvature, and the manifold is then of compact type, i.e. equivalent to the quotient of two real Lie groups, U and $\widehat{G}$.

It remains to identify the possible choices of real Lie groups $U, \widehat{G}$ in the Cartan classification of compact symmetric spaces, compatible both with the requirement that $\widehat{G}$ belong to an invariance group for random covariance matrices, and the dimensional constraint $\dim U/\widehat{G} = m$.

Together with the condition that $U, \widehat{G}$ be real Lie groups, the dimension constraint implies that

$$U/\widehat{G} \simeq SO(m+1)/(SO(1) \times SO(m)), \ \dim U/\widehat{G} = m,$$

and that the ensemble of covariance matrices $\{A_k^T \cdot A_k\}$ is invariant under the induced action of $SO(m)$. □

Remark 5.7.4. Assume that all entries in compression matrices A_k are i.i.d. Gaussian. Then the eigenvalue distribution of the covariance matrices has the large m-limit ($m \gg 1$) given by the Marchenko-Pastur law (shifted semicircle law). This means that, in the limit $N \gg m \gg 1$, the curvature of the manifold $\mathcal{M}$ becomes independent on the set Σ or m, and approaches a universal limit.

Corollary 5.7.5. *Assume the set $\Sigma = \{x_1, x_2, \ldots, x_n\}$ has the restricted isometric property with compression matrices A_k, i.i.d. such that the eigenvalue distribution for covariance matrices $A_k^T \cdot A_k$ approaches the Marchenko-Pastur distribution. Then the set Σ has the extended restricted isometric property, and it can be associated with an $m-$dimensional sphere in $\mathbb{R}^N$.*

References

[1] D. L. Donoho, *Compressed sensing*, IEEE Trans. Info. Theory, vol. 52, no. 4, April. 2006.

[2] E. J. Candès, *Compressive sampling*, in Proc. International Congress of Mathematicians, Madrid, Spain, 2006, vol. 3.

[3] E. Candès, J. Romberg, and T. Tao, *Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information.*, IEEE Transactions on information theory, vol. 52, pp. 489–509, Feb 2006.

[4] E. Candès and J. Romberg, *Quantitative robust uncertainty principles and optimally sparse decompositions*, Foundations of Comput. Math, vol. 6, no. 2, pp. 227 – 254, 2006.

[5] E. Candès and T. Tao, *Near optimal signal recovery from random projections: Universal encoding strategies?,* IEEE Trans. on Information Theory, vol. 52, no. 12, pp. 5406 – 5425, 2006.

[6] Foucart, S., and Rauhut, H. *A mathematical introduction to Compressed Sensing*, Birkhauser, 2013.

[7] M. A. Davenport, M. F. Duarte, Y. C. Eldar, and G. Kutyniok. Introduction to compressed sensing. In Compressed Sensing: Theory and Applications. Cambridge University Press, 2011.

[8] T. Strohmer. *Measure what should be measured: progress and challenges in compressive sensing*. IEEE Signal Process, 2012.

[9] M. Duarte and Y. Eldar, *Structured compressed sensing: From theory to applications,* IEEE Trans. Sig. Proc., vol. 59, no. 9, pp. 4053–4085, 2011

[10] G. Tang, B. Bhaskar, P. Shah, and B. Recht. *Compressed sensing off the grid.* Preprint, 2012.

[11] E. J. Candès, J. Romberg, and T. Tao. *Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information*. IEEE Trans. Inform. Theory, 52(2):489–509, 2006.

[12] M. Lustig, D. L. Donoho, J. M. Santos, and J. M. Pauly. *Compressed Sensing MRI*. Signal Processing Magazine, IEEE, 25(2):72–82, March 2008.

[13] B. Adcock, A. C. Hansen, B. Roman, and G. Teschke. *Generalized sampling: stable reconstructions, inverse problems and compressed sensing over the continuum*. Advances in Imaging and Electron Physics, 182:187–279,2014.

[14] R. Ward. *Compressive sensing with cross validation*. IEEE Trans Inform Theory, 55(12):5773–5782, 2009.

[15] R. Tibshirani. *Regression shrinkage and selection via the Lasso*. J Roy Statist Soc B, 58(1):267–288, 1996.

[16] D. Donoho and M. Elad. *Optimally sparse representation in general (nonorthogonal) dictionaries via l_1 minimization.* Proc Natl Acad Sci, 100(5):2197–2202, 2003.

[17] E. Candès and T. Tao. *Decoding by linear programming*. IEEE Trans Inform Theory, 51(12):4203–4215, 2005.

[18] Marco F. Duarte, Shriram Sarvotham, *Joint Sparsity Models for Distributed Compressed Sensing*

[19] Yonina C. Eldar,Patrick Kuppinger, *Block-Sparse Signals: Uncertainty Relations and Efficient Recovery, IEEE TRANSACTIONS ON SIGNAL PROCESSING, VOL. 58, NO. 6, JUNE 2010*

[20] R. G. Baraniuk, V. Cevher, M. F. Duarte, and C. Hegde. *Model-based compressive sensing*

[21] T. Strohmer and R. Heath. *Grassmanian frames with applications to coding and communication.* Appl Comput Harmon Anal, 14(3):257–275, 2003.

[22] M. Herman and T. Strohmer. *High-resolution radar via compressed sensing.* IEEE Trans Sig Proc, 57(6):2275–2284, 2009.

[23] R. DeVore. *Deterministic constructions of compressed sensing matrices.* J Complex, 23(4):918–925, 2007.

[24] P. Indyk. *Explicit constructions for compressed sensing of sparse signals.* Proc.ACM-SIAM Symp Discrete Algorithms (SODA), San Franciso, CA, 2008.

[25] R. Walden. Analog-to-digital converter survey and analysis. IEEE J Sel Areas Commun, 17(4):539–550, 1999.

[26] R. Prony. *Essai expérimental et analytique sur les lois de la Dilatabilité des uides élastiques et sur celles de la Force expansive de la vapeur de l'eau et de la vapeur de l'alkool, à différentes températures.* J. de l' École Polytechnique, Floréal et Prairial III, 1(2):2476, 1795. R. Prony is Gaspard Riche, baron de Prony.

[27] C. Carathéodory. *Über den Variabilitätsbereich der Koeffizienten von Potenzreihen, die gegebene werte nicht annehmen.* Math Ann, 64:95–115, 1907.

[28] C. Carathéodory. *Über denVariabilitätsbereich der Fourierschen Konstanten von positiven harmonischen Funktionen.* Rend Circ Mat Palermo, 32:193-217, 1911.

[29] M.Vetterli, P. Marziliano, and T. Blu. *Sampling signals with finite rate of innovation*. IEEE Trans Sig Proc, 50(6):1417–1428, 2002.

[30] D. Donoho. *Compressed sensing.* IEEE Trans. Inform. Theory, 52(4):1289- 1306, 2006.

[31] Yann Traonmilin, Gilles Puy, Rémi Gribonval, Mike E. Davies, *Compressed sensing in Hilbert spaces*, 2017.

[32] Bernhard G. Bodmann, Axel Flinth, Gitta Kutyniok. *Compressed Sensing for Analog Signals*, 2018

[33] B. Adcock, A. C. Hansen, C. Poon, and B. Roman, *Breaking the coherence barrier: a new theory for compressed sensing*, arXiv:1302.0561, 2013.

[34] M. Vetterli and T. Blu, *Sampling signals with finite rate of innovation*, IEEE Trans. Signal Process., vol. 50, no. 6, pp. 1417–1428, 2002.

[35] N. Thaper, S. Guha, P. Indyk, and N. Koudas, *Dynamic multidimensional histograms*, in Proc. ACM SIGMOD Int. Conf. Manag. Data, 2002, pp. 428–439.

[36] A. Bourrier, R. Gribonval, and P. Perez, *Compressive gaussian mixture estimation*, in Proc. IEEE Conf. Acoustics, Speech and Signal Processing, 2013, pp. 6024–6028.

[37] J. C. Robinson, Dimensions, embeddings, and attractors. Cambridge Tracts in Mathematics. Cambridge University Press, 2011.

[38] G. Puy, M. E. Davies, and R. Gribonval. *Recipes for stable linear embeddings from hilbert spaces to* $\mathbb{R}^m$. Transactions on Information Theory, Vol XX, No. X, 2017.

[39] V. Chandrasekaran, B. Recht, P. Parrilo, and A. Willsky. The convex geometry of linear inverse problems. Foundations of Computational Mathematics, 12(6):805-849, 2012.

[40] Y. C. Eldar, P. Kuppinger, and H. Bolcskei. *Block-sparse signals: Uncertainty relations and efficient recovery*. Signal Processing, IEEE Transactions on, 58(6):3042-3054, 2010.

[41] B. Adcock, A. Hansen, B. Roman, and G. Teschke. Generalized sampling: stable reconstructions, inverse problems and compressed sensing over the continuum. Adv. in Imag. and Electr. Phys., 182(1):187-279, 2014.

[42] C. Studer and R. G. Baraniuk. Stable restoration and separation of approximately sparse signals. Applied and Computational Harmonic Analysis, Sept. 2013.

[43] Y. Traonmilin, S. Ladjal, and A. Almansa. Robust multi-image processing with optimal sparse regularization. Journal of Mathematical Imaging and Vision, 51(3):413-429, 2015.

[44] T. Cai and A. Zhang. Sparse representation of a polytope and recovery of sparse signals and low-rank matrices. Information Theory, IEEE Transactions on, 60(1):122-132, Jan 2014.

[45] Y. Traonmilin and R. Gribonval. Stable recovery of low-dimensional cones in hilbert spaces: One rip to rule them all. Applied And Computational Harmonic Analysis, In Press, 2016.

[46] A. Bourrier, M. Davies, T. Peleg, P. Perez, and R. Gribonval. Fundamental performance limits for ideal decoders in high-dimensional linear inverse problems. Information Theory, IEEE Transactions on, 60(12):7928-7946, Dec 2014.

[47] Temlyakov, V.N., Weak greedy algorithms. Adv. Comp. Math., 12(2,3):213-227, 2000.

[48] Duval,V., and Peyre, G., Exact support recovery for sparse spikes deconvolution. Found. Comput. Math., 15(5):1314-1355, 2015.

Appendix A
Reproducing Kernel Hilbert Space

Definition A.0.1. (Reproducing kernel) Let $\mathcal{H}$ be a Hilbert space of $\mathbb{R}$ valued functions defined on a non-empty set X. A function $k : X \times X \to \mathbb{R}$ (k_x) is called a *reproducing kernel* of $\mathcal{H}$ if it satisfies

- $\forall x \in X, k(\cdot, x) \in \mathcal{H}$
- $\forall x \in X, \forall f \in \mathcal{H}, \langle f, k(\cdot, x)\rangle_{\mathcal{H}} = f(x)$ (the *reproducing property*)

In particular, for any $x, y \in X$, $k(x, y) = \langle k(\cdot, x), k(\cdot, y)\rangle_{\mathcal{H}}$

Proposition A.0.2. *(Uniqueness of the reproducing kernel) If it exists, reproducing kernel is unique.*

Definition A.0.3. A Hilber space $\mathcal{H}$ is called Reproducing Kernel Hilbert Space (RKHS) if $\mathcal{H}$ has a reproducing kernel.

Remark A.0.4. A Reproducing Kernel Hilbert Space $\mathcal{H}$ has a unit-norm reproducing kernel if $k(x, x) = 1$ for all $x \in X$.

Definition A.0.5. A function $k : X \times X \to \mathbb{R}$ is called a *kernel* on X if there exists a Hilbert space (not necessarilly a RKHS) $\mathcal{H}$ and a map $\phi : X \to \mathcal{H}$, such that $k(x, y) = \langle \phi(x), \phi(y)\rangle_{\mathcal{H}}$

Corollary A.0.6. *Every reproducing kernel is a kernel. (can take $\phi : x \to k(\cdot, x), k(x, y) = \langle(\cdot, x), k(\cdot, y)\rangle_{\mathcal{H}}$ i.e. RKHS $\mathcal{H}$ is a feature space)*

Definition A.0.7. (Positive definite functions). A symmetric function $h : X \times X \to \mathbb{R}$ is positive definite if $\forall n \geq 1, \forall(a_1, \cdots, a_n) \in \mathbb{R}^n, \forall(x_1, \cdots, x_n) \in X^n$

$$\sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j h(x_i, x_j) \geq 0$$

The function $h(\cdot,\cdot)$ is strictly positive definite if for mutually distinct x_i the equality holds only when all the a_i are zero.

Lemma A.0.8. *Let $\mathcal{H}$ be any Hilbert space (not necessarily an RKHS), X a nonempty set and $\phi : X \to \mathcal{F}$. Then kernel $h(x,y) := \langle \phi(x), \phi(y) \rangle_{\mathcal{H}}$ is a positive definite functions.*

Corollary A.0.9. *Reproducing kernels are positive definite*

$$\text{Reproducing Kernel } \Rightarrow Kernel \Rightarrow \text{ Positive Definite}$$

Lemma A.0.10. *If h is positive definite, then $|h(x_1,x_2)|^2 \leq h(x_1,x_1)h(x_2,x_2)$*

Theorem A.0.11. *(Moore - Aronszajn) Let $k : X \times X \to \mathbb{R}$ be positive definite. There is a unique RKHS $\mathcal{H} \subset \mathbb{R}^X$ with reproducing kernel k*